# KENT STATE UNIVERSITY
# CERTIFICATION OF CURRICULUM PROPOSAL

Preparation Date **4-Dec-18** Curriculum Bulletin _____

Effective Date **Fall 2019** Approved by EPC _____

Department

College **AS - Arts and Sciences**

Degree **MS - Master of Science**

Program Name **Data Science** Program Banner Code

Concentration(s) Concentration(s) Banner Code(s)

Proposal **Establish program**

Description of proposal:

**This is the required Program Development Plan to establish a new major within the Master of Science degree program in Data Science. This is a joint effort between the Departments of Mathematical Sciences and Computer Science.**

Does proposed revision change program's total credit hours? ☒ Yes ☐ No

Current total credit hours: **30** Proposed total credit hours **30**

Describe impact on other programs, policies or procedures (e.g., duplication issues; enrollment and staffing considerations; need; audience; prerequisites; teacher education licensure):

**This is a new program in an emerging field that requires significant knowledge in mathematics and computer science. The courses exist and cooperating programs are able to support its development.**

Units consulted (other departments, programs or campuses affected by this proposal):

**A&S cooperating departments**

---

**REQUIRED ENDORSEMENTS**

AM Tonjes — 12/11/18
Department Chair / School Director

— — / / —
Campus Dean (for Regional Campuses proposals)

Mary Ann Haley — 12/14/18
College Dean (or designee)

— / / —
Dean of Graduate Studies (for graduate proposals)

— / / —
Senior Vice President for Academic Affairs and Provost (or designee)

CS/MATH
2

# DATA SCIENCE MASTER'S DEGREE PROPOSAL

**By**

**Department of Mathematical Sciences & Department of Computer Science**

**Anticipated Start Date: Fall 2020**

1. **Designation of the new degree program, rationale for that designation, definition of the focus of the program and a brief description of its disciplinary purpose and significance.**

Data science is an emerging discipline founded on the synthesis of mathematics and computer science, more exactly, from the synthesis of analysis, statistics, databases, big data, artificial intelligence, numerical analysis, graph theory, and visualization. The expression "data science" is relatively new, but the core activities of a data scientist have existed and been developing over many years. These core activities are the organizing, analyzing, understanding, and presenting of data to reveal new information and knowledge. Databases are used for storing and organizing data; statistics, numerical analysis, big data and artificial intelligence are used for analyzing and understanding data; and graph theory and visualization are used to present revealed information. Recent developments in these areas have resulted in enormous growth in data science and especially in data science applications. This explosive growth in applications is changing the world as we know it.

The focus of this program is on tools, methods, theories, and applications of data science; and the purpose of this program is to enable graduates to effectively and purposefully use data science to address societal, scientific, and industrial problems and to develop the expertise and insights that will allow them to successfully identify and respond to the rapidly evolving opportunities for fruitful applications of data science in the workplace. For the State of Ohio this will contribute to creating a cutting edge workforce for state and regional data-enabled industries, educational institutions, and government agencies.

The proposed MS program distinguishes itself as a highly research integrated academic program in line with the University's 2016 Strategic Priority (#2 Distinctive Kent State). Its curriculum blends integrative domain knowledge with fundamental mathematical and computer science skills and knowledge and requires research grade work to be performed on industrial projects with recruitment and participation by a pool of talented research active experts.

1. **Description of the proposed curriculum including identification of any specializations intended to appear on the student transcript (see Section IV).**

Mathematics and computer science are the foundations of data science and its applicability. Strong mathematics and computer science backgrounds will be required to enter the program. Development of this foundation will enable students to engage in innovative applications and/or theoretical research.

Although we expect students to enter the program with core competencies in mathematics and computer science, flexible background modules will be available to enable students with more limited prior knowledge to prepare to successfully complete the degree within 2 years.

Expected proficiency in mathematics includes the standard lower division calculus and linear algebra curriculum (the equivalent of Math 21001 Linear Algebra or Math 21002 Applied Linear Algebra and Math 22005 Analytic Geometry and Calculus III) together with courses in statistics and discrete mathematics (the equivalent of Math 20011 Decision Making Under Uncertainty and Math 31011 Proofs in Discrete Mathematics or CS 23022 Discrete Structures for Computer Science.) Expected proficiency in



CS/MATH 3

computer science includes courses in data structures and database systems design (the equivalent of CS 23001 Computer Science II – Data Structures and Abstraction and CS 33007 Introduction to Database Systems Design).

Mathematics, science, or engineering majors with a good computer science background will likely have the background described above. Since we anticipate other students will be attracted to a strong and vibrant data science master's program, online instruction and support in necessary background areas will be provided in modules. The program committee will determine which, if any, modules students will be required to complete, based on their academic transcripts and/or professional experience. Required modules must be completed prior to full admission to the program.

The thirty (30) credit hour Data Science curriculum includes six core courses (18 credit hours). In the non-thesis track, students take three elective courses (9 credit hours) and an integrative experience (3 credit hours). In the thesis track, students take two elective courses (6 credit hours) and a thesis (6 credit hours).The Data Science core has three mathematics courses and three computer science courses. The elective courses are taken to develop applied and domain background from various disciplines, and the integrative credits are earned through an internship in industry, a data science project course, or a thesis focused on real-world applications.

