

Automated {

→ .. Essay


[] Scoring 2.0 *with BERT }



Kaggle

Learning Agency Lab

Automated Essay Scoring


Automated  {
→ .. Essay
[] Scoring 2.0 *with BERT }

FINAL GOAL


IMPROVE UPON ESSAY SCORING ALGORITHMS
TO IMPROVE STUDENT LEARNING OUTCOMES

OUR GOAL

TRAINING A SCORING MODEL
FOR STUDENT ESSAYS



	essay_id	full_text	score
14982	dc77aca	I am against that the technology doesn't have ...	1
8928	83d10b5	Mona Lisa, when you hear her name you think of...	1
6175	5c752fb	Venus is the planet closest to Earth. Also the...	1
7961	773c30a	The author explains his idea very well by sayi...	2



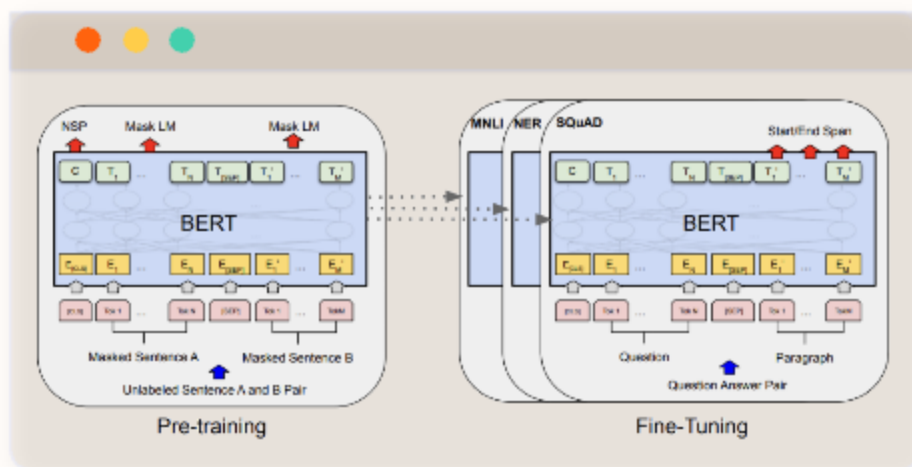
about 24,000 data

essay_id : the unique ID of the essay
full_text : the fully essay response
score : holistic score of the essay on a 1–6 scale



BERT

Bidirectional Encoder Representations from Transformers



The screenshot shows a Jupyter Notebook titled "what is a BERT?.ipynb". The notebook content includes:

- A code cell with the text: `[] an open-source model developed by Google in 2018`
- A text cell with the text: `bidirectional representations from unlabeled text`
- A code cell with the text: `[] Masked LM + Next Sentence Prediction`

The notebook interface includes a toolbar with various icons for text formatting, code execution, and file management. A comment box at the bottom right contains the text: "fine-tuned with just one additional output layer (downstream task)".



ALBERT



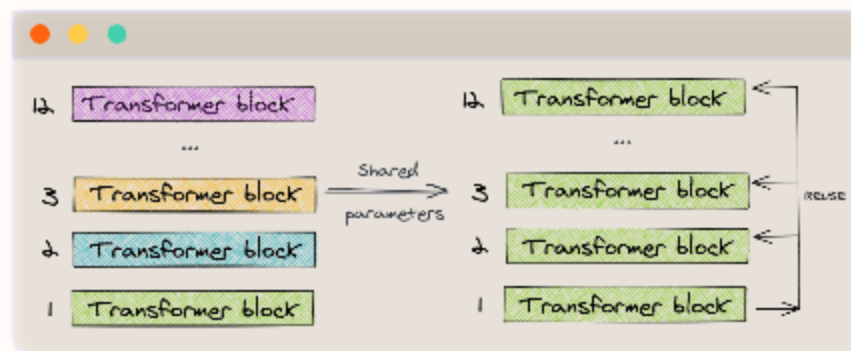
ELECTRA



DeBERTa

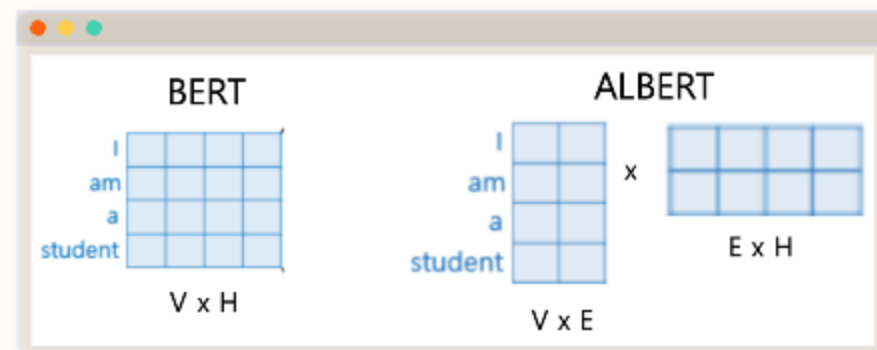
maintain similar performance to BERT
while requiring fewer resources

