1. **Credits:** 3
   a. Class Meets: **Tuesday and Thursday 1:30pm - 2:45pm, Family Life (FL) 109**

2. **Instructor:** Kyumin Lee, (435) 797-8420, kyumin.lee@usu.edu
   Office Hours: **9:30-10:30am T/Th** at MAIN 401D, or by appointment

   **TA:** Hongkyu Choi, tohongs@gmail.com
   Office Hours: **9-10am W/F** at MAIN 422

3. **Textbooks:** No primary textbooks, but we will refer to the following textbooks:
   - Data Mining: Concepts and Techniques. Jiawei Han and Micheline Kamber. Morgan Kaufmann. 2011

   In addition, we will read some research papers from top computer science conferences and journals.

4. **Specific Course Information:**
   a. Course Description: This course introduces research topics in data science and mining, and addresses relevant theoretical foundations, methods, and tools. In particular, we will **focus on research topics in social media mining**. The course topics include, but are not limited to: social media, geography and the social web, crowdsourcing, privacy and security issues related to social systems (e.g., Facebook, Twitter and Google+), etc. Students will learn how to collect, parse and analyze social media data with data mining techniques.
   b. Prerequisites: Grade of C- or better in CS5665, or permission from instructor.

5. **Specific goals for the course**
   a. Course Objectives
      By the end of the semester you will be able to:
      i. Define and explain the key concepts, methods, and tools relevant to data mining
ii. Design, implement and evaluate core algorithms and approaches  
iii. Identify the salient features and apply recent research results in data mining  
iv. Have hands-on experience by performing programming assignments and a project that will reinforce the theoretical aspects covered in lectures.

Mapped objectives in IDEA:

i. Learning fundamental principles, generalizations, or theories  
ii. Learning to apply course material (to improve thinking, problem solving, and decisions)  
iii. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course

b. Student Outcomes:  
   i. An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs  
   ii. An ability to use current techniques, skills, and tools necessary for computing practice

6. Brief list of topics to be covered  
   a. Crowdsourcing  
   b. Geography and the Social Web  
   c. Influence in Social Information Networks  
   d. Social Spam  
   e. Communities  
   f. Duplicate Detection  
   g. Information Diffusion

All course announcements will be posted to the Google Group's course mailing list. If you have a question to discuss with everyone, please post it to the group! If you have a specific question to me, please send me an email.

Grading Policy:  
The course grading policy is as follows:  

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Component</th>
<th>Percentage</th>
<th>Component</th>
<th>Percentage</th>
<th>Component</th>
<th>Percentage</th>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>In-class discussion</td>
<td>10%</td>
<td>Assignments</td>
<td>15%</td>
<td>Paper presentation</td>
<td>20%</td>
<td>Midterm</td>
<td>15%</td>
</tr>
<tr>
<td>15%</td>
<td>Project proposal</td>
<td>15%</td>
<td>Project presentation</td>
<td>20%</td>
<td>Final project deliverable</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The grading scale is A: 93-100, A-: 90-92.9, B+: 87-89.9, B: 83-86.9, B-: 80-82.9, C+: 77-79.9, C: 73-76.9, C-: 70-72.9, D: 60-69.9, F: 0-59.9
Assignments:
There are two programming assignments with/without written assignments. Each assignment is proportion to 5% of your grade. You will have total 2 late days during the semester. You can use up to 2 late days for each assignment without penalty. After you consume the total 2 late days for an assignment, then you will get penalty proportion to extra late days (e.g., 10% off for the next late day, 20% for the next two late days and so on).

For example, you submitted your first assignment 2 days late. You will not get any penalty, but use 2 out of 2 late days. Or if you submit your first assignment 3 days later than due date, you will use 2 late days (again up to 2 late days for an assignment), and get 10% off penalty because of the third late day.

For each assignment, we will NOT accept your solution more than 5 days late.

You may discuss an assignment with your colleague, but you should write a program by yourself and should NOT copy and paste your colleague's program. If you discussed an assignment with your colleague, explicitly report the colleague's name and what you discussed in your submission.

Exam:
The midterm exam is closed book and will be held in class. You may bring one standard 8.5" by 11" piece of paper with any notes you think appropriate or significant (front and back). No electronic devices allowed.

Project:
The most significant portion (50%) of your grade is based on the project, for which you may work in teams of two or three persons. The general goal of research is to contribute new knowledge. Thus, it is important to ask yourself what research question(s) you aim at answering and what challenges you aim at solving.

In general, a project should include both: (i) an empirical evaluation of an algorithm or model on an interesting dataset, in order to better understand these methods, and possibly further improve them; and (ii) a proposal for a new algorithm and model, including a comparison with a baseline.

When picking a topic, try to ask yourself the following questions:

- Is the topic addressing an important problem? Would anyone care about it if you solve the problem?
- To what extent has previous research work addressed this problem? And what remains unknown?
- Do you have any idea at all about how to solve the problem? If not, can you reformulate the problem to make it easier?
- Would you be able to evaluate your solution? That is, how can you demonstrate your solution is good and solves the problem well?

Proposal: The emphasis of the course is on recent and current scholarly material. We will read a number of papers from the past few years and these papers should stimulate you to think
about unexplored avenues of research. For the project proposal, you should pick two or three related papers as the basis for your proposal. These related papers will serve as the research grounding for your proposal. So first in your proposal, you must address these questions (in about 2 or 3 pages, single column, normal fonts and margins):

- What is the key technical content and interesting ideas behind the papers?
- How do the papers relate to the topics presented in the course?
- What are the strengths of the paper? What are its weaknesses? And how might you go about improving on the weaknesses?

