Project Title	**ARRANGE-ICT - PARTNERSHIP FOR ADDRESSING MEGATRENDS IN ICT**
Project Acronym	**ARRANGE-ICT**
Instrument	**ERASMUS+**

Output	Output 5 – Modernised Curricula & Short Programmes, capable to efficiently address cutting-edge market demands
Lead Author (Org)	George Pallis, Marios Dikaiakos (UCY)
Contributing Author(s) (Org)	Dimosthenis Stephanidis (UCY)
Due Date	
Actual Date of Submission	
Version	V1.0

Dissemination Level

X	PU: Public (*on-line platform)
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	01.06.2020	UCY	Initial version

Table of Contents

1. Introduction

ARRANGE-ICT focuses on identifying and recording megatrends developments as raised from market innovators in order to help HEIs follow the rapid pace of emerging industries. The major priority of ARRANGE-ICT is to enhance the quality and relevance of student's knowledge and skills, as the project aims to support Higher Education Institutes (HEIs). This priority is relevant and timely as, the widespread digitalization and the resulting availability of big data, combined with accelerating advances in Artificial Intelligence and Machine Learning, bring significant transformations to economic and social activities, disrupting the landscape of employment and the skills sought to sustain economic activity[1,2,3].

A first key trend is the strategic importance of **digital data** in a wide range of economic and scientific activities[4,5]. As massive amounts of data and ways of analyzing them become available, more aspects of the economy, society, and daily life become dependent on data. The availability of and dependence on data make work across nearly all domains increasingly data driven and create more opportunities and needs for data-driven problem solving and decision making. This trend exposes an ominous *quantitative skills gap* in the workforce, as more employees are increasingly asked to engage in and collaborate on activities pertinent to the PPDAC problem-solving cycle (Problem understanding and definition; Planning of data recording and collection; Data collection, management and cleaning; Analysis of data and hypothesis generation, and Conclusions' derivation and interpretation of data)[6]. Engaging in these activities requires an understanding of and experience with a variety of advanced skills and tools in Information Technology, Statistics, and the application thereof in different industrial and scientific domains. However, as the abundance of data and the proliferation of powerful data-analytics tools expands our reliance on data-driven approaches, their adoption often comes with profound implications, which raise unintended but serious ethical concerns or harmful consequences[7]. This risk demonstrates the necessity to involve *qualitative skills* in problem-solving and decision

[1] *"L 'Avenir du Travail."* Jacques Attali. 2007. Fayard.
[2] *"The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies."* Erik Brynjolfsson and Andrew McAfee. 2014. Norton.
[3] "*The Age of Analytics: Competing in a Data-driven World."* McKinsey & Company, December 2016.
[4] "The New Digital Age." Eric Schmidt and Jared Cohen. 2013. Knopf.
[5] *"The world's most valuable resource is no longer oil, but data."* Economist, May 6, 2017.
[6] *"The Art of Statistics. Learning from Data.*" David Spiegelhalter. 2019. Penguin.
[7] "*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy."* Cathy O'Neil. 2016. Crown.

making for the proper contextualization and thoughtful interpretation of quantitative results[8,9].

Automation is another major trend arising from the advances in digitalization and AI. Automation brings significant productivity gains but will also lead to a massive replacement of humans by robots and intelligent software in many tasks, both physical and cognitive. Consequently, and in contrast to previous industrial revolutions where new technology was associated with long-term job creation and wage growth, several studies predict that the 4th Industrial Revolution will have a large negative effect on jobs, creating significant social problems and exacerbating economic inequality within and between countries. Estimates predict that by 2030, employers will need 20–25% fewer employees, with 30–50% of work activities becoming technically capable of being automated[10]. Automation-induced unemployment is expected to hit both white- and blue-collar workers. At a domestic level, this will exert further pressure to workers who will see their bargaining power, labor rights, wages and job security diminish. At the international front, the capability to manufacture in almost workerless factories at countries with high labor costs could slash the competitive advantage of countries with lower labor costs to attract foreign investment, create jobs, and pursue an export-led path to economic growth[11]. The social and economic transformations brought by automation are, in turn, changing the demand for skills. With manufacturing and certain low-skill tasks increasingly becoming automated, the need for routine cognitive and craft skills is declining, while the demand for information-processing and other high-level cognitive and interpersonal skills is growing. Therefore, and in addition to mastering occupation-specific and IT skills, workers in the 21st century must also have a stock of various "generic" skills, including interpersonal communication, self-management, and the ability to learn, to help them weather the uncertainties of a rapidly changing labor market[12].

---

8 *"Big Data needs Thick Data."* Tricia Wang, May 13, 2013. http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/

9 *"Spreadsheets can't predict behaviour."* Tassos Stassopoulos. Trinetra, July 20, 2020. https://www.trinetra-im.com/post/spreadsheets-cant-predict-behaviour

10 *"Technology and Jobs. Coming to an Office Near You."* Economist, Jan. 18, 2014.

