

UMBC UGC New Course Request: CMSC462 – Introduction to Data Science

Date Submitted: 11/1/2019

Proposed Effective Date: 1/1/2020

	Name	Email	Phone	Dept
Dept Chair or UPD	Jeremy Dixon	jdixon@umbc.edu	5-8866	CSEE
Other Contact	Mohamed Younis	younis@umbc.edu	5-3969	CSEE

COURSE INFORMATION:

Course Number(s)	CMSC462
Formal Title	Introduction to Data Science
Transcript Title (≤30c)	Intro to Data Science
Recommended Course Preparation	
Prerequisite NOTE: Unless otherwise indicated, a prerequisite is assumed to be passed with a "D" or better.	CMSC 341 and (STAT 355, STAT 451, or CMPE 320) each with a grade of C or better
# of Credits Must adhere to the <u>UMBC Credit Hour Policy</u>	3
Repeatable for additional credit?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Max. Total Credits	3 <small>This should be equal to the number of credits for courses that cannot be repeated for credit. For courses that may be repeated for credit, enter the maximum total number of credits a student can receive from this course. E.g., enter 6 credits for a 3 credit course that may be taken a second time for credit, but not for a third time. Please note that this does NOT refer to how many times a class may be retaken for a higher grade.</small>
Grading Method(s)	<input type="checkbox"/> Reg (A-F) <input type="checkbox"/> Audit <input checked="" type="checkbox"/> Pass-Fail

PROPOSED CATALOG DESCRIPTION (Approximately 75 words in length. Please use full sentences.):

Data science is a field that involves data manipulation, analysis, and presentation, all at scale. It's typical for an organization to have a few terabytes of data maintained for different purposes by different business units stored in different formats, and for someone to have an idea about how the data might bring significant additional value. Data scientists are the bridge between the idea and the data and help extract latent value, often uncovering novel insights and novel beneficial ways to use the data in the process.

RATIONALE FOR NEW COURSE:

- Why is there a need for this course at this time?
This course is mandatory in our data science track and has been taught many times as CMSC 491 – Special Topics in Computer Science
- How often is the course likely to be taught?
We will most likely offer it every semester (Fall and Spring).
- How does this course fit into your department's curriculum?
It can be used as a technical elective for all computer science majors. Additionally, it is required for all data science track – computer science majors and it can be used for the new AI/ML track.
- What primary student population will the course serve?
This course will primarily serve Juniors, and Seniors in the CMSC department.
- Why is the course offered at the level (ie. 100, 200, 300, or 400 level) chosen?
The course has significant analysis and design elements for an “Introductory” course. It builds on concepts introduced in CMSC 341- Data Structures, and STAT 355 - Introduction to Probability and Statistics for

Scientists and Engineers with additional considerations for databases, data manipulation, and cloud computing.

- f) Explain the appropriateness of the recommended course preparation(s) and prerequisite(s).
A student is more likely to be successful in this course if they are adequately prepared with programming (CMSC 341 – Data Structures) and statistics (STAT 355 - Introduction to Probability and Statistics for Scientists and Engineers).
- g) Explain the reasoning behind the P/F or regular grading method
Students are most likely to take this course using A-F but on occasion a student could audit it or taking it P-F.
- h) Provide a justification for the repeatability of the course.
This course cannot be repeated for additional credit.

ATTACH COURSE SYLLABUS (mandatory):

CMSC 462: Introduction to Data Science

Prerequisites:

CMSC 341 and (STAT 355, STAT 451, or CMPE 320) each with a B or better.

Instructor:

TBD

Course Description:

Data science is a field that involves data manipulation, analysis, and presentation, all at scale. It's typical for an organization to have a few terabytes of data maintained for different purposes by different business units stored in different formats, and for someone to have an idea about how the data might bring significant additional value. Data scientists are the bridge between the idea and the data and help extract latent value, often uncovering novel insights and novel beneficial ways to use the data in the process.

The goal of this class is to give students hands on experience with all phases of the data science process using real data and modern tools. Topics that will be covered include data formats, loading, and cleaning; data storage in relational and non-relational stores; data analysis using supervised and unsupervised learning, and sound evaluation methods; data visualization; and scaling up with cloud computing, MapReduce, Hadoop, and Spark.

Credits:

Three credits: not repeatable

Learning Outcomes:

At the end of the course, the student will:

- Organize and transmit data using tools such as text files, JSON, or other language specific libraries.
- Demonstrate data management processes including data loading, cleaning, summarization, and outlier detection.
- Analyze data storage options including SQL and NoSQL and where data can be stored and processed including cloud computing options.
- Use statistical methods and modeling to summarize data and identify relationships.
- Explain how to formulate new hypotheses and draw accurate conclusions from data.
- Develop effective visualizations of given data.
- Evaluate ethical and privacy considerations of data sets and apply ethical practices.

Readings:

Baumer, B., Kaplan, D., & Horton, N. J. (2017). *Modern Data Science with R*. Boca Raton: CRC Press, Taylor & Francis Group, CRC Press.

