English version at the end of this document

Ano Letivo	2021-22
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 (*) Curso onde a unidade curricular é opcional
Unidade Orgânica	Instituto Superior de Engenharia
Código da Unidade Curricular	14771138
Área Científica	ENGENHARIA ELECTROTÉCNICA
Sigla	
Código CNAEF (3 dígitos)	481
Contributo para os Objetivos de Desenvolvimento Sustentável - ODS (Indicar até 3 objetivos)	9, 11
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	T; TP	T1; TP1	14T; 28TP

* 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 informáticas análise de dados e AM;

- 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, inlcuindo 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.

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 apredizagem 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. Deteção de anomalias
5.7. Sistemas de recomendação
5.8. Clustering
5.9. Redes neuronais

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 feita por trabalhos práticos.

- Ps - Projetos parciais de acompanhamento da matéria;
- Pf - Projeto final: uso dos conhecimentos adquiridos para inferir / gerar conhecimento de dados fornecidos / recolhidos.

A nota final será calculada pela seguinte fórmula

Classificação final = 0.4 x Ps + 0.6 x Pf.

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.
- João M.F. Rodrigues , Pedro J.S. Cardoso, Jânio Monteiro, & Célia M.Q. Ramos (2020). Smart Systems Design, Applications, and Challenges. IGI Global.

Academic Year	2021-22
Course unit	DATA HANDLING AND MACHINE LEARNING
Courses	ELECTRICAL AND COMPUTER ENGINEERING (*) SPECIALISATION IN INFORMATION TECHNOLOGIES AND TELECOMMUNICATIONS SPECIALISATION IN ENERGY AND CONTROL SYSTEMS (*) Optional course unit for this course
Faculty / School	INSTITUTE OF ENGINEERING
Main Scientific Area	
Acronym	
CNAEF code (3 digits)	481
Contribution to Sustainable Development Goals - SGD (Designate up to 3 objectives)	9, 11
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	T; TP	T1; TP1	14T; 28TP

* 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.

Assessment
The evaluation will be done by assignements / projects:

- Ps - Partial projects for monitoring the follow up of teached subjects;
- Pf - Final project: use of acquired knowledge to infer / generate knowledge from data provided / collected.

The final grade will be calculated by the following formula

Final rating = 0.4 x Ps + 0.6 x Pf.

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.
- João M.F. Rodrigues , Pedro J.S. Cardoso, Jânio Monteiro, & Célia M.Q. Ramos (2020). Smart Systems Design, Applications, and Challenges. IGI Global.