cross – layer parameter sharing



reducing the matrix size
increasing memory efficiency

factorized embedding layer parameterization





Kaggle

Learning Agency Lab

Automated Essay Scoring



ALBERT

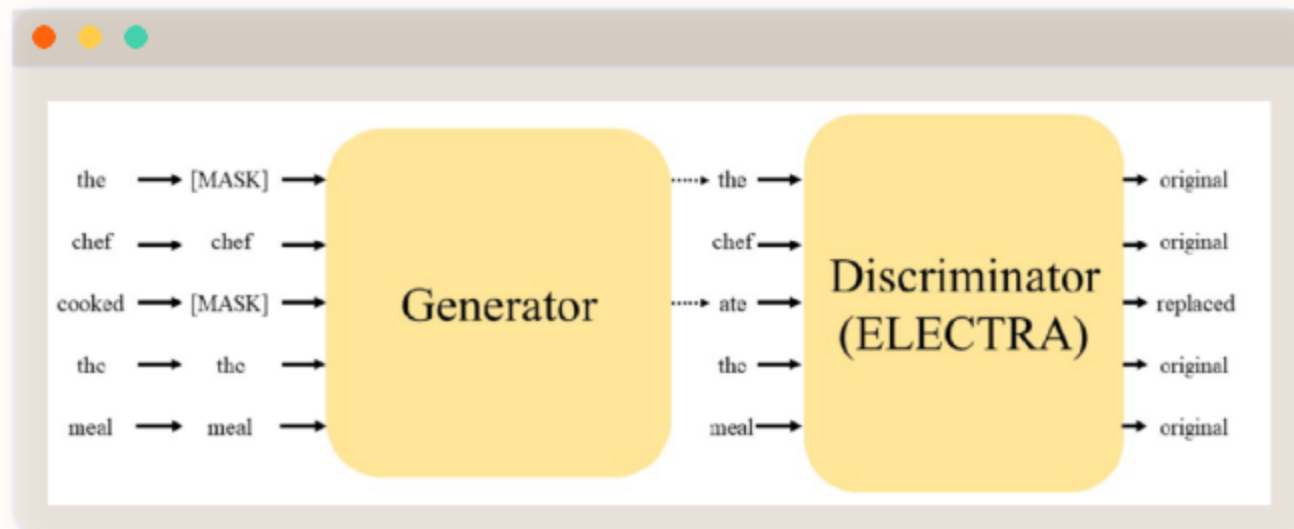


ELECTRA



DeBERTa

a model replacing BERT's pre-learning method with alternative token detection



earring for all tokens



Kaggle

Learning Agency Lab

Automated Essay Scoring



ALBERT



ELECTRA

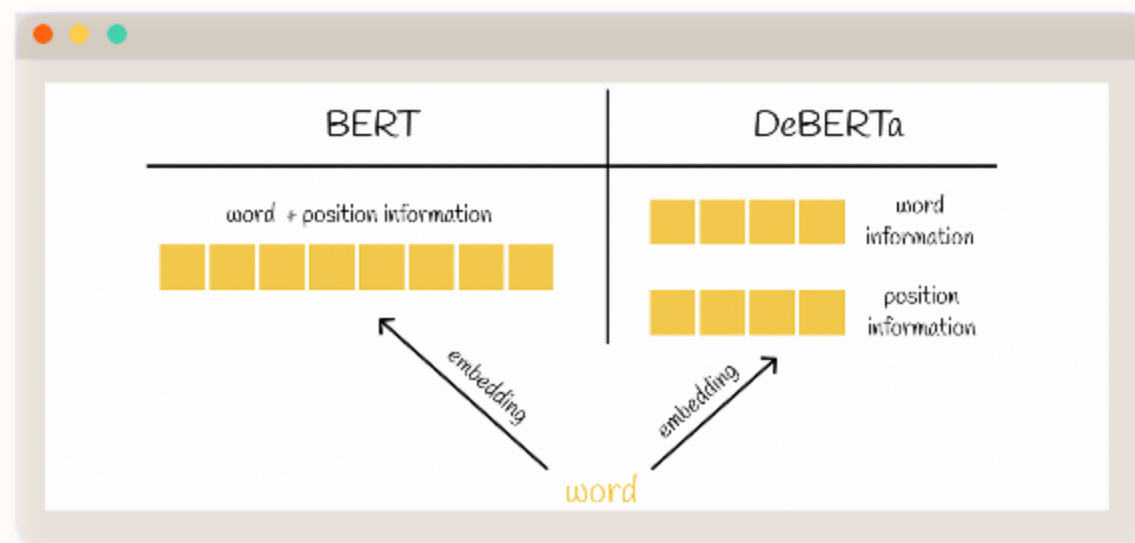


DeBERTa

● disentangled attention mechanism

● enhanced positional encodings

● improved parameter efficiency





Quadratic Weighted Kappa

How to measure agreement on measurement category values among observers



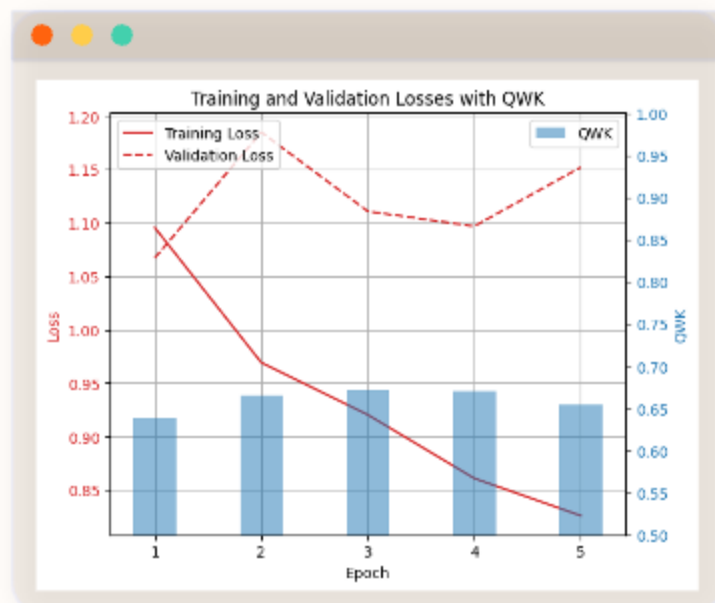


actual score i
predicted score j

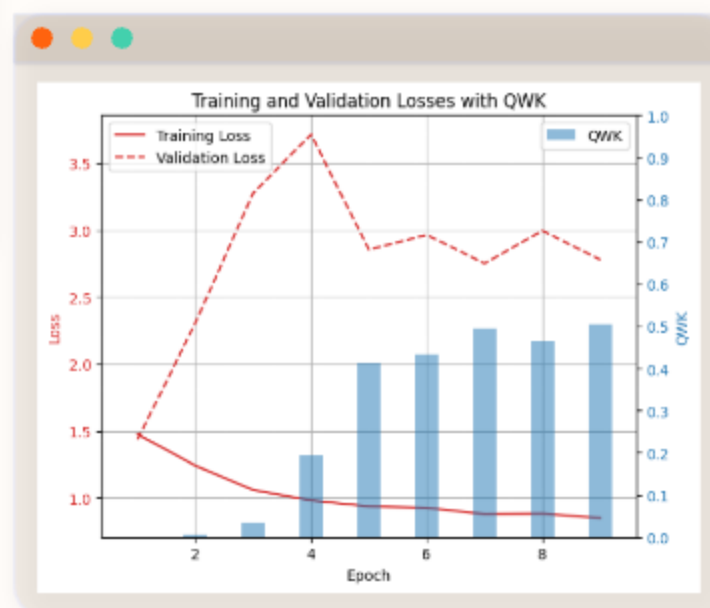
