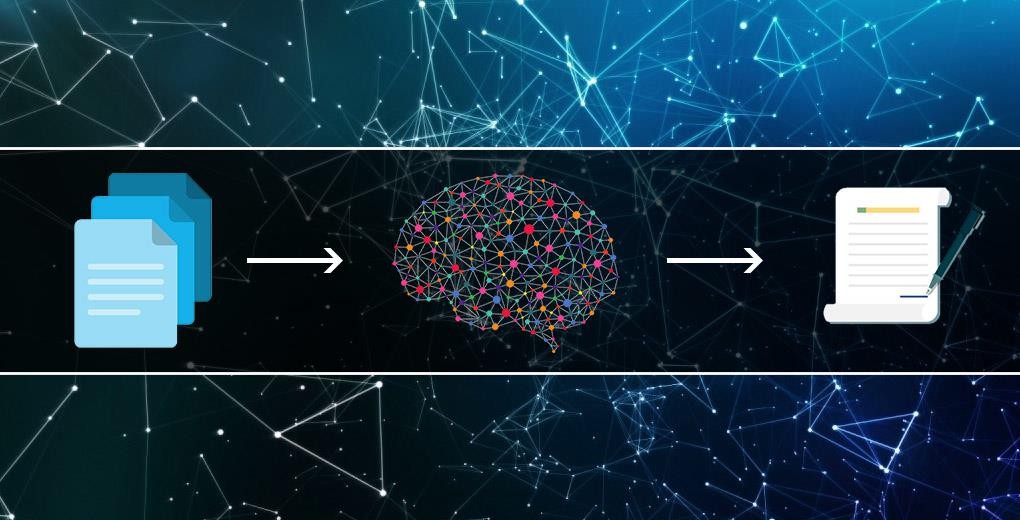
TEXT SUMMARIZATION USING NGRAMS



**TABLE OF CONTENTS**

**ABSTRACT---------------------------------------------------------------------------------------------------------------------- 02**

**INTRODUCTION--------------------------------------------------------------------------------------------------------------- 03**

**ABOUT THE PROJECT-------------------------------------------------------------------------------------------------------- 10**

**TEXT SUMMARIZATION USING N-GRAM PROJECT SURVEY------------------------------------------------------ 11**

**SYSTEM ANALYSIS----------------------------------------------------------------------------------------------------------- 13**

**BLOCK DIAGRAM------------------------------------------------------------------------------------------------------------- 17**

**FLOW CHART------------------------------------------------------------------------------------------------------------------ 18**

**ALGORITHM------------------------------------------------------------------------------------------------------------------- 21**

**CODE SCREEN CODES------------------------------------------------------------------------------------------------------- 25**

**ABSTRACT**

#### This project is about Text Summarization Using N-grams

This project addresses the challenge of efficiently summarizing large volumes of text by developing a text summarization model that leverages N-gram analysis. By extracting and analyzing contiguous sequences of words (N-grams) from textual data, the project aims to generate concise and coherent summaries that retain the essential information from the original content. This approach enhances the effectiveness of information retrieval systems, providing users with quick and accurate summaries, thereby improving user engagement and satisfaction with textual content.

In the rapidly evolving field of natural language processing (NLP), the ability to summarize large texts accurately and efficiently is crucial for providing users with valuable insights. This study focuses on developing and evaluating a model for text summarization that utilizes N-grams to capture the most significant phrases and sentences from a given text. By analyzing a comprehensive dataset of textual information, we aim to enhance the accuracy and coherence of generated summaries, making them more useful for users.

In the domain of text processing, providing users with concise and informative summaries is essential for efficient information consumption. This project aims to develop a predictive model for text summarization using N-grams. By analyzing an extensive dataset and generating N-grams (such as bigrams and trigrams), we seek to identify key phrases and sentences that effectively summarize the original text. Various techniques, including frequency analysis, TF-IDF, and sentence scoring, are employed to build the summarization model.

The performance of these models is evaluated using metrics such as precision, recall, and ROUGE scores. Our results indicate that combining N-gram analysis with statistical techniques significantly enhances the summary quality, with N-grams providing the most informative and coherent summaries. This project not only contributes to improving text summarization systems but also provides deeper insights into the factors influencing summary quality, ultimately enabling more efficient and enjoyable text consumption experiences

### Introduction

#### What is Artificial Intelligence (AI)?

Artificial Intelligence (AI) is a branch of computer science focused on creating systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, understanding natural language, perception, and even exhibiting creativity.

#### Key Areas of AI:

1. **Machine Learning (ML):** A subset of AI where algorithms are used to find patterns or insights in data. The systems "learn" from data and improve over time without being explicitly programmed for specific tasks. Common techniques include supervised learning, unsupervised learning, and reinforcement learning.
2. **Natural Language Processing (NLP):** The ability of a machine to understand, interpret, and generate human language. Applications include chatbots, language translation, sentiment analysis, and speech recognition.
3. **Computer Vision**: The capability of AI systems to interpret and make decisions based on visual input. This includes image and video recognition, object detection, and facial recognition.

