

Code Understanding Report

Generated: 2025-05-06 13:21:57

This report presents automated insights based on large language models and code analysis tools.

File: app.py

Summary

- This function converts a given text into a vector that can be used for machine learning tasks. The text is first transformed into a `TfidfVectorizer` object that is then fitted and transformed into a matrix. To convert the text into vectors, the `Text` parameter is taken and the `TfidfVectorizer` is used. This converts the text into a matrix where each row corresponds to a unique word, and each column corresponds to a unique term.

The result is then returned in a 2-D numpy array.

The `TfidfVectorizer` converts the original text into a - This function `similarity(doc1, doc2)` is used to calculate the cosine similarity between two documents `doc1` and `doc2`. Cosine similarity is a measure used for comparing similarity between two non-zero vectors of an inner product space. In this case, it's used for comparing the documents in a document-term matrix form.

The `cosine_similarity` function is a part of scikit-learn. It computes the cosine of the angle between two non-zero vectors. In simpler words, it measures how similar the documents are in terms of - This script is written in Python language. It seems to compute the similarity between two texts (in this case, they could be documents), then add those scores (to be precise, they're compared) into a resultant list of tuples.

This is basically a function that performs a set of checks to find any similarity scores that are possibly plagiarisms. The results are stored in a variable `plagiarism_results`.

Please note that the function `similarity` needs to be a previously defined function that takes two text vectors as input and returns a similarity score. It looks like

Docstring

- `### Code: def vectorize(Text): return TfidfVectorizer().fit_transform(Text).toarray()`

Docstring:

This function takes a list of text documents and returns a matrix of TF-IDF features.

Parameters: - `Text`: A list of strings, where each string is a document.

Returns: - A sparse - `### Code: def similarity(doc1, doc2): return cosine_similarity([doc1, doc2])`

Docstring:

This function calculates the cosine similarity between two documents.

Parameters: - doc1: The first document. - doc2: The second document.

Returns: The cosine similarity between the two documents. - *### Code: def checkplagiarism():
global svectora for studenta, textvectora in svectora: newvectors = svectora.copy() currentindex =
newvectors.index((studenta, textvectora)) del newvectors[currentindex] for studentb, textvectorb
in newvectors: simscore = similarity(textvectora, textvectorb)[0][1] studentpair =
sorted((studenta, studentb)) score = (studentpair[0], studentpair[1], simscore)
plagiarismresults.add(score) return plagiarismresults*

Docstring:

This function checks for plagiarism in a list of text vectors. It does this by comparing each text vector with every other text vector in the list. The similarity between the two text vectors is calculated using the `similarity` function

Code Quality

Tool: `pylint`

Issues: 1`

```
```text * Module tmpqq1f7pzv C:\Users\nmoha\AppData\Local\Temp\tmpqq1f7pzv.py:2:32:  
E0602: Undefined variable 'TfidfVectorizer' (undefined-variable) C:
\Users\nmoha\AppData\Local\Temp\tmpqq1f7pzv.py:2:4: W0612: Unused variable
'vectorize' (unused-variable)
```

---

```
* Module tmpve6zqogp C:\Users\nmoha\AppData\Local\Temp\tmpve6zqogp.py:2:39: E0602:
Undefined variable 'cosine_similarity' (undefined-variable) C:
\Users\nmoha\AppData\Local\Temp\tmpve6zqogp.py:2:4: W0612: Unused variable
'similarity' (unused-variable)
```

---

[AST Parse Error] expected an indented block after function definition on line 1 (line 2) ```

## Conclusion

This codebase focuses on transforming raw text into vectorized formats, such as matrices, which can then be used for machine learning tasks. It uses cosine similarity as a metric to evaluate similarity between documents, providing a way to understand how similar documents are.