Last Updated 08/23/2021

Master of Science in Data Analytics/Binghamton University

DATA 500: Introduction to Data Analytics

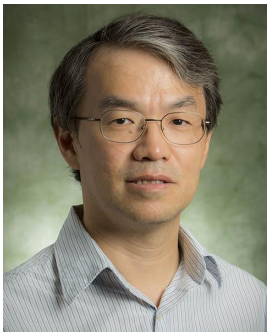
Fall 2021 Syllabus

Instructor Information

Xingye Qiao
qiao@math.binghamton.edu

Office: WH-134
Research Interests: Statistics, Machine Learning, High-dimensional Inference, Big Data Analytics, Precision Medicine

Office hours: Monday 10:30 to 11:30 on Zoom or in-person. Email in advance.



Kenneth Chiu
kchiu@binghamton.edu

Office: N/A
Research Interests: High performance computing, applied machine learning

Office hours: After class, or by appointment.

Welcome

Welcome to Data 500: Introduction to Data Analytics. We hope that you will find this course both intellectually stimulating and relevant to your studies and your future directions as Binghamton MSDA alumni.

Please note that while we will use various technologies, Brightspace will be the gateway of the course. From there you can find links to various sites we use for assignments, notes sharing, recordings, discussion forums, etc.

Preferably all communication will be done on Piazza. General questions about the course contents can be asked so that both the questions and the answers can be visible to all. Private messages can also be sent on Piazza. Questions may be posted anonymously to the other

students. We will respond within 24 hours. We encourage you to use Piazza for general discussions related to data analytics, and share interesting items or topical news items that you may come across. Please follow the Netiquette rules shown below in the syllabus when you communicate on Piazza.

This syllabus provides you with the key information about this course - for the best chance of success, you should read it very carefully and refer to it frequently throughout the semester. It also provides information about your rights and responsibilities - we will assume you have read this before the course commences, and we expect you to refer to it throughout the semester.

Course Information

Description

This course aims to develop foundation skills and knowledge required for data-driven, evidence-based approaches to decision making, especially in the business setting. Since Data Analytics draws techniques and concepts from mathematics, statistics, computer science, and computational sciences, and involves analytic methods, programming tools, data retrieval and management, data visualization, effective communication, and applications in various domains, the course will cover the required background and preview some concepts in these areas, along with case studies that showcase how to make data-driven decisions in the big data era. Ethical issues that arise in data science activities, such as those related to bias, security, privacy, and governance, will also be explored.

Student Feedback

You are strongly encouraged to provide constructive feedback for this course when the Students' Opinions On Teaching (SOOT) survey opens in myCourses. You can also provide feedback directly at SOOT.binghamton.edu.

If you have any concerns about the course during the semester, please contact the course teaching team - preferably early in the semester - so we can discuss your concerns, and make adjustments, if appropriate. You can also discuss your concerns with the Program Director: Manoj K Agarwal (agarwal@binghamton.edu).

Course Delivery Mode

This course will be a traditional in-person course. However, part of the content will be delivered as pre-recorded videos, followed by real-time, in-person discussion. You are expected to watch the videos before the class, and it will be difficult to succeed in the course unless you do so..

Student Learning Outcomes (SLO)

Upon completion of this course, the student will be able to

- **(SLO1)** Explain and apply basic mathematical and statistical concepts that the students will encounter in the MSDA program;
- **(SLO2)** Describe the life cycle of data analytics and explain how the components of data analytics interact and how organizations leverage data analytics for strategic advantage;
- **(SLO3)** Recognize main data analysis tools and techniques from statistics and machine learning and identify the most suitable ones in different settings;
- **(SLO4)** Master and apply core programming skills (Python and/or R) to conduct data collection, data retrieval, data cleaning, data wrangling;
- **(SLO5)** Recognize, quantify, and communicate the uncertainty in the analytic process;
- **(SLO6)** Work effectively and synergistically in teams on data analytics projects;
- **(SLO7)** Communicate the results of data analysis to the management and other decision makers by writing detailed reports and giving professional presentations;
- **(SLO8)** Criticize unethical use of statistical analysis and misinterpretations of data, and examine ethical issues that arise in the analytics process; describe the professional ethic code for data scientists.

Prerequisite(s)

As a minimum, the student should have taken introductory statistics, have mastered one programming language, and have some working knowledge on analyzing and drawing conclusions from data. Recommended mathematics background includes Calculus I and II and Linear Algebra. Recommended statistical background includes probability distributions, independence, mean and variance, conditioning, confidence intervals, estimation, and hypothesis tests.

