

■ Classification Problem

1. 온라인 사용자의 제품 구매 여부를 분류하는 문제
2. 각 row(행)은 사용자를 의미
 - Training 데이터 (10,000명의 사용자), Test 데이터(2,204명의 사용자)
3. 각 column(열)은 입력변수(X1~X12)와 출력변수(y)를 의미
 - X1: 평균 방문 페이지 수
 - X2: 사이트 평균 체류 시간
 - X3: 페이지에서 아무 작업도 안할 확률
 - X4: 특정 페이지에서 사이트를 종료할 확률
 - X5: 기념일과의 유사도 (발렌타인 데이 등)
 - X6: 페이지 방문 월 (1월, 2월 등)
 - X7: 운영체제 종류
 - X8: 브라우저 종류
 - X9: 사용자 거주지
 - X10: 사용자 트래픽 유형
 - X11: 사용자 유형 (재방문, 신규, 기타)
 - X12: 페이지 방문일의 주말 여부
 - y: 구매여부 (구매: 1, 비구매: 0)
4. 제출: 분석 보고서 (~6/6 17시)
 - PDF 파일 제출
 - 파일명: 이름_학번_데마과제
 - 분량: 5page 이내 (양식 X)
 - Test 데이터의 출력변수 6/3 공개
 - 보고서에 Test 분류 성능 (Accuracy, F1-score 등) 포함