In the remainder of the proposal (about 2 pages), I expect you to dig deeply into your proposed work (which should naturally build on the background materials you’ve already covered).

- What is your research question? Clearly define the research problem/question.
- Why is this an interesting question to ask and why would we care about the answer to this question or a solution to the problem?
- Has any existing research work tried to answer the same or a similar question, and if so, what is still unknown?
- How do you plan to work out the answer to the question? (At the proposal stage, you are only expected to have a sketch of your methods.)
- How would you evaluate your solution? That is, how do you plan to demonstrate that your solution/answer is good or is reasonable.
- A rough timeline to show when you expect to finish what. List a couple of milestones.

You should plan to write a proposal of 3 ~ 5 pages. The project proposal is due on February 27 by 11:59pm. Submit your project proposal in PDF with a pithy filename via Canvas. One per team.

Final Project Deliverable: At the end of the semester, you will deliver a final paper and participate in a presentation (maybe with a demo) session.

- The presentation session will be held on April 25 and 27, 2016. 15% of your project grade is based on the project presentation. Your team's grade will be based on the clarity of the presentation itself, audience feedback, and your answers to my questions.
- The final report is due on May 5, 2016 by 11:59pm. You should email me your final report in PDF with a pithy filename. The paper should be in the standard ACM conference template http://www.acm.org/sigs/publications/proceedings-templates (double-column), and should have at least 6 pages if all of your group members are CS6675 students, or at least 8 pages if any of your group members is a CS7675 student.
- The paper counts for 20% of your project grade. You should write your report as if you were writing a conference paper. You should address the same questions as those you have addressed in the proposal, only with more details, especially regarding some of the challenges that you need to solve and your experimental results. You should also include your conclusions from the study and point out how your work can be further extended (i.e., future work).

Add policy:
The last day to add this class is January 30 (5:00 PM). Attending this class beyond that date, without being officially registered, will not be approved by the Dean’s Office. Students must be officially registered for this course. No assignments or tests of any kind will be graded for students whose names do not appear on the class list.
Drop policy:
The last day to drop this class without notation is January 30 (5:00 PM).

Withdrawal Policy and "I" Grade Policy:
Students are required to complete all courses for which they are registered by the end of the semester. In some cases, a student may be unable to complete all of the coursework because of extenuating circumstances, but not due to poor performance or to retain financial aid. The term 'extenuating' circumstances includes: (1) incapacitating illness which prevents a student from attending classes for a minimum period of two weeks, (2) a death in the immediate family, (3) financial responsibilities requiring a student to alter a work schedule to secure employment, (4) change in work schedule as required by an employer, or (5) other emergencies deemed appropriate by the instructor.

Learning Aids:
Lecture notes and other useful information will be available in electronic form on the class's section of the Canvas system. Please check the class's news and notes sections on a regular basis.

The Computer Science Department is a member of the Microsoft’s DreamSpark program. Through this program, students in CS courses can obtain and use a number of Microsoft's operating and software packages. If you are interesting in downloading any of this software for your use, please follow the directions found on the department’s website.

Academic Integrity – “The Honor System”:
Each student has the right and duty to pursue his or her academic experience free of dishonesty. The Honor System is designed to establish the higher level of conduct expected and required of all Utah State University students.

The Honor Pledge: To enhance the learning environment at Utah State University and to develop student academic integrity, each student agrees to the following Honor Pledge: "I pledge, on my honor, to conduct myself with the foremost level of academic integrity." A student who lives by the Honor Pledge is a student who does more than not cheat, falsify, or plagiarize. A student who lives by the Honor Pledge:

- Espouses academic integrity as an underlying and essential principle of the Utah State University community;
- Understands that each act of academic dishonesty devalues every degree that is awarded by this institution; and
- Is a welcomed and valued member of Utah State University.

Plagiarism and Cheating:
Plagiarism includes knowingly "representing, by paraphrase or direct quotation, the published or unpublished work of another person as one's own in any academic exercise or activity without full and clear acknowledgment. It also includes the unacknowledged used of materials prepared by another person or agency engaged in the selling of term papers or other academic materials." The penalties for
plagiarism are severe. They include warning or reprimand, grade adjustment, probation, suspension, expulsion, withholding of transcripts, denial or revocation of degrees, and referral to psychological counseling.

This course adheres to the cheating policy for courses in the Department of Computer Science posted on the bulletin board outside the CS office on the 4th floor of Old Main and posted online at https://cs.usu.edu/policies-forms/cheating-policy.

**Students with Disabilities:**
Students with ADA-documented physical, sensory, emotional or medical impairments may be eligible for reasonable accommodations. Veterans may also be eligible for services. All accommodations are coordinated through the Disability Resource Center (DRC) in Room 101 of the University Inn, (435)797-2444. Please contact the DRC as early in the semester as possible. Alternate format materials (Braille, large print, digital, or audio) are available with advance notice.

**Sexual Harassment:**
Sexual harassment is defined by the Affirmative Action/Equal Employment Opportunity Commission as any "unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature." If you feel you are a victim of sexual harassment, you may talk to or file a complaint with the Affirmative Action/Equal Employment Opportunity Office located in Old Main, Room 161, or call the AA/EEO Office at 797-1266.

**Academic Freedom and Professional Responsibilities (Faculty Code):**
Academic freedom is the right to teach, study, discuss, investigate, discover, create, and publish freely. Academic freedom protects the rights of faculty members in teaching and of students in learning. Freedom in research is fundamental to the advancement of truth. Faculty members are entitled to full freedom in teaching, research, and creative activities, subject to the limitations imposed by professional responsibility. Faculty Code Policy #403 further defines academic freedom and professional responsibilities: USU Policies Section 403.