Design and Evaluation of the MSRT Data Science Program

at Tehran Institute for Advanced Studies

Table of Content

About TEIAS

Tehran Institute for Advanced Studies (TEIAS) was established in 2017 within Khatam University. TEIAS is an independent institute in terms of structure and governance, while closely linked to Khatam University. TeIAS was established with the grand ambition of establishing a world-class higher-education institute in Iran. A group of high-profile world-class academics support this initiative as an advisory board.

TEIAS commenced its official activity by holding a special master's degree program in Economics. The program was highly successful, and, soon TeIAS became the first choice of top professors and students as a higher education institute. As a natural next step, the Data Science master's program was designed and launched at TeIAS in 2020.

1

Introduction

“Data Science” is an interdisciplinary field in which mathematics, probability theory and statistics, and computer science and engineering have a prominent role. The main focus of data science is on scientific theories, algorithmic techniques, and software and computer systems to extract information from data. Data science is an active field of research, which has a direct influence in many areas of science and technology. Establishing active research and educational programs are hence essential for national development. Therefore, training specialists who deeply understand the theoretical underpinnings and practical aspects of this interdisciplinary field and are able to specialize given the local particularities and national demand is imperative for each country.

This document, conducted by TeIAS Computer Science faculty members, presents the principles, processes, and results of the design of a national masters programme in Data Science, presented to and approved by the Iranian Ministry of Science, Research and Education (MSRT). The aim of this document is to review and evaluate the program and seek further advice from experts in order to improve its implementation and prepare future revisions.

The rest of this document is structured as follows. Section 2 explains the objectives of our data science program. Section 3 presents our first step towards this program which was consultation with experts. Section 4 describes the data gathering process followed by Section 5 in which we analyze the collected data. Section 6 presents the MSRT national program as well as its evaluation. Section 7 concludes the document and presents the directions of our future steps.

2 Objectives

2.1. Motivation

The need to process and analyze data is a timely challenge; there exist large data sets with various sources such as meteorology, biological data, traffic data, and financial data. The analysis of such data is challenging due to their high dimensions and heterogenous and ad-hoc structures. Data processing is done based on theoretical methods and using data analysis tools. Responding to these contemporary challenges demands rigorous and thorough education and research in both theoretical underpinnings and practical aspects of it and training experts (both for industrial employment and for academic research) in data science requires a cohesive well-designed educational and research program. Because of this, we set out to define a coherent and dedicated data science masters program for wide-scale national use at various universities.

2.2. Design Principles and Criteria

The goal of our data science masters program is to provide an appropriate education to train students who have a sound understanding of theoretical foundations and practical aspects of data science. The graduates of this field will be able to perform as data analysts and data scientists in industry (both in software development for data analytics and in advanced applications of data analytics in other domains such as computational biology and logistics and mobility) . In addition, the graduates will be able to carry out academic research and contribute to the state of art in this field. Our graduates should have a strong research background so that they can join leading data science research groups, e.g., to continue their education and research as PhD students, both in data science and in other related areas of computer science and engineering.

We used the following two design principles for our program in Data Science:

- covering the **breadth** of **core topics** necessary for research and employability in Data Science; and
- allowing for **customization** and **specialisation** based on the availability of expertise at the individual **universities** and the personal interest of **students**.

To achieve the above-mentioned goals, we followed the following five steps:

1- We **consulted** the experts to define the general framework for our design to devise an approach for designing a well-rounded data science program which is tailored based on our goals and needs;

2- We **gathered** data about the existing data science programs in leading universities, taking geographical and topical diversity into account;

3- We **analyzed** these programs to validate the framework identified in step 1 and identified the key courses that would help us create a well-rounded program that aligns with our two design principles;

4- We **designed** our data science program based on the result of the analysis;

5- We constantly **evaluated** our program and **sought comments** from experts to ensure that it meets the requirements.

The last step creates an active feedback loop and we constantly engage in an iterative process to revise and update our program throughout its life-time. The following sections explain each of these steps in more detail.

3

Consultation with Experts

The foundations of our data science program was initially built based on the consultation with prominent faculty members of the computer science and data science community. For our consultation, we assembled an international advisory board comprising 6 leading scientists at the following universities: Eindhoven University of Technology, The Netherlands; University of Maryland College Park, USA; Monash University, Australia; Oxford University, UK; Radboud University Nijmegen, The Netherlands; and University of Twente, The Netherlands.[1] All members of the committee were tenured faculty members (5 full professors, 1 associate professors) all with more than 15 years experience of research and teaching. We made sure that the advisory board covers the breadth of data science and computer science and is both geographically and topically diverse.