Dietel, P. and Dietel, H. (2020). *Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud*. Pearson. ISBN: 9780135404669

Dietrich, D., Heller, B., Yang, B., & EMC Education Services (Eds.). (2015). *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Indianapolis, IN: Wiley.

Larose, C. D., & Larose, D. T. (2019). *Data Science Using Python and R*. Hoboken, NJ: John Wiley & Sons, Inc.

Course Topics:

Students in Introduction to Data Science will participate by:

- Introduction: Introduction to data science and setting up your environment.
- Background and data acquisition: Types of data and data representations. How to acquire, process, and parse data. Data manipulation, data wrangling, and data cleaning.
- Data engineering and processing: Use SQL, NoSQL, and other storage methods.
- Visualization and basic statistics: Visualization principles and goals. How to communicate results. Visualizing distributions and relationships.
- Statistical analysis: Evaluation, cross-validation, overfitting, clustering, dimensionality reduction, and other practical concerns. Learning predictive models from data.
- Ethical considerations: Examine how data can generate privacy and ethical considerations.

Academic Integrity:

Academic integrity is an important value at UMBC. By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC's scholarly community in which everyone's academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping others to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, but is not limited to, suspension or dismissal.

More information can be found at:
<https://academicconduct.umbc.edu/>

Student Disability Services:

UMBC is committed to eliminating discriminatory obstacles that may disadvantage students based on disability. Services for students with disabilities are provided for all students qualified under the Americans with Disabilities Act (ADA) of 1990, the ADA of 2009, and Section 504 of the Rehabilitation Act who request and are eligible for accommodations. The Office of Student Disability Services (SDS) is the UMBC department designated to coordinate accommodations that would allow students to have equal access and inclusion in all courses, programs, and activities at the University.

If you have a documented disability and need to request academic accommodations for access to your courses, please refer to the SDS website at sds.umbc.edu for registration information and to begin the process, or alternatively you may visit the SDS office in the Math/Psychology Building, Room 212. For questions or concerns, you may contact us through email at disAbility@umbc.edu or phone (410) 455-2459.

If you require accommodations for this class, make an appointment to meet with your instructor to discuss your SDS-approved accommodations.



Department of Computer Science and Electrical
Engineering
University of Maryland, Baltimore County
ITE 325J
1000 Hilltop Circle, Baltimore, MD 21250

jdixon@umbc.edu // p: 410.455.8866
csee.umbc.edu

October 28, 2019

Dear UGC,

Below is a summary of the 7 changes that the Department of Computer Science and Electrical Engineering (CSEE) is proposing regarding the Computer Science program (CMSC).

Our goal is to have all of these go into effect in January 2020.

Thank you,

A handwritten signature in black ink, appearing to read "J. Dixon".

Jeremy Dixon
Undergraduate Program Director – Computer Science
CSEE
UMBC

Num	Name of Change	Change Type	Course Number	Description of Change	Related Files
1	New Study Track - AI/ML	Program Change	XXX	New Study Track	Program_Change_AI-ML.docx Program_Change_CMSC_AI_Track_Details.docx
2	Change to CMSC Tracks	Program Change	XXX	Change requirements for CMSC students who complete multiple tracks.	Program_Change_CMSC_Tracks.docx
3	Change to CMSC Natural Science Requirement	Program Change	XXX	Change natural science requirements for CMSC majors. Going from 12 credits (sequence of BIO1/BIO2 or CHEM1/CHEM2 or PHYS1/PHYS2 + 4 additional credits) to 10-12 credits (sequence of BIO1/BIO2 or CHEM1/CHEM2 or PHYS1/PHYS2 + one lab science to include SCI100 or CHEM102L or PHYS122L or GES286) Additionally, students who transfer from a school with a lab science will have met this lab requirement.	Program_Change_CMSC_Natural_Science_Change.docx
4	Malware Analysis	New Course	CMSC 491 to CMSC 449	New Course - converting CMSC 491 to CMSC 449	NewCourse_CMSC449 - Malware_Analysis.docx NewCourse_CMSC449 - Malware_Analysis_Syllabus.pdf
5	Introduction to Data Science	New Course	CMSC 491 to CMSC 462	New Course - converting CMSC 491 to CMSC 462	NewCourse_CMSC462 - Intro_to_Data_Science.doc NewCourse_CMSC462 - Intro_to_Data_Science_Syllabus.doc
6	Undergraduate Teaching Assistantship	New Course	CMSC 496	New Course	NewCourse_CMSC396 - Undergraduate_Teaching_Assistantship.docx NewCourse_CMSC396 - Undergraduate_Teaching_Assistantship_Syllabus.docx
7	Design and Analysis of Algorithms	Course Change	CMSC 441	Modify prerequisites and at least one of the following: (STAT 355, CMPE 320, or STAT 451) all with a grade of C or better.	Course_Change_CMSC441_Prerequisites.docx