

MGS 3701 / MKT 3950: Data Mining

Welcome to Data Mining Course

Chungil Chae



温州肯恩大学
WENZHOU-KEAN UNIVERSITY

2022/01/01 (updated: 2022-02-15)

Introduction



Instructor Introduction

Chungil Chae



- 채충일, 蔡忠壹, Chad), cchae@kean.edu / cchae@wku.edu.cn, chadchae.github.io
 - Wenzhou-Kean University
 - Management, Business Analytics
 - Assistant Professor
 - Chad, Dr.Chae, Prof.Chae all fine
- Research Interest
 - Social Network Analysis
 - Social Sequence Analysis
 - Human Behavior
 - Knowledge Sharing
 - Learning Experience / Learning Path
- Path
 - Management -> Engineering(HRD in / 29



Class Time & Office Hour

For two weeks, zoom: <https://us02web.zoom.us/j/3550604109>

MKT3950, C102

- Tuesday, Thursday
- 5:30pm - 6:45pm

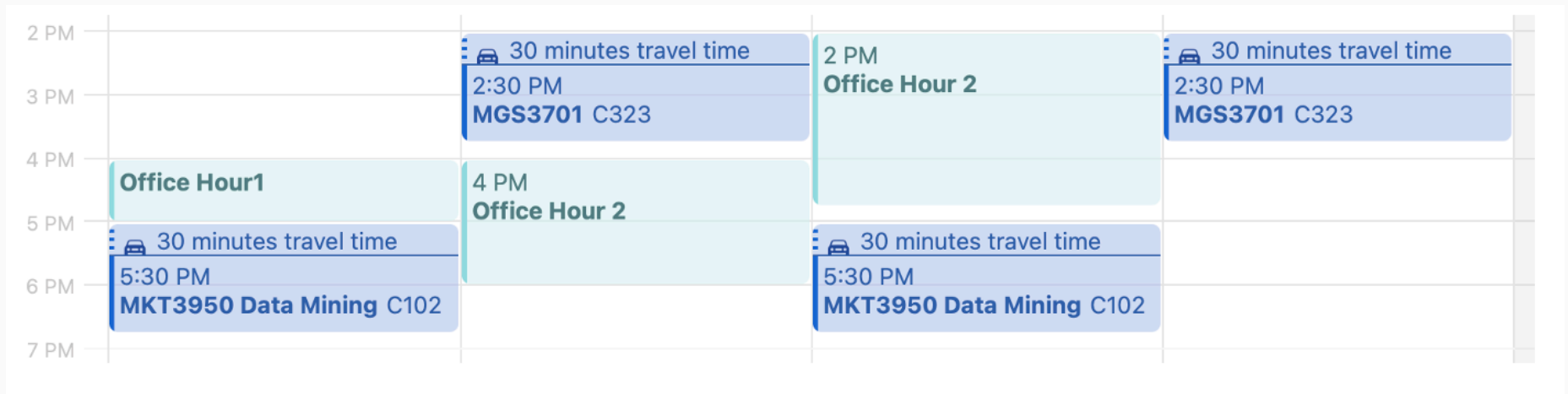
MGS3701, C323

- Wednesday, Friday
- 2:30pm - 3:45pm

- In zoom session, student will asked answering the question, and asked to activate camera
- Two weeks semester started, course will be delivered in-class format only