Based on consultation with our international advisory board and our faculty members, the design of the data science program was started. In this step, we also studied a Data Science program designed by the School of Mathematical Sciences at Sharif University of Technology and consulted its authors for advice.

We identified a number of focus areas and a number of design principles (reported in Section 2.2). This led to an initial draft program. Next, we defined a process to validate this initial program by performing a comprehensive study of similar programs in a number of outstanding universities. In the next section, we report on the data gathering and validation steps.

1. More details about this process and the members of our initial advisory committee can be found in our earlier document (in Farsi): https://teias.institute/beta/wp-content/uploads/2020/10/%D8%AA%DB%8C%D8%A7%D8%B3.pdf
Note that the different members of our advisory board played different roles in the inception of our program and our Data Science group, including serving in the hiring and search committee and defining the general strategy of the group.

4

Data Gathering

4.1. University Selection

We used the existing Data Science programs at the leading schools worldwide as a yardstick to concretize our design principle. Hence, we selected a reasonably large pool of leading universities and analyzed both the core topics covered by their programs and their avenues for specialization. To make our study comprehensive, we picked outstanding universities from various countries based on our consultation with our international advisory board; we also made cautious and informed use of subject ranking in computer science in well-known league tables (such as the Times Higher Education subject ranking in Computer Science). This resulted in the following set of universities which are grouped based on their region:

Europe:

- École polytechnique fédérale de Lausanne (EPFL)
- ETH Zurich
- Delft University of Technology (TU Delft)
- University of Helsinki

United States:

- Stanford University
- Carnegie Mellon University (CMU)
- University of California, Berkeley
- Columbia University
- New York University (NYU)
- University of Southern California (USC)
- University of Massachusetts Amherst (UMass)
- Johns Hopkins University
- University of Rochester
- Emory University

Australia:

- University of Sydney
- University of Melbourne

Canada:

- University of Waterloo
- McGill University
- University of Toronto

4.2. Studying Data Science Programs

After selecting the universities, we extracted and studied their data science programs. For this purpose, we used the program and course pages. To manage the complexity of the data, we used the following techniques:

- We eliminated the elective courses that are not affiliated with data science (e.g., software architecture and networks and systems).
- - We categorized those courses that have slightly different titles (e.g., courses entitled as “Machine Learning”, “Introduction to Machine learning”, or “Fundamentals of Machine learning”) in one class. We were conscious of the content of the courses and made sure that the courses mainly cover similar concepts and can be identified as such. For each category, we chose a more general title that represents them all; for traceability, we did record the original titles as well, which are included in an appendix of this document.
- We did not confine ourselves to titles and studied the curriculum of each course to identify the courses that are closely connected (e.g “Artificial Intelligence”, “Machine Learning”, and “Deep learning”).
- At a more global level, we identified the general areas to which the individual courses belong. Our study resulted in identifying four broad areas as follows:
  - Artificial Intelligence and Machine Learning
  - Data Analysis and Systems
  - Mathematics and Statistics
  - Uncategorized courses (mostly comprising general software engineering, algorithmics, and research methodology)

These techniques made it possible to represent various areas and specializations more concisely for the evaluation of our proposed program.

