

PROJECT: Evaluation Of Descriptive Answers using NLP

- ❖ We have decided to use NLP techniques and a multiple linear regression at the end to evaluate the descriptive answers continuously and follow the procedure below.

REQUIRED PYTHON LIBRARIES:

- ❖ Please install the following libraries before running the code to avoid any errors.
 - o numpy and pandas
 - o spacy
 - o neuralcoref
 - o textblob
 - o re
 - o string
 - o nltk
 - o matplotlib
 - o sklearn
- ❖ In this the installation of neuralcoref and spacy libraries is a bit tricky. So I would like to provide detailed instructions for installing them.
 - o First download visual studio installer and select “Desktop development with C++” to install “Visual Studio Build Tools 2022”.
 - o Now open conda command prompt and install “neuralcoref” by running
pip uninstall neuralcoref
pip install neuralcoref --no-binary neuralcoref
 - o Now install “spacy” library by running below code in command prompt
pip install -U spacy
python -m spacy download en
 - o Now run the below code one by one to avoid any errors in the installation
venv .env
source .env/bin/activate
git clone https://github.com/huggingface/neuralcoref.git
cd neuralcoref
pip install -r requirements.txt
pip install -e .
 - o Finally run python -m spacy download en_core_web_md before importing the spacy library.

PROCEDURE FOLLOWED:

1. Data collection:

- ★ We have chosen the question “What are the various essential characteristics of the cell ?” as a test question and tried to train the algorithm by taking one sample definition as the original answer and wrote about 10 answers conveying almost the exact meaning comprising of following conditions as the student’s answers and their original grading .
 - ❖ Test case 1: Including Synonyms, antonym or wrong statements
 - ❖ Test case 2: Less points.
 - ❖ Test case 3: wrong mapping of functions but includes all.
 - ❖ Test case 4: antonyms and not including all points. With correct mapping of functions.
 - ❖ Test case 5: Various Synonyms. Same text.
 - ❖ Test case 6: Misspelled words and antonyms
 - ❖ Test case 7: Using more pronouns
 - ❖ Test case 8: More points with unnecessary points. [Deduction of score]
 - ❖ Test case 9: Writing facts but irrelevant to the question.
 - ❖ Test case 10: No antonyms, includes all points with different Synonyms. No extra.
- ★ Now for testing we have taken 10 different questions and their original answers and written students answers following exact same conditions as mentioned above.
- ★ Data used to train the algorithm - [link](#)
- ★ Data used to test the developed algorithm - [link](#)

2. Data Cleaning:

- ★ We have decided to perform the required data cleaning before comparing both answers.
- ★ Firstly we have decided to store the text as a pandas DataFrame because of its ease to manipulate required rows and columns.

Original Answer

Historiography has a number of related meanings. Firstly, it can refer to how history has been produced: the story of the development of and practices (for example, the move from short-term biographical narrative toward long-term thematic analysis). Secondly, it can refer to what has been produced: a specific body of historical writing (for example, "medieval historiography during the 1960s" means "works of medieval history written during the 1960s"). Thirdly, it may refer to why history is produced: the philosophy of history. As a meta-level analysis of descriptions of the past, this third conception can relate to the first two in that the analysis usually focuses on the narratives, interpretations, world view, use of evidence, or method of presentation of other historians. Professional historians also debate the question of whether history can be taught as a single coherent narrative or a series of competing narratives.

Student's Answer

History has many relevant implications. First, it may have to do with how history was created: history of development and practice (eg, the transition from short-term biographical stories to long-term thematic analysis). Second, it may refer to what was created: a particular collection of historical works (for example, "Medieval History of the 1960s" is "Works of Medieval History Written in the 1960s". Means). . As an analysis of the meta level of the past, this third concept is the first two, as the analysis usually focuses on stories, interpretations, worldviews, use of evidence, or other historian representations. It may be related. Professional historians also argue whether history can be taught as a single coherent story or as a series of competing stories.

