

Ficha UC

Histórico

I - IDENTIFICAÇÃO / IDENTIFICATION

INSTITUIÇÃO / INSTITUTION: Instituto Superior de Engenharia do Porto

CURSO / DEGREE: Licenciatura em Engenharia Informática

UNIDADE CURRICULAR / COURSE TITLE: Análise de Dados em Informática - ANADI

ANO ESCOLAR / ACADEMIC YEAR: 2023-2024

ANO / YEAR	SEMESTRE / SEMESTER	HORAS-SEMANA / HOURS-WEEK *	ECTS
3	2º Semestre	T: 1.5; TP: 2	4

* T: TEÓRICA / LECTURE; TP: TEÓRICO-PRÁTICA / PRACTICAL; PL: PRÁTICA LABORATORIAL / LABORATORY; TC: TRABALHO DE CAMPO / FIELD WORK; OT: ORIENTAÇÃO TUTORIAL / TUTORIAL

PRÉ-REQUISITOS FORMAIS (PRECEDÊNCIAS) / FORMAL PRECEDENCES:

PORTUGUÊS

ENGLISH

Não tem

None

DOCENTES / PROFESSORS

	NOME / NAME	SIGLA / ACRONYM	HABILITAÇÕES / QUALIFICATIONS
RESPONSÁVEL / RESPONSIBLE	Ana Maria Dias Madureira Pereira	AMD	Doutoramento
	Diogo Marcelo Esterlita Nogueira	DEN	Mestrado
	Elsa Maria De Carvalho Ferreira Gomes	EFG	Doutoramento
	João Emilio Raimundo Carrilho de Matos	JEM	Doutoramento
	Teresa Paula Soares De Araujo	TPA	Doutoramento
OUTROS / OTHERS			

II - PROPÓSITOS, RESUMO, CARATERIZAÇÃO / PURPOSES, OVERVIEW, DESCRIPTION

ENQUADRAMENTO / FRAMEWORK

PORTUGUÊS

ENGLISH

A unidade curricular(UC) de Análise de Dados em Informática faz parte do 3º ano-2º semestre do ciclo de estudos de Engenharia Informática.

Esta UC visa dar conhecimentos estruturantes no planeamento do estudo estatístico no âmbito da Análise Estatística de Dados (AED) para problemas reais no suporte à tomada de decisão. A UC visa fornecer aos alunos conhecimentos e competências na aplicação de tópicos clássicos (classificação e regressão) e uma introdução a tópicos de Aprendizagem Automática (AA) no tratamento de grandes volumes de dados(Big Data) numa perspectiva da Ciência dos Dados.

This course is part of the 3rd year - 2nd semester of Computer Engineering study cycle.

This UC aims to provide structuring knowledge in the planning of the statistical study within the scope of the Exploratory Data Analysis (EDA) for problems in the area of Informatics, particularly related to the support of planning and management of decision making. The UC aims to provide students with knowledge and skills in the application of classic topics (classification and regression) and topics of Machine Learning (ML) in the treatment of Big Data from a Data Science perspective.

CONHECIMENTOS PRÉVIOS ASSUMIDAMENTE ADQUIRIDOS / REQUIRED PREVIOUS KNOWLEDGE

PORTUGUÊS

ENGLISH

Espera-se que o aluno que frequente esta unidade curricular tenha conhecimentos básicos prévios relacionados com Probabilidades e Estatística, incluindo Testes de Hipóteses (MATCP).

It is expected that students who attends this Curricular Unit has previous knowledge on the basic concepts of Probability and Statistics, including Hypothesis Tests (MATCP)

PROPÓSITOS E OBJETIVOS / PURPOSES AND OBJECTIVES

PORTUGUÊS

ENGLISH

Pretende-se que o estudante, após concluir esta UC, reconheça a necessidade de utilizar técnicas de AED e que compreenda alguns dos algoritmos de aprendizagem automática.

