Spring 2024

MGMT 635-102: Data Mining and Analysis

Dantong Yu

MGMT 635-102 - DATA MINING and ANAL FOR MANAGERS

Course Information

Course Number: MGMT 635-102
Course Title: MGMT 635 - DATA MINING&ANAL FOR MANAGERS
Instructor: Dantong Yu
Email: dtyu@njit.edu

Classroom: Hybrid of In-person classroom and Synchronous online (Attention for in-person AND online class is required). We alternate In-person and online classes between odd weeks (1,3,5,7,9,11,15) and the even weeks (2,4,6,8,10,12,14). The in-person classes are marked in black, and synchronous classes are in green.

Time: 6:00-8:50 PM Tuesday.
In-person classes (Weeks 1,3,5,7,9,11,15): KUPF 104
Mid-term Exam (Week 8): KUPF 104
Synchronous classes (2,4,6,10,12,14):
https://njit-edu.zoom.us/j/96796106520?pwd=bHNISGdkUDYvTVZDZUZkeFl1VFNkdz09

Office hours: Tuesday 12:00–1:30 PM:

In-person OFFICE (Weeks 1,3,5,7,9,11,15): CAB 2004
Synchronous online office hours (Week 2,4,6,10,12,14):

Join Zoom Meeting
https://njit-edu.zoom.us/j/91218424195?pwd=SVRsTEVHZ203Ymh4Qm5qYTZFMFdKUT09

Meeting ID: 912 1842 4195
Passcode: 430715

Course Description

This course provides an in-depth study of data mining and analysis, focusing on business applications. As business becomes increasingly complex and globalized, and competitiveness grows in most industries, managers must make better and faster decisions using available data. Data mining is an approach that uses powerful tools and techniques to unlock the value inherent in available organizational and external data. Data mining and analytics now routinely help organizations uncover hidden patterns and correlations in data and leverage these insights to

improve all business decision-making. The prerequisite of this class is basic data analytical skills and some understanding of Computer programming languages (Python or R).

This class is an introduction to the fundamental concepts of data science. You need not be a fluent programmer, but you must have some basic understanding of Python Programming to take this class.

The practice-oriented course develops the required skills to conduct data mining in different industries. Students will better understand the data mining and analysis techniques and gain hands-on knowledge of contemporary tools used for data mining.

This course is designed to enhance students' understanding of key concepts, approaches, and techniques for data mining. The learning materials offer technical depth to help students grasp the workings of data mining technologies. The course covers data mining processes, methods, and techniques, as well as the role and management of data, tools and metrics, text and web mining, sentiment analysis, and integration with Big Data.

Textbook and Materials

Required Textbook

[DSB] Data Science for Business What You Need to Know about Data Mining and Data-Analytic Thinking, By Foster Provost and Tom Fawcett, Publisher: O'Reilly Media. ISBN-13: 978-1449361327, ISBN-10: 1449361323. You can order it from Amazon.

[Chapters 9 and 10 of ISLP] You only need to read Chapter 10 of An Introduction to Statistical Learning with Applications in Python for Deep Learning. We will teach you to use Deep Learning to solve your project and exam questions. The remaining chapters will be supplementary material. Note that this textbook is an online OpenText Book; no Purchase is Needed.

Supplementary Material (Online OpenText Books, No Purchase is Needed)
[MMD] Mining of Massive Datasets: Jure Leskovec, Anand Rajaraman, Jeff Ullman, Download URL: http://www.mmds.org/
[ISLP] An Introduction to Statistical Learning with Applications in Python, Third Edition, Authors: **James**, G., **Witten**, D., **Hastie**, T., **Tibshirani**, R. https://hastie.su.domains/ISLP/ISLP_website.pdf

REQUIRED: Python

Please note that this class is not a language class. It provides some basic help on Python Programming. You do not need to be fluent in Python programming. It will be sufficient if you can read and understand some simple Python code and understand and modify the code as appropriate to fit your project requirements (in the second project). Because we will deliver some course material in Python, you might need to warm up your Python to a moderate level. Getting started earlier is always a good idea!

Students have some Python programming background or want to learn Python and open-source software in data science.

Python: Learning Python in two days (https://developers.google.com/edu/python/)

Scikit Learning: Machine Learning in Python

IPython https://ipython.org/

You can download all the above Python stuff at this site!

From this semester, we will go to Cloud for Python stuff: some reading material (How to import data into Google Cloud (internal link))

Course Outcomes

After taking this course, you should have improved substantially in the following three dimensions. One pragmatic way to think about this is that you will do remarkably better in a data-science-oriented interview, where these three dimensions are the most important. (This is so whether you are the interviewer or the interviewee.) The three dimensions are: first of all, you will approach (business) problems data-analytically. Think carefully & systematically about whether & how data can improve performance to make better-informed decisions for management, marketing, investment, etc. You will also be able to interact competently on the topic of data science and analytics. Know the fundamental principles of data science that are the basis for data mining processes, machine learning algorithms & analytics systems. Understand these well enough to work on data science projects and interact with everyone involved. Envision new opportunities. You will have hands-on experience in mining data and be prepared to follow up on ideas or opportunities and prepare pilot studies. You can

1. Identify business applications of data mining in an organization.

2. Apply the concepts, methodologies, and data mining frameworks to help organizations gain a competitive advantage.
3. Collect, organize, and analyze large data sets for useful business understanding.
4. Utilize techniques in emerging areas of web mining and text mining to provide practical business intelligence.
5. Use data mining tools and techniques to support organizational decision-making that is descriptive, predictive, and prescriptive.
6. Communicate the results of data mining and recommendations of actions to an organization's management effectively.
7. Develop and apply critical thinking, problem-solving, and decision-making skills for data mining and analysis.
8. Attain practical skills in using commercial software to solve real-world data mining problems.