- ★ It is quite obvious that the answers we receive from the students may be different from the original answer because of their own sentence formations, usage of pronouns, synonyms etc.
- ★ So we have decided to use the **"neuralcoref"** package in python to replace the pronouns with the nouns they are intended to represent. This task is commonly referred to as **"Coreference Resolution"** in NLP.

Original Answer (Replacing Pronouns with nouns)

Historiography has a number of related meanings. Firstly, it can refer to how history has been produced: the story of the development of and practices (for example, the move from short-term biographical narrative toward long-term thematic analysis). Secondly, it can refer to what has been produced: a specific body of historical writing (for example, "medieval historiography during the 1960s" means "works of medieval history written during the 1960s"). Thirdly, it may refer to why history is produced: the philosophy of history. As a meta-level analysis of descriptions of the past, this third conception can relate to the first two in that the analysis usually focuses on the narratives, interpretations, world view, use of evidence, or method of presentation of other historians. Professional historians also debate the question of whether history can be taught as a single coherent narrative or a series of competing narratives.

Student Answer (Replacing Pronouns with nouns)

History has many relevant implications. First, it may have to do with how history was created: history of development and practice (eg, the transition from short-term biographical stories to long-term thematic analysis). Second, it may refer to what was created: a particular collection of historical works (for example, "Medieval History of the 1960s" is "Works of Medieval History Written in the 1960s". Means). . As an analysis of the meta level of the past, this third concept is the first two, as the analysis usually focuses on stories, interpretations, worldviews, use of evidence, or other historian representations. It may be related. Professional historians also argue whether history can be taught as a single coherent story or as a series of competing stories.

- ★ Now we tried to convert the entire text into lowercase and also remove any punctuation using **"re and string"** packages in python.

Original Answer (Removing punctuations and all lower case)

historiography has a number of related meanings firstly it can refer to how history has been produced the story of the development of and practices for example the move from shortterm biographical narrative toward longterm thematic analysis secondly it can refer to what has been produced a specific body of historical writing for example medieval historiography during the means works of medieval history written during the it may refer to why history is produced the philosophy of history as a metalevel analysis of descriptions of the past this third conception can relate to the first two in that the analysis usually focuses on the narratives interpretations world view use of evidence or method of presentation of other historians professional historians also debate the question of whether history can be taught as a single coherent narrative or a series of competing narratives

Student Answer (Removing punctuations and all lower case)

history has many relevant implications first it may have to do with how history was created history of development and practice eg the transition from shortterm biographical stories to longterm thematic analysis second it may refer to what was created a particular collection of historical works for example medieval history of the is works of medieval history written in the means as an analysis of the meta level of the past this third concept is the first two as the analysis usually focuses on stories interpretations worldviews use of evidence or other historian representations it may be related professional historians also argue whether history can be taught as a single coherent story or as a series of competing stories

- ★ **"textblob"** python package is used to fix if any typos existed in the student's answer and also the count of typos can be stored, which can be used as a factor in the final evaluation.

```

List of typos
['implications'] ['implication']
no of typos :
1
Original Answer (After Fixing typos)
historiography has a number of related meanings firstly it can refer to how history has been produced the story of the developm
ent of and practices for example the move from shorter biographical narrative toward longer rheumatic analysis secondly it can
refer to what has been produced a specific body of historical writing for example medieval historiography during the means wor
ks of medieval history written during the it may refer to why history is produced the philosophy of history as a metalevel ana
lysis of descriptions of the past this third conception can relate to the first two in that the analysis usually focused on the
narratives interpretations world view use of evidence or method of presentation of other historians professional historians als
o debate the question of whether history can be taught as a single coherent narrative or a series of competing narratives

Student Answer (After Fixing typos)
history has many relevant implication first it may have to do with how history was created history of development and practice
eg the transition from shorter biographical stories to longer rheumatic analysis second it may refer to what was created a part
icular collection of historical works for example medieval history of the is works of medieval history written in the means
as an analysis of the metal level of the past this third concept is the first two as the analysis usually focused on stories in
terpretations worldviews use of evidence or other historian representations it may be related professional historians also arg
ue whether history can be taught as a single coherent story or as a series of competing stories