No final desta UC, o/a estudante deverá ser capaz de:

CO1. Caracterizar o estado da arte das técnicas estatísticas/ferramentas para AED e AA e sua potencial aplicação em engenharia e ciências aplicadas.

CO2. Discutir as técnicas básicas de AED e AA para a conceção de experiências no domínio da Ciência dos Dados.

CO3. Analisar e organizar dados de uma diversidade de fontes.

CO4. Discutir as diferentes técnicas de estatística descritiva e inferencial para implementar a AED.

CO5.. Identificar, selecionar e usar ferramentas de AA adequadas no suporte ao processo da Ciência dos Dados.

CO6. Formular problemas reais no contexto de terminologia de AA e identificação da abordagem mais adequada para a resolução do problema.

CO7. Construção e otimização de modelos relevantes dos dados.

CO8. Avaliar metódica e criticamente o desempenho dos modelos.

CO9. Trabalhar em grupo e produzir relatórios técnicos e artigos científicos, e comunicação oral em português/inglês.

It is intended that the student, after completing this UC, recognizes the need to use EDA techniques and understands some of the machine learning (ml) algorithms.

Especific Objectives

By the end of this course, the student must be able to:

CO1 Characterize the state of the art of statistical techniques/tools for AED and ML and its potential application in engineering and applied sciences.

CO2-Discuss the basic techniques of EDA and ML for the design of experiences in the field of Data Science.

CO3-Analyze and organize data from a variety of sources

CO4-Discuss the different descriptive and inferential statistical techniques to implement EDA

CO5-Identify, select and use appropriate ML tools to support the Data Science process

CO6-Formulate real problems in the context of ML and identify the most appropriate approach to problem solving.

CO7- Construction and optimization of relevant data models

CO8. Evaluate methodical and critically the performance of the models.

CO9-Develop Teamwork and produce technical reports and scientific articles, and oral communication in Portuguese / English.

PROGRAMA / PROGRAMME

PORTUGUÊS

Programa (max. 1000 caracteres) 1.Data Science/BigData/Data Analytics (2h-30%T+70%TP) 2. Análise Exploratória de Dados(12h-30%T+70%TP) - Inferência Estatística - Testes de hipóteses paramétricos e não-paramétricos - Regressão e Correlação 3.Técnicas de Aprendizagem Automática (31h-20%T+80%TP). - Noções e aplicabilidade de técnicas de AA - Classificação das Técnicas de AA (Aprendizagem supervisionada, Aprendizagem não-supervisionada, Aprendizagem por reforço, Deep Learning) - Otimização e Avaliação dos modelos: holdout e K-fold based cross-validation, Métricas de avaliação dos modelos de regressão e de classificação - Técnicas de Aprendizagem Supervisionada para regressão (modelos de regressão linear e árvores de Regressão) - Técnicas de Aprendizagem Supervisionada para classificação (árvores de decisão, redes neurais, kNN, e máquinas de suporte vetorial (SVM))

ENGLISH

1. Data Science / Big Data / Data Analytics (2h-30% T + 70% TP) 2. Exploratory Data Analysis (12h-30% T + 70% TP) - Statistical Inference - Parametric hypothesis tests and Nonparametric hypothesis tests - Regression and Correlation 3. Machine Learning Techniques (31h-20% T + 80% TP). - Notions and applicability of ML techniques - Classification of ML Techniques (Supervised learning, Unsupervised learning, Reinforcement learning, Deep learning) - Optimization and assessment of the models: holdout and K-fold based cross-validation, Evaluation metrics of the regression and classification models - Supervised Learning Techniques for Regression (Linear Regression Models and Regression Trees) - Supervised Learning Techniques for Classification (Decision Trees, Neural Networks, kNN, and Support Vector Machines (SVM))
--

