SVKM'S NMIMS
BTech IT
3rd YEAR
2022-23 BATCH
MACHINE LEARNING

Names: Kunj Jariwala, Pradyumn Mohta, Arushi Shah
Roll nos.: A024, A033, A052

TOPIC NAME : Summarisation Techniques

COLAB LINK
SHARE with ruchik.2020@gmail.com

Bert -
https://colab.research.google.com/drive/1k2-dUDBmcgOv7FfAq9dX6LVeNSzdQ4s3

GPT-2 -
https://colab.research.google.com/drive/1Nwkb23kbZCWpJMWcDTx32zDhcQbpjJrc#scrollTo=aPp90N-drso3

Luhn -
https://colab.research.google.com/drive/198J0uBeMNtcceoMK2pxUVEJkw-3HsHpk

LDA -
https://colab.research.google.com/drive/1ETnrQMEB-QtxPDbv2eVE-DgV-N_9DIoB

LSA -
https://colab.research.google.com/drive/1RejVM_eqZy8TAinueNF3KW5ec5dOOafS

LexRank -
https://colab.research.google.com/drive/1F1Yg8m006tokP1p_X8x92K19wRJs3VYX

TECHNIQUES USED :
Summarization techniques are used to condense large amounts of text or data into shorter summaries while still capturing the most important information. There are different types of summarization techniques including extractive and abstractive summarization.

Extractive summarization involves selecting the most relevant sentences or phrases from the original text and presenting them as a summary. This can be done using various techniques such as frequency analysis, graph-based algorithms, and machine learning models.

Abstractive summarization involves generating a summary by understanding the content of the original text and using natural language generation techniques to create a new summary. This can be done using deep learning models like neural networks and transformer models.

Some popular summarization techniques include TextRank, LSA (Latent Semantic Analysis), LDA (Latent Dirichlet Allocation), GPT (Generative Pre-trained Transformer), and BERT (Bidirectional Encoder Representations from Transformers).

Techniques used by us:-

1. LexRank: This is an unsupervised graph-based approach to extractive summarization that uses the concept of eigenvector centrality to identify important sentences in a text. It creates a similarity matrix based on the pairwise cosine similarity between sentences and then applies a PageRank-style algorithm to assign importance scores to each sentence.
2. BERT: Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained language model that can be fine-tuned for various natural language processing tasks, including summarization. It uses a transformer-based neural network architecture to learn contextual representations of words and can generate abstractive summaries by predicting the most important sentences or phrases in a text.
3. GPT-2: Generative Pre-trained Transformer 2 (GPT-2) is a large-scale transformer-based language model trained on a diverse range of internet text. It can be fine-tuned for various natural language processing tasks, including summarization, by conditioning the model to generate a summary of a given text.
4. Luhn: The Luhn algorithm is an extractive summarization technique that ranks sentences based on the frequency of important words or phrases. It works by calculating the frequency of each word in a text and then assigning a score to each sentence based on the presence of these key terms.
5. LSA: Latent Semantic Analysis (LSA) is a statistical natural language processing technique that uses a matrix factorization method to identify relationships between words and concepts in a text. It can be used for extractive summarization by selecting the sentences that contain the most important latent semantic topics in the original text.
6. LDA: Latent Dirichlet Allocation (LDA) is a topic modeling technique used in natural language processing to identify the underlying topics or themes in a large corpus of text. It works by assuming that each document in the corpus is a mixture of several topics and that each topic is a probability distribution over a set of words. LDA can be used for extractive summarization by identifying the most important topics in a text and selecting the sentences that are most representative of these topics.

```
import pandas as pd

data = pd.read_csv('/content/sample_data/covid.csv')
```

```
import nltk
from nltk.corpus import stopwords
from string import punctuation

nltk.download('stopwords')

stop_words = set(stopwords.words('english') + list(punctuation))
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

