



MESTRADO EM ANÁLISE E ENGENHARIA DE BIG DATA

MASTER ON ANALYSIS AND ENGINEERING OF BIG DATA

MESTRADO EM ANÁLISE E ENGENHARIA DE BIG DATA

MASTER ON ANALYSIS AND ENGINEERING OF BIG DATA

FICHA TÉCNICA

Edição: Departamento de Informática FCT NOVA

Design: Bárbara Teixeira

Impressão: Tipografia Lobão, Lda

Tiragem: 1000 exemplares

Ano: 2018

1

Página 6 | Page 6

Apresentação
Presentation

2

Página 12 | Page 12

Organização
Organization

3

Página 14 | Page 14

O Curso
Program

4

Página 28 | Page 28

Parcerias
Partnerships

1

Apresentação Presentation

Análise e Engenharia de Dados

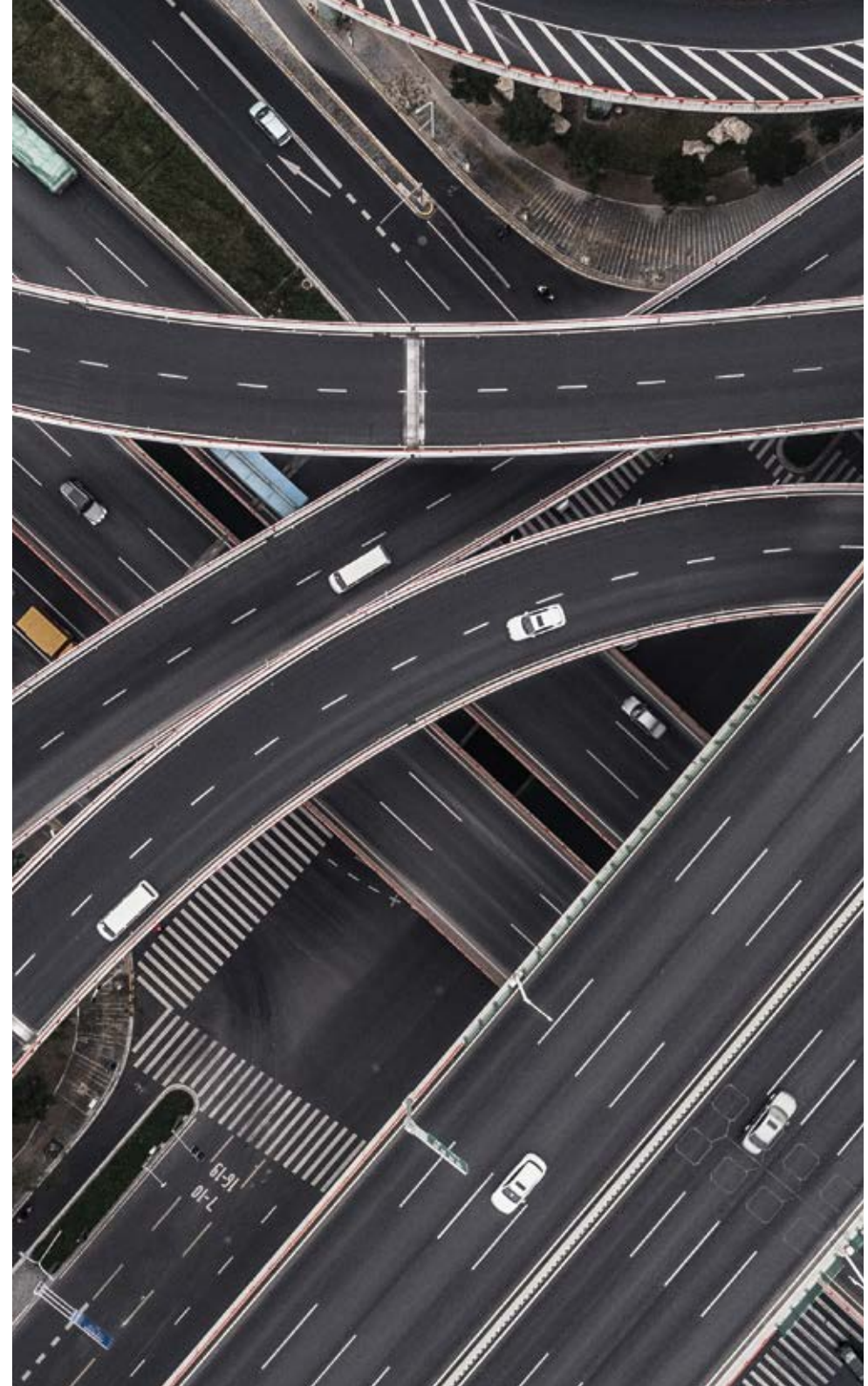
A transformação digital da sociedade, a explosão da Internet e a popularização das redes sociais, geram atualmente dados digitais de volumes gigantescos em setores chave como a saúde, a administração pública, media e comunicação social, marketing e comércio eletrónico, finanças, energia, ambiente e urbanismo, telecomunicações, indústria farmacêutica e bioinformática.

