

---

**Ano Letivo** 2020-21

---

**Unidade Curricular** ANÁLISE DE DADOS E APRENDIZAGEM DE MÁQUINAS

---

**Cursos** ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES (2.º Ciclo) (\*)  
ÁREA DE ESPECIALIZAÇÃO EM TECNOLOGIAS DE INFORMAÇÃO E TELECOMUNICAÇÕES  
ÁREA DE ESPECIALIZAÇÃO EM SISTEMAS DE ENERGIA E CONTROLO  
ÁREA DE ESPECIALIZAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E TELECOMUNICAÇÕES

(\*) Curso onde a unidade curricular é opcional

---

**Unidade Orgânica** Instituto Superior de Engenharia

---

**Código da Unidade Curricular** 14771115

---

**Área Científica**

---

**Sigla**

---

**Línguas de Aprendizagem** Inglês|Português

---

**Modalidade de ensino** Presencial e/ou à distância (em situações excecionais)

---

**Docente Responsável** Pedro Jorge Sequeira Cardoso

DOCENTE	TIPO DE AULA	TURMAS	TOTAL HORAS DE CONTACTO (*)
Pedro Jorge Sequeira Cardoso	PL; T; TP	T1; TP1; PL1	14T; 14TP; 14PL

\* Para turmas lecionadas conjuntamente, apenas é contabilizada a carga horária de uma delas.

ANO	PERÍODO DE FUNCIONAMENTO*	HORAS DE CONTACTO	HORAS TOTAIS DE TRABALHO	ECTS
1º	S2	14T; 28TP	195	7.5

\* A-Anual; S-Semestral; Q-Quadrimestral; T-Trimestral

### Precedências

Sem precedências

### Conhecimentos Prévios recomendados

Conhecimentos básicos em matemática, programação (recomendado), bases de dados (não essencial) e conhecimento de algoritmia (recomendado).

### Objetivos de aprendizagem (conhecimentos, aptidões e competências)

O aluno aprovado na UC será capaz de:

- Projetar e programar aplicações
- Conceber, programar e usar o sistema de gestão de bases de dados relacionais e não-relacionais
- Use os modelos e as funcionalidades de frameworks web para implementar e configurar aplicações RESTful.
- Compreender os conceitos-chave na ciência dos dados, incluindo as aplicações e ferramentas usadas nas ciências dos dados. Compreender conceitos estatísticos e de aprendizagem de máquina. Produzir análises estatísticas
- Planear o uso e aplicar algoritmos de aprendizagem de máquinas a conjuntos de dados, interpretando adequadamente os resultados alcançados, incluindo distinguir entre aprendizagem supervisionada versus não-supervisionada, seleção e avaliação de modelos, e uso de métodos de regressão linear, regressão logística, support vector machines, algoritmos de deteção de anomalias, sistemas de recomendação, algoritmos de agrupamento (clustering) e redes neuronais básicas

## Conteúdos programáticos

### 1. Programação Python

### 2. Bases de dados (BD)

#### 2.1. BD relacionais

##### 2.1.1. Modelo relacional, modelação de entidades e normalização

##### 2.1.2. Programação SQL (operações CRUD)

##### 2.1.3. Acessos a SGBDs

#### 2.2. BD não-relacionais

##### 2.2.1. Introdução ao NoSQL

##### 2.2.2. O exemplo do MongoDB: documentos e coleções, operações CRUD e indexação. Conceito de MapReduce, replicação e fragmentação

##### 2.2.3. Acessos ao MongoDB

### 3. Web frameworks

#### 3.1. Uso de modelos e roteamento

#### 3.2. Configuração e autenticação RESTful

#### 3.3. Implementação de paginação, cache HTTP e segurança

### 4. Análise de dados

#### 4.1. Visualização de dados

#### 4.2. Estatísticas descritivas

#### 4.3. Amostragem de dados

#### 4.4. Ferramentas de análise de dados

### 5. Algoritmos de aprendizagem da máquina

#### 5.1. Aprendizagem supervisionada versus não supervisionada

#### 5.2. Seleção e avaliação de modelos

#### 5.3. Regressão linear

#### 5.4. Regressão logística

#### 5.5. Support vector machines

#### 5.6. Detecção de anomalias

#### 5.7. Sistemas de recomendação

#### 5.8. Clustering

#### 5.9. Redes neurais

---

## Metodologias de ensino (avaliação incluída)

### Metodologia de ensino

- Aulas teóricas - apresentação de conceitos teóricos.
- Aulas práticas / laboratoriais - implementação prática e uso dos conteúdos do curso.

### Avaliação

A avaliação será baseada na implementação de 3 projetos:

- P1 - projeto de programação
- P2 - Implementação de um serviço web para armazenar e recuperar dados de bases de dados relacionais e não-relacionais. Comparação de soluções.
- P3 - Projeto final: Uso da análise de dados e dos métodos de aprendizagem automática para inferir conhecimento de dados fornecidos/recolhidos.

