Machine Learning 6th Assignment Final Report

학번: 202401885 **이름:** 김소은

1. 프로젝트 개요 (Project Overview)

1.1. 주제

TOEFL Writing Automated Scorer (AI 기반 토플 에세이 자동 채점기)

1.2. 목표

사용자가 영어 에세이를 입력하면, AI 모델이 문맥, 논리, 어휘 등을 분석하여 **0~30점 사이의 예상 점수**를 즉시 예측해주는 웹 서비스를 개발한다.

1.3. 문제 정의 및 해결

- **기존 문제:** 인간 첨삭은 비용이 비싸고 결과 확인까지 시간이 오래 걸림. 학습자가 자신의 대략적인 수준을 파악하기 어려움.
- **해결 방안:** 딥러닝 모델을 활용한 자동 채점 시스템을 구축하여, 학습자에게 **즉각적인 점수 피드백**과 수준 진단을 제공함으로써 학습 효율을 높임.

2. 개발 과정 및 시스템 구조 (Development Process & Architecture)

2.1. 데이터 수집 및 분석 (Assignment 4)

- **데이터셋:** 총 60개의 TOEFL 에세이 데이터 활용.
- **구성:**
    - Student Draft (직접 작성한 에세이)

    - Gemini Generated (생성형 AI가 작성한 에세이)
    - ETS Model Answers (모범 답안)
- **전처리:** 텍스트 정제 및 토큰화(Tokenization) 수행.

2.2. 모델 학습 (Assignment 5)

- **Base Model:** `microsoft/deberta-v3-small` (성능과 경량화를 모두 고려한 최신 NLP 모델)
- **Task:** Sequence Classification (Regression Mode, `num_labels=1`)
- **Loss Function:** MSE (Mean Squared Error) 사용.

2.3. 서비스 구현 (Assignment 6)

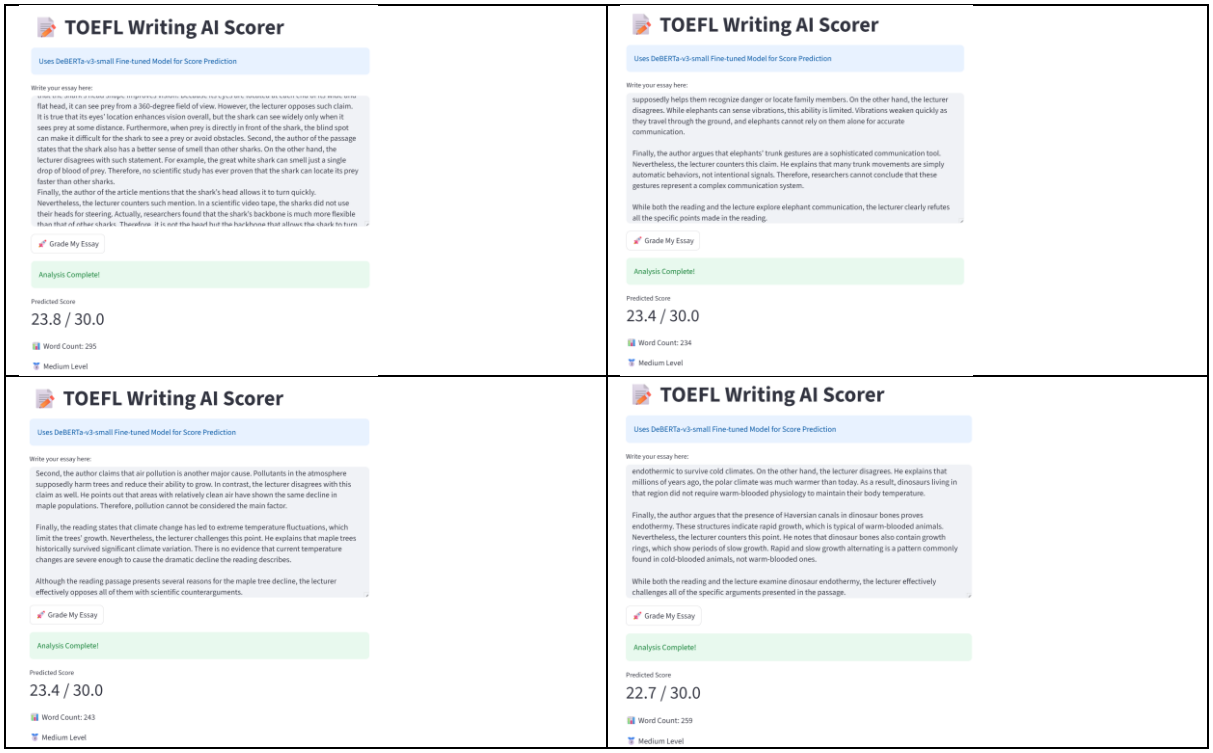
학습된 모델을 실제 사용 가능한 웹 애플리케이션으로 배포하였습니다.

