

네 다음으로는 저희가 해당 task에 대해 고려한 부분을 설명 드리겠습니다.

먼저 저희 고객님들은 본인의 피부 상태가 좋은지 나쁜지를 확실히 알고 싶은 욕구가 강하다고 생각했습니다. 따라서 좋은 피부를 안 좋다고 평가하는 것에 비해 안 좋은 것을 좋게 평가하는 것이 문제가 더 크다고 생각했습니다.

하지만 주어진 데이터에서는 안 좋은 라벨에 대한 데이터가 부족했는데요.

위 두 가지 사항을 고려했을 때, 안 좋은 피부와 좋은 피부를 잘 구분하는 것이 중요하지만, 안 좋은 라벨에 대한 데이터가 부족하기 때문에 모델 평가 시 모델이 해당 task를 잘 수행하고 있는지 확인할 수 있는 정량적인 Metrix가 필요하다고 생각했습니다.

저희가 사용한 Metrix을 설명하기 앞서 Confusion Matrix에 대해 짚고 넘어가도록 하겠습니다. Confusion Matrix는 학습된 모델의 성능을 측정하기 위해 예측값과 실제 라벨을 비교하기 위한 표입니다. (이후 TP, TN, FP, FN 설명)

이제 저희가 사용한 Merix에 대해 설명해드리겠습니다.

저희는 먼저 Accuracy를 평가지표로 사용해 봤는데요. 모델이 전체적인 결과를 잘 예측하는지 확인하는 것이 중요하다고 생각했기 때문에 Accuracy를 평가지표로 사용했습니다. 하지만 Accuracy는 실제 라벨이 한 쪽에 치중되어 있고, 모델이 해당 라벨만 예측할 경우 다른 라벨에 대한 예측은 잘 수행하지 못하지만 Accuracy 점수는 높아지는 문제점을 가집니다.

이러한 Accuracy의 문제점을 해결하기 위해서 Recall을 사용해봤습니다. 단, 저희 task가 multi class대한 classification task이기 때문에 단순히 해당 수식을 적용할 수 없었는데요. 따라서 각 label에 대한 Recall값을 구하고, 해당 Recall값들을 평균 내어주는 방식을 사용했습니다.

이때 저희는 한 가지 의문점이 들었는데요. 만약 해당 이미지의 정답이 0이라면 모델이 1로 예측한 경우랑 4로 예측한 경우는 엄연히 다르다고 생각했습니다. 따라서 인접한 라벨에 대해서는 약간의 가중치를 부여하고 거리가 먼 라벨에 대해서 패널티를 부가하는 Penalty-Recall을 최종 Metric으로 삼았습니다. 이후 모델 실험 및 결과에 대한 부분은 다른 분이 이어서 설명해 주시겠습니다.