$\left[\begin{array}{c} n \times n \text{ histogram matrix} \\ O \\ \text{representing the number of } i \text{ and } j \end{array} \right]$

$\left[\begin{array}{c} \text{weighted matrix} \\ W \\ w_{i,j} = \frac{(i-j)^2}{(N-1)^2} \end{array} \right]$

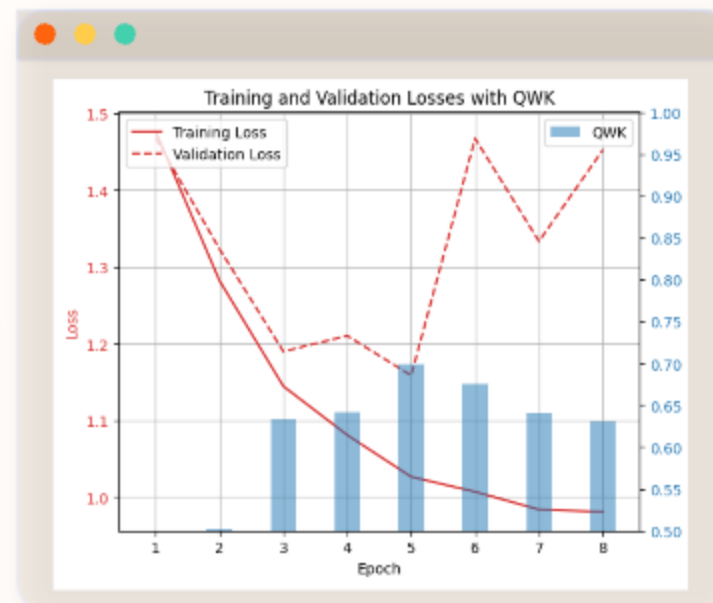
- Normalized so that the sum of the prediction results (E and O) is the same
 - Calculate evaluation index qwk with three matrices W, O, E
- $$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$$



ELECTRA
best qwk : 0.505179



ALBert
best qwk : 0.698652



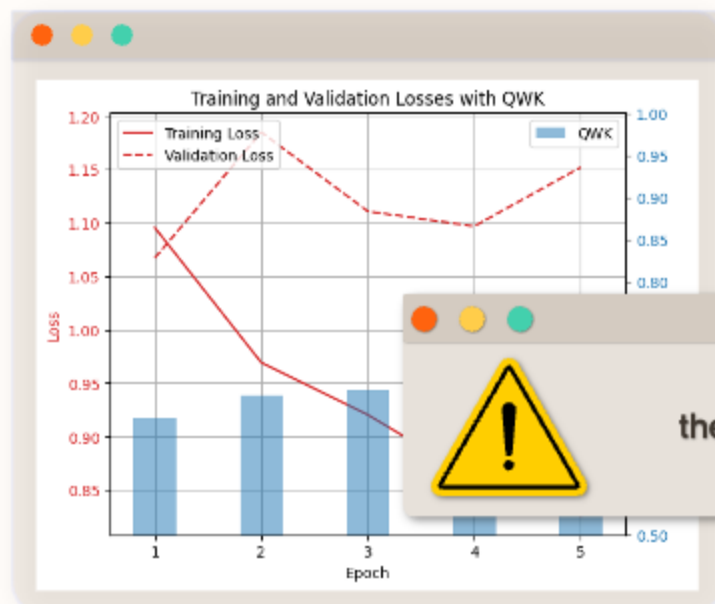
DeBERTa
best qwk : 0.730117



Kaggle

Learning Agency Lab

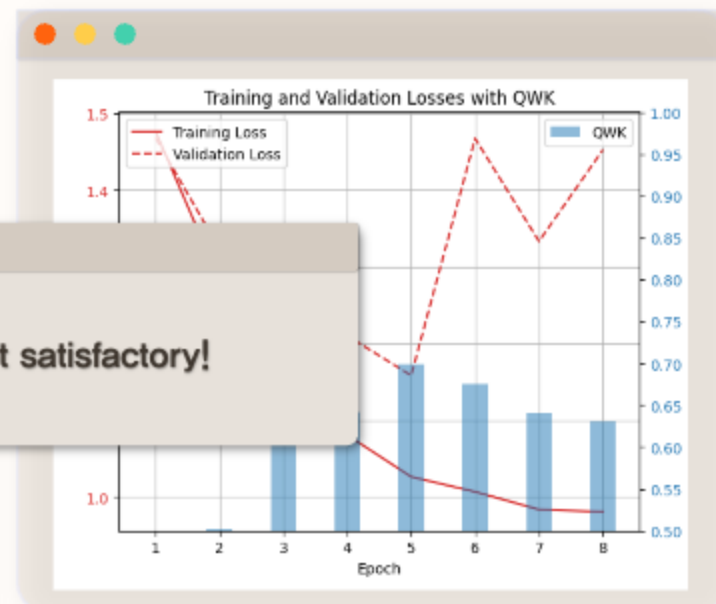
Automated Essay Scoring



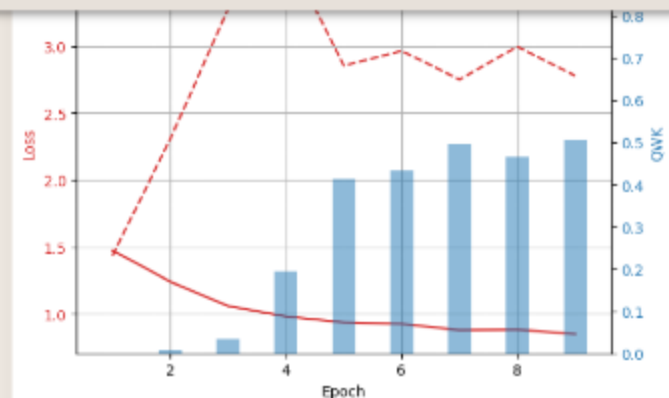
ELECTRA
best qwk : 0.505179



the performance of the classification model was not satisfactory!



