Assignment 2

Problem Statement: Implement Single Pass Algorithm for Clustering of files.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

documets=[
        "The sky is blue and beautiful.",
        "Love this blue and beautiful sky!",
        "The quick brown fox jumps over the lazy dog.",
        "A king's breakfast has sausages, ham, and bacon.",
        "I love green eggs, ham, sausages, and bacon!",
        "The brown fox is quick and the blue dog is lazy!",
        "The sky is very blue and the sky is very beautiful today.",
        "The dog is lazy but the black fox is quick."
]
threshold = 0.3

vectorizer = TfidfVectorizer()
doc_vectorizer = vectorizer.fit_transform(documets)

clusters=[]
clusters.append({
        "representative" : doc_vectorizer[0],
        "members" : [documets[0]]
})

for i,doc_vector in enumerate(doc_vectorizer[1:],start=1):
        max_similarity=0
        best_cluster=None
```

```
        for cluster in clusters:
                similarity = cosine_similarity(doc_vector,cluster["representative"])[0][0]
                if similarity > max_similarity:
                  max_similarity = similarity
                  best_cluster = cluster

        if max_similarity >= threshold:
                best_cluster["members"].append(documets[i])
        else:
                clusters.append({
                        "representative" : doc_vector,
                        "members" : [documets[i]]
                })
for idx, cluster in enumerate(clusters):
         print(f"\n Cluster {idx + 1}:")
         for doc in cluster["members"]:
                print(f" - {doc}")
```

Cluster 1:

- The sky is blue and beautiful.
- Love this blue and beautiful sky!
- The sky is very blue and the sky is very beautiful today.

Cluster 2:

- The quick brown fox jumps over the lazy dog.
- The brown fox is quick and the blue dog is lazy!
- The dog is lazy but the black fox is quick.

Cluster 3:

- A king's breakfast has sausages, ham, and bacon.
- I love green eggs, ham, sausages, and bacon!