9. Learn the practical skills of using open-source machine learning tools (Scikit-learning) to solve real-world data mining problems.

Prerequisites

This course is not a computer science data mining and machine learning class. You do not have to be a fluent programmer to take this course. You need to understand basic statistics (mean, variance, standard error, F-test, p-value), and you are willing to work with some software packages. We do not assign individual projects. Do not worry about your software programming skills. We only have two group projects that need writing, programming, implementation, and evaluation. The key is to team up with someone who has complementary skills and background and work on group projects. Based on your input, I will assign initial groups with mixed experience to ensure sufficient knowledge and skills in each group. You are evaluated based on teamwork and active participation.

Grading

Grade Scale

A	[92,100]
B+	[85, 92)
B	[80, 85)
C+	[75, 80)
C	[65, 75)
F	[0, 65)

Grades will be based on the following task distribution:

Forum Participation	5%
Homework (Quiz)	10%
Lab Assignments	10%
Group Project 1: Credit rating prediction	10%
Group Project 2: Lending club	15%
Midterm (Close Book and Close Notes)	20%
Final (Close Book and Close Notes)	30%

Commitment Requirement

Please note that this course is technical-oriented. You are required to work in groups to implement projects. You are expected to spend as much time as possible practicing the concepts and techniques from the textbook. Machine learning and data mining cover a significant amount of knowledge (some of it is difficult) within one semester. I expect that you will try your best and sometimes work beyond your comfort zone. By taking this online course, you are expected to spend as much effort as you expect in the face-to-face course. Indeed, I do care about what you learn in a semester while paying less attention to the grade! In this class, we will use Python to learn and practice the basic concepts from the textbook. I believe that all students, regardless of whether they are in a business major, management major, or information major, in this new era of data mining and machine learning, must know the Python programming language.

Special Notes on Group Project Assignments

Please note that all projects are GROUP projects. The number of group members is 3. If all members contribute, your actual workload during the semester should be about 1/3 of the tasks in any project.

Here are some reasons why these group assignments are essential for you, besides the fact that they are assigned.

- They allow you to study a topic important to the field of data mining and machine learning. In so doing, you come across many sources of information that may be useful to you later on.
- You will practice synthesizing material and information from many sources in an original manner that should form a well-written and persuasive document.

- Since these assignments are in a group/project setting, you will get useful experience in working as a team, which involves, among others, sharing responsibilities, getting to a consensus, and compromising.
- Since the completed assignment has a due date, you will learn how to work with team members and meet deadlines.

Students should form teams of three or four members (four members are the maximum) as soon as possible for the group assignments. You may use the forum created in the course canvas site to interact and form groups. Once this is done, have one individual – who will be the team leader and contact person – post the names of the group members in that forum. I take group assignments very seriously. All members of the group must participate and contribute equally to all aspects of the assignments. The group will receive one grade for the project, so all members must agree on all the details of the assignment before it is submitted.

Course Schedule

Dates	**Module**
W1 (01/16)	Course Overview and Introduction to Data Mining
W2 (01/23)	Introduction to Predictive Modeling (Chapters: DSB-2, RDM-2)
W3 (01/30)	The Data Mining Process, Example: CRISP-DM (Chapter: RDM-3)
W4 (02/06)	Supervised Segmentation, Example: Decision Tree (Chapters: DSB-3, RDM-4, ISLP-8)
W5 (02/13)	Discriminant Functions (Chapters: DSB-4, RDM-5)
W6 (02/20)	Overfitting and Its Avoidance (Chapter: DSB-5)
W7 (02/27)	Support Vector Machine (ISLP-Chapter 9)
W8 (03/05)	**Midterm** (Close Book and Close Notes) **[The mid-term exam will include questions that are similar to knowledge point checking quizzes and some simple essay (concept) questions; please be sure you provide answers to essay questions because they usually have more points and partial credits]**
W9 (03/12)	**Spring Break**
W10 (03/19)	Decision Analytic Thinking (Chapter DSB-7)
W11 (03/26)	Visualizing Model Performance (Chapter DSB-8)
W12 (04/02)	Evidence and Probabilities, Example: Naïve Bayes Model (Chapter: DSB-9)
W13 (04/09)	Deep Learning (ISLP-Chapter 10, Simple Neural Networks MLP and CNN)
W14 (04/16)	(04/29 online because we run Friday schedule) Deep Learning Continue ((ISLP-Chapter 10), Advanced ANN, CNN, RNN, Transformer (Foundation of Chat GPT))
W15 (04/23)	Text Analytics (Chapters DSB-10)

W16 (04/30) Similarity and Nearest Neighbors (Chapter DSB-6), Unsupervised Data Mining, Example: Clustering (Chapter DSB-6, Chapter: MMD-7)

TBA **Final Exam** (Close Book and Close Notes)

Late Work and Make-Up Exams

All assignments are expected when due, as stated in your syllabus.