#### What is Machine Learning (ML)?

Machine Learning (ML) is a subset of Artificial Intelligence (AI) that involves the development of algorithms and statistical models that enable computers to learn from and make predictions or decisions based on data. Instead of being explicitly programmed to perform a specific task, machine learning systems improve their performance over time as they are exposed to more data.

#### Key Concepts in Machine Learning:

1. **Data:** The foundational element for ML, which includes the information the system learns from. Data can be structured (like databases) or unstructured (like text and images).
2. **Algorithms:** These are the mathematical and computational procedures used by ML systems to learn from data. Examples include decision trees, neural networks, and support vector machines.
3. **Model:** A model is the output generated by the ML algorithm after training on data. It represents the learned patterns and can be used to make predictions.
4. **Training:** The process of feeding data into an ML algorithm to help it learn. During training, the model adjusts its parameters to minimize errors.
5. **Testing**: Evaluating the model's performance on a separate set of data that was not used during training to assess its accuracy and generalization.
6. **Features:** Individual measurable properties or characteristics of the data used by the model for learning.
7. **Labels:** In supervised learning, labels are the known outcomes or categories associated with the training data.

Types of Machine Learning:

1. **Supervised Learning:**
   * The algorithm learns from labeled data, where the input data and the corresponding correct output are provided.
   * Common applications include classification (e.g., spam detection) and regression (e.g., predicting house prices).
2. Unsupervised Learning:
   * The algorithm learns from unlabeled data, identifying patterns and structures within the data.
   * Common applications include clustering (e.g., customer segmentation) and dimensionality reduction (e.g., principal component analysis).
3. Semi-Supervised Learning:
   * A combination of supervised and unsupervised learning, where the algorithm learns from a small amount of labeled data and a larger amount of unlabeled data.
4. Reinforcement Learning:
   * The algorithm learns by interacting with an environment, receiving feedback in the

Key Techniques and Algorithms:

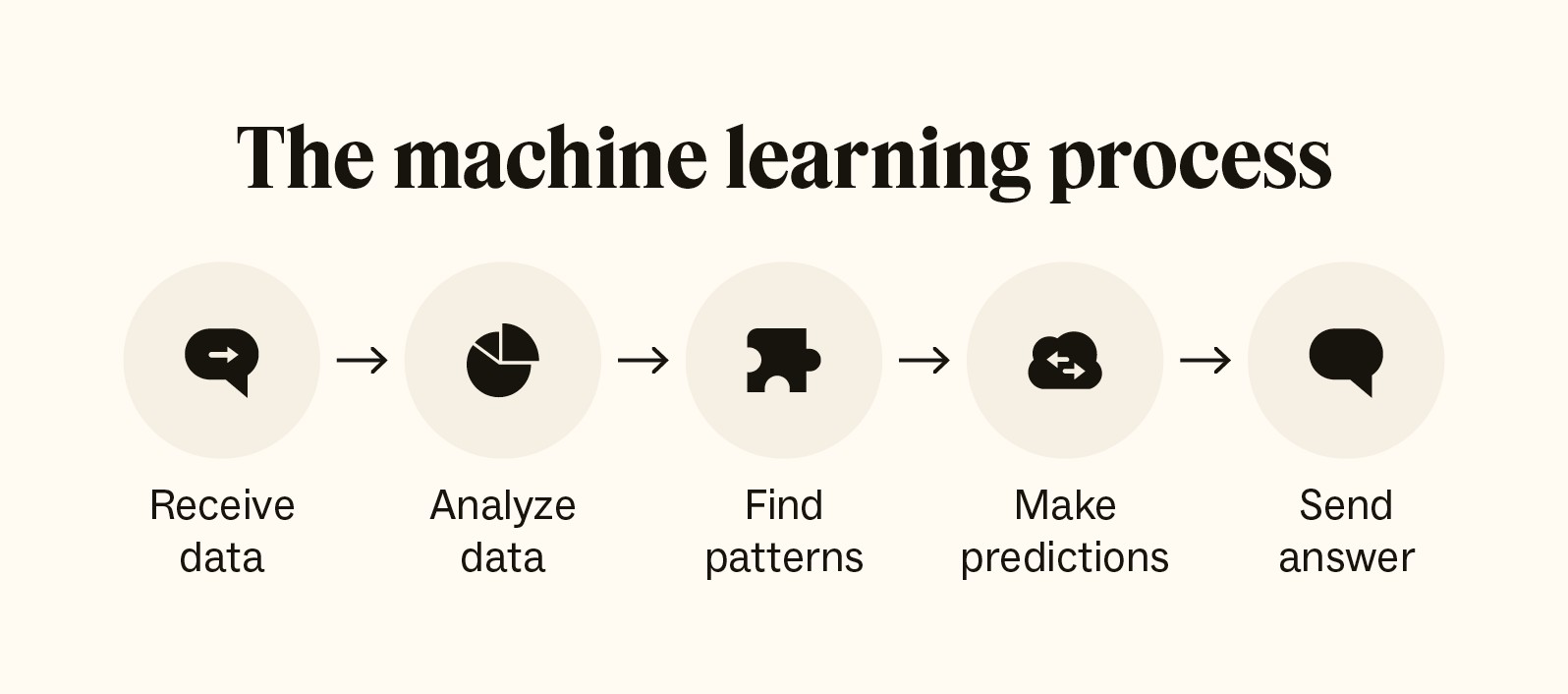
* **Linear Regression**: Used for predicting a continuous outcome based on one or more input variables.
* **Logistic Regression**: Used for binary classification problems.
* **Decision Trees:** A tree-like model used for classification and regression tasks.
* **Random Forests:** An ensemble method that uses multiple decision trees to improve accuracy and prevent overfitting.
* **Support Vector Machines (SVM):** A classification technique that finds the optimal hyperplane separating different classes.
* **Neural Networks:** A set of algorithms modeled after the human brain, used for complex pattern recognition tasks.
* **K-Means Clustering:** An unsupervised learning algorithm for partitioning data into clusters.

