

GitHub Link:

https://github.com/nehabaddam/Feature_Engineering
(https://github.com/nehabaddam/Feature_Engineering)

Part 1: Executing the text summarization code as in the article

1. Import all necessary libraries

This article is for text summarization using python. Please implement their code first. Then, try to apply their code to your data (5 long or short articles). You might notice that their code did not have the text cleaning except the stop-words. Please refer to the text cleaning methods used in ICE-1 and add appropriate text cleaning methods to the text summarization code. Then, apply the modified code to your data again.

```
In [18]: from nltk.corpus import stopwords
from nltk.cluster.util import cosine_distance
import numpy as np
import networkx as nx
```

2. Generate clean sentences

```
In [19]: def read_article(file_name):
file = open(file_name, "r")
filedata = file.readlines()
article = filedata[0].split(". ")
sentences = []

for sentence in article:
    print(sentence)
    sentences.append(sentence.replace("[^a-zA-Z]", " ").split(" "))
sentences.pop()

return sentences
```

```
In [ ]:
```

3. Similarity matrix

```
In [20]: def sentence_similarity(sent1, sent2, stopwords=None):
    if stopwords is None:
        stopwords = []

    sent1 = [w.lower() for w in sent1]
    sent2 = [w.lower() for w in sent2]

    all_words = list(set(sent1 + sent2))

    vector1 = [0] * len(all_words)
    vector2 = [0] * len(all_words)

    # build the vector for the first sentence
    for w in sent1:
        if w in stopwords:
            continue
        vector1[all_words.index(w)] += 1

    # build the vector for the second sentence
    for w in sent2:
        if w in stopwords:
            continue
        vector2[all_words.index(w)] += 1

    return 1 - cosine_distance(vector1, vector2)
```

```
In [21]: def build_similarity_matrix(sentences, stop_words):
    # Create an empty similarity matrix
    similarity_matrix = np.zeros((len(sentences), len(sentences)))

    for idx1 in range(len(sentences)):
        for idx2 in range(len(sentences)):
            if idx1 == idx2: #ignore if both are same sentences
                continue
            similarity_matrix[idx1][idx2] = sentence_similarity(sentences[idx1], sentences[idx2])

    return similarity_matrix
```

In []:

4. Generate Summary Method

```
In [22]: def generate_summary(file_name, top_n=5):
stop_words = stopwords.words('english')
summarize_text = []

# Step 1 - Read text and split it
sentences = read_article(file_name)

# Step 2 - Generate Similarity Matrix across sentences
sentence_similarity_matrix = build_similarity_matrix(sentences, stop_words)

# Step 3 - Rank sentences in similarity matrix
sentence_similarity_graph = nx.from_numpy_array(sentence_similarity_matrix)
scores = nx.pagerank(sentence_similarity_graph)

# Step 4 - Sort the rank and pick top sentences
ranked_sentence = sorted(((scores[i],s) for i,s in enumerate(sentences)),
print("Indexes of top ranked_sentence order are ", ranked_sentence)

for i in range(top_n):
    summarize_text.append(" ".join(ranked_sentence[i][1]))

# Step 5 - Offcourse, output the summarize text
print("Summarize Text: \n", " ".join(summarize_text))
```

```
In [23]: # Let's begin  
generate_summary( "Article.txt", 2)
```

In an attempt to build an AI-ready workforce, Microsoft announced Intelligent Cloud Hub which has been launched to empower the next generation of students with AI-ready skills

Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services

As part of the program, the Redmond giant which wants to expand its reach and is planning to build a strong developer ecosystem in India with the program will set up the core AI infrastructure and IoT Hub for the selected campuses. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset

This will require more collaborations and training and working with AI. That's why it has become more critical than ever for educational institutions to integrate new cloud and AI technologies

The program is an attempt to ramp up the institutional set-up and build capabilities among the educators to educate the workforce of tomorrow." The program aims to build up the cognitive skills and in-depth understanding of developing intelligent cloud connected solutions for applications across industry. Earlier in April this year, the company announced Microsoft Professional Program In AI as a learning track open to the public

The program was developed to provide job ready skills to programmers who wanted to hone their skills in AI and data science with a series of online courses which featured hands-on labs and expert instructors as well

This program also included developer-focused AI school that provided a bunch of assets to help build AI skills.

