Big Data

Predictive Analytics for Big Data

Lecturer: Lucrezia Noli

Language

English

Course description and objectives

Big Data is the hot new buzzword in IT circles. The proliferation of digital technologies with digital storage and recording media has created massive amounts of diverse data, which can be strategically exploited by companies in all sectors. Big Data, which can take up terabytes and petabytes of storage space in diverse formats including text, video, sound, images, and more, is actually analyzed with the same methods as more traditional data are, with the only difference being the tools required to support the specific characteristics defining Big Data.

The course gives an overview of the Big Data phenomenon, focusing on how to extract value from Big Data using predictive analytics techniques.

Upon successful completion of this course, students should be able to:

- Understand the Big Data phenomenon, and the differences between Big Data and Traditional Data
- Understand the potential use of Data in a corporate environment
- Understand the use of Predictive Analytics tools to be used for extracting valuable information from raw data.

Audience

The course is open to all Bocconi students. In particular it is targeted at:

- All those who want to understand what Big Data really is, and how to exploit it
- All those who want to gain practical, analytical skills and confidence with the Data Scientist Toolkit

Prerequisites

Basic statistical knowledge: univariate and bivariate descriptive statistics.

A basic knowledge of programming languages such as R or Python is a plus.

Duration

12 hours

Calendar

Lecture	Date	Time	Room
1	Mon 01/04/2019	18.00 - 19.30	Info 6
2	Thu 04/04/2019	18.00 - 19.30	Info 6
3	Wed 10/04/2019	18.00 - 19.30	Info AS05
4	Sat 13/04/2019	09.30 - 12.45	Info 6
5	Fri 03/05/2019	18.00 - 19.30	Info 6

Syllabus of the course

Lecture	Topics
1	**Introduction to Big Data and Predictive Analytics** - Big data: definition and taxonomy - Predictive Analytics: definitions and the CRISP-DM methodology *Setting up the Knime environment*
2	**Business Understanding and Data Preparation** - CRISP-DM: how to efficiently create a Predictive Analytics Model - Data preparation: the most important task of Data Scientists - Data exploration: understanding the dataset *Exercise*

2

Università
Bocconi
MILANO

Lecture	Topics
3	**Predictive Analytics Techniques**
	- Supervised vs. unsupervised / black-box vs. rule-based models
	- Predictive Analytics Applications: when to use which model
	- Predictive Analytics Algorithms: characteristics and taxonomy
	Exercise
4	**Predictive Analytics Algorithms**
	- From the Linear Regression to the Neural Network
	- From Neural Networks to Deep Learning
	Exercise
5	**Classification Algorithms**
	- Data preparation for Classification Tasks
	- Setting up a Classification Model
	- Model performance & evaluation
	Exercise
6	**Text analytics**
	- Text analytics
	- Natural Language Processing
	- Sentiment analysis
	Exercise

Software used

Knime, R, Rstudio

Suggested bibliography

A. Rezzani, Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati, Apogeo Education, 2013

Karau H., Konwinski A., Learning Spark: Lightning-Fast Big Data Analysis, O'Really, 2015

Available seats

This activity is limited to **110** participants. Registrations cannot be carried out once this number has been reached or after closing of the registration period.