Applications of Machine Learning:

* Healthcare: Disease diagnosis, personalized treatment plans, and medical image analysis.
* **Finance:** Credit scoring, fraud detection, and algorithmic trading.
* **Retail:** Recommendation systems, inventory management, and customer segmentation.
* **Transportation:** Autonomous vehicles, traffic prediction, and route optimization.
* **Entertainment:** Content recommendation on streaming platforms, video and music suggestions.

Challenges in Machine Learning:

* **Data Quality:** ML models require high-quality, relevant data for accurate predictions.
* **Overfitting and Underfitting:** Ensuring the model generalizes well to new data without being too specific or too simple.
* **Interpretability**: Understanding how complex models, especially deep learning models, make decisions.
* **Scalability:** Handling large datasets and computational requirements efficiently.
* **Bias and Fairness:** Addressing biases in the data and ensuring fair outcomes.

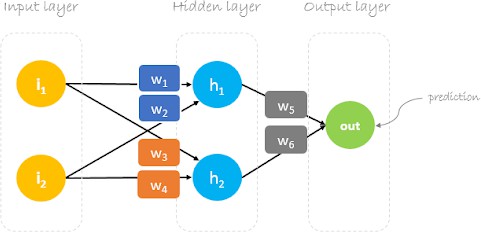


#### What is Deep Learning (DL)?

Deep Learning (DL) is a subset of Machine Learning (ML) that focuses on using neural networks with many layers (hence "deep") to model complex patterns in data. These neural networks, known as deep neural networks (DNNs), are designed to mimic the structure and function of the human brain, allowing them to learn from large amounts of data.

Key Concepts in Deep Learning:

1. **Neural Networks:**
   * Neurons: Basic units of a neural network that take inputs, apply a function (usually nonlinear), and produce an output.
   * Layers: Neural networks consist of multiple layers:
     + Input Layer: The first layer that receives the input data.
     + Hidden Layers: Intermediate layers where data transformation occurs. The depth of the network refers to the number of hidden layers.
     + Output Layer: The final layer that produces the prediction or classification.
2. Activation Functions:
   * Functions applied to the output of each neuron to introduce nonlinearity into the model, enabling it to learn complex patterns.
   * Common activation functions include ReLU (Rectified Linear Unit), Sigmoid, and Tanh.
3. Backpropagation:
   * A training algorithm used to minimize the error by adjusting the weights of the neurons. It involves calculating the gradient of the loss function with respect to each



weight by the chain rule, propagating errors backward from the output to the inputlayer.

1. Loss Function:
   * A function that measures the difference between the predicted output and the actual target. The goal is to minimize this loss during training.
   * Common loss functions include Mean Squared Error (MSE) for regression tasks and Cross-Entropy Loss for classification tasks.
2. Optimization Algorithms:
   * Methods used to adjust the weights of the network to minimize the loss function.
   * Common optimization algorithms include Stochastic Gradient Descent (SGD), Adam, and RMSprop.

#### What is Natural Language Processing (NLP)?

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. It involves the development of algorithms and models that enable machines to understand, interpret, generate, and respond to human language in a meaningful way. NLP combines computational linguistics, computer science, and statistical methods to process and analyze large amounts of natural language data.

Key Concepts in NLP:

1. **Tokenization:**
   * The process of breaking down text into smaller units called tokens, such as words, phrases, or sentences.
2. Parsing:
   * Analyzing the grammatical structure of a sentence to understand the relationships between words and phrases. This includes syntactic parsing (analyzing sentence structure) and semantic parsing (extracting meaning from text).
3. Named Entity Recognition (NER):
   * Identifying and classifying named entities in text, such as people, organizations, locations, dates, and other proper nouns.
4. Part-of-Speech (POS) Tagging:
   * Assigning parts of speech (e.g., nouns, verbs, adjectives) to each word in a sentence.
5. Sentiment Analysis:
   * Determining the sentiment or emotional tone of a piece of text, such as positive, negative, or neutral.
   * Machine Translation: Automatically translating text from one language to another.
6. * Automatically translating text from one language to another.
7. Text Classification:
   * Assigning predefined categories or labels to text based on its content. Common applications include spam detection and topic classification.
8. Speech Recognition:
   * Converting spoken language into written text.
9. Language Generation:
   * Creating human-like text based on a given input, such as text summarization, chatbots, and content creation.

