



UNIVERSITÀ DI PAVIA

Anno Accademico 2020/2021

BIG DATA ANALYSIS

Anno immatricolazione	2020/2021
Anno offerta	2020/2021
Normativa	DM270
SSD	SECS-S/01 (STATISTICA)
Dipartimento	DIPARTIMENTO DI SCIENZE ECONOMICHE E AZIENDALI
Corso di studio	INTERNATIONAL BUSINESS AND ENTREPRENEURSHIP - MANAGEMENT INTERNAZIONALE E IMPRENDITORIALITÀ
Curriculum	Digital Management
Anno di corso	1°
Periodo didattico	Primo Semestre (28/09/2020 - 22/12/2020)
Crediti	9
Ore	66 ore di attività frontale
Lingua insegnamento	English
Tipo esame	SCRITTO E ORALE CONGIUNTI
Docente	CERCHIELLO PAOLA (titolare) - 9 CFU
Prerequisiti	Knowledge of basic concepts of Statistics like inference, confidence interval, test of hypothesis, simple regression model. For the coding part, it is recommended to have some familiarity with Python language, or similar software like R, even through on-line courses (ex. Coursera).
Obiettivi formativi	Knowledge of the most relevant statistical methods for large data set analysis. The student will learn how to run a complete work flow of analysis by employing Python software. From the collection and management of large data set, through the choice of the most appropriate models to the final interpretation and contextualization of the results.
Programma e contenuti	The aim of this course is to study and apply the most relevant statistical models in the analysis of large data set.

	<p>The perspective is mainly applicative: choosing and applying suitable models to exploit the whole informative content of (large) data set with a particular attention to the correct and contextualized interpretation of the final results. Moreover, a focus will be set on some frameworks for the management of large data set like MapReduce for data clustering.</p> <p>The course will be held with the interactive employment of open source software like Python to learn practically the complete analysis work-flow. A particular emphasis will be given to social network data, textual data, business-financial case studies.</p> <p>Some of the models that will be covered are: Naive Bayes Classifier, Latent Dirichlet Analysis, Clustering algorithm, Penalized regression Support Vector Machines.</p>
Metodi didattici	<p>The class integrates theoretical lectures with practicals based on Python to learn how to implement and analyze the most appropriate models according to the available data.</p> <p>A tutor will help students weekly in acquiring all the necessary theoretical and practical knowledge.</p> <p>All the material used during the lectures (slides, script, data) will be available on Kiro platform.</p>
Testi di riferimento	<p>1) Coelho and Richert, Building Machine Learning Systems with Python, Second edition, Packt Publishing.</p> <p>2) Introduction to Python for Econometrics, Statistics and Data Analysis Kevin Sheppard, pdf version available</p>
Modalità verifica apprendimento	<p>Midterm about the first explained models.</p> <p>Final project of analysis with real case studies.</p>
Altre informazioni	<p>Some lectures will be thought by experts of the big data field.</p>
Obiettivi Agenda 2030 per lo sviluppo sostenibile	<p>Gli obiettivi</p>