```

- ★ Each word of the text is lemmatized with its POS tag using “**nlTK**” packages like **punct**, **averaged_perceptron_tagger** and **wordnet**.

```

Original Answer (Lemmatized)
historiography have as number of related meaning firstly it can related to how history have be produce the history of the devel
opment of and practice for example the move from short biographical narrative toward longer rheumatic analysis secondly it can
related to what have be produce as specific body of historical writing for example medieval historiography during the meaning w
ork of medieval history writing during the it may related to why history be produce the philosophy of history as as metalevel a
nalysis of description of the past this third conception can related to the firstly two in that the analysis usually focus on t
he narrative interpretation world view practice of evidence or method of presentation of other historian professional historian
also debate the question of whether history can be teach as as single coherent narrative or as series of compete narrative

Student Answer (Lemmatized)
history have many relevant implication first it may have to do with how history be create history of development and practice e
g the transition from short biographical story to longer rheumatic analysis second it may refer to what be create a particular
collection of historical work for example medieval history of the be work of medieval history write in the mean as an analysis
of the metal level of the past this third concept be the first two as the analysis usually focus on story interpretation worldv
iews use of evidence or other historian representation it may be relate professional historian also argue whether history can b
e teach as a single coherent story or as a series of compete story

```

- ★ Upon researching we also found some advanced text preprocessing techniques like **Chunking**, **Relationship extraction**, **Collocation extraction**, **Named entity recognition**, which are a bit complex for our understanding. So it wasn't included in our model.

3. Organizing the Data :

- ★ We have two suitable types of organizing data: **Corpus** and **Document-Term matrix**.
- ★ The data we got after the preprocessing is stored as **Corpus**, which refers to a collection of text.

	Person	text
0	Teacher	historiography have as number of related meani...
1	student	history have many relevant implication first i...

- ★ We have decided to convert it into a “**Document-Term matrix**”, where every row represents a different document and every column represents a different word(Here we have also excluded common stop words like a, which add no additional meaning to the given text) . It is commonly achieved by using “**scikit-learn’s CountVectorizer**”.

	analysis	argue	biographical	body	coherent	collection	complete	concept	conception	create	...	teach	transition	use	usually	view	work	world	world
0	3	0	1	1	1	0	1	0	1	0	...	1	0	0	1	1	1	1	1
1	3	1	1	0	1	1	1	1	0	2	...	1	1	1	1	0	2	0	0

2 rows × 64 columns

4. Exploratory Data Analysis:

- ★ Now we have decided to calculate the percentage of common words between original and student answers, which can be used as a parameter in final grading.
- ★ It is important to know that this overall similarity doesn't give us the proper comparison.

Original Text Sentences

['historiography have as number of related meaning', 'firstly it can related to how history have be produce the history of the development of and practice for example the move from short biographical narrative toward longer rheumatic', 'analysis secondly it can related to what have be produce as specific body of historical writing for example medieval historiography during the meaning work of medieval history writing during the it', 'may related to why history be produce the philosophy of history as', 'a s metalevel analysis of description of the past this third conception can related to the firstly two in that the analysis usually focus on the narrative interpretation world view practice of evidence or method of presentation of', 'other historian professional historian also debate the question of whether history can be teach as as single coherent narrative or as series of', 'complete', 'narrative']

Student Text Sentences

['history have many relevant implication', 'first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis', 'second it may refer to what be create a particular collection of historical work for example medieval history of the be work of medieval history write in the mean as', 'a n', 'analysis', 'of the metal level of the past this third concept be the first two as the analysis usually focus on story interpretation worldviews use of evidence or other historian representation it may be relate', 'professional historian also argue', 'whether history can be teach as a single coherent story or as a series of complete story']

