Proposal for a New Certificate in Data Science

I. Title: *Certificate in Data Science*

II. Academic year of intended implementation: *Fall 2018*

III. Name of the department and college submitting the proposal.

This is an interdisciplinary program jointly proposed by the
Department of Mathematics and Statistics, College of Natural Sciences
and
Department of Computer Science, College of Engineering, Computer Science, and Construction Management.

A. Identify the unit which will have primary responsibility for the certificate.

The Department of Mathematics and Statistics, and the Department of Computer Science will hold joint responsibility for this certificate.

B. Identify the level of the certificate (i.e., undergraduate or post-baccalaureate).

Undergraduate

IV. Statements on questions of need and demand.

A. Relation of the program to the University Strategic Plan

The certificate in Data Science embraces all six university strategies, however, this proposal is most aligned with strategic priorities #1, #3 and #4.

1. Believing in the primacy of learning, we will continue to develop high-quality learning environments both inside and outside the classroom.

3. Believing in the wise use of new technologies in learning and teaching, we will continue to provide the technology, the related training, and the support needed to create high quality learning environments both inside and outside of the classroom.

4. Believing in the value of service to others, we will continue to serve the educational, cultural, and economic needs of Northern California.

The proposed certificate introduces students to data science by requiring a set of interdisciplinary classes: applied statistics classes, quantitative classes and computer science (programming and database) classes. The certificate provides students with the skills needed to effectively perform data-driven discovery and extract value and competitive intelligence for their field or organization. This certificate cumulates with a High Impact Practice of a student-designed and led project that aims to provide a service in the form of a data product to a local business, researcher, or community member. The skills developed through these courses may be used in pursuing a major, a graduate degree or in their professional careers.

It is increasingly necessary for practitioners and business professionals to be able to have substantial background in data science to be able to turn data into actionable knowledge. Data science plays an important role in many disciplines such as business, biology, informatics, agriculture, etc. Offering a certificate program in Data Science is important to the standing of the Campus as a regional leader in quality education and professional service.

The certificate is designed so that students will gain experience with skills such as inspecting, cleaning, transforming, and modeling data, unstructured data and; computational statistics; pattern recognition; data mining /predictive analytics; machine learning; data visualization; and programming. This certificate will complement a student's academic major and prepare them for a career in which data analysis is required. Students who complete this certificate will be more appealing to employers and graduate schools seeking individuals with data science background.

a. Rationale for seeking a certificate rather than another type of program (degree, option, minor).

The academic and industry need for students trained in Data Science is an immediate one. This certificate proposal is being developed as an intermediate outcome in a longer-term plan for building Chico State's Data Science Capacity.

- *Minor: The total number of units proposed for the certificate is greater than the maximum allowed number of units for a minor.*
- *Option: Options under the major require the student to take all core courses for that major.*
- *Degree: At this time there is not sufficient support or demand for a full four-year degree program. The results from an interest survey conducted in Spring 16 indicated low (14%) interest in a four-year program.*
- *Graduate Degree: Graduate students are at the optimal level to learn Data Science Topics as an enhancement to their current field of study. A Graduate program in Data Science is also under consideration but no formal plan has been developed.*

b. Need for the proposed certificate.

Data Science is at the critical intersection of Statistics, Mathematics, Computer Science and an application field such as Business or Biology. Data scientists have the computational skills to extract different types and quantities of data from multiple sources, ensure consistency of the data, create visualizations and build data products for the information consumer to use. They have the statistical knowledge to build mathematical models and ensure the validity of the data and results. They also have the domain knowledge to gain domain specific insights from the data, and to communicate the results of their work in a verbal or dashboard-like format to non-technical stakeholders. To have this level of proficiency strong skills are needed in Mathematics, Statistics and Computer Science domains.

i. Identify other CSU campuses with the proposed certificate.

No other CSU offers an undergraduate certificate in Data Science. Here are some closely related programs.

- *CSU Fullerton offers a graduate certificate in Data Science*
- *California Polytechnic – San Louis Obispo offers a Cross Disciplinary Minor in Data Science*
- *CSU San Jose offers an Online Business Analytics certificate*

ii. Identify neighboring institutions with the proposed certificate.

No physical neighboring institutions have this proposed certificate.
Here are some examples of popular online certification programs.

- *Johns Hopkins / Coursera*
- *University of Washington*

iii. Identify differences, if any, between these programs and the proposed program.

- *CSU Fullerton: is a graduate certificate.*
- *Cal Poly's degree is a minor only accessible to current Statistics and Computer Science Majors.*
- *Online certificate programs: These self-support programs are arguably less rigorous and with the exception of JHU do not count towards a student's degree progress.*

c. Identify other closely related curricula currently offered by the campus.

The Minor in Computer Science, the Minor in Statistics, and the Minor in Applied Statistics are all programs related to this certificate.

i. Explain the impact the proposed certificate will have on these programs.