## ABOUT THE PROJECT

#### Project Definition

The Text Summarization Using N-grams project aims to address the challenge of efficiently summarizing large volumes of text by leveraging N-gram analysis. In an age where information overload is prevalent, providing concise and coherent summaries of textual content is crucial for effective communication and quick understanding. This project focuses on developing a model that uses N-grams to identify and extract key phrases and sentences, enabling the generation of meaningful summaries. The primary goal is to enhance the readability and utility of large texts by producing accurate and contextually relevant summaries.

### Proposed Solution

To tackle this challenge, we propose using N-gram analysis combined with statistical text processing techniques as the core approach for our summarization model. N-grams, which are contiguous sequences of words (e.g., bigrams, trigrams), will be extracted from the text to capture significant patterns and phrases. We will employ various methods such as frequency analysis, Term Frequency- Inverse Document Frequency (TF-IDF), and sentence scoring to identify the most important parts of the text.

The summarization model will be developed to analyze the N-grams in the context of the entire document, ranking sentences based on their relevance and contribution to the overall summary. The performance of the model will be evaluated using metrics such as ROUGE scores, precision, and recall to ensure the quality and coherence of the generated summaries. By integrating N-gram analysis with these techniques, we aim to create a robust summarization tool that enhances text comprehension and usability.

#### Objective

The objective of this project is to develop a text summarization model that effectively utilizes N-grams to produce concise and informative summaries of large texts. By leveraging N-gram analysis and statistical methods, the project aims to improve the efficiency of information retrieval and enhance the readability of textual content. This will enable users to quickly grasp key points and insights from extensive documents, ultimately improving their ability to process and understand information.

### Text Summarization Using N-grams Project Survey

#### Theoretical Background

The Text Summarization Using N-grams project is grounded in the principles of natural language processing (NLP) and statistical text analysis. Text summarization involves generating concise and coherent summaries from larger texts, which is essential for effective information retrieval. N-gram analysis, a key technique in NLP, involves extracting contiguous sequences of N words from text data to capture significant patterns and context. By analyzing these N-grams, we can identify and rank important phrases and sentences, which aids in producing meaningful summaries. The use of N-grams allows the summarization model to effectively capture key aspects of the text and improve summary

##### quality.

Existing Systems with Drawbacks

Current text summarization systems employ various techniques, including extractive and abstractive methods. Extractive summarization selects and compiles significant sentences or phrases from the original text, while abstractive summarization generates new sentences that convey the text's main ideas. Existing systems often rely on advanced machine learning models, such as deep learning-based approaches or complex algorithms, to perform summarization.

Drawbacks of Existing Systems:

1. **Complexity**: Advanced models, especially those using deep learning, can be computationally intensive and challenging to interpret, making it difficult to understand how summaries are generated.
2. **Resource-Intensive**: Deep learning models often require significant computational resources and large amounts of training data, which may not be feasible for all applications.
3. **Inconsistent Quality**: The quality of summaries can vary significantly depending on the model and the dataset, leading to inconsistencies in summary coherence and relevance.
4. **Data Dependency**: Models relying on extensive training data may struggle with generating summaries for texts that differ from the data they were trained on.

Proposed System with Features

The proposed system utilizes N-gram analysis combined with statistical text processing techniques for text summarization. N-grams are used to capture important phrases and patterns within the text, which are then used to generate summaries. The system leverages frequency analysis, Term Frequency-Inverse Document Frequency (TF-IDF), and sentence scoring to identify and rank the most relevant parts of the text.

Features of the Proposed System:

1. **N-gram Extraction**: The system extracts N-grams (e.g., bigrams, trigrams) to identify significant phrases and patterns in the text.
2. **Frequency Analysis**: N-gram frequencies and TF-IDF scores are used to determine the importance of phrases and sentences in the context of the entire document.
3. **Sentence Ranking**: Sentences are scored and ranked based on their relevance and contribution
4. **Coherence and Conciseness**: The summarization model aims to produce summaries that are both coherent and concise, providing users with meaningful and readable content.

Advantages of the Proposed System

1. **Enhanced Interpretability**: The use of N-grams and statistical techniques offers clear insights into how summaries are generated, making the results more understandable and actionable.
2. **Improved Summary Quality**: By analyzing N-grams and incorporating statistical measures, the model can generate summaries that accurately reflect the main points of the text.
3. **Computational Efficiency**: N-gram analysis combined with statistical methods is computationally efficient and can handle large texts effectively, making it suitable for real-time applications.
4. **Adaptability**: The model can be easily adapted to different types of texts and updated with new data, allowing for continuous improvement in summary quality.
5. **Simplicity**: Compared to more complex models, the N-gram-based approach is relatively straightforward to implement and understand, facilitating easier integration into existing systems.

In summary, the proposed N-gram-based text summarization system aims to improve the efficiency and quality of text summarization by leveraging N-gram analysis and statistical techniques. This approach provides a transparent and effective solution for generating meaningful and coherent summaries from large volumes of text.

## SYSTEM ANALYSIS

System analysis is conducted to study and evaluate the text summarization system using N-grams to ensure that all components work efficiently and meet their objectives. This process involves defining the functional and non-functional requirements to guide the development and implementation of the system.