```
def luhn_score(sentence):
    words = [word for word in sentence.lower().split() if word not in stop_words]
    word_freq = {}
    for word in words:
        if word not in word_freq:
            word_freq[word] = 0
        word_freq[word] += 1

    max_freq = max(word_freq.values())
    score = 0
    for word in word_freq:
        score += (word_freq[word] / max_freq)

    return score
```

```
import nltk

nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

```
def summarize(text, n):
    sentences = nltk.sent_tokenize(text)
    scores = [(i, luhn_score(sentence)) for i, sentence in enumerate(sentences)]
    scores = sorted(scores, key=lambda x: x[1], reverse=True)
    top_n = [sentences[i] for i, score in scores[:n]]

    return ' '.join(top_n)
```

```
text = data['text'][0]
summary = summarize(text, 3)
print(summary)
```

```
COVID-19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence in
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import TruncatedSVD
import pandas as pd
import nltk
nltk.download('stopwords')
nltk.download('punkt')
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

# Load the dataset
data = pd.read_csv('/content/sample_data/covid.csv')

# Preprocess the text
stop_words = set(stopwords.words('english'))
def preprocess(text):
    tokens = word_tokenize(text.lower())
    tokens = [token for token in tokens if token.isalpha() and token not in stop_words]
    return ' '.join(tokens)
data['clean_text'] = data['text'].apply(preprocess)

# Apply LSA summarization on the dataset
n_components = 5
vectorizer = TfidfVectorizer(use_idf=True, max_df=0.5)
X = vectorizer.fit_transform(data['clean_text'])
lsa = TruncatedSVD(n_components=n_components, algorithm='arpack')
lsa.fit(X)
terms = vectorizer.vocabulary_.keys()

for i, text in enumerate(data['text']):
    print(f"Original Text {i}:")
    print(text)
    print()
    print(f"Summary {i}:")
    sentence_scores = lsa.transform(vectorizer.transform([data['clean_text'][i]]))
    sentence_scores = [s[0] for s in sentence_scores]
    summary_sentences = [sent for idx, sent in enumerate(data['clean_text'][i].split('.')) if sentence_scores[idx] > 0.1]
    summary = ' '.join(summary_sentences)
    print(summary)
    print()
```

```
Original Text 0:
COVID-19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence in

Summary 0:
also known coronavirus highly infectious disease rapidly spread across world since emergence late caused virus primarily affects respira

Original Text 1:
The origins of COVID-19 are believed to be linked to a wet market in Wuhan, China, where the first cases were reported in December 2019.

Summary 1:
origins believed linked wet market wuhan china first cases reported december virus quickly spread throughout china beyond march world he

Original Text 2:
The impact of COVID-19 has been felt across all aspects of life, from healthcare to the economy. The healthcare system has been particul

Summary 2:
impact felt across aspects life healthcare economy healthcare system particularly hard hit hospitals healthcare workers overwhelmed numb

Original Text 3:
The pandemic has also had a significant economic impact, with many businesses forced to close or reduce their operations. Unemployment r

Summary 3:
pandemic also significant economic impact many businesses forced close reduce operations unemployment rates soared governments provide f

Original Text 4:
Despite the challenges posed by COVID-19, there have also been positive developments in the fight against the virus. Researchers around

Summary 4:
despite challenges posed also positive developments fight virus researchers around world working tirelessly develop vaccines treatments

Original Text 5:
In addition to vaccines, there have also been significant improvements in testing and tracing capabilities. Rapid diagnostic tests are r

Summary 5:
addition vaccines also significant improvements testing tracing capabilities rapid diagnostic tests widely available making easier ident

Original Text 6:
It is important to note that while vaccines and treatments are promising developments, they are not a silver bullet. The global response
```

```
Summary 6:
important note vaccines treatments promising developments silver bullet global response require ongoing efforts contain spread virus mea

Original Text 7:
In conclusion, COVID-19 has had a profound impact on the world, affecting every aspect of life. While the virus continues to pose signif

Summary 7:
conclusion profound impact world affecting every aspect life virus continues pose significant challenges reasons optimistic future devel

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
import pandas as pd
import gensim
from gensim import corpora
from gensim.models.ldamodel import LdaModel
from gensim.utils import simple_preprocess

# Load the data
data = pd.read_csv('/content/sample_data/covid.csv')

# Preprocess the text
def preprocess(text):
    # Convert the text to lowercase
    text = text.lower()

    # Remove punctuation and numbers
