



Syllabus

Course Information	<p>MIS 431: Data Mining for Business Applications Location: Distance Education Course Website: https://gmudatamining.com</p>
Instructor	<p>David Svancer Office Hours - Available through Slack</p>
Course Description	<p>Data mining—the art of extracting useful information from large amounts of data—is of growing importance in today’s world. The amount of data flowing from, to, and through enterprises is enormous, and growing rapidly—more rapidly than the capabilities of organizations to use it. Enterprises are trying to make effective use of the abundance of data to which they have access: to make better predictions, decisions, and strategies. Therefore, managers now need to know about the possibilities and limitations of data mining. This course introduces data mining problems and tools to enhance managerial decision making. Students will learn how to ask the right questions and how to draw inferences from data by using the appropriate data mining tools. This course will enable students to approach business problems data-analytically, envision data mining opportunities in organizations, and follow up on ideas or opportunities that present themselves.</p>
Course Objectives	<p>Upon completion of the course, students will be able to:</p> <ol style="list-style-type: none"> 1. Describe the cross-industry standard process for data mining (CRISP-DM) and its application across various industries in business. 2. Execute a data mining project, from setting analytical goals to communication of results. 3. Apply modern software tools to prepare data, visualize complex datasets, and build machine learning models. 4. Assess the accuracy of different machine learning techniques in the context of specific business objectives and datasets. 5. Utilize advanced programming techniques and applied statistics to complete a data mining project and present business recommendations that highlight opportunities for extracting business value. 6. Recommend solutions to a high-level business problem using data mining algorithms.

<p>Course Methodology</p>	<p>The class format will combine reading, lectures, presentations, and other learning tools. The class will be interactive and require every student to be engaged in the classroom discussion and assignments. In addition to the lectures, screencasts and timely completion of assignments, every student will be expected to be an active participant and a dedicated individual applying what you learn to every element of the course work.</p>
<p>Required Textbooks</p>	<p>Required Text</p> <p><i>R for Data Science</i> (https://r4ds.had.co.nz/)</p> <p><i>An Introduction to Statistical Learning</i> (http://www-bcf.usc.edu/~gareth/ISL/)</p> <p><i>Statistical Inference via Data Science</i> (https://moderndive.com/)</p>
<p>Computer Requirements</p>	<p>Hardware: You will need access to a Windows or Macintosh computer with at least 2 GB of RAM and access to a fast and reliable broadband internet connection (e.g., cable, DSL). A larger screen is recommended for better visibility of course material. You will need speakers or headphones to hear recorded content and a headset with a microphone is recommended for the best experience. For the amount of Hard Disk Space required taking a distance education course, consider and allow for:</p> <ol style="list-style-type: none"> 1. the storage amount needed to install any additional software and 2. space to store work that you will do for the course. <p>If you consider the purchase of a new computer, please go to Patriot Tech to see recommendations.</p> <p>Software: Many courses use Blackboard as the learning management system. You will need a browser and operating system that are listed compatible or certified with the Blackboard version available on the myMason Portal. See supported browsers and operating systems. Log in to myMason to access your registered courses. Some courses may use other learning management systems. Check the syllabus or contact the instructor for details. Online courses typically use Acrobat Reader, Flash, Java, and Windows Media Player, QuickTime and/or Real Media Player. Your computer should be capable of running current versions of those applications. Also, make sure your computer is protected from viruses by downloading the latest version of Symantec Endpoint Protection/Anti-Virus software for free here.</p> <p>Students owning Macs or Linux should be aware that some courses may use software that only runs on Windows. You can set up a Mac computer with Boot Camp or virtualization software so Windows will also run on it. Watch this video</p>

about using Windows on a Mac. Computers running Linux can also be configured with virtualization software or configured to dual boot with Windows.

Note: If you are using an employer-provided computer or corporate office for class attendance, please verify with your systems administrators that you will be able to install the necessary applications and that system or corporate firewalls do not block access to any sites or media types.

Course-specific Software:

RStudio Cloud

Open source data science and statistical computing software available through a cloud-based application.

An R environment will be set-up for students in MIS 431 and an access link will be provided through Blackboard.

