

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6001 INTRODUCTION TO BIG DATA

3 US CREDITS

The course addresses students, who are new to Data Science and interested in understanding the Big Data techniques. It provides the opportunity to gain expertise with the terminology and the core concepts behind big data problems, applications, and systems. Along with the theoretical aspects, the course provides hands-on experience with a distributed processing system, as well as, a brief introduction to NOSQL databases.

ITC 6002 EXPLORING AND ANALYZING DATA

3 US CREDITS

The course focuses on procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of statistics which apply to analyzing data. It also examines probabilistic inference for generative models of inference together with standard techniques in pattern recognition. Topics covered include: Descriptive and inferential statistics, sampling, probability spaces, queuing theory, stochastic processes, mathematical modeling, experimental design, parametric and non-parametric tests, regression, clustering, Markovian and Bayesian networks.

ITC 6003 APPLIED MACHINE LEARNING

3 US CREDITS

The course provides a broad introduction to the key ideas of machine learning, a rapidly growing field which resides at the intersection of computer science and statistics, and is concerned with finding useful patterns in data. Emphasis is given on intuition and practical examples, which covers a wide range of real life implementations, including personalized product recommendations and natural language processing.

ITC 6004 DATA VISUALIZATION

3 US CREDITS

The course combines the science of data visualization with the art of graphic design, and introduces ways to accurately and effectively communicate complex information. Students are exposed to techniques of presenting complex ideas in easily accessible and understandable manners, by transforming data into visual graphics, such as charts, bar graphs, scatterplots, and heatmaps. The course also offers hands-on experience through exercises, which allow students to explore the types of data in use today, learn how people perceive different graphical displays, and create visual presentations that foster impact on the audiences.

ITC 6005 SEMANTIC WEB

3 US CREDITS

The course provides a broad introduction to core ideas of the Semantic Web layer, an intellectual milestone for the evolution of the web and the management of data from heterogeneous sources, based on "The Web of Meaning" vision, which sets the basis for significant developments in technical prerequisites and business requirements. Students are exposed to a multidimensional overview of the constitutional elements of the semantic web, namely: RDF, OWL, Query, Logic, Proof, Trust. The emphasis is on the integration of core semantic web ideas to real world problems and on the capacity of students to vision real world application by adopting ontological engineering and extensive reasoning capabilities in data.

ITC 6006 BIG DATA IN BUSINESS

3 US CREDITS

The course provides a broad qualitative introduction to Big Data and Data Analytics, a paradigm in Web Science, which enables the transformation of massively produced raw data into knowledge and wisdom capable of supporting smart decision making, innovative services, new business models, innovation and entrepreneurship. Students are exposed to the business potential of related research towards technology driven innovation and sustainable development. Emphasis is given to the integration of Big Data Research to Innovation and Entrepreneurship, with well-structured and justified business plans for real world startups, in several domains of human activity including but not limited to Health, Bioinformatics, Life Sciences, Entertainment, Education, Public Service, Government, Security, Finance, Sustainability, etc.

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6107 STORING AND RETRIEVING DATA

3 US CREDITS

Prerequisites

ITC 6001 INTRODUCTION TO BIG DATA

This course prepares students to deal with large-scale collections of data as objects to be stored, searched over, selected, and transformed for use. Emphasis is placed both on the background theory and the practical application of information retrieval, as well as, database design and management, data extraction, transformation and loading for data warehouses, and operational applications.

ITC 6008 SEARCH ENGINES AND WEB MINING

3 US CREDITS

This course provides a comprehensive introduction to the theory and implementation of algorithms for organizing and searching large text collections. The first half of the course examines text search engines for enterprise and web environments; an open-source engine such as Indri can be used as a working example. The second half of the course explores text mining techniques such as recommender systems, clustering, and categorization. The course strives for a hands-on experience in document ranking, evaluation, and classification into browsing hierarchies, as well as other related topics.

ITC 6009 MACHINE VISION IN DATA SCIENCE

3 US CREDITS

This course emphasizes large-scale machine learning and large-scale vision in a combined cross-talk between the two fields. The goal is to enlighten machine learning techniques on large-scale vision problems, to inform about new developments on large-scale learning, and to identify unique challenges and opportunities in the combined fields.