Indexes of top ranked_sentence order are [(0.15083257041122708, ['Envisioned', 'as', 'a', 'three-year', 'collaborative', 'program', 'Intelligent', 'Cloud', 'Hub', 'will', 'support', 'around', '100', 'institutions', 'with', 'AI', 'infrastructure', 'course', 'content', 'and', 'curriculum', 'developer', 'support', 'development', 'tools', 'and', 'give', 'students', 'access', 'to', 'cloud', 'and', 'AI', 'services']), (0.13161201335715553, ['The', 'company', 'will', 'provide', 'AI', 'development', 'tools', 'and', 'Azure', 'AI', 'services', 'such', 'as', 'Microsoft', 'Cognitive', 'Services', 'Bot', 'Services', 'and', 'Azure', 'Machine', 'Learning. According', 'to', 'Manish', 'Prakash', 'Country', 'General', 'Manager-PS', 'Health', 'and', 'Education', 'Microsoft', 'India', 'said', '"With', 'AI', 'being', 'the', 'defining', 'technology', 'of', 'our', 'time', 'it', 'is', 'transforming', 'lives', 'and', 'industry', 'and', 'the', 'jobs', 'of', 'tomorrow', 'will', 'require', 'a', 'different', 'skillset']), (0.11403047674961148, ['Earlier', 'in', 'April', 'this', 'year', 'the', 'company', 'announced', 'Microsoft', 'Professional', 'Program', 'In', 'AI', 'as', 'a', 'learning', 'track', 'open', 'to', 'the', 'public']), (0.10721749759953525, ['In', 'an', 'attempt', 'to', 'build', 'an', 'AI-ready', 'workforce', 'Microsoft', 'announced', 'Intelligent', 'Cloud', 'Hub', 'which', 'has', 'been', 'launched', 'to', 'empower', 'the', 'next', 'generation', 'of', 'students', 'with', 'AI-ready', 'skills']), (0.10404298514456578, ['As', 'part', 'of', 'the', 'program', 'the', 'Redmond', 'giant', 'which', 'wants', 'to', 'expand', 'its', 'reach', 'and', 'is', 'planning', 'to', 'build', 'a', 'strong', 'developer', 'ecosystem', 'in', 'India', 'with', 'the', 'program', 'will', 'set', 'up', 'the', 'core', 'AI', 'infrastructure', 'and', 'IoT', 'Hub', 'for', 'the', 'selected', 'campuses']), (0.10031366655994461, ['That's', 'why', 'it', 'has', 'become', 'more', 'critical', 'than', 'ever', 'for', 'educational', 'institutions', 'to', 'integrate', 'new', 'cloud

```
d', 'and', 'AI', 'technologies']], (0.10001137283486655, ['The', 'program',
'is', 'an', 'attempt', 'to', 'ramp', 'up', 'the', 'institutional', 'set-up',
'and', 'build', 'capabilities', 'among', 'the', 'educators', 'to', 'educate',
'the', 'workforce', 'of', 'tomorrow.'], 'The', 'program', 'aims', 'to', 'build',
'up', 'the', 'cognitive', 'skills', 'and', 'in-depth', 'understanding',
'of', 'developing', 'intelligent', 'cloud', 'connected', 'solutions', 'for',
'applications', 'across', 'industry']], (0.09916750119894319, ['This', 'will',
'require', 'more', 'collaborations', 'and', 'training', 'and', 'working',
'with', 'AI']), (0.09277191614415067, ['The', 'program', 'was', 'developed',
'to', 'provide', 'job', 'ready', 'skills', 'to', 'programmers', 'who', 'wanted',
'to', 'hone', 'their', 'skills', 'in', 'AI', 'and', 'data', 'science', 'with',
'a', 'series', 'of', 'online', 'courses', 'which', 'featured', 'hands-on',
'labs', 'and', 'expert', 'instructors', 'as', 'well']])]
```

Summarize Text:

Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset

Part 2: Executing the text summarization code using 5 long/short articles

```
In [24]: generate_summary( "Article1.txt", 2)
```

The Intelligent Cloud Hub program launched by Microsoft aims to empower the next generation of students with AI-ready skills

Over a period of three years, the program will collaborate with around 100 institutions, providing them with AI infrastructure, course content and curriculum, developer support, and development tools

Students will also gain access to cloud and AI services. As part of the program, Microsoft will set up the core AI infrastructure and IoT Hub in selected campuses, expanding its reach and establishing a strong developer ecosystem in India

The company will offer AI development tools and Azure AI services, including Microsoft Cognitive Services, Bot Services, and Azure Machine Learning.

Indexes of top ranked_sentence order are [(0.4127608009480484, ['Students', 'will', 'also', 'gain', 'access', 'to', 'cloud', 'and', 'AI', 'services. As', 'part', 'of', 'the', 'program,', 'Microsoft', 'will', 'set', 'up', 'the', 'core', 'AI', 'infrastructure', 'and', 'IoT', 'Hub', 'in', 'selected', 'campuses,', 'expanding', 'its', 'reach', 'and', 'establishing', 'a', 'strong', 'developer', 'ecosystem', 'in', 'India']), (0.3340969025584788, ['The', 'Intelligent', 'Cloud', 'Hub', 'program', 'launched', 'by', 'Microsoft', 'aims', 'to', 'empower', 'the', 'next', 'generation', 'of', 'students', 'with', 'AI-ready', 'skills']), (0.2531422964934725, ['Over', 'a', 'period', 'of', 'three', 'years,', 'the', 'program', 'will', 'collaborate', 'with', 'around', '100', 'institutions,', 'providing', 'them', 'with', 'AI', 'infrastructure,', 'course', 'content', 'and', 'curriculum,', 'developer', 'support,', 'and', 'development', 'tools'])]

Summarize Text:

Students will also gain access to cloud and AI services. As part of the program, Microsoft will set up the core AI infrastructure and IoT Hub in selected campuses, expanding its reach and establishing a strong developer ecosystem in India. The Intelligent Cloud Hub program launched by Microsoft aims to empower the next generation of students with AI-ready skills

```
In [25]: generate_summary( "Article2.txt", 2)
```

Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, emphasized the transformative power of AI and the changing skillset required for future jobs

Educational institutions need to integrate cloud and AI technologies, and this program aims to enhance institutional capabilities and empower educators to educate the workforce of tomorrow

The goal is to develop cognitive skills and a deep understanding of building intelligent cloud-connected solutions across various industries. In addition to the Intelligent Cloud Hub program, Microsoft had previously introduced the Microsoft Professional Program in AI

This program, open to the public, provides job-ready skills in AI and data science through a series of online courses with hands-on labs and expert instructors

It also includes an AI school focused on helping developers build AI skills. Overall, Microsoft's initiatives in India and globally reflect its commitment to preparing individuals and institutions for the AI-driven future by providing the necessary infrastructure, tools, and training.

