

## 결과 보고서

### [1] 데이터 분석

데이터 분석에 사용할 데이터는 중고등학교 2학년의 데이터만을 필요로 하므로 전체 데이터에서 중  
고등학교 2학년의 데이터만 추출하여 사용하였습니다. 또한, 평가 지표에 따른 최종 점수는 데이터  
에 포함되어 있는 점수가 사용하지 않고, 가중평균을 사용하여 다시 계산하였습니다. 데이터 처리의  
편의성을 위해 정수형 표현으로 바꾸어서 데이터를 처리하였습니다.

변수명	의미
Student_grade_group	1 : 중등 , 2 : 고등
Student_educated	0 : False, 1 : True
Essay_type	1 : 수필형, 2 : 논술형

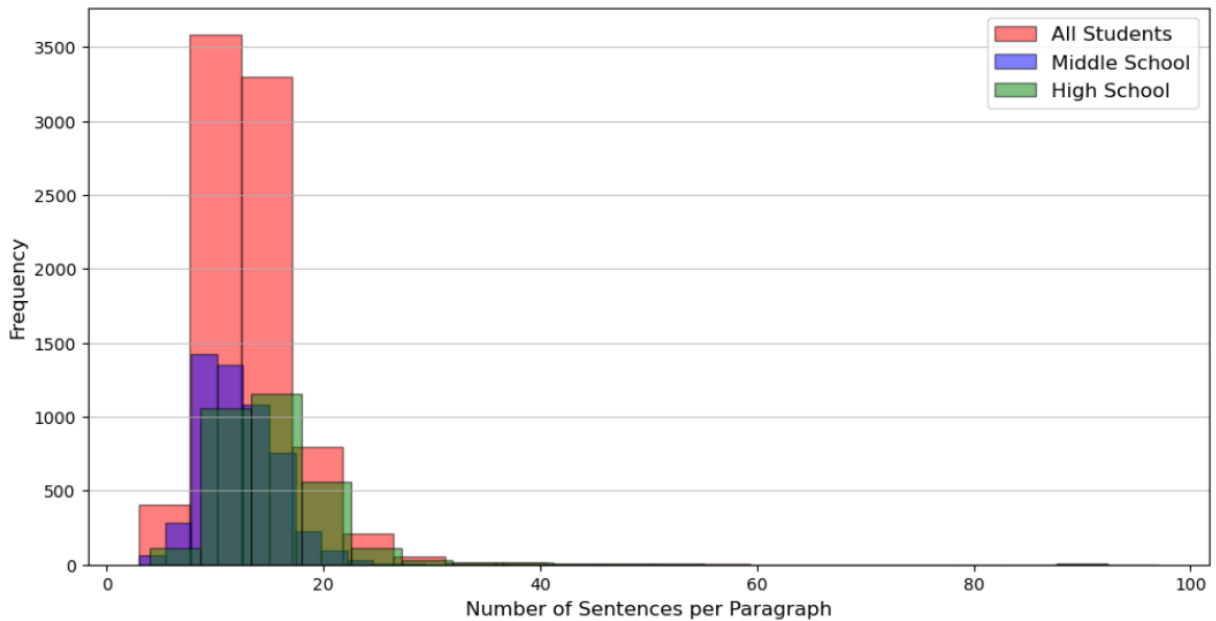
일반적으로 문장의 구분은 마침표를 이용하여 구분하지만, 사용하는 데이터에는 '#@문장구분#' 로  
문장을 나누어서 표시하였으므로, 해당 구분자를 이용하여 문장의 개수를 파악하였습니다.

점수는 크게 'Expression(표현, 이하 Exp)', 'Organization(구조, 이하 Org)', 'Contents(내용, 이하 Con)'으  
로 나눌 수 있으며 각각의 점수의 세부점수의 구성은 아래와 같은 순서를 의미합니다.

표현	의미
Score_exp_1	문법의 정확성
Score_exp_2	단어 사용의 적절성
Score_exp_3	문장 표현의 적절성
Score_org_1	문단 간 구조의 적절성
Score_org_2	문단 내 구조의 적절성
Score_org_3	구조의 일관성
Score_org_4	분량의 적절성
Score_con_1	주제의 명료성
Score_con_2	설명의 구체성
Score_con_3	프롬프트 독해력
Score_con_4	사고의 창의성

#### 1.1 중고등학생별 문장 개수의 분포

중고등학생이 작성한 문장 갯수의 분포입니다. 전체적으로 13개 정도의 문장을 작성하였습니다.  
중학생 보다는 고등학생이 문장을 더 많이 작성한 것으로 나타났습니다.



조금 더 구체적으로 살펴보기 위해 Type을 나누어서 살펴보면 다음과 같습니다.

		mean	
student_group		1	2
type			
1	1	11.853555	14.862069
	2	12.345168	15.869091

‘수필형’ 글보다는 ‘논술형’ 글에서 문장 사용이 더 많음을 볼 수 있으며, 이전 그래프와 마찬가지로, 두 가지 유형에서 모두 고등학생이 작성한 문장의 수가 더 많음을 확인할 수 있습니다.