Specification

**Functional Requirements**

The following are the functional requirements of our text summarization system using N-grams:

* **Text Preprocessing**: The system must include functionality for preprocessing text, including tokenization, stopword removal, and normalization.
* **N-gram Extraction**: The system should be able to extract N-grams (e.g., bigrams, trigrams) from the input text to identify significant phrases and patterns.
* **Frequency Analysis**: The system must perform frequency analysis and compute Term Frequency-Inverse Document Frequency (TF-IDF) scores to evaluate the importance of different N-grams.
* **Sentence Scoring and Ranking**: The system should score and rank sentences based on their relevance and contribution to the overall summary, using the N-gram and statistical analysis results.
* **Summary Generation**: The system must generate coherent and concise summaries based on the ranked sentences, ensuring that the most important information is included.
* **User Interface**: The system should provide an interface for users to input text and receive generated summaries, making it accessible and user-friendly.

Non-Functional Requirements

The following are the non-functional requirements of our text summarization system using N-grams:

* **Maintainability**: The system should be designed to facilitate easy updates and maintenance. Code should be modular and well-documented to accommodate future changes and enhancements.
* **Robustness**: The system must be robust enough to handle various types of text inputs, including different lengths and formats. It should be able to manage errors and exceptions gracefully.
* **Reliability**: The system should consistently perform text summarization tasks with high accuracy and reliability. It must handle input text correctly and generate meaningful summaries without significant errors.
* **Size**: The size of the application should be optimized to ensure efficient performance. The system should be lightweight and avoid excessive memory usage to maintain high efficiency.
* **Speed**: The system should provide fast processing times for text summarization. Given the efficiency of N-gram analysis, the system should be able to handle large texts and generate

In summary, the text summarization system using N-grams is designed to efficiently preprocess text, extract meaningful N-grams, analyze their importance, and generate coherent summaries. It must meet both functional and non-functional requirements to ensure high performance, maintainability, and user satisfaction.

Software Requirements

One of the most difficult tasks is that, the selection of the software, once system requirement is known that is determining whether a particular software package fits the requirements.

|  |  |
| --- | --- |
| **Programming Language** | **Python** |
| **Technology** | **Jupyter** |
| **Operating System** | **Windows 11** |
| **Browser** | **Google Chrome** |

Hardware Requirements

The selection of hardware is very important in the existence and proper working of any software. In the selection of hardware, the size and the capacity requirements are also important.

|  |  |
| --- | --- |
| **Processor** | **Intel Core** |
| **RAM Capacity** | **4GB** |
| **Hardisk** | **512 GB** |
| **I/O Devices** | **Keyboard, Mouse, Monitor** |

### Module Description for Text Summarization Using N-grams

For implementing the text summarization system using N-grams, the project is divided into the following modules:

1. Data Collection & Pre-processing
2. N-gram Extraction & Frequency Analysis
3. Summary Generation
4. Evaluation & Visualization
5. Data Collection & Pre-processing

**Data Collection**: Data collection involves gathering textual data from various sources, such as articles, reports, or web pages. This data serves as the raw input for summarization.

**Pre-processing**: Data pre-processing involves transforming raw text into a format suitable for analysis. This includes:

* + **Tokenization**: Splitting text into individual words or tokens.
  + **Stopword Removal**: Eliminating common words that do not contribute significant meaning (e.g., "the", "and").
  + **Normalization**: Converting text to a consistent format, such as lowercasing, removing punctuation, and stemming or lemmatizing words.

Python Libraries:

* + **NLTK**: The Natural Language Toolkit (NLTK) provides tools for text processing tasks like tokenization, stopword removal, and normalization.
  + **Spacy**: Spacy is another NLP library used for efficient tokenization and lemmatization.

1. N-gram Extraction & Frequency Analysis

**N-gram Extraction**: This module involves extracting N-grams (e.g., bigrams, trigrams) from the pre- processed text. N-grams are contiguous sequences of N words that capture patterns and context within the text.

**Frequency Analysis**: After extracting N-grams, perform frequency analysis to determine the importance of each N-gram. This involves calculating:

* + **Term Frequency (TF)**: The number of times an N-gram appears in the text.
  + **Term Frequency-Inverse Document Frequency (TF-IDF)**: A statistical measure that evaluates the importance of an N-gram in the context of the entire corpus.

Python Libraries:

* + **Pandas**: For handling and analyzing data in DataFrames, including frequency calculations.
  + **Scikit-learn**: Provides utilities for calculating TF-IDF scores.

1. Summary Generation

**Sentence Scoring & Ranking**: Based on N-gram frequencies and TF-IDF scores, score and rank

**Summary Compilation**: Combine the selected sentences into a coherent summary that captures the essential information from the original text.

Python Libraries:

* + **NLTK/Spacy**: For scoring and ranking sentences based on N-gram importance.
  + **Gensim**: May be used for additional text summarization techniques and comparison.

1. Evaluation & Visualization

**Evaluation**: Assess the quality of generated summaries using metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation), which compares the overlap of N-grams between the generated summary and reference summaries.

