Report on the Establishment of Academic Programming Offered Through Distance Education

Master of Science in Data Mining
and
Graduate Certificate in Data Mining

Central Connecticut State University
September, 2007
Descriptive Information

I. Individuals with Institutional Responsibility for the academic quality of distance education programming are:
   Dr. Paulette Lemma, Assoc. V.P. for Academic Affairs, Dean, School of Graduate Studies
   Phone: (860)832-2363 Email: lemma@ccsu.edu

   Dr. Susan Pease, Dean, School of Arts and Sciences
   Phone: (860)832-2605 Email: pease@ccsu.edu

II. Questions on the content of this report should be referred to:
   Dr. Paulette Lemma, Assoc. V.P. for Academic Affairs, Dean, School of Graduate Studies
   Phone: (860)832-2363 Email: lemma@ccsu.edu

   Dr. Daniel Larose, Associate Professor, Mathematics; Coordinator of Data Mining Program
   Phone: (860)832-2862 Email: larosed@ccsu.edu

III. Information about CCSU’s NEASC reaccreditation can be found on the web at:
   http://www.ccsu.edu/NEASC/Default.htm

   Information about CCSU’s Graduate Programs can be found on the web at:
   http://www.ccsu.edu/grad

   Information about CCSU’s Data Mining program can be found on the web at:
   http://www.ccsu.edu/datamining

   To determine if online learning is for you, technical requirements, view frequently asked questions, determine cost, Help Desk access and more, visit Online CCSU at:
   http://www.ccsu.edu/onlineccsu

   To learn about Blackboard Vista, the course management system used by CCSU, including system requirements, visit:
   http://www.ccsu.edu/vista/webct_vista_help.htm

   For Faculty, Student and Help Desk resources, visit:
   http://www.ccsu.edu/vista/

IV. CTDLC help services can be accessed from the web at:
   http://www.ctdlc.org/help/index.cfm

   Course information through CTDLC is listed at:
   http://www.ctdlc.org/courseoffer/search.cfm

V. CCSU’s consortial partner is CTDLC (CT Distance Learning Consortium). Previously, CCSU was in a consortium with CSUS. Although this still exists for specific purposes, CCSU has a contractual relationship with CTDLC for help desk services, security, backups etc.
NEASC Report Narrative
Master of Science in Data Mining/Graduate Certificate in Data Mining
Following NEASC Guidelines on Preparing Reports
for Academic Programming Offered Through Distance Education

Narrative

CCSU offers a Master of Science in Data Mining, with all core courses available for online delivery. We believe that this program is the only Master of Science in Data Mining in the world available online. Students from around the world have taken our courses, including France, Argentina, Nigeria, Australia, and others, as well as most of the 50 States. The primary rationale for the online delivery mode is enrollment: It is questionable whether a sufficient student base exists in the central Connecticut area to support the running of the data mining curriculum on a regular basis. However, all courses have been approved for both online and on-campus delivery. The course content is the same for both delivery modes and many are offered on campus.

The MS in Data Mining (www.ccsu.edu/datamining) was licensed by the Board of Governors of the Connecticut Department of Higher Education (DHE) in 2001, and was accredited in 2003. Recently, the DHE approved a revision of the MS in Data Mining, and also approved a new Graduate Certificate in Data Mining.

MS in Data Mining (36 credits)
(All courses 3 credits unless otherwise indicated.)

Core Courses (27 credits)
- Stat 416 Mathematical Statistics II
- Stat 521 Introduction to Data Mining (4 credits)
- Stat 522 Data Mining Methods (4 credits)
- Stat 523 Applied Data Mining (4 credits)
- Stat 525 Web Mining
- Stat 526 Data Mining for Genomics and Proteomics
- Stat 527 Text Mining
- Stat 570 Applied Multivariate Analysis

Capstone: Thesis Course
- Stat 599 Thesis (3 credits)

Elective Courses (6 credits)
Choose two of:
- Stat 455 Experimental Design
- Stat 529 Current Issues in Data Mining
- Stat 551 Applied Stochastic Processes
- Stat 567 Linear Models
- Stat 575 Mathematical Statistics III
- CS 570 Topics in Artificial Intelligence: Machine Learning
- CS 580 Topics in Advanced Database: Data Mining
- Other appropriate graduate courses, with permission of advisor.
Applicants to the Master of Science in Data Mining program are expected to have completed, or be in the process of completing, Math 221: Calculus II; Stat 315: Mathematical Statistics I; and a second semester course in undergraduate statistics. Students may be admitted on condition that they complete these prerequisite courses with a grade of B or better.

The Graduate Certificate in Data Mining is a non degree program, consisting of 18 credits. The certificate was designed to appeal to mid-career professionals, looking to retool for the global information economy. Many of these students do not have the time or inclination to complete a 36-credit program, making the Graduate Certificate more appropriate for these students. However, once they complete the certificate, students may elect to apply to the MS program. If they are admitted, all of the credits may be counted toward the master’s degree program.

Applicants to the Graduate Certificate program in Data Mining are expected to have completed, or be in the process of completing, a second semester course in undergraduate or graduate statistics. Students may be admitted on condition that they complete this prerequisite course with a grade of B or better.

**Admission Criteria**

Students must hold a bachelor’s degree from a regionally accredited institution of higher education. A minimum undergraduate GPA of 3.00 on a 4.00 point scale (where A is 4.0), or its equivalent, and good standing (3.00 GPA) in all post-baccalaureate course work is required. Conditional admission may be granted to candidates with undergraduate GPA’s of at least 2.40, conditioned on them earning grades of B or better in their first three core courses in the program.

In addition to the materials required by the School of Graduate Studies, three sets of additional materials are required to be considered for acceptance: (1) A formal application essay of 500 – 1000 words that focuses on (a) academic and work history, (b) reasons for pursuing the Master of Science or Graduate Certificate in Data Mining, and (c) future professional aspirations (the essay will also be used to demonstrate a command of the English language); (2) A detailed, itemized letter explaining whether and how the candidate has fulfilled each of the program prerequisites that applicants are expected to have completed or be in the process of completing; (3) Two letters of recommendation. It is preferred but not required that one letter be written to attest to the applicant’s academic background and one letter be written to attest to professional background.

---

**Graduate Certificate in Data Mining (18 credits)**

**Core Courses (12 Credits)**
- Stat 521 Introduction to Data Mining (4 credits)
- Stat 522 Data Mining Methods and Models (4 credits)
- Stat 523 Applied Data Mining (4 credits)

**Elective Courses (6 Credits)**
Choose two of:
- Stat 525 Web Mining
- Stat 526 Data Mining for Genomics and Proteomics
- Stat 527 Text Mining
- Stat 529 Current Issues in Data Mining
- Some other graduate-level data mining or statistics course, with approval of program coordinator.
About the Students

About half of the students already have master’s degrees; 15% of the proportion in the program hold the terminal degree of a Ph.D. For example, in a single section of Stat 525 (Web Mining) in summer 2005, there were 3 Ph.D.’s out of 15 students. In addition, the diversity of student background is remarkable. For example, in Stat 521 in Fall 2005, students had degrees in Business Administration, Economics, Biochemistry and Molecular Biology (Ph.D.), Mathematics, Physics, Applied Math and Statistics, Chemistry, Technology Management, Computer Applications, Information Systems, and French. About half of the data mining students reside outside of Connecticut.

Graduates

There have been five graduates of the MS in data mining. They defended their theses, titled as follows:

- Rafiqul Islam: Knowledge Discovery in Microarray Data (December 2004).
  - Rafiqul is now working for a genomics company in New Jersey.
- James Steck: NETPIX: A Method of Feature Selection Leading to Accurate Sentiment-Based Classification Models (May 2005).
  - James is now working for a bank in the state of Washington. James resided in Washington state and was the first student to complete his course work completely online. He flew to CCSU in May 2005 to defend his thesis.
- Eric Tayor: Comparing Unsupervised Multivariate Normal Cluster Results between Datasets and Consolidating Similar Clusters (May 2005).
  - Eric is currently working for Pratt and Whitney in Connecticut.
- Steven Barbee: The Discovery by Data Mining of Rogue Equipment in the Manufacture of Semiconductor Devices (May 2007)
  - Steve has recently retired from IBM, and is now working for SPSS.
- Kathleen Alber: Identifying Patterns of Potentially Preventable Emergency Department Utilization by American Children (May 2007)
  - Kathleen is taking care of a husband (recovering from cancer) and four children at home, and plans to return to the workforce soon.

Note the diversity of the theses: DNA-data, text-based data, cluster-data, semiconductor data, and hospital patient data.

Active Matriculants

We currently have 42 matriculants who are active (having taken a course since fall 2005). Below is a graph of the credits that these matriculants have completed. There are six students with 27 – 30 credits completed who should be ready to graduate within the next year.
The MS in data mining welcomes about eight new active matriculants per year.

What is Data Mining?

According to the Gartner Group, “Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.”

Employer Demand

As early as 1984, in his book Megatrends, John Naisbitt observed that “We are drowning in information but starved for knowledge.” The problem today is not that there is not enough data
and information streaming in. We are in fact inundated with data in most fields. Rather, the problem is that there are not enough trained human analysts available who are skilled at translating all of this data into knowledge, and thence up the taxonomy tree into wisdom. Industry is only beginning to realize how data mining can be used to leverage their existing data into actionable and profitable results. The data mining wave is still building.

A quick search of the phrase “data mining” on Monster.com on June 28, 2007 found 1446 positions nationwide. A sample position is the following, from Deloitte and Touche: Human Capital Actuarial AQS Senior Consultant, Hartford. The position announcement stated, “We employ data mining analyses to determine our clients’ key success drivers”, and required the following: a minimum of “2 – 5 years data mining and predictive modeling experience”, “extensive knowledge of tools for data mining and statistics (SAS, SPSS, MATLAB)”, and finally an “advanced degree in statistics, actuarial science, artificial intelligence, or data mining (emphasis added).”

The demand for graduates with data mining expertise far outstrips the supply as indicated by the following quote from Information Week that reports: “the hunt is on for data mining talent. Almost three-fourths of IT staff members involved in data mining said they've been contacted by headhunters in the past year; for data mining managers, it's almost nine out of ten. ‘Data mining and analysis people are hard to find,’ confirms Sandy Sully, VP and CIO of Xilinx Inc., a semiconductor company in San Jose, Calif., with 104 IT employees. Sully hires data mining personnel on a per-diem basis through an agency, and the rates can exceed $1,000 a day per worker.” (National Salary Survey, Information Week, 1999; website: www.informationweek.com/731/salsurvey.htm)

Enrollment Information

701 enrollments (matriculated and non matriculated students) have occurred in the last seven years. (See Tables 1 through 4.)
<table>
<thead>
<tr>
<th>Semester</th>
<th>Course</th>
<th>Course Title</th>
<th>Professor</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer 2001</td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>9</td>
</tr>
<tr>
<td>Fall 2001</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larosev</td>
<td>16</td>
</tr>
<tr>
<td>Spring 2002</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods</td>
<td>Daniel Larose</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>CS 501*</td>
<td>Foundations of Computer Sci</td>
<td>Joan Calvert</td>
<td>7</td>
</tr>
<tr>
<td>Summer 2002</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods</td>
<td>Daniel Larose</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CS 501*</td>
<td>Foundations of Computer Sci</td>
<td>Joan Calvert</td>
<td>7</td>
</tr>
<tr>
<td>Fall 2002</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Stat 523</td>
<td>Applied Data Mining</td>
<td>Daniel Larose</td>
<td>9</td>
</tr>
<tr>
<td>Spring 2003</td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>5</td>
</tr>
<tr>
<td>Summer 2003</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods</td>
<td>Daniel Larose</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Stat 525</td>
<td>Web Mining</td>
<td>Daniel Larose</td>
<td>5</td>
</tr>
<tr>
<td>Fall 2003</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Stat 523</td>
<td>Applied Data Mining</td>
<td>Daniel Larose</td>
<td>18</td>
</tr>
<tr>
<td>Spring 2004</td>
<td>CS 570</td>
<td>Topics in Artificial Intelligence: Machine Learning</td>
<td>Zdravko Markov</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods</td>
<td>Daniel Larose</td>
<td>12</td>
</tr>
<tr>
<td>Summer 2004</td>
<td>Stat 570</td>
<td>Applied Multivariate Analysis</td>
<td>Daniel Miller</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Stat 570</td>
<td>Applied Multivariate Analysis</td>
<td>Daniel Miller</td>
<td>2</td>
</tr>
<tr>
<td>Fall 2004</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Stat 523</td>
<td>Applied Data Mining</td>
<td>Daniel Larose</td>
<td>9</td>
</tr>
<tr>
<td>Spring 2005</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Stat 525</td>
<td>Web Mining</td>
<td>Daniel Larose</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>CS 570</td>
<td>Topics in Artificial Intelligence: Machine Learning</td>
<td>Zdravko Markov</td>
<td>2</td>
</tr>
<tr>
<td>Summer 2005</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods and Models</td>
<td>Daniel Larose</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>21</td>
</tr>
<tr>
<td>Fall 2005</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Stat 525</td>
<td>Web Mining</td>
<td>Daniel Larose</td>
<td>15</td>
</tr>
<tr>
<td>Spring 2006</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Roger Bilisoly</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods</td>
<td>Daniel Larose</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Stat 526</td>
<td>Data Mining for Genomics and Proteomics</td>
<td>Darius Dzuba</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Stat 570</td>
<td>Applied Multivariate Analysis</td>
<td>Daniel Miller</td>
<td>6</td>
</tr>
<tr>
<td>Summer 2006</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>8</td>
</tr>
<tr>
<td>Fall 2006</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Larose</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Stat 533</td>
<td>Applied Data Mining</td>
<td>Daniel Larose</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Stat 526</td>
<td>Data Mining for Genomics and Proteomics</td>
<td>Darius Dzuba</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Krishna Saha</td>
<td>15</td>
</tr>
<tr>
<td>Spring 2007</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Krishna Saha</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Stat 522</td>
<td>Data Mining Methods and Models</td>
<td>Daniel Larose</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Stat 527</td>
<td>Text Mining</td>
<td>Roger Bilisoly</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Stat 416</td>
<td>Mathematical Statistics II</td>
<td>Daniel Miller</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Stat 570</td>
<td>Applied Multivariate Analysis</td>
<td>Daniel Miller</td>
<td>10</td>
</tr>
<tr>
<td>Summer 2007</td>
<td>Stat 521</td>
<td>Introduction to Data Mining</td>
<td>Daniel Miller</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>CS 580</td>
<td>Topics in Advanced Database: Data Mining</td>
<td>Zdravko Markov</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td>701</td>
</tr>
</tbody>
</table>
Table 2. Core Enrollment Summary by Semester

<table>
<thead>
<tr>
<th>Semester</th>
<th>Courses</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer 2001</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Fall 2001</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Spring 2002</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>Summer 2002</td>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>Fall 2002</td>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>Spring 2003</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Summer 2003</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>Fall 2003</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>Spring 2004</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>Summer 2004</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>Fall 2004</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>Spring 2005</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>Summer 2005</td>
<td>3</td>
<td>43</td>
</tr>
<tr>
<td>Fall 2005</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>Spring 2006</td>
<td>5</td>
<td>65</td>
</tr>
<tr>
<td>Summer 2006</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Fall 2006</td>
<td>4</td>
<td>57</td>
</tr>
<tr>
<td>Spring 2007</td>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td>Summer 2007</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>701</td>
</tr>
</tbody>
</table>

Table 3. Core Enrollment Summary by Course

<table>
<thead>
<tr>
<th>Course</th>
<th>Course Name</th>
<th>Offerings</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stat 521</td>
<td><em>Introduction to Data Mining</em></td>
<td>17</td>
<td>247</td>
</tr>
<tr>
<td>Stat 522</td>
<td><em>Data Mining Methods</em></td>
<td>8</td>
<td>81</td>
</tr>
<tr>
<td>Stat 523</td>
<td><em>Applied Data Mining</em></td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>Stat 525</td>
<td><em>Web Mining</em></td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>Stat 416</td>
<td><em>Mathematical Statistics II</em></td>
<td>9</td>
<td>96</td>
</tr>
<tr>
<td>Stat 570</td>
<td><em>Applied Multivariate Analysis</em></td>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>Stat 526</td>
<td><em>Data Mining for Genomics and Proteomics</em></td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Stat 527</td>
<td><em>Text Mining</em></td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>CS 501</td>
<td><em>Foundations of Computer Sci</em></td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>CS 570</td>
<td><em>Topics in Artificial Intelligence: Machine Learning</em></td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>CS 580</td>
<td><em>Topics in Advanced Database: Data Mining</em></td>
<td>8</td>
<td>104</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>61</td>
<td>701</td>
</tr>
</tbody>
</table>

Table 4. Core Enrollment Summary by Professor

<table>
<thead>
<tr>
<th>Developer / Instructor</th>
<th>Offerings</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel Larose</td>
<td>29</td>
<td>349</td>
</tr>
<tr>
<td>Daniel Miller</td>
<td>14</td>
<td>131</td>
</tr>
<tr>
<td>Zdravko Markov</td>
<td>10</td>
<td>115</td>
</tr>
<tr>
<td>Joan Calvert</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Roger Bilisoly</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>Darius Dziuda</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Krishna Saha</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>701</td>
</tr>
</tbody>
</table>
(a) Institutional Mission

**Standard 1.1** The mission of the institution defines its distinctive character, addresses the needs of society and identifies the students the institution seeks to serve, and reflects both the institution’s traditions and its vision for the future. The institution’s mission provides the basis upon which the institution identifies its priorities, plans its future and evaluates its endeavors; it provides a basis for the evaluation of the institution against the Commission’s Standards.

The MS in Data Mining is consistent with and serves to fulfill the mission and purpose of CSU, CCSU, and Graduate Studies at CCSU. The mission of the Connecticut State University System emphasizes “the ability to analyze data [and] use information technology.” The MS in Data Mining fulfills this goal through its objectives of introducing students to the concepts and techniques of data mining, acquainting students with the wide range of data mining applications, such as customer relationship management, and helping students achieve a deep familiarity with state-of-the-art software packages for discovering meaningful patterns in large data sets. Part of the CSU System mission is to “forge strategic relationships with businesses and community organizations to develop fully the human capital within CSU and Connecticut.” The MS in Data Mining fulfills this goal through its objective of establishing a symbiotic relationship with area businesses, working closely with them to enhance the skill base of Connecticut’s workforce. The MS also reaches businesses throughout the US and many other countries. In fact, most of our students and graduates are full-time employees.

The mission of CCSU promotes “intellectual achievement and professional competence” of our students. The MS in data mining fulfills this goal through its objectives of fostering an understanding of which problems can be solved by data mining methods, an appreciation that the wrong methodology applied to a certain problem may be worse than useless, of making sure that the students understand the mathematical and statistical foundation that these techniques rest upon, and of familiarizing students with an understanding of how and why these techniques work, rather than blindly applying them.

Finally, the mission statement for the Graduate School at CCSU encourages us to “provide students with the knowledge and skills to make contributions to their discipline and to the rapidly changing world. [Thus] our graduates develop mastery in their field and become lifelong learners and leaders within their respective professions.” The MS in Data Mining fulfills these goals through its objectives of introducing students to powerful data mining techniques such as decision trees, association rules, clustering, statistical modeling, Bayesian classification, k-nearest neighbors, regression trees, naïve Bayes, optimization, genetic algorithms; of making students familiar with text data mining, and the software applications used to perform text classification and knowledge extraction, introducing students to data mining techniques for web-based data, and data mining for genomics and proteomics.

(b) Planning and Evaluation

**Standard 2.5** The institution has a system of periodic review of academic and other programs that includes the use of external perspectives.

**Learning Outcomes**

The program is expected to comply with the annual review of all programs as based on the assessment plan submitted to Academic Affairs. This plan contains the program’s expected learning outcomes with the expectation that at least one learning outcome is assessed each
academic year. Students completing the Master of Science in Data Mining will have the following learning outcomes.

1. Be able to approach data mining as a process, by demonstrating competency in the use of CRISP-DM, the Cross-Industry Standard Process for Data Mining, including the business understanding phase, the data understanding phase, the exploratory data analysis phase, the modeling phase, the evaluation phase, and the deployment phase;
2. Be proficient with leading data mining software, including WEKA, Clementine by SPSS, and the R language;
3. Understand and apply a wide range of clustering, estimation, prediction, and classification algorithms, including $k$-means clustering, BIRCH clustering, and Kohonen clustering, classification and regression trees, the C4.5 algorithm, logistic Regression, $k$-nearest neighbor, multiple regression, and neural networks;
4. Understand and apply the most current data mining techniques and applications, such as text mining, mining genomics data, and other current issues.
5. Understand the mathematical statistics foundations of the algorithms outlined above.

In 2006-2007, these learning outcomes were assessed. All Data Mining students met (75%) or exceeded expectations (25%) for the first outcome. Five percent failed to meet expectations for Outcomes 2 through 4 while 80% met and 15% exceeded expectations. Outcome 5 had 15% failing to meet expectations and 75% who met and 10% who exceeded expectations. Faculty are investigating ways to help students who are struggling with specific statistical concepts.

Assessment

Students’ progress throughout the program is assessed through instruments that reflect real-world problems. Students learn data mining by doing data mining. Students are required to analyze case studies using large real-world data sets, investigate the best data mining methods for arriving at a solution to the business problem posed, and clearly communicate, in an executive report, the predictions made, and conclusions drawn and interpreted.

The project assignments are structured so that each of the above learning outcomes is explicitly tested and evaluated. One unique facet of the assessment is that the students must make actual predictions, using their best models, and that these predictions are graded. This faithfully reflects the real-world environment, where no one knows whether one’s predictions are accurate or not until, for example, the customer either responds or does not respond.

Students’ models are evaluated for efficacy, based on the cost-benefit table derived from the data. Student models must maximize the ratio of benefits (e.g., donations, or positive hits) versus costs. Student model performance is compared against the baseline performance, which serves as a benchmark. These baseline models may take the form of null models or send-to-everyone models. Model performance which falls short of baseline model performance is considered inadequate. Each assessment is project-specific, for example, the highest total amount of money collected from a set of potential donors in a charity database, or the optimal proportion of contacts to make in a direct-mail marketing problem.

However, it is not enough simply to get good results. The program philosophy is that your company is not paying you so that you alone understand the results, but rather so that you can communicate the results to your peers and superiors. This is why we spend a lot of time developing strong report-writing skills. Students are, thus, as well prepared as possible to handle their real-world data mining assignments.
Finally, the thesis capstone experience assesses the ability of the student to bring all of their knowledge to bear on a significant business or research problem. A committee of three faculty members reviews the thesis before acceptance; the student is required to make a formal presentation of the thesis on campus. It also must be approved by the Graduate School Dean.

(c) Organization and Governance

Standard 3.8 The institution’s academic leadership is directly responsible to the chief executive officer, and in concert with the faculty is responsible for the quality of the academic program. The institution’s organization and governance structure assure the integrity and quality of academic programming however and wherever offered. Off-campus, continuing education, distance education, international, evening, and week-end program are clearly integrated and incorporated into the policy formation, and academic oversight, and evaluation system of the institution.

Prior to spring 2007, the software use of Vista blackboard was administered through the Central Connecticut State University System (CSUS) Office for the four institutions, of which one is Central Connecticut State University. CCSU is now in partnership with CTDLC (CT Distance Learning Consortium). Blackboard Learning System will be upgraded to Vista 4.2 for full student use in spring 2008. At CCSU, academic policy is administered through the Provost or Academic Affairs office; course set up and registration is administered through the Enrollment/Continuing Education Office. The software contact is Information Technology and Media Center, which also maintains network and server security.

The rules that apply to on ground academic courses are the same that apply to distance and off campus courses.

(d) Educational Programming

Standard 4.2 Through its system of academic administration and faculty participation, the institution demonstrates an effective system of academic oversight, assuring the quality of the academic program wherever and however it is offered.

The MS in Data Mining program has a faculty coordinator, Daniel T. Larose, Ph.D., who conducts program oversight. He also designed and developed the program.

Standard 4.4 The institution publishes the learning goals and requirements for each program. Such goals include the knowledge, intellectual and academic skills, and methods of inquiry to be acquired. In addition, if relevant to the program, goals include creative abilities and values to be developed and specific career-preparation practices to be mastered.

Both the MS program and Graduate Certificate requirements are published in the Graduate catalog, the graduate website, www.ccsu.edu/grad, as well as on the program website, www.ccsu.edu/datamining.

Standard 4.5 Degree programs have a coherent design and are characterized by appropriate breadth, depth, continuity, sequential progression, and synthesis of learning.

All core courses are offered by the Department of Mathematical Sciences, and all but one are at the 500-level. The core curriculum progresses steadily from the introductory level of data mining through the intermediate to advanced methods, including current issues in data mining. Several of the core courses are based on the Cross-Industry Standard Process for Data Mining (CRISP –
DM), which is industry-neutral, tool-neutral, application-neutral. CRISP-DM was developed in 1996 by analysts representing Daimler-Chrysler, SPSS, and NCR. CRISP provides a non-proprietary and freely available standard process for fitting data mining into the general problem solving strategy of a business or research unit. According to CRISP – DM, a given data mining project has a life cycle consisting of the following six adaptive phases:

- Business or Research Understanding Phase
- Data Understanding Phase
- Data Preparation Phase
- Modeling Phase
- Evaluation Phase
- Deployment Phase

**Standard 4.6** The institution ensures that students use information resources and information technology as an integral part of their education. The institution provides appropriate orientation and training for use of these resources, as well as instruction and support in information literacy and information technology appropriate to the degree level and field of study.

Clearly, since data mining lies near the heart of the information revolution, the MS in Data Mining helps CCSU to ensure information literacy for our students and their employers.

**Standard 4.7** Students completing an undergraduate or graduate degree program demonstrate collegiate-level skills in the English language.

Report-writing is the major component of several courses in the program. It is made clear to students that employers are paying them to share their research with their peers and superiors in easy-to-understand terms. Thus, communication of the results is of nearly equal importance as the results themselves.

**Standard 4.8** The institution develops, approves, administers, and on a regular cycle reviews its degree programs under effective institutional polices that are implemented by designated bodies with established channels of communication and control. Faculty have a substantive voice in these matters.

Quality of the program and course curriculum is enhanced through a faculty driven University curricular process. A formal program review was conducted by the Department of Higher Education (DHE) in 2001 at which time Data Mining was granted licensure. Accreditation through the DHE came in 2003.

The MS Data Mining program also was revised in the same academic year of 2007, changing its curricular expectations for students. The rationale for increasing the number of credits for completing the program as well as changing the curricular offerings was based on the rapid expansion of the data mining cutting-edge field. Faculty believe that the revision significantly enhances the preparation of the students to the latest techniques and methodologies in a variety of applications, such as text mining and genomics and proteomics.

**Standard 4.9** The institution undertakes academic planning and evaluation as part of its overall planning and evaluation to enhance the achievement of institutional mission and program objectives. These activities are realistic and take into account stated goals and perspective and assessment of their effectiveness. Additions and deletions of programs are consistent
with institutional mission and capacity, faculty expertise, student needs, and the availability
of sufficient resources required for the development and improvement of academic
programs. The institution allocates resources on the basis of its academic planning, needs,
and objectives.

As stated above, each graduate program is expected to conduct an annual review that is submitted
to the Dean, School of Graduate Programs as well as to Academic Affairs/Director of
Institutional Research and Assessment. Faculty use the results of program assessments to assess
the progress of their students and to determine if curricular changes are needed.

**Standard 4.10** Institutions undertaking the initiation of degrees at a higher level, off-campus
programs, programs that substantially broaden the scope of the academic offerings, distance
learning programs, academic programs overseas, or other substantive change demonstrate
their capacity to undertake such initiatives and to assure that the new academic programming
meets the standards of quality of the institution and the Commission’s Standards and
policies. The institution recognizes and takes account of the increased demands on resources
made by programs offered at a higher degree level.

CCSU has demonstrated its capacity to undertake distance learning in that the MS in Data Mining
has now successfully graduated five students. One of these students took all his courses from
Washington State; he came to campus for the first time to defend his thesis.

**Standard 4.29** The institution’s degrees and other forms of academic recognition are
appropriately named, following practices common to American institutions of higher
education in terms of both length and content of the programs.

All courses offered as part of the MS and certificate in Data Mining follow the traditional CSU
schedule of Fall, Spring, and Summer. All courses are approved for both online and on-campus
delivery.

**Standard 4.30** The institution offers required and elective courses as described in publicly
available print and electronic formats with sufficient availability to provide students with the
opportunity to graduate within the published program length.

The courses are published in the Graduate catalog, as well as on the program website,
www.ccsu.edu/datamining. Courses are offered with sufficient regularity to allow students the
opportunity to graduate within a reasonable period.

**Standard 4.32** The evaluation of student learning or achievement and the award of credit are
based upon clearly stated criteria that reflect learning objectives and are consistently and
effectively applied. They are appropriate to the degree level at which they are applied.

The criteria for success for courses are provided to the students through course syllabi. These
criteria reflect the learning objectives for the program. Data Mining requires advanced skills and
knowledge; all course syllabi are appropriate for graduate level. In addition course additions or
revisions must be approved by the School of Graduate Studies Dean, the graduate curriculum
committee and the graduate school committee. Final approval comes from the Faculty Senate.

**Standard 4.33** The award of credit is based on policies developed and overseen by the faculty
and academic administration. There is demonstrable academic content for all experiences for
which credit is awarded, including study abroad, internships, independent study, and service
learning. Credit awards are consistent with the course content, appropriate to the field of study, and reflect the level and amount of student learning. No credit toward graduation is awarded for pre-collegiate level or remedial work designed to prepare the student for collegiate study.

All courses are developed by the faculty, following the policies enforced by the University curricular process. Credit hours awarded are based on the content covered and the number of contact hours required.

**Standard 4.36** Faculty, with administrative support, ensure the academic integrity of the award of grades, where applicable, and credit for individual courses. The institution works to prevent cheating and plagiarism as well as to deal forthrightly with any instances in which they occur.

All courses follow the standards in CCSU’s academic integrity policies, which are published on: [http://www.ccsu.edu/AcademicIntegrity/](http://www.ccsu.edu/AcademicIntegrity/). The policy for academic misconduct is also included in the graduate catalog.

**Standard 4.37** The institution offering programs and courses for abbreviated or concentrated time periods or via distance learning demonstrates that students completing these programs or courses acquire levels of knowledge, understanding, and competencies equivalent to those achieved in similar programs offered in more traditional time periods and modalities. Programs and courses are designed to ensure an opportunity for reflection and for analysis of the subject matter and the identification, analysis and evaluation of information resources beyond those provided directly for the course.

The courses offered online are equivalent in every respect except delivery method with the same courses offered on campus. The course descriptions are the same, and each must cover the same content. This is also true for courses that run intensively during the summer, for 5, 8, or 10 weeks. No program courses are offered in a time frame shorter than five weeks, in order to allow opportunity for reflection and analysis of the subject matter.

**Standard 4.38** Courses and programs offered for credit off campus, through technologically mediated instruction, or through continuing education, evening or weekend divisions are consistent with the education objectives of the institution. Such activities are integral parts of the institution and maintain the same academic standards as courses and programs offered on campus. They receive sufficient support for instructional and other needs. Students have ready access to and support in using appropriate learning resources. The institution maintains direct and sole responsibility for the academic quality of all aspects of all programs and assures adequate resources to maintain quality.

All courses offered online for the MS in Data Mining are integral parts of CCSU, and maintain the same academic standards as courses offered on campus. In the introductory course, Stat 521 Introduction to Data Mining, students have recourse to a Graduate Assistant who helps them learn the required software. The CCSU Library maintains superb online resources for research in data mining, statistics, computer science, mathematics, and other fields.

**Standard 4.39** On-campus faculty have a substantive role in the design and implementation of off-campus programs. Students enrolled in off-campus courses and/or distance learning courses have sufficient opportunities to interact with faculty regarding course content and related academic matters.
All courses in the program were developed by full-time on-campus faculty. CCSU provides a small stipend to support the development of courses for online delivery. Students living within driving distance have the option of visiting campus to meet with the faculty for office hours. Otherwise, faculty are available by phone, via email, or via threaded discussions.

**Standard 4.43** The institution accepts graduate credit in transfer on a strictly limited basis to preserve the integrity of the degree awarded.

Only nine of the 36 program credits may be transferred from another institution per graduate policy, and these only if they have not been used for a degree elsewhere. Courses transferred to Data Mining require approval of the Department of Mathematical Sciences.

**Standard 4.44** The institution implements and supports a systematic and broad-based approach to the assessment of student learning focused on educational improvement through understanding what and how students are learning through their academic program and, as appropriate, through experiences outside the classroom. This approach is based on a clear statement or statements of what students are expected to gain, achieve, demonstrate, or know by the time they complete their academic program. The approach provides useful information to help the institution understand what and how student are learning, improve the experiences provided for students, and assure that the level of student achievement is appropriate for the degree awarded. Institutional support is provided for these activities.

Learning outcomes are assessed using a variety of dimensions, ranging from report writing, through model building to mathematical analysis. All outcomes are ultimately synthesized and measured through the thesis capstone.

**Standard 4.45** The institution’s approach to understanding student learning focuses on the course, program, and institutional level. Data and other evidence generated through this approach are considered at the appropriate level of focus, with the results being a demonstrable factor in improving the learning opportunities and results for students.

See Table 5 for curriculum mapping that demonstrates the application of each learning outcomes.

**Standard 4.46** Expectations for student learning reflect both the mission and character of the institution and general expectations of the larger academic community for the level of degree awarded and the field of study. These expectations include statements that are consistent with the institution’s mission in preparing students for further study and employment, as appropriate.

CCSU’s mission statement includes the statement: “*Central Connecticut State University values the development of knowledge and its application.*” Data Mining is often referred to as “knowledge discovery.” The MS in Data Mining fulfills this goal through its objectives of preparing students for the application of data mining techniques and methodologies in the real world, of exposing students to the types of problems encountered in real-life situations when dealing with large data sets, and of preparing students for the application of data mining techniques and methodologies in their capstone experience course. In its vision statement, CCSU “aspires to ... be global in its perspective and outreach.” Online delivery of the entire MS in Data Mining helps to address the basic problem of access. Now students who must work full time, or care for children, or live at a prohibitive distance, can share in the educational opportunities available to the more traditional student. Further, online access will enable CCSU
to achieve its goal of reaching out to students from around the world. The mission of CCSU includes the expectation “that members of the university will engage in ... basic research and the creation of original works.” Since data mining lies at the cutting edge of several fields, opportunities abound for productive research endeavors.

Table 5. Curriculum Map

<table>
<thead>
<tr>
<th>Be able to approach data mining as a process, by demonstrating competency in the use of CRISP-DM, the Cross-Industry Standard Process for Data Mining, including the business understanding phase, the data understanding phase, the exploratory data analysis phase, the modeling phase, the evaluation phase, and the deployment phase.</th>
<th>Stat 521</th>
<th>Stat 522</th>
<th>Stat 523</th>
<th>Stat 525</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand and apply the most current data mining techniques and applications, such as text mining, mining genomics data, and other current issues.</td>
<td>Stat 526</td>
<td>Stat 527</td>
<td>Stat 529</td>
<td></td>
</tr>
</tbody>
</table>

(e) Faculty

**Standard 5.0** The institution develops a faculty that is suited to the fulfillment of the institution’s mission. Faculty qualifications, numbers, and performance are sufficient to accomplish the institution’s mission and purposes. Faculty competently offer the institution’s academic programs and fulfill those tasks appropriately assigned them.

All faculty teaching courses for the MS in Data Mining are full-time, tenured or tenure-track, on-campus faculty. The following is a listing of the faculty. All are in the Department of Mathematical Sciences except Dr. Markov, who is in the Department of Computer Science. (CV’s will be made available on request.)

- **Daniel T. Larose, Ph.D., Program Coordinator and Professor of Statistics.** Dr. Larose is the author of *Discovering Knowledge from Data: An Introduction to Data Mining* (Wiley, 2005), *Data Mining Methods and Models* (Wiley, 2006), and the co-author (with Dr. Zdravko Markov) of *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage* (Wiley, 2007). He is the author of *Discovering Statistics*, an
undergraduate statistics textbook to be published by W.H. Freeman in 2009. His consulting work includes a $750,000 Phase II grant from the Air Force Office of Research, Storage Efficient Data Mining of High Speed Data Streams. He is the Series Editor for the new Wiley Series on Methods and Applications in Data Mining. Dr. Larose teaches Stat 521, Stat 522, Stat 523, and Stat 525.

- **Daniel S. Miller, Ph.D, Professor of Statistics.** Dr. Miller received his Ph.D. in Statistics from the University of Connecticut in 1989. He has been at CCSU for over 20 years teaching statistics and actuarial science courses of all levels. In 1998 he co-developed and has twice co-taught a self-supported on-line review course for the Casualty Society of Actuaries Part 4A examination. He has collaborated with faculty from other disciplines and the Center for Social Research as a surveying and statistical analysis expert, resulting in publication and expert witness testimony. Dr. Miller teaches Stat 416, Stat 570, and Stat 521.

- **Roger Bilisoly, Ph.D, Assistant Professor of Statistics.** An Assistant Professor of Statistics, Dr. Bilisoly received his Ph.D. in Statistics from The Ohio State University in 1998 and his M.S. in Mathematics from Purdue University in 1986. Prior to coming to CCSU in 2004, he was a senior member of technical staff at Sandia National Laboratories in the Geohydrology Department, where he worked on both geostatistical and optimization projects. His current areas of interest include data mining, analysis of scientific data, and spatial statistics. Dr. Bilisoly teaches Stat 521, Introduction to Data Mining, and has developed a state-of-the-art course on text mining, Stat 527. Dr. Bilisoly is under contract to write a book for Wiley Interscience called An Introduction to Text Mining Using Perl.

- **Darius Dziuda, Ph.D, Assistant Professor of Statistics.** Dr. Darius Dziuda has a Ph.D. in Computer Science and extensive academic and biotech experience in data mining and biomarker discovery. His research and professional activities have focused on efficient data mining of biomedical data sets and on identification of small and accurate multivariate markers for genomics, proteomics, drug discovery, and medical diagnosis and prognosis. He is a consultant in bioinformatics and author of the MbMD data mining software system for biomarker discovery. His recent and ongoing collaborations include research projects with Baylor College of Medicine and Virginia Bioinformatics Institute. Dr. Dziuda teaches Stat 527, Data Mining for Genomics and Proteomics, a cutting-edge course which he developed, as well as a state-of-the-art course in biomarker discovery. He is also under contract to write a book for Wiley Interscience titled Data Mining for Genomics and Proteomics.

- **Zdravko Markov, Ph.D., Professor of Computer Science.** Dr. Zdravko Markov has an M.S. in Mathematics and Computer Science and a Ph.D. in Artificial Intelligence. He has been teaching and doing research in the area of Machine Learning and Data Mining for more than 10 years. Dr. Markov has published three textbooks (two in Machine Learning) and more than 40 research papers in conference proceedings.
18

and journals. He has been teaching at CCSU for 10 years in the areas of Computer Architecture and Design, Computing and Communication technology, Machine Learning and Data Mining. His graduate courses are offered in two graduate programs at CCSU - Computer Information Technology and Data Mining. He is co-author (with Dr. Daniel Larose) of *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage* (Wiley, 2007).

- **Krishna Saha, Ph.D., Assistant Professor of Statistics.**
  The newest member of the data mining faculty, Krishna received his Ph.D. from the University of Windsor. He teaches Stat 521 and is developing materials for performing data mining using the *R* language.

**Standard 5.2** The preparation and qualifications of all faculty are appropriate to the field and level of their assignments. Qualifications are measured by advanced degrees held, evidence of scholarship, advanced study, creative activities, teaching abilities, and relevant professional experience, training, and credentials.

*See 5.0*

**Standard 5.3** There are an adequate number of faculty whose time commitment to the institution is sufficient to assure the accomplishment of class and out-of-class responsibilities essential for the fulfillment of institutional mission and purposes. Responsibilities of teaching faculty include instruction and the systematic understanding of effective teaching/learning processes and outcomes in courses and programs for which they share responsibility; additional duties may include such functions as student advisement, academic planning, and participation in policy-making, course and curricular development, research, and institutional governance.

From 2001 – 2005, there were only three faculty members teaching for the data mining program, Drs. Larose, Miller, and Markov. In 2005, a new faculty line was added in support of the data mining program; the new faculty hire, Dr. Darius Dziuda, added a new dimension to the program by developing courses in data mining for proteomics and genomics as well as biomarker discovery. A second statistics faculty member, Roger Bilisoly, started teaching data mining in 2006; he taught text mining for the first time in 2007. Finally, a third statistics faculty member, Krishna Saha, started teaching data mining in 2007. There are now six full-time tenured or tenure-track faculty members teaching for the data mining program. At present Dr. Larose, the program coordinator serves as advisor to all students in the program.

**Standard 5.4** The institution employs an open and orderly process for recruiting and appointing its faculty. Faculty participate in the search process for new members of the instructional staff. The institution ensures equal employment opportunity consistent with legal requirements and any other dimensions of its own choosing; compatible with its mission and purposes, it addresses its own goals for the achievement of diversity of race, gender and ethnicity. Faculty selection reflects the effectiveness of this process and results in a variety of intellectual backgrounds, and training. Each prospective faculty member is provided with a written contract that explicitly states the nature and term of the initial appointment and, when applicable, institutional considerations that might preclude or limit future appointments.
All faculty searches follow the ground rules/policies set down by the University.

**Standard 5.7** Faculty assignments and workloads are consistent with the institution’s mission and purposes. They are equitably determined to allow faculty adequate time to provide effective instruction, advise and evaluate students, contribute to program and institutional assessment and improvement, continue professional growth, and participate in scholarship, research, creative activities and service compatible with the mission and purposes of the institution. Faculty workloads are reappraised periodically and adjusted as institutional conditions change.

All faculty carry a 12-credit load per semester, in addition to their research, creative activities, advising, and professional activities, as set down in the collective bargaining agreement. Some release time is available due to grants, contractual agreements, research time, and dean allocation.

**Standard 5.8** The institution avoids undue dependence on part-time faculty, adjuncts, and graduate assistants to conduct classroom instruction. Institutions that employ a significant proportion of part-time, adjunct, clinical or temporary faculty assure their appropriate integration into the department and institution and provide opportunities for faculty development.

No part-time faculty or adjuncts teach for the data mining program.

**Standard 5.10** Faculty are demonstrably effective in carrying out their assigned responsibilities. The institution employs effective procedures for the regular evaluation of faculty appointments, performance, and retention. The evaluative criteria reflect the mission and purposes of the institution and the importance it attaches to the various responsibilities of faculty, e.g., teaching, advising, assessment, scholarship, creative activities, research, and professional and community service. The institution has equitable and broad-based procedures for such evaluation applying to both full and part-time faculty, in which its expectations are stated clearly and weighed appropriately for use in the evaluative process.

The quality of the duties carried out by the faculty is assessed through the tenure and promotion process, as set down in the collective bargaining agreement. Tenured full professors continue to face sexennial assessments for the remainder of their careers. Faculty are assessed in the four main areas of (a) credit load activity, including teaching, (b) creative activity, including research, (c) service to the Department and the University, and (d) professional activity.

The data mining faculty are especially strong in curricular development and creative activity. For example, four of the six data mining faculty have written or are writing books on data mining.

**Standard 5.12** The institution provides its faculty with substantial and equitable opportunities or continued professional development throughout their careers. Such opportunities are consistent with and enhance the achievement of the institution’s mission and purposes. Faculty accept the obligation to take advantage of these opportunities and otherwise take the initiative in ensuring their continued competence and growth as teachers, scholars, and practitioners.

CCSU provides the faculty with a travel budget of $1200 per year per faculty member to attend research conferences, on condition that the faculty member is presenting at the conference. There is some reassigned time for research available from both CCSU and the CSU system office, but
there is strong competition for this resource. Also available are internal grants that provide funding for professional development and for curriculum development.

(g) Physical and Technological Resources

Standard 8.1  The institution’s physical and technological resources, including classrooms, laboratories, network infrastructure, materials, equipment, and buildings and grounds, whether owned or rented, are commensurate with institutional purposes. They are designed, maintained, and manage at both on-and off-campus sites in a manner that serves institutional needs. Proper management, maintenance, and operation of all physical facilities, including student housing provided by the institution, are accomplished by adequate and competent staffing.

CCSU Information Technology Services (ITS) Department offers a comprehensive series of talks and workshops on the latest instructional and research technology, so that faculty may keep abreast of the latest developments. ITS also supports the annual licensing of data mining software and statistical software. Crucially, ITS provides the resources to make this software available to students and faculty in remote areas, through the useful Citrix server. (In other words, our students around the world can access a variety of software available on CCSU’s server.)

Standard 8.2  Classrooms and other facilities are appropriately equipped and adequate in capacity. Classrooms and other teaching spaces support teaching methods appropriate to the discipline. Students and faculty have access to appropriate physical, technological, and educational resources to support teaching and learning.

A large proportion of CCSU’s classrooms are computer classrooms, with computers available for all students, and an enhanced computer work station for the instructor. Faculty may make special requests for software to be loaded in particular classrooms. Overall, the data mining program is satisfied with the level of technological support it receives from the University and from ITS in particular.

Standard 8.4  The institution undertakes physical resource planning linked to academic and student services, support functions, and financial planning. It determines the adequacy of existing physical and technological resources and identifies and plans the specified resolution of deferred maintenance needs. Space planning occurs on a regular basis as part of physical resource evaluation and planning, and is consistent with the mission and purposes of the institution.

Every year, ITS polls the faculty to determine their needs, in particular, asking which classrooms should be converted to computer classrooms, and which software is needed. The CCSU Library offers access to a wide range of resources that are, importantly for the Data Mining program, available online. Full-text databases include Ebsco Host and Academic Search Premier, among many others. Faculty-Library liaisons canvass faculty to determine which new books should be bought, and which databases should be added.