		EPFL	ETH	TUDelft	Helsinki	Stanford	CMU	Berkeley	Columbia	NYU	USC	UMASS	Johns Hopkins	Rochester	Emory	Sydney	Melbourne	Waterloo	Mcgill	Utoronto
AI & Machine Learning	Artificial Intelligence			✓	✓	✓			✓		✓	✓		✓	✓			✓		
	Natural Language Processing	✓	✓			✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			
	Computer vision	✓	✓	✓	✓		✓		✓			✓		✓	✓					
	Speech Processing	✓		✓					✓	✓										
	Machine learning	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Deep learning	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓		✓				
	Neural Networks	✓	✓			✓	✓		✓			✓								
	Statistical Learning Theory		✓			✓										✓		✓		
Data Analysis and Systems	Data Management		✓	✓							✓			✓			✓	✓	✓	
	Database Systems							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Distributed Systems	✓	✓	✓	✓							✓		✓	✓		✓	✓	✓	
	Cloud Computing Architecture		✓		✓			✓								✓	✓			
	Data Analysis	✓	✓	✓			✓		✓			✓	✓	✓		✓	✓	✓	✓	
	Introduction to Data Science				✓			✓	✓	✓			✓			✓		✓		
	Big Data	✓	✓		✓				✓	✓	✓	✓	✓	✓						✓
	Data Visualization	✓		✓	✓			✓			✓	✓	✓			✓	✓	✓		
	Data Mining					✓	✓		✓		✓		✓	✓	✓	✓			✓	
	Information Retrieval		✓	✓					✓		✓	✓			✓	✓			✓	
	Data Analysis in Complex Networks	✓		✓																
	Behind the Data: Humans & Values							✓												
	Security and privacy	✓	✓	✓				✓							✓	✓	✓			
Mathematics and Statistics	Foundations of Data Science	✓	✓																	✓
	Mathematics for Data Science	✓	✓		✓					✓	✓						✓		✓	
	Probability and Statistics	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
	Optimization	✓	✓			✓	✓		✓	✓	✓	✓	✓				✓	✓	✓	
	Foundations of Graphical Models						✓		✓				✓				✓			
	High-Dimensional Statistics		✓		✓												✓			
	Time Series Analysis	✓	✓				✓		✓	✓		✓		✓		✓			✓	
	Stochastic Systems		✓			✓						✓	✓	✓			✓			
	Signal Processing	✓	✓									✓								
	Causal Inference		✓			✓		✓	✓	✓		✓		✓				✓	✓	
Other	Algorithms	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓			
	Performance Analysis	✓	✓	✓	✓						✓									
	Design of Experiments		✓			✓						✓								
	Computational complexity	✓							✓					✓						
	Python programming																			✓
	Information Theory	✓	✓	✓					✓			✓								
	Software Engineering			✓								✓								

Table 1. Courses in the data science programs of different universities where bold check marks indicate mandatory courses

The collected and analyzed data is presented in Table 1. The first column represents the consolidated title of courses; each university is represented in an individual column and a check mark denotes that the consolidated course in that row is offered in their data science program. A bold check mark indicates that the course is mandatory in the corresponding program. The specific titles of courses at different universities are recorded in Appendix 1.

5

Analyzing Data Science Programs

We analyzed the programs regarding the two designed principles mentioned in Section 2.2. To this end, we identified the core courses along with program features that increase customizability. The result of this analysis is provided in this section and guides us in the design of the MSRT National Program in the next section.

5.1. Core Courses

We used the percentage of the universities that present a given course as a metric to identify it as a core subject, i.e., the subjects with higher percentages are more likely to be at the core of the data science program. Table 2 shows the results of calculating these percentages in the descending order.

Course	PCT.	Course	PCT.
Machine learning	94.7%	Security and privacy	36.8%
Probability and Statistics for Data Science	89.4%	Mathematics for Data Science	36.8%
Algorithms	73.6%	Neural Networks	31.5%
Natural Language Processing	68.4%	Stochastic Systems	31.5%
Deep learning	63.1%	Cloud Computing Architecture	26.3%
Database Systems	63.1%	Performance Analysis	26.3%
Data Analysis	63.1%	Information Theory	26.3%
Optimization	63.1%	Speech Processing	21.0%
Distributed Systems	52.6%	Statistical Learning Theory	21.0%
Big Data	52.6%	Foundations of Graphical Models	21.0%
Data Visualization	52.6%	Foundations of Data Science	15.7%
Artificial Intelligence	47.3%	High-Dimensional Statistics	15.7%
Computer vision	47.3%	Signal Processing	15.7%
Data Mining	47.3%	Design of Experiments	15.7%
Time Series Analysis	47.3%	Computational Complexity	15.7%
Causal Inference	47.3%	Modeling & Data Analysis in Complex Networks	10.5%
Information Retrieval	42.1%	Software Engineering	10.5%
Data Management	36.8%	Behind the Data: Humans and Values	5.2%
Introduction to Data Science	36.8%	Python programming	5.2%

Table 2. The percentages of universities that present each course

5.2. Customizability

To analyze the customization potential in different programs, we took advantage of our categorization of courses into four main areas in Table 1. Given these areas and their consolidated courses, we calculated the percentage of consolidated courses in each area offered in any particular program. It is reasonable to assume that the programs that cover a higher percentage of courses in each and every area, are more customizable and flexible and can be used as a model to evaluate our program. Table 3 shows these percentages. As the last area (uncategorized course) contains a very wide range of courses, it is not included in this table.