11 *"Universities as Knowledge Platforms in a Period of Disruption."* Marios D. Dikaiakos. Invited talk *Conference on The Role of the University in the Crisis of Forced Displacement: Ethics, Innovation, and Immersive Learning*, Boston University, September 26, 2019. https://medium.com/swlh/universities-as-knowledge-platforms-in-a-period-of-disruption-c68062edf7f1

12 *"OECD Skills Outlook 2019: Thriving in a Digital World."* OECD. OECD Publishing, Paris, 2019. https://doi.org/10.1787/df80bc12-en

Another important trend that arises in late-20th / early-21st century is the effort of numerous countries around the world to promote activities in **innovation-driven entrepreneurship**. Such efforts are typically fostered by the aspiration to adopt Silicon Valley's culture and replicate its economic impact. This aspiration seems reasonable: numerous economic studies have demonstrated that technological change drives up income levels, there is a strong relationship between high levels of intellectual property creation and GDP growth, and there is a positive impact of innovation on business productivity and performance[13]. Indeed, as the rate of scientific and technological progress accelerates, and as their impact on economic and social activities expands, the capability of people to swiftly adapt to and take advantage of new conditions and developments by engaging in innovative and creative entrepreneurship becomes a strategic priority for nations that wish to encourage and enable the imagination and extras of their people, helping them develop creative capacities, abilities to spark new ideas, and to start up new industries[14]. Nevertheless, the convergence of the unique conditions that led to the creation and successes of Silicon Valley's entrepreneurial ecosystem cannot easily be replicated at a different place and at a different time. Training, however, in innovative thinking and in entrepreneurial skills can help people from different backgrounds to identify and pursue opportunities and take risks in different contexts, leading to new business ventures or social enterprises with a global or local impact.

Within this broad context, Data Science is quickly becoming a field of central importance to the strategy of modern organizations and has become megatrend in ICT. Today, data science has become the holy grail of employability skills. The importance and opportunities inherent in data science are clear (see http://cra.org/data-science/). As stated in a recent article at CACM[15], "*Higher-education institutions across U.S. recognize that data science is a critical skill for 21st-century research and a 21st-century workforce. In higher education, data science curricula have two audiences: new professionals in data science, and scientists and*

[13] *"Knowledge, Networks and Nations: Global scientific collaboration in the 21st century."* Royal Society, March 2011.
[14] *"Thank you for Being Late."* Thomas Friedman. 2016. Farrar, Straus and Giroux.
[15] *"Realizing the potential of data science."* Francine Berman, Rob Rutenbar, Brent Hailpern, Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Margaret Martonosi, Padma Raghavan, Victoria Stodden, and Alexander S. Szalay. 2018. Commun. ACM 61, 4 (April 2018), 67–72. DOI:https://doi.org/10.1145/3188721

professionals who need data science skills to contribute to other ICT fields". In 2011, McKinsey Global Institute called Big Data as the Next Frontier. From then to now, "*Data Science has evolved to form an integral part of digital transformation and technological innovation for the next world. This implies that not only is the demand for data scientists is growing everyday but also they receive high salaries due to high market demand. Every industry from the fields of business, finance, government, healthcare, social networking, and technology, are looking for people with data science skills*" (see https://www.analyticsinsight.net/top-10-universities-in-usa-offering-ph-d-in-data-science/).

In this deliverable, we design a modernized curriculum in Data Science. This curriculum is expected to be delivered by the University of Cyprus. The methodology that we followed to design the modernized curriculum in Data Science can be summarized as follows:

- Identifying the skills that students should develop while pursuing a post-graduate program in Data Science
- Studying successful related curricula in highly respected academic institutions all over the world (e.g., Georgia Tech and New York University in the USA, the University of Amsterdam and Bocconi University in Europe).
- Designing the modernized curriculum in Data Science
- Implementing the modernized curriculum in Data Science in University of Cyprus

The following Sections present in detail the methodology that has been followed.

2. Identifying the skills

Data science is inherently interdisciplinary. Working with data requires the knowledge and understanding of a variety of skills and concepts associated with the fields of Statistics, Computer Science and Management Science. A graduate with a Data Science degree should be prepared to interact with data at all stages of an investigation and will be expected to work within a team environment.

Specifically, the need for an inter-departmental program in Data Science springs exactly from this need for scientists and managers of different specializations to think across disciplines:

- Business analytics specialists must understand the underlying principles of the tools they employ (need for a statistics perspective) and often find themselves dealing with practical implementation issues, especially when dealing with the increasing volumes of data that modern organizations generate (need for computational science skills).
- Computational data science specialists must, at the end of the day, care about the utilitarian -- economic or otherwise -- impact of their work and be able to clearly communicate its value to all relevant stakeholders (need for a business analytics perspective). At the same time these specialists are also not immune from misusing the tools of their trade if they do not sufficiently understand the fundamental science and underlying mathematics of the field (need for a statistics perspective).
- Statistics data science specialists are often tasked with solving complex and multifaceted theoretical problems that often require the careful engineering of data analytics stacks capable of processing data of very high throughput (need for a computational science perspective), while they too must often deliver results within a framework of organizational priorities that rewards business value and often places utility and other practical considerations above precision or methodological purity (need for a business analytics perspective).