- ★ So we went one step further and decided to find a similarity score by listing the nouns to choose sentences and find individual sentence similarities.
- ★ We also stored a list of missing nouns in the student's text which can be used in final evaluation.

```
0:['firstly it can related to how history have be produce the history of the development of and practice for example the move from short biographical narrative toward longer rheumatic', ['first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis', 'second it may refer to what be create a particular collection of historical work for example medieval history of the be work of medieval history write in the mean as', 'first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis']],
overall similarity:0.9735369974293147
1:['analysis secondly it can related to what have be produce as specific body of historical writing for example medieval historiography during the meaning work of medieval history writing during the it', ['second it may refer to what be create a particular collection of historical work for example medieval history of the be work of medieval history write in the mean as', 'first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis']],
overall similarity:0.9698571811048565
2:['may related to why history be produce the philosophy of history as', ['first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis']],
overall similarity:0.9582719589528044
3:['as metalevel analysis of description of the past this third conception can related to the firstly two in that the analysis usually focus on the narrative interpretation world view practice of evidence or method of presentation of', ['first it may have to do with how history be create history of development and practice eg the transition from short biographical story to longer rheumatic analysis', 'of the metal level of the past this third concept be the first two as the analysis usually focus on story interpretation worldviews use of evidence or other historian representation it may be relate']],
overall similarity:0.9604702293147324
4:['other historian professional historian also debate the question of whether history can be teach as as single coherent narrative or as series of', ['whether history can be teach as a single coherent story or as a series of complete story', 'professional historian also argue', 'of the metal level of the past this third concept be the first two as the analysis usually focus on story interpretation worldviews use of evidence or other historian representation it may be relate']],
overall similarity:0.9067023273233802
```

List of missing nouns

['historiography', 'number', 'meaning', 'move', 'historiography', 'meaning', 'body', 'writing', 'philosophy', 'metalevel', 'description', 'conception', 'world', 'view', 'method', 'presentation', 'question']

5. Feature Engineering:

- ★ From the data, we extracted 3 features that determine the overall score for the given answer.

a. Sentence Similarity Score :

This Score considers the presence of synonyms, antonyms and wrong mapping of verbs, adj, adv of corresponding nouns

```
[0.9735369974293147, 0.9698571811048565, 0.9582719589528044, 0.9604702293147324, 0.9067023273233802]  
similarity_score = 0.9537677388250175
```

b. fraction_of_new_topics:

missing_words_count - total no of missing words(distinct) in student's

texttotal_words_student_count - total no of words(distinct) in student's text

total_words_original_count - total no of words(distinct) in original text

$$\text{fraction_of_new_topics} = (\text{total_words_student_count} - \text{total_words_original_count} + \text{missing_words_count}) / \text{total_words_original_count}$$

```
common_word_count = 24 and total_words_count_student = 43 and % = 0.5581395348837209
```

```
common_words = ['analysis', 'biographical', 'coherent', 'compete', 'development', 'evidence', 'example', 'focus', 'historian',  
'historical', 'history', 'interpretation', 'longer', 'medieval', 'past', 'practice', 'professional', 'rheumatic', 'series', 'short', 'single', 'teach', 'usually', 'work']
```

```
total_words_student = ['analysis', 'argue', 'biographical', 'coherent', 'collection', 'compete', 'concept', 'create', 'development', 'evidence', 'example', 'focus', 'historian', 'historical', 'history', 'implication', 'interpretation', 'level', 'longer', 'mean', 'medieval', 'metal', 'particular', 'past', 'practice', 'professional', 'refer', 'relate', 'relevant', 'representation', 'rheumatic', 'second', 'series', 'short', 'single', 'story', 'teach', 'transition', 'use', 'usually', 'work', 'worldviews', 'write']
```

```
missing_words_student = ['body', 'conception', 'debate', 'description', 'firstly', 'historiography', 'meaning', 'metalevel', 'method', 'narrative', 'number', 'philosophy', 'presentation', 'produce', 'question', 'related', 'secondly', 'specific', 'view', 'world', 'writing']
```

```
total_words_original = ['analysis', 'biographical', 'body', 'coherent', 'compete', 'conception', 'debate', 'description', 'development', 'evidence', 'example', 'firstly', 'focus', 'historian', 'historical', 'historiography', 'history', 'interpretation', 'longer', 'meaning', 'medieval', 'metalevel', 'method', 'narrative', 'number', 'past', 'philosophy', 'practice', 'presentation', 'produce', 'professional', 'question', 'related', 'rheumatic', 'secondly', 'series', 'short', 'single', 'specific', 'teach', 'usually', 'view', 'work', 'world', 'writing']
```

```
common_word_count = 24  
total_words_student_count = 43  
missing_words_count = 21  
total_words_original_count = 45
```

c. fraction_of_topics_missed :

