HON 385-01                                    2019 Spring
Term
(Is Big Data Enough? A Conceptual Exploration of Data Science)

Instructor: Dr. Domenico Napoletani
E-mail: napoleta@chapman.edu
Office Hours: Mondays and Wednesdays, 12:00-12:45 and 14:30-15:15, or by
appointment. DeMille Hall, room 150E.

Catalog Description: Prerequisite: acceptance to the University Honors Program, or
consent of instructor. (Offered as needed.) 3 credits. In this course we will explore the
computational, mathematical and philosophical concepts underpinning the use of large
collections of data to solve problems. We will ask whether it is possible to preserve a role
for our reason, when so much of what we understand and what we decide is ultimately
shaped by data-driven algorithms.

Course Learning Outcomes:

By the end of this course, students will have:

 Obtained a structured knowledge of philosophical, mathematical and computational
concepts relating to data collection and analysis.
 Developed the ability to rigorously and creatively analyze and assess the core algorithms
and ideas of data science.
 Sharpened their research skills by exploring how the reliance on large and unstructured
collections of data has an impact on what it means to understand natural phenomena.

Honors Program Learning Outcomes:
Upon completing a course in the University Honors Program students will have:
 a. Obtained a starting point for integrative exploration of the development of
cultures and intellectual achievements through a variety of disciplinary and
interdisciplinary perspectives;
 b. Sharpened their ability to critically analyze and synthesize a broad range of
knowledge through the study of primary texts and through engagement in active
learning with fellow students, faculty, and texts (broadly understood);
c. Understood how to apply more integrative and interdisciplinary forms of understanding in the advancement of knowledge and in addressing complex challenges shaping the world;
d. Developed effective communication skills, specifically in the areas of written and oral exposition and analysis.

Content:

The course will be comprised of two parts: an introduction to the core ideas and methods of data science; and a philosophical analysis of its methodology.

We will introduce the mathematical ideas that are at the basis of data science, i.e. the use of algorithms to solve problems on the basis of intensive data collection. The focus will be on understanding the fundamental ideas on which most data science methods are based, and on being able to decide which specific method is relevant for a given problem.

We will explore the ways models can be fit on data, and the key role of optimization in finding such models. The techniques and ideas we will introduce are exemplary of data science and include: linear regression, logistic regression and support vector machines; regularization techniques and the curse of dimensionality; nearest neighbor methods, hierarchical and k-means clustering; neural networks and deep learning.

However, rather than considering data science as a collection of loosely related techniques, we will take the view that there is a coherent set of methodological and philosophical principles on which data science can be based. We will study the structure of these principles and the impact of data science on our modalities of understanding reality. This philosophical approach will bring to the fore not only the potential of data science, but its limits as well, and will suggest a narrow path for a future science that is still driven by us, and not ultimately shaped only by our machines.

Current Required Texts:

There are no required texts. The following papers will be an integral part of our analysis and the starting point of the philosophical discussion.


Other relevant texts that will be used for some topics are:


**Instructional strategies:**

The course will include daily, in-depth discussions, based on a variety of readings. Mathematical and computational topics will be introduced by lectures. A period of guided written reflection will be set aside at the end of each class. Moreover, students will write an extensive and rigorously argued analysis of a specific topic agreed with the instructor by the fifth week of class. They will also present their research in class and moderate the resulting discussion.

**Methods of Evaluation:**

Assessment of student performance will be based on the following items:

**In-class discussions (25% of the grade).** Each student is expected to be directly involved in the discussion during each meeting. Because of this, attendance is required and at most two justified absences are permitted. For each additional absence, 2.5% of the grade will be deducted from the overall grade.

**Weekly written reflections (20% of the grade):** Written reflections (2 or 3 pages, 12pt font, 1” margin and double spaced) will be assigned every Wednesday and are due the following Monday. These reflections will take the form of creative and accurate syntheses of the topics discussed during the week, as well as of commentaries on specific problems and topics.

**Midterm evaluation (30% of the grade).** There will be a two-weeks, take-home exploration of the key conceptual and scientific ideas of data science methods. The emphasis will be on making sure that students have developed the ability to manipulate and generalize the key ideas presented in the course and to understand how to apply them to realistic scenarios. **The midterm exam will be assigned on Monday March 25th and is due Monday April 8th at the beginning of class.**
Final presentation (25% of the grade). By the end of the fifth week of class, students need to confirm a topic for their presentation, related to the main themes of the course and chosen among a wide selection of given topics. They will prepare for a 10 minutes presentation and 25 minutes Q and A session on their chosen topic. Students will be assessed based on the accuracy and creativity of their approach to the topic; the ease and effectiveness of their answers during the discussion; and the breadth and depth of their understanding of the themes of the course. Presentations will be scheduled for the last week of class, and for the day of the final exam.

Chapman University Academic Integrity Policy: Chapman University is a community of scholars which emphasizes the mutual responsibility of all members to seek knowledge honestly and in good faith. Students are responsible for doing their own work, and academic dishonesty of any kind will not be tolerated anywhere in the university.

Students with Disabilities Policy: In compliance with ADA guidelines, students who have any condition, either permanent or temporary, that might affect their ability to perform in this class are encouraged to inform the instructor at the beginning of the term. The University, through the Center for Academic Success, will work with the appropriate faculty member who is asked to provide the accommodations for a student in determining what accommodations are suitable based on the documentation and the individual student needs. The granting of any accommodation will not be retroactive and cannot jeopardize the academic standards or integrity of the course.

Last revised: 01/26/2019