

# Deepfake Detection Model Evaluation Report

# 딥페이크 탐지 모델 성능 분석 보고서

## ## (1) 모델 개요

본 보고서는 딥페이크 탐지 모델의 성능을 평가하기 위해 수집된 데이터를 기반으로 작성되었습니다. 모델은 가짜(Fake)와 진짜(Real) 콘텐츠를 분류하는 데 초점을 맞추고 있으며, 정확도, 정밀도(precision), 재현율(recall), F1 점수(f1-score) 등의 지표를 통해 성능을 측정합니다.

## ## (2) 성능 요약

모델의 전체 정확도는 89.70%로, 이는 모델이 전체 데이터에서 89.70%의 정확도로 올바른 예측을 했음을 의미합니다. 클래스별 성능은 다음과 같습니다:

### - \*\*Fake\*\*:

- 정밀도: 93.36%
- 재현율: 89.03%
- F1 점수: 91.14%
- 지원 수: 237

### - \*\*Real\*\*:

- 정밀도: 84.88%
- 재현율: 90.68%
- F1 점수: 87.69%
- 지원 수: 161

### - \*\*매크로 평균\*\*:

- 정밀도: 89.12%
- 재현율: 89.86%
- F1 점수: 89.42%

### - \*\*가중 평균\*\*:

- 정밀도: 89.93%

- 재현율: 89.70%
- F1 점수: 89.75%

### ## (3) 분석 및 해석

모델은 가짜 콘텐츠를 탐지하는 데 있어 높은 정밀도(93.36%)를 보이고 있으며, 이는 모델이 가짜 콘텐츠를 올바르게 식별하는 데 매우 효과적임을 나타냅니다. 그러나 재현율(89.03%)은 상대적으로 낮아, 일부 가짜 콘텐츠가 잘못 분류될 가능성이 있음을 시사합니다. 반면, 진짜 콘텐츠에 대한 정밀도(84.88%)는 낮은 편이며, 이는 진짜 콘텐츠를 가짜로 잘못 분류할 위험이 있음을 의미합니다. 전반적으로, 모델의 성능은 균형 잡힌 편이나, 특정 클래스에서의 성능 차이가 존재합니다.

### ## (4) 개선 방향

모델의 성능을 향상시키기 위해 다음과 같은 개선 방향을 제안합니다:

1. \*\*데이터 증강\*\*: 다양한 유형의 딥페이크와 진짜 콘텐츠를 포함한 데이터셋을 확장하여 모델의 학습 범위를 넓힐 필요가 있습니다.
2. \*\*모델 아키텍처 개선\*\*: 최신 딥러닝 기술을 적용하여 모델의 구조를 개선하고, 더 깊고 복잡한 네트워크를 통해 성능을 향상시킬 수 있습니다.
3. \*\*하이퍼파라미터 튜닝\*\*: 모델의 하이퍼파라미터를 최적화하여 성능을 극대화할 수 있습니다.
4. \*\*클래스 불균형 해결\*\*: 클래스 간의 불균형 문제를 해결하기 위해 오버샘플링 또는 언더샘플링 기법을 사용할 수 있습니다.

### ## (5) 결론 요약

본 딥페이크 탐지 모델은 89.70%의 높은 정확도를 기록하며, 가짜 콘텐츠 탐지에서 우수한 성능을 보이고 있습니다. 그러나 진짜 콘텐츠에 대한 정밀도가 상대적으로 낮아, 향후 개선이 필요합니다. 제안된 개선 방향을 통해 모델의 성능을 더욱 향상시킬 수 있을 것으로 기대됩니다. 이러한 발전은 딥페이크 탐지의 신뢰성을 높이고, 다양한 분야에서의 활용 가능성을 확대하는 데 기여할 것입니다.