When creating an account on <https://rstudio.cloud>, **students must use their GMU e-mail address.**

Available at the following link: <https://rstudio.cloud/>

RStudio Cloud Guide: <https://rstudio.cloud/learn/guide>

DataCamp

DataCamp (<https://datacamp.com>) offers interactive R and Python courses on topics in data science, statistics, and machine learning. Students can learn through short video tutorials and interactive exercises from within their web browser. Courses on DataCamp do not require any software installation to complete.

Students in MIS 431 have been granted access to all DataCamp courses for a 6-month period. DataCamp will serve as a tool for MIS 431 students to enable learning programming concepts through hands-on interactive exercises.

In MIS 431, students are required to complete three DataCamp courses during the semester (150 points towards the final grade). Each course requires approximately 4 to 6 hours to complete.

	<p>Links to the description of the courses are provided below.</p> <p><i>Introduction to R</i> https://www.datacamp.com/courses/free-introduction-to-r</p> <p><i>Introduction to the Tidyverse</i> https://www.datacamp.com/courses/introduction-to-the-tidyverse</p> <p><i>Working with Data in the Tidyverse</i> https://www.datacamp.com/courses/working-with-data-in-the-tidyverse</p> <p>Students will receive an access link to join the MIS 431 team in DataCamp through Blackboard. After clicking the link, students will first be prompted to create an account. Students must use their GMU e-mail address to enroll (username@gmu.edu) and add their name as it appears in GMU records.</p> <p>Slack</p> <p>Slack (https://slack.com/) is a tool for collaboration where students can interact with their peers as well as the professor throughout the course.</p> <p>Students will receive an access link to join the MIS 431 Slack group.</p>
Course Website	<p>Blackboard 9.1 will be used for this course. You can access the site at http://mymasonportal.gmu.edu. Login and click on the "Courses" tab. You will see MIS 431 course. NOTE: Username and passwords are the same as your Mason email account. You must have consistent access to an internet connection in order to complete the assignments in this course through Blackboard (http://mymason.gmu.edu). Note the technology requirements for School of Business in your Blackboard course menu—it contains details of minimum technology requirements.</p> <p>The course website is at the following link: https://gmudatamining.com</p>
Participation	<p>Learning can only happen when you are playing an active role. It is important to place more emphasis on developing your insights and skills, rather than transmitting information. Knowledge is more important than facts and definitions.</p>

	<p>It is a way of looking at the world, an ability to interpret and organize future information. An active learning approach will more likely result in long-term retention and better understanding because you make the content of what you are learning concrete and real in your mind.</p> <p>Although an active role can look differently for various individuals, it is expected in this class that you will work to explore issues and ideas under the guidance of the professor and your peers. You can do this by reflecting on the content and activities of this course, asking questions, striving for answers, interpreting observations, and discussing issues with your peers.</p>
<p>Rules and Expectations</p>	<p>In correspondence/communication students will be expected to:</p> <ul style="list-style-type: none"> a) Be professional and respectful in correspondence b) Make reasonable requests of the instructor. We will be happy to clarify course material and answer legitimate questions; however, please exhaust other information sources (e.g., syllabus, Blackboard) for answering your question before contacting me and remember, "Poor planning on your part does not constitute an emergency on my part" <p>Regarding honesty in work students will be expected to:</p> <ul style="list-style-type: none"> a) Review the University integrity and honesty policies in the student handbook for guidelines regarding plagiarism and cheating (summarized below). I will gladly clarify my stance on any questionable or "grey area" issues you may have. b) Refrain from dishonest work as it will receive a minimum penalty of zero on the assignment and a maximum penalty of a zero for the course with a report to the Honor committee. The GMU Honor Code requires that faculty submit any suspected Honor Code violations to the Honor Committee. Therefore, any suspected offense will be submitted for adjudication.
<p>Mason Honor Code</p>	<p>The complete Honor Code is as follows:</p> <p><i>To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University community and with the desire for greater academic and personal achievement, we, the student members of the university community, have set forth this honor code: Student members of the George Mason University community pledge not to cheat, plagiarize, steal, or lie in matters related to academic work.</i></p> <p><i>(From the Catalog – catalog.gmu.edu)</i></p>
<p>Cheating Policy</p>	<p>Any form of cheating on an activity, project, or exam will result in zero points earned.</p> <p>"Cheating" includes, but is not limited to, the following: reviewing others' exam</p>