MATERIAL E FERRAMENTAS DE ENSINO MAIS IMPORTANTE / MOST IMPORTANT STUDING MATERIAL AND TOOLS

PORTUGUÊS

Material de Ensino - Textos e Slides, Ana Madureira e João Matos - Douglas C. Montgomery, Design and Analysis of Experiments, 8th edition. John Wiley & Sons, New York, 2013. - Wes Mckinney, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition, https://wesmckinney.com/book/ , O'Reilly Media, 2022. - A. Subasi, Practical Machine Learning for Data Analysis Using Python, Academic Press, ISBN 978-0-12-821379-7, 2020. Ferramentas de ensino/aprendizagem - Moodle (http://moodle.isep.ipp.pt/) - Linguagem de Programação: Python - Ambiente desenvolvimento: Jupyter Notebook (https://www.anaconda.com/download) - Repositório de código (e.g. Bitbucket, GitHub)
--

ENGLISH

Most important studying material and tools Main Teaching Material - Course Texts and Slides (in English), Ana Madureira and João Matos - Douglas C. Montgomery, Design and Analysis of Experiments, 8th edition. John Wiley & Sons, New York, 2013. - Wes Mckinney, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition, https://wesmckinney.com/book/ , O'Reilly Media, 2022. - A. Subasi, Practical Machine Learning for Data Analysis Using Python, Academic Press, ISBN 978-0-12-821379-7, 2020. Teaching/Learning Tools - Moodle (http://moodle.isep.ipp.pt/) - Programming Language: Python - Development environment: Jupyter Notebook (https://www.anaconda.com/download) - Code Repository (e.g. Bitbucket, GitHub)

MATERIAL DE ENSINO COMPLEMENTAR / SUPPLEMENTARY STUDING MATERIAL

PORTUGUÊS

- I. Chatterjee, Machine Learning and Its Application: A Quick Guide for Beginners, Bentham Science Publishers, ISBN 978-1-68108-940-9, 2021. - Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists. 4th edition, Elsevier Academic Press, 2009. -Tom Mitchell, Machine Learning. McGraw-Hill, 1997. - Christopher Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
--

ENGLISH

- I. Chatterjee, Machine Learning and Its Application: A Quick Guide for Beginners, Bentham Science Publishers, ISBN 978-1-68108-940-9, 2021. - Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists. 4th edition, Elsevier Academic Press, 2009. -Tom Mitchell, Machine Learning. McGraw-Hill, 1997. - Christopher Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
--

METODOLOGIA ENSINO-APRENDIZAGEM / TEACHING-LEARNING METHODOLOGY

PORTUGUÊS

Nas aulas T serão usados o método expositivo e interrogativo e sempre que adequado serão usadas técnicas do método ativo. Nas aulas TP serão usados preferencialmente técnicas do método ativo(trabalho de grupo, estudo de casos e aprendizagem baseada em problemas). As estratégias usadas visam ajudar o aluno a assumir uma atitude de aprendizagem ativa, colaborativa e responsável, trabalho persistente e de aplicação de espírito crítico na análise e resolução de problemas.
--

ENGLISH

In Lectures(T) classes, expositive and interrogative method will be used and when appropriate active methods (brain storming, case studies). In Practical (TP) classes, active learning techniques will be used, preferentially, as teamwork, case studies and problem-based learning. The used strategies aimed at helping the student to assume an active learning attitude, collaborative, responsible, persistent work and application of critical analysis on the problem resolution.
--

DISTRIBUIÇÃO PERCENTUAL ESTIMADA DOS CONTEÚDOS / ESTIMATED PERCENTAGE DISTRIBUTION OF THE CONTENTS

PORTUGUÊS / ENGLISH

COMPONENTE CIENTÍFICA / SCIENTIFIC COMPONENT	50
COMPONENTE TECNOLÓGICA / TECHNOLOGICAL COMPONENT	40
COMPONENTE CONTEXTO ENVOLVENTE / SURROUNDING CONTEXT COMPONENT	10