DeBERTa
best qwk : 0.730117



ALBert
best qwk : 0.698652



How to solve the problem?

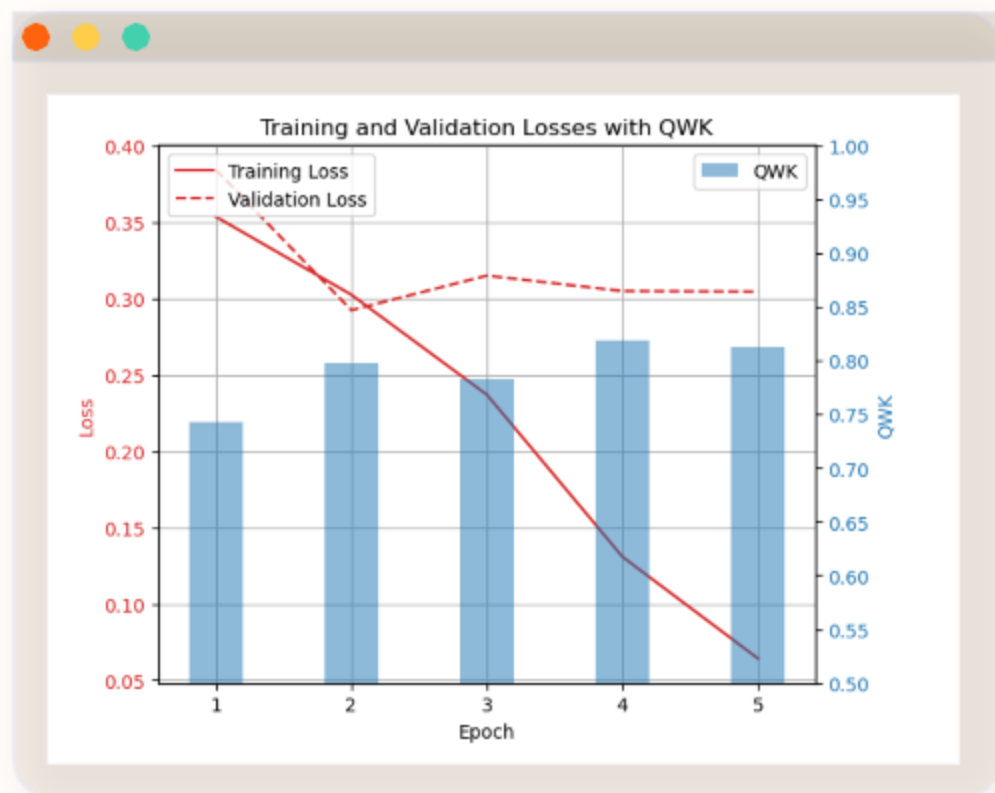
classification



regressor

- score is an integer in the range of 1 to 6 points
- number of labels = 6
- loss : cross entropy

- score is an order scale
- number of labels = 1
- loss : MSE
- dropout part



A table titled "Regressor" comparing three BERT models based on their best training loss, best validation loss, and best QWK score. The table has four columns: Model, Best Training Loss, Best Validation Loss, and Best QWK. The rows are ALBert, ELECTRA, and DeBERTa. DeBERTa is highlighted with orange text.

Model	Best Training Loss	Best Validation Loss	Best QWK
ALBert	0.018100	0.349295	0.788330
ELECTRA	0.196400	0.341242	0.782262
DeBERTa	0.017700	0.297137	0.818309

INTERMEDIATE RESULTS

verify that the regression model has higher performance
among the three BERT model, DeBERTa has the highest performance



Kaggle

Learning Agency Lab

Automated Essay Scoring

	essay_id	full_text	score
14982	dc77aca	I am against that the technology doesn't have ...	1
8928	83d10b5	Mona Lisa, when you hear her name you think of...	1
6175	5c752fb	Venus is the planet closest to Earth, Also the...	1
7961	773c30a	The author explains his idea very well by sayi...	2



