CONNECTICUT BOARD OF REGENTS FOR HIGHER EDUCATION
Connecticut State Colleges & Universities
*APPLICATION FOR **MODIFICATION OF ACCREDITED PROGRAM***

SECTION 1: GENERAL INFORMATION

Institution: Central Connecticut State University	Date of Submission to CSCU Office of the Provost:
Most Recent NEASC Institutional Accreditation Action and Date: Fifth-year interim report accepted 11/07/2013	

Original Program Characteristics

CIP Code No. 110401
Title of CIP Code: Information Science/Studies
Name of Program: Data Mining
Degree: Title of Award *(e.g. Master of Arts)* **Master of Science**
Certificate: *(specify type and level)*
Date Program was Initiated: Fall semester, 2001
Modality of Program: On ground **X** Online Combined
If "Combined", % of fully online courses?
Total # Cr the Institution Requires to Award the Credential *(i.e. include program credits, GenEd, other)*: 33

Original Program Credit Distribution

Cr in Program Core Courses: **27**
Cr of Electives in the Field: **6**
Cr of Free Electives:
Cr Special Requirements *(include internship, etc.):*
Total # Cr in the Program *(sum of all #Cr above)*: **33**
From "Total # Cr in the Program" above, enter #Cr that are part of/belong in an already approved program(s) at the institution: **33**

Type of Program Modification Approval Being Sought (mark all that apply):

- Licensure and Accreditation (specify whether New Certificate, Minor, Option, Concentration, or Other)
- X Significant Modification of Courses/Course Substitutions*
- Offering of Program at Off-Campus Location (specify new location)
- Offering of Program Using an Alternate Modality (e.g. from on ground to online)
- X Change of Degree Title or Program Title

Modified Program Characteristics

Name of Program: Data Science
Degree: Title of Award *(e.g. Master of Arts)* **Master of Science**
Certificate[1]: *(specify type and level)*
Program Initiation Date: **Fall semester, 2019**
Modality of Program: On ground **X** Online Combined
If "Combined", % of fully online courses?
Total # Cr the Institution Requires to Award the Credential *(i.e. include program credits, GenEd, other)*: **31**
Other: **Student choice increased with the creation of five different tracks: four specialized and one general track. Also, we wish to change the CIP code to 30.3001 (Computational Science) to better reflect the content of our program.**

Modified Program Credit Distribution

Cr in Program Core Courses: **19**
Cr of Electives in the Field: **12**
Cr of Free Electives:
Cr Special Requirements *(include internship, etc.):*
Total # Cr in the Program *(sum of all #Cr above)*: **31**
From "Total # Cr in the Program" above, enter #Cr that are part of/belong in an already approved program(s) at the institution: **19 core credits and two elective courses.**

***Significant** is defined as "more than 15 credits in a previously approved undergraduate degree program or more than 12 credits in a previously approved graduate degree program.

CSCU REVIEW STATUS *(For System Office Use Only - please leave blank)*

Notes regarding Application:
Log of Steps Toward Approval:
Date of Approval:

[1] If creating a Certificate program from existing courses belonging to a previously approved baccalaureate/associate degree program, enter information about that program in the "Original Program" section.

Date for Inclusion in BOR-ASA Meeting Package:
Comments:
Conditions for Approval (if any)

SECTION 1: GENERAL INFORMATION (continued)

If program modification is concurrent with discontinuation of related program(s), please list for such program(s):
Program Discontinued: Data Mining M.S. CIP: 110401 OHE#: 009328 Accreditation Date: 2001
Phase Out Period **2019-2020** Date of Program Termination **Summer, 2020**

Institution's Unit *(e.g. School of Business)* and Location *(e.g. main campus)* Offering the Program: **Engineering, Science, and Technology**

Other Program Accreditation:

- If seeking specialized/professional/other accreditation, name of agency and intended year of review:
- If program prepares graduates eligibility to state/professional license, please identify:

