연세-네이버클라우드 데이터사이언스 교육과정

**과목명**: 서포트벡터머신(SVM)

**데이터**: 폴란드 회사의 부도 여부 데이터로 5년 동안 관찰된 64개의 금융 특성변수를 바탕으로 이후 1년간 부도 발생 여부(1:여, 0:부)를 나타냄. 부도가 발생한 기업의 경우 2000~2012년의 데이터를 수집하고, 부도가 발생하지 않은 기업의 경우는 2007~2013년의 데이터를 수집함

**목적**: 금융 특성변수를 바탕으로 기업의 부도 여부를 예측

1. 해당 데이터를 불러오시오.
2. 범주변수의 비대칭 문제를 확인하고, SMOTE를 이용하여 이를 해결하시오.
3. 훈련용 데이터와 테스트용 데이터을 7:3으로 분리하시오.
4. 표준화 전처리 및 SVM 모형(linear, sigmoid, rbf, poly 커널) 학습 과정을 하나의 단일 과정인 파이프라인으로 설정하시오.
5. rbf 커널 SVM의 모수 C값을 10-겹 교차검증을 이용해 최적화하고, 검증곡선을 시각화하시오.
6. 그리드 서치를 이용하여 최적의 SVM 모형을 구하고, 훈련용 데이터에 대해 최적화된 모형을 학습한 뒤 테스트용 데이터에 대해 정확도를 구하시오.