## 1.2 그룹과 글쓰기 유형의 관계

각 그룹의 점수가 문장 길이와 어떤 관계가 있는지 살펴보기 위해 pandas의 pivot table을 이용하여 표로 나타내었습니다. 문장의 길이는 해당 그룹에 속한 학생들이 작성한 문장 길이의 평균을 사용하였습니다.

type	student_group	1	2	1	2	type	student_group	1	2	1	2
score_exp_1	1	2	1	2	score_con_1	1	2	1	2		
1	15.400000	7.200000	22.000000	16.000000	0	11.500000	32.000000	NaN	NaN		
2	11.970817	14.685654	12.401451	15.807980	1	12.090909	14.263158	8.000000	13.000000		
3	11.752903	15.041116	12.315651	15.912186	2	11.933333	14.453012	11.967480	17.441441		
type	1	2			3	11.827551	15.110482	12.420986	15.543396		
student_group	1	2	1	2	type	1	2	1	2		
score_exp_2	1	2	1	2	score_con_2	1	2	1	2		
1	13.277778	26.000000	12.772727	NaN	0	NaN	NaN	12.345168	NaN		
2	12.206696	14.144000	12.359797	15.042781	1	12.588235	16.222222	NaN	12.500000		
3	11.711099	15.047305	12.336638	16.068343	2	11.905840	14.578431	NaN	15.821721		
type	1	2			3	11.821016	15.001346	NaN	15.889895		
student_group	1	2	1	2	type	1	2	1	2		
score_exp_3	1	2	1	2	score_con_3	1	2	1	2		
1	11.847826	13.473684	12.212121	13.000000	0	11.853555	14.862069	NaN	NaN		
2	11.963216	14.440454	12.442167	15.552017	1	NaN	NaN	8.900000	15.600000		
3	11.757347	15.269608	12.314484	15.984116	2	NaN	NaN	12.000000	15.198324		
type	1	2			3	NaN	NaN	12.514857	16.023688		
student_group	1	2	1	2	type	1	2	1	2		
score_org_1	1	2	1	2	score_con_4	1	2	1	2		
1	10.173913	13.909091	10.400000	14.000000	0	6.272727	7.000000	5.720000	9.428571		
2	11.966292	14.444251	12.600470	15.671545	1	8.750000	7.000000	9.677966	10.100000		
3	11.798931	15.297391	12.244373	15.967841	2	10.663580	11.500000	10.152249	12.650000		
type	1	2			3	12.129941	14.954104	12.750742	16.024758		
student_group	1	2	1	2							
score_org_2	1	2	1	2							
0	NaN	NaN	12.345168	NaN							
1	12.588235	16.222222	NaN	12.500000							
2	11.905840	14.578431	NaN	15.821721							
3	11.821016	15.001346	NaN	15.889895							
type	1	2									
student_group	1	2	1	2							
score_org_3	1	2	1	2							
1	11.647059	15.000000	10.000000	18.250000							
2	11.960396	14.430556	12.303779	15.641473							
3	11.797232	15.118949	12.373031	15.945907							
type	1	2									
student_group	1	2	1	2							
score_org_4	1	2	1	2							
0	6.272727	7.000000	5.720000	9.428571							
1	8.750000	7.000000	9.677966	10.100000							
2	10.663580	11.500000	10.152249	12.650000							
3	12.129941	14.954104	12.750742	16.024758							

Score\_org\_1과 score\_org\_4 그리고 score\_con\_4에서는 문장의 개수가 많을수록 더욱 좋은 점수를 받았음을 확인할 수 있습니다. 특히 문장의 개수와 직접적으로 관련이 있는 score\_org\_4에서 가장 두드러지게 표현되었습니다.

Score\_exp\_3와 Score\_con\_2에서 고등학생이 작성한 문장의 개수가 많을수록 좋은 점수를 받았음을 확인할 수 있었습니다. 문장을 많이 사용할수록, 자신의 생각을 더 자세하게 표현할 수 있다는 점이 작용한 것 같습니다. 반대로 중학생의 경우에는 그러한 효과가 미비한 것을 확인할 수 있습니다.

### 1.3 그룹 별 글쓰기 유형별 최종 점수

	중학생	고등학생
수필형	26.560042	26.066112
논술형	26.909883	27.094961

중학생과 고등학생의 최종 점수를 글쓰기 타입에 따라 나타내었습니다. 가장 점수가 높은 그룹은 고등학생이 작성한 논술형 글이 가장 점수가 높았으며, 반대로 가장 점수가 낮은 그룹은 고등학생이 작성한 수필형 글로 나타났습니다.