RESULTADOS EXPETÁVEIS / OUTCOMES

PORTUGUÊS

Resultados expectáveis (em conformidade com os critérios EUR-ACE)
Knowledge and Understanding #KU CO1, CO2, CO4
Engineering Analysis (#EA) CO2, CO3, CO4, CO5, CO6
Engineering Design (#ED) CO3, CO4, CO6, CO7, CO8
Investigations (#IN) CO1, CO2,CO4, CO5
Engineering Practice (#EP) CO3, CO5, CO6, CO7,CO8
Making Judgments (#MJ) CO5, CO6, CO8
Communication and Team-working (#CT) CO9
Lifelong Learning (#LL) CO1, CO2,CO4, CO5

ENGLISH

Outcomes (according with the EUR-ACE criteria)
Knowledge and Understanding #KU CO1, CO2, CO4
Engineering Analysis (#EA) CO2, CO3, CO4, CO5, CO6
Engineering Design (#ED) CO3, CO4, CO6, CO7, CO8
Investigations (#IN) CO1, CO2,CO4, CO5
Engineering Practice (#EP) CO3, CO5, CO6, CO7,CO8
Making Judgments (#MJ) CO5, CO6, CO8
Communication and Team-working (#CT) CO9
Lifelong Learning (#LL) CO1, CO2,CO4, CO5

III - PROCEDIMENTOS DE AVALIAÇÃO / EVALUATION PROCEDURES

TIPO DE AVALIAÇÃO / EVALUATION TYPE

AVALIAÇÃO DURANTE O PERÍODO LETIVO COM AVALIAÇÃO FINAL FACULTATIVA / EVALUATION DURING THE TERM WITH OPTIONAL FINAL EVALUATION

Os estudantes têm a possibilidade de realizar toda a avaliação durante o período letivo. Caso obtenham aprovação, estão dispensados da realização de avaliação final. Se isso não acontecer e os mínimos das componentes não repetíveis forem atingidos, têm ainda a possibilidade de realizar avaliação final, nas épocas normal e de recurso.

Students are given the possibility to carry out all items of evaluation throughout the term. If approved, students are dismissed from the final evaluation. If this does not happen, and all non repeatable moments minimums are met, students still have the possibility to be evaluated at the end of the term.

AVALIAÇÃO DURANTE O PERÍODO LETIVO / EVALUATION DURING THE TERM

NÚMERO DE MOMENTOS NÃO REPETÍVEIS (NR) / NUMBER OF NON REPEATABLE EVALUATION MOMENTS (NR)	0
NÚMERO DE MOMENTOS REPETÍVEIS (R) / NUMBER OF REPEATABLE EVALUATION MOMENTS (R)	2

PESOS E MÍNIMOS / WEIGHTS AND MINIMUMS

COMPONENTES / COMPONENTS	TIPO / TYPE	PESO / WEIGHT (%)	MÍN / MIN
M1 - R Trabalho Prático 1	R	45	-
M2 - R Trabalho Prático 2	R	55	-
MÉDIA PONDERADA MOMENTOS REPETÍVEIS / WEIGHED AVERAGE REPEATABLE MOMENTS		100	-
NOS EXAMES, PERMITE A REPETIÇÃO INDIVIDUALIZADA DOS MOMENTOS DE AVALIAÇÃO E EXAME PARCIAL (SE APLICÁVEL)? / IN THE EXAMS, THE INDIVIDUALIZED REPETITION OF EVALUATION MOMENTS AND PARTIAL EXAM (IF APPLICABLE) IS ALLOWED?			
NORMAL / 1 ST EXAM: <input checked="" type="radio"/> SIM / YES <input type="radio"/> NÃO / NO RECURSO / 2 ND EXAM: <input checked="" type="radio"/> SIM / YES <input type="radio"/> NÃO / NO			
EXAME GLOBAL / GLOBAL EXAM	EG	100	-