IMPORTANT NOTE REGARDING MIDTERM AND FINAL EXAMS

Online Course Exams and Proctoring

NJIT policy requires that all midterm and final exams must be proctored, regardless of delivery mode, in order to increase academic integrity. Note that this does not apply to essay or authentic based assessments. Effective beginning Fall semester 2020, students registered for a fully online course section (e.g., online or Hyflex mode) must be given the option to take their exam in a completely online format, with appropriate proctoring.

In this course, you are required to use Respondus Monitor to ensure academic integrity for exams. See below for more information on Respondus.

Using Respondus LockDown Browser and a Webcam for Online Exams

Respondus LockDown Browser is a locked browser for taking assessments or quizzes in Canvas. It prevents you from printing, copying, going to another URL, or accessing other applications during a quiz. If a Canvas quiz requires that LockDown Browser be used, you will not be able to take the assessment or quiz with a standard web browser. You may be required to use LockDown Browser with a webcam (Respondus Monitor), which will record you during an online exam.

The webcam can be built into your computer or can be the type that plugs in with a USB cable. Watch this short video (Links to an external site.) to get a basic understanding of the LockDown Browser and the webcam feature. A student Quick Start Guide (PDF) (Links to an external site.) is also available.

1. Download and install LockDown Browser from this link: http://www.respondus.com/lockdown/download.php?id=264548414 (Links to an external site.)

2. Once your download has finished, locate the “LockDown Browser” shortcut on the desktop and double-click it. (For Mac users, launch “LockDown Browser” from the Applications folder.)
3. You will be brought to the Canvas login page within the LockDown Browser.
4. Under “My Courses,” click on the course in which you have to take the exam that requires the LockDown Browser.
5. After you enter the course, find the exam and click on it.
6. A confirmation prompt will appear. Click the “Start attempt” button. Once a quiz has been started with LockDown Browser, you cannot exit until the “Submit all and finish” button is clicked.
7. If you are required to use a webcam (Respondus Monitor), you will be prompted to complete a Webcam Check and other Startup-Sequence steps.

Software and Hardware Requirements

We will use Python as the programming language for the data mining tasks in this course. There are many options for the Python coding platform. Two of the most popular choices are Jupyter Notebook and Google Colaboratory. You may use either.

Sometimes, you must use Word processing and presentation software, such as MS Word and PowerPoint, found in Microsoft Office. You will also need to be comfortable with various aspects of using the Internet, such as: Search engines, Newsgroups, E-mail, Ability to download files

To view certain media elements in this course, you must have several browser plug-ins such as Shockwave, Flash, and Adobe Acrobat on your computer. Use the links in the course to download and install the appropriate software application.

Important: With regards to plug-ins, ensure you are using the most recent version of each plug-in you require. View the hardware and software requirements for this course.

Academic Integrity:

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found in NJIT Academic Integrity Code.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu."

Request for Exception:

When a student invokes extenuating circumstances for any reason (late withdrawal from a course, request for a make-up exam, request for an Incomplete grade, request for accommodation due to illness – COVID related or other –), the student should be sent to the Dean of Students Office. The Dean of Students will be making the determination of whether extenuating circumstances exist or not and will notify the instructor accordingly. Instructors should never request or accept medical or other documents from students; such documents need to be submitted by the student to the Dean of Students.

Recording Class Session:

I will record all of our classes. I will record blackboard, projection screen, and voice. I will not record any student images in my classroom. Class recordings can be a valuable resource that students can use to 1) Review and improve their understanding of the course material 2) Watch recordings of classes that they were unable to attend. This can be particularly useful for students who miss a class due to technical difficulties or illness.

For the hybrid course, attendance is mandatory for both in-person and online classes. Attendance cannot be earned by watching video recordings afterward, and skipping class without approval will result in an absence.

The student with Disabilities Codes

NJIT adheres to section 504 of the Rehabilitation Act (ADA) of 1990. Appropriate accommodations are provided at no cost to the student. If you have any questions or would like additional information, please contact Dr. Phyllis Bolling, Center for Counseling and Psychological Services (C-CAPS), Campbell Hall, (entry-level), room 205, (973) 596-3420. For further information, visit the Student Disability Services website.

Student Services and Support

Students may contact the IST Service Desk with any technical questions. Questions or problems can be submitted via web form by going to **https://servicedesk.njit.edu** and clicking on the "Report your issue online" link. You may also call the IST Service Desk with any questions at 973-596-2900. NJIT passwords may be changed using the Global Password Change mechanism. You will need to know your current UCID and UCID password. Questions can be referred to as 973-596-2900.