FAIRMONT STATE UNIVERSITY™

Curriculum Committee Approved on March 17, 2020

MEMORANDUM

TO: Faculty Senate

FROM: Susan Ross

DATE: March 26, 2020

SUBJECT: Curriculum Proposal #19-20-16 Data Science (r1)

The intent of the proposal is to create a new minor in Data Science (Appendix A). We have specifically designed this minor to be accessible to any interested Fairmont State student. Even for students with no mathematical or computer science background, this minor will prepare students for the modern data-driven workforce. Data Science is quickly becoming one of the most important fields in today's high-tech economy and will give our students an edge in the marketplace.

cc: Richard Harvey
           Cheri Gonzalez
           Laura Ransom
           Lori Schoonmaker
           Mahmood Hossain
           Bob Niichel

Proposal Number: **#19-20-16**

School/Department/Program: Science and Technology / Computer Science and Math

Preparer/Contact Person: Dr. Mahmood Hossain, Professor of Computer Science
                                 Dr. Bob Niichel, Assistant Professor of Mathematics

Telephone Extension: 4967

Date Originally Submitted:

Revision:

Implementation Date Requested   Fall 2020

I. PROPOSAL

The intent of the proposal is to create a new minor in Data Science (Appendix A). We have specifically designed this minor to be accessible to any interested Fairmont State student. Even for students with no mathematical or computer science background, this minor will prepare students for the modern data-driven workforce. Data Science is quickly becoming one of the most important fields in today's high-tech economy and will give our students an edge in the marketplace.

II. DESCRIPTION OF THE PROPOSAL

A. Deletion of courses/credits from program <u>Not Applicable</u>

B. Addition of courses/credits to program

COMP/MATH 2541 Fundamentals of Data Science (3)
COMP/MATH 2542 Applied Data Science (3)
COMP/MATH 4510 Data Science Capstone (3)

Total hours added: <u>9</u>

C. Provision for interchangeable use of course(s) with program(s)

Aside from the courses listed above, this minor will use courses from the existing Computer Science and Mathematics courses.

D. Revision of course content <u>Not Applicable</u>

E. Other changes to existing courses

To accommodate the new minor, the prerequisite for COMP 4450 (Introduction to Data Mining) will have added prerequisites:

COMP 4450: PR COMP 2270 AND MATH 2562 OR MATH/COMP 2541 AND MATH/COMP 2542

F. Creation of new course(s)

COMP/MATH 2541 *Fundamentals of Data Science* (Appendix B)

COMP/MATH 2542 *Applied Data Science* (Appendix B)

COMP/MATH 4510 *Data Science Capstone* (Appendix B)

G. Attach an itemized summary of the present program(s) affected <u>Not Applicable</u>

III. RATIONALE FOR THE PROPOSAL

A. Quantitative Assessment

Glassdoor recently named Data Scientist as the best job in America for 2019,[1] the third straight year it has claimed that honor. It also notes that there are 6500 job openings nationwide, with a base median salary of $108,000. Also on their list of the top 50 jobs were #8 Data Engineer with 4,700 openings and a base median salary of $100,000, as well as #31 Data Analyst with 5,400 openings and $60,000 base median salary.

The August 2018 LinkedIn Workforce Report[2] reports a nationwide deficit of 151,000 people with data science skills overall. A quick search on LinkedIn revealed the following information about job openings:

Search Term	Job openings "near WV"	Job openings "near U.S."
Data Analyst	71	33,000+
Data Science Specialist	16	5,000+
Data Scientist	33	23,000+
Total	**120**	**61,000+**

Bloomberg also reports that job postings for data scientists rose 75% from January 2015 to January 2018.[3]

The Bureau of Labor Statistics also reports that data science is a growing component of many businesses. In their report *Working with Big Data*,[4] they cite the need for data analysts in the fields of business, e-commerce, government, healthcare, science, social networking, and telecommunications (as well as a few other areas). Data analysis has become a powerful way to gain a competitive edge in the marketplace.

Although this is a minor, the above information conveys the growing importance of data science. We believe that giving students from a variety of different areas of study a broad knowledge of data science will help them interact with their more specialized colleagues at work. In addition math and computer science majors will be much better prepared to take jobs as data scientists or analysts with this minor.

B. Qualitative Assessment

As far as added costs are concerned, the minor will require three new courses, including a capstone. Likely these will be taught as overloads for professors early on, with a new faculty member being added later if the demand is strong. Many popular software packages used in Data Science (e.g. Python and R) are open-source and freely available.

IV. PROPOSAL AFFECTING OTHER COLLEGES/SCHOOLS   Not Applicable

V. PROPOSAL AFFECTING GENERAL STUDIES   Not Applicable

VI. ADDITIONAL COMMENTS

Based on the demand and the hype surrounding this field, we think this minor could be a good way to attract and recruit new students to our other programs (math and computer science) as well as the university at large. We think the minor is an attractive addition to virtually any major.