- CBPM B223
- Tuesday 4-5pm, Wednesday 4-6pm, Thursday 2-5pm

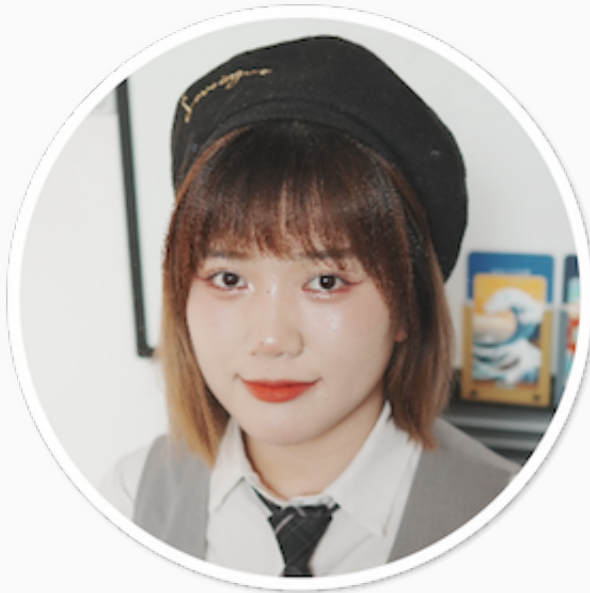




Teaching Assistant

MKT 3950

- Yuxi Huang (Rosie)



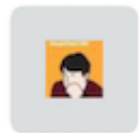
MGS 3701

TBA



Resource

Wechat Group



MKT3950 MGS3701 DM
spring 2022



LMS

- <http://142.93.124.81/apps/moodle/login/index>
- ID: your first part of email (e.g., cchae, cchae@kean.edu)
- PW: 1111 (after login, you need to change it)

A screenshot of a Moodle login page. At the top is a hexagonal logo with a stylized person's head and colorful lines, with the text 'chadchae' above it. Below the logo are two input fields: 'Username / email' and 'Password'. To the right of the 'Password' field is a link that says 'Forgotten your username or password?'. Below the 'Username / email' field is a checkbox labeled 'Remember username'. Below the 'Password' field is a message that says 'Cookies must be enabled in your browser' with a small icon. A blue 'Log in' button is at the bottom. Below the login form is a section titled 'Is this your first time here?' with a message that says 'Please, ignore email authentication. All user registration will be processed after site manager approved'.

Course Introduction



- Data is one of the most valuable and strategic assets that organizations have.
- Data mining is a continuous process that is crucial for many organizations to discover and extract information and patterns from their data to make business-related decisions using data-driven approaches.
- The objective of the course is
 - to familiarize the students with the data mining process and applications from a managerial perspective
 - to help discover information and knowledge in large data sets to support the business needs.

Prerequisite(s): MGS 2150



House Rule

- **I Will NOT ACCEPT any requests for changing exam or class schedule due to other class activities**
- Don't try negotiate with your instructor
- Academic Integrity (<https://wku.edu.cn/en/org/student-affairs/code-of-conduct/>)



- NO excuse for absent due to individual schedule, except:
 - Health issue
 - Family issue
 - Emergency that WKU alert

Category	Penalty
On time	100% score for attendance
5 - 10 mins late	50% score for attendance
More than 10 mins	10% score for attendance
Absent	No score for attendance
More than 3 times absent	F



In-class

- Turn-off your mobile/cell phone
- No computer and digital device allowed except lab time (refer to course schedule)

In Exam

- Lateness is not allowed (Please be in class 5 mins before exam start)
- Once exam started, you can't leave until you submit your answer sheet - NOT allowed any digital devices
- Student Card
- Pen/Eraser