A Análise e Engenharia de Big Data centra-se no desenvolvimento de metodologias para lidar com grandes volumes de dados, com o objetivo de extrair informação útil. Adquirir, armazenar, gerir e organizar os dados são as questões que se colocam nesta área de estudo. Assiste-se a uma crescente procura e empregabilidade de especialistas nesta temática que irão desempenhar nos próximos anos um papel cada vez mais interventivo e gerador de valor nos processos de gestão e inovação em todos os domínios da indústria e dos serviços.

Big Data Analysis and Engineering

The digital transformation of society, the explosion of the Internet and the popularization of social networks, led to the generation of huge volumes of digital data in key sectors such as health, public administration, media and social communication, marketing and e-commerce, finance, energy, environment and urbanism, telecommunications, pharmaceutical industry and bioinformatics.

Big Data Analysis and Engineering focuses on the development of methodologies to handle large volumes of data, in order to extract useful information. Acquiring, storing, managing and organizing data are the issues that arise in this area of study. There is a growing demand and employability of specialists in this area who will play an increasingly intervening and value-creating role in management and innovation processes in all areas of industry and services in the coming years.





Objectives

The Master on Analysis and Engineering of Big Data is aimed at training analysts, project development leaders and innovation experts in the emerging field of Data Science and Data Engineering.

The course develops competences for processing and analysing large volumes of data by advanced computational and mathematical methods, and in methodologies for searching and finding answers to management, monitoring, and optimisation processes, or for extracting knowledge, trends, correlations, or forecasts through automatic machine learning.

The course also includes curricular units that expose students to various case studies, where institutions (companies, public institutes, research centers, etc) that obtain and / or manage large volumes of data wish to extract knowledge and useful information from such data, favoring the identification of Dissertation themes.

The objectives of the course are aligned with the “National Digital Competence Initiative e.2030”, in the axes of specialization (item qualification and creation of added value in economics) and research (big data item).

Objetivos

O Mestrado em Análise e Engenharia de Big Data tem como objetivo a formação de analistas, líderes de desenvolvimento de projeto e especialistas em inovação no campo emergente da Ciência de Dados e Engenharia de Dados.

O curso desenvolve competências de processamento e análise de grandes volumes de dados por métodos computacionais e matemáticos avançados, em metodologias para procurar e encontrar respostas necessárias aos processos de gestão, monitorização, e otimização, ou para extrair conhecimento, tendências, correlações, ou previsões, em particular através de aprendizagem automática.

O curso inclui ainda unidades curriculares que expõem os estudantes a vários casos de estudo, em que as instituições (empresas, institutos públicos, centros de investigação, etc) que obtêm e/ou administram grandes volumes de dados pretendem retirar conhecimento e informação útil desses dados, favorecendo a identificação de temas de Dissertação.

Os objetivos do curso estão alinhados com a “Iniciativa Nacional Competências Digitais e.2030”, nos eixos da especialização (item “qualificação e criação de maior valor acrescentado na economia”) e de investigação (item “big data”).

Destinatários e Perfis Profissionais

O curso destina-se a candidatos com formação certificada ao nível de um primeiro ciclo de estudos (licenciatura ou equivalente), possuidores de bases matemáticas e de programação sólidas.

Os especialistas na área de Big Data são especialmente procurados pelas empresas e instituições onde são gerados ou são consumidos grandes volumes de dados.

Para além das empresas parceiras da FCT NOVA no âmbito do curso, podem referir-se várias empresas e instituições nas áreas da saúde, administração pública, comércio eletrónico e marketing, finanças, energia, meio ambiente e urbanismo, telecomunicações, media e comunicação social, indústria farmacêutica ou biotecnológica, entre outras, interessadas em profissionais com este perfil.

Como exemplos de Perfis Profissionais muito procurados, podem-se referir áreas de especialização como as seguintes, nas quais o estudante do curso se pode focalizar através da escolha de unidades curriculares adequadas, em aconselhamento com a comissão científica do curso.

Chief Data Officer

Data Analyst

Big Data Visualizer

Big Data Solutions Architect

Big Data Engineer

Big Data Researcher

Data Warehouse Manager

Data Architect

Data Mining Analyst

Target Audience and Professional Roles

Candidates should have certified competences at the level of a first cycle of studies, with strong mathematical and programming bases.

Specialists in the area of Big Data are especially sought after by companies and institutions where large volumes of data are generated or consumed.

In addition to the companies that collaborate in the organization of the master's degree, there are several companies and institutions in the areas of health, public administration, e-commerce and marketing, finance, energy, environment and urbanism, telecommunications, media and social networks, pharmaceutical industry or biotechnology, among others, which are interested in professionals with this profile.

As examples of high demand professional profiles, we may refer to specialization areas such as those below. The students have the opportunity of focusing their study plan in one or more of such profiles by selecting appropriate elective courses, under advice of the program coordinating committee.