The Certificate in Data Science will have a significant number of courses that overlap with all three of these Minor programs, but will serve a different group of students than those pursuing any of the listed minors. Some students may prefer more breadth and less depth in their minor field of study and choose the Data Science certificate over a traditional Minor. We expect this to be limited to few students.

ii. Explain how current programs do not meet the proposed certificate's objectives.

There are no existing courses that present blended material from both Computer Science and Statistics in an application-based approach. The goal is to use computational and statistical theory and methods to produce a domain specific data product that companies and researchers in other disciplines can use to make decisions with. Student Learning Objectives for the Certificate can be found in Appendix A.

d. Student demand for the program.
   i. Give evidence of serious student interest in the proposed certificate.

A campus-wide Data Science Interest Survey was conducted during Spring 2016. Over half (59%) of the 136 respondents were undergraduate students (n=81). These students indicated that they were interested in taking 1-2 additional courses (44%), a Minor (33%), more computationally and statistically rigorous courses (28%) or an Undergraduate Certificate (22%).

The full results from this survey can be found in the appended document: DS Interest Survey 2016.pdf.

ii. Estimated number of students seeking the certificate
   1. in the year of initiation 15
   2. after three years 30
   3. after five years 50
   4. Describe methodology for developing these estimates.

18 undergraduate and 1 graduate student indicated an immediate interest in a certificate. Assuming that fewer than this would be early enough in their academic career to complete the certificate without adding time to degree, this would put the estimate for enrollment in the first year of initiation at 15.

Next we consider that those who are interested in Data Science are likely to be existing Statistics or Computer Science majors and therefore will already have many of the courses required for the Certificate completed (n=43).

Of these 43, an additional 15 students who were not interested in the undergraduate certificate reported other interest in Data Science related curriculum (Minor, Major, Grad certificate or MS). If a Certificate program were available they may be interested in pursuing that since other degree programs are not available. This shows evidence of longer term interest and commitment. Adding half of these 15 to the initial 15 and adding a conservative number of 5 new students per year provides an estimate of 30 students seeking the certificate after three years (either initiating or in the process of taking classes).

After five-years we expect word to have spread campus wide about the program and have steadily increasing interest of approximately 10 new students per year, resulting in approximately 50 students seeking the certificate after five years.

iii. Estimate the number of certificates awarded
   1. in the year of initiation 9
   2. after three years 18
   3. after five years 30
   4. Describe methodology for developing these estimates.

Upon successful passing of this proposal, the first year of initiation of this certificate program would be in AY 2018-19. At that time all courses needed for the program will have been run at least one time as Special Topics and ready for the process of becoming official courses. Any student that has taken these courses as Special Topics courses would be eligible to earn the certificate in Data Science.

Of the courses that teach skills related to this certificate program, the Introduction to R (1 unit CR/NC) has been offered 3 times so far and has an average completion rate of 40%. The S16 run of Introduction to Data Science (MATH 398 - 3 units, graded) started with n=13 enrolled and saw three students drop before census, a 77% retention rate.

Assuming a program retention rate of 60% and current capacity to supervise student projects we could expect to award approximately 9 certificates in the year of initiation, 18 by the end of year 3, and 30 by the end of year 5. If our enrollment and retention estimates are correct we would be awarding approximately 6 certificates per academic year.

e. Identify professional uses for the proposed certificate.

Students who complete this certificate will be vastly more appealing to employers. Upon completion of the certificate program, students will be able to:

- *Understand the complete data analytics project lifecycle*
- *Understand the tools companies are using to acquire, store, and process data*
- *Apply the various statistical analysis and modeling tools*
- *Apply appropriate analytic techniques and tools to analyze big data*
- *Frame/reframe challenges and translate them into analytical framework*
- *Present data results to managers and non-technical audiences using data visualization*

Here are a small few examples of where Data Science is used in the professional field..

- *Hewlett-Packard Enterprises has an entire Data Science Team*
- *Google: Data scientist job posting (10/25/16), As a Data Scientist, you will evaluate and improve Google's products. You will collaborate with a multi-disciplinary team of engineers and analysts on a wide range of problems. This position will bring analytical rigor and statistical methods to the challenges of measuring quality, improving consumer products, and understanding the behavior of end-users, advertisers, and publishers.*
- *Searching Glassdoor.com for Data Scientist resulted in 2,433 full time jobs. One of the 101 internships was from Cardinal Innovations Healthcare Solutions for a Data Science Analyst Intern. This listing contained the following requirements: "…research oriented educational and/or practical experience and approaches, statistical competence, experience in population health management techniques and methods, and data science / analytics experience. While deep programming skills are not required, hands-on experience with handling diverse datasets, programming tools, analytical methods and algorithms, and capabilities is required."*
- *Another internship from the same search was from the CIA. This posting specified that "Bachelor's degree and experience with applied quantitative research working with real world data, either through thesis research, internships, or work experience. Applicants should have demonstrated creativity, initiative, and leadership abilities." The pay scale for this posting started at $61k.*

The proposed certificate directly prepares students for careers in Data Science while not taking away from the depth of their primary field of study. This provides the student with greater opportunity for an entry level position while still allowing them to pursue a graduate level degree in their primary field or in Data Science directly: 88% of big data professionals have advanced degrees. References and links to additional professional and academic references justifying the need for a Data Science program at Chico State can be found in Appendix B.