---

[1] https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm
[2] https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018?trk=lilblog_08-20-18_data-scientists-America-great_tl&cid=70132000001AyziAAC
[3] https://www.bloomberg.com/news/articles/2018-05-18/-sexiest-job-ignites-talent-wars-as-demand-for-data-geeks-soars
[4] https://www.bls.gov/careeroutlook/2013/fall/art01.pdf

Minor in Data Science
Proposed Program

Required Major Courses		HRS
MATH 1550	Applied Statistics	3
COMP 1110	Introduction to Programming	3
MATH/COMP 2541	Fundamentals of Data Science	3
MATH/COMP 2542	Applied Data Science	3
COMP 4450	Introduction to Data Mining	3
MATH/COMP 4510	Data Science Capstone	3
TOTAL HOURS FOR MINOR		**18**

Learning Outcomes and Assessments

Outcomes	*Direct Assessment*
Effectively use data from different sources and efficiently warehouse data	Student Project
Interpret the meaning of data in ways that other non-specialists can understand and use, including creating visualizations	Student Project
Analyze data to draw conclusions and participate effectively in data-driven decision making	Student Project

Note: MATH 1550 and COMP 1110 are currently offered every semester. The proposed courses listed as 2541 and 2542 will be offered once a year. COMP 4450 will be offered every other year, at least until there is substantial interest in the minor. This schedule is flexible enough that any student can begin the minor in their freshman year and finish by the end of their senior year.

APPENDIX B

COMP/MATH 2541 Fundamentals of Data Science 3 hrs. This course covers mathematical structures crucial for data scientists. Topics include set theory, essentials of matrices, graphs and trees, and a treatment of some useful probability concepts. All topics will be paired with appropriate software. PR: MATH 1550 and COMP 1110.

Detailed Course Outline

- Set Theory (include implementation in R) (2 weeks)
    - Basic concepts (intersections, unions, complements)
    - Set builder notations
    - Subset, element of
    - Normalizing
    - Plyr, dPlyr
    - Discretization, substitution
- Matrices (5 weeks)
    - Representation and indexing
    - Addition, matrix multiplication
    - Covariance matrices
    - Design matrices and linear regression
    - ANOVA
    - Eigenvalues and principal component algorithm
- Graphs and trees (5 weeks)
    - Definition, nodes, vertices
    - Types of graphs
    - Traversal
        - Recursion
        - Algorithms
    - Adjacency matrices
    - Trees in R
        - Growing and pruning
        - Classification and Regression trees
        - Forests
- Additional Probability concepts (3 weeks)
    - Review of basic Probability
    - CDFs
    - Other classical random variables
    - Bayes' Theorem
    - Applications

Outcome	Direct Assessment	Satisfactory performance standard
Use statistical software to implement statistical techniques.	Exam/Homework	A class average of 70% or more on associated questions.
Use data science methods involving matrices.	Exam/Homework	A class average of 70% or more on associated questions.
Use data science methods involving graphs and trees.	Exam/Homework	A class average of 70% or more on associated questions.
Use appropriate probabilistic terminology to describe data science problems and techniques.	Exam/Homework	A class average of 70% or more on associated questions.

COMP/MATH 2542 Applied Data Science 3 hrs. This course provides an overview of data science, introduces the different steps involved in data science, and provides the students hands-on experience with tools and methods. Topics include data acquisition and cleaning, data processing, data organization and management, data analysis, data visualization, data warehousing, and data security/privacy. PR: MATH 1550 and COMP 1110.

<u>Detailed Course Outline</u>

- Introduction
    - o Steps involved in data science
    - o Social implications
- Data Collection
    - o Types of Data
    - o Data cleaning
    - o Data reduction
    - o Data transformation
- Data Organization
    - o Arrays
    - o Linked lists
- Data Management
    - o Database Concepts
    - o Queries
    - o Data Manipulation
- Data Analysis
    - o Basic concepts
    - o Overview of Methods
- Data Visualization
- Data Warehousing
- Data Security/Privacy

Outcome	Direct Assessment	Satisfactory performance standard
Describe the components of data analysis.	Exam/Homework	A class average of 70% or more on associated questions.
Collect and preprocess data from different sources.	Exam/Homework	A class average of 70% or more on associated questions.
Utilize appropriate data structures to represent and manipulate data.	Exam/Homework	A class average of 70% or more on associated questions.
Store, access, and manipulate data using a database.	Exam/Homework	A class average of 70% or more on associated questions.
Create effective visualizations of given data.	Exam/Homework	A class average of 70% or more on associated questions.

COMP/MATH 4510 Data Science Capstone 3 hrs. The focus of this course is on preparing students for the workforce through a large-scale project generated either internally or through contacts with local or regional industry. In the process, students will learn about advanced Data Science topics like big data analysis, distributed data analytics, data warehousing, predictive modeling, and simulations. PR: COMP 4450.

Note: This will be an individual directed study, and detailed outline, outcomes, and assessments are not provided.

Subject: RE: Creating new courses for the catalog
Date: Thursday, December 12, 2019 at 2:46:47 PM Eastern Standard Time
From: Gonzalez, Cheri
To: Hossain, Mahmood
Attachments: image002.jpg, image003.png

Yes you may use these course numbers.

Cheri

Cheri L. Gonzalez
University Registrar
Fairmont State University
304-367-4112
Cheri.Gonzalez@fairmontstate.edu

From: Hossain, Mahmood
Sent: Thursday, December 05, 2019 10:04 AM
To: Gonzalez, Cheri <Cheri.Gonzalez@fairmontstate.edu>
Subject: Creating new courses for the catalog
Importance: High

Dear Cheri,

We are submitting a curriculum proposal where we are creating three new cross-listed courses. These are:

COMP/MATH 2541
COMP/MATH 2542
COMP/MATH 4510

Could you please let us know if we can use these course numbers?

Regards,

Mahmood

Mahmood Hossain, Ph.D.
Chair, Department of Computer Science and Math
Professor of Computer Science
Fairmont State University
201G Engineering Tech Building
1201 Locust Avenue

Fairmont, WV 26554
Voice: (304) 367-4967