Automated Essay Grading for ELs: Sprint 2

Sally Shin

EC601 Fall 2022

CONTENTS

PROJECT GOALS

AIMS

- Kaggle competition
- Grade essays from English Learners while taking into account their skill levels
- Explore transformer NLP architecture model through the project

GRADING CRITERIA

- **Cohesion**: How sentences or paragraphs flow naturally
- **Syntax**: Correct order of words in phrases and sentences
- **Vocabulary**: Number of unique words
- **Phraseology**: How different phrases may present
- **Grammar**: Correct usage of punctuation, word order, word usage, among others
- **Conventions**: mechanics of writing, notably capitalizations and punctuations

USERS

- Teachers who may need help in grading essays from students of all backgrounds
- Writers who wants to test out automated grading essays for their short essays

PREVIOUS GOALS

Accomplished/Touched upon

- Literature Review - 2 important papers detailed
- Check dataset - Ran initial data analysis & closer look at training dataset

Did not accomplish

- Detailing each grading criteria

BASIC DATASET INSIGHTS

- Training set has 3911 entries, each with 6 different supervised scores
- Median number of characters: 2173
- Median number of words: 402

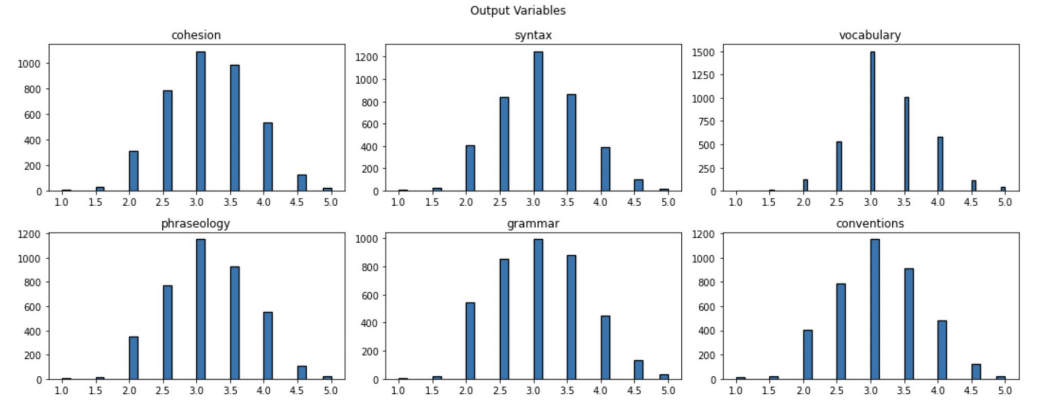
	text_id	full_text	cohesion	syntax	vocabulary	phraseology	grammar	conventions
0	0016926B079C	I think that students would benefit from learn...	3.5	3.5	3.0	3.0	4.0	3.0
1	0022683E9EA5	When a problem is a change you have to let it ...	2.5	2.5	3.0	2.0	2.0	2.5
2	00299B378633	Dear, Principal\n\nIf u change the school poli...	3.0	3.5	3.0	3.0	3.0	2.5
3	003885A45F42	The best time in life is when you become yours...	4.5	4.5	4.5	4.5	4.0	5.0
4	0049B1DF5CCC	Small act of kindness can impact in other peop...	2.5	3.0	3.0	3.0	2.5	2.5

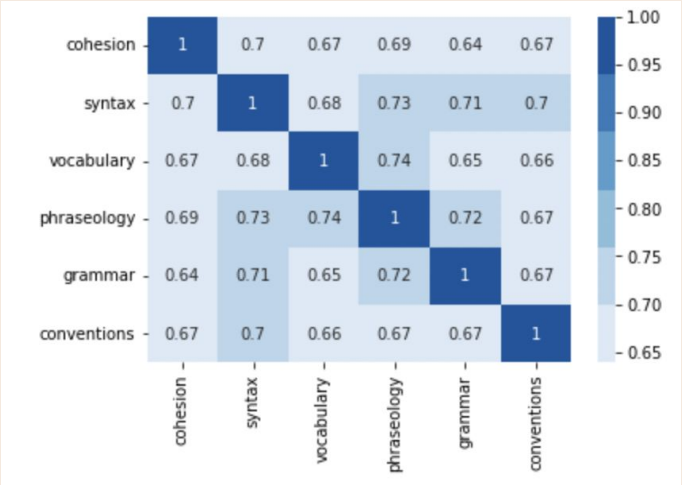
```
"Dear Principal,\r\n\r\nOur school should have a community center. The reasons why, are so students can learn what ou
r community needs, how to make our community better place, and why is community important for students to know. Its a
great to have a community center to know how we can make things better.\r\n\r\nStudents think community center takes
their time away. but they have to learn what our community needs. students will participate in a group of students ma
king a list what our community needs, therefore students will learn what our community needs! students will present t
heir list of things our community needs! due to that students will be giving extra credit for the ones who have low g
rades!\r\n\r\nSome students don't participate because their friends say its waste of time. it would not be waste of t
ime when you get to know how our community can be a better place for us. students should know that the program is abo
ut our own lives, because if our community is bad well our lives are going to be bad. due to that students will want
to participate and will want to make our community a better place for us.\r\n\r\nsome student might say why is the co
mmunity important anyways, were fine nothing is wrong. but when get to know what our community is, they would want to
know about and would want to participate and know why is our community important. students will receive a good grade
if they participate in this community because is like an extra credit. due to that their friends would want to enjoy
the program because their friend told them about it.\r\n\r\nso principal the community center will be a great place f
or students to learn lots of things. so i guess this will be a good idea for our school. students will learn what our
community needs, how to make our community a better place, and why is our community important!\r\n\r\nSincerely, STUD
ENT_NAME"
```

