San José State University  
College of Business  
BUS 235C Data Mining  
Fall 2015

**Instructor:**  
Dr. David Czerwinski

**Office Location:**  
Business Tower 753

**Email:**  
david.czerwinski@sjsu.edu

**Office Hours:**  
By appointment.

### Course Description

This course surveys a range of state-of-the-art data mining techniques. The focus of the course is on developing the intuition behind the algorithms and hands-on learning with actual data sets. Students will learn and use the R statistical package.

### Course Web Site

This course takes place on Canvas at [http://sjsu.instructure.com/](http://sjsu.instructure.com/). Your Canvas login is the same as your MySJSU login.

Some students have reported having problems using Canvas with Internet Explorer. I'd encourage you to use a different web browser such as Chrome, Firefox, Opera, or Safari.

Your password expires every six months without warning. Don't let an expired password cause you an inconvenience during the semester – change your password *today* so you don't have to worry about it for the rest of the semester. You can change your password at [http://its.sjsu.edu/services/sjsuone/](http://its.sjsu.edu/services/sjsuone/).

### Technical Help

For technical help, you can access any of the following resources.

SJSU Information Technology Support Services (ITSS) help desk: Password resets, log-in issues, email support, account requests: [http://www.sjsu.edu/at/hd/support/](http://www.sjsu.edu/at/hd/support/)


You can also submit a help ticket at  
[https://isupport.sjsu.edu/ecampus/ContentPages/Incident.aspx](https://isupport.sjsu.edu/ecampus/ContentPages/Incident.aspx)
Course Goals and Student Learning Objectives

Upon completion of this course students will:

- Be able to apply predictive data mining techniques including logistic regression, classification trees, and ensemble methods.
- Be able to gain business insight by applying unsupervised machine learning techniques including clustering analysis and association rules.
- Be able to assess the performance of data mining models.
- Get a sense of the range of applications of these techniques.
- Gain hands-on experience through realistic problems and a course project.

Business Analytics Certificate Learning Outcomes

Upon completion of the Certificate Program in Business Analytics, students will have:

- an understanding of the breadth of the discipline of business analytics
- basic depth of knowledge for the various aspects of business analytics
- competencies associated with undertaking custom-designed research
- an understanding of the role of custom-designed research in business planning
- competencies associated with data mining*
- an understanding of the role of data mining in business planning*
- an understanding of the importance of integratively assimilating data from diverse sources (i.e., consumer, competition, macro market, and company data)
- an understanding of the role of business metric dashboards in business planning

* Indicates a learning outcome directly supported by this course.

Suggested Texts

You are not required to purchase a textbook for this course. The lectures and demonstrations are intended to be self-contained. However, many students find it beneficial to have a text to refer to and some people learn best by reading. If you’d like to purchase a textbook, our lectures will closely follow the presentation in Data Mining for Business Intelligence, 2nd Edition by G. Shmueli, N. Patel, and P. Bruce. John Wiley & Sons, Hoboken, New Jersey. 2010. ISBN 978-0-470-52682-8. Of the many data mining books out there, I find this book to be the most readable and it presents the mathematics at a reasonable level. The only caveat is that it is accompanied by software, XLMiner, that we will not be using because it is not widely used in industry.

We will be using the R statistical software package. I will introduce the basics of R and cover the details of using R to perform the specific data mining techniques we will learn. There is much more to R than we will cover, and if you would like to dig deeper or gain a broader understanding of R, there are many books about it, including some nice free references. I will post two on the class website as PDFs. Using R for Data Analysis and Graphics: Introduction, Code and Commentary by J. H. Maindonald provides a good
introduction to the basic features of R and will be familiar to those of you who took Introduction to Business Analytics. *R and Data Mining: Examples and Case Studies* by Y. Zhao includes example R code for many data mining methods. (There are also more expansive paid versions of both books that you can purchase from any online bookstore.)

**Assignments and Grading Policy**

The course grade will be based on homework assignments (50%), a final project (40%), and class participation (10%). A description of the final project will be posted on the class website.

**Homework Guidelines**

Homework will be due at 11pm on Sunday at the end of each week when assigned and should be turned in through Canvas. I monitor the class discussion list to answer questions most actively on weekdays, so I encourage you to start the homework during the week if your schedule permits. I will also monitor the class discussion over the weekend, but not as frequently. If you know in advance that you will have difficulty turning in an assignment on time please notify me as soon as possible.

Homework should be submitted as a single file, in either Word or pdf format.

