

## Data Analysis Report

[ Due: 2019-06-22(토) 오후 11시 59분 ]

**주제: 평소 관심이 있는 분야 혹은 주제에 대해 학기 간 습득한 다음의 데이터 분석 알고리즘을 적용하여 보자!**  
Linear regression, Logistic regression, kNN, Naive Bayes, Association rules, Collaborative Filtering, Clustering

본 과제는 개별 프로젝트로 데이터 분석의 전 과정을 직접 주도하여 결론 도출에 이를 수 있는 능력을 배양하는데 주목적을 둔다. 데이터 분석의 과정은 다음과 같다.

[문제 정의] - [데이터 확보] - [데이터 전처리] - [알고리즘 적용] - [결과 해석 및 시각화]

■ 다음 사항을 읽고 Data Analysis Report를 작성하여 블랙보드에 제출하라.

- (1) Report 형식으로 제출하라. 문서 프로그램으로 작성(코드, 결과물, 시각화 자료 등은 캡처본 삽입)하여 제출하거나, .py 내 Markdown을 사용하여 작성 후 PDF 변환하여 제출 할 것을 추천한다.
- (2) 단순 캡처본 제출 혹은 코드만 단순하게 나열되어 있는 .py파일 제출은 0점 처리한다.
- (3) 적어도 3개 이상의 알고리즘을 사용하여야 한다.
- (4) Linear/Logistic regression 중 하나는 반드시 사용하여야 한다.
- (5) Supervised/Unsupervised Learning(PCA 포함)을 각각 적어도 하나씩 활용하여야 한다.
- (6) 데이터 분석의 목적인 **문제 정의**와 **결과 해석 및 활용방안**이 반드시 명확히 기재되어야 한다.
- (7) 각 코드 별로 목적 및 사유를 주석으로 기재해야 한다.
- (8) 데이터 셋은 여러 개를 사용하여도 무방하다.
- (9) 최대 점수는 10점이며 Collaborative Filtering을 사용할 경우 가산점(+1)을 부여한다.
- (10) 블랙보드에 기재 된 제출양식을 준수하라.

■ 데이터 확보에 참고 할 수 있는 URL은 다음과 같다.

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- U.S. Government's Open Data: <https://www.data.gov/>
- 공공데이터포털: <http://data.go.kr/>