UNIVERSITÀ DI PISA

BIG DATA ANALYTICS

LUCA PAPPALARDO

Anno accademico	2020/21
CdS	DATA SCIENCE AND BUSINESS INFORMATICS
Codice	599AA
CFU	6

Moduli	Settore/i	Tipo	Ore	Docente/i
BIG DATA ANALYTICS	INF/01	LEZIONI	48	FOSCA GIANNOTTI LUCA PAPPALARDO

Obiettivi di apprendimento

Conoscenze

In our digital society, every human activity is mediated by information technologies, hence leaving digital traces behind. These massive traces are stored in some, public or private, repository: phone call records, movement trajectories, and social media records are all examples of Big Data, a novel and powerful "social microscope" to understand the complexity of our societies. The analysis of big data sources is a complex task, involving the knowledge of several technological and methodological tools. This course has three objectives:

1. introducing to the emergent field of big data analytics and social mining;

2. introducing to the technological scenario of big data, like programming tools to analyze big data, query NoSQL databases, and perform predictive modeling;

3. guide students to the development of a reproducible big data analytics project, based on the analysis of big real-world datasets.

Modalità di verifica delle conoscenze

The assessment of the course consists of: (i) the presentation of a paper in the literature showing a real-world example of big data analytics project; (ii) the development, during the course, of a big data analytics project in team with other students; (iii) an oral exam where the student describes the developed project, the motivations behind it and the results obtained. During the oral exam the student must be able to demonstrate knowledge of the course contents and be able to discuss the topics thoughtfully and with propriety of expression.

Capacità

The student will be able to develop complex big data analytics projects, i.e., to pre-process and clean big and complex data, to perform predictive analytics, and to interpret and communicate the results of predictive models.

Prerequisiti (conoscenze iniziali)

Good knowledge of Data Mining, Machine Learning, Databases and Programming.

Programma (contenuti dell'insegnamento)

Module 1: Big Data Analytics and Social Mining
In this module, analytical methods and processes are presented thought exemplary cases studies in challenging domains, organized according to the following topics:

- The Big Data Scenario and the new questions to be answered
- Sport Analytics:
    1. Soccer data landscape and injury prediction
    2. Analysis and evolution of sports performance
- Mobility Analytics
    1. Mobility data landscape and mobility data mining methods

      2. Understanding Human Mobility with vehicular sensors (GPS)
      3. Mobility Analytics: Novel Demography with mobile-phone data
- Social Media Mining
      1. The social media data landscape: Facebook, Linked-in, Twitter, Last_FM
      2. Sentiment analysis. example from human migration studies
      3. Discussion on ethical issues of Big Data Analytics
- Well-being&Now-casting
      1. Nowcasting influenza with retail market data
      2. Predicting well-being from human mobility patterns
- Paper presentations by students

Module 2: Big Data Analytics Technologies
This module will provide to the students the technologies to collect, manipulate and process big data. In particular the following tools will be presented:

- Python for Data Science
- The Jupyter Notebook: developing open-source and reproducible data science
- MongoDB: fast querying and aggregation in NoSQL databases
- GeoPandas: analyze geo-spatial data with Python
- Scikit-learn: programming tools for data mining and analysis
- M-Atlas: a toolkit for mobility data mining

Module 3: Laboratory for Interactive Project Development
During the course, teams of students will be guided in the development of a big data analytics project. The projects will be based on real-world datasets covering several thematic areas. Discussions and presentation in class, at different stages of the project execution, will be performed.

- Data Understanding and Project Formulation
- Mid Term Project Results
- Final Project results

Bibliografia e materiale didattico

Several research papers will be provided to the aim of discuss new trends and developments on Big Data application scenarious. Some basilar white papers and reference books are the followings

1. F Giannotti, D Pedreschi, A Pentland, P Lukowicz, D Kossmann, J Crowley, D Helbing. **A planetary nervous system for social mining and collective awareness**. The European Physical Journal Special Topics 214 (1), 49-75, 2012

2. M Batty, KW Axhausen, F Giannotti, A Pozdnoukhov, A Bazzani, M Wachowicz. **Smart cities of the future.** The European Physical Journal Special Topics 214 (1), 481-518, 2012

Agrawal et al. Challenges and Opportunities with Big Data 2011-1 (2011). Cyber Center Technical Reports. Paper

1. http://docs.lib.purdue.edu/cctech/1

Data, data everywhere. The Economist, Special Report on Big Data, February 2010.

Data Science for Business -- Foster Provost, Tom Fawcett, Publisher: O'Reilly Media

SOCIAL MEDIA E SENTIMENT ANALYSIS L'EVOLUZIONE DEI FENOMENI SOCIALI ATTRAVERSO LA RETE Ceron Andrea; Curini Luigi; Iacus Stefano 2014

Selection of papers of applied data science project will available on course website for student presentations and interactive discussions.

Following technologies will be introduced and used within the project development:

1. The Jupyter Notebook: for developing open-source and reproducible data science
2. MongoDB: fast querying and aggregation in NoSQL databases
3. GeoPandas: analyze geo-spatial data with Python
4. Scikit-learn: programming tools for data mining and analysis
5. M-Atlas: a toolkit for mobility data mining

Ultimo aggiornamento 26/07/2021 15:04