Relationship to Other Courses

This course is the first course in the Master of Science in Data Analytics program. It gives students an overview of data analytics and brings all the students up to speed with the basic tools and techniques necessary in the program.

Topic Outline/Schedule

Note that on Tuesday, Oct. 12, Thursday classes will meet. Tuesday classes will NOT meet on Oct. 12. This may be a bit confusing, but this schedule has been designed to provide 14 class meetings for each day of the week over the course of the semester, thereby assuring that our classes have sufficient contact hours to meet credit-hour standards. Note that we have different classrooms for Tuesday and Thursday. Hence, ***on Oct. 12 we should meet in CW-212 instead of the usual room.***

Last Updated 08/23/2021

Date	Day	Day	AM	AM Classroom	PM	PM Classroom
8/24	Tue	1	Introduction	CW 202	Python	CW 206
8/25	Wed	2	Matrix Algebra	CW 214	Python	CW 107
8/26	Thur	3	Calculus	CW 212	Python	CW 206
8/30	Mon	4	Probability	CW 214	NumPy	CW 107
8/31	Tue	5	Multivariate RV	CW 202	NumPy	CW 206
9/1	Wed	6	Num. Linear Algebra	CW 214	NumPy	CW 107
9/2	Thur	7	Classical Stat.	CW 212	Pandas	CW 206
9/9	Thur	8	Statistical Thinking	CW 212		
9/14	Tue	9	Pandas	CW 202		
9/21	Tue	10	Sampling & Bootstrap	CW 202		
9/23	Thur	11	Pandas	CW 212		
9/28	Tue	12	Cause and Effect	CW 202		
9/30	Thur	13	Pandas	CW 212		
10/5	Tue	14	Regression	CW 202		
10/7	Thur	15	RegEx	CW 212		
10/12	Tue	16	Machine Learning	CW 212 (sp loc)		
10/19	Tue	17	Web-Scrapping	CW 202		
10/21	Thur	18	Classification	CW 212		
10/26	Tue	19	Ethics	CW 202		
10/29	Thur	20	Outroduction	CW 212		

Course Requirements and Resources

Textbooks

There is no single text in data science and data analytics. We will use a variety of materials, drawn from the following:

- Mathematics and Statistics
    - (MML) Mathematics for Machine Learning, by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Published by Cambridge University Press. https://mml-book.github.io/book/mml-book.pdf
    - (PSDS) *Practical Statistics for Data Scientists: 50 Essential Concepts*, by Peter Bruce and Andrew Bruce (ISBN-13: 978-1491952962) As the name suggests, this book covers the very essential statistical concepts.
    - (IP) Introduction to probability by Blitzstein, J. K., & Hwang, J. (2019). Chapman and Hall/CRC. Companion Harvard course with YouTube videos; Free book.
- Data Science Overview
    - (MDSR) Modern Data Science with R, by Ben Baumer, Daniel T. Kaplan, and Nicholas J. Horton. https://mdsr-book.github.io/
    - (PTDS) *Principles and Techniques of Data Science*. https://www.textbook.ds100.org/
- Python

    - (PDA) *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython,* 2nd Edition, by McKinney (ISBN-13: 978-1491957660) This book is a must-have for Python data analytic types. It covers all the necessary extensions to Python needed for data.
    - (IMLP) *Introduction to Machine Learning with Python: A Guide for Data Scientists* 1st edition, by Müller and Guido. ISBN-13: 978-1449369415. This book is for beginners and covers basic topics in detail.
- R
    - (r4ds) *R for Data Science* by Garrett Grolemund and Hadley Wickham. Free online book at: https://r4ds.had.co.nz/
- Ethics
    - (WMD) *Weapons of math destruction: How big data increases inequality and threatens democracy*. By O'Neil, C. Crown.
    - (DF) *Data feminism*. By D'ignazio, C. and Klein, L.F. MIT press. https://data-feminism.mitpress.mit.edu/

Note: MML, PTDS, IP, r4ds, MDSR, and DF are all free online books. Binghamton University students have free access to PSDS and MSDR online via the libraries. Check the Binghamton Libraries website. DF, IMLP, PDA and WMD have been put on the course reserve and can be picked up at the Bartle Library frontdesk. You may check out a book for one day at a time.

