

Projeto Da.Re.

Intellectual Output 2: Modulo de Data Science e Recursos de Formação Aberta Síntese e tradução na língua Portuguesa

A nossa pesquisa identificou uma lacuna na oferta educativa na área de Data Science que não é satisfeita pelas universidades ou campos de formação, nomeadamente, o que chamamos de pessoa-ponte (**bridge person**), alguém que combina conhecimento dos negócios de uma organização com conhecimento e compreensão de Data Science, de forma a criar uma 'ponte' entre pessoas não técnicas no negócio com os cientistas dos dados altamente qualificados capazes de agregar valor ao negócio.

Nós acreditamos que os papéis da **Bridge Person** e do **Chefe Data Analytics Officer** identificados pelo Da.Re. O projecto pode ter um impacto significativo na melhoria da aplicação da ciência dos dados às empresas e organizações europeias.

A nossa solução é permitir uma educação adicional adequada no domínio da Data Science para:

- Funcionários de empresas que possuem conhecimento especializado sobre negócios e pouco sabem sobre ciência de dados (Através desta carreira profissional eles podem se tornar Chief Data Analytics Officer (ou similar) na empresa (depois de um tempo));
- Alunos de pós-graduação que querem trabalhar como cientista de dados nos negócios e estão motivados para aprender novos tópicos técnicos sobre o valor de sua posição futura. (uma possível carreira profissional que eles podem seguir Chief Data Analytics Officer (depois de um tempo));
- Pessoas muito experientes nos negócios que têm pouco tempo, mas querem saber como a ciência de dados pode agregar valor a seus negócios e como dar os primeiros passos em direção a ela.

Um esboço do Da.Re. Programa. Incluído no I.O.2 e os seus anexos

O programa do Da.Re. tem duas partes: 80 horas de educação on-line, seguidas de 70 horas de ensino presencial.

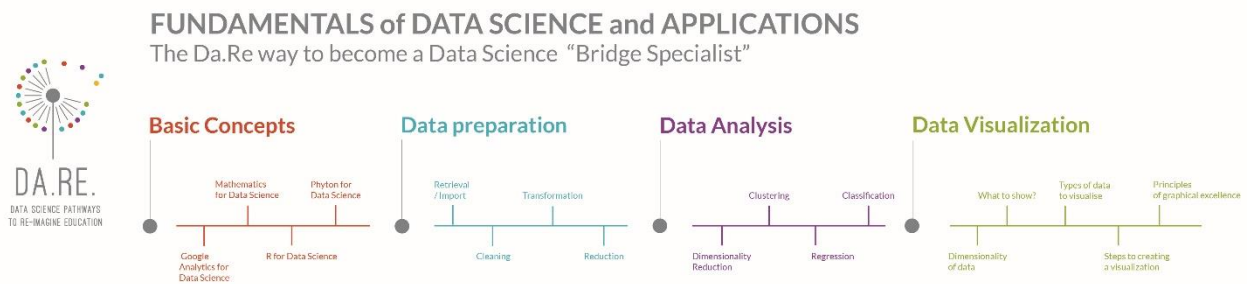
A idéia é fornecer aos alunos através da formação on-line o conhecimento técnico útil e as habilidades necessárias para lidar com a parte presencial do curso.

Combinando as duas fórmulas de treino Da.Re. este consegue fornecer os melhores elementos contidos nos MOOCs e BOOTCAMPs, a fim de fornecer novas ferramentas educativas, úteis e sustentáveis para serem disseminadas na Europa.

Os estudos de caso fornecidos para a parte na presencial do curso são:

1. A pegada de energética de uma micro-rede elétrica;
2. A análise estacionária para sistemas industriais;
3. Gestão de Comunicação Empresarial;
4. Análise de séries temporais para análise biomédica.

A parte on-line do curso é representada pela seguinte imagem:



O curso está estruturado em 5 módulos diferentes, 4 teóricos e um prático. Este último consiste nos estudos de caso incluídos no I.O.2 como anexos.

A parte teórica contida no texto da I.O.2 é a seguinte:

BASIC CONCEPTS (Conceitos Básicos)

Neste módulo, os alunos aprenderão alguns dos conceitos fundamentais da ciência de dados. Isso começa com a aprendizagem de como o Google Analytics pode fornecer uma ampla variedade de informações sobre o uso de websites, incluindo muitos dados sobre s utilizadores e como eles navegam em sites. Em seguida, você será apresentado ao Python e ao uso de bibliotecas de ciência de dados padrão, incluindo NumPy, Pandas, Scikit-learn e Matplotlib. Então você aprenderá sobre o pacote R que fornece uma ampla variedade de métodos para análise estatística de dados. O módulo terminará com uma visão geral de vários tópicos matemáticos usados em ciência de dados.

DATA PREPARATION (Preparação dos Dados)

A preparação de dados assume um papel crítico na vida dos cientistas de dados. Durante estes módulos, vários aspectos fundamentais da preparação dos dados para posterior análise devem ser dados particularmente, os que são cruciais para as aplicações dadas com estudos de casos reais. Desta forma, tópicos como carregamento de dados, amostragem, extração de recursos e transformada de Fourier devem ser dados. No final deste módulo, espera-se que você seja capaz de preparar com sucesso seus dados para uma análise eficiente.

DATA ANALYSIS (Análise de Dados)

O módulo de análise de dados é a porta de entrada para o mundo da aprendizagem de máquina (Machine learning, ML). Neste módulo, as diversas faces do ML, como redução de dimensionalidade, classificação, previsão e clusterização, serão reveladas. Eles são apresentados de forma teórica simples com um olho em exemplos práticos; implementações simples usando Python são também fornecidas. No final do módulo,

you will be asked to apply appropriately the various paradigms and interpret correctly the results.

DATA VISUALIZATION (Visualização dos Dados)

Now that we have the data, what do we do with it? In this module, we will focus on ways to use visualization to answer this question. In particular, we will learn how to display data to help us get information and communicate that information to other people. We will present the theory that underpins the use of effective visualizations. We will do practical exercises using R, both on sample data and on data provided in real life studies of case studies chosen to visualize it in different ways.

Pré-requisitos de Conhecimento e Background necessários para o curso

- Level 6 of education or higher, for example a diploma in any area
- Being able to count, read simple equations, graphs and tables
- Literate and able to write reports with illustrative graphs
- Good research skills, finding and synthesizing information
- Interest in data patterns and in measures that affect businesses
- Good self-study and time management skills
- Good team working skills.

Os resultados de aprendizagem

The general learning outcomes of I.O.2 are the following:

- A clear understanding of the data life cycle
 1. Preparation
 2. Analysis
 3. Visualization
- Practical ability to solve real problems that companies face
- The ability to go beyond the data life cycle, creating added value for the organization
- Ability to organize and review a data life cycle in an organization. More precisely, to identify and select existing and non-existing competencies in the organization, create a team and structure the work to pass through an established pipeline:
 1. Identification of the problem
 2. Preparation of the necessary data
 3. Analysis of the data
 4. Visualization of the data
 5. Solution of the problem.

The content of the five modules of the course is available for download on the Da.Re. website.