T: PROVAS ESCRITAS (TESTES, QUESTÕES, ...) / WRITTEN TESTS;
L: TRABALHOS DE LABORATÓRIO OU DE CAMPO / LAB WORKS, OR FIELDWORKS;
P: PROJETO, ESTÁGIO, DISSERTAÇÃO / PROJECT, THESIS;
R: RELATÓRIOS, TRABALHOS DE PESQUISA, TRABALHOS APLICADOS / PAPERS, APPLIED WORKS, ACADEMIC WORKS;
A: APRESENTAÇÕES / PRESENTATIONS;
EG: AVALIAÇÃO GLOBAL EM PERÍODO DE EXAMES (EXAMES OU OUTROS) / GLOBAL EVALUATION AT THE END OF THE TERM (EXAMS OR OTHERS);
MG: MELHORIA GLOBAL / GLOBAL IMPROVEMENT;
CF: CLASSIFICAÇÃO FINAL / FINAL CLASSIFICATION

FÓRMULA(S) DE CÁLCULO DA CLASSIFICAÇÃO FINAL / FINAL CLASSIFICATION FORMULA(S) (CF)

CF = M1*0.45 + M2*0.55
CF = EG

OBSERVAÇÕES / COMMENTS

PORTUGUÊS

A avaliação da UC é suportada num Trabalho Prático, com 2 iterações, com um peso de 100% na nota final da UC. A avaliação consiste na realização de 1 Trabalho Prático em grupo, de realização extra-aulas (submetido no Moodle), com defesa/apresentação obrigatória, em grupo e individual. Instrumentos de avaliação: M1 - Trabalho Prático iteração 1 (Análise Exploratória de Dados e Inferência Estatística), realização extra-aulas, defesa obrigatória, peso de 45% na nota final M2 - Trabalho Prático iteração 2 (Aprendizagem Automática), realização extra-aulas, apresentação e defesa obrigatória, peso de 55% na nota final. Cada iteração do trabalho prático só será considerada terminada após apresentação e discussão, em grupo e individual.

ENGLISH

The CU assessment is supported in a Practical Work, with 2 iterations, with a weight of 100% in the final grade of the CU. The assessment consists of 1 Practical Work in teamwork, extra-classes (submitted in Moodle), with mandatory defense/presentation, in group and individual. Evaluation tools: M1 - Practical Work iteration 1 (Exploratory Data Analysis and Statistical Inference), extra-classes, mandatory defense, weight of 45% in the final grade M2 - Practical work iteration 2 (Machine Learning), extra-classes, mandatory presentation and defense, weight of 55% in the final grade. Each iteration of the practical work will only be considered completed after presentation and discussion, in group and individual.
--

MELHORIA DE NOTA DE AVALIAÇÃO / GRADE IMPROVEMENT

MELHORIA GLOBAL / GLOBAL IMPROVEMENT (MG):

EXAME / EXAM

CF = MG

NOTA BIOGRÁFICA DO RESPONSÁVEL DA UNIDADE CURRICULAR / BIOGRAPHICAL NOTE OF RESPONSIBLE OF THE COURSE

PORTUGUÊS

A docente responsável pela disciplina é doutorada em Engenharia de Produção e Sistemas - Inteligência Artificial e Escalonamento Distribuído pela Universidade do Minho em 2003, Mestre em Engenharia Electrotécnica e de Computadores na área de especialização em Informática Industrial pela FEUP em 1996 e licenciada em Engenharia Informática Industrial pelo ISEP em 1993. Exerce funções de Professor Coordenador com Agregação no Departamento de Engenharia Informática do ISEP/IPP e é Diretora do grupo de I&D ISRC. Foi chair da secção Portuguesa da IEEE (2015-2017). Foi secretária da região 8 do IEEE (2019-2020). Co-fundadora do Portuguese Chapter of IEEE Computational Intelligence Society e do Portuguese Chapter of IEEE Systems, Man, and Cybernetics Society. Tem publicado, no últimos anos mais de 150 artigos científicos em conferências, revistas e capítulos de livros internacionais.