Indexes of top ranked_sentence order are [(0.33814158420715, ['The', 'goal', 'is', 'to', 'develop', 'cognitive', 'skills', 'and', 'a', 'deep', 'understanding', 'of', 'building', 'intelligent', 'cloud-connected', 'solutions', 'across', 'various', 'industries. In', 'addition', 'to', 'the', 'Intelligent', 'Cloud', 'Hub', 'program,', 'Microsoft', 'had', 'previously', 'introduced', 'the', 'Microsoft', 'Professional', 'Program', 'in', 'AI']), (0.22492198977425926, ['This', 'program,', 'open', 'to', 'the', 'public,', 'provides', 'job-ready', 'skills', 'in', 'AI', 'and', 'data', 'science', 'through', 'a', 'series', 'of', 'online', 'courses', 'with', 'hands-on', 'labs', 'and', 'expert', 'instructors']), (0.22051739357123604, ['Educational', 'institutions', 'need', 'to', 'integrate', 'cloud', 'and', 'AI', 'technologies,', 'and', 'this', 'program', 'aims', 'to', 'enhance', 'institutional', 'capabilities', 'and', 'empower', 'educators', 'to', 'educate', 'the', 'workforce', 'of', 'tomorrow']), (0.21641903244735428, ['Manish', 'Prakash,', 'Country', 'General', 'Manager-PS,', 'Health', 'and', 'Education,', 'Microsoft', 'India,', 'emphasized', 'the', 'transformative', 'power', 'of', 'AI', 'and', 'the', 'changing', 'skillset', 'required', 'for', 'future', 'jobs'])]

Summarize Text:

The goal is to develop cognitive skills and a deep understanding of building intelligent cloud-connected solutions across various industries. In addition to the Intelligent Cloud Hub program, Microsoft had previously introduced the Microsoft Professional Program in AI. This program, open to the public, provides job-ready skills in AI and data science through a series of online courses with hands-on labs and expert instructors


```
In [26]: generate_summary( "Article3.txt", 2)
```

Newsroom AI, a UK-based company, has introduced a new platform designed to assist news publishers in delivering faster and personalized user experiences similar to Facebook newsfeeds

The platform functions by decoupling content management systems from the user-facing experience and employs algorithmic delivery models, machine learning, and natural language processing to gain insights into user preferences such as content history, location, time of day, and expressed interests

This enables the platform to offer unique content experiences to individual users, catering to the growing demand for content diversity. The technology has been tested by Newsroom AI for over 10 months in collaboration with various digital publishers, resulting in significant improvements such as up to 400% increases in time spent on site and up to six times higher average revenue per user

The platform includes a built-in exchange module that facilitates immediate content trading between like-minded publishers

Additionally, it offers features commonly found in mobile apps, such as swiping between sections and articles, as well as instant loading of article pages.

Indexes of top ranked_sentence order are [(0.27600316301228256, ['This', 'enables', 'the', 'platform', 'to', 'offer', 'unique', 'content', 'experiences', 'to', 'individual', 'users,', 'catering', 'to', 'the', 'growing', 'demand', 'for', 'content', 'diversity.The', 'technology', 'has', 'been', 'tested', 'by', 'Newsroom', 'AI', 'for', 'over', '10', 'months', 'in', 'collaboration', 'with', 'various', 'digital', 'publishers,', 'resulting', 'in', 'significant', 'improvements', 'such', 'as', 'up', 'to', '400%', 'increases', 'in', 'time', 'spent', 'on', 'site', 'and', 'up', 'to', 'six', 'times', 'higher', 'average', 'revenue', 'per', 'user']), (0.2556436326248097, ['The', 'platform', 'functions', 'by', 'decoupling', 'content', 'management', 'systems', 'from', 'the', 'user-facing', 'experience', 'and', 'employs', 'algorithmic', 'delivery', 'models,', 'machine', 'learning,', 'and', 'natural', 'language', 'processing', 'to', 'gain', 'insights', 'into', 'user', 'preferences', 'such', 'as', 'content', 'history,', 'location,', 'time', 'of', 'day,', 'and', 'expressed', 'interests']), (0.2541228080616135, ['The', 'platform', 'includes', 'a', 'built-in', 'exchange', 'module', 'that', 'facilitates', 'immediate', 'content', 'trading', 'between', 'like-minded', 'publishers']), (0.21423039630129417, ['Newsroom', 'AI,', 'a', 'UK-based', 'company,', 'has', 'introduced', 'a', 'new', 'platform', 'designed', 'to', 'assist', 'news', 'publishers', 'in', 'delivering', 'faster', 'and', 'personalized', 'user', 'experiences', 'similar', 'to', 'Facebook', 'newsfeeds'])]

Summarize Text:

This enables the platform to offer unique content experiences to individual users, catering to the growing demand for content diversity. The technology has been tested by Newsroom AI for over 10 months in collaboration with various digital publishers, resulting in significant improvements such as up to 400% increases in time spent on site and up to six times higher average revenue per user. The platform functions by decoupling content management systems from the user-facing experience and employs algorithmic delivery models, machine learning, and natural language processing to gain insights into user preferences such as content history, location, time of day, and expressed interests

```
In [27]: generate_summary( "Article4.txt", 2)
```

Mihai Fanache, the founder and CEO of Newsroom AI, stated that the platform helps publishers save resources by covering topics outside their newsroom's expertise, such as local news and stock markets

It also serves as a distribution platform for independent reporters and specialized bloggers through simple RSS integration

According to Newsroom AI, publishers can significantly expedite their product development cycles, reducing them from months to just 48 hours with the assistance of this platform.