~~im~~ balanced data



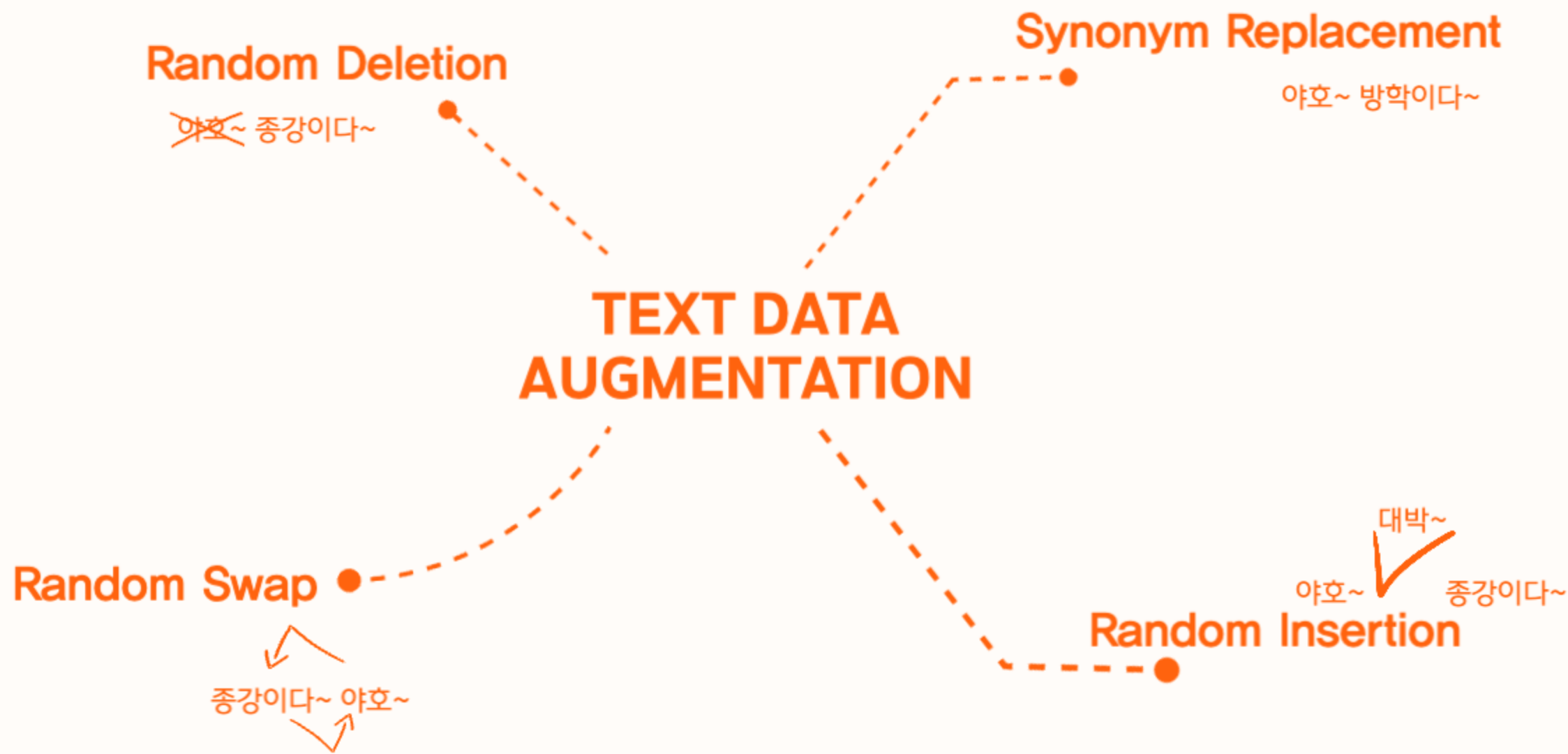
text data
augmentation



Kaggle

Learning Agency Lab

Automated Essay Scoring





Kaggle

Learning Agency Lab

Automated Essay Scoring



Original Sentence : This is an example sentence for random insertion.

Augmented Sentence : This is an example sentence for random sentences insertion.

TEXT DATA AUGMENTATION

Random Insertion

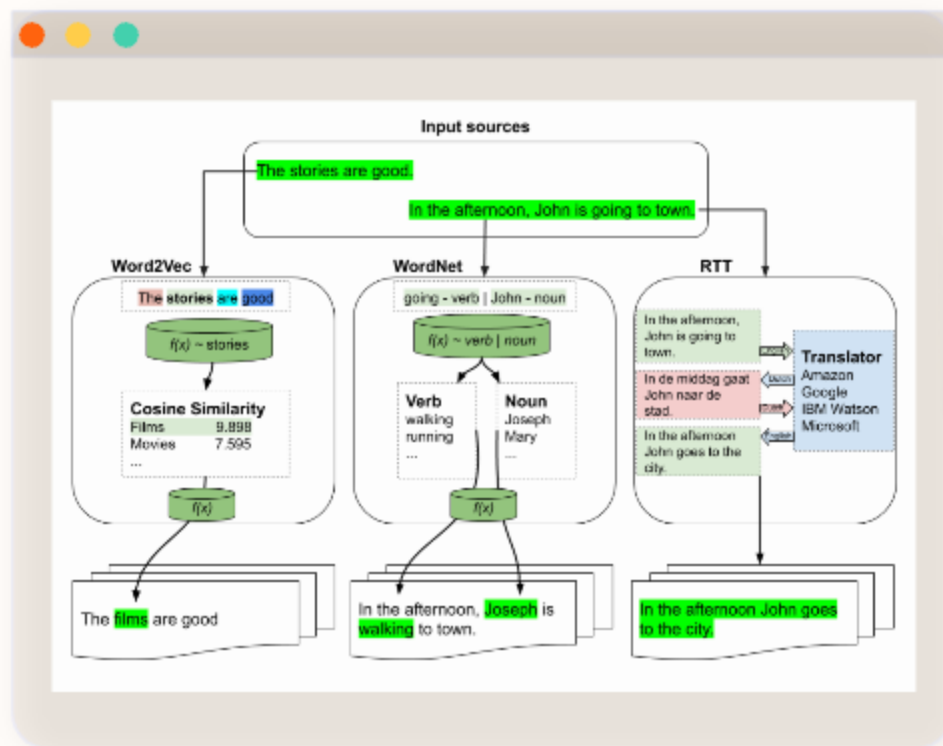


Original Sentence : This is an example sentence for random insertion.

