Instructor Contact Information

Instructor: Michael Hahsler
Office: TBD
Phone: (214) 768-8878
E-mail: mhahsler@lyle.smu.edu
Office hours: TBD or by appointment

Course Information

Course website: http://michael.hahsler.net/SMU/CS7331/
For further information, assignments and assignment submission: http://canvas.smu.edu/

Course Description

Analytics is based on collecting, managing, exploring and acting on large amounts of data and has become a source of competitive advantage for many organizations. This course introduces the core data-mining techniques (classification, association analysis and cluster analysis) used for predictive analytics and the 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 mining process for descriptive and predictive analytics.
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 stakeholders, 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></td>
<td>Intro + Analytics and the Data Mining Process</td>
<td>Ch 1</td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td></td>
<td>Data</td>
<td>Ch 2</td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td></td>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td></td>
<td>Clustering</td>
<td>Ch 7</td>
<td>P1 due</td>
</tr>
<tr>
<td>Week 5</td>
<td></td>
<td>Clustering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td></td>
<td>Classification</td>
<td>Ch 3</td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td></td>
<td>Classification</td>
<td>Ch 4</td>
<td>P2 due</td>
</tr>
<tr>
<td>Week 8</td>
<td></td>
<td>Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td></td>
<td>Association Analysis</td>
<td>Ch 5</td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td></td>
<td>Association Analysis</td>
<td></td>
<td>P3 due</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>25% per project</td>
</tr>
<tr>
<td>Peer Evaluation</td>
<td>5% per project</td>
</tr>
<tr>
<td>Online Participation</td>
<td>3% per project</td>
</tr>
<tr>
<td>(Canvas discussion board)</td>
<td>+ 1% for introduction</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 on three 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 points as stated above.
**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](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 [2019-2020 University Undergraduate Catalogue](https://www.smu.edu/Provost/SASP/DASS) under “Excused Absences”)

- **Student Academic Success Programs:** Students needing assistance with writing assignments for SMU courses may schedule an appointment with the Writing Center through Canvas. Students wishing support with subject-specific tutoring or success strategies should contact SASP, Loyd All Sports Center, Suite 202; 214-768-3648; [https://www.smu.edu/sasp](https://www.smu.edu/sasp).