

Course: EST 371 – Data Science Management
Department: Department of Technology and Society
Semester: Spring 2020
Class schedule: TBA (location TBA)
Professor: Sira Maliphol (TBA smaliphol@sunykorea.ac.kr)
Office: TBA
Office hours: Drop-in hours: TBA
By appointment, email in advance



COURSE DESCRIPTION

The concepts of big data, data science, cloud computing, and data visualization for technology management are introduced. They are explored in the context of the digital network revolution, characterized by social media, Internet of Things (IoT), and mobile devices as sources of big data. Services that use large amounts of data and the statistical and software tools that enable them are emphasized. The underlying networking infrastructure is explained as a function of cloud computing. The case studies focus on information and communications technologies for sustainable development (ICT4D).

Prerequisites: AMS 161 or MAT 132 or MAT127; CSE 114; U3 or U4 standing

STUDENT LEARNING OUTCOMES

This course emphasizes and supports the following student learning objectives:

- Distinguish between key concepts such as data science, artificial intelligence, machine learning, neural networks, and deep learning.
- Understand the importance of key performance indicators (KPIs) for designing data science solutions in various industries.
- Classify the different roles of the data scientist.
- Assess emerging job classifications in data science and extended fields, e.g. AI.
- Distinguish the capabilities of different programming languages for data science.
- Understand how R can be used for data analysis.
- Describe how various programming languages can be used for big data and other methodologies.
- Propose how data science can be used in various organizations and industries
- Distinguish the difference between descriptive, predictive, and prescriptive analytics.
- Understand the different aspects of predictive analytics: regression, clustering, classification, association, text and web mining.
- Understand how data is being used to monitor behaviors and create prediction products
- Differentiate between various data priorities and strategies.
- Understand the importance of ethics in data science

TEXTBOOKS

(BBB) Salganik, M.J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Online access: <https://www.bitbybitbook.com/>

(RDS) Golemund, G. and Wikham, H. 2017. *R for Data Science*. New York, NY: O'Reilly Media. Online access: <https://r4ds.had.co.nz/>

(IDS) Cielen, D., Meysman, A.D.B, and M. Ali. 2016. *Introducing Data Science*. Manning Publications. Online Access: <https://livebook.manning.com/book/introducing-data-science/>

(DDS) O'Neil, C. 2013. *Doing Data Science*. New York, NY: O'Reilly Media. (Optional)

EVALUATION

Participation (10%): Class participation includes contributions to class discussion. Note: Attendance is necessary for participation. See [student attendance policy](#).

Case study (20%): Assigned groups of 2-3 students will provide a new proposed case study for data science used for ICT4D. For assignment #2, students must provide a brief proposal of their case study including the topic, the motivation, and possible data sources that may be used to understand the case.

Assignments (35%): Assignments (except #2; see case study) will provide hands-on work of data science using R. Assignments may be done individually or in groups (no more than 3 students per group). All names of students in groups should be listed on the assignments on submission. Each assignment is worth 5 points. Late assignments will not be accepted once they have been reviewed in class.

Final exam (35%): In class final exam.

GRADING

| | | | | |
|------------|------------|------------|------------|-----------|
| A = 93-100 | B+ = 86-89 | B- = 80-82 | C = 73-75 | D = 60-69 |
| A- = 90-92 | B = 83-85 | C+ = 76-79 | C- = 70-72 | F = 0-59 |

CLASS POLICIES

Student conduct must follow university policies. Excused absences must be notified in advance unless emergency related. [See university policies](#).

You will need access to a computer to complete assignments. Laptops are permitted in class for notetaking, assignments, and other classroom-related activities. Laptops are encouraged for use during lab sessions.

Changes to the syllabus and course may become necessary. As instructor, I reserve the right to modify this syllabus. Any changes will be communicated to the entire class clearly and in writing.