Chief Data Officer

Data Analyst

Big Data Visualizer

Big Data Solutions Architect

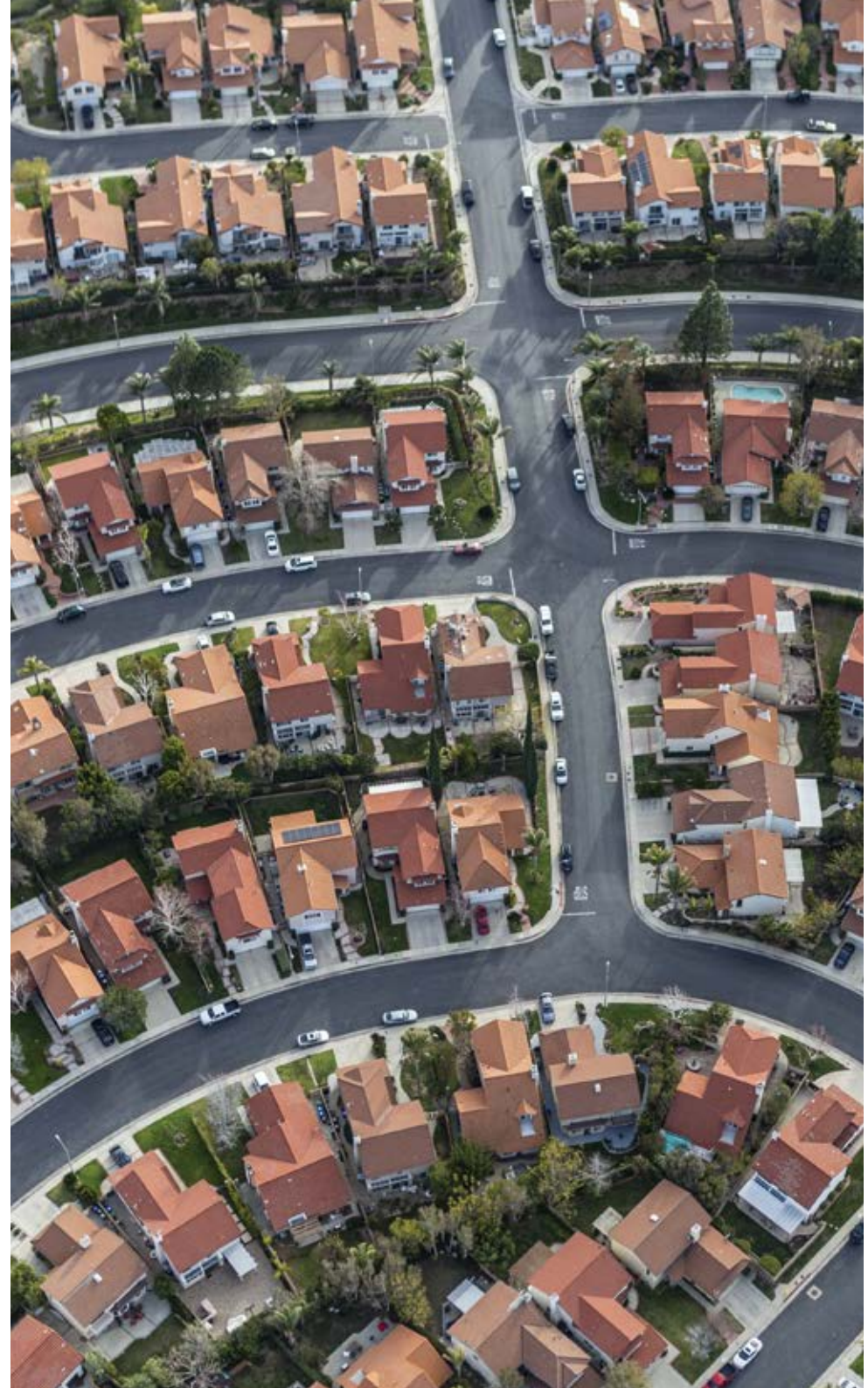
Big Data Engineer

Big Data Researcher

Data Warehouse Manager

Data Architect

Data Mining Analyst



2

Organização Organisation

O Mestrado em Análise e Engenharia de Big Data é organizado conjuntamente pelo Departamento de Informática e pelo Departamento de Matemática, com a colaboração de várias empresas, nomeadamente na unidade curricular de Seminário e na Dissertação / Projeto de Mestrado.

The Master in Analysis and Engineering of Big Data is jointly organised by the FCT NOVA Department of Informatics and Department of Mathematics, with the collaboration of several companies, namely in the Seminar curricular unit and in the final Master Dissertation / Project.



About the Department of Informatics

The Department of Informatics of FCT NOVA is a pioneering and reference institution in higher education and scientific research in Informatics in Portugal, having already graduated thousands of computer engineers. It is currently responsible for training more than 800 students enrolled in various courses and at various levels of training. The department hosts a research center, NOVA LINCIS, where the research done in the department is coordinated.