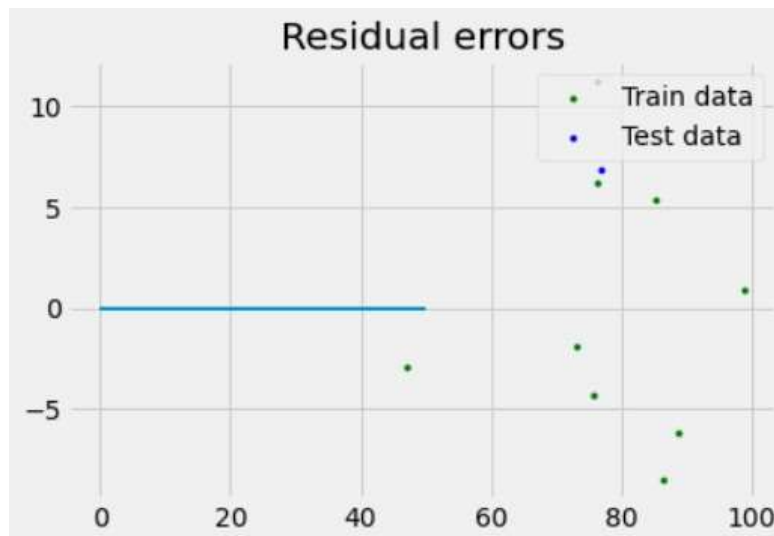
$$(\text{missing_nouns_count} / \text{distinct_nouns_original_count})$$

List of missing nouns

```
['historiography', 'number', 'meaning', 'move', 'historiography', 'meaning', 'body', 'writing', 'philosophy', 'metalevel', 'description', 'conception', 'world', 'view', 'method', 'presentation', 'question']  
no_of_missing_nouns: 17
```

6. Model Development:

- ★ For developing the model we used a training data set containing 10 different answers to a specific question covering various conditions as discussed in the starting.
- ★ We have decided to use a linear regression model to fit a grading scheme based on the 3 features derived earlier and the original grading that we have provided for those 10 answers.
- ★ We have used the “**sklearn**” package and imported “datasets, linear_model and metrics” libraries for creating this linear regression model after splitting the given data set with a test size of 0.1.
- ★ “**matplotlib**” package is used to plot residual errors in both training and testing parts of the training data to provide a basic idea of our model.



- ★ We have obtained coefficients corresponding to 3 features and the intercept and obtained the final linear equation as below.

```
Coefficients: [[102.25013008 -73.81543911 -23.0687793 ]]  
Intercept: [15.09902547]
```

- $\text{predicted_score} = 15.09902547 + (102.25013008 * \text{similarity_score}) + (-73.81543911 * \text{fraction_of_topics_missed}) + (-23.0687793 * \text{fraction_of_new_topics})$.

Final Check with the Test Data:

- ★ We have tested this model on the test data set that we developed and obtained the following results.
- ★ It has provided very few marks (around 30) when facts irrelevant to the question is provided in a sample answer, graded around 50 to 60 when less points are provided or used wrong statements in between. And provided fairly higher marks

(around 90) when the entire matter conveys the same meaning as the original answer despite using several synonyms and pronouns.

★ The above sample grading is reasonably good and satisfactory.

Instructions for using own sample answers:

★ Just replace the **first row** of original_answer and student_answer columns with your sample original and student answers in the **test_data excel file** provided and run the code.

(OR)

★ Add those answers in the row below 10 sample answers which we have provided and **change the row number from 0 to 10 in the code** while loading.

THANK YOU