Indexes of top ranked_sentence order are [(0.5, ['Mihai', 'Fanache,', 'the', 'founder', 'and', 'CEO', 'of', 'Newsroom', 'AI,', 'stated', 'that', 'the', 'platform', 'helps', 'publishers', 'save', 'resources', 'by', 'covering', 'topics', 'outside', 'their', 'newsroom's', 'expertise,', 'such', 'as', 'local', 'news', 'and', 'stock', 'markets']), (0.5, ['It', 'also', 'serves', 'as', 'a', 'distribution', 'platform', 'for', 'independent', 'reporters', 'and', 'specialized', 'bloggers', 'through', 'simple', 'RSS', 'integration'])]

Summarize Text:

Mihai Fanache, the founder and CEO of Newsroom AI, stated that the platform helps publishers save resources by covering topics outside their newsroom's expertise, such as local news and stock markets. It also serves as a distribution platform for independent reporters and specialized bloggers through simple RSS integration

```
In [28]: generate_summary( "Article5.txt", 2)
```

McCartney, 80, told the BBC that the technology was used to separate the Beatles' voices from background sounds during the making of director Peter Jackson's 2021 documentary series, "The Beatles: Get Back." The "new" song is set to be released later this year, he said

Jackson was "able to extricate John's voice from a ropery little bit of cassette and a piano," McCartney told BBC radio

"He could separate them with AI, he'd tell the machine 'That's a voice, this is a guitar, lose the guitar'." "So when we came to make what will be the last Beatles record, it was a demo that John had that we worked on," he added

"We were able to take John's voice and get it pure through this AI so then we could mix the record as you would do

It gives you some sort of leeway." McCartney didn't identify the name of the demo, but the BBC and others said it was likely to be an unfinished 1978 love song by Lennon called "Now and Then." The demo was included on a cassette labeled "For Paul" that McCartney had received from Lennon's widow, Yoko Ono, the BBC reported

McCartney described AI technology as "kind of scary but exciting," adding: "We will just have to see where that leads."

Indexes of top ranked_sentence order are [(0.2862548473054012, ['Jackson', 'was', '"able', 'to', 'extricate', 'John's', 'voice', 'from', 'a', 'ropery', 'little', 'bit', 'of', 'cassette', 'and', 'a', 'piano,', 'McCartney', 'told', 'BBC', 'radio']), (0.23191225686999012, ['It', 'gives', 'you', 'some', 'sort', 'of', 'leeway."', 'McCartney', "didn't", 'identify', 'the', 'name', 'of', 'the', 'demo,', 'but', 'the', 'BBC', 'and', 'others', 'said', 'it', 'was', 'likely', 'to', 'be', 'an', 'unfinished', '1978', 'love', 'song', 'by', 'Lennon', 'called', '"Now', 'and', 'Then."', 'The', 'demo', 'was', 'included', 'on', 'a', 'cassette', 'labeled', '"For', 'Paul"', 'that', 'McCartney', 'had', 'received', 'from', 'Lennon's', 'widow,', 'Yoko', 'Ono,', 'the', 'BBC', 'reported']), (0.20075611289231773, ['McCartney,', '80,', 'told', 'the', 'BBC', 'that', 'the', 'technology', 'was', 'used', 'to', 'separate', 'the', 'Beatles"', 'voices', 'from', 'background', 'sounds', 'during', 'the', 'making', 'of', 'director', 'Peter', 'Jackson's', '2021', 'documentary', 'series,', '"The', 'Beatles:', 'Get', 'Back."', 'The', '"new"', 'song', 'is', 'set', 'to', 'be', 'released', 'later', 'this', 'year,', 'he', 'said']), (0.1780648679288586, ['"We', 'were', 'able', 'to', 'take', 'John's', 'voice', 'and', 'get', 'it', 'pure', 'through', 'this', 'AI', 'so', 'then', 'we', 'could', 'mix', 'the', 'record', 'as', 'you', 'would', 'do']), (0.10301191500343265, ['"He', 'could', 'separate', 'them', 'with', 'AI,', "he'd", 'tell', 'the', 'machine', '"That's", 'a', 'voice,', 'this', 'is', 'a', 'guitar,', 'lose', 'the', 'guitar\'.."', '"So', 'when', 'we', 'came', 'to', 'make', 'what', 'will', 'be', 'the', 'last', 'Beatles', 'record,', 'it', 'was', 'a', 'demo', 'that', 'John', 'had', 'that', 'we', 'worked', 'on,', 'he', 'added'])]

Summarize Text:

Jackson was "able to extricate John's voice from a ropery little bit of cassette and a piano," McCartney told BBC radio. It gives you some sort of leeway." McCartney didn't identify the name of the demo, but the BBC and others said it was likely to be an unfinished 1978 love song by Lennon called "Now and Then." The demo was included on a cassette labeled "For Paul" that McCartney had received from Lennon's widow, Yoko Ono, the BBC reported

Part 3: refer to the text cleaning methods used in ICE-1 and add appropriate text cleaning

methods to the text summarization code.

```
In [46]: import re
from nltk.corpus import stopwords
from nltk.cluster.util import cosine_distance
import numpy as np
import networkx as nx

def read_clean_article(file_name):
    file = open(file_name, "r")
    filedata = file.readlines()
    article = filedata[0].split(". ")
    sentences = []

    for sentence in article:
        print(sentence)
        cleaned_sentence = re.sub(r"^[a-zA-Z\s]", "", sentence) # Remove special characters
        cleaned_sentence = cleaned_sentence.lower() # Convert to lowercase
        cleaned_sentence = re.sub(r"\s+", " ", cleaned_sentence) # Remove extra spaces
        sentences.append(cleaned_sentence.split(" "))
        # Text cleaning for each sentence

    sentences.pop()

    return sentences

def generate_clean_summary(file_name, top_n=5):
    stop_words = stopwords.words('english')
    summarize_text = []

    # Step 1 - Read text and split it
    sentences = read_clean_article(file_name)

    # Step 2 - Generate Similarity Matrix across sentences
    sentence_similarity_matrix = build_similarity_matrix(sentences, stop_words)

    # Step 3 - Rank sentences in similarity matrix
    sentence_similarity_graph = nx.from_numpy_array(sentence_similarity_matrix)
    scores = nx.pagerank(sentence_similarity_graph)

    # Step 4 - Sort the rank and pick top sentences
    ranked_sentence = sorted(((scores[i], s) for i, s in enumerate(sentences)),
                             key=lambda x: x[0], reverse=True)
    print("Indexes of top ranked sentence order are ", ranked_sentence)

    for i in range(top_n):
        summarize_text.append(" ".join(ranked_sentence[i][1]))

    # Step 5 - Offcourse, output the summarize text
    print("Summarize Text: \n", ". ".join(summarize_text))
```

```
In [47]: generate_clean_summary( "Article.txt", 2)
```