5. Resources
    a. List the faculty members for the required courses in the program.

The list of courses includes core courses from Mathematics, Statistics and Computer Science. Thus the list of faculty members that can teach most courses in the list of required courses is essentially all existing Mathematics, Statistics and Computer Science faculty (tenured, tenure-track, and adjunct faculty for Calculus).

The list below contains tenured and tenure-track faculty members who have expressed interest, and have the capabilities to teach the new Data Science topic specific courses.

- *Robin Donatello, DrPH 2013, Biostatistics.*
  Assistant Professor in the Department of Mathematics and Statistics
- *Todd Gibson, PhD 2009, Computational Bioscience.*
  Associate Professor in the Department of Computer Science.
- *Kathy Gray, PhD 2007, Mathematics.*
  Professor in the Department of Mathematics and Statistics.
- *Essia Hamouda El Hafsi, PhD 2006, Computer Science.*
  Assistant Professor in the Department of Computer Science.
- *Edward Roualdes, PhD 2015, Statistics.*
  Assistant Professor in the Department of Mathematics and Statistics.
- *Patrick Donnelly, PhD 2015, Computer Science.*
  Assistant Professor in the Department of Computer Science

b. List the faculty members for the elective courses in the program.

- *Sergei Fomin, PhD 1982, Physics and Mathematics,*
  Professor in the Department of Mathematics and Statistics
- *Vladimir Rosenhaus, PhD 1983, Physics and Mathematics,*
  Professor in the Department of Mathematics and Statistics
- *Thomas Mattman, PhD 2000, Mathematics,*
  Professor in the Department of Mathematics and Statistics
- *Colette Calmelet, PhD 1987, Applied Mathematics,*
  Associate Professor in the Department of Mathematics and Statistics
- *Todd Gibson, PhD 2009, Computational Bioscience.*
  Associate Professor in the Department of Computer Science.
- *Elena Harris, PhD 2010, Computer Science,*
  Assistant Professor in the Department of Computer Science
- *Jaime Raigoza, PhD 2013, Computer Science,*
  Assistant Professor in the Department of Computer Science

c. List the resources needed to sustain the program for the first five years, including cost and funding source.

Resources needed to fund this certificate over the next five years are all for faculty instruction. All estimates for increases in existing course load is based on our estimates of certificates awarded / year.

- *6 AWTU per year would be needed to teach the Introduction (385 – Fall) and Advanced (485 – Spring) Data Science topics courses.*

- *CSCI 490 and CINS 490 are existing project-based Capstone courses where the instructor receives .33 AWTU/student. Assuming half of those pursuing the Certificate would already be taking these capstone courses, we would estimate that an additional 3 students per year would need project supervision through MATH 490 or independent study such as MATH 499. This would require 1 AWTU of faculty support.*
- *The addition of a few more students per semester in CSCI 111, CSCI 211 and MATH 314, would result in an increasing rolling waitlist that could be handled with an additional section of each of these courses once every two years.*
- *Most elective courses are either near capacity (CSCI 344/ CINS 370/ CSCI 580) or have not been offered in a while due to low enrollment (CSCI 582/MATH 344). Additional Data science students could provide necessary enrollment numbers for classes that have historically not been able to make. It is estimated that one extra section (or course) should be offered every other year.*

These estimates for additional funding requirements were based on examining trends in historical enrollments of courses that are likely to be impacted by this Certificate. Numbers were obtained from the posted course schedule online. Each row represents a separate section, entries are # enrolled / (# enrolled + # empty seats).

Required Courses	*S15*	*F15*	*S16*	*F16*	*S17*
CSCI 111	*63/63* *70/70*	*60/60* *61/61* *87/90*	*60/60* *55/60*	*61/65* *53/60*	*59/60* *50/50*
CSCI 211	*90/102*	*49/50* *38/50*	*49/50*	*47/50* *24/50*	*44/60* *15/60*
MATH 314	*34/35*	*-*	*40/40*	*27/35*	*35/35*
CSCI 490	*15/17*	*14/30*	*25/32*	*12/30*	*34/35*
CINS 490	*12/17*	*6/30*	*9/20*	*5/30*	*5/20*

Elective Courses	*S15*	*F15*	*S16*	*F16*	*S17*
CSCI 344	*27/35*	*-*	*40/40*	*-*	*31/31*
CINS 370	*78/79*	*38/40*	*62/68*	*37/40*	*35/35*
CSCI 580	*-*	*39/50*	*-*	*35/35*	*34/34*
*CSCI 582**	*13/40*	*-*	*-*	*-*	*-*
MATH 344†	*-*	*-*	*-*	*-*	*-*
MATH 461	*12/25*	*-*	*7/25*	*-*	*17/25*
MATH 480	*11/25*	*-*	*12/25*	*-*	*11/25*

** The S15 offering of CSCI 582 was a Special Topics 598 course. This course was subsequently approved as a new course that semester but has not had a successful offering since then.*
†Math 344 hasn't had a successful offering since Fall 2011 (Enrollment 7/25).

