

**USC VITERBI SCHOOL OF ENGINEERING INFORMATICS PROGRAM**  
**INF 250: Introduction to Data Informatics**  
*Syllabus*  
*Monday, Wednesday 1 – 3 PM (4 Units)*  
*Room TBD*

Instructor: Dr. Seon Ho Kim  
Email: [seonkim@usc.edu](mailto:seonkim@usc.edu)  
Office: PHE304  
Phone: 213.740.2483

Teaching Assistant: TBD  
Email: TBA  
Office: TBA  
Phone: TBA

**Instructor's Office Hours:**

Wednesday 3:00 p.m. to 5:00 p.m. in PHE304. Other hours by appointment only. Students are advised to make appointments with the professor ahead of time in any event and be specific with the subject matter to be discussed. Students should also be prepared for their appointment by bringing all applicable materials and information.

**Course Description:**

This is an overview course to give undergraduate students a broad understanding of Informatics topics, i.e., basic concepts and applications of Informatics. Topics include data representation and modeling, the role of the data scientist, the data lifecycle, high level introduction to machine learning, analyzing and exploring big data with real world use cases. This class will be primarily individual and/or team study, with assigned readings, homework assignments, midterm exams, and a final.

**Catalogue Description:**

Fundamentals of data informatics: representation of data and knowledge, role of a data scientist, data storage/processing/analytics, machine learning, big data, and data visualization.

**Corequisites:**

ITP 115 Programming in Python

**Recommended Preparation:**

This is an introductory Data Informatics course for undergraduate students with limited technical backgrounds in computing. ITP 115 is listed as a corequisite so that students have some basic programming skills in Python. A basic understanding of engineering and/or technology is recommended.

**Learning Objectives:**

Learning objectives for students are:

1. Understanding the methods to represent data and knowledge in computing.
2. Understanding the methods to store, index, and retrieve data on computers.
3. Understanding data lifecycle, the role of the data scientists, and data modeling.
4. Basic understanding of machine learning algorithms.
5. Basic understanding of data-centric decision making and problem solving on computers.
6. Introduction to visualization of large scale datasets.

## Methods of Teaching:

The primary teaching methods will be lectures, discussion, and case studies. Basic concepts and principles will be lectured and several real world case studies will be introduced for discussion. Some basic mathematical and computational analysis using R will be introduced and practiced in the class.

There will be two mid-term exams and a comprehensive final exam. Students will be required to complete six homework assignments which should be done individually. There will be no laboratory assignments, and no special computing facility, hardware or software will be necessary for this course.

## Class Communication:

Blackboard at USC will be used for class communication.

## Grading Schema:

Final: 30%

Mid-Terms: 25%

Homework Assignments: 40%

Participation: 5%

---

Total 100%

Grades will range from A through F. The following is the breakdown for grading:

94-100 =A

90-93 =A-

87-89 =B+

84-86 =B

80-83 =B-

77-79 = C+

74 - 76 = C

70 - 73 = C-

67 - 69 = D+

64 - 66 = D

60 - 63 = D-

Below 60 is an F

## Books and Readings:

### *Books (some selected chapters):*

J. Stanton, **An Introduction to Data Science**, Syracuse University, 2012.

<https://itunes.apple.com/us/book/introduction-to-data-science/id529088127?mt=13>

C. O'Neil, R. Schutt, **Doing Data Science: Straight Talk from the Frontline**, O'Reilly Media, Oct. 2013. <http://shop.oreilly.com/product/0636920028529.do?sortby=publicationDate>

**Assigned Reading List: can be added later**

1. Instructor's Lecture Notes
2. D. P. Groth and J. K. MacKie-Mason, "Why an Informatics Degree?," Communications of the ACM, vol. 53, pp. 26-28, February 2010.
3. USGS Data Management Data Lifecycle Overview  
<http://www.usgs.gov/datamanagement/why-dm/lifecycleoverview.php>
4. T. H. Davenport and D. Patil, "Data scientist: the sexiest job of the 21st century," Harvard business review, vol. 90, pp. 70-77, 2012.
5. P. Fox and J. Hendler, "Changing the equation on scientific data visualization," Science(Washington), vol. 331, pp. 705-708, 2011.
6. D. A. Keim, "Information visualization and visual data mining," Visualization and Computer Graphics, IEEE Transactions on, vol. 8, pp. 1-8, 2002.
7. J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, pp. 107-113, 2008.
8. Challenges and Opportunities with Big Data,  
<http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

**Course Schedule**

Week	Topic	Readings	Deliverables
1	Introduction to Data Informatics 1	Lecture Note, O'Neil Ch 1, Reading 2, 4	
2	Introduction to Data Informatics 2	Lecture Note, Reading 3	
3	Data and Knowledge Representation	Lecture Note Stanton Ch 1, 2, 4	HW1
4	Overview of R	Stanton Ch 3	
5	Data Wrangling and Formats	Lecture Note	HW2
6	Data Storage and Database	Lecture Note Stanton Ch 5	Midterm 1
7	Statistical Inference and Data Science Process	O'Neil Ch 2 Stanton Ch 6, 7	HW3
8	Introduction to Machine Learning	Lecture Note O'Neil Ch 3	
9	Extracting Meaning from Data	O'Neil Ch 7	HW4
10	Use Case Study 1	TBD	Midterm 2
11	Introduction to Big Data	Lecture Note, Reading 8	

12	Big Data Engineering and Parallel Processing	Lecture Note	HW5
13	Hadoop and MapReduce	O'Neil Ch 14, Reading 7	
14	Introduction to Data Visualization	Reading 5, 6	HW6
15	Use Case Study 2	TBD	

### Students with Disabilities:

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. Your letter must be specific as to the nature of any accommodations granted. DSP is located in STU 301 and is open 8:30 am to 5:30 pm, Monday through Friday. The telephone number for DSP is (213) 740-0776.

### Statement on Academic Conduct and Support Systems

#### Academic Conduct

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Section 11, Behavior Violating University Standards <https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/>. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct/>.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity <http://equity.usc.edu/> or to the Department of Public Safety <http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us>. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men <http://www.usc.edu/student-affairs/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage <http://sarc.usc.edu/> describes reporting options and other resources.

#### Support Systems

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute <http://dornsife.usc.edu/ali>, which sponsors courses and workshops specifically for international graduate students. The Office of Disability Services and Programs [http://sait.usc.edu/academic-support/centerprograms/dsp/home\\_index.html](http://sait.usc.edu/academic-support/centerprograms/dsp/home_index.html) provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially

declared emergency makes travel to campus infeasible, USC Emergency Information <http://emergency.usc.edu/> will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.

**Return of Course Assignments:**

Returned paperwork, unclaimed by a student, will be discarded after a year and hence, will not be available should a grade appeal be pursued following receipt of his/her grade.

**Emergency Preparedness/Course Continuity in a Crisis:**

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.