	Artificial Intelligence and Machine Learning	Data Analysis and Systems	Mathematics and Statistics
EPFL	75.0%	46.1%	60.0%
ETH	75.0%	46.1%	90.0%
TUDelft	62.5%	53.8%	0.0%
Helsinki	50.0%	46.1%	30.0%
Stanford	62.5%	7.6%	40.0%
CMU	62.5%	15.3%	40.0%
Berkeley	37.5%	46.1%	20.0%
Columbia	87.5%	46.1%	50.0%
NYU	50.0%	23.0%	50.0%
USC	25.0%	46.1%	30.0%
UMASS	75.0%	46.1%	60.0%
Johns Hopkins	25.0%	46.1%	40.0%
Rochester	62.5%	46.1%	40.0%
EMORY	50.0%	38.4%	10.0%
Sydney	50.0%	61.5%	20.0%
Melbourne	25.0%	53.8%	60.0%
Waterloo	37.5%	46.1%	30.0%
Mcgill	12.5%	46.1%	40.0%
Utoronto	12.5%	7.69%	20.0%

Table 3. The percentage of courses that are covered from each area by each university

6

MSRT Program and Its Evaluation

We used the data gathered in Section 4 and the subsequent analysis in Section 5 to evaluate and review the design of our nation-wide data science program proposed to the Iranian Ministry of Science, Research, and Technology (MSRT). In this section we describe this program, called henceforth the MSRT program, and evaluate it based on our defined principles.

6.1. MSRT Program

The courses in the MSRT program are selected based on our design principles while taking the leading data science programs into consideration. Tables 4, 5, 6, and 7 show our proposed program where each table corresponds to one of the main areas identified in Table 1. The first column includes the consolidated course titles as in Table 1. The MSRT column represents the specific course titles we gave to the corresponding course in our program (also taking local variation of course names into account) where the bold names indicate mandatory courses. The "overall" column is the percentage of reviewed universities (worldwide) that present each course (from Table 3). The last column indicates the percentage of universities at which the corresponding course is mandatory. This data is used in the next section to verify the correlation of our program (both in terms of mandatory and optional courses) with the core data science subjects at the reviewed universities and evaluate it against our design principles.

Course Title	MSRT	Overall	Mandatory
Artificial Intelligence and Machine Learning			
Artificial Intelligence		47.3%	26.3%
Natural Language Processing	Natural Language Processing	68.4%	15.7%
Computer Vision	Computer Vision	47.3%	5.2%
Speech Processing		21.0%	0.0%
Machine Learning	**Machine Learning**	94.7%	68.4%
Deep learning	Deep Learning	63.1%	10.5%
Neural Networks	Neural Networks	31.5%	10.5%
Statistical Learning Theory	Statistical Learning	21.0%	10.5%
Covered Courses	75.0%		

Table 4. "Artificial Intelligence and Machine Learning" area courses in the MSRT program

Course Title	MSRT	Overall	Mandatory
Data Analysis and Systems			
Data Management		36.8%	10.5%
Database Systems	Advanced Database	63.1%	31.5%
Distributed Systems	Distributed Systems	52.6%	15.7%
Cloud Computing Architecture	Cloud Computing	26.3%	5.2%
Data Analysis	**Applied Data Analytics**	63.1%	26.3%
Introduction to Data Science		36.8%	21.0%
Big Data	Big Data Analytics	52.6%	21.0%
Data Visualization	Data Visualization	52.6%	26.3%
Data Mining	Advanced Data Mining	47.3%	15.7%
Information Retrieval		42.1%	15.7%
Modeling and Data Analysis in Complex Networks	Graph Mining	10.5%	0.0%
Behind the Data: Humans and Values		5.2%	0.0%
Security and privacy	Information Security	36.8%	5.2%
Covered Courses	69.2%		