Summary of anticipated funding needed for the first five years: Average of 12.4 AWTU / year

Year	*DS Courses 385, 485*	*MATH 490*	*CSCI 111, CSCI 211, MATH 314*	*Other Electives*
1 (18-19)	*6*	*1*	*3*	
2 (19-20)	*6*	*1*	*6*	*3*
3 (20-21)	*6*	*1*	*3*	
4 (21-22)	*6*	*1*	*6*	*3*
5 (22-23)	*6*	*1*	*3*	

Funding for additional faculty time would be split between the Colleges of NS and ECC.

d. Additional support resources required, including source of support.

No additional funding for non-instructional purposes is needed.

6. Curriculum
   Note: Proposed curriculum should take advantage of courses already offered in other departments when subject matter would otherwise overlap or duplicate existing course content.
   a. Total number of units required for certificate.

36-38 units

Most required and many elective courses count towards major degree programs that students initially interested in this certificate are likely to already be taking. This certificate would add 16-18 units to the Statistics major, and 9 to the Computer Science major.

b. Special criteria for admission and/or continuation (if applicable).

Students must have completed MATH 385 or CSCI 385 with a C or before applying to the certificate program.

c. Explanation of any special program characteristics (e.g., terminology, credit units required, types of coursework, etc.).

Students admitted to the program will be assigned a faculty advisor from either the CS or MATH departments. There will be at least one faculty advisor in each department to advise and supervise students in this certificate program. This will result in a small increase in the number of advisees for the advising faculty members. The Statistics major is relatively small, so this increase will not have a major impact on the current level and quality of advising.

d. List all new courses for the proposed program.
    i. Course number and title, Units of credit, Prerequisites, Proposed catalog description, Mode of course delivery if other than regular.

Curriculum details for the new Data Science courses can be found in Appendix C: Data Science at Chico State – A three course sequence. The course dependency map can be found as Appendix D.

- *MATH 385 / CSCI 385: Introduction to Data Science*
  3 units.
  Prereq: MATH 109 or MATH 120, and CSCI 111 or MATH 130 or MATH 230.

  <u>Proposed catalog description:</u> Data Science is the science of learning from data in order to gain useful predictions and insights. The course will provide an overview of the wide area of data science, with a particular focus on to the tools required to store, clean, manipulate, visualize, model, and ultimately extract information from various sources of data. Topics include: The analytics life cycle, data extraction, integration and modeling in R/Python, data visualization, relational databases and SQL, text processing and sentiment analysis. Emphasis is placed on reproducible research, code sharing, version control, and communicating results to a non-technical audience. This is a 3 hour discussion.

  MATH 385 and CSCI 385 are intended to be "cross-listed" or equivalent courses. This course has been jointly designed by faculty from both Computer Science and Statistics, and will be offered in each department on an alternating/rotating schedule (i.e. F1 / F2)

  An initial run of the Introduction to Data Science course was conducted as MATH 398 in Spring 2016 by Dr. Donatello. The course was revised by Dr. Donatello and Dr. Hamouda jointly during Fall 2016, and offered as CSCI 398 by Dr. Hamouda in Spring 2017.

- *MATH 485 / CSCI 485: Advanced Topics in Data Science*
  3 units.
  Prerequisite: MATH 385, MATH 314, CSCI 211, MATH 456 (co-requisite)

  <u>Proposed catalog description:</u> How to be a successful Data Scientist: Overview and challenges, project management and effective team collaboration tools and methods, workflow automation. Ethics of predictive analytics and privacy and open data. Reporting and dissemination of research using interactive dashboards and web-publishing. Identifying and applying appropriate methods to answer a business or research need. Introduction to current scalable technologies to handle Big Data. Introduction to advanced statistical analysis for Data Science: Sample topics include: Non-Parametric techniques, Simulation methods, Network analysis, Experimental Design, Predictive modeling using Supervised and Unsupervised Machine Learning techniques, Modeling and mapping of Geospatial data. MATH 485 and CSCI 485 are equivalent courses.

MATH 485 and CSCI 485 are also intended to be "cross-listed" or equivalent courses offered in both departments in an alternating schedule (i.e. S1/ S2). The course content will be jointly designed between Computer Science and Mathematics and A Special Topics offering of this course is scheduled for Spring 2018.

Not all topics in the proposed course description can be covered in a single semester in great depth. This is about breadth and exposure to different analytical tools and methods, not about depth and theory. Many topics are covered in greater theoretical depth in other elective classes such as AI and Graph Theory.

- *MATH 490: Data Science Capstone*
  1-3 units.
  Prerequisite: MATH 485, Senior standing, approved project, enrollment in the Data Science Certificate Program.

  <u>Proposed catalog description:</u> Students will work independently to provide a service in the form of a data product to a local business, researcher, or community member. Students provide status reports at weekly meetings, and present their finished project to a group of peers at the end of the semester in an appropriate venue such as at an undergraduate seminar series or poster symposium. Existing capstone courses such as CSCI 490 or CINS 490 or Independent study projects through CSCI 499 or MATH 499 can substitute providing an approved Data Science project.

