강의계획서(Syllabus)

개설학과 : 정보미디어MBA	학년도 학기 : 2020년도 2 학기
학수번호 : IM 694 (한국어강의)	교과목명 (국문) 고급 비즈니스 애널리틱스 (영문) Advanced Business Analytics
강의-실험・실습-학점 : 3-**0-3**	담당교수명 : 이혜선 (Lee, Hyeseon)
연구실No :	전화번호 : 010-2807-3122 (hyelee@postech.ac.kr)

1. 강의목표

고급 비즈니스 애널리틱스는 통계 지식, 컴퓨팅 기술, 도메인기반 지식을 결합한 통합적 데이터과학 분석을 목적으로 한다. 데이터 애널리틱스를 위한 데이터시각화, 데이터탐색, 예측모델링을 kaggle, github의 공유데이터를 활용하고 데이터분석 기술을 습득한다. 불균형데이터의 샘플링, 변수간 인과구조, 시각화기술, 지도/비지도학습, 빅데이터의 특징추출 기법 및 자동화 기계학습을 포함한다. 미래데이터 플랫폼, 데이터과학의 윤리, 개인정보에 대한 인식과 더불어 인접학문 혹은 현업데이터 캡스톤 프로젝트 수행하면서 탐색, 분석, 모델링과정의 연구능력을 배양한다

2. 강의개요

현업 시장에서의 탐구문제를 찾고, 그 문제를 해결하기 위한 데이터를 수집하고, 분석하고 모델링하는 과정을 학습한다. 고급 비즈니스 애널리틱스 프로그래밍은 오픈소스 R을 활용한다 (Python 사용 가능).

- 데이터 추출 기술 (데이터추출, 데이터변환, 데이터정제)
- 데이터과학과 데이터시각화 (시각화분석, 다차원그래픽, 공간 데이터)
- 통계적 모델링
- 지도/비지도학습 기법, 예측모형의 평가기법
- 하이퍼파라미터 최적화 (자동화 기계학습)
- 데이터 특징기반 기술 (선형, 비선형, 인공지능기법)

2. 선행필수과목

기초통계, 데이터분석 관련 과목 (회귀분석, 데이터마이닝, 비즈니스 애널리틱스)

3. 성적

출석 (퀴즈포함)	10%
과제 (격주)	30%
중간고사 혹은 프로젝트	30%
기말 프로젝트	30%
Total	100

4. 교재

- Benjamin D. Baumer, Daniel T. Kaplan, and Nicholas J. Horton, Modern Data Science with R, 2017

[참고도서]

- Mastering machine learning with R, England Birmingham, 2017
- 전치혁, 데이터마이닝기법과 응용, 한나래출판사, 2012

5. 강의내용

주	내용	
1	Data Science, Business Analytics	Data Science, Open source, Github,
2	Advanced Business Analytics	Overfitting, Random sampling
3	Computing tools for R	Intro in R programming (basic script) Data import, export, merging
4	Data Wrangling	Data explanatory analysis using public data (kaggle data source)
5*	Data Visualization	R Graphics (histogram, scatterplot, boxplot)
6	Data preparation	missing, duplicated, near-zero variance, high correlation
7	Statistical Modeling	Sample statistics, Bootstrap, Outlier Statistical modeling : explaining variation
8	Midterm	
9	Supervised learning : Classifiers	Decision Tree, Support vector machine K-NN
10	Supervised learning : Bagging & Boosting	Random Forest Tuning parameters
11	Evaluating methods	Cross-validation Confusion matrix
12	Unsupervised learning : Clustering	Parallelizing hierarchical clustering
13	Feature learning	Feature extraction method Linear/nonlinear/deep learning
14	Unbalancing data	Synthetic Minority Over-sampling
15	Automation machine learning	Turning parameter
16	Final	Final project presentation

6. 기타사항

▶ 공유데이터, 오픈소스의 활용을 위해 R프로그래밍을 사용하고, 데이터는 kaggle의 경영경제 관련 데이터를 활용한다. Python사용자는 Python을 사용해도 됩니다.

▶ 강의내용의 분석기법에 대한 R프로그래밍 코드 제공

▶ 참고 : R프로그래밍 온라인 무료강좌 (https://pabi.smartlearn.io/) 참고

3기 오픈기간내 (8.24-10.18) 온라인으로 개별 수강할 수 있음

빅데이터분석과 R프로그래밍(기초I, 기초II), 머신러닝기법과 R 프로그래밍(I, II)

▶ 과제제출 : 늦은제출 (1일까지 10%, 2일까지 20%, 3일 이후 30% off)