In an attempt to build an AI-ready workforce, Microsoft announced Intelligent Cloud Hub which has been launched to empower the next generation of students with AI-ready skills

Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services

As part of the program, the Redmond giant which wants to expand its reach and is planning to build a strong developer ecosystem in India with the program will set up the core AI infrastructure and IoT Hub for the selected campuses. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset

This will require more collaborations and training and working with AI. That's why it has become more critical than ever for educational institutions to integrate new cloud and AI technologies

The program is an attempt to ramp up the institutional set-up and build capabilities among the educators to educate the workforce of tomorrow." The program aims to build up the cognitive skills and in-depth understanding of developing intelligent cloud connected solutions for applications across industry. Earlier in April this year, the company announced Microsoft Professional Program in AI as a learning track open to the public

The program was developed to provide job ready skills to programmers who wanted to hone their skills in AI and data science with a series of online courses which featured hands-on labs and expert instructors as well

This program also included developer-focused AI school that provided a bunch of assets to help build AI skills.

Indexes of top ranked_sentence order are [(0.15830140629821587, ['envisioned', 'as', 'a', 'threeyear', 'collaborative', 'program', 'intelligent', 'cloud', 'hub', 'will', 'support', 'around', 'institutions', 'with', 'ai', 'infrastructure', 'course', 'content', 'and', 'curriculum', 'developer', 'support', 'development', 'tools', 'and', 'give', 'students', 'access', 'to', 'cloud', 'and', 'ai', 'services']), (0.12498385558314998, ['the', 'company', 'will', 'provide', 'ai', 'development', 'tools', 'and', 'azure', 'ai', 'services', 'such', 'as', 'microsoft', 'cognitive', 'services', 'bot', 'services', 'and', 'azure', 'machine', 'learningaccording', 'to', 'manish', 'prakash', 'country', 'general', 'managerps', 'health', 'and', 'education', 'microsoft', 'india', 'said', 'with', 'ai', 'being', 'the', 'defining', 'technology', 'of', 'our', 'time', 'it', 'is', 'transforming', 'lives', 'and', 'industry', 'and', 'the', 'jobs', 'of', 'tomorrow', 'will', 'require', 'a', 'different', 'skillset']), (0.11651121979738954, ['as', 'part', 'of', 'the', 'program', 'the', 'redmond', 'giant', 'which', 'wants', 'to', 'expand', 'its', 'reach', 'and', 'is', 'planning', 'to', 'build', 'a', 'strong', 'developer', 'ecosystem', 'in', 'india', 'with', 'the', 'program', 'will', 'set', 'up', 'the', 'core', 'ai', 'infrastructure', 'and', 'iot', 'hub', 'for', 'the', 'selected', 'campuses']), (0.11371803970751426, ['earlier', 'in', 'april', 'this', 'year', 'the', 'company', 'announced', 'microsoft', 'professional', 'program', 'in', 'ai', 'as', 'a', 'learning', 'track', 'open', 'to', 'the', 'public']), (0.10957772461312007, ['the', 'program', 'is', 'an', 'attempt', 'to', 'ramp', 'up', 'the', 'institutional', 'setup', 'and', 'build', 'capabilities', 'among', 'the', 'educators', 'to', 'educate', 'the', 'workforce', 'of', 'tomorrow', 'the', 'program', 'aims', 'to', 'build', 'up', 'the', 'cognitive', 'skills', 'and', 'indepth', 'understanding', 'of', 'developing', 'intelligent', 'cloud', 'connected', 'solutions', 'for', 'applications', 'across', 'industry']), (0.10164086

```
276038953, ['in', 'an', 'attempt', 'to', 'build', 'an', 'aiready', 'workforc
e', 'microsoft', 'announced', 'intelligent', 'cloud', 'hub', 'which', 'has',
'been', 'launched', 'to', 'empower', 'the', 'next', 'generation', 'of', 'stud
ents', 'with', 'aiready', 'skills']], (0.09217104184840202, ['thats', 'why',
'it', 'has', 'become', 'more', 'critical', 'than', 'ever', 'for', 'educational',
'institutions', 'to', 'integrate', 'new', 'cloud', 'and', 'ai', 'technolo
gies']], (0.09210528311814742, ['the', 'program', 'was', 'developed', 'to',
'provide', 'job', 'ready', 'skills', 'to', 'programmers', 'who', 'wanted', 't
o', 'hone', 'their', 'skills', 'in', 'ai', 'and', 'data', 'science', 'with',
'a', 'series', 'of', 'online', 'courses', 'which', 'featured', 'hands-on', 'la
bs', 'and', 'expert', 'instructors', 'as', 'well']], (0.09099056627367119,
['this', 'will', 'require', 'more', 'collaborations', 'and', 'training', 'an
d', 'working', 'with', 'ai']))]
```

Summarize Text:

envisioned as a threeyear collaborative program intelligent cloud hub will support around institutions with ai infrastructure course content and curriculum developer support development tools and give students access to cloud and ai services. the company will provide ai development tools and azure ai services such as microsoft cognitive services bot services and azure machine learning according to manish prakash country general manager ps health and education microsoft india said with ai being the defining technology of our time it is transforming lives and industry and the jobs of tomorrow will require a different skillset

In [48]: `generate_clean_summary("Article1.txt", 2)`

The Intelligent Cloud Hub program launched by Microsoft aims to empower the next generation of students with AI-ready skills