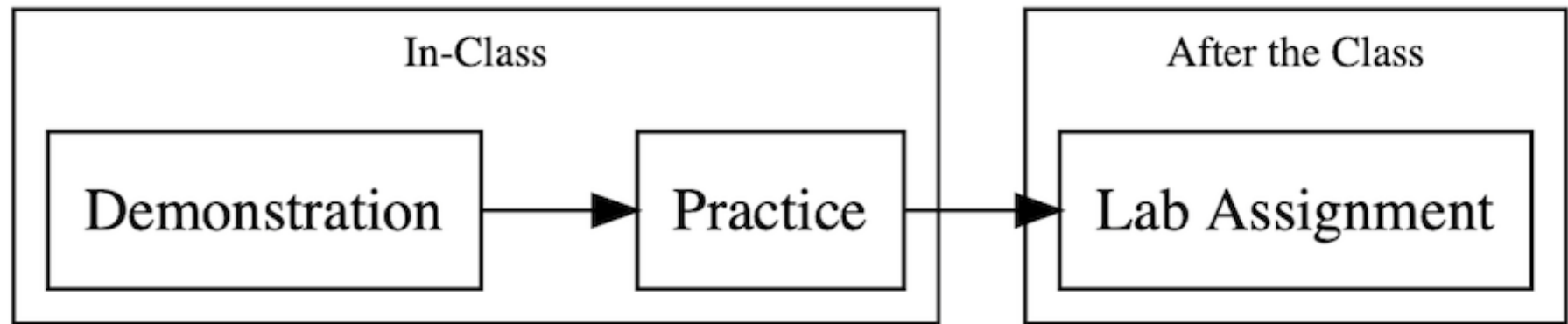
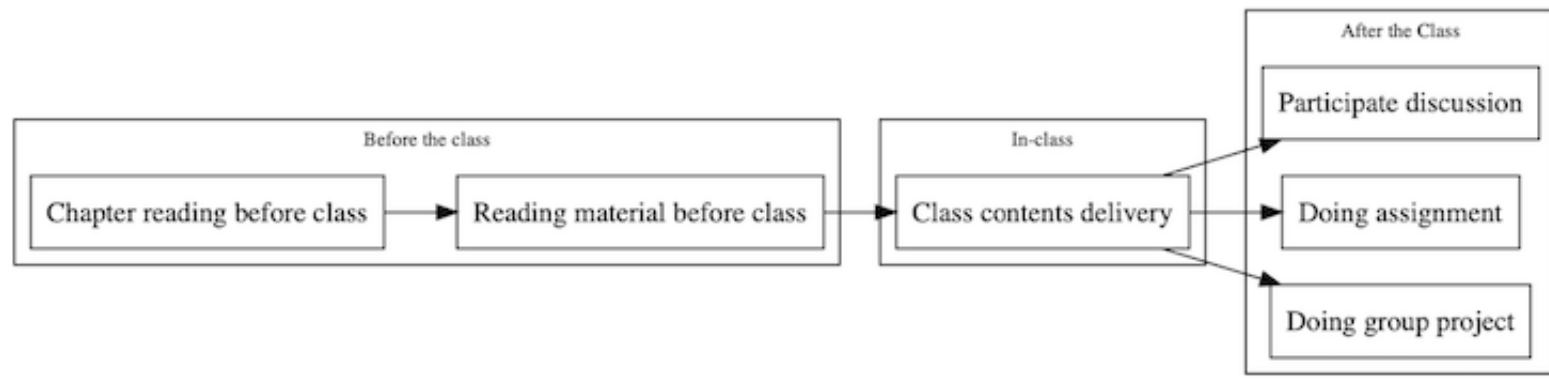


Textbook

- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2018). Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). Data mining for business analytics: concepts, techniques and applications in Python. John Wiley & Sons.



Class Model





Submission

- Single space
- PDF format from markdwon generated document (markdown > word > pdf also fine) only
- File name ([ID_Chinese Name in Eng_ChapterX_Summary.pdf)
 - **NO SPACE**
 - file format is critically important
 - (1111111_ChungilChae_assignment1.pdf)
- Submit to chadlab LMS
- Due: Sunday midnight a week before new chapter started (refer to the course schedule)

Task



Task

- Each task, exam, activity, and group task is 100 point score base but different percentage

Task	Category	Percentage
Attendance * 32	Activity	1%
Proposal	Ind	2%
Data Collection Plan	Ind	2%
Final Proposal	Ind	2%
Project Portfolio	Group	3%
Final Report (report, data, code, video)	Group	10%
Poter Session	Group Task	10%
Exam1	Exam	20%
Exam2	Exam	20%
Exam3	Exam	20%
Total	1000 Points	100%



Attendance

- Attendance and class participation is a basic element of the learning experience.
- Attendance will be assessed by a random attendance check.
- For the class participation, this semester, students are asked to submit 5 mins in-class essay (one or two paragraphs of reflection or impression, and interpretation of what they learned) and 10 discussion questions in online blackboard (one answer, one comment on the other student's answer).
- Also, by the course instructor's observation, some extra class participation will be scored.