Other Readings

- (IDSSS) *An Introduction to Data Science,* by Jeffrey Saltz and Jeffrey Stanton (ISBN-13: 978-1506377537) It's indeed an introduction - a book for beginners (such as people with little background in statistics or programming) who want to learn about data science.
- (DSB) *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking,* by Foster Provost and Tom Fawcett (ISBN: 978-1449361327) Purely business-oriented. This book will convince you why you want to choose data science and analytics as your career.
- (PDSH) *Python Data Science Handbook: Essential Tools for Working with Data,* by Jake VanderPlas (ISBN-13: 978-1491912058) It is a handbook, meaning that it does not get into the fundamentals much. The machine learning case studies are pretty nice, providing a good overview of the scikit-learn API and effective patterns for machine learning problems.

Other Requirements

- Internet service: Brightspace, Piazza (Privacy Policy), GradeScope (Privacy Policy), Google G Suite (including Google Colab).
- Hardware: Laptop or desktop computer
- Software:
    - Python 3 and the entire Anaconda suite of tools. (This is open source and runs on all major operating systems.) We will use Google Colab for running our

example code, but we would like to keep Anaconda as a backup option in case it has some hiccup (for example, when there is no internet connection.)

Course Learning Resources

Your course learning resources are available on the Brightspace course site or other web site. You will find the following resources:

- Recordings of some lectures
- Lecture Slides
- Readings
- Exercises and solutions
- Discussions
- Additional relevant materials for some of the topics

COVID-19 related policy

Binghamton University follows the recommendations of public health experts to protect the health of students, faculty, staff and the community at large. Safeguarding public health depends on each of us strictly following requirements as they are instituted and for as long as they remain in force. Health and safety standards will be enforced in this course.

Face coverings and other safety measures

Current rules require **everyone to wear a face covering that completely covers both the nose and mouth while indoors** (unless they are eating or alone in a private space like an office). **A face shield is not an acceptable substitute.** Classroom safety requirements will continue to be based on guidance from public health authorities and will be uniformly applied across campus. If these requirements change, a campus-wide announcement will be made to inform the University.

- **If you forget your face covering or it does not meet these requirements, you will be asked to leave the room immediately.** You may not return until you meet the requirement.
- If you miss a graded assessment due to being asked to leave the classroom for not having a proper face covering, **you will receive a zero on the assessment**.
- If you do not meet the face covering requirement and refuse to leave the room when directed, the instructor will immediately cancel the remainder of the class session and inform the dean's office, which will work with the Student Records office to **issue a failing grade ("F") for the course** regardless of when in the semester the incident occurs. The dean's office will also inform the Office of Student Conduct.
- If a student's refusal to comply is a second offense, the Office of Student Conduct may recommend **dismissal from the University**.

If the rules for health and safety measures change, the campus will be notified and the new requirements will take effect.

The University recommends and supports swift action and clear consequences if a student's non-compliance risks the safety of others. The academic and course-removal sanctions listed here are provided because the Provost's Office considers them to be valid responses if a student puts the safety of others at risk. Non-compliance with safety requirements constitutes a public health risk and a disruption of the learning experience.

Illness/quarantine

Students should stay home if they are ill and seek prompt medical evaluation if they experience symptoms of COVID-19. Early case finding will benefit the entire campus. If you have to be absent due to illness or quarantine, please notify the instructor team, who will work with you to identify solutions for you to gain access to the learning materials.

Credit Hours and Expectations

This course is a 3-credit course, which means that students are expected to do at least 126 hours of course-related work in total. This includes about 36 hours for in-class lectures and 90 hours completing assigned readings, studying for tests and examinations, preparing written assignments, and other course-related tasks.

To be successful in this course, you must:

- Read all materials in preparation for your classes, and follow up each with further study and research on the topic;
- Start your assessment tasks well ahead of the due date;
- Read or listen to all feedback carefully, and use it in your future work.

Assessment

The SLOs that each assessment is linked to are shown in the parenthesis.

- Labs. **(SLO1 to SLO7)**
- Homework Assignments. **(SLO1 to SLO7)**
- Coding Assignments. **(SLO1 to SLO7)**
- Readings and discussion on data ethics. **(SLO8)**
- Quizzes/questions: We will give a few in-class quizzes, in-video quizzes, or simply ask people to answer questions during the lecture to check the learning. You will receive the full 5% quiz score in the final grade if you can pass or answer at least 70% of all the quizzes/in-class questions given. Otherwise, the score you receive will be proportional to the quizzes/questions you do pass/answer. You will have unlimited attempts to in-video quizzes. For in-class questions or quizzes, we mark you as passed if you make a genuine effort to answer the questions. Our hope is that you can actively participate in the learning process without worrying too much about the grade.