	<p>papers, having ANY resources utilized when not allowed, collaborating with another student during an individual assignment.</p> <p>If you have questions about when the contributions of others to your work must be acknowledged and appropriate ways to cite those contributions, please talk with the professor or utilize the GMU writing center.</p>
Plagiarism and the Internet	<p>Copyright rules also apply to users of the Internet who cite from Internet sources. Information and graphics accessed electronically must also be cited, giving credit to the sources.</p> <p>This material includes but is not limited to e-mail (don't cite or forward someone else's e-mail without permission), newsgroup material, information from Web sites, including graphics. Even if you give credit, you must get permission from the original source to put any graphic that you did not create on your web page. Shareware graphics are not free. Freeware clipart is available for you to freely use. If the material does not say "free," assume it is not.</p> <p>Putting someone else's Internet material on your web page is stealing intellectual property. Making links to a site is, at this time, okay, but getting permission is strongly advised, since many Web sites have their own requirements for linking to their material. Review the Honor Code here.</p>
Individuals with Disabilities	<p>Students with documented disabilities should contact the Office of Disability Services (703) 993-2474) to learn more about accommodations that may be available to them.</p> <p><i>(From the 2019-2020 Catalog – catalog.gmu.edu)</i></p>
Academic Integrity and Inclusivity	<p>This course embodies the perspective that we all have differing perspectives and ideas and we each deserve the opportunity to share our thoughts. Therefore, we will conduct our discussions with respect for those differences. That means, we each have the freedom to express our ideas, but we should also do so keeping in mind that our colleagues deserve to hear differing thoughts in a respectful manner, i.e. we may disagree without being disagreeable. http://oai.gmu.edu/</p>
Student Privacy Policy	<p>George Mason University strives to fully comply with FERPA by protecting the privacy of student records and judiciously evaluating requests for release of information from those records.</p> <p>Please see George Mason University's student privacy policy https://registrar.gmu.edu/students/privacy/</p>
E-Mail Policy	<p>Web: masonlive.gmu.edu</p> <p>Mason uses electronic mail to provide official information to students. Examples include notices from the library, notices about academic standing, financial aid information, class materials, assignments, questions, and instructor feedback.</p>

	<p>Students are responsible for the content of university communication sent to their Mason e-mail account and are required to activate that account and check it regularly.</p> <p>Students are also expected to maintain an active and accurate mailing address to receive communications sent through the United States Postal Service.</p> <p><i>(From the 2017-18 Catalog – catalog.gmu.edu)</i></p>
Course Grading & Evaluation	<p>Grades will be assigned as follows:</p> <p>A : 93.00 - 100%</p> <p>A- : 89.50 - 92.99%</p> <p>B+ : 86.50 - 89.49%</p> <p>B : 81.50 - 86.49%</p> <p>B- : 79.50 - 81.49%</p> <p>C+ : 77.50 - 79.49%</p> <p>C : 69.50 - 77.49%</p> <p>C- : 64.50 - 69.49%</p> <p>D : 59.50 - 64.49%</p> <p>F : 0 - 59.49%</p>
Writing Intensive Requirement	<p>This course has been approved by the Faculty Senate Writing Across the Curriculum Committee to fulfill all/in part the Writing Intensive requirement in the Business Analytics Concentration. It does so through the course projects where students will be writing executive summaries of their data analysis and modeling results within a particular business context. The data analysis project will be completed through a draft/feedback/revision process. The first draft will be due 11-08-2020; I will provide commentary on the draft, and the revised draft will be due on 11-15-2020.</p>
Class Participation (10 Points)	<p>Students are required to create at least one post to the # <i>introductions</i> Slack channel where they briefly describe their major, career goals, and skills they look forward to learning in MIS 431.</p>
Quiz (15 Points)	<p>One Quiz will be administered through Blackboard and will test the students' understanding of machine learning algorithms.</p>
DataCamp Courses (150 Points)	<p>Students are required to complete the 3 DataCamp courses listed in the Software requirements section of the syllabus. Successful completion of each course is worth 50 points. To get credit, students must upload their DataCamp Statement of Accomplishment in PDF format to the MIS 431 Blackboard site.</p>
Homework Assignments (75 Points)	<p>Students will complete 3 programming homework assignments.</p>

Midterm Data Analysis Project (100 Points)	Students will complete a data analysis project utilizing RStudio Cloud and the programming techniques that were covered in the first 7 sessions of the course
Final Project (150 Points)	Students will complete a final project utilizing RStudio Cloud and the programming techniques that were covered in the first 13 sessions of the course. The final project will be a complete implementation of the data mining process to a business problem and will build upon the skills developed in the midterm data analysis project.
<p>Need Help?</p> <p>Utilize the “Course Q&A” discussion forum or email your instructor directly.</p>	

Expect to work 10-15 hours per week on learning material and course assignments for this course.