ENGLISH

The CU responsible professor has a Ph.D. in Production and Systems Engineering - Artificial Intelligence and Distributed Scheduling - form University of Minho in 2003. She got his Master degree in Electrical and Computers Engineering - Industrial Informatics, in 1996, from FEUP, and BSc degree in Computer Engineering in 1993 from ISEP. She was Chair of IEEE Portugal Section (2015-2017). She was IEEE Region 8 Secretary (2019-2020). Co-founder of the Portuguese Chapter of IEEE Computational Intelligence Society and Portuguese Chapter of IEEE Systems, Man, and Cybernetics Society. She became IEEE Senior Member in 2010. Currently she is Coordinator Professor with Habilitation at the Institute of Engineering/Polytechnic of Porto (ISEP/IPP) and Director of the ISRC Research Group. In the last few years, she was author of more than 150 scientific papers in international scientific conference proceedings, journals and books.

IV - INFORMAÇÃO PARA A3ES / INFORMATION FOR A3ES

DEMONSTRAÇÃO DA COERÊNCIA DOS CONTEÚDOS PROGRAMÁTICOS COM OS OBJETIVOS DA UNIDADE CURRICULAR / DEMONSTRATION OF THE SYLLABUS COHERENCE WITH THE CURRICULAR UNIT'S OBJECTIVES

PORTUGUÊS

Apresenta-se a relação entre as secções dos conteúdos programáticos e as competências/objetivos:

1. Data Science/BigData/Data Analytics (CO1,CO2)
2. Análise Exploratória de Dados - AED CO1, CO2, CO3, CO6, CO7, CO8,CO9)
3. Técnicas de Aprendizagem Automática (CO1,CO2,CO3,CO4,CO5,CO6,CO7,CO8,CO9).

ENGLISH

It presents the relation between the sections of the programmatic contents (syllabus) and the competences/objectives to be developed by students:

1. Data Science / BigData / Data Analytics (CO1, CO2)
2. Exploratory Data Analysis - EDA (CO1, CO2, CO3, CO6, CO7, CO8,CO9)
3. Machine Learning Techniques (CO1, CO2, CO3, CO4, CO5, CO6, CO7, CO8,CO9).

DEMONSTRAÇÃO DA COERÊNCIA DAS METODOLOGIAS DE ENSINO COM OS OBJECTIVOS DE APRENDIZAGEM DA UNIDADE CURRICULAR / DEMONSTRATION OF THE COHERENCE BETWEEN THE TEACHING METHODOLOGIES AND THE LEARNING OUTCOMES

PORTUGUÊS

As aulas T terão como principal foco a sistematização de conceitos teóricos relacionados com técnicas de estatística descritiva e inferencial na AED na análise estatística de problemas oriundos da Engenharia Informática. As aulas teórico-práticas(TP) terão uma forte componente tecnológica, com uso intensivo da linguagem de programação Phyton, para apoiar a dinâmica e objetivos da aula. Preferencialmente, as aulas devem ser em laboratórios com PC's e portáteis. Considerando que o principal objetivo da UC é proporcionar aos alunos competências para analisar e organizar os dados a partir de uma diversidade de fontes, o uso de técnicas de estatística descritiva e inferencial para implementar a AED e o desenho/especificação de testes/experiências para avaliar a análise de desempenho de abordagens baseadas em AA, considera-se que a forma mais eficaz para avaliar as competências adquiridas é através de trabalhos práticos com dinamização do trabalho em equipa.

ENGLISH

The T classes will have as main focus the systematization of theoretical concepts related to descriptive and inferential statistical techniques in EDA for statistical analysis of problems coming from Computer Engineering field. The Practical (TP) classes will have a strong technological component with intensive use of programming language Phyton to support the class dynamics. Preferentially, classes should be on labs with PC and laptops. Considering that the main purpose of the Curricular Unit is to provide students competences to analyze and organize data from a diversity of sources, use techniques of descriptive and inferential statistics to implement EDA and design/conduct experiments to evaluate performance analysis of ML based approaches it is considered that the more effective to assess acquired skills is trough Teamwork Projects.