    text = gensim.parsing.preprocessing.strip_numeric(gensim.parsing.preprocessing.strip_punctuation(text))

    # Tokenize the text
    tokens = gensim.utils.simple_preprocess(text)

    # Remove stop words
    stop_words = gensim.parsing.preprocessing.STOPWORDS
    tokens = [token for token in tokens if token not in stop_words]

    # Stem the tokens
    stemmer = gensim.parsing.PorterStemmer()
    tokens = [stemmer.stem(token) for token in tokens]

    # Join the tokens back into a string
    clean_text = ' '.join(tokens)

    return clean_text

data['clean_text'] = data['text'].apply(preprocess)

# Create a dictionary and corpus for LDA modeling
doc_clean = [doc.split() for doc in data['clean_text']]
dictionary = corpora.Dictionary(doc_clean)
doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

# Train the LDA model
num_topics = 5
lda = LdaModel(doc_term_matrix, num_topics=num_topics, id2word=dictionary, passes=50)

# Print the topics
for i, topic in lda.show_topics(num_topics=num_topics, formatted=False):
    print('Topic {}: {}'.format(i, ', '.join([word for word, _ in topic])))

# Get the topic distribution for each document
topic_dist = [lda.get_document_topics(doc) for doc in doc_term_matrix]

# Get the dominant topic for each document
dominant_topic = [max(topics, key=lambda x: x[1]) for topics in topic_dist]

# Add the dominant topic to the dataframe
data['dominant_topic'] = [topic[0] for topic in dominant_topic]
data['dominant_topic_prob'] = [topic[1] for topic in dominant_topic]
```

```
Topic 0: pandem, busi, impact, challeng, govern, signific, individu, provid, economi, help
Topic 1: trace, test, identifi, develop, improv, infect, quickli, vaccin, individu, signific
Topic 2: covid, spread, contain, effort, global, viru, diseas, origin, emerg, affect
Topic 3: world, covid, organ, china, global, viru, pandem, ongo, respons, pose
Topic 4: vaccin, healthcar, treatment, develop, viru, us, covid, requir, measur, distanc
```

```
!pip install sumy
import pandas as pd
import nltk
from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from sumy.parsers.plaintext import PlaintextParser
from sumy.summarizers.lex_rank import LexRankSummarizer
from sumy.nlp.tokenizers import Tokenizer

# Load the dataset
df = pd.read_csv('/content/sample_data/covid.csv')

# Preprocess the text data
corpus = []
stemmer = PorterStemmer()
for i in range(len(df)):
    text = df.iloc[i]['text']
    sentences = sent_tokenize(text)
    stem_sentences = []
    for sent in sentences:
        words = word_tokenize(sent)
        stem_words = [stemmer.stem(word.lower()) for word in words]
        stem_sentences.append(' '.join(stem_words))
    corpus.append(' '.join(stem_sentences))

# Create a document-term matrix
vectorizer = CountVectorizer()
doc_term_matrix = vectorizer.fit_transform(corpus)

# Calculate the cosine similarity matrix
similarity_matrix = cosine_similarity(doc_term_matrix)

# Apply LexRank algorithm
summarizer = LexRankSummarizer()
parser = PlaintextParser.from_string(" ".join(df['text'].values.tolist()), Tokenizer("english"))
summary = summarizer(parser.document, sentences_count=3)

# Print the summary
for sentence in summary:
    print(sentence)
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: sumy in /usr/local/lib/python3.9/dist-packages (0.11.0)
Requirement already satisfied: pycountry>=18.2.23 in /usr/local/lib/python3.9/dist-packages (from sumy) (22.3.5)
Requirement already satisfied: docopt<0.7,>=0.6.1 in /usr/local/lib/python3.9/dist-packages (from sumy) (0.6.2)
Requirement already satisfied: requests>=2.7.0 in /usr/local/lib/python3.9/dist-packages (from sumy) (2.27.1)
Requirement already satisfied: nltk>=3.0.2 in /usr/local/lib/python3.9/dist-packages (from sumy) (3.8.1)
Requirement already satisfied: breadability>=0.1.20 in /usr/local/lib/python3.9/dist-packages (from sumy) (0.1.20)
Requirement already satisfied: chardet in /usr/local/lib/python3.9/dist-packages (from breadability>=0.1.20->sumy) (4.0.0)
Requirement already satisfied: lxml>=2.0 in /usr/local/lib/python3.9/dist-packages (from breadability>=0.1.20->sumy) (4.9.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.9/dist-packages (from nltk>=3.0.2->sumy) (4.65.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.9/dist-packages (from nltk>=3.0.2->sumy) (1.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.9/dist-packages (from nltk>=3.0.2->sumy) (8.1.3)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.9/dist-packages (from nltk>=3.0.2->sumy) (2022.10.31)
Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from pycountry>=18.2.23->sumy) (67.6.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests>=2.7.0->sumy) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests>=2.7.0->sumy) (1.26.15)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests>=2.7.0->sumy) (2.0.12)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests>=2.7.0->sumy) (2022.12.7)
In this essay, we will explore the origins of COVID-19, its global impact, and the efforts being made to contain and treat the disease.