At least six of the elective credit hours must be taken at the 60000 level. Course descriptions are in Appendix A.

| REQUIRED CORE COURSES (18 HOURS) | | |
|---|---|---|
| MATH 50015 | Applied Statistics | 3 Hours |
| MATH 50024 | Computational Statistics | 3 Hours |
| MATH 50028 | Statistical Learning | 3 Hours |
| CS 63015 | Data Mining Techniques | 3 Hours |
| CS 63016 | Big Data Analytics | 3 Hours |
| CS 63017 | Big Data Management | 3 Hours |
| **Core Total** | | **18 Hours** |

A first group of electives consists of courses in mathematics and computer science that complement or go beyond the core. A second group allows students to explore the applications of data science in diverse areas such as biological sciences, geography, psychology, sociology, and translation science. Elective opportunities will broaden as more data science related courses are developed in the College of Arts and Sciences and across campus.

| ELECTIVE COURSES (9 HOURS, 6 HOURS FOR THESIS TRACK) | | |
|---|---|---|
| MATH 50011 | Probability Theory and Applications | 3 Hours |
| MATH 50051 | Topics in Stochastic Processes and Applications | 3 Hours |
| MATH 50059 | Stochastic Actuarial Models | 3 Hours |
| MATH xxxxx | Advanced Computational Statistics | 3 Hours |
| MATH xxxxx | Advanced Statistical Learning | 3 Hours |
| CS 57206 | Data Security and Privacy | 3 Hours |
| CS 63005 | Advanced Database Systems Design | 3 Hours |
| CS 63018 | Probabilistic Data Management | 3 Hours |
| CS 63100 | Computational Health Informatics | 3 Hours |
| CS 64201 | Advanced Artificial Intelligence | 3 Hours |



CS/MATH 4

| | |
|---|---|
| CS 64402 Multimedia Systems and Biometrics | 3 Hours |
| CS 67302 Information Visualization | 3 Hours |
| CS 6XXX Cloud Infrastructure | 3 Hours |
| BSCI 60103 Biological Statistics | 3 Hours |
| GEOG 59070 Geographic Information Science | 4 Hours |
| GEOG 59072 Geographic Information Science and Health | 3 Hours |
| GEOG 59075 GIS Applications for Social Problems | 3 Hours |
| GEOG 59078 GIS and Environmental Hazards | 3 Hours |
| GEOG 59080 Advanced Geographic Information Science | 3 Hours |
| PSYC 61651 Quantitative Statistical Analysis I | 3 Hours |
| PSYC 61654 Quantitative Statistical Analysis II | 3 Hours |
| CRIM 56803 Information and Cyber Security | 3 Hours |
| SOC 62217 Multivariate Techniques in Sociology | 4 Hours |
| SOC 62218 Advanced Data Analysis | 4 Hours |
| TRST XXX1 Cognitive Processing, Text Analysis and Data Sciences I | 3 Hours |
| TRST XXX2 Cognitive Processing, Text Analysis and Data Sciences II | 3 Hours |
| **Elective Total** | 6-9 Hours |
| INTEGRATIVE EXPERIENCE (3-HOURS, 6 HOURS FOR THESIS TRACK) | |
| MATH/CS 6XXXX Data Science Research Project | 3 Hours |
| MATH/CS6XXXX Data Science Internship | 3 Hours |
| MATH/CS 67199 Thesis | 6 Hours |
| **Integrative Experience Total** | **3-6 Hours** |
| **PROGRAM TOTAL** | **30 Hours** |

2. **Description of a required culminating, or integrated learning, experience.**

The culminating experience can be fulfilled through a research project, an optional internship., or a thesis. Students must present a report or their thesis to their supervisory committee in an open forum, and the committee must approve the document and the presentation.

3. **Administrative arrangements for the proposed program: department and school or college involved.**

The M.S. in Data Science degree program will be jointly sponsored by the Departments of Mathematical Sciences and Computer Science in the College of Arts and Sciences. The program committee will consist of a coordinator from Mathematical Sciences, a coordinator from Computer Science and a third member from one of the partner disciplines. The coordinators will serve three-year renewable terms, and the partner discipline representative will serve a one-year renewable term. The partner discipline representative will rotate among departments.

The supervisory committee for each culminating experience will consist of at least one faculty member from mathematics, one from computer science, and one from mathematics, computer science, or an appropriate partner discipline.

4. **Evidence of need for the new degree program, including the opportunities for employment of graduates. This section should also address other similar programs in the state addressing this need and potential duplication of programs in the state and region.**



CS/MATH 5

There is very high employment demand for highly qualified data scientists, and this field is consistently rated at the top of job rankings. Starting salaries and job satisfaction are high.