- Example text
- Shows \n or \r or \t characters in some essays
    - Need to be removed during data preprocessing

BASIC SCORING INSIGHTS

- Overall an even spread of each grading categories
- Median of 3.0 for all of the categories
- Highest correlation with Vocabulary and Phraseology
- Other higher correlations with Syntax and Phraseology & Grammar and Phraseology





LITERATURE REVIEW: AES using efficient transformer-based NL models

Ormerod et al.

- Goal: To achieve significant results with modest computational and memory budget
- Transformers score highest on GLUE (General Language Understanding Evaluation) benchmarks
  - But with high computational and memory resource costs
- Considered models Albert (12M), Reformer (16M), Electra (14M), and Mobile-BERT (24M) on different set of student essays from Kaggle
  - All have different tricks to be more efficient
  - Linear layer applied to make embedding size more manageable
- Conclusion: Lesser but very comparable quadratic weighted kappa (QWK) compared to larger models

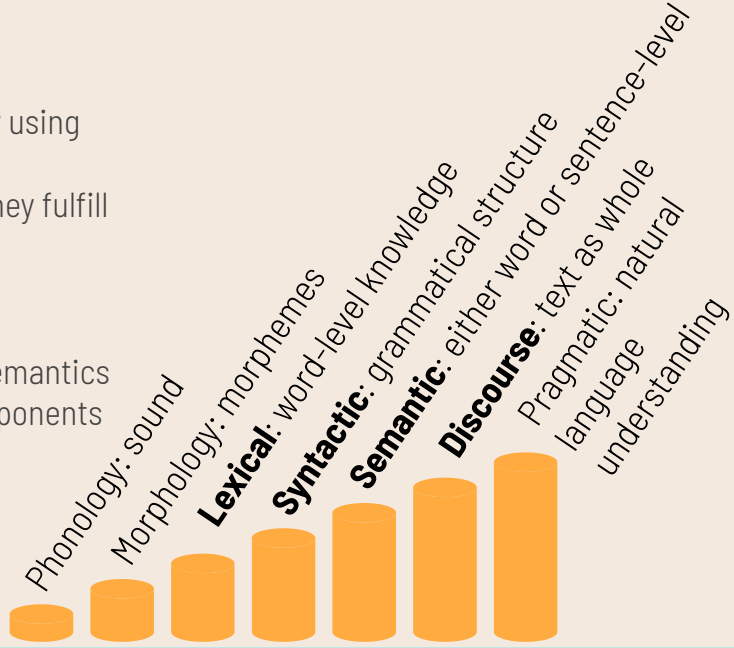
	Essay Prompt							
	1	2	3	4	5	6	7	8
Rater score range	1-6	1-6	0-3	0-3	0-4	0-4	0-12	5-30
Resolved score range	2-12	1-6	0-3	0-3	0-4	0-4	2-24	10-60
Average Length	350	350	150	150	150	150	250	650
Training examples	1783	1800	1726	1772	1805	1800	1569	723

QWK scores	**Prompt**								AVG
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	QWK
EASE	0.781	0.621	0.630	0.749	0.782	0.771	0.727	0.534	0.699
LSTM	0.775	0.687	0.683	0.795	0.818	0.813	0.805	0.594	0.746
LSTM+CNN	0.821	0.688	0.694	0.805	0.807	0.819	0.808	0.644	0.761
LSTM+CNN+Att	0.822	0.682	0.672	0.814	0.803	0.811	0.801	0.705	0.764
BERT(base)	0.792	0.680	0.715	0.801	0.806	0.805	0.785	0.596	0.758
XLNet	0.777	0.681	0.693	0.806	0.783	0.794	0.787	0.627	0.743
BERT + XLNet	0.808	0.697	0.703	0.819	0.808	0.815	0.807	0.605	0.758
R^2BERT	0.817	0.719	0.698	0.845	0.841	0.847	0.839	0.744	0.794
BERT + Features	0.852	0.651	0.804	0.888	0.885	0.817	0.864	0.645	**0.801**
Electra (small)	0.816	0.664	0.682	0.792	0.792	0.787	0.827	0.715	0.759
Albert (base)	0.807	0.671	0.672	0.813	0.802	0.816	0.826	0.700	0.763
Albert (large)	0.801	0.676	0.668	0.810	0.805	0.807	0.832	0.700	0.763
Mobile-BERT	0.810	0.663	0.663	0.795	0.806	0.808	0.824	0.731	0.762
Reformer	0.802	0.651	0.670	0.754	0.771	0.762	0.747	0.548	0.713
Electra+Mobile-BERT	0.823	0.683	0.691	0.805	0.808	0.802	0.835	0.748	0.774
Ensemble	0.831	0.679	0.690	0.825	0.817	0.822	0.841	0.748	**0.782**