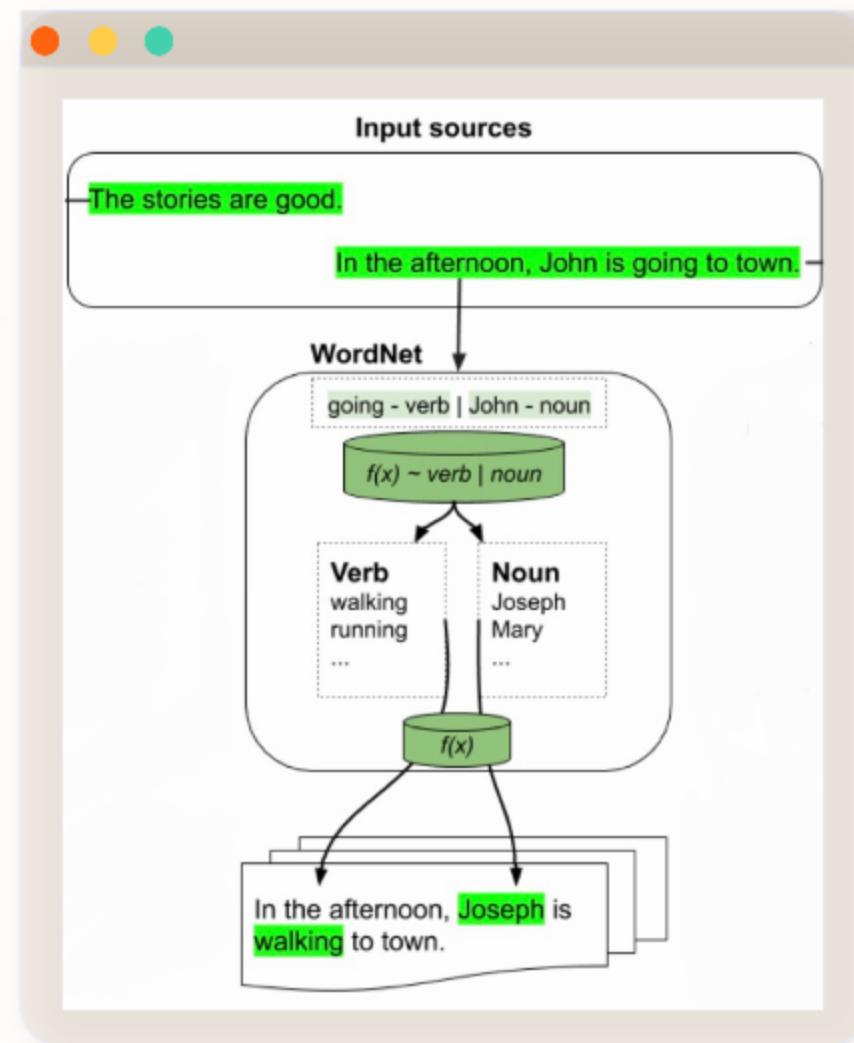
Augmented Sentence : example This is an example random sentence for random sentences insertion insertion.

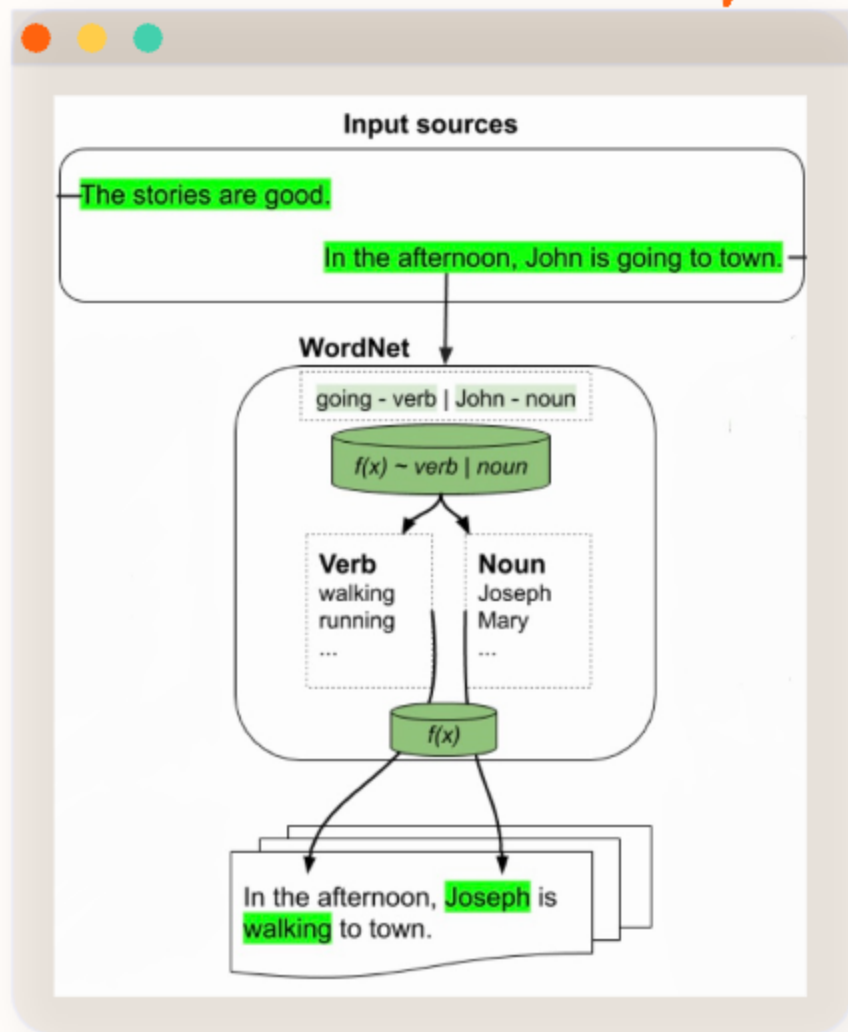


problem : affect the score



Wordnet-Based Synonym Augmentation





Wordnet-Based Synonym Augmentation

- **SELECT REPLACEABLE WORDS**
like verbs, nouns, and the combination of them
- The new sentence is constructed by replacing the selected verb or noun with **THEIR SYNONYMS**
- The algorithm has options to choose to augment using either verbs or nouns or even a combination