The Postgraduate Data Science program aspires to:

- Offer students the opportunity to acquire deep knowledge, hands-on experience, and research expertise in one or more fields of Data Science.
- Prepare graduates able to pursue careers in positions of responsibility in either academia or industry, where they will effectively drive the development and application of new methods and ideas.
- Offer the students with education in the widest sense of the term, and cultivate the desire for continuous learning, which, in turn, leads to maturity and develops the facilities for independent and critical thinking.
- Help its graduates to acquire a deep understanding of Data Science, both as a science, and in terms of its more general applications and effects on society.
- Prepare its graduates so that, whichever career path they choose, their studies will have given them the necessary grounding to keep abreast of the incredibly rapid scientific and technological developments in Computer Science.

- Provide transferable skills in scientific and technical communication (both written and spoken), gathering and collecting actionable information, collaborative work, creativity and innovative thinking.

3. Studying successful related curricula

Data Science programs have exploded in academics as university administrators have rushed to meet the demand. The website http://datascience.community/colleges currently lists 620 programs in Data Science, analytics and related fields at over 200 universities around the world. The vast majority of these are master's degree and certificate programs offered both traditionally and online. It is thus not surprising that inter-departmental programs in Data Science have been introduced by highly respected academic institutions all over the world (e.g., Georgia Tech and New York University in the USA, the University of Amsterdam and Bocconi University in Europe). These programs retain an overall structure that closely matches the structure proposed herein, which is in light of the fact that the fundamental logic of the need to bring statistics, computer science, and business perspectives together is common across all such programs.

A successful postgraduate Data Science program should provide the following assets:

- Comprehensive multidisciplinary program; Students will become well-versed in a variety of tools, perspectives and approaches, to be able to identify the most appropriate methods and models to use to solve each specific case. Compared to pure masters programs in one of the three disciplines, graduates of the Data Science master will benefit from a foundation and functional knowledge in all three disciplines (Computer Science, Mathematics and Statistics, Business and Public Administration) and this will put them in an advantageous position for employment in industry. At the same time, they will be able to pursue their preferred discipline through the choice of one of the three offered tracks (Computer Science Track/ Statistics Track/Business Analytics Track).
- Students will gain the ability to use different types of data in order to make complex predictions and computations at scale. In light of the continuous growth of data all around the globe, the question of how we can use data to gain valuable insights is more important than ever. How can we extract relevant information from massive amounts of data? In which way can computers learn

from experience to make intelligent decisions? These questions are key to the specialized Master's program in data science.

- Students will have the opportunity to be exposed to other key aspects of relevance to the processing and monetization of data, such as privacy, ethics, and innovative business models.
- Students will have first-hand access to cutting-edge research and to exciting industrial internship and career opportunities in various fields such as finance, health care, insurance, pharmaceutical, telecoms, tourism, travel and transportation etc.

4. Designing the modernized curriculum in Data Science

The specialized curriculum in Data Science, offered in collaboration with the **Department of Computer Science**, the **Department of Mathematics and Statistics** as well as the **Department of Business and Public Administration**, provides a high quality education. It is important to note that existing Data science programs are being sited in departments and schools of computer science, information science, mathematics and statistics, and management. Many of the most successful, particularly at the undergraduate level, represent university-wide coalitions frequently sponsored by interdisciplinary institutes, rather than by a particular department or school.

The modernized curriculum in Data Science will be an 1,5 year postgraduate program fully taught in English. Research in the field of data science requires solid skills in statistics, managing and storing massive amounts of data, business analytics, as well as the ability to develop efficient algorithms for data analysis. These techniques are employed in complex applications in business, engineering and science. Part of the program is the **Capstone project in Data Science**, where students tackle specific and practical problems of interdisciplinary applications. In this course students engage in all tasks - from the process modelling to the implementation and validation of data science techniques. A key aspect of this course will be the industrial engagement with industry and public sector participation.

The program has been designed to offer 3 tracks (Computational Science Track / Statistics Track / Business Analytics Track). The first two semesters will be core courses, while in the third semester the students will select 3 selective courses. The capstone project will begin in the summer semester (after the end of the second semester).

What ties these programs, and the modernized curriculum together, is a set of common core courses, and an eventual meeting of all specialties during the final capstone project. This common core of courses is a set of graduate level courses whose purpose is to level the field across students of different backgrounds and to build a solid core on which the individual tracks are built.

Depending on the background of the students, it is natural that some students will find some courses more challenging than others, while their colleagues with different undergraduate backgrounds will rate course difficulty in reverse. This is to be expected by the very design of the program, and is even true for the multidisciplinary programs offered internationally, such as those mentioned above.

Each of the common core courses has been designed around the abilities of an attentive student with a different undergraduate background than the course's primary field. The faculty of the common core courses have put together additional bibliography as well as a set of resources for students who may find their courses more challenging that intended, and the faculty will hold two extra office hours per week during the semester, on top of the regular three office hours that all other courses observe.

Target audience: new professionals in data science, and scientists and professionals who need data science skills to contribute to other fields. The primary focus for this programme is to support local industry and ensure staff have the relevant skills to apply data science in their workplace. In this context, we plan to offer a tiered approach to education and training. This will include individual modules (called *short programmes*) that can be taken separately and provide certificates.

Language: The programme will be in English.

