

데이터마이닝 406.429
2012년 2학기
서울대학교 산업공학과

1. 목표: 본 강좌에서는 학부생들에게 데이터마이닝의 핵심 개념과 실제 응용사례를 소개한다. 특히, 연관분석, 군집화, 회귀분석, 의사결정나무, k-NN, 신경회로망과 같은 핵심적인 기법을 공부하고, 이 기법들을 XLMiner 패키지로 실습을 하며, 실제 비즈니스 문제를 데이터마이닝 문제로 변환하고, 데이터 전처리 및 모델링을 수행하고, 평가하는 기말 프로젝트를 수행한다.
2. 시간과 장소: 매주 월,수 3:30 pm- 4:45 pm (39-B119)
3. 담당교수: 조성준 (39 동 302 호, zoon@snu.ac.kr, <http://dmlab.snu.ac.kr/~zoon>), 면담시간은 월,수 2:30~3:30 또는 예약
4. 조교: 박은정, 양지원 (39 동 339, 880-7025)
5. 교재
 - A. Data Mining for Business Intelligence : Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner 2nd edition, Galit Shmueli, Nitin R.Patel and Peter C.Bruce, 2010, Wiley, ISBN 978-0-470-52682-8 (번역본: 비즈니스인텔리전스를 위한 데이터마이닝, 이엔비플러스, 2012)
 - B. 뉴머러티, The Numerati, Stephen Baker, 2010
 - C. The Economist, Feb 25, 2010

6. 평가:

Task	Weight	Evaluation
개념 및 모델 이론	55%	퀴즈: 5 회 * 3 = 15 시험: 2 회 * 20 = 40
문제해결, 모델링, 패키지 사용, 팀 워	45%	기말 프로젝트: 제안 15, 중간 15, 기말 15
Total	100%	

7. 홈페이지: <http://dmlab.snu.ac.kr/~zoon/2011dm.html>

8. 퀴즈와 시험

- A. 퀴즈는 일정이 표시된 월요일 강의시작과 함께 당 주 reading assignment 부분에서 출제됨. 목적은 용어 이해와 개념 이해. 총 7 회 가운데 좋은 점수 5 개를 더함.
- B. 시험은 2 회 출제됨. 문제는 중요 개념, 수식 및 응용에서 출제됨.

9. 기말 프로젝트

- A. 3 인 1 조로 구성되는 팀 프로젝트이며, 배경이 다른 학생들을 같은 조로 구성함.
- B. 데이터마이닝을 적용하여 해결할 수 있는 비즈니스 문제를 도출하고, 데이터

소스를 구하고, 이를 데이터마이닝 문제로 셋업하고, 모델을 구축평가하고, 최종적으로 비즈니스 적용에 대해 소개한다.

- C. 제안서 (최대 ppt 10 장) Business Background, Business Problem, Data Mining Problem, How to obtain Data, Models to use, Expected Data Mining Results, Expected Business Implications
- D. 중간 보고서 (최대 ppt 20 장) Proposal + Data Exploration, Model used, Data Mining Results
- E. 최종 보고서 (최대 ppt 20 장) Progress Report + Enhanced Results, Business Implications, Future Work.

10. 자기 소개서

- A. A4 용지에 자기 이름, 사진, 학과, 학년, 이메일 주소
- B. 데이터마이닝 수강 이유, 배우고자 하는 내용, 및 장래 인생 계획과의 관계
- C. 그 외 기타 자기 소개

11. Delay Penalty: 20% off (same day), 50% (next day) and 100% (after that)

12. 일정, 상황에 따라 변경 가능

No.	Date	Topic	Reading Assignment
1	9/3	Introduction	E:The data deluge
2	9/5	Introduction, Overview of DM Process	Ch 1 and 2
3	9/10	데이터마이닝 사례 (3인 발표)	N:1,2 E:Data, data everywhere, All too much, Needle in a haystack
4	9/12	Data Visualization	Ch 3, 자기소개서 due
5	9/17	Dimension Reduction	Ch 4, quiz
6	9/19	Association Rules	Ch 13
7	9/24	Cluster Analysis	Ch 14, quiz
8	9/26	Cluster Analysis	Ch 14
9	10/1	XLMiner, SAS Eminer	추석 (9.26 보강)
10	10/3	데이터마이닝 사례 (3인 발표)	N:3,4 E:A different game, Clicking for gold (10.10 보강)
11	10/8	Evaluating Classification and Predictive Performance	Ch 5, quiz, 출장 (조교)
12	10/10	Multiple Linear Regression	Ch 6
13	10/15	Exam 1	
14	10/17	XLMiner, SAS Eminer	
15	10/22	Term Project Proposal	
16	10/24	Term Project Proposal	
17	10/29	k-Nearest Neighbors, Naïve Bayes	Ch 7, Ch 8, quiz
18	10/31	데이터마이닝 사례 (3인 발표)	N:5,6 E: The open society, Show me
19	11/5	Classification and Reg trees	Ch 9, quiz
20	11/7	Classification and Reg trees	
21	11/12	Logistic Regression	Ch 10, quiz
22	11/14	Logistic Regression	
23	11/19	Neural Network	Ch 11, quiz
24	11/21	Neural Network, XLMiner, SAS Eminer	
25	11/26	Term Project Progress Presentation	
26	11/28	Term Project Progress Presentation	
27	12/3	Exam 2	
28	12/5	데이터마이닝 사례 (3인 발표)	N:7, 멧음말, E: New Rules for big data, Handling the cornucopia
29	12/10	Term Project Final report	
30	12/12	Term Project Final report	12/14 종강