Round-Trip Translation

ORIGINAL LANGUAGE

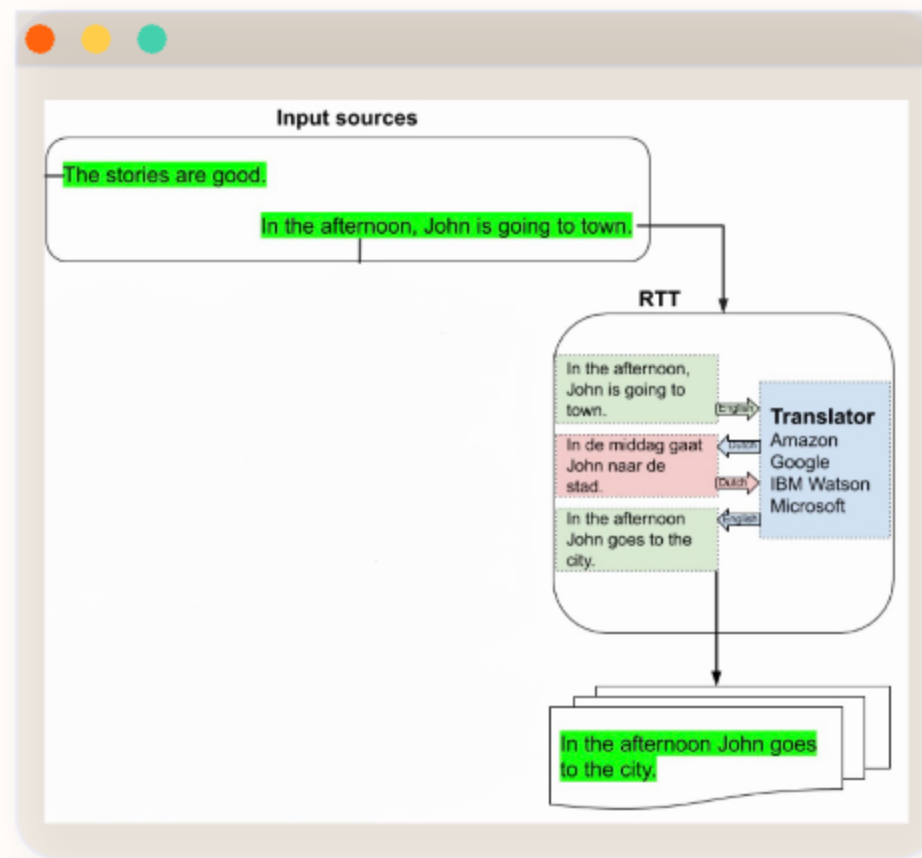
야호~ 종강이다~

FORWARD TRANSLATION

Yay~ It's the end of the semester~

BACK TRANSLATION

이야~ 이제 학기말이다~



TEXT DATA AUGMENTATION RESULT

```
class Augmenter:
    def __init__(self, method='wordnet', **kwargs):
        """
        Augmenter 클래스 초기화
        method: 사용할 증강 기법 ('wordnet', 'word2vec', 'translate')
        kwargs: 각 증강 기법의 인스턴스 생성 시 필요한 추가 매개변수
        """
        self.method = method.lower()
        if self.method == 'wordnet':
            self.augmenter = Wordnet(v=kwargs.get('v', True),
                                     n=kwargs.get('n', False),
                                     runs=kwargs.get('runs', 1),
                                     p=kwargs.get('p', 0.5))

        elif self.method == 'translate':
            self.augmenter = Translate(src=kwargs.get('src', 'en'),
                                       to=kwargs.get('to', 'fr'))

        else:
            raise ValueError("지원하지 않는 증강 기법입니다. 'wordnet', 'translate' 중 하나를 선택하세요.")

    def augment_text(self, text, **kwargs):
        """
        주어진 텍스트를 증강하는 메소드
        text: 증강할 원본 텍스트
        kwargs: 증강 시 필요한 추가 매개변수
        """
        if self.method == 'wordnet':
            top_n = kwargs.get('top_n', 10)
            augmented_text = self.augmenter.augment(text, top_n=top_n)
        elif self.method == 'translate':
            augmented_text = self.augmenter.augment(text)
        else:
            raise ValueError("지원하지 않는 증강 기법입니다. 'wordnet', 'translate' 중 하나를 선택하세요.")

        # 개행 문자 제거
        augmented_text = re.sub(r'\n', ' ', augmented_text)
        return augmented_text
```

Wordnet

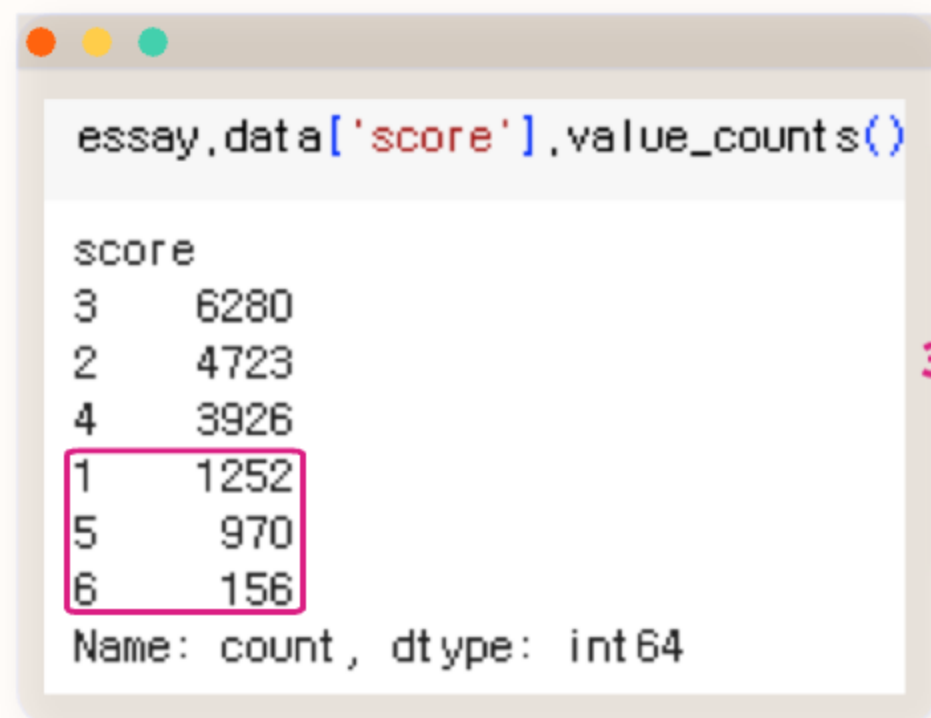
I am a scientist at nasa that is discussing the "face" on mars,
iodine will comprise explaining how the "face" is a ground form.