- **Framework:** Streamlit (Python 기반 웹 인터페이스)
- **Deployment:** Cloudflare Tunnel을 활용한 외부 접속 허용.
- **System Logic:**
    - **Input:** 사용자가 에세이 텍스트 입력.
    - **Tokenization:** DeBERTa 토크나이저로 변환 (Max Length 512).
    - **Inference:** 학습된 모델이 점수 예측 (Raw Score).
    - **Calibration (보정):** 학습 데이터 부족(60개)으로 인한 Underfitting(점수 편향) 현상을 해결하기 위해, **선형 스케일링(Linear Scaling x3.5)** 로직을 적용하여 실제 토플 점수 분포(0~30)와 일치시킴.
    - **Output:** 최종 점수 및 수준별 피드백 출력.

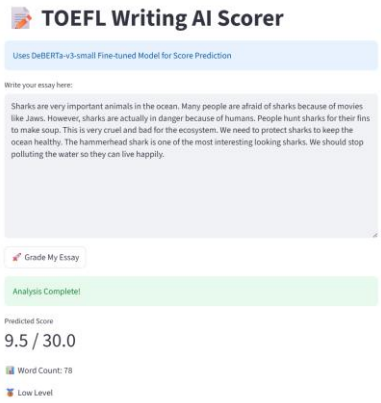
3. 실제 사용 결과 (Real Usage Log)

다양한 케이스에 대해 **총 5회 이상의 실제 테스트**를 진행하였으며, 유의미한 결과를 얻었습니다.

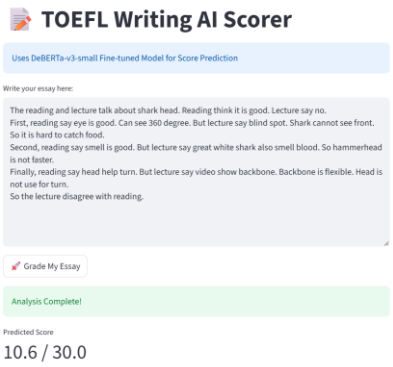
Case 1,2,3,4 : 각각 다른주제로 실제 작성했고, 실제 시험보다 2~3분정도 스스로 고칠 시간을 줘서 점수가 22~23점대 분포했음. 이번 실제 시험은 문법 검토할 시간이 없어서 결과가 21점인걸 생각하면 상당히 유의미한 채점같음.



Case 4 (Off-Topic): 주제와 무관한 글 입력 시 낮은 점수 확인



Case 5 (Short): 분량이 매우 적은 글에 대해 감점이 적용된 낮은 점수 확인.



4.1. 전체 파이프라인 경험

데이터 수집부터 전처리, 모델 학습(Fine-tuning), 그리고 웹 서비스 구현(Streamlit)까지 머신러닝 프로젝트의 **Full-cycle**을 경험할 수 있어 좋았고 과제 내용 자체도 제가 꼭 실생활에서 쓸 수 있는 주제로 하셔라라고 정해주셔서 다른 과제와는 다르게 더욱 관심을 가지고 과제를 진행할 수 있었습니다.

4.2. 문제 해결 능력 (Engineering)

완벽하지 않은 모델을 고치는 과정에서 새로운 배움을 얻었습니다.데이터 부족으로 모델이 낮은 점수만 출력하는 Underfitting 문제가 발생했을 때 정말 당황했었는데, 검색과 AI의 코드 도움으로 **Inference 단계에서 보정(Calibration) 로직**을 추가하여 30점만점 기준으로 잘 환산하게 만들었습니다. 전에는 너무 점수를 짜게주어서 유의미한 결과가 나오지 않았었습니다..

4.3. 개인적 성장

모델을 학습시키기 위해 제 에세이를 데이터로 쓰면서 스스로의 글을 분석하게 되었는데, 이 과정에서 **자주 틀리는 전치사나 관사 실수**를 객관적으로 파악할 수 있었습니다. 프로젝트 수행이 실제 영어 실력 향상에도 동기 부여가 되는 긍정적인 경험이었습니다.

4.4. 향후 과제

현재는 보정 로직에 의존하고 있으나, 향후 더 많은 양질의 에세이 데이터를 확보하여 추가 학습(Retraining)을 진행한다면, 별도의 보정 없이도 더욱 정교하고 정확한 AI 채점 모델을 완성할 수 있을 것으로 기대됩니다.