* **Subject : -**

PHQ-9 test implementation using BLEU (Bilingual Evaluation Understudy)

Score (A Dynamic Approach).

* **Short Description : -**

To get the depression level of the user, we used a PHQ-9 questionnaire. And according to the answers of each question, we are going to give that answer a rating from 0 to 3. And after adding all the scores together we get the final depression score and we can categorize the depression level of the user. Now to get the score for each sentence we used a concept of BLEU score.

* **PHQ - 9 Form : -**

1. Form link : -

* https://drive.google.com/file/d/1dmkuMM72Pq0qOlVpQn1DvX-G6AdcO29L/view?usp=sharing.

1. Number of questions : -

* Total Number of questions are :- 9.

1. Questions : -
2. How often have you found interest or pleasure in doing things?
3. Are you feeling down or depressed or hopeless?
4. How often do you feel sleepy?
5. How often do you feel tired? Do you feel energetic after eating?
6. How much is your food intake? How much do you eat everyday?
7. How often do you feel yourself being responsible for everything happening around you?
8. Do you find it difficult to concentrate on your studies?
9. Do you feel anxious while talking to people?
10. How many times have you thought of taking your own life or ending yourself?

* **Description : -**

1. The Bilingual Evaluation Understudy Score, or BLEU for short, is a metric for evaluating a generated sentence to a reference sentence.
2. A perfect match results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0.
3. The reference sentences must be provided as a list of sentences where each reference is a list of tokens. The candidate sentence is provided as a list of tokens.
4. Example (Calculating the BLEU score for sentence): -

from nltk.translate.bleu\_score import sentence\_bleu

reference = [['this', 'is', 'a', 'test'], ['this', 'is' 'test']]

candidate = ['this', 'is', 'a', 'test']

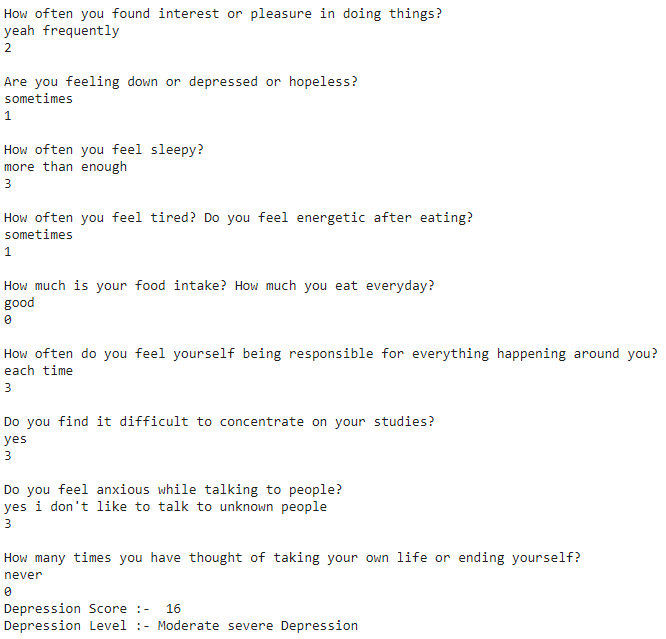
Score = sentence\_bleu(reference, candidate)

print(score)

>> 1.0

1. Now, we need to calculate the score with respect to the answer of the user to the particular question in the PHQ-9 questionnaire.
2. Following is the criteria for which we required a BLEU score to calculate the depression level of the user.
   1. Not at all
   2. Several Days
   3. More than half of the days
   4. Nearly Every Day
3. For this, we created one corpus for each criteria and the references to calculate the BLEU score is as follows : -
   1. reference\_1 = [[["not", "at", "all"], ["no"], ["enough"], ["sufficient"], ["never"]]]
   2. reference\_2 = [[["several", "days"],["few", "days"], ["sometimes"], ["rarely"], ["barely"]]]
   3. reference\_3 = [[["more", "than", "half", "of", "the", "day"], ["mostly"], ["many", "times"], ["very", "frequently"], ["very", "often"], ["many", "of", "the", "days"]]]
   4. reference\_4 = [[["everyday"], ["each", "day"], ["moderate"], ["yes"], ["daily"], ["always"], ["everytime"], ["more", "than", "enough"]]]
4. When the user answers the questions following steps are going to happen with the entered answer.
   1. It will first be pre-processed and all the characters are converted to lowercase.
   2. After that, we store all the words in the answer to one list.
   3. If the length of list is 1 then we simply add “a” at the start of the list. (Disadvantage of BLEU score explained later.)
   4. Now, we calculate BLEU score for each reference and which ever has max value that is our targeted criteria.
   5. If criteria = 1 then add 0 to the depression score. If criteria = 2 then add 1 to the depression score. If criteria = 3 then add 2 to the depression score. If criteria = 4 then add 3 to the depression score.
   6. After asking the entire questionnaire, we have our depression score.
   7. If the depression score is in range 1 to 4 then depression level is “Minimal”. If the depression score is in range 5 to 9 then depression level is “Mild”. If the depression score is in range 10 to 14 then depression level is “Moderately severe”. If the depression score is in range 20 to 27 then depression level is “Severe”.
   8. We can have a maximum depression score of 27 and a minimum depression score is 0.

* **Output of Test Result : -**



* **Advantages of the BLEU Score : -**
  1. It is quick and inexpensive to calculate.
  2. It is easy to understand.
  3. It is language independent.
  4. It correlates highly with human evaluation.
  5. It has been widely adopted.
* **Disadvantages of the BLEU Score : -**
  1. When we answer in one word, the BLEU score will not work as efficiently as we thought of.
  2. BLEU score concept fails when the user answer has length less than 4.
* **Solution for above Disadvantages : -**
  1. We used a method 4 of smoothing function to handle this problem.
  2. It will automatically reassign the weights to each ngram.
  3. Example : -



* 1. And when we have length of the list of words of user answer equal to one then we just add “a” at the start of the list to avoid this problem and we get our desired output.
* **Code Link : -**

1. We used “Google colaboratory” to execute the code.
2. Link of the code : - https://colab.research.google.com/drive/1aIKeZ486B1hNROyskJES8u4hkZB\_pDwt?usp=sharing.