



## CALL FOR APPLICATIONS FOR ADMISSION

### GRADUATE STUDY PROGRAM "MASTER OF SCIENCE in DATA SCIENCE"

#### Full Time and Part Time Programs 2018-2019

Data Science is the study of data through computational and statistical techniques, in order to answer questions, develop explanatory and predictive models, perform analyses and communicate the results in revealing ways. Data science draws from a wide variety of disciplines such as computer science, artificial intelligence, statistics, economics, and operations research. It applies quantitative methods to uncover relationships in data drawn from business, medicine, financial, social or other domains. It is a key driver of improvements to all aspects of business, including strategy, operations, marketing, finance, and human resource management.

The Master of Science in Data Science, the first of its kind in Greece, offers students in-depth focus in data science while allowing them to tailor it their particular interests. Students will be interacting with diverse faculty members and other students, given the opportunity to complete innovative data science projects and be exposed to industry needs and real-life data science challenges.

The program focuses on computation and quantitative techniques and offers students new opportunities for building sustainable competitive advantage through data analysis. The full-time program consists of an academic year of taught courses and a 3 month-long Analytics Capstone project that enables students to work on a real-world data-intensive problem using the tools and skills learned in the program. The part-time program consists of two academic years of taught courses.

The program is offered jointly by the Departments of Informatics and Statistics in the School of Information Sciences and Technology of the Athens University of Economics and Business.

#### The Athens University of Economics and Business

The Informatics Department has been in existence, in its present form, since 1984 and is focused on providing innovative undergraduate and postgraduate education, along with research for the information and computing professions. Each year, it welcomes approximately 200 undergraduate and over 100 graduate students. Faculty members have over 20 years of academic teaching experience on average and collectively have contributed more than 1.000 research publications, which have attracted over 10.000 references from other researchers worldwide. More than half the faculty have been faculty members in leading American and other European Universities.

The Department of Statistics was founded in 1989. Its purpose is to provide innovative and high quality undergraduate and postgraduate education in probability and statistics. Every year 120 undergraduate students are admitted, and two postgraduate courses are run, with excellent student to faculty ratios. The department has an international reputation for development of statistical methodology that has grown from its history of significant contributions to research and teaching in statistics.

Athens University of Economics and Business (AUEB) was founded in 1920. It is considered one of the most competitive universities, at the European level, in the fields of Economics, Business Administration, Informatics, Statistics, Marketing, Accounting and Finance. AUEB was the first Greek University to establish postgraduate studies. Today it enrolls over 2000 students in 35 part-time and full-time Master's level postgraduate programs with duration of 1 to 2 years. It is the first university in Greece to receive the distinction of Excellence, according to the internationally accepted EFQM (European Foundation of Quality

Management) Excellence Model, and it has also received the corresponding “Ever to Excel” Greek distinction. AUEB is by far the most international of Greek universities: It has the largest ratio of Erasmus students to its active student population, and a large number of undergraduate and postgraduate students participating in the Erasmus and Erasmus+ programs. It hosts one of the most active branches of AIESEC, through which it provides valuable opportunities for internships abroad.

## Target audience

For the full-time program: Excellent recent graduates, or midcareer professionals looking to upskill, with strong technical/math skills, from engineering, mathematics, statistics, finance, economics, operations research, and computer science. Prospective students get excited about gathering, measuring and analyzing information and want to focus on quantitative, computational methods to unlock the potential of data assets to reveal patterns, make predictions and improve decision-making. All applicants should have demonstrated academic success as evidenced by undergraduate and graduate courses and grades.

For the part-time program: Early- and Mid-career professionals (at least 2 years full time professional: experience required) wanting to face the challenge of understanding and exploiting the deluge of data in their organizations. Any professional (in private or public sector) with a mandate to gather, measure and analyze information. Professionals especially in business consulting, retail banking, market research, quantitative marketing, IT, Business Intelligence, finance, operations as well as managers focused on using data to extract business value. Recent programming experience and facility with basic mathematical concepts and quantitative techniques are necessary.

