

2.1 아래의 문제를 해결하기 위해 데이터마이닝 기법을 이용하려고 한다. 지도학습과 비지도 학습 중 어느 기업을 사용하는 것이 보다 적절한지 구분하시오.

- a. 과거 소비자들의 인구통계학적, 금융 데이터를 바탕으로 새로운 소비자들의 대출 여부를 결정 -> 교사학습
- b. 온라인 서점에서 기존 구매 고객들의 구매 패턴을 파악하여 책을 추천 -> 비교사학습
- c. 기존에 알고 있는 네트워크 데이터 패킷 상태 정보를 이용하여 새로운 네트워크 데이터 패킷의 상태를 분류(예: 바이러스, 해커 공격) -> 교사학습
- d. 비슷한 성향을 가지고 있는 고객군 파악 -> 비교사학습
- e. 파산한 회사와 그렇지 않은 회사들의 재정상태 데이터를 기반으로 특정 회사의 파산 여부를 예측 -> 교사학습
- f. 수리이력 정보를 바탕으로 항공기 수리 시간을 추정 ->교사학습
- g. 우편번호 스캔 결과 정보를 바탕으로 우편물 자동 분류
-> 비교사학습(지역별 분류)/교사학습(우편번호 인식훈련)
- h. 보인의 구매물품과 다른 고객들의 과거 구매 이력 정보를 바탕으로 할인 쿠폰의 발행 여부 판단
-> 교사학습(쿠폰을 발행 여부)/ 비교사 학습(구매 패턴에 따라 발행할 쿠폰 선택)

2.2 검증데이터와 평가데이터 그리고 훈련데이터의 차이를 서술하시오.

학습데이터 : 모델 구축 시 사용되는 데이터를 의미한다. 일반적으로 여러 모델을 비교할 경우 동일한 학습데이터를 사용한다.

검증데이터 : 학습데이터로부터 구축된 여러 모델의 성능을 비교할 때 사용된다.

평가데이터 : 평가데이터는 구축된 모델이 미래에 수집될 새로운 데이터에 대해 얼마만큼 예측/분류 성능을 보일지 평가할 때 쓰이는 데이터이다.