About the Department of Mathematics

The Department of Mathematics is a teaching and research unit of FCT NOVA, where multiple activities in the field of Mathematics are developed, ranging from teaching and research to promotion and dissemination, through lifelong learning and the provision of services to community. The Department hosts a research center, CMA, where research is coordinated in the department.

Sobre o Departamento de Informática

O Departamento de Informática é uma unidade de ensino e investigação da FCT NOVA, pioneira e de referência no ensino superior e na investigação científica em Informática em Portugal, tendo já formado milhares de engenheiros informáticos. Atualmente é responsável pela formação de mais de 800 alunos inscritos em vários cursos e em diversos níveis de formação. O Departamento hospeda um centro de Investigação, Laboratory for Computer Science and Informatics (NOVA LINCIS) onde é coordenada a investigação desenvolvida no mesmo.

Sobre o Departamento de Matemática

O Departamento de Matemática é uma unidade de ensino e investigação da FCT NOVA, onde são desenvolvidas múltiplas atividades no domínio da Matemática que vão do ensino e da investigação à promoção e divulgação, passando pela formação ao longo da vida e pela prestação de serviços à comunidade. O Departamento hospeda um centro de investigação, Centro de Matemática e Aplicações (CMA), onde é coordenada a investigação feita no mesmo.

3

O Curso Program

O Mestrado em Análise e Engenharia de Big Data (MAEBD) é um curso de segundo ciclo que visa a educação de analistas, líderes de desenvolvimento de projeto e especialistas em inovação no campo emergente da Ciência de Dados e Engenharia, que está atualmente a criar novas oportunidades para a gestão de processos e otimização de produtos em todos os campos de atividade. Sendo uma colaboração equilibrada entre o Departamento de Informática e Departamento de Matemática da FCT NOVA, o MAEBD reúne competências computacionais e matemáticas fundamentais para a resolução de problemas baseada na análise de grandes volumes de dados, proporcionando uma compreensão dos princípios e fundamentos científicos e tecnológicos dos métodos atuais e emergentes.

The Master in Big Data Analytics and Engineering (MAEBD) is a second cycle program that targets the education of analysts, project development leaders and innovation specialists in the emergent field of Data Science and Engineering, which is nowadays creating novel opportunities for process management and product optimization in every field of activity. Building on an evenly balanced collaboration between the FCT NOVA Department of Mathematics, and Department of Informatics, the MAEBD brings together fundamental computational and mathematical skills for large scale data driven problem solving while providing a principled understanding of scientific and technological underpinnings of current and emerging methods.

The rapidly expanding area of Big Data includes Data Science topics such as: evaluation and quality; prediction; risk analysis and decision; large graph and network analysis; longitudinal analysis; machine learning topics such as feature selection, neural networks, and dimensionality reduction; data warehousing topics such as information search and retrieval, data and knowledge mining and acquisition; and Data Engineering and Computing Infrastructures topics such as big data analytics, stream processing, and cloud computing. It finds applications in general business, health, e-commerce, finance, energy, environment and urbanism, telecom, pharmaceuticals, media, social networks, bioinformatics, forecasting, marketing, general process and product management and innovation in all fields of industry and services.

A área de Big Data, em rápida expansão, inclui temas de ciência de dados, tais como: avaliação e qualidade; previsão; análise de risco e decisão; análise de grafos e redes de grande dimensão; análise longitudinal; vários tópicos de aprendizagem automática, tais como a seleção de características, redes neuronais e redução de dimensionalidade; tópicos de “Data Warehousing” tais como pesquisa e recuperação da informação, mineração e aquisição de dados e conhecimentos; e ainda tópicos de Engenharia de Dados e Sistemas Computacionais, tais como análise de grandes volumes de dados, processamento de fluxo de dados (“streams”), e computação na nuvem (“cloud computing”). Estes temas encontram aplicação na gestão de atividades em geral, e em particular na saúde, comércio eletrónico, finanças, energia, meio ambiente e urbanismo, telecomunicações, indústria farmacêutica e bioinformática, media e redes sociais, previsão, marketing, gestão de produtos e processos e inovação em todos os domínios da indústria e dos serviços.





The overall structure of the MAEBD curriculum accommodates a variety of graduate profiles between Data Science and Data Engineering, while ensuring a strong core on key science, engineering and technology principles.

The first year is composed of formal courses. In the first semester the students are offered a mandatory block of 5 courses, 3 in the CS / Informatics area and 2 in the Mathematics area. In the second semester there is a compulsory 3 ECTS requirement in Entrepreneurship, common to all second cycle programs at FCT NOVA, and a 3 ECTS Seminar course, organized in collaboration with partner companies (where students learn from invited practitioner speakers about relevant scenarios of big data engineering and analytics in various societal domains).

The student is also required to select 4 elective courses. From these 4 courses at least one must belong to each of the two main scientific areas of the program (either Mathematics or Computer Science and Informatics), and one from “Bloco Livre FCT”, including courses of all scientific areas of FCT NOVA, yearly defined by the Scientific Council of FCT NOVA. The third and fourth semester are dedicated to the preparation and elaboration of the master project / thesis curricular unit, typically developed in collaboration with an external partner entity providing some real world scenario and problem, but which may also accommodate research oriented thesis.