ECTS: 90

Educational Prerequisites. Successful applicants to the master programme of Data Science will come from many different undergraduate backgrounds, including degrees in Statistics, Computer Science, Mathematics, Engineering, Economics, Business, Biology, Physics and Psychology. Regardless of degree, we will require specific and substantial knowledge of certain mathematical competencies, and some training in programming and basic computer science. To be considered for the program, students will be required to have completed the following (or equivalents):

- Fundamental Mathematical concepts (eg basic calculus, linear algebra, etc)
- Intro to Computer Science (or an equivalent programming course): We have no set requirements as regards specific languages, but we generally expect academic and/or professional experience with programming languages such as Python, Scalla, R, etc.
- One of Probability, Statistics, engineering, or econometrics course.

Intended learning outcomes:

By the time they graduate, the students are expected to:

- Acquire deep knowledge and research expertise in one or more fields of Data Science through the classes outlined in Annex 2.
- Master powerful tools, for addressing a wide range of topics.
- Acquire statistical skills at an advanced level.
- Obtain familiarity with basic concepts in other Natural and/or Social Sciences, pertinent to data-driven discovery.
- Demonstrate in-depth understanding of a breadth of disciplines in Data Science, and be largely familiar with the dominant research directions and cutting-edge problems.
- Exhibit versatility and innovative thinking in addressing and managing open questions in a variety of contexts, as an essential asset for careers in research, industry, commerce, education and the public sector.
- Develop transferable skills such as: oral and written scientific communication, near fluent use of scientific English, use of information/communication technology, organization and planning of group work.

- Acquire experience of independent work, ideally so in the context of class research projects.
- Identify and assess the needs of an organization for a data science task
- Collect and manage data to devise solutions to data science tasks.
- Interpret data science analysis outcomes.
- Communicate data science-related information effectively in various formats to appropriate audiences.
- Value and safeguard the ethical use of data in all aspects of their profession.
- Transform findings from data resources into actionable business strategies.

5. Implementing the modernized curriculum in Data Science in University of Cyprus

The following paragraphs present the course plan with a short description of the courses of the Data Science Curriculum that is planned to be implemented by University of Cyprus (expected to start on September 2021).

The faculty members that have designed the Data Science programme in University of Cyprus are: Sergios Agapiou (Assistant Professor, Mathematics and Statistics Dpt.), Anastassia Baxevani (Associate Professor, Mathematics and Statistics Dpt.), Marios Dikaiakos (Professor, Computer Science Dpt.), Konstantinos Fokianos (Professor, Mathematics and Statistics Dpt.), Panos Markopoulos (Assistant Professor, Business and Administration Dpt.), Christos Nicolaides (Lecturer, Business and Administration Dpt.), George Pallis (Associate Professor, Computer Science Dpt.).

The programme has been submitted for accreditation by the Cyprus Agency of Quality Assurance and Accreditation in Higher Education (CYQAA), which is the competent Authority responsible for ensuring the quality of higher education in Cyprus and for the support of the processes provided by the relevant Legislation, for the continuous improvement and upgrading of higher education institutions and their programs of study.

SEMESTER COURSE PLAN

EPL: COMPUTER SCIENCE COURSES

MAS: STATISTICS COURSES

BUS: MANAGEMENT/BUSINESS COURSES

Year 1 – Fall

Course Title	ECTS
TOTAL ECTS	**28**
EPL: Introduction to Data Science and Analytics This course will examine how data analysis technologies can be used to improve decision-making. The aim is to study the fundamental principles and techniques of data science, and we will examine real- world examples and cases to place data science techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science. In addition, this course will work hands-on with the Python programming language and its associated data analysis libraries.	8
MAS: Probability and Statistics for Data Science This is a theoretical course covering fundamentals topics of probability and statistics in the context of data science with its inherent challenges. This course will start with a review of fundamental probability, covering topics like random variables, their distribution functions, expected values, conditioning on certain events and independence. The students will be acquainted with certain families of probability distributions and then will learn how to estimate certain quantities of interest from observations. A range of properties of estimators will be studied, including sufficiency, unbiasedness and consistency which enable the evaluation of their	8

quality. The students will also learn how to introduce different types of hypotheses, how to construct tests for their hypotheses, as well as how to compare between tests and how to construct confidence intervals for their estimators.

MAS: Simulations and Data Analysis 8

The students will be introduced to the R programming language, a programming language that was specifically developed for analyzing data, and is today widely used in most organizations that conduct data analysis. The students will learn how to explore datasets in R, using basic visualization tools and summary statistics, how to run different kinds of regressions and analyses, and how to perform statistical inference in practice, for example how to test certain hypotheses regarding the data or how to compute confidence intervals for quantities of interest. The students will also learn how to use R in order to conduct simulations, an extremely useful tool that can fulfill a wide range of analytical tasks. Simulation techniques covered will include Monte Carlo, importance sampling and rejection sampling. Finally, the students will learn how to estimate the precision of computed sample statistics using resampling methods. The course uses a hands on approach, with nearly half the work done in the lab.