**Participation Grade**

Participation is worth 10% of your grade. I’ve set the bar low for obtaining full participation points: you must ask or answer five questions on the discussion list over the course of the class. Each question you ask or answer will be worth 2 points, up to a maximum of a full 10 points towards your participation score. Of course, I encourage you to participate more than that, and expect many of you will contribute much more than five posts.

**Classroom Protocol**

Your command of the data mining material and the development of critical thinking skills will be facilitated in the context of online screencasts and online discussions. Online screencasts will consist of lectures that introduce data mining concepts and workshops that demonstrate how to use R to accomplish data mining tasks. Discussions will be conducted asynchronously on Canvas. You are expected to contribute your answers, thoughts, and questions and engage with your classmates in a civilized but intellectually rigorous discussion. I am also available to meet one-on-one by phone, video chat, or in-person.
Dropping and Adding
Students are responsible for understanding the policies and procedures about add/drops, academic renewal, etc. Information on add/drops are available at http://www.sjsu.edu/advising/faq/index.htm#add
Information about late drop is available at http://www.sjsu.edu/aars/policies/latedrops/. Students should be aware of the current deadlines and penalties for adding and dropping classes.

University Policies
Academic integrity
Students should know the University’s Academic Integrity Policy that is available at http://www.sa.sjsu.edu/download/judicial_affairs/Academic_Integrity_Policy_S07-2.pdf.
Your own commitment to learning, as evidenced by your enrollment at San Jose State University and the University’s integrity policy, require you to be honest in all your academic course work. Faculty members are required to report all infractions to the office of Student Conduct and Ethical Development. The website for Student Conduct and Ethical Development is available at http://www.sa.sjsu.edu/judicial_affairs/index.html
 Instances of academic dishonesty will not be tolerated. Cheating on exams or plagiarism (presenting the work of another as your own, or the use of another person’s ideas without giving proper credit) will result in a failing grade and sanctions by the University. For this class, all assignments are to be completed by the individual student unless otherwise specified. If you would like to include in your assignment any material you have submitted, or plan to submit for another class, please note that SJSU’s Academic Policy F06-1 requires approval of instructors.

Campus Policy in Compliance with the American Disabilities Act
If you need course adaptations or accommodations because of a disability, or if you need to make special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Presidential Directive 97-03 requires that students with disabilities requesting accommodations must register with the DRC (Disability Resource Center) to establish a record of their disability.

College of Business Program Goals:
(Not all program learning goals are covered in every course)

1. Business Knowledge
   • Understand basic business principles and demonstrate discipline-specific competencies as applied to local and global environments.

2. Communication
   • Communicate ideas clearly, logically, and persuasively in oral and written format, using technology appropriately.
3. Ethical Awareness
   • Recognize, analyze, and articulate solutions to ethical issues that arise in business.

4. Leadership, Teams and Diversity
   • Comprehend the challenges and opportunities of leading and working in diverse teams and environments.

5. Critical Thinking
   • Comprehend, analyze, and critically evaluate complex and unstructured qualitative and quantitative business problems, using appropriate tools and technology.

6. Innovation
   • Recognize, analyze, and articulate strategies for promoting creativity and innovation.
## Course Schedule

*The schedule is subject to change with fair notice.*

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics and Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>August 17–23</td>
<td><strong>Getting Started with R, Model Measures</strong>&lt;br&gt;Read the syllabus and watch lecture videos&lt;br&gt;Install R&lt;br&gt;Turn in Homework 1</td>
</tr>
<tr>
<td>2</td>
<td>August 24–30</td>
<td><strong>Linear Regression and Logistic Regression</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 2</td>
</tr>
<tr>
<td>3</td>
<td>August 31 – September 6</td>
<td><strong>Classification and Regression Trees</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 3</td>
</tr>
<tr>
<td>4</td>
<td>September 7 – 13</td>
<td><strong>K-Nearest Neighbors</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 4</td>
</tr>
<tr>
<td>5</td>
<td>September 14–20</td>
<td><strong>Ensemble Methods</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 5</td>
</tr>
<tr>
<td>6</td>
<td>September 21 – 27</td>
<td><strong>Clustering</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 6</td>
</tr>
<tr>
<td>7</td>
<td>September 28 – October 4</td>
<td><strong>Association Rules</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Homework 7</td>
</tr>
<tr>
<td>8</td>
<td>October 5 – 11</td>
<td><strong>Big Data</strong>&lt;br&gt;Watch lecture videos&lt;br&gt;Turn in Final Project</td>
</tr>
</tbody>
</table>