O MAEBD acomoda vários perfis de pós-graduação entre Ciência de Dados e Engenharia de Dados, com um forte núcleo de fundamentos da ciência, engenharia e tecnologia.

O primeiro ano é preenchido com cursos formais, com um bloco obrigatório no primeiro semestre incluindo 5 unidades curriculares, 3 na área de Informática e 2 na área de Matemática. No segundo semestre são exigidos 3 ECTS em Empreendedorismo, comum a todos os programas de segundo ciclo na FCT NOVA e 3 ECTS na unidade curricular Seminário, organizada em colaboração com as empresas parceiras (onde se apresentam cenários relevantes para a engenharia e análise de grandes volumes de dados em vários domínios da sociedade).

São ainda requeridas 4 unidades curriculares opcionais, de escolha livre a partir de uma oferta adequada, das quais pelo menos uma deve pertencer a cada uma das áreas científicas do programa (i.e. Informática e Matemática) e uma ao Bloco Livre FCT, que inclui unidades curriculares de todas as áreas científicas da FCT NOVA. Este bloco de unidades curriculares é definido anualmente pelo Conselho Científico da FCT NOVA.

O terceiro e quarto semestres são dedicados à dissertação, normalmente desenvolvida em colaboração com uma entidade externa, parceira do curso, que forneça um problema e cenário do mundo real, mas que pode igualmente ter o perfil de uma dissertação de investigação orientada.

Funcionamento

As aulas são lecionadas na FCT NOVA, Campus da Caparica. O horário de funcionamento do mestrado é diurno. No entanto para permitir a frequência de alunos trabalhadores, está previsto que:

- As aulas do semestre ímpar tenham lugar ao final do dia (todos os dias da semana das 16:00 às 20:00), e que no semestre par, a maioria das unidades curriculares sejam também disponibilizadas com este horário, embora algumas unidades curriculares opcionais possam decorrer noutros horários.
- O curso possa ser realizado em tempo parcial, ao longo de quatro anos, podendo os alunos realizar metade das unidades curriculares previstas em cada semestre, distribuindo-se os custos das propinas de forma proporcional em relação à frequência normal em full-time.

ÁREA CIENTÍFICA SCIENTIFIC AREA	SIGLA ACRONYM	ECTS OBRIGATÓRIOS MANDATORY ECTS	ECTS MÍNIMOS OPTATIVOS MINIMUM OPTIONAL ECTS
Informática Computer Science and Informatics	I	18	6
Matemática Mathematics	M	12	6
Competências Complementares Professional Skills	CC	3	0
Qualquer Área Científica Any Scientific Area	QAC	0	6
Informática ou Matemática Mathematics or Computer Science and Informatics	M/I	63	6
Total Total		96	24

Operation

The master's classes are given in daytime working hours in the FCT NOVA Campus of Caparica. However, to facilitate the attendance of employed students, it is expected that:

- First semester classes take place at the end of the day (every week day from 4 to 8 pm), and that in the second semester most courses are also scheduled at this time, although some electives may take place at other times.
- The course can be concluded in part-time regime (four years) completing half of the curricular units planned in each semester and support just half of the tuition fees per semester.



Program Structure

The Master's program lasts two academic years, the first year being devoted to the pedagogical component and the second year exclusively dedicated to the dissertation. The first semester consists of five compulsory Curricular Units designed to create a common basic formation for students from different backgrounds. The second semester includes two compulsory UCs (Seminar and Entrepreneurship) and four options chosen by the student from among a block of UCs that will define their specialization profile, (one of them can be chosen from a block offered by the different department of FCT NOVA).

In these options, students must choose at least one UC from the scientific areas of Computer Science and Informatics, and Mathematics and at most one UC, from a block of free electives, that includes courses from all master programs at FCT NOVA .

Estrutura Curricular

O Mestrado decorre em dois anos letivos, sendo o primeiro ano dedicado à componente letiva e o segundo ano exclusivamente dedicado à dissertação. O primeiro semestre consiste em cinco Unidades Curriculares (UCs) obrigatórias destinadas a criar formações de base comum para alunos de diferentes proveniências. O segundo semestre inclui duas UCs obrigatórias (Seminário e Empreendedorismo) e quatro opcionais escolhidas pelo aluno de entre um bloco de UCs que irão definir o seu perfil de especialização.

Nestas opções, os alunos devem escolher pelo menos uma UC das áreas científicas de Informática e Matemática e no máximo uma UC do bloco livre constituído por UCs de vários cursos de Mestrado da FCT NOVA.