Table 5. "Data Analysis and Systems" area courses in the MSRT program

Course Title	MSRT	Overall	Mandatory
Mathematics and Statistics			
Foundations of Data Science		15.7%	15.7%
Mathematics for Data Science Probability and Statistics for Data Science	**Mathematics for Data Science**	36.8% 89.4%	63.1%
Optimization	Optimization for Data Science	63.1%	31.5%
Foundations of Graphical Models	Probabilistic Graphical Models	21.0%	10.5%
High-Dimensional Statistics	High-Dimensional Probability for Data Science	15.7%	5.2%
Time Series Analysis	Time Series	47.3%	0.0%
Stochastic Systems		31.5%	5.2%
Signal Processing	Compressive Sensing	15.7%	0.0%
Causal Inference	Causal Inference	47.37%	10.5%
Covered Courses	70.0%		

Table 6. "Mathematics and Statistics" area courses in the MSRT program

Course Title	MSRT	Overall	Mandatory
Uncategorized			
Algorithms	Advanced Algorithms	73.6%	42.1%
Performance Analysis		26.3%	0.0%
Design of Experiments		15.7%	5.2%
Computational complexity		15.7%	5.2%
Python programming		5.2%	0.0%
Information Theory		26.3%	0.0%
Software Engineering	Software Synthesis	10.5%	0.0%

Table 7. Other (uncategorized) courses included in the MSRT program

6.2. Program Evaluation

In this section, we evaluate the MSRT program with respect to the defined design principles. The first principle is that the program should cover the core topics of data science. It is reasonable to consider well-known courses (e.g machine learning) along with the courses that are presented in a majority of reviewed data science programs as the core courses. As it can be seen in Tables 4-6, our MSRT program features all such courses. Moreover, the courses that are defined as mandatory at most of the reviewed universities, are mandatory in our program as well.

The second criterion is if the program is customizable and can be tailored based on the specialization of expertise in the individual universities and the interest of students. To evaluate this criteria, we examine the percentage of covered courses from each category. If a program is sufficiently customizable, it should cover an acceptable number of courses of in each area. Tables 4-6 show that we achieved this goal as our program covers more than about 70% of the courses in each area.

7

Conclusions

In this document, we presented our method to design a data science program for the Iranian Ministry of Science, Research, and Technology (MSRT). We set out two principle design decisions: namely covering the core topics of data science and being customizable to the specific expertise of the faculty and the interest of the students. We specified our methodology to define such a program based on consultation with an international advisory board. We used different sources such as experts' advice and the data science programs in outstanding universities. We design our program based on our initial design principles and using these sources. Subsequently, we quantitatively evaluated the outcome by measuring how much it includes the core courses of data science and how much it allows for customization in different areas. The current evaluation results show that the proposed program meets the requirements.

Our next step is to present this program back to domain experts and solicit detailed feedback in order to plan our next revision rounds. In this process, we will also be interviewing our own students and faculty and feed this information back into the revision process.

Appendix 1

In this section we provide more details on the studied data science programs. Each section presents the courses of the main categories. The universities are divided based on their region.

Area1: Artificial Intelligence and Machine Learning

	EPFL	ETH	TUDelft	Helsinki
Artificial Intelligence			Artificial Intelligence Techniques	Introduction to Artificial Intelligence
Natural Language Processing	Introduction to natural language processing	Natural Language Understanding		
Computer vision	Computer vision	Computer Vision	Advanced Digital Image Processing	Computer Vision
Speech Processing	Automatic speech processing		Digital Audio and Speech Processing	
Machine learning	Machine learning	Advanced Machine Learning	Machine Learning	Introduction to Machine Learning
Deep learning	Deep learning	Deep Learning	Deep Learning	Deep Learning
Neural Networks	Artificial neural networks	Neural Network Theory		
Statistical Learning Theory		Statistical Learning Theory		
	75.0%	**75.0%**	**62.5%**	**50.0%**

Table 8. "Artificial Intelligence and Machine Learning" area courses in the European universities

	USC	UMASS	Johns Hopkins	Rochester	EMORY
Artificial Intelligence	Foundations of Artificial Intelligence	Artificial Intelligence		Artificial Intelligence	Artificial Intelligence
Natural Language Processing		Natural Language Processing	Semantic Natural Language Processing	Natural Language Processing	Natural Language Processing
Computer vision		Intelligent Visual Computing		Advanced Topics in Computer Vision	Digital Image Processing
Speech Processing					
Machine learning	Machine Learning	Machine Learning	Introduction to Machine Learning	Machine Learning	Machine Learning
Deep learning		Reinforcement Learning		Deep Learning	
Neural Networks		Neural Networks: A Modern Intro.			
Statistical Learning Theory					
	25.0%	**75.0%**	**25.0%**	**62.5%**	**50.0%**