Over a period of three years, the program will collaborate with around 100 institutions, providing them with AI infrastructure, course content and curriculum, developer support, and development tools

Students will also gain access to cloud and AI services. As part of the program, Microsoft will set up the core AI infrastructure and IoT Hub in selected campuses, expanding its reach and establishing a strong developer ecosystem in India

The company will offer AI development tools and Azure AI services, including Microsoft Cognitive Services, Bot Services, and Azure Machine Learning.

```
Indexes of top ranked_sentence order are [(0.4310615905954554, ['students',
'will', 'also', 'gain', 'access', 'to', 'cloud', 'and', 'ai', 'servicesas',
'part', 'of', 'the', 'program', 'microsoft', 'will', 'set', 'up', 'the', 'cor
e', 'ai', 'infrastructure', 'and', 'iot', 'hub', 'in', 'selected', 'campuse
s', 'expanding', 'its', 'reach', 'and', 'establishing', 'a', 'strong', 'devel
oper', 'ecosystem', 'in', 'india']), (0.2971573092823531, ['the', 'intelligen
t', 'cloud', 'hub', 'program', 'launched', 'by', 'microsoft', 'aims', 'to',
'empower', 'the', 'next', 'generation', 'of', 'students', 'with', 'aiready',
'skills']), (0.2717811001221912, ['over', 'a', 'period', 'of', 'three', 'year
s', 'the', 'program', 'will', 'collaborate', 'with', 'around', 'institution
s', 'providing', 'them', 'with', 'ai', 'infrastructure', 'course', 'content',
'and', 'curriculum', 'developer', 'support', 'and', 'development', 'tools'])]
```

Summarize Text:

students will also gain access to cloud and ai services as part of the program microsoft will set up the core ai infrastructure and iot hub in selected campuses expanding its reach and establishing a strong developer ecosystem in india. the intelligent cloud hub program launched by microsoft aims to empower the next generation of students with aiready skills

```
In [49]: generate_clean_summary( "Article2.txt", 2)
```

Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, emphasized the transformative power of AI and the changing skillset required for future jobs

Educational institutions need to integrate cloud and AI technologies, and this program aims to enhance institutional capabilities and empower educators to educate the workforce of tomorrow

The goal is to develop cognitive skills and a deep understanding of building intelligent cloud-connected solutions across various industries. In addition to the Intelligent Cloud Hub program, Microsoft had previously introduced the Microsoft Professional Program in AI

This program, open to the public, provides job-ready skills in AI and data science through a series of online courses with hands-on labs and expert instructors

It also includes an AI school focused on helping developers build AI skills. Overall, Microsoft's initiatives in India and globally reflect its commitment to preparing individuals and institutions for the AI-driven future by providing the necessary infrastructure, tools, and training.

Indexes of top ranked_sentence order are [(0.3259684320300158, ['the', 'goal', 'is', 'to', 'develop', 'cognitive', 'skills', 'and', 'a', 'deep', 'understanding', 'of', 'building', 'intelligent', 'cloudconnected', 'solutions', 'across', 'various', 'industriesin', 'addition', 'to', 'the', 'intelligent', 'cloud', 'hub', 'program', 'microsoft', 'had', 'previously', 'introduced', 'the', 'microsoft', 'professional', 'program', 'in', 'ai']), (0.24959056262294213, ['this', 'program', 'open', 'to', 'the', 'public', 'provides', 'jobready', 'skills', 'in', 'ai', 'and', 'data', 'science', 'through', 'a', 'series', 'of', 'online', 'courses', 'with', 'handson', 'labs', 'and', 'expert', 'instructors']), (0.24522888168522244, ['educational', 'institutions', 'need', 'to', 'integrate', 'cloud', 'and', 'ai', 'technologies', 'and', 'this', 'program', 'aims', 'to', 'enhance', 'institutional', 'capabilities', 'and', 'empower', 'educators', 'to', 'educate', 'the', 'workforce', 'of', 'tomorrow']), (0.17921212366181974, ['manish', 'prakash', 'country', 'general', 'managerps', 'health', 'and', 'education', 'microsoft', 'india', 'emphasized', 'the', 'transformative', 'power', 'of', 'ai', 'and', 'the', 'changing', 'skillset', 'required', 'for', 'future', 'jobs'])]

Summarize Text:

the goal is to develop cognitive skills and a deep understanding of building intelligent cloudconnected solutions across various industriesin addition to the intelligent cloud hub program microsoft had previously introduced the microsoft professional program in ai. this program open to the public provides jobready skills in ai and data science through a series of online courses with handson labs and expert instructors


```
In [50]: generate_clean_summary( "Article3.txt", 2)
```

Newsroom AI, a UK-based company, has introduced a new platform designed to assist news publishers in delivering faster and personalized user experiences similar to Facebook newsfeeds

The platform functions by decoupling content management systems from the user-facing experience and employs algorithmic delivery models, machine learning, and natural language processing to gain insights into user preferences such as content history, location, time of day, and expressed interests

This enables the platform to offer unique content experiences to individual users, catering to the growing demand for content diversity. The technology has been tested by Newsroom AI for over 10 months in collaboration with various digital publishers, resulting in significant improvements such as up to 400% increases in time spent on site and up to six times higher average revenue per user

The platform includes a built-in exchange module that facilitates immediate content trading between like-minded publishers

Additionally, it offers features commonly found in mobile apps, such as swiping between sections and articles, as well as instant loading of article pages.

