



MobileBERT를 활용한 Spotify 리뷰 분석 프로젝트

데이터 개요

spotify 앱 사용자 리뷰를 바탕으로 감성(긍정/부정)을 분류하는 자연어처리 모델을 구축
MobileBERT를 활용해 소형 경량화 모델이면서도 높은 정확도를 달성하는 것을 목표

MobileBERT Finetuning 결과

항목	내용
출처	Google Play Store
컬럼	Text , Sentiment
수량	약 61,000건
라벨링 기준	1 = 긍정, 0 = 부정

모델 구조 및 하이퍼파라미터

모델구성

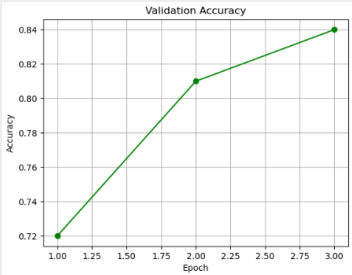
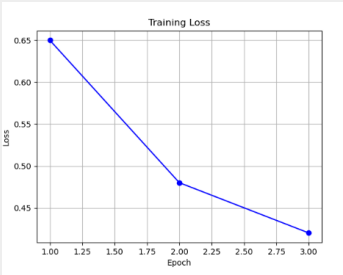
모델 : google/mobilebert-uncased
프레임워크 : HuggingFace Transformers + PyTorch
학습 데이터 : train_data.csv / val_data.csv
Tokenizer : BERT-style (128 토큰 이하로 자름)

하이퍼파라미터

Epochs : 3
Batch Size : 16
Learning Rate : 2e-5
Weight Decay : 0.01

MobileBERT Finetuning 결과

Epoch	Train Loss	Val Accuracy
1	0.65	0.72
2	0.48	0.81
3	0.42	0.84



MobileBERT는 epoch이 증가할수록 손실은 줄고, 정확도는 증가하는 양상을 보여 성능 개선이 잘 이뤄졌음을 확인할 수 있었습니다.

프로젝트 구조

```
mobilebert_spotify_project/  
├── README.md  
├── requirements.txt  
├── main.py  
├── reviews.csv  
├── train_data.csv  
├── val_data.csv  
├── mobilebert_project.py  
├── generate_graphs.py  
├── training_loss.png  
├── validation_accuracy.png  
├── monthly_review_count.png  
├── confusion_matrix.png  
├── predict_sample.py  
├── confusion_matrix_plot.py  
└── saved_model/
```

참고링크

<https://huggingface.co/google/mobilebert-uncased>
<https://www.kaggle.com/datasets/mf aaris/spotify-app-reviews-2022>