The final project report must be submitted to and approved by the independent study faculty mentor, with copy provided to the Data Science advisor at the completion of the study. Student will work with the Data Science advisor to display their finished work in a public setting.

ii. Identify the new courses needed to initiate the program.

- *Introduction to Data Science (CSCI 385 / MATH 385)*
- *Advanced Data Science (CSCI 485 / MATH 485)*

iii. Identify the new courses needed during the first two years after implementation.

- *Data Science capstone (MATH 490)*

e. List all required courses for the program.

Lower Division: 22 units

MATH 120 – Analytic Geometry and Calculus, 4 units,
Prerequisites: Completion of ELM requirement; both MATH 118 and MATH 119 (or high school equivalent); a score that meets department guidelines on a department administered calculus readiness exam must be achieved by those who claim high school equivalence.

MATH 121 – Analytic Geometry and Calculus, 4 units,
Prerequisite: MATH 120

CSCI 111 – Programming and Algorithms I, 4 units,
Prerequisites: Completion of ELM requirement

CSCI 211 – Programming and Algorithms I, 4 units,
Prerequisites: CSCI 111 with a grade C or higher

1 course selected from:

MATH 235 – Elementary Linear Algebra, 3 units,
Prerequisite: MATH 121.

CSCI 217 or MATH 217 – Discrete Mathematics, 3 units,
Prerequisite: Completion of ELM, CSCI 111 with a grade C or higher, MATH 119 (or equivalent).

Upper Division: 11-13 units

MATH 314 – Probability and Statistics for Science and Technology, 4 units,
Prerequisite: MATH 121

CSCI 385 or MATH 385 – Introduction to Data Science, 3 units,
Prereq: MATH 109 or MATH 120, and CSCI 111 or MATH 130 or MATH 230.

MATH 456 – Applied Statistical Methods II, 3 units,
Prerequisite: MATH 315 or MATH 314

CSCI 485 or MATH 485 – Advanced Topics in Data Science, 3 units.
Prerequisite: CSCI/MATH 385, CSCI 211, MATH 456 (Can be taken concurrently)

MATH 490 – Data Science Capstone, 1-3 units.
Prerequisite: MATH 485, Senior standing, approved project, Enrollment in the Data Science Certificate Program. Existing capstone courses such as CSCI 490 or CINS 490 or Independent study projects through CSCI 499 or MATH 499 can substitute providing an approved Data Science project

f. List all elective courses for the program.

This list is intended to serve as a list of current courses that are related to Data Science in either theory or application. Students pursuing this certificate in its early stages are likely to come from existing Computer Science, Mathematics, and Statistics majors. The courses listed would be able to serve as an upper division elective for these students' major programs as well. As the Data Science program grows this list can be revised to include appropriate courses from other disciplines.

1 course selected from:

CSCI 344 - Shell Programming, 3 units,
Prerequisite: CSCI 144 or CSCI 211 with a grade C or higher.

CINS 370 - Introduction to Databases, 3 units,
Prerequisite: CSCI 211 with a grade C or higher.

CSCI 580 - Introduction to Artificial Intelligence, 3 units,
Prerequisite: CSCI 311 and CSCI 217 or MATH 217 with a grade C or higher.

CSCI 582 - Bioinformatics, 3 units,
Prerequisite: CSCI with a grade C- or higher; concurrent enrollment or prior completion of either MATH 105 of MATH 314.

MATH 344 – Combinatorial Mathematics and Graph Theory, 3 units
Prerequisite: MATH 121, MATH 330

MATH 461- Numerical Analysis, 3 units
Prerequisites: MATH 235, MATH 260 (may be taken concurrently), completion of the computer literacy requirement.

MATH 480 – Mathematical Modeling, 3 units
Prerequisites: MATH 235, MATH 260.

g. Explain provisions for articulation of the proposed certificate with community college courses.

All lower division courses in both Mathematics and Computer Science are already articulated.

h. Complete catalog copy, including admission and completion requirements.

The Certificate in Data Science

Course Requirements for the Certificate: 36-38 units

Data Science is at the three-way intersection of Statistics, Computer Science and a third domain such as Business or Biology. Data scientists have the computational skills to extract different types and quantities of data from multiple sources, ensure consistency of the data, create visualizations and build data products for the information consumer to use. They have the statistical knowledge to build mathematical models and ensure the validity of the data and results. They also have the domain knowledge to gain domain specific insights from the data, and to communicate the results of their work in a verbal or dashboard-like format to non-technical stakeholders. This certificate provides students with the necessary proficiency in both the Statistics and Computer Science domains and a framework to apply these skills in an interdisciplinary manner.

Students must have completed Introduction to Data Science (MATH/CSCI 385) with a grade of at least "C" before applying for admission to the program.

The following courses, or their approved transfer equivalents, are required of all candidates for this certificate. A grade of at least "C" must be earned in each course of the certificate program.