(As applicable, the documentation in this request should addresses the standards of the identified accrediting body or licensing agency)

Institutional Contact for this Proposal: **Daniel Larose, PhD** Title: **Professor** Tel.: **860-832-2862** e- mail: **larosed@ccsu.edu**

SECTION 2: BACKGROUND, RATIONALE AND NATURE OF MODIFICATION

(Please Complete Sections as Applicable)

Background and Rationale *(Please provide the context for and need for the proposed modification, and the relationship to the originally approved program)*

As applicable, please describe:

- How does the program address CT workforce needs and/or the wellbeing of CT society/communities? *(Succinctly present as much factual evidence and evaluation of stated needs as possible.)*

In 2018, Bloomberg called data scientist "the hottest job in America." Business Insider called it "The best job in America right now." Glassdoor.com rated it the best job in the world in 2018 for the third year in a row. The Harvard Business Review called data scientist "The sexiest job in the 21st century."

LinkedIn's August 2018 Workforce Report found that there is a severe shortage of qualified data scientists, with America short 151,171 people with data science skills overall. "The need is particularly acute in New York City (short 34,032 people with data science skills)." The January 2019 report from Indeed.com, a leading jobs website, showed a 29% increase in demand for data scientists year-over-year, and a 344% increase since 2013.

Many of these jobs are right here in Connecticut. A March 20, 2019, search of Indeed.com found 1,729 data science jobs within 100 miles of Hartford, CT, 339 of which were entry level. Monster.com found over 3,000 data scientist jobs within 100 miles of Hartford. Most salary estimates were six figures.

Our Data Mining M.S. degree program began in 2001, and based on our experiences, we have redesigned this degree to remove eight credits of core courses to add five tracks: four tracks allow students to specialize in areas pertinent to Connecticut employers, and the fifth general track allows students to take classes from all other tracks. Doing the following job searches with Indeed.com restricted to Connecticut produced many hits for searches done in March, 2019. Data scientists specializing in the first track (bioinformatics, the analysis of biological data) are sought after by Pfizer, Yale New Haven Hospital, Techne Life Science, UConn Health Center, the Jackson Laboratory Cancer Center. Scientists in the second track (text analytics) appeal to business such as Travelers, Gartner, and Priceline.com. The third track (clustering) appeals to ESPN, Gartner, and Cigna. Finally, the fourth track (computational techniques) is sought after by The Hartford,

Federal Bureau of Investigation (New Haven office), Unilever, and Pratt & Whitney. Clearly, graduates from all tracks would be in demand in Connecticut.

- How does the program make use of the strengths of the institution (*e.g. curriculum, faculty, resources)* and of its distinctive character and/or location?

The online Data Mining M.S. degree was the first such program in the world. Building on our past 18 years of experience, we propose to reduce the number of core credits from 27 to 19 in order to add four specialized tracks, which are (1) based on the research specialties of current faculty and (2) reflect the needs of Connecticut employers. On the second point, as noted above, job searches on Indeed.com restricted to Connecticut produce ads from well-known companies such as Pfizer, Yale New Haven Hospital, Travelers, Gartner, ESPN, Cigna, Pratt & Whitney, and Unilever among others. Hence, we are leveraging the expertise of our faculty while meeting present employer demand in Connecticut.

- Please describe any transfer agreements with CSCU institutions that will become instituted as a result of the approval of this program. *(Please highlight details in the Quality Assessment portion of this application, as appropriate.)*

  n/a

- Please indicate what similar programs exist in other institutions within the CSCU System, and how unnecessary duplication is being avoided.

As mentioned earlier, CCSU was the first university in the world to offer a completely online Master of Science degree in data mining. Essentially, we have been doing data science since 2001. But back then, the term "data science" had not been invented yet. Today, the term "data mining" has evolved into something rather pejorative, similar to "data dredging" or 'fishing", which involves looking for (largely) spurious results in datasets to confirm a post hoc prediction. Thus, the name change.