One Elective Course (See list of Elective Courses, below) 4

Note: Two Elective courses (total 8 ECTS) can be replaced by one 8 ECTS course from the tracks (Computer Science Track, Statistics Track, Business Analytics Track)

Year 1 – Spring

Course Title	Credits
TOTAL CREDITS	**28**
MAS: Statistical Learning Students will acquire the knowledge to conduct statistical analysis on a variety of data sets using a wide range of modern computerized methods. The students will learn how to recognize which tools are needed to analyze different types of datasets, how to apply these tools in each case, and how to employ diagnostics to assess the quality of their results. They will learn about statistical models, their complexity and their relative benefits depending on the available data. Some of the tools that the students will come to learn well include linear simple and multiple regression, nearest neighbors methods, shrinkage methods (ridge, lasso), dimension reduction methods (principal components), logistic regression, linear discriminant analysis, tree-based methods and model selection algorithms with criterion or by resampling techniques. The focus of the course will be less on theory and more on providing the students with as much intuition as possible and acquantining them with as many methods as possible. The course will make substantial use of the R statistical programming language and its libraries.	8
EPL: Big Data Management Big Data requires the storage, organization, and processing of data at a scale and efficiency that go well beyond the capabilities of conventional information technologies. In this course, we will study the state of the art in big data management: we will learn about algorithms, techniques and	8

tools needed to support big data processing. In addition, this course will examine real applications that require massive data analysis and how they can be implemented on Big Data platforms. The course will consist of lectures based both on textbook material and scientific papers. It will also include programming assignments that will provide students with hands-on experience on building data-intensive applications using existing Big Data platforms.

BUS: Business Analytics Applications 8

This course presents knowledge and skills for applying business analytics to managerial decision-making in corporate environments. Topics include descriptive analytics (techniques for categorizing, characterizing, consolidation, and classifying data for conversion into useful information for the purposes of understanding and analyzing business performance), predictive analytics (techniques for detection of hidden patterns in large quantities of data to segment and group data into coherent sets in order to predict behavior and trends), prescriptive analytics (techniques for identification of best alternatives for maximizing or minimizing business objectives). Students will learn how to use data effectively to drive rapid, precise, and profitable analytics-based decisions. The framework of using interlinked data-inputs, analytics models, and decision-support tools will be applied within a proprietary business analytics shell and demonstrated with examples from different functional areas of the enterprise.

One Elective Course (See list of Elective Courses, below) 4

Note: Two Elective courses (total 8 ECTS) can be replaced by one 8 ECTS course from the tracks (Computer Science Track, Statistics Track, Business Analytics Track)

Year 1 – Summer

Course Title	Credits
TOTAL CREDITS	**5**
Capstone Project in Data Science (1st Phase)	**5**
The purpose of the Capstone Project is for the students to apply theoretical knowledge acquired during the Data Science program to a project involving actual data in a realistic setting. During the project, students engage in the entire process of solving a real-world data science project: from collecting and processing actual data, to applying suitable and appropriate analytic methods to the problem. Both the problem statements for the project assignments and the datasets originate from real-world domains similar to those that students might typically encounter within industry, government, non-governmental organizations (NGOs), or academic research. Within this course, we will promote placement and secondment opportunities in the local industry.	

Year 2 – Fall

Course Title	Credits
TOTAL CREDITS	**29**
Computer Science Track/ Statistics Track/Business Analytics Track Course	8
Computer Science Track/ Statistics Track/Business Analytics Track Course	8

Computer Science Track/ Statistics Track/Business Analytics Track Course	8
Capstone Project in Data Science (2nd Phase)	5

The purpose of the Capstone Project is for the students to apply theoretical knowledge acquired during the Data Science program to a project involving actual data in a realistic setting. During the project, students engage in the entire process of solving a real-world data science project: from collecting and processing actual data, to applying suitable and appropriate analytic methods to the problem. Both the problem statements for the project assignments and the datasets originate from real-world domains similar to those that students might typically encounter within industry, government, non-governmental organizations (NGOs), or academic research. Within this course, we will promote placement and secondment opportunities in the local industry.

Computer Science Track

EPL: Data Management (8 ECTS)

This course broadly introduces database systems, including the relational data model, query languages, database design, index and file structures, query processing and optimization, concurrency and recovery, transaction management and database design. Students acquire hands-on experience in working with database systems and in building web-accessible database applications.

EPL: Natural Language Processing (8 ECTS)

How should human languages be understood and analyzed? This course will examine modern computational approaches based on representation learning for understanding, processing and using human language. These include neural network-based deep learning methods and vector-space models of word meaning, and

together will give you the tools to build state-of-the art models for hard language understanding tasks like translation.

EPL: Information Retrieval (8 ECTS)

The objective of this course is to examine the main computer science principles that lie behind Google and other search engines. To this end, we will provide an introduction to Information Retrieval (IR), which is described as the science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational stand-alone or hypertextually-networked such as the World Wide Web.

EPL: Deep Learning (8 ECTS)

In this course, students will learn the foundations of Deep Learning, understand how to build neural networks, and learn how to lead successful machine learning projects. Students will learn about Convolutional networks, RNNs, LSTM, Adam, Dropout, BatchNorm, Xavier/He initialization, and more.