Time's Money section (http://time.com/money/5114734/the-50-best-jobs-in-america-and-how-much-they-pay/) reports that data scientist is the top ranked 2018 job in a list compiled by *Glassdoor* based on equal weighting of base salary, job satisfaction ratings, and number of job openings. The high demand for data scientists is noted in multiple sources, including:

"Data scientist roles have grown over 650 percent since 2012, but currently 35,000 people in the US have data science skills, while hundreds of companies are hiring for those roles - even those you may not expect in sectors like retail and finance - supply of candidates for these roles cannot keep up with demand." [https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report, 8 October 2018]

"By 2020, the number of jobs for all US data professionals will increase by 364,000 openings to 2,720,000 according to IBM." [https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#1b9ec2847e3b, 8 October 2018]

Data science jobs are particularly important in the north-east Ohio economy, with its heavy concentration of insurance, energy, and medical industries. Supporting documents are included in Appendix B. Statewide there are several programs in the related area of Business Analytics:

a) BGSU – MS in Analytics b) CWRU – Weatherhead's Master of Science in Management – Business Analytics, c) KSU – Master of Science in Business Analytics, d) OSU has a Specialized Master in Business – Business Analytics e) OSU has a full-time MBA in Data Analytics, f) Xavier University – Master of Science in Customer Analytics, g) Wittenberg University – Master of Science in Analytics, h) UC – Master of Science in Business Analytics.

These programs are not direct competitors with the proposed M.S. in Data Science program, which is based more heavily in mathematics, computer science and allied sciences than the programs listed above. It emphasizes research in data science and data science applications and will zero-in on filling industry's dire scarcity of technical workforce where advanced STEM data analytics skills are needed.

Bowling Green State University has recently established an M.S. in Data Science degree, which is a collaboration between their Departments of Mathematics and Statistics, Computer Science, and Applied Statistics and Operations Research. While there are clear commonalities with our proposed program, Bowling Green's program is differently and more tightly focused, with coursework confined to the collaborating departments.

**5. Prospective enrollment.**

Students will be recruited from mathematics, computer science, engineering, and science programs at Kent State and other universities in the north-east Ohio region. Initially, we aim to enroll ten students. In subsequent years we aim to increase enrollment steadily to reach a program total of 40 students (20 per year) in year 4.

**6. Special efforts to enroll and retain underrepresented groups in the given discipline.**

Kent State has extensive UG programs/scholarship programs for first-generation college students and under-represented groups. A degree in data science can readily be made accessible, with pre-requisites being completed as required or elective components of many majors, especially STEM or Business



CS/MATH 6

majors. A career in data science can be highly empowering. Efforts to advertise the program to various UG groups, including PELL grant recipients, will be made. Collaborations with the Office of Diversity, and will work in collaboration with the Office of Diversity will be established. A mentorship program will be established including faculty and professionals from underrepresented groups.

7. **Availability and adequacy of the faculty and facilities available for the new degree program.**
Faculty Resources: The faculty needed for fully implementing the curriculum will be in place by fall semester 2019. Over the last several years, Mathematical Sciences and Computer Science have hired (or are hiring) six tenure track faculty with expertise in data science, with intent to develop academic programs in data science and related areas and in support of KSU's strategic priorities. A top university priority is "to drive innovation, idea generation and national distinction through top-tier academic and research programs including the recruitment and support of talented faculty and staff." The necessary tenure track hiring[1] has been strategically planned and has now reached critical mass. Courses have been carefully updated or have been developed in the last 1-3 years by these faculty. Program faculty are research active in data science, and some have active grants in areas related to big data. We anticipate no need for additional faculty to start this degree program

Other Resources: Laboratories, facilities, computing infrastructure and specialized staff: Since, 2012 the Department of Computer Science, with support from the Department of Mathematical Sciences and the College of Arts and Sciences,, has operated a Data and Computing Center with advanced storage, virtual machine, and computing resources needed to support various data and computing intensive programs. There are also three research engineers trained and with experience in science data center and virtual machine support operations. Five required labs including Big Data Intelligence Lab, Visual Analytics Lab, Data & Information Security and Privacy Research Lab, and Cloud Computing Labs are also in place. Further, the CS Networking group is working with Ohio Research Network (OARNet), so that these resources will integrate with the vast US national big-data resources - including national labs facilities, DOE and state supercomputing facilities, NSF repositories, industrial cloud service providers (Amazon, IBM, Microsoft, etc.) with extremely high bandwidth. The required equipment, network infrastructure, and specialized staff resources will be abundantly available to the proposed research intensive Data Science program and its students, researchers, and affiliates to build a premium program.

8. **Need for additional facilities and staff and the plans to meet this need.**
In year three, it will need to be determined if the program should be capped at forty students, or if plans should be made for further growth.

9. **Projected additional costs associated with the program and evidence of institutional commitment and capacity to meet these costs.**
Advertisement, recruitment and program management costs will be borne by the two sponsoring departments from their general budget, with support from A&S.

---

[1] The pool of faculty comprises experts ranging from mathematicians and statisticians engaged in big data research, supercomputing software/systems scientists of big data computing for climate science, and experimental nuclear physics (from Los Alamos National lab), a distinguished hire on multi-lingual unstructured corpus data analytics (from a renowned European Translation Research Lab), to scientists from related industry (Cleveland Clinic and Progressive Corp).



CS/MATH 7