WCSU and ECSU do not have any programs in data science. SCSU has new programs offering a B.S. and an M.S. in data science. SCSU emphasizes that their M.S. degree will be a Professional Science Masters (PSM), which means the degree caters to students who want to acquire "professional skills" (i.e., business skills). By contrast, CCSU's data science M.S. is focused on developing programming skills, using software packages, and understanding the underlying statistical, computational, and database aspects of data mining/data science. Thus, there is room for both programs in the state of Connecticut to satisfy the already high and growing demand.

- Please provide a description/analysis of employment prospects for graduates of this proposed program.

To reiterate some of the rationale: Data scientists are in demand in Connecticut. There are between 1,700 and 3,000 data science jobs within 100 miles of Hartford, CT. Most salary estimates were six figures. Data scientists specializing in bioinformatics are sought after by Pfizer, Yale New Haven Hospital, Techne Life Science, UConn Health Center, and the Jackson Laboratory Cancer Center. Scientists in text analytics appeal to business such as Travelers, Gartner, and Priceline.com. Scientists completing the clustering track would appeal to ESPN, Gartner, and Cigna. Scientists with computational specialization are sought after by The Hartford, Federal Bureau of Investigation (New Haven office), Unilever, and Pratt & Whitney.

Description of Modification *(Please provide a summary of the modifications to curriculum, admissions or graduation requirements, mode of delivery,etc., and concisely describe how the institution will support these changes.*

Admission requirements have changed in one way – reducing the number of recommendation letters to one rather than two. Note that the M.S. is still entirely online. The main difference is the addition of four tracks so that students can specialize, and the fifth track allows students to pick classes from different tracks or from electives. The rest of the changes have been made to the curriculum as detailed below.

The existing data mining M.S. has six required core courses (four credits each) plus the capstone (three credits), and totals 33 credits. This proposal reduces the required number of core courses to four. The two classes dropped from the core were STAT 526, Data Mining for Genomics and Proteomics, and STAT 527, Text Mining. Both of these were expanded into a track consisting of three 4-credit courses. In addition, two more tracks were developed: Advanced Methods, and Computational Methods (the latter overseen by the Computer Science Department.) Consequently, the M.S. degree has dropped from 33 credits to 31, so the latter is a little less expensive for students, and the number of classes has dropped from 9 to 8 (in three of the tracks), so most students will be able to graduate a little quicker.

Finally, although the data science M.S. is adding new classes, these classes can be handled by existing faculty. First, the computer science track consists of classes that currently run in their department, so there is no additional teaching load. Second, although the three specialized tracks add new classes, existing faculty can cover these at current student levels because we will change from scheduling all the same classes every year to scheduling some classes odd years and others even years. (We do not expect this to delay student graduation, as most students attend part time, and the Program Advisor may substitute courses for students on a deadline.) Third, currently our classes are typically run between 50% and 100% capacity, so we can handle a growing population of students for several years with existing faculty. Fourth, even large numbers of new students could be handled by hiring one or more adjuncts to teach current faculty's introductory-level statistics courses. Details on this are given in Section 3 of this form.

Description of Resources Needed *(As appropriate please summarize faculty and administrative resources, library holdings, specialized equipment, etc. Details to be provided in the next section, as appropriate)*

Given that we are modifying an existing online M.S. program, the main concern is covering new classes for three of the proposed specialized tracks (the fourth track is based on currently running classes in the Computer Science Department, and the fifth track will consist of classes taught in the other tracks, so it doesn't require more faculty load). However, instead of offering the same courses every year, which is done for our current program, we have developed a two-year cycling of classes, which allows the current faculty to teach the larger number of classes. We will need to hire one adjunct instructor to teach introductory-level statistics course, as reflected in the pro-forma budget below.