LITERATURE REVIEW: Classification of NLP Techniques

Zhao et al.

- Descriptive paper on different NLP techniques over time
- Focused on classifying and analyzing different types and scopes by using linguistic analysis levels
- Broad range of techniques (57 total) with brief description & task they fulfill
- Most notable analyses:
  - Frequency analysis: frequencies of words
  - Similarity analysis: numerical relatedness of different texts
  - Rule-based analysis: rules & patterns to analyze syntax & semantics
  - Syntactic analysis: represent relationship of its logical components
  - Semantic analysis: label semantically relevant components
    - Identify meaning of words and phrases in context
    - Relationship between words



IN PROGRESS for Next Sprint

- Configuring working environment for transformer training
  - Ongoing process
- Understanding workflow of NLP training
- Defining how to modulate each grading categories
  - Still need more literature review
  - Need to look into different corpora that can analyze these criteria

RESOURCES

- Dataset explorations:
  - https://www.kaggle.com/code/riteshsinha/feedback-prize-data-exploration
  - https://www.kaggle.com/code/ryanluoli2/simple-text-analysis-to-start-with

- Literature review:
  - Ormerord et al. Automated essay scoring using efficient transformer-based language models
  - Zhao et al. Classification of Natural Language Processing Techniques for Requirements Engineering

Instructions for use

In order to use this template, you must credit **Slidesgo** by keeping the Credits slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs:
https://slidesgo.com/faqs

Fonts & colors used

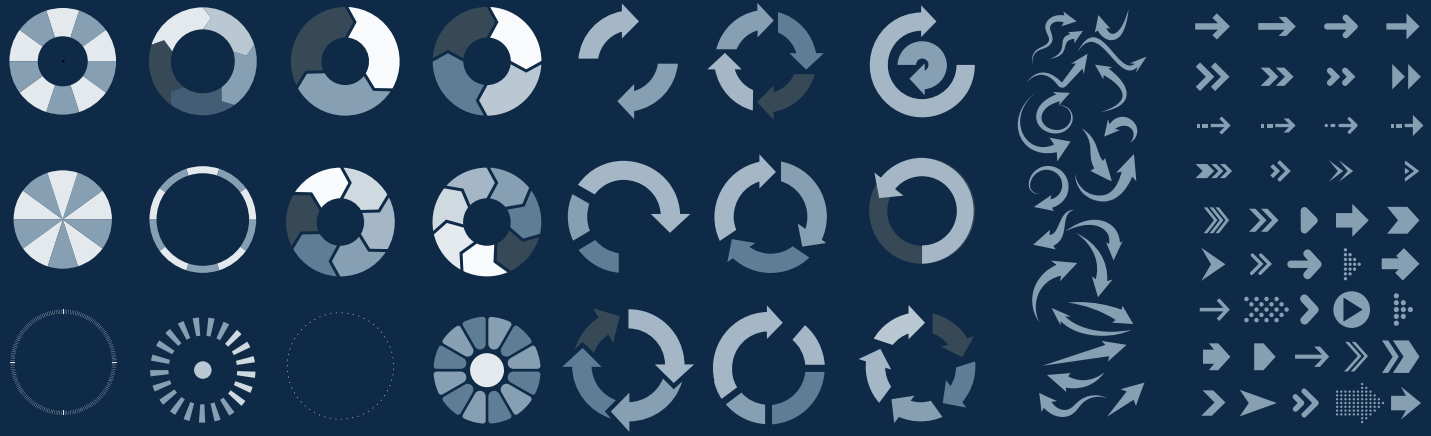
This presentation has been made using the following fonts:

Denk One
(https://fonts.google.com/specimen/Denk+One)

Barlow Semi Condensed
(https://fonts.google.com/specimen/Barlow+Semi+Condensed)

Use our editable graphic resources...

You can easily resize these resources keeping the quality. To change the color just ungroup the resource and click on the object you want to change. Then click on the paint bucket and select the color you want. Don’t forget to group the resource again when you’re done.



...and our set of editable icons

You can resize these icons keeping the quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

Business Icons

Avatar Icons

Creative Process Icons

Educational Process Icons

Help & Support Icons

Medical Icons

Nature Icons

Performing Arts Icons

SEO & Marketing Icons

Teamwork Icons


slidesgo