Indexes of top ranked_sentence order are [(0.30264574996977345, ['this', 'enables', 'the', 'platform', 'to', 'offer', 'unique', 'content', 'experiences', 'to', 'individual', 'users', 'catering', 'to', 'the', 'growing', 'demand', 'for', 'content', 'diversity', 'the', 'technology', 'has', 'been', 'tested', 'by', 'newsroom', 'ai', 'for', 'over', 'months', 'in', 'collaboration', 'with', 'various', 'digital', 'publishers', 'resulting', 'in', 'significant', 'improvements', 'such', 'as', 'up', 'to', 'increases', 'in', 'time', 'spent', 'on', 'site', 'and', 'up', 'to', 'six', 'times', 'higher', 'average', 'revenue', 'per', 'user']), (0.246766226941331, ['the', 'platform', 'includes', 'a', 'built in', 'exchange', 'module', 'that', 'facilitates', 'immediate', 'content', 'trading', 'between', 'likeminded', 'publishers']), (0.22769289512797633, ['the', 'platform', 'functions', 'by', 'decoupling', 'content', 'management', 'systems', 'from', 'the', 'userfacing', 'experience', 'and', 'employs', 'algorithmic', 'delivery', 'models', 'machine', 'learning', 'and', 'natural', 'language', 'processing', 'to', 'gain', 'insights', 'into', 'user', 'preferences', 'such', 'as', 'content', 'history', 'location', 'time', 'of', 'day', 'and', 'expressed', 'interests']), (0.22289512796091934, ['newsroom', 'ai', 'a', 'uk based', 'company', 'has', 'introduced', 'a', 'new', 'platform', 'designed', 'to', 'assist', 'news', 'publishers', 'in', 'delivering', 'faster', 'and', 'personalized', 'user', 'experiences', 'similar', 'to', 'facebook', 'newsfeeds'])]

Summarize Text:

this enables the platform to offer unique content experiences to individual users catering to the growing demand for content diversity the technology has been tested by newsroom ai for over months in collaboration with various digital publishers resulting in significant improvements such as up to increases in time spent on site and up to six times higher average revenue per user. the platform includes a builtin exchange module that facilitates immediate content trading between likeminded publishers

```
In [51]: generate_clean_summary( "Article4.txt", 2)
```

Mihai Fanache, the founder and CEO of Newsroom AI, stated that the platform helps publishers save resources by covering topics outside their newsroom's expertise, such as local news and stock markets

It also serves as a distribution platform for independent reporters and specialized bloggers through simple RSS integration

According to Newsroom AI, publishers can significantly expedite their product development cycles, reducing them from months to just 48 hours with the assistance of this platform.

Indexes of top ranked_sentence order are [(0.5, ['mihai', 'fanache', 'the', 'founder', 'and', 'ceo', 'of', 'newsroom', 'ai', 'stated', 'that', 'the', 'platform', 'helps', 'publishers', 'save', 'resources', 'by', 'covering', 'topics', 'outside', 'their', 'newsrooms', 'expertise', 'such', 'as', 'local', 'news', 'and', 'stock', 'markets']), (0.5, ['it', 'also', 'serves', 'as', 'a', 'distribution', 'platform', 'for', 'independent', 'reporters', 'and', 'specialized', 'bloggers', 'through', 'simple', 'rss', 'integration'])]

Summarize Text:

mihai fanache the founder and ceo of newsroom ai stated that the platform helps publishers save resources by covering topics outside their newsrooms expertise such as local news and stock markets. it also serves as a distribution platform for independent reporters and specialized bloggers through simple rss integration

In [52]: `generate_clean_summary("Article5.txt", 2)`

McCartney, 80, told the BBC that the technology was used to separate the Beatles' voices from background sounds during the making of director Peter Jackson's 2021 documentary series, "The Beatles: Get Back." The "new" song is set to be released later this year, he said

Jackson was "able to extricate John's voice from a ropery little bit of cassette and a piano," McCartney told BBC radio

"He could separate them with AI, he'd tell the machine 'That's a voice, this is a guitar, lose the guitar'." "So when we came to make what will be the last Beatles record, it was a demo that John had that we worked on," he added

"We were able to take John's voice and get it pure through this AI so then we could mix the record as you would do

It gives you some sort of leeway." McCartney didn't identify the name of the demo, but the BBC and others said it was likely to be an unfinished 1978 love song by Lennon called "Now and Then." The demo was included on a cassette labeled "For Paul" that McCartney had received from Lennon's widow, Yoko Ono, the BBC reported

McCartney described AI technology as "kind of scary but exciting," adding: "We will just have to see where that leads."

Indexes of top ranked_sentence order are [(0.24039784741494385, ['jackson', 'was', 'able', 'to', 'extricate', 'johns', 'voice', 'from', 'a', 'ropery', 'little', 'bit', 'of', 'cassette', 'and', 'a', 'piano', 'mccartney', 'told', 'bbc', 'radio']), (0.20207066289647646, ['we', 'were', 'able', 'to', 'take', 'johns', 'voice', 'and', 'get', 'it', 'pure', 'through', 'this', 'ai', 'so', 'then', 'we', 'could', 'mix', 'the', 'record', 'as', 'you', 'would', 'do']), (0.1912305844270659, ['mccartney', 'told', 'the', 'bbc', 'that', 'the', 'technology', 'was', 'used', 'to', 'separate', 'the', 'beatles', 'voices', 'from', 'background', 'sounds', 'during', 'the', 'making', 'of', 'director', 'peter', 'jacksons', 'documentary', 'series', 'the', 'beatles', 'get', 'back', 'the', 'new', 'song', 'is', 'set', 'to', 'be', 'released', 'later', 'this', 'year', 'he', 'said']), (0.1870831649762321, ['he', 'could', 'separate', 'them', 'with', 'ai', 'hed', 'tell', 'the', 'machine', 'thats', 'a', 'voice', 'this', 'is', 'a', 'guitar', 'lose', 'the', 'guitar', 'so', 'when', 'we', 'came', 'to', 'make', 'what', 'will', 'be', 'the', 'last', 'beatles', 'record', 'it', 'was', 'a', 'demo', 'that', 'john', 'had', 'that', 'we', 'worked', 'on', 'he', 'added']), (0.17921774028528137, ['it', 'gives', 'you', 'some', 'sort', 'of', 'leeway', 'mccartney', 'didnt', 'identify', 'the', 'name', 'of', 'the', 'demo', 'but', 'the', 'bbc', 'and', 'others', 'said', 'it', 'was', 'likely', 'to', 'be', 'an', 'unfinished', 'love', 'song', 'by', 'lennon', 'called', 'now', 'and', 'then', 'the', 'demo', 'was', 'included', 'on', 'a', 'cassette', 'labeled', 'for', 'paul', 'that', 'mccartney', 'had', 'received', 'from', 'lennons', 'widow', 'yoko', 'ono', 'the', 'bbc', 'reported'])]