In addition to vaccines, there have also been significant improvements in testing and tracing capabilities.
The global response to COVID-19 will require ongoing efforts to contain the spread of the virus through measures such as social distanci
```

```
import pandas as pd
import torch
from transformers import BertTokenizer, BertModel, BertForMaskedLM

# Load the dataset
data = pd.read_csv('/content/sample_data/covid.csv')

# Load the BERT tokenizer
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')

# Load the BERT model for sequence-to-sequence classification
model = BertForMaskedLM.from_pretrained('bert-base-uncased')

# Set the device to run the model on
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
model.to(device)

# Apply summarization on the dataset
for text in data['text']:
    input_ids = tokenizer.encode(text, return_tensors='pt', max_length=512).to(device)
    summary_ids = model.generate(input_ids, num_beams=4, max_length=100, early_stopping=True)
    summary = tokenizer.decode(summary_ids.squeeze(), skip_special_tokens=True)
    print(summary)
```

```
Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertForMaskedLM: ['cls.seq_relationship.weight
- This IS expected if you are initializing BertForMaskedLM from the checkpoint of a model trained on another task or with another archit
- This IS NOT expected if you are initializing BertForMaskedLM from the checkpoint of a model that you expect to be exactly identical (i
Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly trunca
covid - 19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence i
the origins of covid - 19 are believed to be linked to a wet market in wuhan, china, where the first cases were reported in december 201
the impact of covid - 19 has been felt across all aspects of life, from healthcare to the economy. the healthcare system has been partic
the pandemic has also had a significant economic impact, with many businesses forced to close or reduce their operations. unemployment r
despite the challenges posed by covid - 19, there have also been positive developments in the fight against the virus. researchers arour
in addition to vaccines, there have also been significant improvements in testing and tracing capabilities. rapid diagnostic tests are r
it is important to note that while vaccines and treatments are promising developments, they are not a silver bullet. the global response
in conclusion, covid - 19 has had a profound impact on the world, affecting every aspect of life. while the virus continues to pose sign
```

```
!pip install transformers
import torch
from transformers import pipeline, set_seed
import pandas as pd

# Load the data
data = pd.read_csv('/content/sample_data/covid.csv')

# Set random seed for reproducibility
set_seed(42)

# Instantiate the pipeline for summarization
summarizer = pipeline("summarization")

# Define a function to generate summaries using GPT
def generate_summary_gpt(text, num_sentences=2):
    # Summarize the text using GPT
    summary = summarizer(text, max_length=120, min_length=30, do_sample=False, num_beams=2, length_penalty=0.6)
    # Extract the top N sentences from the summary
    top_sentences = sorted(summary[0]['summary_text'].split('. '), key=len)[:num_sentences]
    # Join the top sentences to create the final summary
    final_summary = '. '.join(top_sentences)
    return final_summary

# Generate summaries using GPT
for i, row in data.iterrows():
    summary = generate_summary_gpt(row['text'], num_sentences=3)
    print(f"Article {i+1} summary:\n{summary}\n")
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.9/dist-packages (4.27.4)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in /usr/local/lib/python3.9/dist-packa
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.9/dist-packages (from transfor
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.9/dist-packages (from transformers)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.9/dist-packages (from transformers
Requirement already satisfied: huggingface-hub<1.0,>=0.11.0 in /usr/local/lib/python3.9/dist-packages (
Requirement already satisfied: requests in /usr/local/lib/python3.9/dist-packages (from transformers) (
Requirement already satisfied: filelock in /usr/local/lib/python3.9/dist-packages (from transformers) (
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.9/dist-packages (from transf
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.9/dist-packages (from transformers
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.9/dist-packages (fr
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests->t
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from re
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from reque
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (fro
No model was supplied, defaulted to sshleifer/distilbart-cnn-12-6 and revision a4f8f3e (https://huggin
Using a pipeline without specifying a model name and revision in production is not recommended.
```

