SI 699 004 WN 2019

SI 699-004 – Big Data Analytics
Thursday 8:30-11:30am, North Quad 1265
Instructor: Qiaozhu Mei (qmei@umich.edu); Office Hours: By Appointment

The big data analytics mastery course will require students to demonstrate mastery of data collection, processing, analysis, retrieval, mining, visualization, and prediction. Students will synthesize methods from information retrieval, statistical data analysis, data mining, machine learning, and other big-data related fields. Students will work on semester-long projects that deal with industry-scale data sets and solve real-world problems. Aligned with best industry practices, students will be expected to work in a fast-paced, collaborative environment and to demonstrate independence and leadership. Students must be able to create and use tools to handle very large transactional, text, network, behavioral, and/or multimedia data sets.
By taking this course, the students are expected to:
1.Transform a real world scenario into a data mining problem by identifying the input, the output, and different types of data needed to generate the output.
2.Formulate the problem (input -> output) as one of the typical data analysis tasks, including but not limited to pattern extraction, visualization, classification, clustering, ranking, prediction, and anomaly detection.
3.Design experiments to judge/measure the success of the data analysis task.
4.Identify the state-of-the-art algorithms and tools for particular type of data and the data analysis task.
5.Collect data needed for the analysis task from various sources.
6.Write programs to manipulate raw data into correct formats needed for the analysis.
7.Use statistical and visualization tools to describe the properties of the collected data.
8.Deal with data at scale (by writing programs and using correct tools).
9.Setup, configure, customize, and execute the state-of-the-art big-data analytics tools and conduct the experiments. Implement algorithms if nothing exists off-the-shelf.
10.Validate, summarize, and present the analysis results.
11.Draw correct conclusions from the analysis results.
12.Disseminate results of big-data analytics to the broader community in the format of social media posts, websites, etc.

Prior to enrolling in this course, students should manage the basic concepts, theory, and techniques about data structures, data mining algorithms, probability distributions, statistical tests, data visualization, statistical data analysis, and data-intensive computation. Depending on the particular projects, students should also be familiar with concepts, theory, and techniques about particular data types and application domains, such as natural language processing, information retrieval, network science, social media, financial markets, and/or information visualization. All these knowledge and foundations can be obtained through combinations of the required and recommended courses.

Prior to taking this course, students are expected to have competency in programming, data manipulation, statistical data analysis, data mining algorithms, working with unix environments, and configuring and using state-of-the-art data mining and statistical analysis tools.
Students should also be familiar with common practices of managing individual and team projects, such as version control (e.g., git), project documentation (e.g., wiki), and progress tracking (e.g., Trello).
Students must complete the following MSI courses:
SI501, SI507, SI618 (data manipulation and data exploration, or equivalent), SI544 (statistics and data analysis, or equivalent), SI671 (data mining)
and at least two of the following (or equivalent):
SI561 (natural language processing), SI608 (networks), SI649 (information visualization), SI650 (information retrieval)
Students will do semester long projects that are of real interest of the industry, using real-world, large-scale data, and demanding state-of-the-art techniques and tools of big-data analytics. The course will simulate real data analytics projects in industry: students will form small teams (2 per team normally) to solve selected real problems that can be formulated as big-data analytics tasks. Instructor will provide a few sample problems and large-scale data sets. Students are encouraged to propose their own problems and data sets, which needs to be approved by the instructors before Week 3.
Every team should meet with the instructor and/or the GSI weekly to discuss project progress (which may or may not be scheduled during the class sessions). Feedback of the project progress will be updated on Trello and documented in the course Wiki. In-class sessions will be used to discuss issues that are related to all students/teams (such as tutorials of tools and algorithms). In particular, the first two weeks of class will be dedicated for team building and the introduction to computational environments, datasets, and sample projects. All projects and teams will be finalized in week 3 and presented in front of the class. A halfway project presentation will be held in class in week 7 or 8, and a final presentation of the projects will be held in the last week of class. In other weeks, teams do not need to attend in-class sessions unless for weekly progress report meetings. Members of individual teams must meet at least once every week other than the progress meeting and use different channels of coordination throughout the semester.
Every team will be assigned to a different project so that there is no direct competition among teams. Project documentation will be kept up-to-date on the course Wiki and be shared among all teams. Teams are encouraged to make peer critiques during the opening presentations and halfway presentations and to provide suggestions and help to each other throughout the semester.
In class sessions, every team will give a brief about their progress, and then they will collaborate with their team members and/or meet with the instructor.
A rough schedule of the projects is as follows:
Week 1: Introduction to computational environments, datasets, and sample projects. Team building
Week 2: Team building and project selection
Week 3: Team finalized. Project design finalized and presented in class
Week 8: Halfway report submitted, reviewed, and presented in class
Week 13: Project results submitted and reviewed.
Week 14: Final presentations and result dissemination.

—-

About Academic Integrity:

Unless otherwise specified in an assignment all submitted work must be your own, original work. Any excerpts, statements, or phrases from the work of others must be clearly identified as a quotation, and a proper citation provided. Any violation of the School's policy on Academic and Professional Integrity (stated in the Master's and Doctoral Student Handbooks) will result in serious penalties, which might range from failing an assignment, to failing a course, to being expelled from the program. Violations of academic and professional integrity will be reported to UMSI Student Affairs. Consequences impacting assignment or course grades are determined by the faculty instructor; additional sanctions may be imposed by the assistant dean for academic and student affairs.

About Students with Disabilities:

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way we teach may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate accommodations. SSD (734-763-3000; ssd.umich.edu/) (Links to an external site.)Links to an external site. *typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. I will treat any information that you provide in as confidential a manner as possible.*

About Student Mental Health and Wellbeing

The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and https://caps.umich.edu/ during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (734) 764-8320 and https://www.uhs.umich.edu/mentalhealthsvcs, or for alcohol or drug concerns, see www.uhs.umich.edu/aodresources.

For a listing of other mental health resources available on and off campus, visit: http://umich.edu/~mhealth/ (Links to an external site.)Links to an external site.