

✳ Escola Superior de Tecnologia de Tomar

Ano letivo: 2021/2022

Mestrado em Análítica e Inteligência Organizacional

Mestrado, 2º Ciclo

Plano: Despacho n.º 11262/2016 - 19/09/2016

Ficha da Unidade Curricular: Arquitetura de Sistemas Analíticos

ECTS: 8; Horas - Totais: 216.0, Contacto e Tipologia, T:20.0; TP:20.0; PL:12.0;

OT:32.0; S:4.0; O:10.0;

Ano | Semestre: 1 | A

Tipo: Obrigatória; Interação: Presencial; Código: 39323

Área Científica: Tecnologias de Informação e Comunicação

Docente Responsável

Ricardo Nuno Taborda Campos

Professor Adjunto

Docente(s)

Ricardo Nuno Taborda Campos

Professor Adjunto

Objetivos de Aprendizagem

Esta UC tem por objetivo introduzir os alunos à extração e ao processamento de informação a partir de fontes de dados não estruturadas. No final da unidade curricular o aluno deverá ser capaz de entender e aplicar métodos de análise de dados textuais e da web

Objetivos de Aprendizagem (detalhado)

Esta unidade curricular tem por objetivo introduzir os alunos à extração e ao processamento de informação a partir de fontes de dados não estruturadas. No final da unidade curricular o aluno deverá ser capaz de entender e aplicar métodos de análise de dados textuais e da web, bem como saber usar ferramentas de text mining e de processamento da linguagem natural para a extração e a análise de informação com recurso à linguagem de programação Python.

Ao concluir esta unidade o estudante deverá:

1. entender o que é a ciência dos dados, ter a noção das tendências atuais e conhecer as várias

fases de um projeto de data science.

2. entender a importância da linguagem de programação Python no contexto da ciência dos dados;
3. conhecer a arquitetura de um sistema de extração de informação e dominar o processo de coleta de dados a partir de fontes de dados não estruturadas;
4. compreender e saber usar ferramentas de text mining e de processamento da linguagem natural para o processamento de informação e a extração de conhecimento;
5. saber importar, manipular, transformar, relacionar, analisar e armazenar dados numéricos com recurso às principais frameworks de data science;
6. saber importar, manipular, transformar, relacionar, analisar e armazenar dados textuais obtidos a partir de diferentes fontes de dados com recurso às principais frameworks de data science;
7. saber conduzir uma análise exploratória de dados com recurso à utilização de funções de visualização e ao uso das principais frameworks de data science.

Conteúdos Programáticos

1. Introdução à Ciência dos Dados
2. Python na Ciência dos Dados
3. Extração de Informação e Aquisição de Dados
4. Text Mining e Processamento da Linguagem Natural
5. Manipulação e Análise de dados numéricos
6. Manipulação e Análise de de dados textuais
7. Análise Exploratória de Dados e Visualização

Conteúdos Programáticos (detalhado)

1. Introdução à Ciência dos Dados
 - 1.1. Visão geral
 - 1.2. Tendências atuais
 - 1.3. Análise de dados e analítica
 - 1.4. Competências de um cientista de dados
 - 1.5. Arquitetura de um projeto de data science
2. Python na Ciência dos Dados
 - 2.1. Porquê usar Python?
 - 2.2. História do Python
 - 2.3. Características
 - 2.4. Vantagens
 - 2.5. Jupyter Anaconda
3. Extração de Informação e Aquisição de Dados
 - 3.1 Fontes e tipos de dados (estruturados; semi-estruturados; não estruturados)
 - 3.2. Aquisição de dados a partir de Ficheiros (texto, imagens, pdfs, word, html, csv, json)
 - 3.3. Aquisição de dados textuais a partir da web
 - 3.3.1 Pacotes de software (newspaper3k, Wikipedia, Facebook, Twitter, Google)
 - 3.3.2. APIs
 - 3.3.3. Web Scraping

3.3.4. Web Crawling

3.4. Arquivos da Web

4. Text Mining & Processamento da Linguagem Natural (NLP)

4.1. Text Mining

4.1.1. Definição

4.1.2. Arquitetura

4.1.3. Aplicações e Áreas de Investigação

4.2. Processamento da Linguagem Natural

4.2.1. Introdução e desafios

4.2.2. Pipeline

4.2.3. Pre-processamento de dados textuais (Parsing; Divisão de frases; Tokenização; Normalização; Part-of-Speech (POS), Reconhecimento de Entidades (NER); Stemming)

