Instructor Contact Information

Instructor: Michael Hahsler  
E-mail: mhahsler@lyle.smu.edu  
Office hours: see http://michael.hahsler.net

Course Information

This course will be held as a virtual course and meets at the scheduled times via Zoom. See Canvas (http://canvas.smu.edu/) for details and the Zoom meeting IDs.

Course website: http://michael.hahsler.net/SMU/CS7331/

Course Description

Introduces data mining techniques (classification, association analysis, and cluster analysis) used in analytics. All material covered is reinforced through hands-on experience using state-of-the-art tools to design and execute data mining processes. All material covered will be reinforced through hands-on experience using the R programming environment.

Prerequisites:

• Introduction to programming: CS 1342 (any programming language)  
• Probability and statistics: CS 4340/EMIS 3340/STAT 4340 or CS 7370/EMIS 7370  
• Databases: EMIS 3309 or CS 3330 (undergraduates only)

Learning Outcomes

1. Students will be able to explain and use the knowledge discovery on databases (KDD) process.
2. Students will be able to use R for data preparation, data exploration and analysis, and predictive modeling.
3. Students will understand and be able to apply the core data mining methods of
   • Classification  
   • Association Analysis  
   • Cluster Analysis
4. Students will be able to conduct a complete data mining project including research, data preparation, and reporting the results.
Unofficial Learning Goals

The course consists of four major projects, where you will be asked to answer given questions using provided data and the tools covered in class. However, the aim is that by the end of this course you can

1. formulate your own questions and clearly explain why answers to these questions are significant,
2. communicate the value of these answers to stake-holders, and
3. identify and use the appropriate data and techniques to obtain the answers.

Textbook


We will cover chapters 1-9 in this course.

Course Topics and Class Outline

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Lecture</th>
<th>Reading</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Intro +</td>
<td>Intro + Analytics and the Data Mining Process</td>
<td>Ch 1</td>
<td>P1 due</td>
</tr>
<tr>
<td>Week 2</td>
<td>Data</td>
<td>Ch 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Data</td>
<td>Ch 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Data</td>
<td>Ch 8</td>
<td>P1 due</td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Data, Clustering</td>
<td>Ch 8</td>
<td>P1 due</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Clustering</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>Clustering</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 8</td>
<td>Spring break</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>Classification</td>
<td>Ch 8</td>
<td>P2 due</td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>Classification</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Classification</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>Classification</td>
<td>Ch 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 13</td>
<td>Association Analysis</td>
<td>Ch 6</td>
<td>P3 due</td>
<td></td>
</tr>
<tr>
<td>Week 14</td>
<td>Association Analysis</td>
<td>Ch 6</td>
<td>P3 due</td>
<td></td>
</tr>
<tr>
<td>Week 15</td>
<td>Association Analysis, Advanced Topics</td>
<td>Web Site</td>
<td>P4 due</td>
<td></td>
</tr>
</tbody>
</table>
Course Work and Grading

It is expected that each student will keep up with the reading as outlined above. Additional materials may be referenced in class as needed. The course grade will be determined based on the four projects and peer evaluation.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects*</td>
<td>23% each</td>
</tr>
<tr>
<td>Peer Evaluation</td>
<td>2% each</td>
</tr>
<tr>
<td>Online Participation</td>
<td>Up to 1% bonus per project</td>
</tr>
<tr>
<td>(Canvas discussion board)</td>
<td></td>
</tr>
</tbody>
</table>

Note: There is no midterm or final examination! Your entire grade depends on your performance on the four projects. The projects are hard work and designed to take several weeks each to complete. So, start early and start strong!

* Students enrolled in the graduate section will have additional project requirements.

Projects

You will work in teams of two students (distance students may elect to work alone) on four projects throughout the semester. The result of each project is a detailed project report (following the categories of the CRISP-DM framework). You may share code, comments and ideas between teams (using the discussion board on canvas), but each team needs to write their own report independently! The report needs to be submitted via Canvas before the due date. Late submissions of up to 3 days are possible, but you will lose 10% per day.

You can expect a B (up to 90%) if you address all questions in the project description adequately. For a better grade you need to present exceptional work (e.g., asking and answering your own questions, in-depth explanation why one method works better than another, developing special preprocessing, insightful visualization, using and explaining a method not covered in class).