NOME NAME	ÁREA AREA	ECTS ECTS	NOME NAME	ÁREA AREA	ECTS ECTS
1º SEMESTRE 1 ST SEMESTRE			3º/4º SEMESTRE 3 RD /4 TH SEMESTER		
Aprendizagem Automática Machine Learning	INF	6	Dissertação Master Thesis	INF	60
Estatística Multivariada Multivariate Statistics	MAT	6	BLOCO DE ESPECIALIZAÇÃO SPECIALIZATION TIER		
Estatística Numérica Computacional Computational Numerical Statistics	MAT	6	Algoritmos para Redes Complexas Algorithms for Complex Networks	INF	6
Recuperação de Informação Information Retrieval	INF	6	Análise de Grandes Grafos Analysis of Large Graphs	MAT	6
Sistemas para Processamento de Big Data Big Data Processing Systems	INF	6	Aprendizagem com Dados Não Estruturados Learning from Non Structured Data	INF / MAT	6
2º SEMESTRE 2 ND SEMESTRE			Decisão e Risco Decision and Risk	MAT	6
Empreendedorismo Entrepreneurship	EG	3	Métodos Bayesianos Bayesian Methods	MAT	6
Seminário Seminar	INF / MAT	3	Otimização Linear Linear Optimisation	MAT	6
Bloco de Especialização Specialization Tier	INF	6	Otimização Não-Linear Nonlinear Optimisation	MAT	6
Bloco de Especialização Specialization Tier	MAT	6	Processamento de Streams Stream Processing	INF	6
Bloco de Especialização Specialization Tier	INF / MAT	6	Prospecção e Análise de Dados Data Analytics and Mining	INF	6
Bloco Livre Free Tier	QQ	6	Visualização Interativa de Dados Interactive Data Visualisation	INF	6

Descrição das Unidades Curriculares

Curricular Units

Algoritmos para Redes Complexas

(Opcional, 6 ECTS, INF)

Nesta unidade curricular pretende-se identificar as noções fundamentais que caracterizam a estrutura de uma rede, para poder analisar redes concretas do mundo real, calculando as métricas apropriadas, comunidades, etc., e extrair informação/insights dos dados. Pretende-se ainda estudar alguns algoritmos eficientes usados na análise de redes complexas, compreender alguns modelos comportamentais de redes e simular um processo de propagação numa rede.

Análise de Grandes Grafos

(Opcional, 6 ECTS, MAT)

Nesta unidade curricular serão estudadas estruturas de redes e grafos de grandes dimensões, abordando-se questões como a conexidade, centralidade e comunidades, previsão de processos que poderão ocorrer em grafos grandes e previsão do crescimento de grandes grafos reais.

Aprendizagem Automática

(Obrigatória, 6 ECTS, INF)

A crescente abundância de informação torna cada vez mais importante a extração automática de conhecimento. Esta unidade curricular cobre os fundamentos da aprendizagem automática, incluindo diferentes modelos de classificação e regressão e o problema do sobre ajustamento na aprendizagem supervisionada, e ainda algoritmos de aglomeração e extração de atributos em aprendizagem não supervisionada.

Algorithms for Complex Networks

(Elective, 6 ECTS, INF)

In this course, we identify the fundamental notions that characterize the structure of a network, to be able to analyze concrete networks of the real world, calculating appropriate metrics, communities, etc., and extracting information / insights from the data. It is also intended to study some efficient algorithms used in the analysis of complex networks, to understand some network behavioral models and to simulate a propagation process in a network.

Analysis of Large Graphs

(Elective, 6 ECTS, MAT)

This curricular unit studies structures of large networks and graphs, addressing issues such as connectivity, centrality and communities; prediction of processes that may occur in large graphs; and prediction of the growth of large real graphs.

Machine Learning

(Mandatory, 6 ECTS, INF)

The increasing abundance of information makes automatic knowledge extraction increasingly important. This curricular unit covers the fundamentals of automated learning, including different classification and regression models, and the over-adjustment problem in supervised learning, and clustering and attribute extraction algorithms in unsupervised learning.

Learning from Non Structured Data

(Elective, 6 ECTS, INF/MAT)

Most digital information is not structured, requiring a laborious process to extract knowledge of video, sound, images or free text using traditional analytical techniques. This curricular unit presents recent deep learning techniques that allow direct dealing with this data, including deep neural networks, convolution and recurrent, deep generator models, pre-training and practical examples of application with the Tensorflow library.

Decision and Risk

(Elective, 6 ECTS, MAT)

This course addresses topics in decision making, game theory, negotiation models and strategic interaction between the players in the decision-making process. Different concepts and models of risk are studied.

Entrepreneurship

(Mandatory, 3 ECTS, EG)

This course is intended to motivate students for entrepreneurship and the need for technological innovation. It covers a list of topics and tools that are important for new venture creation as well as for the development of creative initiatives within existing enterprises.

Aprendizagem com Dados Não Estruturados

(Opcional, 6 ECTS, INF/MAT)

A maior parte da informação digital não está estruturada, exigindo um processo laborioso para extrair conhecimento de vídeo, som, imagens ou texto livre usando técnicas tradicionais de análise. Esta unidade curricular apresenta técnicas recentes de aprendizagem profunda (deep learning) que permitem lidar diretamente com estes dados, incluindo redes neuronais profundas, de convolução e recorrentes, modelos geradores profundos, pré-treino e exemplos práticos de aplicação com a biblioteca Tensorflow.

