|  |  |  |  |
| --- | --- | --- | --- |
| **팀명** | AI제조 팀 | **이름** | 남도형 |
| **일시** | 8/24 | **장소** | 온라인 |
| **프로젝트명** | 머신러닝을 활용한 스마트폰 불량품탐지 | | |
| **개별활동**  **내용** | 이상치를 얼마나 잘 탐지하는 지와 정상 데이터를 얼마나 잘 분류하는지를 시각화했다. 위 곡선은 ROC곡선이고 이진 분류 모델의 성능을 평가하는 데 사용된다. 해당 모델의 ROC 곡선 아래 면적이 0.47로, 모델의 이상치 탐지 능력이 무작위 수준보다는 높지만 아직 개선할 필요가 있었다.    Isolation Forest는 tree기반으로 구현되며 랜덤으로 데이터를 split하여 모든 관측치를 고립시키며 구현된다. Isolation Forest의 컨셉은 ＂각 관측치를 고립시키는 것은 이상치가 정상 데이터보다 쉽다”이다. 첫 번째 그림은 정상 데이터를 분리하는 경우이다. 컨셉에서와 같이 정상 데이터를 분리하기 위해서는 7번의 split이 필요하다 하지만 오른쪽 그림인 이상치를 분리하기 위해서는 약 3번의 split만이 필요하다. 이처럼 Isolation Forest는 랜덤포레스트가 의사결정나무를 여러 번 반복하여 앙상블 하듯이, iTree를 여러 번 반복하여 앙상블을 하는 방법으로 학습한다. | | |
| **향후추진 계획** | Iforest 모델으로 나온 결과를 시각화하여 결과보고서를 작성할 예정이다. | | |