**Visualization**: Create visualizations to illustrate the performance of the summarization model and the distribution of N-grams in the text.

Python Libraries:

* + **Matplotlib**: For plotting graphs and visualizations, such as N-gram distributions and summary performance metrics.
  + **Seaborn**: For creating more complex and aesthetically pleasing visualizations.

Implementation Tools:

* + **NLTK (Natural Language Toolkit)**: Provides tools for text processing, including tokenization, stopword removal, and frequency analysis.
  + **Scikit-learn**: Used for TF-IDF calculations and other machine learning tasks related to text processing.
  + **Gensim**: An additional library for text summarization and topic modeling, useful for comparison and enhancement.
  + **Pandas**: Manages and processes data efficiently, handling the frequency analysis and text data manipulation.
  + **Matplotlib/Seaborn**: Visualizes results, such as summary quality and N-gram distributions, to understand and communicate findings.

## BLOCK DIAGRAM:

Block diagram illustrating the text summarization process using n-grams:

Sentence Scoring

- Score Sentences Based on N-gram Frequencies

Sentence Ranking and Thresholding

-Rank Sentences by Score

- Select Top-Scoring Sentences

Frequency Analysis

- Calculate N-gram Frequencies

N-gram Generation

- Unigrams, Bigrams, Trigrams

Text Preprocessing

-Tokenization

-Stop Word Removal

-Stemming and Lemmatization

Input Text Document

Generated Summary

Summary Extraction

- Extract and Concatenate

-Selected Sentences

###### Detailed Description:

1. **Input Text Document:** The original text document that needs to be summarized.
2. Text Preprocessing:
   * Tokenization: Splitting the text into individual words or tokens.
   * Stop Word Removal: Removing common words that do not add significant meaning.
   * Stemming and Lemmatization: Reducing words to their root or canonical forms.
3. **N-gram Generation**: Creating sequences of contiguous words (n-grams) from the preprocessed text. This can include unigrams, bigrams, trigrams, etc.
4. **Frequency Analysis**: Calculating the frequency of each n-gram to identify the most common and significant ones.
5. **Sentence Scoring**: Evaluating the importance of each sentence based on the presence and frequency of n-grams. Sentences with higher scores are considered more relevant.
6. **Sentence Ranking and Thresholding**: Ranking sentences based on their scores and applying a threshold to select the top-scoring sentences for the summary.
7. **Summary Extraction**: Extracting and concatenating the selected top-scoring sentences to form the final summary.
8. **Generated Summary**: The final output, which is a concise summary of the original text document.

This block diagram provides a visual representation of the steps involved in creating a text summarization system using n-grams, highlighting the flow from the input document to the generated summary.

###### FLOW CHART:

Start

│

▼

Input Text Document

│

▼

Text Preprocessing

│

|── Tokenization

│

|── Stop Word Removal

│

|── Stemming and Lemmatization

│

▼

N-gram Generation

│

|── Generate Unigrams

│

|── Generate Bigrams

│

└── Generate Trigrams

│

▼

Frequency Analysis

│

|── Calculate N-gram Frequencies

│

▼

Sentence Scoring

│

|── Score Sentences Based on N-gram Frequencies

│

▼

Sentence Ranking and Thresholding

│

|── Rank Sentences by Score

│

|──Apply Threshold to Select Top-Scoring Sentences

│

▼

Summary Extraction

│

|── Extract and Concatenate Top-Scoring Sentences

│

▼

Evaluation (ROUGE Metrics and Human Evaluation)

│

▼

Generated Summary

│

▼

End

# Algorithm: N-gram Text Generation

#### Input:

Article Text (article\_text): The input text from which N-grams will be extracted. N-gram Size (words): The number of words in each N-gram (e.g., 3 for tri-grams).

Number of Sequences (num\_sequences): Number of text sequences to generate (e.g., 5). Maximum Length (max\_length): Maximum length of each generated sequence (e.g., 100 words).

#### Output:

Generated Text Sequences: Each sequence is based on the N-grams and generated text from the given article.

#### Libraries Required:

NLTK (Natural Language Toolkit): For tokenizing the text. Import: import nltk

Random: For selecting random starting points and next words. Import: import random

#### Steps:

Text Tokenization:

Action: Convert article\_text into a list of tokens (words). Library Function: nltk.word\_tokenize(article\_text) Result: words\_tokens - List of words from the article.

Build N-gram Dictionary:

Initialize: Create an empty dictionary ngrams.

##### Loop:

Iterate over the indices from 0 to len(words\_tokens) - words.

Extract: Form the N-gram sequence seq from words\_tokens[i:i+words]. Check: If seq is not already in ngrams:

Initialize ngrams[seq] as an empty list.

Append: Add the next word words\_tokens[i + words] to ngrams[seq].

Result: ngrams - A dictionary mapping each N-gram sequence to a list of possible next words. Generate Text Sequences:

Repeat for num\_sequences:

Initialization:

Select: Randomly choose a starting position number in words\_tokens.

Form: Initial sequence curr\_sequence from words\_tokens[number:number + words]. Initialize: output with curr\_sequence.

Loop for up to max\_length:

Check: If curr\_sequence is not in ngrams:

Break the loop.