Downloading (…)lve/main/config.json: 100%	1.80k/1.80k [00:00<00:00, 53.8kB/s]
Downloading pytorch_model.bin: 100%	1.22G/1.22G [00:11<00:00, 115MB/s]
Downloading (…)okenizer_config.json: 100%	26.0/26.0 [00:00<00:00, 808B/s]
Downloading (…)olve/main/vocab.json: 100%	899k/899k [00:00<00:00, 6.69MB/s]
Downloading (…)olve/main/merges.txt: 100%	456k/456k [00:00<00:00, 4.24MB/s]

```
Your max_length is set to 120, but you input_length is only 98. You might consider decreasing max_lengt
Your max_length is set to 120, but you input_length is only 79. You might consider decreasing max_lengt
Article 1 summary:
It is caused by the SARS-CoV-2 virus and primarily affects the respiratory system ..  COVID-19, also kr

Your max_length is set to 120, but you input_length is only 79. You might consider decreasing max_lengt
Article 2 summary:
The virus quickly spread throughout China and beyond, and by March 2020, the World Health Organization
```

OUTPUT COMPARISN:

1. Luhn:

COVID-19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence in late 2019. It is caused by the SARS-CoV-2 virus and primarily affects the respiratory system, leading to symptoms such as fever, cough, and difficulty breathing. In this essay, we will explore the origins of COVID-19, its global impact, and the efforts being made to contain and treat the disease.

2. LSA:

Summary 0:

also known coronavirus highly infectious disease rapidly spread across world since emergence late caused virus primarily affects respiratory system leading symptoms fever cough difficulty breathing essay explore origins global impact efforts made contain treat disease

Summary 1:

origins believed linked wet market wuhan china first cases reported december virus quickly spread throughout china beyond march world health organization declared global pandemic since infected millions people worldwide claimed lives hundreds thousands

Summary 2:

impact felt across aspects life healthcare economy healthcare system particularly hard hit hospitals healthcare workers overwhelmed number patients requiring treatment governments implemented range measures try contain spread virus including lockdowns social distancing widespread use masks

Summary 3:

pandemic also significant economic impact many businesses forced close reduce operations unemployment rates soared governments provide financial support individuals businesses help survive effects pandemic economy still unknown clear recovery slow challenging

Summary 4:

despite challenges posed also positive developments fight virus researchers around world working tirelessly develop vaccines treatments several breakthroughs recent months vaccines approved emergency use many countries mass vaccination programs underway

Summary 5:

addition vaccines also significant improvements testing tracing capabilities rapid diagnostic tests widely available making easier identify isolate infected individuals contact tracing apps also developed help identify contain outbreaks quickly

Summary 6:

important note vaccines treatments promising developments silver bullet global response require ongoing efforts contain spread virus measures social distancing improved hygiene practices

Summary 7:

conclusion profound impact world affecting every aspect life virus continues pose significant challenges reasons optimistic future development vaccines treatments provides hope able control eventually overcome pandemic however require ongoing efforts collaboration individuals governments organizations around world ensure successful response global crisis.

3. LDA:
Topic 0: pandem, busi, impact, challeng, govern, signific, individu, provid, economi, help
Topic 1: trace, test, identifi, develop, improv, infect, quickli, vaccin, individu, signific
Topic 2: covid, spread, contain, effort, global, viru, diseas, origin, emerg, affect
Topic 3: world, covid, organ, china, global, viru, pandem, ongo, respons, pose
Topic 4: vaccin, healthcar, treatment, develop, viru, us, covid, requir, measur, distanc

4. LexRank:
In this essay, we will explore the origins of COVID-19, its global impact, and the efforts being made to contain and treat the disease.
In addition to vaccines, there have also been significant improvements in testing and tracing capabilities.
The global response to COVID-19 will require ongoing efforts to contain the spread of the virus through measures such as social distancing, mask-wearing, and improved hygiene practices.