EPL: Web Analytics and Mining (8 ECTS)

The Web Analytics and Mining course covers the areas of web analytics, text mining, web mining, and practical application domains. The web analytics part of the course studies the metrics of web sites, their content, user behavior, and reporting. Google analytics tool is used for collection of web site data and doing the analysis. The text mining module covers the analysis of text including content extraction, string matching, clustering, classification, and recommendation systems. The web mining module studies how web crawlers process and index the content of web sites, how search works, and how results are ranked. Application areas mining the social web and game metrics will be extensively investigated.

EPL: Cloud Computing (8 ECTS)

Cloud Computing is a large-scale distributed computing paradigm which has become a driving force for information technology over the past several years. The exponential growth data size in scientific instrumentation/simulation and social media has triggered the wider use of cloud computing services. This course covers topics and technologies related to Cloud Computing and their practical implementations. This course focuses on learning emerging issues related to Cloud computing technology. The objectives are: Understand various basic concepts related to cloud computing technologies; Understand the architecture and concept of different cloud models: IaaS, PaaS, SaaS; Understand big data analysis tools and techniques; Understand the underlying principle of cloud virtualization, cloud storage, data management and data visualization; Understand different cloud programming platforms and tools Be familiar with cloud programming using Google's 'Go' programming language; Have details knowledge on reading and writing in cloud storage; Be familiar with application development and deployment using cloud platforms; Create application by utilizing cloud platforms such as Google app Engine and Amazon Web Services (AWS); Learn to develop scalable applications using AWS features; Learn basic concepts of MapReduce programming models for big data analysis on cloud.

EPL: Data Visualisation (8 ECTS)

Data visualisation is an important visual method for effective communication and analysing large datasets. Through data visualisations we are able to draw conclusions from data that are sometimes not immediately obvious and interact with the data in an entirely different way. This course will provide students with an informative introduction to the methods, tools and processes involved in visualising big data. Topics will include: Introduction to visualisation; Information visualisation; Scientific visualisation; Visualisation tools; Design approaches for visualisation; Visualisation for communication. We will use a variety of tools so that students become comfortable engaging with different software and confident trialing new packages to find those that best meet your needs. This includes R, Tableau and D3.js. By the end of the semester, we aim for students to feel comfortable designing and developing visual stories with data.

BUS: Information Networks (8 ECTS)

This course focuses on how the social, economic, technological and natural systems are connected, and how the study of *networks* sheds light on these connections. Topics include: how to model the formation of social and economic networks; understand and measure certain patterns of real world networks; identify, quantify and model how opinions, fads, political movements and diseases spread through interconnected systems and measure the robustness and fragility of them. We will bring together models and techniques from economics, sociology, math, physics, statistics and computer science to answer these questions.

Statistics Track

MAS: Bayesian Statistics (8 ECTS)

Subjective probability, Bayes rule, prior and posterior distributions, conjugate and non-informative priors, pointwise estimation and credible intervals, hypothesis testing, introduction to Bayesian decision analysis, introduction to empirical Bayes analysis, introduction to Markov chain Monte Carlo techniques.

MAS: Survey Sampling (8 ECTS)

Survey design, sampling and nonsampling errors, simple random sampling, stratified sampling, systematic sampling, cluster sampling, ratio estimators, regression estimators, determination of optimal sample size, bias in survey sampling, modern techniques of survey sampling.

MAS: Multivariate Analysis (8 ECTS)

Random vectors, measures of center and variation in multivariate moments. Multivariate normal distribution. Tests for normality. Estimation of the mean vector and the variance analysis, independence, multivariate –covariance matrix. Wishart and Hotelling distributions. Statistical inference. Union – Intersection Test. Confidence regions. Multivariate analysis of variance and multivariate regression analysis. Least squares method and Wilks distribution. Analysis of covariance. Principal components, Factor analysis, Discriminant analysis, Cluster analysis.

MAS: Computational Statistics (8 ECTS)

Numerical linear algebra: Multiple regression, Cholesky decomposition, diagnostics and collinearity, principal components and eigenvalue problems. Nonlinear statistical methods: Maximum likelihood estimation, Newton-Raphson and related methods, multivariate data and the Newton Raphson method, optimization techniques (unconditional and under constraints) EM algorithm. Numerical Integration and Approximation: Newton-Coates method, spline interpolation, Monte Carlo integration, general approximation methods. Probability Density Estimation: Histogram, linear and non linear smoothing, splines. Bootstrap.

MAS: Time Series Analysis (8 ECTS)

Stochastic processes, weak and strong stationarity. Trend and seasonal behavior of time series. Sample autocorrelation function and sample partial autocorrelation function. Prediction. Parametric families of stochastic processes. ARMA, ARIMA and SARIMA models. Properties, estimation and examples. ARCH and GARCH processes, properties of estimators and examples.

Business Analytics Track

BUS: Managing Business Processes with Information Systems & Analytics (8 ECTS)

This course provides students the key tools to analyze and improve business processes in organizations, with an emphasis on the service sector. This is achieved by bringing together key ideas from the fields of information systems, business analytics, and business process design and management. The course introduces the fundamental types of information systems, including enterprise-wide systems (ERP, SCM, CRM), and the basic principles of supporting business strategy with Information Systems. The students will learn how to use information systems to support their organization's business processes, and how to use business analytics and business process modeling techniques to inform key decisions during Business Process Re-engineering. The students will be introduced to different business analytics systems in fields such as marketing, retail, supply-chain management, e-commerce, etc. and

will learn how to measure business process performance through appropriate metrics and frameworks (e.g. the Balanced Scorecard approach)

BUS: Quantitative and Qualitative Decision-Making (8 ECTS)

This course explores decision making and policy formulation in organizations. Includes goal setting and the planning process, rational models of decision making, effective combination of qualitative and quantitative data (e.g. triangulation, complementarity etc.) with respect to the goal set, evaluation of alternatives, prediction of outcomes, cost-benefit analysis, decision trees, uncertainty and risk assessment, and procedures for evaluation of outcomes.

BUS: Financial Concepts (8 ECTS)

Introduction to the concepts, methods and problems of accounting and financial analysis. Includes accounting principles, measurement and disclosure issues, financial statement analysis, time value of money, cash flow projection and analysis, capital budgeting and project evaluation, bond and equity valuation, cost of capital and capital structure.

BUS: Web Analytics for Business (8 ECTS)

Explore web analytics, text mining, web mining, and practical application domains. The web analytics part of the course studies the metrics of websites, their content, user behavior, and reporting. The Google analytics tool is used for collection of website data and doing the analysis. The text mining module covers the analysis of text including content extraction, string matching, clustering, classification, and recommendation systems. The web mining module presents how web crawlers process and index the content of web sites, how search works, and how results are ranked. Application areas mining the social web and game metrics will be extensively investigated.

BUS: Data Mining for Business Analytics (8 ECTS)

Enterprises, organizations and individuals are creating, collecting, and using massive amount of structured and unstructured data with the goal to convert the information into knowledge, to improve the quality and the efficiency of their decision-making process, and to better position themselves to the highly competitive marketplace. Data mining is the process of finding, extracting, visualizing and reporting useful information and insights from both small and large datasets with the help of sophisticated data analysis methods. It is part of the business analytics, which refers to the process of leveraging different forms of analytical techniques to achieve desired business outcomes through requiring business relevancy, actionable insight, performance management, and value management. The students in this course will study the fundamental principles and techniques of data mining. They will learn how to apply advanced models and software applications for data mining. Finally, students will learn how to examine the overall business process of an organization or a project with the goal to understand (i) the business context where hidden internal and external value is to be identified and captured, and (ii) exactly what the selected data mining method does.

BUS: Information Networks (8 ECTS)

This course focuses on how the social, economic, technological and natural systems are connected, and how the study of *networks* sheds light on these connections. Topics include: how to model the formation of social and economic networks; understand and measure certain patterns of real world networks; identify, quantify and model how opinions, fads, political movements and diseases spread through interconnected systems and measure the robustness and fragility of them. We will bring together models and techniques from economics, sociology, math, physics, statistics and computer science to answer these questions.

BUS: Project Management using Analytical Tools (8 ECTS)

This course examines the project management process with a focus on business analytics techniques to overcome the pitfalls and obstacles that frequently occur during a typical project. Designed for business leaders responsible for implementing projects, as well as beginning and intermediate project managers. Includes topics on

planning and scheduling issues, costing and budgeting, staffing and organizing, project management methodologies, and the use of data to inform the project manager's decisions throughout the project's lifecycle

EPL: Data Visualisation (8 ECTS)

Data visualisation is an important visual method for effective communication and analysing large datasets. Through data visualisations we are able to draw conclusions from data that are sometimes not immediately obvious and interact with the data in an entirely different way. This course will provide students with an informative introduction to the methods, tools and processes involved in visualising big data. Topics will include: Introduction to visualisation; Information visualisation; Scientific visualisation; Visualisation tools; Design approaches for visualisation; Visualisation for communication. We will use a variety of tools so that students become comfortable engaging with different software and confident trialing new packages to find those that best meet your needs. This includes R, Tableau and D3.js. By the end of the semester, we aim for students to feel comfortable designing and developing visual stories with data.

Elective Courses

C4E: Data Entrepreneurship (4 ECTS)

Existing business models change and new business models are created through the use of data analytics and business intelligence. The purpose of this course is to introduce new business models arising by the use of data analytics, and to provide students with the ability to develop new data-driven business processes in more traditional organizations. To do so, the students will learn how to balance diverse issues from the fields of analytics, business, entrepreneuring, law, ethics, etc. The course makes extensive use of case studies and provides many examples of existing enterprises and how their business model was transformed by the use of business analytics, in the fields of customer acquisition, recommended systems (e.g. Amazon/Netflix), healthcare (e.g. 23andme), targeted/personalised advertising (e.g. Facebook), logistics (e.g. Maersk), etc.

LAW: Ethics of Data Science (4 ECTS)

The course is designed to build students' ethical imaginations and skills for collecting, storing, sharing and analyzing data derived from human subjects including data used in algorithms. The course provides historical background to understand the tenets of informed consent, discrimination, and privacy. Using case study design, students will explore current applications of quantitative reasoning in organizations, algorithmic transparency, and unintended automation of discrimination via data that contains biases rooted in race, gender, class, and other characteristics.