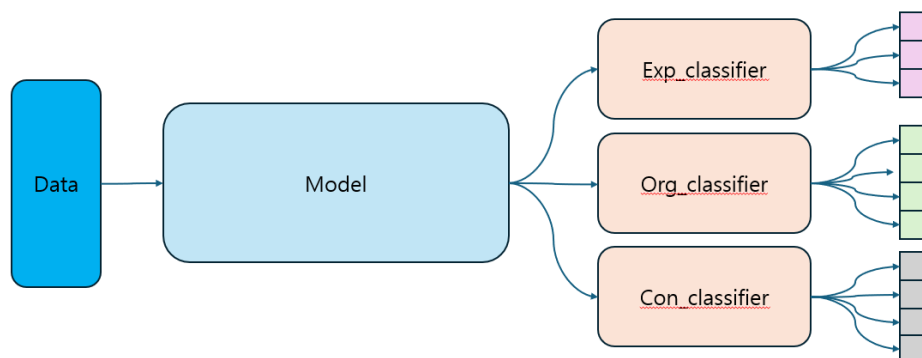
### 1.4 독서량과 최종 점수와의 관계

type	1		2	
student_group	1	2	1	2
student_reading				
1	26.462461	26.262620	26.656402	26.953664
2	26.652546	25.945737	26.858929	27.199257
3	26.538667	26.190262	27.354573	27.250365

1.3에서 수행했던 분석에 독서량을 추가하여 살펴보았습니다. 수필형 글짓기에서는 독서량 증가에 따른 점수 향상의 관계는 약하지만, 논술형 글짓기에서는 중학생과 고등학생 모두 독서량 증가에 따른 점수의 향상이 나타났습니다.

## [2] 모델 및 개선사항

전체적인 분류 모델의 구조는 다음과 같습니다.



위 모델에서 사용한 Pre-trained model은 'roberta-base'를 사용하였습니다. RoBERTa model의 출력은 3가지의 분류기의 입력으로 들어가게 됩니다. 각각의 분류기는 exp, org, con score을 예측합니다. 이렇게 구성한 이유는 각 대분류에 포함된 소분류 평가 지표는 서로 비슷한 특성을 가지고 있을 것이라고 가정하였기 때문입니다.

모델의 주요 parameter는 다음과 같습니다.

Batch\_size : 32

Max\_len = 128

Epoch = 5

Learning\_rate = 2e-5

Optimizer = AdamW

Loss = MSE\_loss(exp) + MSE\_loss(org) + MSE\_loss(Con)

## 2.1 평가 지표 및 모델의 성능

평가 지표는 micro F1 score을 사용하였습니다. Micro F1 score는 소분류 각각의 예측값을 전체적으로 고려한다는 특징이 있으므로, 해당 평가지표를 사용하였습니다. 또한 Micro F1 score 특성 상, 특정 소분류에 불균형이 존재하는 경우에도 유연하게 대처할 수 있습니다.

모델의 성능은 다음과 같습니다.

Model	Overall F1 score	Exp F1 score	Org F1 score	Con f1 score
BERT	0.705	<b>0.789</b>	0.773	0.705
all-MiniLM-L6-v2	0.606	0.697	0.727	0.605
Distilbert	0.678	0.456	0.705	0.678
<b>RoBERTa</b>	<b>0.712</b>	0.697	<b>0.9100</b>	<b>0.712</b>

RoBERTa 모델이 전체적인 성능이 가장 좋았음을 확인할 수 있었습니다. 특히, Org 점수에 대해 예측을 잘 하는 모습을 보여줍니다. BERT 모델의 경우, 전체적인 점수는 RoBERTa와 비슷하지만, Exp 점수에 대해 예측을 잘 하는 모습을 보여줍니다. 마지막으로 Con 점수에 대해서는 예상보다는 낮은 f1 score을 보여주었습니다.

## [3] 한계 및 개선사항

1. 입력 텍스트 길이의 한계: RoBERTa model의 입력 토큰이 512로 제한되므로, Essay와 같이 입력 텍스트가 매우 긴 경우에 대해 전체 텍스트를 입력할 수 없으므로 정보의 손실에 대처하기 어렵습니다. 이런 경우에는 Sliding Window를 적용한 LongFormer의 적용을 고려해보거나 입력 시퀀스에 N-gram을 적용하여 데이터를 늘리는 방법등을 고려해 볼 수 있습니다.

2. 한국어 모델의 부족: 본 모델에서는 한국어로 학습한 모델이 아닌 다국어로 학습한 모델을 사용하였습니다. 한국어 데이터를 이용하여 사전 학습된 모델의 성능이 더욱 좋을 것이라고 예상되지만, 대부분의 최신 모델이 다국어를 지원한다는 점과 back translation등의 기법을 통하여 영어로 치환하여 문제를 해결할 수 있다는 점을 고려하여 다국어 언어 모델로 선택하였습니다.

3. 각 항목별 단일 모델 개발: 제안한 모델은 오직 하나의 Model을 이용하여 Multilabel classification을 수행하였습니다. 하지만, 각각의 대분류의 특성에 맞는 모델을 개발하여 합치는 방식을 이용하면 개별 예측 점수도 높아질 것으로 예상되며, 이에 따라 전체적으로 모델의 성능이 좋아질 것으로 예상됩니다.