Category	Penalty
On time	100% score for attendance
5 - 10 mins late	50% score for attendance
More than 10 mins	10% score for attendance
Absent	No score for attendance
More than 3 times absent	F



Assignment

Lab Assignment (individual)

- In every lab session, there will be a lab assignment
- Submit your rmarkdown or jupyter notebook file
- File name ([ID_Chinese Name in Eng_lab_assignmentX.pdf])
 - (1111111_ChungilChae_lab_assignment1.pdf)



Initial Proposal (individual)

- 1 page single space
- State motivation, purpose and plan for your analytics project - PDF format from markdown generated document only

Data Collection Plan (individual)

- 1 page single space
- State your plan for data collection or generation with example

Final Proposal (individual)

- up to 5 pages single space
- Revise your initial proposal with data collection plan
- Add your choice of method from chapter 4 to 21, reasoning, and procedure



Final Project

Final Project Portfolio (Group)

- Create a github web portfolio
- Portfolio website should include your team's final project

Final Project Report (Group)

- Student will be asked to work as team and for the final project. - Choose analytics subject, data, and method in team discussion - Using knowledge and skills in class
- Final project includes many tasks and product: – Final report – Code that used for analytics – Data – Promotion video – Poster



Code

- In case you using R/Python, you should submit r/python script separately - In case you using Tableau, you should submit full tbx file and data
- Code and product should be reproducible

Data

- Students should submit data file with project code
- In a case the data is located in internet, your code should includes a code for load data file from internet Promotion Video
- 5 mins promotion video for your poster session
- Explain your project shortly and encourage other to come to your poster session
- Using media lab
- Upload video to 'Youtube' or 'bilibili'



Exams

- Three exams (exam1,2,3 [see course schedule]) - 25 ~ 40 multiple choice / one page essay
- All exams are in-class exam in class time frame - No open book
- Students allows: – One sheet of hand-writing summary

Three non-cumulative exams will be given during the term. The exams may consist of multiple choice, true/false, and short answer essay. The exams will primarily cover the textbook materials, lectures, and discussions. However, they may also include required readings, videos, selected cases (if any), and classroom activities. Students are expected to take exams on the scheduled dates. Makeup exams will not be given unless the absence is excused by the instructor prior to the exam date. You are expected to be on time for all exams, and extra time will not be allocated to any student who arrives late. Students who are registered with Disability Services and who may need individual arrangements should schedule a meeting with the instructor to work out the details. No student will be excused from taking any exam, except for valid documented medical reasons, or family emergency, and any of those situations should be attested to by a medical practitioner or other suitably qualified professional. The exam schedule is available at the beginning of the term so make



Grading

- Round up from decimal point
- **No negotiation unless clear miss grading found**

Grade	Percentage
A	>93%
A-	90% ~ 93%
B+	87% ~ 89%
B	84% ~ 86%
B-	80% ~ 83%
C+	77% ~ 79%
C	70% ~ 76%
D	60% ~ 69%
F	< 60%



Policy and Integrity Professionalism

Policy

- Laptops, phones, PDA's, and similar electronic devices are prohibited in class. Please make sure these are turned off and put away before class begins. No email, texting, tweeting, et al, or web surfing will be allowed in this class out of respect for others.
- Attendance is expected and required. It will be your responsibility to obtain class notes and/or handouts from your classmates or me.
- Students arrive on time. On time, arrival ensures that classes are able to start and finish at the scheduled time. On time arrival shows respect for both fellow students and faculty and it enhances learning by reducing avoidable distractions.
- Students are fully prepared for each class. Much of the learning takes place during classroom discussions. When students are not prepared, they cannot contribute to the overall learning process. This affects not only the individual, but their peers who count on them, as well.
- All assignments are due on the date indicated unless prior approval has been granted. Late assignments will receive a minimum of a one-grade reduction.