A nota final será calculada pela seguinte fórmula

Classificação final =  $0.2 \cdot P1 + 0.3 \cdot P2 + 0.5 \cdot P3$

---

## Bibliografia principal

- Andreas C. Mueller and Sarah Guido (2016). Introduction to Machine Learning with Python. O'Reilly Media, Inc.
- Fabrizio Romano (2015). Learning Python. Packt Publishing.
- Luis Pedro Coelho & Willi Richert (2015). Building Machine Learning Systems with Python. Packt Publishing.
- Mark Summerfield (2008), Programming in Python 3: A Complete Introduction to the Python Language. Addison-Wesley Professional.
- Gareth Dwyer (2016). Flask By Example. Packt Publishing.
- Amol Nayak(2014). MongoDB Cookbook. Packt Publishing.
- Albert Lukaszewski(2010). MySQL for Python. Packt Publishing.

**Academic Year** 2020-21

**Course unit** DATA ANALYSIS AND MACHINE LEARNING

**Courses** ELECTRICAL AND ELECTRONICS ENGINEERING (\*)  
BRANCH SPECIALISATION IN INFORMATION TECHNOLOGIES AND TELECOMMUNICATIONS

(\*) Optional course unit for this course

**Faculty / School** INSTITUTE OF ENGINEERING

**Main Scientific Area**

**Acronym**

**Language of instruction** English/Portuguese

**Teaching/Learning modality** Classroom-based and/or distance learning (on exceptions)

**Coordinating teacher** Pedro Jorge Sequeira Cardoso

Teaching staff	Type	Classes	Hours (*)
Pedro Jorge Sequeira Cardoso	PL; T; TP	T1; TP1; PL1	14T; 14TP; 14PL

\* For classes taught jointly, it is only accounted the workload of one.

---

#### Contact hours

T	TP	PL	TC	S	E	OT	O	Total
14	28	0	0	0	0	0	0	195

T - Theoretical; TP - Theoretical and practical ; PL - Practical and laboratorial; TC - Field Work; S - Seminar; E - Training; OT - Tutorial; O - Other

---

#### Pre-requisites

no pre-requisites

---

#### Prior knowledge and skills

Basic knowledge in mathematics, programming (recommended), databases (not essencial) and knowledge of algorithm (recommended).

---

#### The students intended learning outcomes (knowledge, skills and competences)

The student approved in the UC should be able to:

- Design and program simple applications in python
- Conceive, program and use the relational and non-relational database management system
- Use web frameworks to deploy and configure RESTful applications.
- Understand key concepts in data science, including the applications and tools used in data science. Understand statistical concepts and machine learning. Produce statistical analysis.
- Plan, use and apply machine learning algorithms to data sets, interpreting adequately the results achieved, including: distinguishing between supervised versus unsupervised learning, use selection and evaluation of models, and use of linear regression, logistic regression, support vector machines, anomaly detection algorithms, recommendation systems, clustering algorithms and basic neural networks.

## Syllabus

### 1. Python Programming

#### 2. Databases (BD)

##### 2.1. Relational BDs

###### 2.1.1. Relational model, entity modeling and normalization

###### 2.1.2. SQL programming (CRUD operations)

###### 2.1.3. Access to DBMSs

##### 2.2. Non-relational DBs

###### 2.2.1. Introduction to NoSQL

###### 2.2.2. The MongoDB example: documents and collections, CRUD operations, and indexing. MapReduce concept, replication and fragmentation

###### 2.2.3. MongoDB accesses

### 3. Web frameworks

#### 3.1. Use of models and routing

#### 3.2. RESTful configuration and authentication

#### 3.3. HTTP paging, caching, and security implementation

### 4. Data analysis

#### 4.1. Data visualization

#### 4.2. Descriptive Statistics

#### 4.3. Data sampling

#### 4.4. Data Analysis Tools

### 5. Machine Learning Algorithms

#### 5.1. Supervised versus unsupervised learning

#### 5.2. Selection and evaluation of models

#### 5.3. Linear Regression

#### 5.4. Logistic regression

#### 5.5. Support vector machines

#### 5.6. Detection of anomalies

#### 5.7. Recommendation systems

#### 5.8. Clustering

#### 5.9. Neural networks

---

## Teaching methodologies (including evaluation)

### Teaching Methodology

Theoretical classes - presentation of theoretical concepts.

Practical / laboratory classes - practical implementation and use of course contents.

### Evaluation

The evaluation is supported in 3 projects:

P1 - programming project in the Python language.

P2 - Implementation of a web service to store and retrieve data from relational and non-relational databases. Comparison of solutions.

P3 - Final Project: Use of data analysis and automatic learning methods to infer knowledge of data provided / collected.

The final grade will be calculated by the following formula

Final classification =  $0.2 * P1 + 0.3 * P2 + 0.5 * P3$

---

## Main Bibliography

- Andreas C. Mueller and Sarah Guido (2016). Introduction to Machine Learning with Python. O'Reilly Media, Inc.
- Fabrizio Romano (2015). Learning Python. Packt Publishing.
- Luis Pedro Coelho & Willi Richert (2015). Building Machine Learning Systems with Python. Packt Publishing.
- Mark Summerfield (2008). Programming in Python 3: A Complete Introduction to the Python Language. Addison-Wesley Professional.
- Gareth Dwyer (2016). Flask By Example. Packt Publishing.
- Amol Nayak(2014). MongoDB Cookbook. Packt Publishing.
- Albert Lukaszewski(2010). MySQL for Python. Packt Publishing.