Retrieve: List of possible next words from ngrams[curr\_sequence]. Select: Randomly choose the next word from possible\_words.

Update: Append the selected word to output.

Update: Adjust curr\_sequence to include the newly added word. Print: Output the generated text sequence.

#### Conditions:

Text Length: Ensure words\_tokens has enough length for forming N-grams. Empty N-grams: If ngrams is empty, no text can be generated.

Max Length: Limit text length to max\_length words to prevent excessive output.

#### Notes:

Randomness: The output text will vary between runs due to random selection of starting points and next words.

#### Dataset: Any textual dataset or corpus can be used as article\_text.

This algorithm provides a framework for generating text sequences based on N-grams, utilizing tokenization and dictionary-based lookups to produce coherent and contextually relevant outputs.

# Flow Chart Of Algorithm

ms



For each of num sequences



Initalize output with curr\_sequence



For up to max length

import nltk import random import string

text\_file = open ("dataset.txt","r",encoding="utf-8") article\_text=text\_file.read()

text\_file.close()

article\_text=article\_text.lower()

article\_text= article\_text.translate(str.maketrans('', '', string.punctuation))

nltk.download('punkt')

[nltk\_data] Downloading package punkt to /root/nltk\_data... [nltk\_data] Unzipping tokenizers/punkt.zip.

True

ngrams = {} words = 3

*# Now word tokenization should work* words\_tokens = nltk.word\_tokenize(article\_text) for i in range(len(words\_tokens)-words):

seq = ' '.join(words\_tokens[i:i+words]) if seq not in ngrams.keys():

ngrams[seq] = [] ngrams[seq].append(words\_tokens[i+words])

print ("ngrams dictionary")

for key, value in ngrams.items() : print (key, value,)

ngrams dictionary

mr narendra damodardas ['modi']

narendra damodardas modi ['is', 'was', 'was'] damodardas modi is ['the']

modi is the ['present', 'glorious'] is the present ['and']

the present and ['15th'] present and 15th ['indian'] and 15th indian ['prime'] 15th indian prime ['minister'] indian prime minister ['he'] prime minister he ['has'] minister he has ['been']

he has been ['serving']

has been serving ['our'] been serving our ['nation'] serving our nation ['since'] our nation since ['26th'] nation since 26th ['may'] since 26th may ['2014'] 26th may 2014 ['from']

may 2014 from ['the'] 2014 from the ['year'] from the year ['2001'] the year 2001 ['to'] year 2001 to ['2014'] 2001 to 2014 ['before'] to 2014 before ['taking']

2014 before taking ['over'] before taking over ['delhi'] taking over delhi ['he'] over delhi he ['served'] delhi he served ['the']

he served the ['role'] served the role ['of'] the role of ['honourable'] role of honourable ['chief']

of honourable chief ['minister']

honourable chief minister ['of'] chief minister of ['gujarat'] minister of gujarat ['he']

of gujarat he ['is'] gujarat he is ['a']

he is a ['member', 'prominent', 'beacon'] is a member ['of']

a member of ['the']

member of the ['parliament'] of the parliament ['mp'] the parliament mp ['who']

parliament mp who ['represents'] mp who represents ['the']

who represents the ['city'] represents the city ['of'] the city of ['varanasi'] city of varanasi ['he'] of varanasi he ['is'] varanasi he is ['the']

he is the ['leader', 'first'] is the leader ['of']

the leader of ['the'] leader of the ['popular'] of the popular ['bharatiya']

the popular bharatiya ['janata']

popular bharatiya janata ['party'] bharatiya janata party ['bjp'] janata party bjp ['in']

party bjp in ['the'] bjp in the ['2014']

in the 2014 ['general']

the 2014 general ['election'] 2014 general election ['bjp'] general election bjp ['led'] election bjp led ['by']

bjp led by ['narendra'] led by narendra ['modi'] by narendra modi ['gained'] narendra modi gained ['the'] modi gained the ['majority'] gained the majority ['in'] the majority in ['the'] majority in the ['lok']

in the lok ['sabha'] the lok sabha ['this'] lok sabha this ['was'] sabha this was ['the'] this was the ['first'] was the first ['such'] the first such ['major'] first such major ['win'] such major win ['for'] major win for ['a']

win for a ['political'] for a political ['party'] a political party ['since']

political party since ['1984'] party since 1984 ['all'] since 1984 all ['about'] 1984 all about ['narendra'] all about narendra ['modi'] about narendra modi ['early'] narendra modi early ['life'] modi early life ['prime'] early life prime ['minister'] life prime minister ['narendra']

prime minister narendra ['modi', 'modi', 'modis']

minister narendra modi ['was', 'decided'] narendra modi was ['born', 'designated'] modi was born ['in', 'on']

was born in ['a']

born in a ['lowermiddleclass'] in a lowermiddleclass ['family'] a lowermiddleclass family ['at']

lowermiddleclass family at ['vadnagar'] family at vadnagar ['gujarat']

at vadnagar gujarat ['he'] vadnagar gujarat he ['had'] gujarat he had ['a']

he had a ['keen']

