Libera Università Internazionale
degli Studi Sociali Guido Carli

LUISS

Syllabus

Academic Year	2021/2022
Program	Data Science and Management
course	Big Data and Smart Data Analytics
Term	I semester
Year	2
SSD	SECS-S/01
Credits	6

INSTRUCTIONAL GOALS	As datasets grow to Petabyte scale, traditional analysis models and computation paradigms become obsolete. The course will focus on fundamental algorithmic, statistical, and programming issues posed by big-data analytics, tackling major problems and techniques for extracting knowledge from massive amounts of data. By the end of the course the students will gain an understanding of the theory and computing of modern methods for big data analytics, with particular emphasis on advanced statistical methods and algorithms for mining massive and noisy datasets as well as rapidly changing streams of data.
INTENDED LEARNING OUTCOMES They describe what a learner is expected to know, understand and be able to demonstrate after completion of a learning path.	**Knowledge and understanding**: Upon successful completion of the course, the students will be familiar with data mining problems and techniques, advanced statistical methods, as well as computational models and frameworks for analyzing and extracting insights from massive, possibly distributed or rapidly changing amounts of data at a large scale. **Applying knowledge and understanding**: After this course, the students will be able to proficiently develop innovative big data solutions, based on sound statistical and algorithmic techniques, in different application domains. They will be also able to implement the proposed solutions on top of industry-standard frameworks, e.g., Hadoop and Spark, in order to tackle real-world problems such as those typically faced by big tech companies. **Making judgements**: Throughout the entire course, students will be invited to assess critically strengths and weaknesses of all different methods and tools presented in class. After this course, they will be able to analyze different solutions to big data problems and to demonstrate an in-depth, critical understanding of the scope and challenges of different data-driven analytics techniques. **Communications Skills**: This course will give the students the possibility to acquire and to understand major terms and concepts so as to communicate effectively their ideas, findings, proposals, analysis, and critical reasoning in the area of data-driven analytics. A special emphasis will be given to oral presentations and pitches in project group works. **Learning skills**: This course will provide the students with the ability to learn cutting-edge design and analysis tools and to apply them to real-world data analytics problems. The method of study will make the students able to break down complex problems arising in specific applications into manageable pieces and to apply different patterns in order to design rigorous and documentable solutions. A strong emphasis will be given to the direct application of the techniques and tools covered in this course to complex problems that are typical of today's data-driven industry.
Pre-requisites	Basic knowledge of probability and statistics, fundamental algorithms, and computer programming skills (preferably using the Python programming language).

Course content	• Defining big data: the five V's (volume, velocity, variety, veracity, and value). From big data to actionable insights: smart data. • Data sources. Statistical features of big data. Selectivity bias, coverage, missing data, noise. • Pattern discovery: similarity search, dimensionality reduction, association rules and causality analysis. • Sampling and estimators: traditional approaches, data stream algorithmics. • Statistical algorithms for predictive analytics. Recommender systems. • Analytics at scale: open-source frameworks for analyzing big data
Reference Books	Recommended books: • Mining of Massive Datasets. Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman. Second edition. • The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Second edition. Lecture notes, research papers, and other course material will be made available on the e-learning platform.
Teaching Methods	The lectures will follow an experimental and problem-driven approach and will be complemented by hands-on programming sessions, that will guide the students on the use of good analytics practices and industry-standard programming frameworks. The goal for the class is to be broad and to touch upon a variety of statistical and algorithmic techniques, introducing standard practices as well as cutting-edge research topics in this area.
Assessment	There will be a written midterm exam, a written final exam and a project. In the midterm and final exams students are required to demonstrate that: • they have acquired a deep understanding on how to develop innovative solutions for big data problems; • they are able to analyze different solutions to big data problems; • they have acquired an in-depth, critical understanding of the scope and challenges of different techniques for big data problems. Midterm and final will count for 30% of the grade each. Students that will not take the midterm and final during the course are required to take an oral exam after the course, where they are required to demonstrate the same skills described above. In the project students are required to demonstrate that they are able to: • break down complex problems arising in specific applications into manageable pieces and to apply different patterns in order to design rigorous and documentable solutions; • implement their big data solutions on top of industry-standard frameworks (e.g., Hadoop and Spark); • analyze and assess critically strengths and weaknesses of different tools and techniques for big data problems. The project will count for 40% of the grade. The overall assessment will take into account the level of knowledge and understanding of techniques for approaching big data problems acquired by the students; their capacity for thinking creatively, innovatively and critically; their capacity to design and evaluate solutions for big data problems, making critical judgements about these; and their capacity to present effectively findings and conclusions.