The admissions committee considers the totality of a candidate’s experience, skills, personality and potential to reach a decision, aiming for a diverse class of motivated students who can most benefit from and contribute to our rigorous program of study.

## Application process and admission requirements

The application period for the MSc in Data Science for this academic year (FT: 2018-2019/PT: 2018-2020) is as follows: **July 23rd, 2018 to August 11th, 2018**. The admissions committee may review submitted applications at any time and send acceptance/rejection letters earlier than the respective deadline. Acceptance letters will be sent out at the latest by August 30<sup>th</sup>.

Each application is required to include the following:

- Completed and signed application form with photo
- Copy of all university degrees/diplomas received
- Copy of transcripts of grades in Greek or English. Accepted candidates must submit official transcripts
- Certificate of equivalence for degrees from foreign Universities, issued by DOATAP/Hellenic NARIC (or proof that an application for certification has been filed -- admission is contingent on submission of certificate by the Program enrollment date)
- Proof of knowledge of English: Certificate of Proficiency in English from U. of Michigan/ Cambridge, or other equivalent
- GRE scores (if available)
- Two recommendation letters (to be sent via post to the Program Administrator or via email to [datascience@aueb.gr](mailto:datascience@aueb.gr))
- Proofs of employment (optional for full time program, required for part time program)
- CV in English



Applications are accepted until August 11th, 2018. Places are limited.

**Applications will be filled online at <http://e-graduate.applications.aueb.gr/> and printed copy must be submitted to the secretary of the program during the application period at the following address:**

Secretariat for Graduate Programs, Informatics Department, Athens University of Economics and Business, Evelpidon 47A & Lefkados, Athens 11362 Greece, 7<sup>th</sup> Floor, Office 707 (Monday-Friday 12:00-18:00, tel.: (+30) 210-82.03.860, e-mail: [salexandri@aueb.gr](mailto:salexandri@aueb.gr), Ms. Simone Alexandri).

For clarifications and any other information, interested parties may contact the Secretariat or the Director via e-mail or phone. Information about the program can be found at <http://datascience.aueb.gr/>.

The Program does not discriminate on the basis of race, color, religion, national origin, sex, sexual orientation, gender identity, age or disability. Our nondiscrimination policy applies to all phases of the admission and scholarship process, and to all aspects of the educational programs and activities.

## Program Structure

The Full Time (FT) program is a 1-year program. Students need to complete 60 units of coursework, of which 46 units of core courses and 14 units of electives. After completion of coursework, students undertake an intense 3-month integrated Capstone Project in collaboration with industry, with joint academic and industrial supervision, for 15 units. Attendance of lectures and laboratory sessions is mandatory.

The Part Time (PT) program is a 2-year program. Students need to complete 75 units of coursework, of which 46 units of core courses and 29 units of electives. Students can under special circumstances replace 15 units of coursework with an integrated Capstone Project in collaboration with industry, or a faculty-supervised research thesis, with Director approval. Required classes take place twice a week, 6:00-9:00pm. Attendance of lectures and laboratory sessions is mandatory.

Full courses are worth 5-7 units, half courses are worth 3 units. Before the beginning of classes, students are required to complete 1-3 preparatory courses in Statistics, Mathematics, and Computer Science, as decided by the Admissions Committee. Each course comprises 5 3-hour lectures and a final exam.

## Tuition Fees

### Full-time program: €6000

€2000 payable by October 1<sup>st</sup>, 2018

€3000 payable by February 11<sup>th</sup>, 2019

€1000 payable by April 1<sup>st</sup>, 2019

### Part-time program: €7500

€2000 payable by October 1<sup>st</sup>, 2018

€3000 payable by February 9<sup>th</sup>, 2019

€2500 payable by April 1<sup>st</sup>, 2019

Tuition fees are non-refundable. A limited number of merit-based scholarships is available. More details can be found in the Program Regulations Handbook.

