|  |  |  |  |
| --- | --- | --- | --- |
| **팀명** | AI제조 팀 | **이름** | 남도형 |
| **일시** | 7/29 | **장소** | 온라인 |
| **프로젝트명** | 머신러닝을 활용한 스마트폰 불량품탐지 | | |
| **개별활동**  **내용** | "SMOTE"는 "Synthetic Minority Over-sampling Technique"의 약어로, 불균형한 데이터 셋에서 주로 사용되는 샘플링 기법이다. 불균형한 데이터셋은 하나의 클래스가 다른 클래스들보다 훨씬 더 많은 샘플을 가지고 있을 때 발생한다. 이런 경우, 학습 알고리즘은 흔히 소수 클래스의 패턴을 올바르게 학습하지 못할 수 있다.  SMOTE는 소수 클래스의 샘플을 합성하여 데이터셋에 다양성을 더하는 방식으로 작동한다. 이를 통해 소수 클래스의 샘플 수를 증가시켜, 다수 클래스와 소수 클래스 사이의 균형을 조정하는 데 도움을 줍니다. SMOTE의 주요 아이디어는 소수 클래스의 샘플들 사이에 새로운 가상의 샘플들을 만들어내는 것인데, 이를 “합성”이라고 한다.    이론을 학습 후 data를 가지고 SMOTE를 적용하여 데이터 오버샘플링을 진행했다.  SMOTE 알고리즘은 다음과 같은 방식으로 작동한다.   1. 소수 클래스에서 임의의 샘플을 선택한다. 2. 선택된 샘플과 그 주변의 샘플들을 이용하여 새로운 합성 샘플을 생성한다.  * 샘플들을 벡터 공간상에서 연결하고, 두 샘플 사이의 거리를 이용한다. * 임의의 가중치를 사용하여 새로운 합성 샘플의 위치를 결정한다.  1. 새로운 합성 샘플을 데이터셋에 추가한다. 2. 위의 과정을 원하는 샘플 수에 도달할 때까지 반복한다.   SMOTE를 사용하면 소수 클래스의 샘플 수가 증가하므로, 학습 알고리즘이 소수 클래스를 더 잘 이해하고 학습할 수 있다. 이는 모델의 성능을 향상시키고, 불균형한 데이터셋에서 더 효과적인 결과를 얻을 수 있게 해준다.  주의할 점은, 과도하게 합성 샘플을 생성하여 다수 클래스와 소수 클래스 간의 잘못된 관계를 만들지 않도록 하는 것이 중요하다. | | |
| **향후추진 계획** | 응용 프로젝트를 위해 다양한 모델을 학습할 예정이다. | | |