EECS 731: DATA SCIENCE  
FALL 2017

Professor: Nicole Beckage
Office: Eaton 3022
Office Hours: TR 11:00am – 12:00pm, and by appointment
E-Mail: beckage@ku.edu
Course Website: piazza.com/ku/fall2017/eecs731
http://ittc.ku.edu/~beckage/2017ds/index.html

Lecture Hours and Venue: TR 2:30pm – 3:45pm in Learned 1136

Description:
Data Science draws on tools from computer science and statistics to answer disciplinary scientific questions through integration of different analysis techniques and types of data. Because data science uses tools from computer science and statistics, this course aims to introduce students to analyzing and modeling data. We begin by understanding how to think about data science: How can we evaluate results from data science? What is the experimental design of a particular data science problem, is it valid? We then consider statistical and machine learning approaches that facilitate learning from data. We discuss limitations of modeling and how to construct predictive and explanatory models of data. A large component of this course is based on the process and evolution of a research question and will be project focused.

This graduate-level course will focus on a research question from a KU faculty. Teams of approx. 5 students will work through the course of the semester on an in-depth data driven research project. These projects have a wide variety of topics and require self-directed learning. Students will present a technical lecture on the main methodology for the project half-way through the semester. At the end of the semester, students must submit a 8-10 page research paper and give a research presentation. Students will be evaluated based on the quality of their research and their ability to serve as a member of a research team. While the project component is the largest part of this class, homeworks will provide hands-on experience with data munging, data analysis, and the use of data in real world situations. Lectures will provide scaffolding and background information that will support the project component of this course.

Prerequisites: Probability, statistics and analysis of algorithms are assumed prerequisites of this course. This course also assumes programming ability. We will use R for assignments. It is advised that students are familiar, or can quickly become familiar, with this language. This course additionally requires regular access to a computer. If you do not have regular access to a computer, you should get access to the EECS computer lab.
Textbook: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

Objectives: After successful completion of this course, students should be able to:

- Understand experimental design and model verification/validation
- Use the scientific method and statistical knowledge to evaluate claims
- Perform exploratory statistical analysis
- Learn new modeling methods
- Effectively communicate the outcome of data analysis
- Write and present novel research results

Grade Breakdown:

Homework: There will be 4 programming homework assignments throughout the semester covering the more applied topics of data science. Homework will be graded based on the correctness as well the clarity of the solution. I expect students not to copy, refer to, or look at online or other solutions in preparing their answers. It will be considered plagiarism and academic misconduct to intentionally refer to online solutions or to seek solutions that are otherwise easily available online.

Students will also be asked to post on Piazza regularly with responses to assigned readings and to comment on in-class discussions.

Research Project: The main focus of this course is a research project. The project is based on a data set and scientific question of faculty at KU. Teams will include approx. 5 people and will last through the course of the semester.

- Technical Presentation: Research team must give a 60 minute in-depth technical lecture on an analysis method that is relevant to the research project. Lecture topics must be approved by the professor. This assignment will consist of an in-class presentation and a set of lecture notes. You must describes the algorithm/method/framework in appropriate, including optimization procedures, model assumptions, and real-world use cases. You should discuss applications and limitations. The use of primary sources from initial publications of the method/framework as well as how it is used today are to be included.

- Research Presentation: The semester will conclude with research teams giving a presentation on the work they have done throughout the semester. We will use the last few classes as well as the final slot. You must attend the final class meeting which will be held Wednesday December 13th from 1:30-4:00pm.

- Research Paper: Each group must submit a final research paper. The paper is due at 1:25 pm on Wednesday December 13th. Before submission of the final research paper, a complete draft will be workshopped by the class approximately 2 weeks before the end of the semester.

- Group Evaluation: Throughout the course of the semester and for each major deliverable, research teams will provide self-evaluation and team evaluations of progress and collaboration. As a group, weekly summaries of progress be posted on Piazza starting Sept 12.
**Course Grading:** Your course grade will be calculated based on the breakdown below. No extra credit will be available. **+/- grading will be used.**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attend 3 Research Talks (2 in EECS, 1 in Eng.)</td>
<td>5%</td>
</tr>
<tr>
<td>Group Evaluations (overall)</td>
<td>5%</td>
</tr>
<tr>
<td>Code Base</td>
<td>5%</td>
</tr>
<tr>
<td>Participation and Professionalism</td>
<td>10%</td>
</tr>
<tr>
<td>Homeworks and Piazza</td>
<td>15%</td>
</tr>
<tr>
<td>Technical Lecture (includes group eval. at 3%)</td>
<td>20%</td>
</tr>
<tr>
<td>Research Presentation (includes group eval. at 3%)</td>
<td>20%</td>
</tr>
<tr>
<td>Research Paper (includes group eval. at 3%)</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Course Logistics:**

**Assignment Submission:** Assignments are due at 11:55pm central time via blackboard according to the date posted on the course web page. Each person has 3 free late days to be used during the course. Late days can only be used in increments of one day, so both 1 minute late and 10 hours late count as a whole late day. After late days are used, late assignments will get maximally half credit. Assignments more than 2 days late will not be graded unless late days are used. Late days can only be used for homeworks and evaluations, not for presentations or papers. Each calendar day counts as a late day.

**Academic Misconduct:** All work submitted for credit must be the student’s own and is subject to the provisions of KU policies. Students should review the university policy on Academic conduct at: [http://policy.ku.edu/governance/](http://policy.ku.edu/governance/).

**Accommodation Procedure:** The Academic Achievement and Access Center (AAAC) coordinates academic accommodations and services for all eligible KU students with disabilities. If you have a disability for which you wish to request accommodations and have not contacted the AAAC, please do so as soon as possible. They are located in 22 Strong Hall and can be reached at 785-864-4064 (V/TTY). Information about their services can be found at [http://www.access.ku.edu](http://www.access.ku.edu). Please contact me privately in regard to your needs in this course.

**KU Writing Center:** At the KU Writing Center, you can work one-on-one with writing consultants trained to talk with you about writing. You can meet face-to-face, receive feedback via e-mail, or schedule a videoconference. Consultants are coaches who work with you on any type of assignment—essays, research papers, reports, presentations—at any point of completeness or incompleteness. To learn more, browse to [http://writing.ku.edu](http://writing.ku.edu).

**Nondiscrimination:** The University of Kansas prohibits discrimination on the basis of race, color, ethnicity, religion, sex, national origin, age, ancestry, disability, status as a veteran, sexual orientation, marital status, parental status, retaliation, gender identity, gender expression, and genetic information in the University’s programs and activities. Please contact the University’s Title IX Coordinator at IOA@ku.edu with any inquiries.

**Religious Observances:** Should a session for this course conflict with your mandated religious observance, please contact me at the beginning of the semester so that we can find a mutually acceptable solution. Students are responsible for initiating discussion with the instructor.