

# R을 활용한 빅데이터 분석

## ▣ 교육 목표

- 데이터 분석에 대한 개념을 이해하고, 오픈소스 분석도구인 R을 활용하여 빅데이터 분석을 실습한다.

## ▣ 교육 대상

- R 초급자 (R 처음 접하신 분, R 사용 2년 미만인 분)
- 현업에서 데이터 분석업무 담당자

## ▣ 세부 교육내용

구 분	시 간	교 육 내 용
1일차 (데이터 전처리)	09:00 ~ 13:00 (4시간)	○ 오픈소스 분석 도구 R - R 분석의 필요성(특징, 장/단점) - R 설치 (R Base, R Studio) - R 기본 함수 실습 - 데이터 Read/Write/Save/Load
	14:00 ~ 18:00 (4시간)	○ 데이터 전처리 (data munging) - 데이터 핸들링 기초 - tidyverse package 실습 - filter / select / group_by / summarise / arrange / mutate / join 함수 등 - 데이터 전처리 전과정 실습
2일차 (시각화)	09:00 ~ 13:00 (4시간)	○ Visualization I - base 그래프 그리기 - base plot vs. ggplot2 - ggplot2 package 실습 - 시각화를 통한 통찰력 도출
	14:00 ~ 18:00 (4시간)	○ Visualization II (기타 유용한 그래프) - 변수별 상관관계 그래프 (correlation graph) - 센서 데이터 그래프 (scatter plot) - heat map / mosaic plot
3일차 (통계)	09:00 ~ 13:00 (4시간)	○ 기초 통계 - 개념/용어 이해 - 분포 (r, d, p, q 함수 이해) - t-test - 분산분석
	14:00 ~ 18:00 (4시간)	○ 회귀분석 - 정의 / 개념 / 활용방안 정리 - 회귀모형 모델링 (단순회귀 / 다항회귀 / 다중회귀) - 로지스틱 회귀 모델링
4일차 (머신러닝)	09:00 ~ 13:00 (4시간)	○ 기계학습 I - 회귀문제 (r-part, random forest, gbm, neural network) - 분류문제 (r-part, random forest, gbm, neural network)
	14:00 ~ 18:00 (4시간)	○ 기계학습 II - 군집분석 (H-clustering / K-means) - 연관성 분석 (association rules)