Computer Vision has become ubiquitous in our society, with applications in search, image understanding, apps, mapping, medicine, drones, self-driving cars etc. Core to many of these applications are visual recognition tasks such as image classification, localization and detection. At the same time, the emergence of "big data" has brought a paradigm shift throughout computer science. Computer vision is no exception. The explosion of images and videos on the Internet and the availability of large amounts of annotated data have created unprecedented opportunities and fundamental challenges on scaling up computer vision. Over the past few years, machine learning on big data has become a thriving field with a plethora of theories and tools developed. Meanwhile, large-scale vision has also attracted increasing attention in the computer vision community.

ITC 6010 NATURAL LANGUAGE PROCESSING

3 US CREDITS

The course explores fundamental concepts and ideas in natural language processing (NLP), otherwise known as computational linguistics, and aims to develop an in-depth understanding of both algorithms for processing linguistic information and the underlying computational properties of natural languages. Students are exposed to word-level, syntactic, and semantic processing from both a linguistic and an algorithmic perspective, led by current research and tools in the field. The course focuses on modern quantitative techniques in NLP -- using large corpora, statistical models for acquisition, disambiguation, and parsing -- and the construction of representative systems. Topics to cover are: natural language understanding, morphological processing, lexicon, tagger, converter, parser, word sense disambiguation, word and sentence embeddings, deductive approaches to interpretation, machine translation and language acquisition.

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6420 SAS PLATFORM FOR BUSINESS ANALYTICS 3 US CREDITS

Prerequisites

- ITC 6001 INTRODUCTION TO BIG DATA
- ITC 6002 EXPLORING AND ANALYZING DATA
- ITC 6003 APPLIED MACHINE LEARNING
- ITC 6004 DATA VISUALIZATION AND COMMUNICATION

The course aims to provide practical skills in Business Analytics and Data mining with SAS tools. In particular, it covers Data Management using the SAS Enterprise Guide, Statistical Analysis, Data Mining using the SAS Enterprise miner and Visualization using the SAS Visual Analytics.

In addition, the course aims at a Joint Certificate in Business Analytics and Data Mining that will be provided by SAS after the successful completion of a separate examination.

ITC 6421 CAPSTONE PROJECT

3 US CREDITS

Prerequisites

- ITC 6107 STORING AND RETRIEVING DATA
- ITC 6008 SEARCH ENGINES AND WEB MINING
- ITC 6420 SAS PLATFORM FOR BUSINESS ANALYTICS
- ITC 6109 MACHINE VISION IN DATA SCIENCE

The capstone project is undertaken by students working individually or collaboration and in consultation with an instructor who acts as their capstone advisor. The capstone project must convey a sound solution to a practical problem. The capstone project is a significant analytical, design and implementation piece work and provides an opportunity for students to draw on their methodological, analytical and substantive learning in a comprehensive written study in the field of data mining and big data.

ITC 6230 ADVANCED MACHINE LEARNING

3 US CREDITS

The course provides exposure to advanced techniques that extract useful information in the presence of incomplete, or noisy data. Also, there is emphasis on not only predicting a single value but a sequence of values. Finally, distributed techniques are considered for scaling-up the methods to the needs of real world problems. In particular, the course provides a background on Bayesian methods, Markov models and distributed processing.

ITC 6440 THESIS

3 US CREDITS

Prerequisites

- ITC 6008 SEARCH ENGINES AND WEB MINING
- ITC 6230 ADVANCED MACHINE LEARNING
- ITC 6107 STORING AND RETRIEVING DATA
- ITC 6109 MACHINE VISION IN DATA SCIENCE

The thesis is undertaken by students working individually and in consultation with an instructor who acts as their thesis advisor. The thesis must convey scholarly and/or professional analysis informed by appropriate application of methodology. The thesis is a significant analytical piece of work and provides an opportunity for students to draw on their methodological, analytical and substantive learning in a comprehensive written study in the field of data mining and big data.