UNIVERSITÀ DI PAVIA

Anno Accademico 2020/2021

DATA SCIENCE AND BIG DATA ANALYTICS

Anno immatricolazione	2020/2021
Anno offerta	2020/2021
Normativa	DM270
SSD	ING-INF/05 (SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI)
Dipartimento	DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
Corso di studio	INDUSTRIAL AUTOMATION ENGINEERING - INGEGNERIA DELL'AUTOMAZIONE INDUSTRIALE
Curriculum	PERCORSO COMUNE
Anno di corso	1°
Periodo didattico	Secondo Semestre (08/03/2021 - 14/06/2021)
Crediti	6
Ore	60 ore di attività frontale
Lingua insegnamento	English
Tipo esame	ORALE
Docente	NOCERA ANTONINO (titolare) - 6 CFU
Prerequisiti	Basic Knowledge on database management systems and on data manipulation.
Obiettivi formativi	This course aims at identifying and provide an in-depth of the main tools and systems for mining big data and, more in general, for working in the current scenario of data science. At the end of this course the student must be able to: understand data, formulate hypothesis on data, transform and model data, define a methodology for the analysis, test and confirm results. The student will learn basic concepts of cloud computing along with the best practice to being able to operate in this context. He will learn the main architectures for processing big-datasets, such as MapReduce. He will understand how to work with NoSQL databases to face the main

	issues in the big data realm.
Programma e contenuti	The lifecycle of a data science project. Python libraries for data manipulation and data preparation. Using the main data mining techniques with Python. Overview of the R language as a tool for data science and data mining. The Big Data paradigm and the main issues in this context. Main Cloud architectures for Big Data. Open Stack as an example of open source solution to Cloud Computing. Hadoop, HDFS and Map Reduce. Notes on Apache Spark. NoSQL Databases and MongoDB. Social Network Analysis as a case study.
Metodi didattici	This course is organized in lectures, laboratory and cooperative learning. Lectures are used to present theoretical concepts and all the notions about this course. During the lectures, the student will also understand how to apply these notions. Laboratory is used as a mean to allow the student to apply the concepts and techniques shown during the lectures to real-world case studies. Finally, this course leverages cooperative learning, group working and brainstorming. This allows for the development of many transversal skills, such as: team working capabilities, conflict management, and the capability to acquire and exploit different ideas from a team.
Testi di riferimento	1. Data Science and Big Data Analytics - Discovering, Analyzing, Visualizing and Presenting Data. Wiley. 2. Big Data Fundamentals – Concepts, Drivers & Techniques. Prentice Hall, 2015. 3. Data Mining - Practical Machine Learning Tools and Techniques. Elsevier. 4. Data Mining - Concept and Techniques. Elsevier. 5. Notes provided by Professor
Modalità verifica apprendimento	The assessment consists of an oral discussion about a group project work each student is involved in. The student is in charge of preparing a report about his project work. During the oral discussion the report presented by the student will bel used as a mean to go in-depth in the theoretical concepts used therein. To prepare the report, the student will have to use the tools, introduced during lectures, to extract knowledge form real-life datasets. During the assessment the student must prove a good knowledge of the

	main concepts introduced in this course, to be able to handle the lifecycle of a data science project and to know the main architectures, tools and NoSQL solutions to work in the data analytics and Big Data contexts. The assessment will carefully consider the level of expertise in the use of the tools, the ability of the student to build projects adopting these tools, the level of understanding of the notions taught in this course, the methodological rigor and appropriateness of the technical vocabulary.
Altre informazioni	
Obiettivi Agenda 2030 per lo sviluppo sostenibile	$lbl_legenda_sviluppo_sostenibile