Athens, 23/07/2018

The Head of the Department of Informatics

The Rector

Professor Nikolaos G. Malevris

Professor Emmanouil A. Giakoumakis



## Additional Information - Curriculum

### Core courses:

#### Probability and Statistics for data analysis (6 units)

Basic principles of Probabilities. Basic theorems in Probability e.g. law of large numbers, the Central Limit theorem etc. Common probability distributions. Principles of statistics. Data summarization. Statistical inference and causality, Experimental design and sampling methods, Estimation and hypothesis testing. Bootstrap and variants.

#### Practical Data Science (6 units)

The course gives students a set of practical skills for handling data that comes in a variety of formats and sizes, such as texts, spatial and time series data. These skills cover the data analysis lifecycle from initial access and acquisition, modeling, transformation, integration, querying, application of statistical learning and data mining methods, and presentation of results. (The course is hands-on, using python, in iPython interactive computing framework.)

#### Large Scale Data Management (6 units)

Methods and techniques for database design and management, operational data management and transaction processing, data warehouse creation, and information retrieval. New approaches for storage and querying (column stores, NewSQL) will be discussed and experimented upon. Management of large scale structured and unstructured data in different information systems environments.

#### Machine Learning and Computational Statistics (7 units)

Introduction to the basic ideas of statistical learning models (supervised and unsupervised learning). Model selection, feature selection and cross-validation. Linear regression and logistic regression. Generalized linear models. K-nearest neighbor classification, Bayes and naive Bayes classifiers. Kernel Discriminant Analysis and Support Vector Machines. Unsupervised learning methods. Clustering using k-means and mixtures models. The EM algorithm. Dimensionality reduction using PCA, probabilistic PCA, factor analysis and independent component analysis.

#### Numerical optimization and Large Scale Linear Algebra (6 units)

Floating point arithmetic; Stability of numerical algorithms; Norms; Fundamentals of matrix theory; Solution of systems of linear equations: direct methods, error analysis, structured matrices; Iterative methods for linear equations and least squares; Eigenanalysis; important matrix factorizations and their algorithms. Application to case studies.

#### Data visualization and communication (6 units)

Communicating clearly and effectively about the patterns we find in data is a key skill for a successful data scientist. Visualizations are graphical depictions that can improve comprehension. Collaborative filtering Visualizations will be paired with verbal analyses and reporting. Different tools will be used to transform data and create visualizations, including Python, Google Charts, Tableau, and Spotfire. Assignments will give students experience with reporting on complex patterns and results with graphics and prose.



### **Legal, ethical and policy issues in data science (3 units)**

Discusses issues of privacy, surveillance, security, classification, discrimination and decisional autonomy from a legal, ethical, and policy perspective (whether business or public policy). Areas of relevance include health, marketing, employment, law enforcement, and education.

### **Text engineering and analytics (6 units)**

Language models, text normalization. Applying feature extraction, classification, sequence labeling algorithms (e.g., PCA, naive Bayes, logistic regression, SVMs, HMMs, CRFs) to texts (for document classification, entity recognition etc.). Parsing (CKY, Earley, probabilistic CFGs). Semantics (logic-based, distributional, word embeddings, sense disambiguation) and discourse analysis (co-reference, rhetorical relations). Machine translation. Information extraction (incl., relation extraction) and sentiment analysis. Question answering. Text summarization. Concept-to-text generation. Speech recognition fundamentals.

### **Electives (indicative list):**

#### **Data mining (6 units)**

Data-oriented techniques for extracting patterns from data. Association rules, decision trees. Collaborative filtering and recommendation algorithms Finding similar items and frequent items. Mining data streams. Mining social network graphs. Mining for Web advertising. Implementing machine learning schemes.

#### **Bayesian Statistics and simulation methods (6 units)**

Bayesian inference. Simulation and random number generation. Markov models and hidden Markov models. Probabilistic graphical models. Bayesian statistical methods, Markov chain Monte Carlo, Metropolis-Hastings algorithm, Gibbs sampling, sequential Monte Carlo methods, approximate Bayesian computation.

#### **Advanced Large Scale Data Management (5 units)**

Distributed and parallel data-oriented computation and transaction processing. Integration and management of large scale structured and unstructured data in different information systems environments.

