CH5019 Course Project

-P.Praneeth Reddy

PROBLEM STATEMENT

Develop a Machine Learning algorithm to grade a student's answer based on a given standard answer.

CONSTRAINTS

➢ Two paragraphs will be given as inputs .
➢ One paragraph will be the correct answer and the other is the student's answer which is to be graded.
➢ We will return the percentage of similarity between the two answers.

METHOD USED

● We cannot make the machine understand the input and grade even by training it with a lot of training sets.
● Hence we choose to develop an algorithm which works almost accurately for all test sets.
● We are going to use a famous algorithm called the **"Cosine Similarity Algorithm".**
● Along with the Cosine similarity algorithm we will also be using **TF - IDF .**

Cosine Similarity

● Cosine similarity measures the similarity between two vectors of an inner product space.
● This is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction.
● Now we will apply this algorithm to measure the similarity between two paragraphs.

What is the principle behind Cosine Similarity ?

- Cosine similarity measures the cosine of the angle between two vectors in the multi-dimensional space to find similarity between two vectors.

How to apply Cosine Similarity for paragraphs ?

- For applying cosine similarity , we need to convert the paragraphs (strings) into vectors.
- A paragraph can be represented by thousands of attributes, each recording the frequency of a particular word . Thus, each paragraph is an object represented by what is called a "Term-frequency vector".
- For example : Let sentence1 contain the word 'cricket' 3 times , 'hockey' 1 time , 'win' 1 time and 'lose' 0 times and sentence2 contain the word 'cricket' 2 times , 'hockey' 0 times , 'win' 2 times and 'lose' 1 time. Now we can represent sentence1 as a vector (3 , 1 , 1 , 0) and sentence2 as (2 , 0 , 2 , 1).

Why only Cosine Similarity ?

The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance because of the size (like, the word 'cricket' appeared 50 times in one document and 10 times in another) they could still have a smaller angle between them. Smaller the angle, higher the similarity.

What is TF - IDF ?

- TF-IDF stands for " Term frequency-inverse document frequency " .
- Term frequency works by looking at the frequency of a *particular term* you are concerned with relative to the document.
- Term frequency can be measured in multiple ways , we prefer Logarithmically scaled frequency [log(1 + count_of_word)].
- Inverse document frequency looks at how common or uncommon a word is amongst the text.
- IDF is calculated as follows where *t* is the term (word) we are looking to measure the commonness of and *N* is the number of documents (*d*) in the Text (*D*).

$$idf\ (t,\ D) = log\ \left(\frac{N}{count\ (d \in D{:}t \in d)} \right)$$

- The denominator is simply the number of documents in which the term, **t**, appears in.
- TF-IDF is the importance of a term is inversely related to its frequency across documents.
- TF gives us information on how often a term appears in a document and IDF gives us information about the relative rarity of a term in the collection of documents. By multiplying these values together we can get our final TF-IDF value.

$$tf\ idf\ (t,\ d,\ D) = tf\ (t,\ d)\ .\ idf\ (t,\ D)$$

What is the need of using the IDF ?

The reason we need IDF is to help correct for words like "of", "as", "the", etc. since they appear frequently in English . Thus by taking inverse document frequency, we can minimize the weighting of frequent terms while making infrequent terms have a higher impact.

THE CODE

- We will be using the Scikit-learn python library to find TF-IDF and cosine similarity of the given texts.
- **Step 1** : We will import TfidfVectorizer and cosine_similarity functions from the scikit-learn library.
- **Step 2** : The code takes the input paragraphs and adds them to a list and passes the list to the similarity function .
- **Step 3** : The list which is passed down , first undergoes tf-idf function and then the obtained vector undergoes the cosine similarity function and the cosine of the vectors is obtained.
- **Step 4** : Now we print out the cosine value multiplied by 100 to get similarity of the paragraphs in percentage.

SCREENSHOT OF THE CODE

```python
1   from sklearn.feature_extraction.text import TfidfVectorizer
2   from sklearn.metrics.pairwise import cosine_similarity
3
4   def similarity(answers):
5
6       vectorizer = TfidfVectorizer()
7
8       vector = vectorizer.fit_transform(answers)
9
10      cosine_similarities = cosine_similarity(vector[0],vector[1])
11
12      percentage = round(float(cosine_similarities) * 100 , 4)
13
14      print("The student's answer is " + str(percentage) + "% similar to the original answer" )
15
16  correct_answer = input("Enter the correct answer : ")
17  student_answer = input("Enter the student's answer : ")
18  answers = [correct_answer , student_answer]
19
20  similarity(answers)
```

TEST CASES

1. Comparing 2 joint sentences in reverse order.

```
Enter the correct answer : My name is Bruce Wayne and I am Batman
Enter the student's answer : I am batman and my name is Bruce Wayne
The student's answer is 100.0% similar to the original answer
```