ADDITIONAL COURSE SYLLABUS STATEMENTS

1. Disabilities Act. If you have a physical, psychological, medical or learning disability that may impact your course work, please contact the Department of Student Affairs, Campus Building A, Room 207, (032) 626-1190. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

2. Academic Integrity. Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Plagiarism and representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at <http://www.stonybrook.edu/uaa/academicjudiciary/>

The school regards any act of academic dishonesty as a major violation punishable by severe penalties, including dismissal from the University. University policy requires that instructors and GAs and TAs report all suspected cases of academic dishonesty to the appropriate Academic Judiciary Committee, which is empowered to take strong action against violators. Under no circumstances will the University permit cheating of any kind.

3. Critical Incident Management. SUNY Korea expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Department of Academic Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn.

COURSE SCHEDULE

| Week | Topic (Tentative) | Class Preparation (Assignment due) |
|---------------|---|--|
| 1 (Feb 24) | Introduction <ul style="list-style-type: none"> • What is data science? • Installing tools used for class | |
| 2 (Mar 2) | Why Data Science <ul style="list-style-type: none"> • What enables data science? • What are uses of data science? • Reading <ul style="list-style-type: none"> ○ BBB Ch 1: Introduction. ○ RDS Ch 1: Introduction ○ RDS Ch 2: Explore – Introduction | Email me with your contact information |
| 3 (Mar 9) | Data Visualization <ul style="list-style-type: none"> • What is data visualization? • What is it good for? • Reading <ul style="list-style-type: none"> ○ RDS Ch 3: Data visualization. | Assignment #1 |
| 4 (Mar 16) | The Data <ul style="list-style-type: none"> • What are KPIs? • Reading <ul style="list-style-type: none"> ○ BBB Ch 2. Observing data | Assignment #2 |
| 5 | Data Transformation <ul style="list-style-type: none"> • Reading | Assignment #3 |

| | | |
|-----------------------|---|--------------------|
| (Mar 23) | <ul style="list-style-type: none"> ○ RDS Ch 4: Workflow: Basics ○ RDS Ch 5: Data transformation | |
| 6 (Mar 30) | <p>Descriptive Data Analysis</p> <ul style="list-style-type: none"> • What do we know about the data? • Reading <ul style="list-style-type: none"> ○ R Companion Handbook | Assignment #4 |
| 7 (Apr 6) | <p>Exploratory Data Analysis</p> <ul style="list-style-type: none"> • Reading <ul style="list-style-type: none"> ○ RDS Ch 6: Workflow: scripts ○ RDS Ch 7: Exploratory Data Analysis | Assignment #5 |
| 8 (Apr 13) | <p>Working with the Data</p> <ul style="list-style-type: none"> • How is data stored? • Reading <ul style="list-style-type: none"> ○ RDS Ch 9: Wrangle Introduction ○ RDS Ch 10: Tibbles ○ RDS Ch 11: Data Import | Assignment #6 |
| 9 (Apr 20) | <p>Cleaning the Data</p> <ul style="list-style-type: none"> • Dirty data • Reading <ul style="list-style-type: none"> ○ RDS Ch 12: Tidy Data | Assignment #7 |
| 10 (Apr 27 /May 4) | <p>Applications</p> <ul style="list-style-type: none"> • Reading <ul style="list-style-type: none"> ○ Cases of data science and ICT4D (TBD) ○ ITU | Work on case study |
| 11 (May 11) | <p>Research Questions</p> <ul style="list-style-type: none"> • Reading <ul style="list-style-type: none"> ○ BBB Ch 3: Asking Questions | Work on case study |
| 12 (May 18) | <p>Experiments</p> <ul style="list-style-type: none"> • Reading <ul style="list-style-type: none"> ○ BBB Ch 4: Running Experiments | Work on case study |
| 13 (May 25) | <p>Using Data Science</p> <ul style="list-style-type: none"> • What is data science, big data, and artificial intelligence? • Complexity • Reading <ul style="list-style-type: none"> ○ IDS Ch 1: Data Science in a Big Data World ○ BBB Ch 5. Mass Collaboration ○ UNDP Section 1 | Work on case study |
| 14 (Jun 1) | <p>Ethics</p> <ul style="list-style-type: none"> • Problems that the data cannot solve • Reading <ul style="list-style-type: none"> ○ BBB Ch. 6. ○ UNSDG | Case study due |
| 15 (Jun 8) | Reading period | |
| Finals (Jun 15) | Final exam, in-class | |