Grading

Assignment Name	Percent of Total
Lab Assignments	30%
Homework Assignments	30%
Coding Assignments	30%
Data Ethics	5%
Quizzes & Attendance	5%

Grading Scheme

Grade	Percent
A	94 – 100%
A-	90 - <94%
B+	87 - <90%
B	83 - <87%
B-	80 - <83%
C+	77 - <80%
C	73 - <77%
C-	70 - <73%
D	60 - <70%
F	<60%

Accessing Grades

Students can check their grades using the Brightspace course site.

Extra Credit (5 points maximum)

This course was offered for the first time in Fall 2020. Your feedback will make this course better for you and future students. Give 5+ pieces of feedback/corrections to course material.

- typos/spelling/grammar (there are likely many)

- rewording awkward/unclear language
- improve visualizations
- point out explanations you found effective/could be better

Send the feedback (whenever it becomes available) to the course instructors publicly on Piazza (using the *feedback* tag; you can hide your identity, but the content should be visible to all). We will mark the feedback down and make adjustments to your grade at the end of semester if you contribute 5 pieces or more. It is important to use the *feedback* tag so that it is easy for us to do the book-keeping.

Please keep an eye on the feedback as sometimes it helps to correct an error or clarify a muddy point. If a similar feedback has been made, then a duplicate feedback will not be counted for the extra credit purpose.

Study Groups

In an organization, you will often work in a data science team instead of work individually. Moreover, our students often come with very different academic backgrounds and working experiences. We therefore encourage you to study in a group setting.

The instructors will divide the students into six study groups for the first and second homework assignments. Within each group, a coordinator will be responsible for scheduling and hosting group study meetings. The instructors will check with the coordinators to see how the groups are going. We require each group to meet at least once for each homework assignment. Additional meetings are optional and up to the group to decide.

Although study groups are encouraged, each student should write or code his/her own answer to the homework problems. In particular, nearly identical answers are viewed as unauthorized collaboration which is a serious violation of academic dishonesty.

The instructors will not assign groups for the rest of the homework. Students will form the groups or teams on their own.

Course Policies

Penalties for Late Work and Requests for Extensions

Late submissions are allowed up to one week. For each day the submission is late, the grade will be reduced by 10%. For example, a homework that is 3 days late will only receive 70% of the real grade should you submit the work on time.

Make-up Exam Policy (not applicable in this class)

Students who cannot take their exams as scheduled because of a religious conflict, documented serious illness, or compelling, unexpected circumstances may appeal for a make-up exam. The following are examples that are not considered compelling reasons to grant

a make-up exam: lack of preparation, negligence, misinformation, or planned vacations and other events.

Absences Due to Religious Holidays

If you anticipate being absent because of any religious observance, please notify me in writing at least one week in advance. We will work together to accommodate.

Attendance & Participation

Class attendance and participation will be enforced by the in-video quizzes, the in-class quizzes and questions the instructor asks during class. They will form 5% of the final grade.

Although it is not required, most students send their professor a brief email to explain their absence in advance.

Class participation is a very important part of the learning process in this course. Quality contributions and insights possess one or more of the following properties:

- Offers a different and unique, but relevant, perspective;
- Contributes to moving the discussion and analysis forward;
- Builds on other comments;
- Transcends the “I feel” syndrome. That is, it includes some evidence, argumentation, or recognition of inherent tradeoffs. In other words, the comment demonstrates some reflective thinking.

We will use our assessment of your participation to manage borderline grades. While your participation grade is subjective, it will not be random or arbitrary. And, clearly, more frequent quality comments are better than less frequent quality comments.

Activities on the online forum are also part of the class participation.

Use of Mobile Devices

As research on learning shows, unexpected noises and movement automatically divert and capture people's attention, which means you are affecting everyone’s learning experience if your cell phone, tablet, laptop, etc. makes noise or is visually distracting during class.

For this reason, we

- ask you to turn off your mobile devices;
- allow you to take notes on your laptop, but you must turn the sound off so that you do not disrupt other students' learning.

If you are doing anything other than taking notes on your laptop, please sit in the back row so that other students are not distracted by your screen.

Academic Integrity

Binghamton University provides explicit guidelines in the Student Academic Honesty Code (see the University Bulletin - Academic Policies and Procedures for All Students). Unless specified otherwise in the syllabus, I expect the work you submit for grading to be yours and yours alone. Not acknowledging another's work with proper references, taking credit for someone else's work, letting your work appear in another student's paper, or fabricating "results" are grounds for failing the assignment and/or the course. If you have any questions about what constitutes plagiarism or cheating, please ask me.

Netiquette

Unfortunately, there are instances where disrespectful and inappropriate behavior can occur in a course, even in an online setting. It is important to remember to be respectful of your classmates and their ideas. Here are the Netiquette Rules students are expected to follow in this course. Based on “15 Rules of Netiquette for Online Discussion Boards” by Touro’s College Online Education Department.

