|  |  |  |  |
| --- | --- | --- | --- |
| **팀명** | AI제조 팀 | **이름** | 남도형 |
| **일시** | 7/31 | **장소** | 온라인 |
| **프로젝트명** | 머신러닝을 활용한 스마트폰 불량품탐지 | | |
| **개별활동**  **내용** | Class imbalance(클래스 불균형)에 대해서 학습했다. 이는 지도 학습에서 발생하는 한 가지 문제이다. 학습 데이터셋의 각 클래스에 속하는 데이터 포인트들의 분포가 불균형하게 나타나는 상황을 의미한다. 즉, 하나 이상의 클래스가 다른 클래스들보다 훨씬 더 많은 데이터를 가지고 있거나 그 반대인 경우를 말한다. 이러한 클래스 불균형은 다양한 문제를 야기할 수 있다.    극심한 class imbalance가 존재하는 경우, 대부분의 classification 모델은 다수 클래스를 잘 분류해내도록 decision boundary가 설정된다. 또 위 그림과 같이 accuracy는 높지만, recall / f1-score는 작아질 수 밖에 없다.    Classification model은 특정 관측치가 어떤 클래스에 속할지에 대한 확률을 계산한다. 확률이 높은 클래스로 해당 관측치의 클래스를 예측하게 된다.  이를 해결하기 위해 리샘플링, 클래스 가중치, 합성 샘플링, 다양한 모델을 적용하는 방법이 있다.  우선 리샘플링을 통해 소수 클래스의 데이터를 증가시켜 데이터셋의 클래스 분포를 균형있게 만들 수 있다. 오버샘플링은 소수 클래스의 데이터를 복제하거나 유사한 데이터를 생성하여 클래스를 균형있게 만든다.  다음으로 클래스 가중치는 모델 학습 시 소수 클래스에 높은 가중치를 부여하여 소수 클래스의 오분류를 패널티로 간주하도록 할 수 있다. | | |
| **향후추진 계획** | 응용 프로젝트를 위해 다양한 모델을 학습할 예정이다. | | |