

캡스톤 중간발표 자료조사

배경 및 목적

THEELEC

≡ 전체 최신뉴스 반도체 디스플레이 배터리 통신·부품 완성품 자동차

HOME > 중국산업동향

TSMC, 감광액 불량 사고로 6200억원 상당 웨이퍼 폐기

이종준 기자 | 승인 2019.02.18 15:39 | 댓글 0

<https://www.thelec.kr/news/articleView.html?idxno=816> : 기사

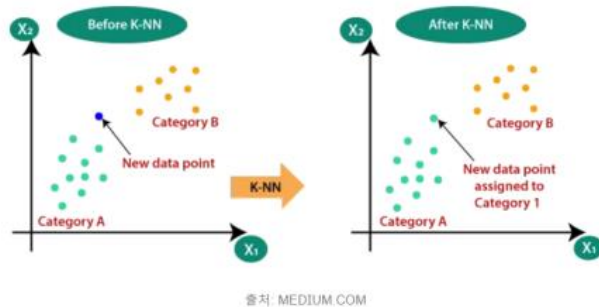
- TSMC : 감광액 이물질로 인한 불량 → 6200억원 상당 웨이퍼 폐기 (2019년 기준 웨이퍼 당 가격 약 69만원)

여러 형태의 공정들을 통해 웨이퍼 상에 결함들이 생길 수 있으며, 이런 공정 결함들은 웨이퍼의 수율을 저하시키는 요인으로 작용됨 → 반도체 제품의 성능 저하로 직결, 웨이퍼의 결함은 특정 영역에서 작동하지 않거나 전체 제품의 수명을 단축시킬 수 있음.

기대효과

불량 웨이퍼를 머신러닝을 통해 걸러냄으로 금전적인 손해 대폭 감소 예상

KNN Imputer



거리기반 분류분석 모델, 머신러닝에서 데이터를 가장 가까운 유사 속성에 따라 분류하여 데이터를 분류하는 기법

K값 정하는 방법 : 교차 검증 등의 방법 사용, 데이터셋이 노이즈가 많은 경우 작은 K값 선택, 데이터셋이 복잡한 패턴 or 클래스 간 경계가 모호한 경우 큰 K값 선택

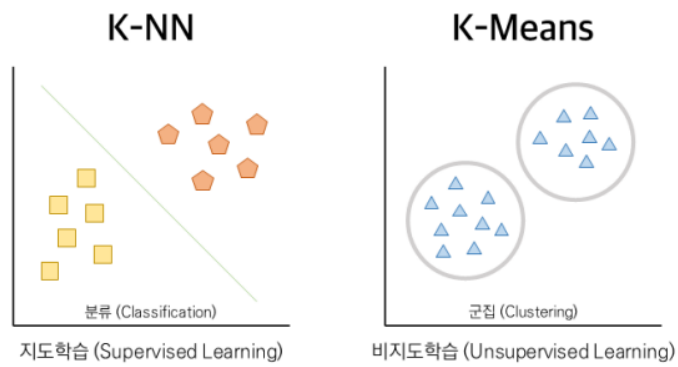
장점

- 1) 간단하고 이해하기 쉬움
- 2) 학습 데이터 분포 고려 X
- 3) 높은 분류 정확도
- 4) 적은 데이터셋에서도 잘 작동
- 5) 다목적 사용 가능

단점

- 1) 계산 복잡도 높음
- 2) 이상치에 민감
- 3) 클래스 불균형

Kmeans



KNN : 미리 레이블링 붙어 있는 데이터들을 학습 후 새로운 데이터에 대해 분류

K-Means : 레이블을 모르더라도 비슷한 특징을 가진 데이터끼리 묶어주는 군집 수행

KMeans 프로세스

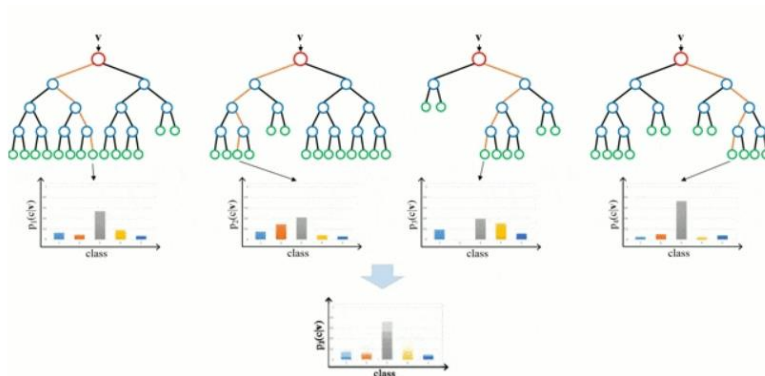
- 1) 데이터셋에 K개의 중심을 임의로 지정
- 2) 각 데이터들을 가장 가까운 중심이 속한 그룹에 할당
- 3) 분산 계산 후 각 클러스트의 새 중심 배치
- 4) 중심이 더 이상 변하지 않을 때까지 2,3번 반복

앙상블 방법이란?

강력한 하나의 모델을 사용하는 대신 보다 약한 여러 개의 모델을 조합하여 더 정확한 예측을 해주는 방법

대표적으로 랜덤포레스트, XGboost, Light Boost 등이 있음

Random Forest



결정나무를 여러 개 만들어서 그 결과들을 종합적으로 고려하여 결론을 도출하는 앙상블 방법

집단지성을 통해 결론을 도출할 때 한쪽에 치우치지 않은 더 좋은 결론을 도출해낼 수 있는 것과 비슷한 원리

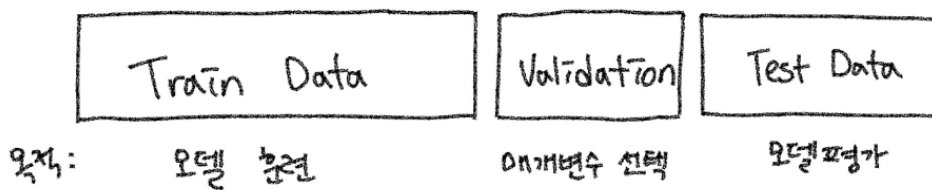
장점

- 1) 분류, 회귀문제에 모두 사용 가능
- 2) 결측치를 다루기 쉬움
- 3) 대용량데이터 처리가 쉬움
- 4) 과적합문제 해결
- 5) 특성중요도를 구할 수 있음

XGboost 특징

- 1) 과적합방지가 가능한 규제 포함
- 2) Classification And Regression Tree를 기반 → 분류와 회귀 둘 다 가능
- 3) 조기 종료 제공
- 4) 앙상블 부스팅의 특징인 가중치 부여를 경사하강법으로 함.

CV Gridsearch



관심있는 매개변수들을 대상으로 가능한 모든 조합을 시도해 보며 최적 하이퍼 파라미터 튜닝을 하는 것

일반화 성능을 더 잘 평가하기 위해 훈련 세트와 검증 세트를 한번만 나누지 않고, 교차 검증을 사용해서 각 매개변수 조합의 성능 평가 가능 → 교차 검증과 그리드 서치를 한번에 해주는 모델이 GridSearch CV