Table 9. "Artificial Intelligence and Machine Learning" area courses in the American universities (part 1)

	Stanford	CMU	Berkeley	Columbia	NYU
Artificial Intelligence	Artificial Intelligence: Principles and Techniques			Artificial Intelligence	
Natural Language Processing	Natural Language Processing with Deep Learning	Algorithms for NLP	Natural Language Processing with Deep Learning	Natural language processing	Natural Language Processing
Computer vision		Computer Vision		Computer vision	
Speech Processing				Sparse Signal Modeling	Speech Recognition
Machine learning		Introduction to Machine Learning	Applied Machine Learning	Machine Learning	Machine Learning
Deep learning	Deep Learning	Topics in Deep Learning	Deep Learning in the Cloud and at the Edge	Advanced/ Mathematics of Deep Learning	Deep Learning
Neural Networks	Convolutional Neural Networks for Visual Recognition	Neural Networks for NLP		Neural Networks and Deep Learning	
Statistical Learning Theory	Statistical Learning Theory				
	62.5%	**62.5%**	**37.5%**	**87.5%**	**50.0%**

Table 10. "Artificial Intelligence and Machine Learning" area courses in the American universities (part 2)

	Sydney	Melbourne	Waterloo	Mcgill	Utoronto
Artificial Intelligence			Introduction to Artificial Intelligence		
Natural Language Processing	Natural Language Processing	Natural Language Processing			
Computer vision					
Speech Processing					
Machine learning	Advanced Machine Learning	Statistical Machine Learning	Introduction to Machine Learning	Machine Learning	Machine learning
Deep learning	Deep learning				
Neural Networks					
Statistical Learning Theory	Statistical Learning and Data Mining		Statistical Learning		
	50.0%	**25.0%**	**37.5%**	**12.5%**	**12.5%**

Table 11. "Artificial Intelligence and Machine Learning" area courses in the Australian and Canadian universities

Area 2: Data Analysis and Systems

	EPFL	ETH	TUDelft	Helsinki
Data Management		Data Management Systems	Web-scale Data Management	
Database Systems				
Distributed Systems	Distributed information systems	Principles of Distributed Computing	Distributed Systems	Distributed Data Infrastructures
Cloud Computing Architecture		Cloud Computing Architecture		Cloud and Edge Computing
Data Analysis	Applied data analysis	Systems-on-chip for Data Analytics and Machine Learning	Cyber/Multivariate Data Analytics	
Introduction to Data Science				
Big Data	Systems for data science	Big Data		Introduction to Big Data Management
Data Visualization	Data visualization		Data Visualization	Interactive Data Visualization
Data Mining				
Information Retrieval			Information Retrieval	Introduction to Information Retrieval
Modeling and Data Analysis in Complex Networks	Networks out of control		Modeling and Data Analysis in Complex Networks	
Behind the Data: Humans and Values				
Security and privacy	Information security and privacy	Information Security	Security and Cryptography	
	46.2%	**46.2%**	**53.8%**	**46.2%**

Table 12. "Data Analysis and Systems" area courses in the European universities

	Stanford	CMU	Berkeley	Columbia	NYU
Data Management					
Database Systems			Fundamentals of Data Engineering	Introduction to Databases	Database Systems
Distributed Systems					
Cloud Computing Architecture			Deep Learning in the Cloud and at the Edge		
Data Analysis		Data Analysis		Advanced Data Analysis	
Introduction to Data Science			Introduction to Data Science Programming	Applied Data Science	Introduction to Data Science
Big Data				Big Data Analytics	Big Data
Data Visualization			Data Visualization		
Data Mining	Modern Applied Statistics: Data Mining	Multimedia Databases and Data Mining		Data Mining	
Information Retrieval				Information Processing	
Modeling and Data Analysis in Complex Networks					
Behind the Data: Humans and Values			Behind the Data: Humans and Values		
Security and privacy			Privacy Engineering		
	7.7%	**15.4%**	**46.2%**	**46.2%**	**23.1%**

Table 13. "Data Analysis and Systems" area courses in the American universities (Part 1)