Decisão e Risco

(Opcional, 6 ECTS, MAT)

Nesta unidade curricular são abordadas questões relacionadas com a tomada de decisão, a teoria de jogos, modelos de negociação e interação estratégica entre os intervenientes no processo de tomada de decisão. São também estudados diferentes conceitos e modelos de risco.

Empreendedorismo

(Obrigatória, 3 ECTS, EG)

O curso pretende motivar os alunos para o empreendedorismo e para a necessidade da inovação tecnológica. O programa cobre vários tópicos que são importantes para a adoção de uma cultura aberta aos riscos suscitados em processos de criação de novos produtos ou atividades que exigem características empreendedoras.

Estatística Multivariada (Obrigatória, 6 ECTS, MAT)

Esta unidade curricular foca-se na análise de dados multivariados. Em particular são abordados aspetos relacionados com as principais distribuições estatísticas, inferência para vetores de médias e matrizes de covariância, técnicas para classificação de dados e de redução da dimensionalidade, e a aplicação de diferentes técnicas e métodos estatísticos na análise de (grandes volumes) dados multivariados.

Estatística Numérica Computacional (Obrigatória, 6 ECTS, MAT)

Nesta unidade curricular abordam-se conceitos relacionados com a inferência em dados de grande dimensão. É dada especial atenção aos métodos estatísticos, técnicas de amostragem, reamostragem e simulação.

Métodos Bayesianos (Opcional, 6 ECTS, MAT)

Esta unidade curricular cobre os modelos bayesianos para a análise de estruturas de dados muito complexas. São expostos, entre outros tópicos, os princípios que regem a inferência bayesiana, técnicas de incorporação do conhecimento existente a priori e a sua correspondente incerteza numa distribuição de probabilidade, e ainda a modelação hierárquica para representar e analisar sistemas complexos.

Multivariate Statistics (Mandatory, 6 ECTS, MAT)

This curricular unit focuses on the analysis of multivariate data. In particular, aspects related to the main statistical distributions, inference for means vectors and covariance matrices, techniques for data classification and dimensionality reduction, and the application of different techniques and statistical methods in the analysis of (large volumes) data multivariate.

Computational Numerical Statistics (Mandatory, 6 ECTS, MAT)

This course unit deals with concepts related to inference in large data, with a focus on statistical methods, sampling-resampling and simulation techniques.

Bayesian Methods (Elective, 6 ECTS, MAT)

The course introduces Bayesian models for the analysis of very complex data structures. The principles governing Bayesian inference, techniques for incorporating existing a priori knowledge and the corresponding uncertainty into a probability distribution, as well as hierarchical modelling to represent and analyse complex systems, will be exposed.

Linear Optimisation (Elective, 6 ECTS, MAT)

This course introduces the student to linear optimization. Topics related to modelling, solving and analysing solutions of linear, linear and integer linear programming problems with multiple objectives will be addressed.

Nonlinear Optimisation (Elective, 6 ECTS, MAT)

This course will focus on modelling and solving non-linear optimisation problems. It will cover topics that will allow the student to distinguish easy problems from those of difficult resolution and to use the main methods of nonlinear optimization with the knowledge of their numerical weaknesses. It will also analyse possible resolution approaches for optimization problems of this nature.

Stream Processing (Elective, 6 ECTS, INF)

In recent years there has been a huge increase in the amount of information that is continuously generated (for example, by financial systems, telecommunications networks and public services or sensor networks). This course aims to study the fundamentals, languages and systems for the construction of applications that process data streams, from generic real-time distributed stream processing systems to structured data models to deal with streams.

Otimização Linear (Opcional, 6 ECTS, MAT)

Esta unidade curricular introduz a otimização linear. São abordados tópicos relacionados com a modelação, resolução e análise das soluções de problemas de programação linear, linear inteira e linear com múltiplos objetivos.

Otimização Não-Linear (Opcional, 6 ECTS, MAT)

Esta unidade curricular foca a modelação e resolução de problemas de otimização não linear. Abordará tópicos que permitirão ao aluno distinguir os problemas fáceis dos de difícil resolução e utilizar os principais métodos de otimização não linear com o conhecimento das suas fragilidades numéricas. São também analisadas possíveis abordagens de resolução para problemas de otimização desta natureza.

Processamento de Streams (Opcional, 6 ECTS, INF)

Nos últimos anos tem havido um enorme aumento na quantidade de informação que é continuamente gerada (por exemplo, por sistemas financeiros, redes de telecomunicações e serviços públicos ou redes de sensores). Este curso tem como objetivo estudar os fundamentos, linguagens e sistemas para a construção de aplicações que processam fluxos de dados, desde sistemas genéricos de processamento de fluxos distribuídos em tempo real a modelos de dados estruturados para lidar com fluxos.