11 courses required

MATH 120	Analytic Geometry and Calculus	4	FS
Prerequisites: Completion of ELM requirement; both MATH 118 and MATH 119 (or high school equivalent); a score that meets department guidelines on a department administered calculus readiness exam must be achieved by those who claim high school equivalence.			
MATH 121	Analytic Geometry and Calculus	4	FS
Prerequisites: MATH 120			
MATH 314	Probability and Statistics for Science and Technology	4	FS
Prerequisites: MATH 121			
CSCI 111	Programming and Algorithm I	4	FS
Prerequisites: Completion of ELM requirement.			
CSCI 211	Programming and Algorithm II	4	FS
Prerequisites: CSCI 111 or EECE 135, with a grade of C or better.			
MATH 456	Applied Statistical Methods II	3	S2
Prerequisites: MATH 314 or MATH 315			
MATH 490	Data Science Capstone	1-3	FS
Prerequisites: MATH 485 or CSCI 485, Senor Standing, Approved Project, Enrollment in the Data Science Certificate program. Existing capstone courses such as CSCI 490 or CINS 490 or Independent study projects through CSCI 499 or MATH 499 can substitute providing an approved Data Science project.			

1 course selected from:

MATH 235	Elementary Linear Algebra	3	FS
Prerequisites: MATH 121			
MATH 217	Discrete Mathematical Structures	3	FS
Prerequisites: CSCI 111 with a grade of C or higher.			
CSCI 217	Discrete Mathematical Structures	3	FS
Prerequisites: CSCI 111 with a grade of C or higher.			

1 course selected from:

MATH 385	Introduction to Data Science	3	F2
Prerequisites: MATH 109 or MATH 120, and CSCI 111 or MATH 130 or MATH 230			
CSCI 385	Introduction to Data Science	3	F1
Prerequisites: MATH 109 or MATH 120, and CSCI 111 or MATH 130 or MATH 230			

1 course selected from:

MATH 485	Advanced Topics in Data Science	3	S2
Prerequisites: MATH 385, CSCI 211, MATH 456 (Can be taken concurrently)			
CSCI 485	Advanced Topics in Data Science	3	S1
Prerequisites: MATH 385, CSCI 211, MATH 456 (Can be taken concurrently)			

1 course selected from:

CSCI 344	Shell Programming	3	FS
Prerequisites: Grade of C or higher in either CSCI 144 or CSCI 211.			
CINS 370	Introduction to Databases	3	FS
Prerequisites: CSCI 211 with a grade of C or higher.			
CSCI 580	Artificial Intelligence	3	FS
Prerequisites: CSCI 311 with a grade C or higher.			
CINS 582	Bioinformatics	3	SP
Prerequisites: CSCI 211 with a grade of C or higher.			
MATH 344	Combinatorial Mathematics and Graph Theory	3	FA
Prerequisites: MATH 121, MATH 330			
MATH 461	Numerical Analysis	3	SP
Prerequisites: MATH 235, MATH 260 (may be taken concurrently), completion of the computer literacy requirements.			
MATH 480	Mathematical Modeling	3	SP
Prerequisites: MATH 235, MATH 260			

Attach the Undergraduate Program Signature form or the Graduate Program Signature form to the front of the proposal and submit to Undergraduate Education or the Office of Graduate Studies after all department and college reviews are complete

Appendix A. Student Learning Outcomes

This Appendix contains the learning outcomes as stated on the academic program review pages for both the Math and Computer Science departments. For each objective we state the level at which this certificate would meet that objective (I – Introductory, P – Proficient, M – Mastery). For the Math SLO's, the objectives are met at a mastery level according to the level of mastery achieved in relevant classes as determined by relevant classes according to the published curriculum Matrix.

Mathematics

http://www.csuchico.edu/apr/program_portfolios/natural-sci/math/learning_outcomes.shtml

Learning Objective	Class where objective met
Goal 1: General Content	
Demonstrate basic skills and conceptual understanding of differential, integral, and multivariable calculus.	P (121)
Demonstrate basic skills and conceptual understanding as relating to fundamental mathematical objects introduced in our degree core, such as, sets, functions, equations, vectors, and matrices.	I (121)
Demonstrate basic understanding of probability and statistics, relevant to their option.	P (314)
Demonstrate more technical skills and more indepth and broader conceptual understanding in core mathematical areas (such as, analysis, geometry/topology, algebra, applied math, statistics), relevant to their option.	M (314, 456, **485**)
Goal 2: Problem Solving	
Interpret and translate problems into appropriate mathematical language.	M (314, 456, **485**)
Solve problems by applying appropriate strategies and interpreting the results.	M (314, 456, **485**, 499)
Goal 3: Communication	
Demonstrate the ability to effectively and accurately write on mathematical topics relevant to their mathematics option and appropriate to their audience.	M (456, **385, 485**, 499)
Demonstrate the ability to effectively and accurately speak on mathematical topics relevant to their mathematics option and appropriate to their audience.	M (456, **385, 485**, 499)
Goal 4: Proofs Proficiency	
Students can read mathematical proofs, extract the key ideas used in the proof, and convey the logic behind the proof.	
Students demonstrate the ability to write their own rigorous and logically correct proofs.	
Goal 5: Technology	
Students use technology to manipulate mathematical objects (e.g., functions, equations, data sets, etc.)	M (314, 456, **385, 485,** 499)
Students use technology to conduct mathematical explorations, to model problem scenarios, and to analyze mathematical objects.	M (314, 456, **385, 485,** 499)
Goal 6: Life-long Learner	
Students demonstrate the ability to apply mathematics and statistics to new contexts (e.g. in other classes, the workplace, graduate school, classes they teach).	M (499)
Students recognize and appreciate the role that mathematics can play in their futures and in society in general.	M (**485,** 499)