4.2.4. Soluções comerciais e frameworks de NLP em Python

5. Manipulação e Análise de Dados Numéricos com Numpy

5.1. Importação e Exportação de Dados

5.2. Inspeção de Dados (tipos de dados e estatísticas de sumário)

5.3. Criação de Arrays uni-dimensionais e multi-dimensionais

5.4. Manipulação de Arrays

5.5. Operações Matemáticas

6. Manipulação e Análise de Dados Textuais com Pandas Dataframes

6.1. Tipos de Dados (Numéricos: Contínuos e Discretos; Categóricos: Nominais e Ordinais)

6.2. Importação e Exportação de Dados (incluindo de elevada dimensão – big data)

6.3. Limpeza e Transformação de Dados (Data Wrangling/Munging)

6.3. Inspeção de Dados (tipos de dados e estatísticas de sumário)

6.4. Séries de dados em Pandas

6.5. Criação de Dataframes

6.6. Manipulação de Dataframes (Seleção, Projeção, Pesquisa e Modificação)

6.7. Operações de Concatenação

6.8. Operações de Ordenação

6.9. Operações de Agregação

6.10. Operações de Junção

6.11. Filtros

6.12. HTML Display

7. Análise Exploratória de Dados e Visualização com Matplotlib

7.1. Introdução ao matplotlib

7.2. Personalização e formatação de Gráficos

7.3. Galeria de Plots (Plots de linhas, barras, circulares, dispersão, histogramas, séries temporais, semi-log e log-log)

7.4. Subplots

7.5. Partilha de gráficos

Metodologias de avaliação

Avaliação periódica: Projeto (40%) + Frequência (60%)

Participação obrigatória nos vários elementos de avaliação com nota mínima de 6 valores (em cada um)

Avaliação Final: Exame (100%)

Requisitos de admissibilidade à frequência e ao exame:

A entrega do projeto fora do prazo previsto ou com uma nota inferior a 6 valores implica a reprovação automática do aluno impossibilitando-o de se propor a exame.

Os alunos ficam também automaticamente reprovados e excluídos de exame no caso de não atingirem um mínimo de 70% de assiduidade às aulas (exceto trabalhadores estudantes);

Software utilizado em aula

Jupyter Anaconda / Python

Moodle

Estágio

Não Aplicável

Bibliografia recomendada

- Nelli, F. (2018). *Python Data Analytics: With Pandas, NumPy, and Matplotlib* . 2, Roma, Itália. US
- Corrêa, E. (2020). *Data Wrangling para Ciência de Dados* (pp. 1-223). Alura. Brasil
- Gomes, D. e Demidova, E. e Winters, J. e Risse, T. (2021). *The Past Web. Exploring Web Archives* . Springer. Holanda
- Provost , F. e Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* . O'Reilly Media. USA

Coerência dos conteúdos programáticos com os objetivos

Os objetivos de aprendizagem da unidade curricular são atingidos através da realização de um conjunto de exercícios e trabalhos práticos permitindo desta forma que os alunos solidifiquem as competências adquiridas. Os conteúdos programáticos estão em coerência com os objetivos da unidade curricular, atendendo a que:

- 1 vs 1
- 2 vs 2
- 3 vs 3
- 4 vs 4
- 5 vs 5
- 6 vs 6
- 7 vs 7

Metodologias de ensino

Exposição dos conteúdos programáticos aos alunos com recurso ao método expositivo e demonstrativo. Análise e resolução de casos práticos. Os conhecimentos adquiridos serão avaliados através de um projeto de grupo e uma prova de avaliação

Coerência das metodologias de ensino com os objetivos

Os objetivos de aprendizagem são atingidos através do acompanhamento dos estudantes no decurso da realização dos exercícios práticos e do projeto permitindo desta forma que os alunos solidifiquem as competências adquiridas.

Língua de ensino

Português

Pré-requisitos

Não Aplicável

Programas Opcionais recomendados

Não Aplicável

Observações

Os conteúdos programáticos da UC serão lecionados, potenciando uma educação de qualidade (ODS4) e a construção de cidades e comunidades sustentáveis (ODS11). Serão ainda desenvolvidos esforços no sentido de promover parcerias para a implementação dos objetivos da UC (ODS17).

Objetivos de Desenvolvimento Sustentável:

- 4 - Garantir o acesso à educação inclusiva, de qualidade e equitativa, e promover oportunidades de aprendizagem ao longo da vida para todos;
- 11 - Tornar as cidades e comunidades inclusivas, seguras, resilientes e sustentáveis;
- 17 - Reforçar os meios de implementação e revitalizar a parceria global para o desenvolvimento sustentável;

Docente responsável