Prospecção e Análise de Dados (Opcional, 6 ECTS, INF)

Esta unidade curricular centra-se nas áreas de Análise de Dados e Text Mining. O módulo de Análise de Dados cobre tópicos de pré-processamento, redução de dimensionalidade, algoritmos de clustering particional e clustering espectral (crisp vs. fuzzy), validação e interpretação de clustering. O módulo de Text Mining foca a extração de Informação Relevante, a análise simbólica e estatística dos textos, a classificação de documentos, a construção automática dos seus descritores e a distribuição das palavras em contexto de Big Data.

Recuperação de Informação (Obrigatória, 6 ECTS, INF)

Como pesquisar informações médicas para ajudar a decidir o diagnóstico de um paciente? Como pesquisar todas as notícias relacionadas com um desastre natural? Este curso aborda esses problemas técnicos fornecendo aos alunos uma compreensão do design e da implementação dos mecanismos de pesquisa. Os alunos irão dominar conceitos fundamentais de Recuperação de Informação: representação de informação, indexação, consulta e classificação por relevância.

Seminário (Obrigatória, 3 ECTS, INF/MAT)

Nesta unidade curricular apresenta vários casos de estudo, em que as instituições (empresas, institutos públicos, centros de investigação, etc) obtêm e/ou administram grandes volumes de dados e pretendem retirar conhecimento

Data Analytics and Mining (Elective, 6 ECTS, INF)

This course focusses on Data Analysis and Text Mining. The Data Analysis module covers topics of pre-processing, dimensionality reduction, partitioning clustering algorithms and spectral clustering (crisp vs. fuzzy), validation and interpretation of clustering. The Text Mining module covers extraction of Relevant Information, the symbolic and statistical analysis of the texts, the classification of documents, the automatic construction of its descriptors and the distribution of words in the context of Big Data.

Information Retrieval (Mandatory, 6 ECTS, INF)

How to research medical information to help decide a patient's diagnosis? How to search for all news related to a natural disaster? This course addresses these technical issues by providing students with an understanding of the design and implementation of search engines. Students will master fundamental concepts of Information Retrieval: information representation, indexing, query and classification by relevance.

Seminar (Mandatory, 3 ECTS, INF/MAT)

In this course students are exposed to various case studies, in which several institutions (companies, public institutions, research centres, etc.) obtain and / or manage large volumes of data and want to take advantage

from the knowledge and useful information extracted from these data. Students should examine these case studies, presented by representatives of these institutions, in order to analyse the possibilities of extraction and use of this knowledge to improve the decisions to be taken.

Big Data Processing Systems (Mandatory, 6 ECTS, INF)

This course unit offers a general perspective of large-scale data processing. The main existing solutions for the storage and distributed processing of data are reviewed. Regarding the processing tools studied, special emphasis is given to the supported programming models. In the hands-on component of the UC, a project is developed that addresses a non-trivial problem, using the Hadoop ecosystem suite of tools, where the Apache Spark system stands out.

Interactive Data Visualisation (Elective, 6 ECTS, INF)

This curricular unit presents the theoretical and practical bases for the conception, use and evaluation of modern systems for interactive visualization of data. In addition to the fundamentals of Visualization, the main techniques of visualization of multivariate data, geo-spatial data, and time-dependent data are presented.

e informação útil desses dados. Os alunos deverão analisar estes casos de estudo, apresentados por responsáveis dessas instituições, de forma a analisar as possibilidades de extração desse conhecimento, bem como da sua utilização para o melhoramento das decisões a serem tomadas.

Sistemas para Processamento de Big Data (Obrigatória, 6 ECTS, INF)

Esta unidade curricular oferece uma perspetiva geral do processamento de dados em grande escala. São revistas as principais soluções atualmente existentes para o armazenamento e processamento distribuído de dados. No que se refere às ferramentas de processamento estudadas, é dada especial ênfase aos modelos de programação suportados. Na componente prática da UC, é desenvolvido um projeto que aborda um problema não trivial, usando a suite de ferramentas do ecossistema Hadoop, onde se destaca o sistema Apache Spark.

Visualização Interativa de Dados (Opcional, 6 ECTS, INF)

Esta unidade curricular apresenta as bases teóricas e práticas para a conceção, uso e avaliação de sistemas modernos para visualização interativa de dados. Além dos fundamentos da Visualização são apresentadas as principais técnicas de visualização de dados multivariados, dados geo-espaciais, e dados dependentes do tempo.

4

Parcerias Partnerships

O Mestrado em Análise e Engenharia de Big Data conta com a colaboração de um conjunto de empresas e instituições, que, em parceria com o curso, fornecem casos de estudo e oportunidades de desenvolvimento de projetos conjuntos que podem servir de base à elaboração de Dissertações.

The Master in Big Data Analytics and Engineering counts on the collaboration of a group of companies and institutions that, in partnership with the course, provide case studies and development opportunities for joint collaborative projects that can serve as a basis for the elaboration of Dissertations.



alTRAN



Deloitte.



NEWVISION®





Departamento de Informática
Departamento de Matemática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

Campus da Caparica
2829-516 Caparica

(+351) 212 948 536
maebd.coordenador@fct.unl.pt
www.di.fct.unl.pt