Unless otherwise stated, all assignments are due by the end of the week in which they are assigned. For the purposes of this course, a week is defined as **beginning at 12:01 am each Monday EST, and ending at 11:59 pm on the following Sunday EST.**

To help you manage your schedule and time to complete the assignments in this course, please follow the recommended timeline below. If you have a question or concern or encounter a problem about an assignment, please contact me immediately so we can discuss and work out a resolution.

Week	Lessons	Assignments Due (by Sunday 11:59 pm)
Week 1 08-24	Lesson 1: <ul style="list-style-type: none"> Introduction to Data Mining Introduction to R Programming 	<ul style="list-style-type: none"> Register and access course software <ul style="list-style-type: none"> DataCamp RStudio Cloud Slack Course Participation <ul style="list-style-type: none"> # introductions post to Slack
Week 2 08-31	Lesson 2: <ul style="list-style-type: none"> Data Mining Process (CRISP-DM) Intermediate R programming 	<ul style="list-style-type: none"> DataCamp <ul style="list-style-type: none"> Introduction to R
Week 3 09-07	Lesson 3: <ul style="list-style-type: none"> Introduction to data analysis with <i>dplyr</i> 	<ul style="list-style-type: none"> Quiz - Blackboard
Week 4	Lesson 4:	<ul style="list-style-type: none"> DataCamp

09-14	<ul style="list-style-type: none"> Intermediate data analysis with <i>dplyr</i> 	<ul style="list-style-type: none"> Introduction to the <i>Tidyverse</i>
Week 5 09-21	Lesson 5: <ul style="list-style-type: none"> Joining related data frames Reshaping and pivoting data with <i>tidyr</i> 	
Week 6 09-28	Lesson 6: <ul style="list-style-type: none"> Data Visualization with <i>ggplot2</i> 	<ul style="list-style-type: none"> DataCamp <ul style="list-style-type: none"> Working with Data in the <i>Tidyverse</i>
Week 7 10-05	Lesson 7: <ul style="list-style-type: none"> Probability and Descriptive Statistics 	
Week 8 10-12	Lesson 8: <ul style="list-style-type: none"> Inferential Statistics and Resampling 	<ul style="list-style-type: none"> Homework #1
Week 9 10-19	Lesson 9: <ul style="list-style-type: none"> Model Training Process <ul style="list-style-type: none"> Data Preprocessing and Feature Engineering for Machine Learning Cross Validation 	
Week 10 10-26	Lesson 10: <ul style="list-style-type: none"> Linear Regression <ul style="list-style-type: none"> One predictor Multiple predictors 	
Week 11 11-02	Lesson 11: <ul style="list-style-type: none"> Introduction to Classification <ul style="list-style-type: none"> Logistic Regression Assessing Model Fit <ul style="list-style-type: none"> F-1 score, Precision, Recall, ROC, AUC 	<ul style="list-style-type: none"> Data Analysis Project
Week 12 11-09	Lesson 12: <ul style="list-style-type: none"> Discriminant Analysis <ul style="list-style-type: none"> LDA, QDA K-Nearest Neighbors (KNN) Intro to Hyperparameter Tuning 	<ul style="list-style-type: none"> Homework #2 – Executive Summary revision for the Data Analysis Project
Week 13 11-16	Lesson 13: <ul style="list-style-type: none"> Decision Trees and Random Forests Hyperparameter tuning with grid search 	<ul style="list-style-type: none"> Homework # 3
Week 14 11-23	Thanksgiving Holiday	Enjoy time with your loved ones

Week 15 11-30	Lesson 14: <ul style="list-style-type: none">• Introduction to Unsupervised Learning<ul style="list-style-type: none">○ Principal Components Analysis○ K-means Clustering	<ul style="list-style-type: none">• Final Project
------------------	---	---