2. Comparing Direct speech and Indirect speech.

```
Enter the correct answer : He said , " I had already left "
Enter the student's answer : He said he had already left
The student's answer is 94.8683% similar to the original answer
```

3. Comparing the plot of Naruto from two different websites.

```
Enter the correct answer : Moments prior to Naruto Uzumaki birth, the Nine-Tailed Fox, attacked Konohagaku
re, the Hidden Leaf Village, and wreaked havoc. In order to put an end to the Nine-Tailed Fox rampage, the
 leader of the village, the Fourth Hokage, sacrificed his life and sealed the monstrous beast inside the n
ewborn Naruto. Now, Naruto is a hyperactive and knuckle-headed ninja still living in Konohagakure. Shunned
 because of the Nine-Tailed Fox inside him, Naruto struggles to find his place in the village, while his b
urning desire to become the Hokage of Konohagakure leads him not only to some great new friends, but also
some deadly foes.
Enter the student's answer : Twelve years before the start of the series, the Nine-Tailed Fox attacked Kon
ohagakure destroying much of the village and taking many lives. The leader of the village, the Fourth Hoka
ge, sacrificed his life to seal the Nine-Tailed Fox into a newborn, Naruto Uzumaki. Orphaned by the attack
, Naruto was shunned by the villagers, who out of fear and anger, viewed him as the Nine-Tailed Fox itself
. Though the Third Hokage outlawed speaking about anything related to the Nine-Tailed Fox, the children —
taking their cues from their parents — inherited the same animosity towards Naruto. In his thirst to be ac
knowledged, Naruto vowed he would one day become the greatest Hokage the village had ever seen.
The student's answer is 70.6529% similar to the original answer
```

4. Comparing the plot of two different anime.

```
Enter the correct answer : Following the successful end of the Fourth Shinobi World War, Konohagakure has been enjoying a period of peace, prosperity, and
 extraordinary technological advancement. This is all due to the efforts of the Allied Shinobi Forces and the village's Seventh Hokage, Naruto Uzumaki. No
w resembling a modern metropolis, Konohagakure has changed, particularly the life of a shinobi. Under the watchful eye of Naruto and his old comrades, a n
ew generation of shinobi has stepped up to learn the ways of the ninja. Boruto Uzumaki is often the center of attention as the son of the Seventh Hokage.
Despite having inherited Naruto's boisterous and stubborn demeanor, Boruto is considered a prodigy and is able to unleash his potential with the help of s
upportive friends and family. Unfortunately, this has only worsened his arrogance and his desire to surpass Naruto which, along with his father's busy lif
estyle, has strained their relationship. However, a sinister force brewing within the village may threaten Boruto's carefree life. New friends and familia
r faces join Boruto as a new story begins in Boruto: Naruto Next Generations.
Enter the student's answer : Yuu Izumi leads a high school life filled with one mishap after another. No matter how improbable the situation, unfortunate
events strike him at every turn. In possession of such terrible luck, Izumi enters his second year with a single wish in mind—to spend more time with his
affectionate girlfriend, Micchon Shikimori. Cute, athletic, and caring, Shikimori is immensely popular at their school. But since they began dating a year
 ago, Izumi has witnessed a surprising side to his otherwise adorable girlfriend: when the need arises, she turns incredibly cool! His misfortunes are eas
ily avoided when she is there to protect him with an awe-inspiring look on her face. Charming in every way, she never ceases to make his heart skip a beat
. Unfortunate as he may be, Izumi is sure to see his days of bad luck end thanks to the cute yet cool Shikimori.
The student's answer is 26.7175% similar to the original answer
```

5. Two non-related sentences

```
Enter the correct answer : praneeth is a boy
Enter the student's answer : lahari is a girl
The student's answer is 20.1993% similar to the original answer
```

6. Sentences with different meanings

```
Enter the correct answer : i love you
Enter the student's answer : i hate you
The student's answer is 33.6097% similar to the original answer
```