Original Sentence : I am a scientist at NASA that is discussing the "face" on mars,
I will be explaining how the "face" is a land form.

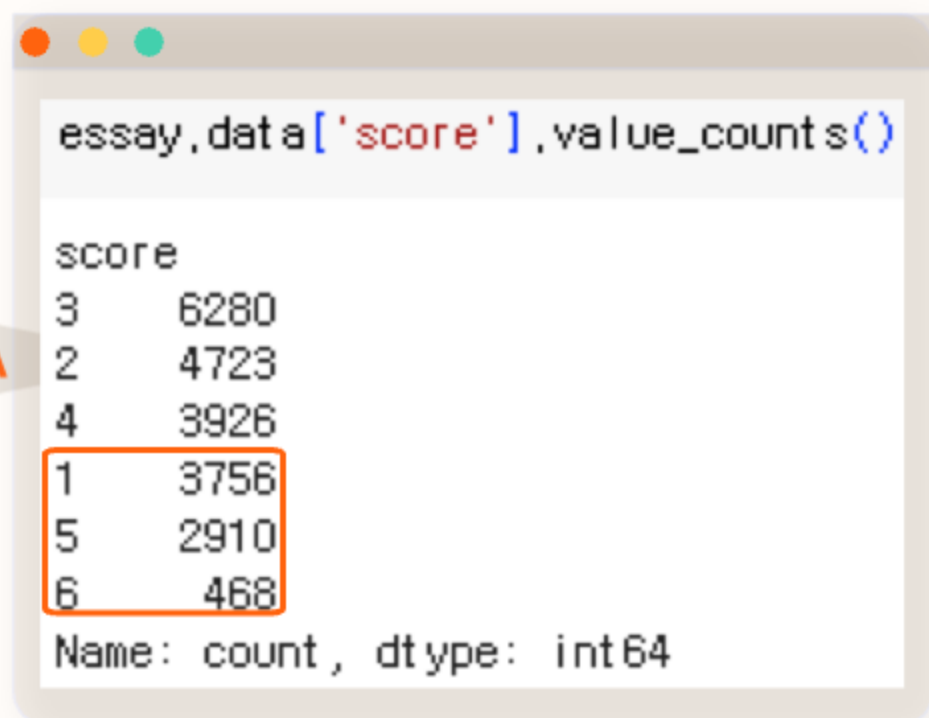
Translate

i am scientist from nasa who discusses the "face" no mars,
i will explain how the "face" is a terrestrial form.

TEXT DATA AUGMENTATION RESULT




3X MORE DATA





FINAL RESULTS



[17650/17650 4:33:57, Epoch 10/10]			
Epoch	Training Loss	Validation Loss	Qwk
1	0.328100	0.295461	0.892667
2	0.265100	0.298933	0.900289
3	0.193900	0.251835	0.907659
4	0.110500	0.258230	0.907451
5	0.058500	0.275791	0.905832
6	0.030600	0.244650	0.912932
7	0.016200	0.244928	0.911339
8	0.010100	0.242523	0.914721
9	0.004300	0.244368	0.911536
10	0.002100	0.242474	0.912705

MODEL - DEBERTA	BEST TRAINING LOSS	BEST VALIDATION LOSS	BEST QWK
NOT AUGMENTED	0.018100	0.349295	0.788330
AUGMENTED	0.010100	0.242523	0.914721

BEST QWK

0.788330

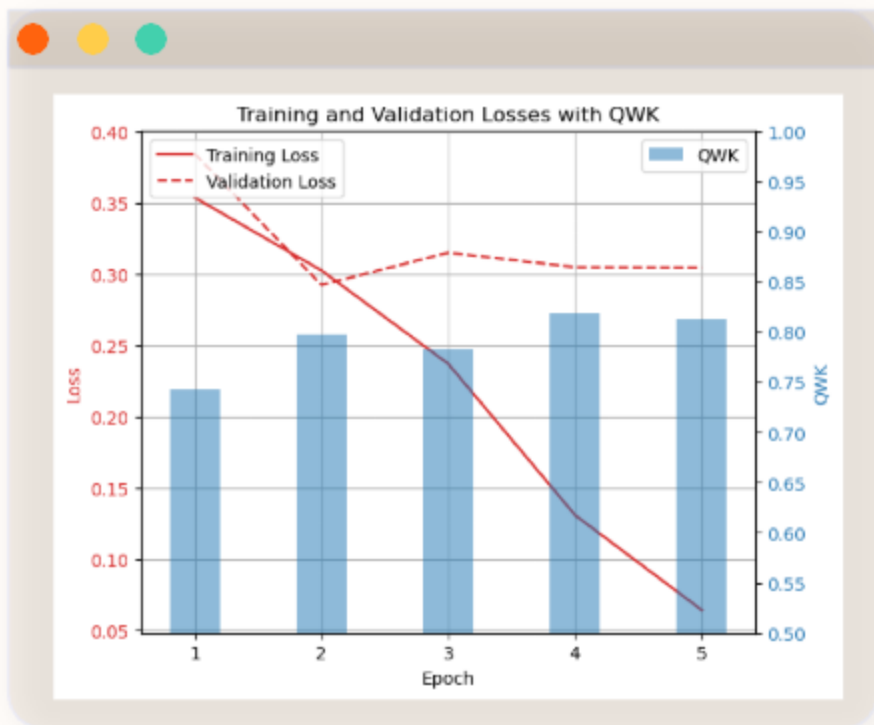
0.914721



Kaggle

Learning Agency Lab

Automated Essay Scoring



REGRESSOR DEBERTA MODEL

+

ORIGINAL DATA



REGRESSOR DEBERTA MODEL

+

AUGMENTED DATA



PROBLEM 1

Poor performance of classification model



Using regressor model

PROBLEM 2

Data imbalance



Text augmentation

COMPLEMENT



Add word2vec
for text data augmentation



Use other LLM models

THANK == {

→ .. YOU

[] ^ _ ^ u ^ n u n *