#### **Big Data Systems and techniques (3 units)**

Cloud services, engineering issues, stream processing, graph processing, Cassandra, Dremel, Pregel, Storm, parallel data mining systems (Graph Lab, Mahout).

#### **Statistical methods for Big data (3 units)**

Small n large p problems, regularizations, model and variable selection techniques, LASSO, elastic net. Multiplicity. Graphical Models. Techniques for sparse matrices and graphical LASSO. Compressed sensing.

#### **Time series and Forecasting methods (3 units)**

Basic principles, autocorrelation and autocovariance, Holt-Winters method, AR, ARMA, ARIMA models. Regression models, ARCH – GARCH, volatility models.



### **Optimization (5 units)**

Convex and semidefinite optimization (Convex sets and functions, Problems, duality, unconstrained and constrained minimization), Combinatorial optimization (Branch and bound, tabu search, Simulated annealing), Multivariate function optimization (e.g. gradient descent). Linear Programming (Formulations, Algorithms).

### **Advanced customer analytics (6 units)**

Overview of data mining techniques for sales and marketing: clustering, classification, dimensionality reduction, sequence modeling. Techniques for Customer Segmentation. Churn management. Cross-/Up-sell Campaign Targeting. Next Best Action. Marketing Mix optimization. Omni-Channel Optimization. Loyalty Analytics. Basket Analysis

### **Data Science for medicine (3 units)**

Introduction to epidemiological methods: bias, confounding, sample size. Survival analysis: hazard functions, parameter inference. Methods for categorical data. Analysis of contingency tables, risk assessment in retrospective and prospective studies

### **Information retrieval (3 units)**

Text vocabulary, automatic indexing, inverted files, fast inversion algorithm, index compression. Evaluation of information retrieval systems. Information retrieval models (Boolean model, vector space model, probabilistic retrieval model), latent semantic indexing. Computing scores, result ranking. Crawling. Link analysis. Search engine architecture and systems issues.

### **Recommender Systems (3 units)**

Neighborhood-based collaborative filtering. Model-based collaborative filtering. Content-based recommender systems. Knowledge-Based recommender Systems. The cold-start problem. Direct vs implicit signals.

### **Data Science Challenge (5 units)**

Introduction to the DS methods including data preprocessing, feature selection & engineering, machine learning, graph/text mining and visualization. Introduction to a specific data challenge and its domain specificities presented as a Kaggle competition. The best solutions are presented to the class.

### **Introduction to Quantitative Finance and Financial Risk Management (5 units)**

Basic financial instruments and associated fundamental concepts: time value of money, interest rates and fixed income securities; Simple derivatives: Futures, Forwards and Interest Rate Swaps; Options and the Black-Scholes framework. Statistical measures and error metrics of different distributions. Value at Risk (VaR), Expected Shortfall; Methodologies for VaR calculation; Credit risk and the Basel II capital requirements.





## ACADEMIC CALENDAR 2018-2019

### PREPARATORY COURSES (4 weeks)

START	MONDAY	3/9/2018
END	FRIDAY	28/9/2018

### 1st TEACHING PERIOD

START	MONDAY	1/10/2018
END	FRIDAY	7/12/2018

### EXAMS

START	MONDAY	17/12/2018
END	FRIDAY	21/12/2018

### WINTER HOLIDAY BREAK

START	MONDAY	24/12/2018
END	FRIDAY	4/1/2019

### 2nd TEACHING PERIOD

START	MONDAY	7/1/2019
END	FRIDAY	15/3/2019

### EXAMS

START	FRIDAY	22/3/2019
END	FRIDAY	29/3/2019

### 3rd TEACHING PERIOD

START	TUESDAY	1/4/2019
END	FRIDAY	14/6/2019

### SPRING HOLIDAY BREAK

START	THURSDAY	25/4/2019
END	WEDNESDAY	1/5/2019

### EXAMS

START	MONDAY	24/6/2019
END	FRIDAY	29/6/2019