	USC	UMASS	Johns Hopkins	Rochester	EMORY
Data Management	Foundations of Data Management			Advanced Topics in Data Management	
Database Systems	Database Systems	Database Design and Implementation	Principles of Database Systems	Introduction to Databases	Database Systems
Distributed Systems		Distributed and Operating Systems		Parallel and Distributed Systems	Advanced Computer Systems
Cloud Computing Architecture					
Data Analysis		Business Intelligence and Analytics	Statistical Methods and Data Analysis	Data Analysis	
Introduction to Data Science			Data Science		
Big Data	Advanced Big Data Analytics	Advanced Systems for Big Data Analytics	Big Data Processing Using Hadoop	Big Data	
Data Visualization	Information Visualization	Data Visualization and Exploration	Data Visualization		
Data Mining	Foundations and Applications of Data Mining		Data Mining	Data Mining	Data Mining
Information Retrieval	Information Retrieval and Web Search Engines	Information Retrieval			Information Retrieval
Modeling and Data Analysis in Complex Networks					
Behind the Data: Humans and Values					
Security and privacy					Data Privacy and Security
	46.2%	**46.2%**	**46.2%**	**46.2%**	**38.5%**

Table 14. "Data Analysis and Systems" area courses in the American universities (Part 2)

	Sydney	Melbourne	Waterloo	Mcgill	Utoronto
Data Management		Knowledge Management Systems	Principles of Data Management and Use	Knowledge Management	
Database Systems	Advanced Data Models	Advanced Database Systems	Database Systems Implementation	Database Design & Development	
Distributed Systems		Distributed Systems	Distributed Systems	Distributed Systems	
Cloud Computing Architecture	Cloud Computing	Cluster and Cloud Computing			
Data Analysis	Data Analytics and Business Intelligence	Analysis of High-Dimensional Data	Exploratory Data Analysis	Data Analysis & Report Writing	
Introduction to Data Science	Principles of Data Science		Fundamentals of Computer Science for Data Science		
Big Data	Machine Learning and Data Mining			Data Mining	
Data Visualization	Visual Analytics	Information Visualisation	Data Visualization		
Data Mining	Machine Learning and Data Mining			Data Mining	
Information Retrieval	Multimedia Retrieval			Information Retrieval	
Modeling and Data Analysis in Complex Networks					
Behind the Data: Humans and Values					
Security and privacy	Information Security Management	Cryptography and Security			
	61.5%	**53.8%**	**46.2%**	**46.2%**	**7.7%**

Table 15. "Data Analysis and Systems" area courses in the Australian and Canadian universities

Area 3: Mathematics and Statistics

	EPFL	ETH	TUDelft	Helsinki
Foundations of Data Science	Foundations of Data Science	Mathematics of Information		
Mathematics for Data Science	Mathematics of data: from theory to computation	Mathematics of Data Science		Scientific Computing
Probability and Statistics for Data Science	Statistics for data science	Fundamentals of Mathematical Statistics		Statistical Data Science/ Computational Statistics
Optimization	Optimization for machine learning	Optimization for Data Science		
Foundations of Graphical Models				
High-Dimensional Statistics		High-Dimensional Statistics		High Dimensional Statistics
Time Series Analysis	Time series	Time Series Analysis		
Stochastic Systems		Stochastic Systems		
Signal Processing	Mathematical foundations of signal processing	Mathematical Methods of Signal Processing		
Causal Inference		Causality		
	60.0%	**90.0%**	**0.0%**	**30.0%**

Table 16. "Mathematics and Statistics" area courses in the European universities

	Stanford	CMU	Berkeley	Columbia	NYU
Foundations of Data Science					
Mathematics for Data Science					Mathematical Tools for Data Science
Probability and Statistics for Data Science	Theory of Statistics	Probability & Mathematical Statistics	Statistics for Data Science	Probability and Statistics	Probability and Statistics for Data Science
Optimization	Optimization	Convex Optimization		Optimization	Optimization and Computational Linear Algebra
Foundations of Graphical Models		Probabilistic Graphical Models		Foundations of Graphical Models	
High-Dimensional Statistics					
Time Series Analysis		Regression Analysis		Linear Regression/Time Series Analysis	Probabilistic Time Series Analysis
Stochastic Systems	Stochastic Methods in Engineering				
Signal Processing					
Causal Inference	Topics in Causal Inference		Experiments and Causal Inference	Causal Inference	Inference and Representation
	40.0%	**40.0%**	**20.0%**	**50.0%**	**50.0%**