Peer Evaluation

For each project you will be assigned two reports from other students. The evaluation will be done via Canvas and the other student will not see who evaluated their report. You will be graded on the completeness and quality of the reviews you write, not on the reviews your work receives!

Online Participation

Frequent participation on the classes discussion board (on Canvas) will earn you up to 1% bonus per project.
**Attendance Policy**
Students are expected to attend class regularly. **If a student is absent from class, it is that student's responsibility to make arrangements with the professor to make up any work missed or to ensure that assignments are submitted on time or early.** Late assignments will not be accepted except in extreme instances. Any assignments that will be missed (including those due to university-sanctioned events) must be completed before the due date. This includes quizzes and homework assignments.

**Academic Ethics and Collaboration**
You may work together to get ideas and write code, but you are expected to create, edit and print your own assignments. If you submit work which is not completely yours then you will receive a 0 on the particular assignment. In severe cases, you will receive an F in the course and may be brought in front of the SMU Honor Council. It is your responsibility to know and understand the University's Honor Code and the expectations for collaboration in this course.

**Additional Information**

- **Disability Accommodations:** Students needing academic accommodations for a disability must first register with Disability Accommodations & Success Strategies (DASS). Students can call 214-768-1470 or visit http://www.smu.edu/Provost/SASP/DASS to begin the process. Once approved and registered, students will submit a DASS Accommodation Letter to faculty through the electronic portal DASS Link and then communicate directly with each instructor to make appropriate arrangements. Please note that accommodations are not retroactive and require advance notice to implement.

- **Religious Observance:** Religiously observant students wishing to be absent on holidays that require missing class should notify their professors in writing at the beginning of the semester, and should discuss with them, in advance, acceptable ways of making up any work missed because of the absence. ([https://www.smu.edu/StudentAffairs/Chaplain/ReligiousHolidays](https://www.smu.edu/StudentAffairs/Chaplain/ReligiousHolidays)).

- **Excused Absences for University Extracurricular Activities:** Students participating in an officially sanctioned, scheduled University extracurricular activity should be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work. (See [2020-2021 SMU Undergraduate Catalog](https://www.smu.edu/StudentAffairs/Chaplain/ReligiousHolidays) under “Enrollment and Academic Records/Excused Absences.”)

- **Syllabus Statement for Pregnant and Parenting Students:** Accommodations for pregnant and parenting students: Under Title IX students who are pregnant or parenting may request academic adjustments by contacting Elsie Johnson ([elsiej@smu.edu](mailto:elsiej@smu.edu)) in the Office of the Dean of Students, or by calling 214-768-4564. Students seeking assistance must schedule an appointment with their professors as early as possible, present a letter from the Office of the Dean of Students, and make appropriate arrangements. Please note that academic adjustments are not retroactive and, when feasible, require advance notice to implement.

- **Covid-19 Attendance Statement:** Students who are experiencing COVID-19 symptoms or who have been notified through contact tracing of potential exposure and need to self-
quarantine or isolate must follow the protocols laid out in SMU’s Contact Tracing Protocol. To ensure academic continuity, students in these situations will not be penalized and will be provided appropriate modifications to assignments, deadlines, and testing. Please also note that SMUFlex classes might, in rare circumstances, go remote for two-week periods to accommodate COVID-related issues. To ensure these necessary accommodations, affected students must:

- Provide as much advance notification as possible to the instructor about a change in circumstances. Students must notify their instructor about a potential absence as well as plans for a return to class. For cases in which students test positive for COVID-19, they should fill out a CCC form at this link.
- Communicate promptly with the instructor to establish, as necessary, alternative assignments and/or changes to deadlines and exams. Students are then responsible for meeting the expectations laid out in these alternative arrangements.
- Continue participation in class via Zoom, as health circumstances permit. Attend class regularly, when not in a situation outlined above, in accordance with safety measures laid out by SMU CAN in the Pledge to Protect (including wearing masks, maintaining social distancing, and cleaning personal space after class). In-person participation in SMUFlex classes is required on students’ assigned red/blue rotation days except in cases when students are experiencing illness, are in self-quarantine or in isolation.
- Students facing multiple or extended COVID-19-related absences or illness can work with the Office of the Dean of Students to consider options such as fully remote learning or medical withdrawal.

- This policy, aligned with the SMU Honor Code and the SMU Pledge to Protect, relies on mutual trust and respect between students and faculty to ensure safety, academic integrity, and instructional continuity.