Summarize Text:

jackson was able to extricate johns voice from a ropery little bit of cassette and a piano mccartney told bbc radio. we were able to take johns voice and get it pure through this ai so then we could mix the record as you would do

In []:

1. What are the two main strategies used in text summarization?

Text summarising is the process of extracting essential information from a certain text content. The goal of text summarization is to offer a brief and logical summary of the original text that highlights the most important aspects of it. Summaries can help users quickly understand the

true meaning of a lengthy article, generate previews for search results, or provide a succinct overview of a document.

Two main approaches to text summarization :

1.Extraction Summarization :

This method chooses key sentences to create a summary. It will weigh and rank the most important parts of texts based on their relevance and similarity using statistical or ML techniques. The selected sentences are combined and a summary is generated with these statements.

Extractive summarization preserves the original wording and structure of the sentences, it may lack coherence if the generated words are not correctly connected.

2.Abstractive Summarization:

It generates a summary by understanding the meaning of the text and generating new summarized sentences.

This approach uses natural language processing techniques, such as language models or deep learning models, to assess the input text, interpret its semantic representation, and generate a more human like summary.

Abstractive summarization allows for more flexibility and can result in more cohesive and meaningful summaries but this method is more complicated.

2. Which feature is used in the text summarization code? Explain how to calculate it.

The feature used in the text summarization code is "sentence similarity measurement". The approach computes sentence similarity using the cosine distance metric.

We generate the text summarization code using the feature sentence similarity as shown below:

1. Text Processing: Firstly, The stop words are removed. We can apply other text cleaning methods as well.
2. Sentence Tokenization: Then, we tokenize the sentence i.e we divide it into words or tokens.
3. Similarity Matrix: Then, we generate a similarity matrix. We build vector for sentences. Here we have used cosine similarity to find the similarities between the sentences. The algorithm computes cosine similarity by measuring the cosine of the angle between two vectors. Greater the value for cosine similarity, two vectors are more similar.
4. Ranking Sentences: Now we rank sentences in the similarity matrix based on their similarity score.
5. Sort Rank and pick top sentences: Now we sort all the sentences according to the rank and have selected the top sentences based on the rank.
6. Print the output: Finally, we print the summarized text.

These steps are used to summarize the text.

3. what is the similarity measurement method used in this code?

In this code cosine_distance function from the "nltk.cluster.util" package has been used to find the similarity measurement between the sentences.

In the text summarization, each sentence is represented as a vector, where the elements correspond to the number of times the word appears in the text.

Cosine similarity is used to compare each pair of the vectors. It is an abstractive method that calculates and reports the cosine of the angle between the vectors. Its value ranges from -1 (dissimilar) to 1 (identical/similar). The value of 0 indicates that the vectors are independent of each other. Higher cosine similarity scores indicate greater content similarity.

The function in code uses `1 - cosine_distance(vector1, vector2)` to convert the distance into a similarity score, a higher value indicates higher similarity between two sentences.

4. We know in ICE-1, TF-IDF is used as the text feature. Can we use it in this code?

Yes, we can use TF-IDF in this code as the text feature.

Term Frequency (TF) - Inverse Document Frequency (IDF):

The TF-IDF is used to calculate the importance of a phrase in a document, by considering both the frequency of a term in a document (TF) and the rarity of the phrase in the entire document collection (IDF).

The frequency with which a phrase appears in a particular document is determined by this measure. Words that appear frequently in a document are more likely to be significant to that document, according to the theory.

Inverse document frequency evaluates the significance of a phrase throughout the whole document collection, Words that appear only infrequently throughout the collection are thought to be more useful and informative than those that appear frequently.

We need to change the `build_similarity_matrix` method to use TF-IDF as a text feature in this code. TF-IDF values for each phrase in each sentence would be calculated instead of a basic term frequency representation.

5. Compare the outputs above. Are they the same or not? Please analyze the comparison result.

Output 1: Envisioned as a three-year collaborative program, Intelligent Cloud Hub will support around 100 institutions with AI infrastructure, course content and curriculum, developer support, development tools and give students access to cloud and AI services. The company will provide AI development tools and Azure AI services such as Microsoft Cognitive Services, Bot Services and Azure Machine Learning. According to Manish Prakash, Country General Manager-PS, Health and Education, Microsoft India, said, "With AI being the defining technology of our time, it is transforming lives and industry and the jobs of tomorrow will require a different skillset

Output 2:

envisioned as a threeyear collaborative program intelligent cloud hub will support around institutions with ai infrastructure course content and curriculum developer support development tools and give students access to cloud and ai services. the company will provide ai development tools and azure ai services such as microsoft cognitive services bot services and azure machine learningaccording to manish prakash country general managerps health and education microsoft india said with ai being the defining technology of our time it is transforming lives and industry and the jobs of tomorrow will require a different skillset

The above outputs are not identical, they differ by the text cleaning processes that we have employed.

The punctuations are removed, the Upper case letters are converted to lower case, and the special character are removed. The seconds output is more optimized and summarized and is

In []: