

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 are interested in understanding Big Data management and analysis techniques. Prior programming experience in Python is necessary to successfully complete the course.

Along with the theoretical aspects, the course provides hands-on experience with programming language libraries for data management, and query languages for relational 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, stochastic processes, mathematical modeling, experimental design, clustering, hypothesis testing, forecasting, parametric and non-parametric tests, time series analysis, regression analysis and parametric modeling.

ITC 6103 APPLIED MACHINE LEARNING

3 US CREDITS

Prerequisites ITC 6001 Introduction to Big Data

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 KNOWLEDGE GRAPHS

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.

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6107 BIG DATA ARCHITECTURES

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.

ITC6045 Project Management in Data Science

3 US CREDITS

The course provides the fundamentals in Project Management (PM) focused on Data Science. The course combines planning, organizing, resourcing and monitoring all the steps of a Data Science project. Students are exposed to PM techniques of managing complex Data Science projects and better understand how companies manage those projects considering various stakeholders needs, financial constraints, etc. The course also offers hands-on experience through exercises.

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 problem segmentation in the field of machine vision. The goal is to master processing techniques and enlighten machine learning techniques on vision problems, to inform about new developments on 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. 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.

ITC 6110 NATURAL LANGUAGE PROCESSING

3 US CREDITS

Prerequisites: ITC6003 Applied Machine Learning

The course explores fundamental concepts and ideas in natural language processing (NLP), and develops 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.

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6119 BUSINESS FORECASTING

-3 US CREDITS

Prerequisites:

ITC 6003 APPLIED MACHINE LEARNING

Forecasting is a critical process that is conducted by organizations across all industries and sectors, such as telcos, manufacturing, consumer packaged goods, retail, pharmaceutical, government and related institutions, banks, and others. The above organizations, in order to become more efficient and more effective in their operations, forecast every day various measures such as interest rates, demand, energy consumption, GDP, revenues, network traffic, exchange rates and many more. This course aims to teach students fundamental concepts about quantitative forecasting (time series and econometric methods) and how these can be applied in practice using state of the art software i.e., Python and SAS Forecast Server.

ITC 6230 DEEP LEARNING

3 US CREDITS

Prerequisites:

ITC 6002 EXPLORING AND ANALYZING DATA

ITC 6003 APPLIED MACHINE LEARNING

The course focuses on Deep Neural Network Cell architectures. Deep learning is a branch of machine learning concerned with the development and application of modern neural networks. Deep learning algorithms using modern NN architectures extract layered high-level representations of data in a way that maximizes performance on a given task. We will cover their theory, their applications, and their training.

In particular, the course provides a background on NN architectures starting with Feed Forward and Recursive Neural Networks (FFNN, RNN) and the derivation of Backpropagation as an organized computation of Gradient Descent for their training from data. We address limitations stemming from the vanishing and exploding gradients problems. Then we move to more complex architectures: Convolutional Neural Networks (CNN), a variety of Sequence modeling NN like Long Short Term Memory NN (LSTM), Gated, Gated Recurrent Units (GRU), Transformers, Reformers. We also discuss Attention mechanisms to increase long term learning in Sequence NN's. We revisit Back Propagation and explain how the latest NN cell architectures solve or mitigate the vanishing/exploding gradients problem of traditional RNN's. We finally cover optimization techniques adapted to NN training, hyperparameter tuning and explore Python libraries like Pytorch and Tensorflow for efficiently program and train NN's.

ITC 6125 MACHINE LEARNING AND APPLICATIONS

3 US CREDITS

Prerequisites:

ITC 6003 APPLIED MACHINE LEARNING

Machine learning algorithms are data analysis methods, which search data sets for patterns and characteristic structures. Typical tasks are the classification of data, automatic regression and unsupervised model fitting. The course will cover methods from a variety of related subjects including Statistics, applied mathematics and more specialized fields, such as pattern recognition and neural computation. Those methods will cover image and speech analysis, medical imaging, bioinformatics, and exploratory data analysis.

MASTER OF SCIENCE (MS) IN DATA SCIENCE

ITC 6421 CAPSTONE PROJECT

3 US CREDITS

Prerequisites:

ITC 6107 BIG DATA ARCHITECTURES

ITC 6008 SEARCH ENGINES AND WEB MINING

ITC 6119 BUSINESS FORECASTING OR ITC 6230 DEEP LEARNING

ITC 6009 MACHINE VISION IN DATA SCIENCE

The capstone project is undertaken by students working individually or in 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 6440 THESIS

3 US CREDITS

Prerequisites:

ITC 6107 BIG DATA ARCHITECTURES

ITC 6008 SEARCH ENGINES AND WEB MINING

ITC 6230 DEEP LEARNING

ITC 6009 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.