

데이터마이닝 406.429

2015 년 2 학기

서울대학교 산업공학과

1. 목표: 데이터마이닝은 의사결정 하는데 사용할 Insight 와 Foresight 를 데이터로부터 추출하는 수리 계산적 작업이다. 본 강좌에서는 첫째, 연관분석, 군집화, 회귀분석, 의사결정나무, k-NN, 베이지언, 신경회로망과 같은 핵심적인 기법을 공부하고, 둘째 이 기법들을 R 패키지 등을 활용하여 실습하며, 셋째, 다양한 실제 비즈니스 상황에서 어떠한 기법으로 어떠한 분석을 실시하여 문제를 해결하는지 공부하고, 마지막으로, 실제 비즈니스 문제를 데이터마이닝 문제로 변환하고, 데이터 전처리 및 모델링을 수행하여 인사이트와 포사이트를 추출하고 평가하는 기말 프로젝트를 수행한다. 이를 통해, 향후 대학원 수준의 고급 기법을 공부하는데 기초를 닦으며, 실제 데이터를 분석하여 비즈니스 문제를 해결할 수 있는 구체적인 능력을 함양한다. 또한, 최근 많은 관심을 받고 있는 Artificial Intelligence, Big Data, Internet of Things, Smart Factory, Social Media, Fintech 등과의 관련성에 대해서도 공부한다.

2. 시간과 장소: 매주 월, 수 3:30 pm- 4:45 pm (39 동 325 호)

3. 담당교수: 조성준 (39 동 302 호, zoon@snu.ac.kr, <http://dm.snu.ac.kr/ko/members>), 면담은 이메일로 예약

4. 조교: 심재웅, 김한결 39 동 339 호, 880-7025

5. 교재: 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) [N] The Numerati, Stephen Baker, 2010 [E] The Economist, Feb 25, 2010

6. 평가:

Task	Weight	Evaluation
예습	10	퀴즈: 5 회 * 2 = 10
이론적 이해	40	시험: 2 회 * 20 = 40
Formulation / 사례 발표 및 토론 참여	10	부교재 및 Formulation
프로젝트 주제 제안	15	임팩트, 데이터, 마이닝 계획
전처리, 모델링, 패키지, 결과 해석, 팀 워	25	프로그레스 10, 파이널 15

7. 홈페이지: <http://dm.snu.ac.kr/dm2015>

8. 퀴즈와 시험

A. 퀴즈는 일정이 표시된 월요일 강의시작과 함께 해당 주 reading assignment 에서 출제됨. 목적은 용어 이해와 개념 이해. 총 7 회 가운데 점수가 좋은 5 회를 최종 성적에 반영함.

B. 시험은 2 회 출제됨. 문제는 중요 개념, 수식 및 응용에서 출제됨.

9. 프로젝트

A. 주제 제안은 개인적으로 하고, 이 가운데 좋은 주제를 골라 3~4 인 1 조로 팀 구성. 좋은 주제란 비즈니스 임팩트가 있고, 데이터의 양과 질이 좋으며, 마이닝이 가능한 가에 따라 판단함.

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 3 장) Business Background, Business Problem, Data Mining Problem, How to obtain Data,

Models to use, Expected Data Mining Results, Expected Business Implications or Business Impact

D. 중간 보고서 (최대 ppt 10 장) Proposal + Data Exploration, Model used, Data Mining Results. 시각적 data exploration 4 페이지, 초기 분석 3 페이지 포함.

E. 최종 보고서 (최대 ppt 20 장) Progress Report + Enhanced Results, Business Implications, Future Work. Enhanced 분석 결과 및 비즈니스 implication 내용 중요.

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

11. 일정 (변경 가능)

주차	일자	교재 (장)	
1	9/2, 7	1, 2	Admin, DM 소개, Overview of DM Process
2	9/9, 14	3,4	Data Visualization, Dimension Reduction, Spotfire 실습
3	9/16, 21	5,6, 10	예측 및 분류, 성능평가, 선형회귀, 로지스틱 회귀
4	9/23, 30	18.5	R 실습 -카탈로그구매반응 숙제, 반응모델링 리뷰
5	10/5, 7	14	Cluster Analysis-비누고객 숙제, 응용 사례,
6	10/12, 14	별도 자료	RFM, 마켓세분화 리뷰 , 포물레이션 연습,
7	10/19, 21		Term Project Proposal (*19 일 75+75)
8	10/26, 28		Term Project 팀 구성 , 자율학습
9	11/2, 4	13,	시험 1 , A Priori 연관분석,
10	11/9, 11	7, 8, 9	k-Nearest Neighbors, Naïve Bayes, Decision trees
11	11/16, 18		Term Project Progress (*18 일 75+75 연속)
12	11/23, 25	10, 11	로지스틱회귀 revisited, Neural Networks
13	11/30, 12/2	11	Neural Networks, 앙상블
14	12/7, 9	별도 자료	조직 및 역할, Big Data, 시험 2
15	12/14, 16		Term Project Final Presentation (*16 일 75+75 연속)

12. 주의사항

강의 시간은 학습을 하고 아이디어를 교환하는, 학생과 교수 모두에게 집중력을 요하는 시간임. 특히, 50 명이 넘는 강의실에서는 타인의 학습 환경을 해치지 않기 위해 상호 존중과 배려가 필요함. 다음과 같은 행위는 타인의 학습 환경을 현격히 훼손하므로, 이를 방지하기 위해 해당 행동을 하는 학생은 최대 성적의 10% 포인트 (예, 85%=> 75%, 60%=>50%) 삭감에 해당하는 페널티를 받게 됨.

- 1) 노트북/패드 컴퓨터, 휴대폰 등의 전자 제품 사용 행위 (단, 실습 시간에 프로그래밍을 위한 노트북 컴퓨터 사용은 허용)
- 2) 수업에 늦게 도착하거나 일찍 나가는 행위
- 3) 수업 도중에 나가거나 들어오는 행위