- Students are conscientious both in giving feedback to the professor and other students through discussions or emails as well as adjusting their attitudes and behaviors in response to others' opinions. Students should review the Student Code of Conduct, as it discusses expectations of appropriate conduct in the classroom: <http://www.kean.edu/KU/Code-of-Conduct>.
- You are responsible for all administrative announcements - No extra-credit projects will be given at anytime during the semester.



Academic Integrity

- Integrity is important in the “real world” and the classroom alike.
- When doing your written graded assignments for this class, you are not permitted to copy material from assignments done by you in another class or assignments of other students who have taken this class or other classes in the past.
- Individual assignments are to be done individually, without collaborating with other students. Cooperation between teams in preparing team assignments is also prohibited.
- It is also a violation of academic integrity to sign in on an attendance sheet for someone who is absent from class, or to otherwise misrepresent one’s lack of attendance at class.
- Clearly, activities such as sharing class notes or discussing in-class materials outside of class are not honor code violations.
- Indeed, such interactions are encouraged because it enhances learning.
- The bottom line is that unethical behavior in any form is not acceptable and maximum penalties for any student violating the University’s policies will be sought.

Please go to the following website (<https://www.kean.edu/media/academic-integrity-policy>) for the detailed information about the “University Policy” on academic integrity.



Course Schedule

1	w1	S1	chapter0	Introduction	Assign1:introduction
2	w1	S2	chapter0	Software installation	
3	w2	S1	chapter1	Introduction	
4	w2	S2	chapter1	Introduction	
5	w3	S1	chapter2	Overview of the Data Mining Process	
6	w3	S2	chapter2	Overview of the Data Mining Process	
7	w4	S1	chapter3	Data Visualization	
8	w4	S2	chapter4	Dimension Reduction	
9	w5	S1	chapter5	Evaluating Predictive Performance	Assign2: proposal(1p)
10	w5	S2			Exam1



11	w6	S1	chapter6	Multiple Linear Regression	
12	w6	S2	chapter7	k-Nearest Neighbors (kNN)	
13	w7	S1	chapter8	The Naive Bayes Classifier	
14	w7	S2	chapter9	Classification and Regression Trees	Assign3: Data collection plan
15	w8	S1	chapter10	Logistic Regression	
16	w8	S2	chapter11	Neural Nets	
17	w9	S1	chapter12	Discriminant Analysis	
18	w9	S2	chapter13	Ensembles and Uplift Modeling	Assign4: Final Proposal
19	w10	S1			Exam2



20	w10	S2	chapter14	AR and CF	
21	w11	S1	chapter15	Cluster Analysis	
22	w11	S2	chapter16	Handling Time Series	
23	w12	S1	chapter17	Regression-Based Forecasting	Assign5: Final Report
24	w12	S2	chapter18	Smoothing Methods	
25	w13	S1	chapter19	Social Network Analytics	
26	w13	S2	chapter20	Text Mining	
27	w14	S1			Exam3
28	w14	S2	chapter21	Cases	Assign6: Project Portfolio
29	w15	S1		Poster Session	
30	w15	S2		Poster Session	

Chapter Video



What is Data Mining?



References

Shmueli, G., P. C. Bruce, P. Gedeck, and N. R. Patel (2019). Data mining for business analytics:

Assignment



Introducing your self

- To do
 - 1 page self introduction(single space)
 - Add your picture (your face should be clearly shown)
- Requirement
 - Picture of your face
 - Your student ID
 - Your name (Chinese name in eng and Eng name)
 - Your introduction
- Format
 - id_name_assignment1.pdf
 - 11111_Chungil_Chae_assignment1.pdf
 - PDF
- Submission
 - Class LMS
- Due: Feb 27, 11:58pm(Bejing Time)