All program courses are taught by full-time faculty. No new library resources will be required beyond what we can obtain through the Department of Mathematical Sciences, as has been the case since 2001. No new computing equipment or software is required, nor is any software licensing. The faculty coordinator reviews applications, fields inquiries about the program, and We have had a Graduate Assistant every semester since 2001, who helps the students navigate the steep learning curve with the new data science software.

To initiate and sustain this growth, we need to devote resources to *marketing.* Until now, there has been no advertising or marketing of our data mining programs. Marketing plans include:

- Partnering with the American Statistical Association to sponsor an annual Data-thon, where teams would compete to solve a business problem using analytics.
- Getting the word out that the textbooks written by our faculty, including *Data Mining and Predictive Analytics* (Larose and Larose, Wiley, 2015) have been adopted by data science programs, worldwide.

Other Considerations

Previous Three Years Enrollment and Completion for the Program being Modified

ACTUAL Enrollment	Fall Term, 2018		Fall Term, Year 2017		Fall Term, Year 2016	
	Full Time	Part Time	Full Time	Part Time	Full Time	Part Time
Transfers In	0	0	0	0	0	0
New Students	0	5	0	5	0	3
Returning Students	0	13	0	12	1	23
ACTUAL Headcount Enrollment	0	18	0	17	1	26
Fall FTE accounted for by Program Majors*	6.5		5.5		11	
Size of Credentialed Group(s) for Given Year	N/A		3		5**	

*FTE is considered 12 credits for graduate students. Part-time students are calculated at 1/3 FTE.

**Graduation rate for AY15-16 was 7 students.

Curriculum Details for a Program Modification *(to be used as appropriate for specific modification request)*[2]

Course Number and Name [3]	**L.O. #**	**Pre-Requisite**	**Cr Hrs**	**Course Number and Name**	**L.O. #**	**Cr Hrs**
Program Core Courses				**Tracks, continued**		
Core: DATA 511, Introduction Data Science		none	4	Track 2: DATA 531, Text Analytics with Information Retrieval		4
Core: DATA 512, Predictive Analytics, Estimation and Clustering		DATA 511	4	Track 2: DATA 532, Text Analytics with Natural Language Programming		4
Core: DATA 513, Predictive Analytics, Classification		DATA 511	4	Track 2: Elective		4
Core: DATA 514, Multivariate Analytics		DATA 511	4	Track 3: DATA 541, Advanced Estimation Methods		4
Track 1: DATA 521, Introduction Bioinformatics			4	Track 3: DATA 542, Advanced Clustering Methods		4
Track 1: DATA 522, Mining Gene and Protein Expression Data			4	Track 3: DATA 543, Advanced Classification Methods		4
Track 1: DATA 525, Biomarker Discovery			4			
Tracks, continued				**Elective Courses in the Field**		
Track 4: CS 508, Distributed Computing			3			
Track 4: CS 570, Topics in Artificial Intelligence			3			
Track 4: CS 580, Topics in Database Systems and Applications			3			
Track 4: CS 563 or CS 525, Algorithms			3			
Track 5: 12 credits worth of classes from the other tracks			12			
Total Other Credits Required to Issue Modified Credential						

[2] **Details of course changes for Community College institutions should be provided with enough detail to introduce necessary changes in the centralized programmatic database for that system.**

[3] Make any detailed annotations for individual courses as needed to understand the curricular modifications taking place

Learning Outcomes - L.O. *(Please list up to seven of the most important student learning outcomes for the program, and any changes introduced)*

1. Approach data science using a scientific approach via a systematic process that avoids expensive mistakes by assessing and accounting for the true costs of making various errors.
2. Apply data science using a systematic process by implementing an adaptive and iterative framework that includes the following phases: research understanding, data understanding, data exploration, data modeling, evaluation, and deployment.
3. Demonstrate proficiency with leading open-source analytics software such as R and Python, as well as commercial platforms such as IBM/SPSS Modeler.
4. Understand and apply a wide range of clustering, estimation, prediction, and classification algorithms.
5. Demonstrate specialized skills in bioinformatics, text analytics, advanced methods, and/or algorithms.

