English version at the end of this document

Ano Letivo	2022-23
Unidade Curricular	ANÁLISE DE DADOS
Cursos	ENGENHARIA INFORMÁTICA (2.º ciclo) (*) (*) Curso onde a unidade curricular é opcional
Unidade Orgânica	Faculdade de Ciências e Tecnologia
Código da Unidade Curricular	14741067
Área Científica	CIÊNCIA DE COMPUTADORES
Sigla	
Código CNAEF (3 dígitos)	481
Contributo para os Objetivos de Desenvolvimento Sustentável - ODS (Indicar até 3 objetivos)	4 8 10
Línguas de Aprendizagem	Português (PT) ou Inglês (ENG)

Modalidade de ensino

Presencial

Docente Responsável António Eduardo de Barros Ruano

DOCENTE	TIPO DE AULA	TURMAS	TOTAL HORAS DE CONTACTO (*)
António Eduardo de Barros Ruano	PL; T	T1; PL1	28T; 28PL

* 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	S1	28T; 28PL	156	6

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

Precedências

Sem precedências

Conhecimentos Prévios recomendados

Conhecimentos de Algoritmos e Estruturas de Dados e de Probabibilidades e Estatítica

Objetivos de aprendizagem (conhecimentos, aptidões e competências)

O aluno, no final da disciplina, deverá ser capaz de:
- Conhecer as etapas principais da ciência dos dados
- Conhecer as técnicas estatísticas mais usadas em análise de dados
- Conhecer as técnicas mais importantes de visualização, pré-processamento e de mineração de dados
- Conhecer os fundamentos da lógica difusa, os modelos de Mamdani e de Sugeno, e os sistemas neuro-difusos; Saber utilizá-los no contexto de aproximação de sistemas e de classificação.
- Conhecer os métodos mais utilizados para seleção e validação de modelos

Conteúdos programáticos

1. O processo da ciência de dados
2. Análise exploratória de dados
    1. Breve revisão da teoria da probabilidade e distribuições comuns
    2. Estatísticas descritivas
    3. Estatística inferencial
3. Técnicas de Visualização de Dados
4. Pré-processamento de dados
    1. Limpeza
    2. Integração
    3. Redução de dimensionalidade
    4. Transformação e Discretização
5. Mineração de Padrões, Associações e Correlações Frequentes
    1. Conceitos básicos
    2. Métodos de Mineração de Itens Frequentes
    3. Métodos de mineração de padrões
    4. Mineração Avançada de Padrões
6. Lidando com incerteza: conjuntos, lógica e sistemas difusos
    1. Fundamentos da Lógica Convencional
    2. Conjuntos Difusos e Lógica Difusa
    3. Lógica e Controlo Difusos
    4. Sistemas Neuro-Difusos
7. Seleção e validação de modelos
    1. Conjuntos de Treino, Teste e Validação
    2. Métodos de validação cruzada e reamostragem
    3. Avaliando o desempenho do modelo
    4. Comparando o desempenho de modelos
    5. Critérios de seleção de modelos

Metodologias de ensino (avaliação incluída)

As aulas serão distribuídas por componentes Teórica (T) e Prática Laboratorial (PL). Nas aulas T será exposta a teoria, e apresentados exemplos de aplicação. Nas aulas PL os estudantes implementarão programas em Python ou Matlab, utilizando os módulos disponíveis em cada uma destas linguagens.
Nesta disciplina haverá um trabalho em grupo e um mini-projeto individual, ambos com um peso de 50% da nota final. Se a nota for igual ou superior a 10, e ambas as componentes forem classificadas com uma nota superior a 8, o aluno passa.

Bibliografia principal

Apontamentos on-line
Cielen, D., A. Meysman, and M. Ali, Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. 2016: Manning Publications.
Myatt, G.J., Making sense of data: a practical guide to exploratory data analysis and data mining. 2007: Wiley-Interscience.
Han, J., M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. 2012: Morgan Kaufmann Publishers.
Moreira, J., A. Carvalho, and T. Horvath, A General Introduction to Data Analytics. 2019: Wiley.

Academic Year	2022-23
Course unit	DATA ANALYTICS
Courses	INFORMATICS ENGINEERING (*) Common Branch (*) Optional course unit for this course
Faculty / School	FACULTY OF SCIENCES AND TECHNOLOGY
Main Scientific Area	CIÊNCIA DE COMPUTADORES
Acronym	
CNAEF code (3 digits)	481
Contribution to Sustainable Development Goals - SGD (Designate up to 3 objectives)	4 8 10
Language of instruction	Portuguese (PT) or English (ENG)
Teaching/Learning modality	Presential

Coordinating teacher António Eduardo de Barros Ruano

Teaching staff	Type	Classes	Hours (*)
António Eduardo de Barros Ruano	PL; T	T1; PL1	28T; 28PL

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

Contact hours

T	TP	PL	TC	S	E	OT	O	Total
28	0	28	0	0	0	0	0	156

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

Knowlegdge of Algorithms and Data Structures, and a a basic course of Probability and Statistics

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

The student, at the end of the course, should be able to:
- Know the main steps of data science
- To know the statistical techniques most used in data analysis
- Know the most important visualization, pre-processing and data mining techniques
- know the fundamentals of fuzzy logic, Mamdani and Sugeno models, and neuro-fuzzy systems; know how to use them in the context of systems approximation and classification.
- Know the most used methods for selection and validation of models

UAlg
UNIVERSIDADE DO ALGARVE

Syllabus

1. The data science process
2. Exploratory data analysis
    1. Brief review of probability theory and common distributions
    2. Descriptive Statistics
    3. Inferential Statistics
    4. Comparative Statistics
3. Data Visualization Techniques
4. Data Preprocessing
    1. Cleaning
    2. Integration
    3. Dimensionality Reduction
    4. Transformation and Discretization
5. Mining Frequent Patterns, Associations, and Correlations
    1. Basic Concepts
    2. Frequent Itemset Mining Methods
    3. Pattern Mining Methods
    4. Advanced Pattern Mining
6. Handling Uncertainty: Fuzzy Sets and Systems
    1. Foundations of Conventional Logic
    2. Fuzzy Sets and Fuzzy Logic
    3. Fuzzy Logic and Control
    4. Neuro-Fuzzy Systems
7. Model selection and validation
    1. Training, Test and Validation sets
    2. Cross-Validation and Resampling Methods
    3. Assessing model performance
    4. Comparing model performance
    5. Model selection criteria

Teaching methodologies (including evaluation)

Classes will be distributed by Theoretical (T) and Laboratory Practice (PL) components. In lectures T the theory will be exposed, and examples of application will be presented. In PL classes students will implement programs in Python or Matlab, using the available modules in each of these languages.
In this discipline there will be a group work and an individual mini-project, both with a weight of 50% of the final grade. If the final grade is equal to or higher than 10, and both components are graded with a mark higher than 8, the student passes.

Main Bibliography

Apontamentos on-line
Cielen, D., A. Meysman, and M. Ali, Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. 2016: Manning Publications.
Myatt, G.J., Making sense of data: a practical guide to exploratory data analysis and data mining. 2007: Wiley-Interscience.
Han, J., M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. 2012: Morgan Kaufmann Publishers.
Moreira, J., A. Carvalho, and T. Horvath, A General Introduction to Data Analytics. 2019: Wiley.