7. Comparing the plot of Your Name from two different websites

```
Enter the correct answer : Mitsuha Miyamizu, a high school girl, yearns to live the life of a boy in the bustling city of Tokyo , a dream that stands in s
tark contrast to her present life in the countryside. Meanwhile in the city, Taki Tachibana lives a busy life as a high school student while juggling his
part-time job and aiming for a future in architecture as an urban planner. One day, Mitsuha awakens in a room that is not her own and suddenly finds herse
lf living the dream life in Tokyo but in Taki body. Elsewhere, Taki finds himself living Mitsuha's life in the humble countryside. In pursuit of an answer
 to this strange phenomenon, they begin to search for one another. Kimi no Na wa. revolves around Mitsuha and Taki actions, which begin to have a dramatic
 impact on each other lives, weaving them into a fabric held together by fate and circumstance.
Enter the student's answer : Mitsuha Miyamizu is a high school girl living in a countryside town . She yearns for a life in Tokyo as she is sick of living
 in the countryside. At the same time, Taki Tachibana, a high school student, lives in Tokyo and has a great interest in architecture, aiming to become an
 urban planner in the near future.One day, Mitsuha dreams of herself as a boy living a life in the city of Tokyo , while Taki dreams of living as a girl i
n a rural town . As time passes by, the two discover that these are not just dreams, but that they actually swap bodies upon falling asleep. Kimi no Na wa
. revolves around Mitsuha and Taki as they experience the supernatural. With both of them working together to deal with the strange phenomenon, how will t
his impact their everyday life
The student's answer is 68.6691% similar to the original answer
```

8. Comparing the plot of Book and Movie of Your Lie in April

```
Enter the correct answer : Kousei Arima is a child prodigy known as the "Human Metronome" for playing the piano with precision and perfection. Guided by a
 strict mother and rigorous training, Kousei dominates every competition he enters, earning the admiration of his musical peers and praise from audiences.
 When his mother suddenly passes away, the subsequent trauma makes him unable to hear the sound of a piano, and he never takes the stage thereafter. Nowad
ays, Kousei lives a quiet and unassuming life as a junior high school student alongside his friends Tsubaki Sawabe and Ryouta Watari. While struggling to
get over his mother's death, he continues to cling to music. His monochrome life turns upside down the day he encounters the eccentric violinist Kaori Miy
azono, who thrusts him back into the spotlight as her accompanist. Through a little lie, these two young musicians grow closer together as Kaori tries to
fill Kousei's world with color.
Enter the student's answer : At a very young age, Kousei Arima was strictly taught how to play the piano and meticulously follow the score by his mother,
to the point where he dominated every competition he entered with ease. He earned the title of "Human Metronome" for performing almost perfectly. Every mu
sician of his age looked up to him. However, after his mother suddenly dies, he became tone-deaf due to the shock and then disappeared, never to be seen o
n stage since. Two years later, Kousei lives a monotonous life with his childhood friends Tsubaki Sawabe and Ryouta Watari supporting him. He continues to
 cling to music, although performing is still an impossibility for him. This is until his unexpected encounter with Kaori Miyazono, a violinist who perfor
ms freely without the dictations of a score. A story of friendship, love, music, and a single lie, Kousei's life begins to change and gain color as Kaori
helps him to take up music again.
The student's answer is 60.7336% similar to the original answer
```

9. Comparing the plot of Komi san can't communicate season 1 to it's book.

```
Enter the correct answer : Hitohito Tadano is an ordinary boy who heads into his first day of high school with a clear plan: to avoid trouble and do his b
est to blend in with others. Unfortunately, he fails right away when he takes the seat beside the school's madonna : Shouko Komi. His peers now recognize
him as someone to eliminate for a chance to sit next to the most beautiful girl in class. Gorgeous and graceful with long, dark hair, Komi is universally
adored and immensely popular despite her mysterious persona. However, unbeknownst to everyone, she has crippling anxiety and a communication disorder whic
h prevents her from wholeheartedly socializing with her classmates. When left alone in the classroom, a chain of events forces Komi to interact with Tadan
o through writing on the blackboard, as if in a one-way conversation. Being the first person to realize she cannot communicate properly, Tadano picks up t
he chalk and begins to write as well. He eventually discovers that Komi's goal is to make one hundred friends during her time in high school. To this end,
 he decides to lend her a helping hand, thus also becoming her first ever friend.
Enter the student's answer : It's Shouko Komi's first day at the prestigious Itan Private High School, and she has already risen to the status of the scho
ol's Madonna. With long black hair and a tall, graceful appearance, she captures the attention of anyone who comes across her. There's just one problem th
ough despite her popularity, Shouko is terrible at communicating with others. Hitohito Tadano is your average high school boy. With his life motto of "rea
d the situation and make sure to stay away from trouble," he quickly finds that sitting next to Shouko has made him the enemy of everyone in his class . O
ne day, knocked out by accident, Hitohito later wakes up to the sound of Komi's "meow." He lies that he heard nothing, causing Shouko to run away. But bef
ore she can escape, Hitohito surmises that Shouko is not able to talk to others easily . In fact, she has never been able to make a single friend. Hitohit
o resolves to help Shouko with her goal of making one hundred friends so that she can overcome her communication disorder
The student's answer is 60.7619% similar to the original answer
```

10. Comparing two sentences of the same meaning with almost no difference.

```
Enter the correct answer : My name is Praneeth
Enter the student's answer : My name is Praneeth Reddy
The student's answer is 81.818% similar to the original answer
```