SECTION 3: RESOURCE AND FINANCIAL CONSIDERATIONS

Two-Year Cost Effectiveness and Availability of Adequate Resources

(Please complete the Pro-Forma Budget – Projected Revenues and Expenditures on the following page. Provide any necessary annotations for the Pro-Forma Budget and other commentary regarding the cost effectiveness and availability of adequate resources for the proposed modification below:

Instead of offering the same courses every year, which is done for the current program, we have developed a two-year cycling of classes, which allows the current faculty to teach the larger number of classes. We do not expect this to delay student graduation, as most students are part time, and the Program Advisor may substitute courses for students on a deadline.

Core Courses for existing Data Mining M.S.	**Cycling**		
STAT 520	Fall		All
STAT 521	Fall	Spring	All
STAT 522		Spring	All
STAT 523	Fall		All
STAT 526	Fall		All
STAT 527		Spring	All
Electives			
STAT 525		Spring	All
STAT 529		Irregular	

The existing data mining M.S. requires the faculty to teach a little more than 8 courses per academic year. The "little more than" is due to STAT 529, which is a topics course that has been used to pilot new courses developed for the data science program. Here is the new course cycling.

Core Courses for proposed Data Science M.S.	**Cycling**		
DATA 511 Introduction to Data Science.	Fall	Spring	All
DATA 512 Predictive analytics: Estimation and Clustering.		Spring	All
DATA 513 Predictive analytics: Classification	Fall		All
DATA 514 Multivariate Analytics.	Fall		All
Bioinformatics Track			
DATA 521 Introduction to Bioinformatics	Fall		Odd
DATA 522 Mining Gene and Protein Expression Data.		Spring	Odd
DATA 525 Biomarker Discovery.	Fall		Even
Text Analytics Track			
DATA 531 Text Analytics with Information Retrieval.		Spring	Odd
DATA 532 Text Analytics with Natural Language Processing.		Spring	Even
Non-core 500-level DATA course			
Advanced Methods Track			
DATA 541 Advanced Estimation Methods.	Fall		Odd
DATA 542 Advanced Clustering Methods.	Fall		Even
DATA 543 Advanced Classification Methods		Spring	Even
Computational Track*			
CS 463 Algorithms		Spring	
CS 508 Distributed Computing		Spring	
CS 525 Advanced Algorithms	Fall		

CS 570 Topics in Artificial Intelligence		Spring	
CS 580 Topics in Database Systems and Applications	Fall		

*CS courses are mostly, but not exclusively online. Most courses have been offered in an online format.

General Data Science Track

12 credits of non-core 500-level DATA and/or CS courses listed within the computational track

Annotations for the pro-forma budget:

[1]We expect enrollments to consist of primarily part-time students because our students are typically employed. We believe that tailoring our program to address the need for training in specialized applications of data science (i.e., the four proposed tracks) will increase enrollment by 20% per year for the next three years, resulting in total enrollments of 22, 26, and 31 part-time students each year.

[2]Tuition reflects 2018-19 current per credit costs; it does not include any other general, online or university fees nor assumes tuition increases for 2019-22. Tuition in this program is $683 per credit regardless of part-time or full-time status. Part-time students are estimated to take 5 credits per semester.

[3]Other revenue reflects the $50 per course fee for online courses. The expected seat count across all online Data Science courses is estimated conservatively at 42, 46, and 52 seats for the next 3 years.

[4]These program expenditures are the same for the embedded OCP in Data Science.

[5]A faculty member receives 3 credits of reassignment per semester to administer the Data Science program, which includes the Master's and the OCP. The cost reflects 12.5% of the faculty member's base salary plus 48% in estimated fringe. Further, the increase across 3 years reflects the 3.7% contractual increase in salary through AY2020 as reflected in Articles 12.3 and 12.4 of the AAUP Contract.