6. Short Programmes

Individual modules (called *short programmes*) can be taken separately and provide certificates. The following programmes have been identified:

Data Analytics Using R Language (8 ECTS)

Participants will be able to: Understand key concepts in data science and their real-world applications; work effectively with R language; familiarize with data management tools.

MAS: Probability and Statistics for Data Science

This is a theoretical course covering fundamentals topics of probability and statistics in the context of data science with its inherent challenges. This course will start with a review of fundamental probability, covering topics like random variables, their distribution functions, expected values, conditioning on certain events and independence. The students will be acquainted with certain families of probability distributions and then will learn how to estimate certain quantities of interest from observations. A range of properties of estimators will be studied, including sufficiency, unbiasedness and consistency which enable the evaluation of their quality. The students will also learn how to introduce different

types of hypotheses, how to construct tests for their hypotheses, as well as how to compare between tests and how to construct confidence intervals for their estimators.

EPL: Big Data Management (8 ECTS)

Big Data requires the storage, organization, and processing of data at a scale and efficiency that go well beyond the capabilities of conventional information technologies. In this course, we will study the state of the art in big data management: we will learn about algorithms, techniques and tools needed to support big data processing. In addition, this course will examine real applications that require massive data analysis and how they can be implemented on Big Data platforms. The course will consist of lectures based both on textbook material and scientific papers. It will also include programming assignments that will provide students with hands-on experience on building data-intensive applications using existing Big Data platforms.

Data Analytics for Business

Participants will be able to: approach business problems data-analytically; think carefully and systematically about whether and how data and business analytics can improve business performance; develop business analytics ideas; analyze data using business analytics software; and generate business insights.

Courses:

EPL: Data Visualisation (8 ECTS)

Data visualisation is an important visual method for effective communication and analysing large datasets. Through data visualisations we are able to draw conclusions from data that are sometimes not immediately obvious and interact with the data in an entirely different way. This course will provide students with an informative introduction to the methods, tools and processes involved in visualising big data. Topics will include: Introduction to visualisation; Information visualisation; Scientific visualisation; Visualisation tools; Design approaches for visualisation; Visualisation for communication. We will use a variety of tools so that students become comfortable engaging with different software and confident trialing new packages to find those that best meet your

needs. This includes R, Tableau and D3.js. By the end of the semester, we aim for students to feel comfortable designing and developing visual stories with data.

BUS: Data Mining for Business Analytics (8 ECTS)

Enterprises, organizations and individuals are creating, collecting, and using massive amount of structured and unstructured data with the goal to convert the information into knowledge, to improve the quality and the efficiency of their decision-making process, and to better position themselves to the highly competitive marketplace. Data mining is the process of finding, extracting, visualizing and reporting useful information and insights from both small and large datasets with the help of sophisticated data analysis methods. It is part of the business analytics, which refers to the process of leveraging different forms of analytical techniques to achieve desired business outcomes through requiring business relevancy, actionable insight, performance management, and value management. The students in this course will study the fundamental principles and techniques of data mining. They will learn how to apply advanced models and software applications for data mining. Finally, students will learn how to examine the overall business process of an organization or a project with the goal to understand (i) the business context where hidden internal and external value is to be identified and captured, and (ii) exactly what the selected data mining method does.

Fundamental Data Analytics

Participants will be able to: prepare and analyze data; learn the different stages of exploratory data analysis and how to create interactive dashboards; familiarize with Python Fundamentals for Data Analysts.

EPL: Introduction to Data Science and Analytics (8 ECTS)

This course will examine how data analysis technologies can be used to improve decision-making. The aim is to study the fundamental principles and techniques of data science, and we will examine real- world examples and cases to place data science techniques in context, to develop data-analytic

thinking, and to illustrate that proper application is as much an art as it is a science. In addition, this course will work hands-on with the Python programming language and its associated data analysis libraries.

EPL: Data Visualisation (8 ECTS)

Data visualisation is an important visual method for effective communication and analysing large datasets. Through data visualisations we are able to draw conclusions from data that are sometimes not immediately obvious and interact with the data in an entirely different way. This course will provide students with an informative introduction to the methods, tools and processes involved in visualising big data. Topics will include: Introduction to visualisation; Information visualisation; Scientific visualisation; Visualisation tools; Design approaches for visualisation; Visualisation for communication. We will use a variety of tools so that students become comfortable engaging with different software and confident trialing new packages to find those that best meet your needs. This includes R, Tableau and D3.js. By the end of the semester, we aim for students to feel comfortable designing and developing visual stories with data.

7. Conclusions

This deliverable can serve as a starting for building new programs and transitioning existing programs in Data Science. In summary, the key points of this study involve:

- Data Science is a fast evolving discipline.
- Data science programs are being sited in departments and schools of computer science, information science, mathematics and statistics, and management. Many of the most successful, particularly at the undergraduate level, represent university-wide coalitions frequently sponsored by interdisciplinary institutes, rather than by a particular department or school.

- A redesign of the curriculum, integrating the elements of mathematical foundations and computational and statistical thinking at all levels will provide a rich and effective series of courses to prepare graduates for a career in Data Science.