Table 17. "Mathematics and Statistics" area courses in the American universities (Part 1)

	USC	UMASS	Johns Hopkins	Rochester	EMORY
Foundations of Data Science					
Mathematics for Data Science	Numerical Methods				
Probability and Statistics for Data Science	Applied Probability	Intro to Probability and Math Statistics	Computational Statistics	Computational Introduction to Statistics	Statistical Methods
Optimization	Theory and Computational Methods for Optimization	Optimization	Introduction to Optimization		
Foundations of Graphical Models			Statistical Models and Regression		
High-Dimensional Statistics					
Time Series Analysis		Regression Modeling		Time Series Analysis and Forecasting in Data Science	
Stochastic Systems		Stochastic Processes in Industrial Engineering	Probability and Stochastic Process	Stochastic Inverse Modeling in Geophysics	
Signal Processing		Intro to Compressive Sensing			
Causal Inference		Introduction to Causal Inference in a Big Data World		Causal Inference	
	30.0%	**60.0%**	**40.0%**	**40.0%**	**10.0%**

Table 18. "Mathematics and Statistics" area courses in the American universities (Part 2)

	Sydney	Melbourne	Waterloo	Mcgill	Utoronto
Foundations of Data Science					Foundations of Data Science
Mathematics for Data Science		Mathematics of Risk		Mathematical Statistics	
Probability and Statistics for Data Science	Computational Statistical Methods	Computational Statistics & Data Science	Statistical Concepts for Data Science		Statistics for Data Science
Optimization		Optimisation for Industry	Optimization for Data Science	Optimization & Optimal Control	
Foundations of Graphical Models		Statistical Modelling for Data Science			
High-Dimensional Statistics		Multivariate Statistics for Data Science			
Time Series Analysis	Predictive Analytics			Time Series Analysis	
Stochastic Systems		Stochastic Calculus with Applications			
Signal Processing					
Causal Inference			Computational Inference	Causal Inference in Biostatistics	
	20.0%	**60.0%**	**30.0%**	**40.0%**	**20.0%**

Table 19. "Mathematics and Statistics" area courses in the Australian and Canadian universities

Area 4: Uncategorized Courses

	EPFL	ETH	TUDelft	Helsinki
Algorithms	Advanced algorithms	Advanced algorithms	Advanced algorithms	Design and Analysis of Algorithms
Performance Analysis	Performance evaluation	Design of Parallel and High-Performance Computing	Performance Analysis	Tools of High Performance Computing
Design of Experiments		Applied Analysis of Variance and Experimental Design		
Computational complexity	Computational complexity			
Python programming				
Information Theory	Information theory and coding	Information Theory	Information Theory	
Software Engineering			Software Verification	

Table 20. The uncategorized area courses in European universities

	Stanford	CMU	Berkeley	Columbia	NYU
Algorithms	Distributed Algorithms and Optimization	Algorithms		Algorithms for Data Science	Fundamental Algorithms
Performance Analysis					
Design of Experiments	Design of Experiments				
Computational complexity				Computational Aspects of Robotics	
Python programming					
Information Theory				Information Theory	
Software Engineering					

Table 21. The uncategorized area courses in the American universities (part 1)

	USC	UMASS	Johns Hopkins	Rochester	EMORY
Algorithms	Analysis of Algorithms	Advanced Algorithms	Algorithms for Data Science	Design and Analysis of Efficient Algorithms	Algorithms
Performance Analysis	High Performance Computing and Simulation				
Design of Experiments		Design of Experiments			
Computational complexity				Computational Complexity	
Python programming					
Information Theory		Applied Information Theory			
Software Engineering		Advanced Software Engineering: synthesis and development;			

Table 22. The uncategorized area courses in the American universities (part 2)

	Sydney	Melbourne	Waterloo	Mcgill	Utoronto
Algorithms		Distributed Algorithms			
Performance Analysis					
Design of Experiments					
Computational complexity					
Python programming					Python Programming
Information Theory					
Software Engineering					

Table 23. The uncategorized area courses in the Australian and Canadian universities

Telephone: **02189174612**
E-mail: **cs@teias.institute**
Website: **https://teias.institute/computer-science-department/**
Telegram: @**pasargadschools**