5. BERT:
covid - 19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence in late 2019. it is caused by the sars - cov - 2 virus and primarily affects the respiratory system, leading to symptoms such as fever, cough, and difficulty breathing. in this essay, we will explore the origins of covid - 19, its global impact, and the efforts being made to contain and treat the disease. influenza influenza is
the origins of covid - 19 are believed to be linked to a wet market in wuhan, china, where the first cases were reported in december 2019. the virus quickly spread throughout china and beyond, and by march 2020, the world health organization declared a global pandemic. covid - 19 has since infected millions of people worldwide and claimed the lives of hundreds of thousands.... the resulting. the virus of the covid virus is currently also being studied and a potential…
these measures include prevention and control measures. prevention include : improved human - di -cing, improved hiv prevention. prevention..... and these vaccines, these are very effective technologies.
in conclusion, covid - 19 has had a profound impact on the world, affecting every aspect of life. while the virus continues to pose significant challenges, there are reasons to be optimistic about the future. the development of vaccines and treatments provides hope that we will be able to control and eventually overcome the pandemic. however, it will require ongoing efforts and collaboration from individuals, governments, and organizations around the world to ensure a successful response to this global crisis. this has...

6. GPT-2:
Article 1 summary:
It is caused by the SARS-CoV-2 virus and primarily affects the respiratory system .. COVID-19, also known as the coronavirus, is a highly infectious disease that has rapidly spread across the world since its emergence in late 2019
Article 2 summary:
The virus quickly spread throughout China and beyond, and by March 2020, the World Health Organization declared a global pandemic .. The origins of COVID-19 are believed to be linked to a wet market in Wuhan, China, where the first cases were reported in December 2019

Article 3 summary:
Governments have implemented a range of measures to try and contain the spread of the virus .. The impact of COVID-19 has been felt across all aspects of life, from healthcare to the economy . The healthcare system has been particularly hard hit, with hospitals and healthcare workers overwhelmed
Article 4 summary:
The long-term effects of the pandemic on the economy are still unknown, but it is clear that the recovery will be slow and challenging . Unemployment rates have soared, and governments have had to provide financial support to individuals and businesses to help them survive .
Article 5 summary:
Researchers around the world are working tirelessly to develop vaccines and treatments . Vaccines have been approved for emergency use in many countries, and mass vaccination programs are underway .
Article 6 summary:
Contact tracing apps have also been developed to help identify and contain outbreaks quickly .. Rapid diagnostic tests are now widely available, making it easier to identify and isolate infected individuals
Article 7 summary:
Global response to COVID-19 will require ongoing efforts to contain the spread of the virus through measures such as social distancing, mask-wearing, and improved hygiene practices .
Article 8 summary:
The development of vaccines and treatments provides hope that we will be able to control and eventually overcome the pandemic . However, it will require ongoing efforts and collaboration from individuals, governments, and organizations around the world to ensure a successful response to this global crisis .

CONCLUSION:

Summarization techniques have become increasingly important in various natural language processing tasks, including document summarization, news summarization, and summarization for chatbots and virtual assistants.

Our project on summarization techniques has provided us with a deep understanding of the various approaches used to generate concise summaries of longer texts. We have explored extraction-based methods such as LexRank, Luhn, and LSA, as well as more advanced models like BERT and GPT-2, which can generate abstractive summaries by rewriting the original text in a more concise form. We also investigated LDA, a topic modeling technique that can be used for extractive summarization.

Our findings suggest that the choice of summarization technique depends on the specific use case and the requirements of the user. Extraction-based methods are useful for generating summaries that retain important details from the original text, while abstractive methods can produce more concise summaries that are easier to read but may omit some information. Moreover, our comparison of different techniques highlights the tradeoff between the quality of the summary and the complexity of the method.

In conclusion, summarization techniques offer a powerful tool for quickly understanding large amounts of information and extracting the most important points from a text. By comparing and contrasting different techniques, we have gained insights into the strengths and weaknesses of each approach and provided guidance on the tradeoffs involved in selecting a particular method. Ultimately, the choice of technique will depend on the needs of the user and the specific context in which the summarization is being applied.

REFERENCES :

https://pypi.org/project/bert-extractive-summarizer/
https://betterprogramming.pub/how-to-summarize-text-with-googles-t5-4dd1ae6238b6
https://www.turing.com/kb/5-powerful-text-summarization-techniques-in-python