[6]Currently, full-time faculty teach 30 credits per academic year within the Data Science program (both Master's and OCP). Across the next 3 years, we expect the number of credits offered by full-time faculty to increase to 24 in FY19, 27 in FY20, and 34 in FY21. We assume half of these credits to be offered each Fall. We scaled the salary of each faculty member teaching in the program to the number of credits they teach across the academic year and divided by 2 to obtain the Fall semester cost. In Fall 2018, 78 student credit hours were generated by students enrolled in the program (MS and OCP). However, the faculty taught 128 total student credit hours in program courses. Thus, the true cost of faculty teaching within the program was scaled to the student credit hours expected for program candidates as a function of total student credit hours expected. The scaling rates projected for the next 3 years based on enrollment increases within the total program (MS and OCP) are .65, .68, and .71, respectively. We should also note that the proportion of program student credit hours for computer science courses will be far below these scaling rates. As such, the estimates for faculty cost are likely overestimates. We estimated fringe at 48%. The increase across years represents a 3.7% change in salary through AY2020 as reflected in Articles 12.3 and 12.4 of the AAUP Contract. (Note: Faculty members teaching within the program and/or their ranks may change within 3 years.)

[7]We assume one additional undergraduate section taught each year by an adjunct at the Level C rate. Fringe benefits are estimated at 31%. Rates and increases reflect Article 12.7.6 of the CSU-AAUP contract.

[8]The cost of a GA is $5400 per year, or $2700 per semester. We also estimate $1000 per semester in marketing costs. These costs are shared with the OCP in Data Science.

[9]The expense of the Data Science program incorporates both the OCP and MS. To scale the true expense of the MS, we took the total revenue of the Data Science program and calculated the proportion of revenue attributable to the MS and then applied the same scaling factor to the expenses.

PRO FORMA Budget - Projected Revenues and Expenditures
(Whole Dollars Only)

PROJECTED Program Revenue[1]	Fall 2019	Fall 2020	Fall 2021
Tuition (do not include internal transfers)[2]	$ 75,130	$ 88,790	$ 105,865
Program-Specific Fees	$ -	$ -	$ -
Other Revenue (Online course fees)[3]	$ 2,000	$ 2,300	$ 2,600
Total Estimated Program Revenue	**$ 77,130**	**$ 91,090**	**$ 108,465**

PROJECTED Program Expenditures*[4]	Fall 2019	Fall 2020	Fall 2021
Administration (Chair or Coordinator)[5]	$ 21,136	$ 21,918	$ 21,918
Faculty (Full-time, total for program)[6]	$ 52,741	$ 65,132	$ 81,883
Faculty (Part-time, total for program)[7]	$ 6,830	$ 7,204	$ 7,204
Support Staff	n/a	n/a	n/a
Library Resources Program	n/a	n/a	n/a
Equipment (List as needed)	n/a	n/a	n/a
Other (e.g. student services) GA[8]	$ 3,700	$ 3,700	$ 3,700
Estimated Indirect Costs (e.g. student services, operations, maintenance)	n/a	n/a	n/a
Total Estimated Program Expenditures	**$ 84,406**	**$ 97,953**	**$ 114,704**

*Note: Capital outlay costs, institutional spending for research and services, etc. can be excluded.

	Fall 2019	Fall 2020	Fall 2021
Total Program Revenue [OCP+MS][9]	$ 88,790	$ 105,865	$ 126,355
Proportion of Total Program Revenue from MS	0.846	0.839	0.839
Total Expenditures Assignable to MS	**$ 71,408**	**$ 82,183**	**$ 96,237**

This PRO FORMA Budget provides reasonable assurance that the proposed program modification can be established and is sustainable. Some assumptions and/or formulaic methodology may be used and annotated in narrative on page 4 of Application.