Computer Science

"*To satisfy ABET CAC 2008-2009 Computer Science program criteria, our graduates must demonstrate* [the following SLO's]*"* Computer Science does not include I/P/M levels of mastery in their curriculum matrix. For each learning outcome listed below we list the class in which the curriculum matrix states this SLO is being practiced in, the additional data science courses where this SLO would be met (**in bold – with an estimated mastery level for that course**).

http://www.csuchico.edu/apr/program_portfolios/engineering_comp-sci_construc-manage/comp-sci/learning_outcomes.shtml

Learning Outcome	Current class participation
1. An ability to apply knowledge of computing and mathematics appropriate to the discipline.	111, 211, 217, 490
2. An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution.	211, 490, **385 (M), 485 (M)**
3. An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs.	211, 490
4. An ability to function effectively on teams to accomplish a common goal.	**385(M), 485(M)**
5. An understanding of professional, ethical, legal, security and social issues and responsibilities.	211, 490, **485 (I)**
6. An ability to communicate effectively with a range of audiences.	217, **385 (P), 485 (M)**
7. An ability to analyze the local and global impact of computing on individuals, organizations, and society.	490
8. Recognition of the need for and an ability to engage in continuing professional development.	217, **385(P), 485(M)**
9. An ability to use current techniques, skills, and tools necessary for computing practice.	111, 211, 217, 490, **385 (M), 485 (M)**
10. An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices. [CS]	111, 211, 490
11. An ability to apply design and development principles in the construction of software systems of varying complexity. [CS]	211, 490

Appendix B. Additional professional and academic references justifying the need for a Data Science program at Chico State.

- National Institutes of Health: Big Data to Knowledge https://datascience.nih.gov/bd2k
- Big Data grants from the National Science Foundation https://www.nsf.gov/cise/bigdata/
- UC Davis Data Science Initiative https://www.nsf.gov/cise/bigdata/
- State of Data Science education in 2016 *https://www.tableau.com/about/press-releases/2016/tableau-study-reveals-state-data-education-2016*
- The American Statistical Association revised guidelines for Undergraduate Programs in Statistical Sciences has an emphasis on Data Science. *https://www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx*
- *Self-Service Data Prep is the Next Big Thing for BI" March 28, 2016 http://www.datanami.com/2016/03/28/self-service-data-prep-next-big-thing-bi/*
- Data Science Job board at Kaggle *https://www.kaggle.com/jobs*
- IBM Watson http://www.ibm.com/analytics/us/en/technology/data-science/
- Hewlett-Packard Enterprise http://www8.hp.com/us/en/software-solutions/big-data-analytics-software.html
- Google https://www.google.com/about/careers/search#!t=jo&jid=43148&
- Civis Analytics: Bringing Data Science to Government https://civisanalytics.com/blog/data-science/2016/04/20/bringing-data-science-to-government/
- Pivotal: Big Data and Data Science in Healthcare https://pivotal.io/big-data-for-healthcare
- *Biotechnology http://www.fiercebiotechit.com/special-reports/10-reasons-why-biotech-needs-big-data*
- *Postdoctoral scholar in water data science at Northwestern University http://criticalzone.org/iml/news/story/job-opening-postdoctoral-scholar-in-water-data-science/*
- *How big data and predictive analytics can help manage climate change: http://www.kdnuggets.com/2015/12/big-data-predictive-analytics-climate-change.html*
- The sheer numbers of masters degree programs in Data Science: http://www.mastersindatascience.org/

Appendix C. Data Science at Chico State: A Three Course Sequence

Course 1: Introduction to Data Science

Description: *Data Science is the science of learning from data in order to gain useful predictions and insights. The course will provide an overview of the wide area of data science, with a particular focus on to the tools required to store, clean, manipulate, visualize, model, and ultimately extract information from various sources of data. Topics include: The analytics life cycle, data extraction, integration and modeling in R/Python, data visualization, relational databases and SQL, text processing and sentiment analysis. Emphasis is placed on reproducible research, code sharing, version control, and communicating results to a non-technical audience. Sample application areas: Bioinformatics, Health Informatics, Institutional Research, Business Analytics, Environmental Science and Renewable Energies*

What is Data Science?