1. Before posting your question to a discussion board, check if anyone has asked it already and received a reply. Just as you wouldn’t repeat a topic of discussion right after it happened in real life, don’t do that in discussion boards either.
2. Stay on topic – Don’t post irrelevant links, comments, thoughts, or pictures.
3. Don’t type in ALL CAPS! If you do, it will look like you’re screaming.
4. Don’t write anything that sounds angry or sarcastic, even as a joke, because without hearing your tone of voice, your peers might not realize you’re joking.
5. Always remember to say “Please” and “Thank you” when soliciting help from your classmates.
6. Respect the opinions of your classmates. If you feel the need to disagree, do so respectfully and acknowledge the valid points in your classmate’s argument. Acknowledge that others are entitled to have their own perspective on the issue.
7. If you reply to a question from a classmate, make sure your answer is accurate! If you’re not 100% sure when the paper is due, DO NOT GUESS! Otherwise, you could really mess things up for your classmates, and they will not appreciate it.
8. If you ask a question and many people respond, summarize all answers and post that summary to benefit your whole class.
9. Be brief. If you write a long dissertation in response to a simple question, it’s unlikely that anyone will spend the time to read through it all.
10. Don’t badmouth others or call them stupid. You may disagree with their ideas, but don’t mock the person.
11. If you refer to something your classmate said earlier in the discussion, quote just a few key lines from their post so that others won’t have to go back and figure out which post you’re referring to.
12. Before asking a question, check the class FAQs or search the internet to see if the answer is obvious or easy to find.

13. Check the most recent comments before you reply to an older comment since the issue might have already been resolved or opinions may have changed.
14. Be forgiving. If your classmate makes a mistake, don't badger him or her for it. Just let it go – it happens to the best of us.
15. Run a spelling and grammar check before posting anything to the discussion board. It only takes a minute and can make the difference between sounding like a fool and sounding knowledgeable.

Campus Help for Students

Disability-Related Equal Access Accommodations Statement

Students wishing to request academic accommodations to insure their equitable access and participation in this course should notify the instructor as soon as they are aware of their need for such arrangements. Authorizations from Services for Students with Disabilities (SSD) are generally required. We encourage you to contact SSD at (607) 777-2686 to schedule an appointment with the Director or Learning Disabilities Specialist. The SSD website (www.binghamton.edu/ssd/) includes information regarding their Disability Documentation Guidelines. The office is located in UU – 119.

University Tutoring Services

UTS offers free tutoring for undergraduate students at Binghamton University. All UTS tutoring appointments must be scheduled online through the my.binghamton.edu portal. Walk-in tutoring is also available for select courses. If you have any questions about UTS, call 607-777-9235, email uts@binghamton.edu, or visit the website: http://www.binghamton.edu/tutoring.

ITS Helpdesk/myCourses Support

Walk-in: Located in the Computer Center first floor lobby.
Call: 607-777-6420; E-mail: helpdesk@binghamton.edu.
https://www.binghamton.edu/its/

Libraries

The Libraries offer a wide variety and range of services including research assistance, instruction, user-friendly interfaces, digital preservation, digital scanners, and resource sharing.
Text: 607-205-8173; Call: 607-777-2345; Email: refquest@binghamton.edu
http://www.binghamton.edu/libraries

Dean of Students

If you are experiencing undue personal or academic stress at any time during the semester or need to talk with someone about a personal problem or situation, I encourage you to seek support as soon as possible. I am available to talk with you about stresses related to your work

in my class. Additionally, I can assist you in reaching out to any one of a wide range of campus resources, including:

1. Dean of Students Office: 607-777-2804
2. Decker Student Health Services Center: 607-777-2221
3. University Police: On campus emergency, 911
4. University Counseling Center: 607-777-2772
5. Interpersonal Violence Prevention: 607-777-3062
6. Harpur Advising: 607-777-6305
7. Office of International Student & Scholar Services: 607-777-2510

University Counseling Center

At some point during their college experience, students may encounter personal, social, or developmental issues that call for assistance beyond the advice provided by friends and family. That's where the University Counseling Center (UCC) can help. The UCC provides a variety of free and confidential counseling services delivered by professional counselors. All currently enrolled Binghamton University undergraduate students, graduate students and affiliated entities are eligible to receive these services free of charge. Services and programs available through the center include individual and group counseling, consultation, referral, and psychoeducational programs. For more information or to make an appointment, visit https://www.binghamton.edu/counseling.