- Data Analytics Lifecycle
  - CRISP-DM https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining
- *The many faces of a data scientist*. Why DS is an interdisciplinary field.
  - https://www.datacamp.com/community/tutorials/data-science-industry-infographic
  - http://ucanalytics.com/blogs/career-data-science-analytics-play-strengths/
  - Data Analyst / Data Architect / Data Engineer / Statistician / Database Administrator / Product Engineer / Business Analyst / Data Science Team Leader
- *Application Areas* With examples and faculty links. Bioinformatics, Business Analytics, Health informatics, Environmental Sciences,
  - Invited speakers (20 minutes) to talk about their application area.

How to be a Successful Data Scientist

- Overview and Challenges
  - Hard skills vs. Soft skills - Finding and respecting the balance.
  - Working with a Data Science Team
- Project Management and Team collaboration tools
- Automating the Data Science Project Workflow
  - Version control using Git/Github

Getting Data

- Reading and writing to files from various types
- Scraping the web for data
- Connecting to API's
- Relational Databases with SQL

Preparing Data for Analysis

- Data types & summary statistics
- Quality control
  - Data editing
  - Identify and deal with missing data
- Transformations
- Data aggregation
- Concepts of Tidy data
  - When and how to reshape
- Dates and times

Data Exploration: Uncovering Valuable Information

- Question Formulation
- Data Visualization & Interpretation
- Unstructured data analysis
  - string manipulation
  - regular expressions
  - word frequency & clouds
  - sentiment analysis
- Cluster identification
  - knn, concept of distance
  - Dendograms, Heatmaps

Sharing the Insights: Reporting and Dissemination

- Reproducible reports using Markdown and LaTeX
- Web-based interactive apps using Shiny
- Creating slide decks (Beamer/Slidify)
- Presenting results to a non-technical audience

Course 2: Advanced Topics in Data Science

Description: *How to be a successful Data Scientist: Overview and challenges, project management and effective team collaboration tools and methods, workflow automation. Ethics of predictive analytics and privacy and open data. Reporting and dissemination of research using interactive dashboards and web-publishing. Identifying and applying appropriate methods to answer a business or research need. Introduction to current scalable technologies to handle Big Data. Introduction to advanced statistical analysis for Data Science: Sample topics include: Non-Parametric techniques, Simulation methods, Network analysis, Experimental Design, Predictive modeling using Supervised and Unsupervised Machine Learning techniques, Modeling and mapping of Geospatial data.*

How to be a Successful Data Scientist, continued...

- Overview and Challenges
  - Hard skills vs. Soft skills - Finding and respecting the balance.
  - Working with a Data Science Team
- Project Management and Team collaboration tools
- Automating the Data Science Project Workflow
  - Automation via Makefiles

Privacy, Ethics and Open Data

- Ethics of building and using predictive models
- Open data in a hostile world

Advanced Analytical Methods

Not all topics on this list can be covered in a single semester in great depth. This is about breadth and exposure to different analytical tools and methods, not about depth and theory. Many topics are covered in greater theoretical depth in other classes such as AI and Graph Theory.

- Choosing the right analytic tool for the job. - Hammer vs saw
- Non-Parametric Analysis
  - An overview of alternative tests for when assumptions just aren't met.
- Mindful and Targeted data collection
  - Identifying target metrics, Designing data collection tools, Analyzing results, transforming results to action.
- Simulation Techniques
  - Bootstrapping and Beyond!
- Predictive Analytics
  - Cross Validation
  - Navigating the Machine Learning Soup of Algorithms
    - Classification & Regression
    - Ensemble Models
  - Quantifying model fit, generalization, predictive ability, uncertainty
    - Regularization: Ridge, Lasso, Bias-variance trade-off
    - Feature Selection & Dimension Reduction
- Network Analysis
  - Introduction to Graph Theory
  - Social Networks, Images, protein interactions
- Mapping and Modeling of Spatially correlated data

Big Data Processing and Analysis
- Different architecture needed: HDFS / Hadoop / Apache Spark
- Methods of connecting and using that architecture: sparklyr, NoSQL
- Assessing scalability of models and methods
- Amazon Cloud Services

Advanced Reporting and Dissemination
- Authoring websites using Github/Jekyll
- Installing R on a server (like AWS)
- Creating and deploying Interactive Dashboards
- JavaScript and D3.js, plotly

Course 3 : Capstone Project in Data Science

Description: *Students will work independently to provide a service in the form of a data product to a local business, researcher, or community member. Students provide status reports at weekly meetings, and present their finished project to a group of peers at the end of the semester in an appropriate venue such as at an undergraduate seminar series or poster symposium. Existing capstone courses such as CSCI 490 or CINS 490 or Independent study projects through CSCI 499 or MATH 499 can substitute providing an approved Data Science project.*

The final project report must be submitted to and approved by the independent study faculty mentor, with copy provided to the Data Science advisor at the completion of the study. Student will work with the Data Science advisor to display their finished work in a public setting.

Appendix D. Certificate Course Dependency Map

Course Dependency map for the Proposed Certificate in Data Science