had a keen ['interest'] a keen interest ['in']

keen interest in ['politics'] interest in politics ['since'] in politics since ['the'] politics since the ['early'] since the early ['days']

the early days ['of'] early days of ['his'] days of his ['childhood'] of his childhood ['after']

his childhood after ['completing'] childhood after completing ['his'] after completing his ['higher'] completing his higher ['education'] his higher education ['in'] higher education in ['his'] education in his ['hometown']

in his hometown ['he'] his hometown he ['decided']

hometown he decided ['to'] he decided to ['join'] decided to join ['rashtriya']

to join rashtriya ['swayamsevak'] join rashtriya swayamsevak ['sangh'] rashtriya swayamsevak sangh ['this'] swayamsevak sangh this ['is'] sangh this is ['popularly']

this is popularly ['known'] is popularly known ['as'] popularly known as ['rss'] known as rss ['in']

as rss in ['our']

rss in our ['country'] in our country ['during'] our country during ['his']

country during his ['earlier'] during his earlier ['ages'] his earlier ages ['of'] earlier ages of ['life'] ages of life ['he']

of life he ['was']

life he was ['headstrong']

he was headstrong ['and'] was headstrong and ['was'] headstrong and was ['not'] and was not ['that']

was not that ['keen'] not that keen ['on'] that keen on ['the'] keen on the ['concept'] on the concept ['of']

the concept of ['marriage'] concept of marriage ['since'] of marriage since ['then'] marriage since then ['he'] since then he ['has']

then he has ['dedicated'] he has dedicated ['his'] has dedicated his ['entire'] dedicated his entire ['life'] his entire life ['to'] entire life to ['his']

life to his ['motherland'] to his motherland ['at'] his motherland at ['the'] motherland at the ['age'] at the age ['of']

the age of ['17', 'knowledge'] age of 17 ['narendra']

of 17 narendra ['modi']

17 narendra modi ['decided'] narendra modi decided ['to', 'to'] modi decided to ['travel', 'create'] decided to travel ['around']

to travel around ['the'] travel around the ['country'] around the country ['and'] the country and ['gain', 'it'] country and gain ['knowledge'] and gain knowledge ['while'] gain knowledge while ['helping']

knowledge while helping ['others'] while helping others ['mr'] helping others mr ['modi'] others mr modi ['is']

mr modi is ['a']

modi is a ['great', 'motivation'] is a great ['admirer']

a great admirer ['of'] great admirer of ['the'] admirer of the ['ideologies']

of the ideologies ['of'] the ideologies of ['swami']

ideologies of swami ['vivekananda'] of swami vivekananda ['he']

swami vivekananda he ['always'] vivekananda he always ['emphasizes'] he always emphasizes ['coming'] always emphasizes coming ['age'] emphasizes coming age ['is'] coming age is ['the']

age is the ['age'] is the age ['of']

age of knowledge ['however'] of knowledge however ['rich'] knowledge however rich ['poor'] however rich poor ['or']

rich poor or ['powerful'] poor or powerful ['a'] or powerful a ['country'] powerful a country ['be'] a country be ['if'] country be if ['they'] be if they ['want']

if they want ['to'] they want to ['move'] want to move ['ahead'] to move ahead ['only']

move ahead only ['knowledge'] ahead only knowledge ['can'] only knowledge can ['lead'] knowledge can lead ['them'] can lead them ['to']

lead them to ['that'] them to that ['path'] to that path ['the'] that path the ['life'] path the life ['story']

the life story ['narendra'] life story narendra ['modi'] story narendra modi ['is']

narendra modi is ['a', 'considered', 'the'] is a motivation ['for']

a motivation for ['every'] motivation for every ['indian'] for every indian ['he']

every indian he ['became'] indian he became ['the'] he became the ['prime']

became the prime ['minister']

the prime minister ['of', 'of']

prime minister of ['india', 'india', 'india', 'india'] minister of india ['after', 'campaigns', 'overtook', 'he'] of india after ['breaking']

india after breaking ['the'] after breaking the ['bar'] breaking the bar ['of'] the bar of ['a']

bar of a ['povertystricken']

of a povertystricken ['teaselling'] a povertystricken teaselling ['boy'] povertystricken teaselling boy ['he'] teaselling boy he ['has']

boy he has ['seamlessly'] he has seamlessly ['become'] has seamlessly become ['a']

seamlessly become a ['developmentoriented'] become a developmentoriented ['leader']

a developmentoriented leader ['narendra'] developmentoriented leader narendra ['damodardas'] leader narendra damodardas ['modi']

damodardas modi was ['born', 'awarded'] was born on ['17th']

born on 17th ['september'] on 17th september ['1950'] 17th september 1950 ['he'] september 1950 he ['is'] 1950 he is ['a']

is a prominent ['figure'] a prominent figure ['who']

prominent figure who ['showed'] figure who showed ['us']

who showed us ['success'] showed us success ['is'] us success is ['not'] success is not ['related'] is not related ['to'] not related to ['the'] related to the ['caste'] to the caste ['system'] the caste system ['it'] caste system it ['doesn'] system it doesn ['’']

it doesn ’ ['t'] doesn ’ t ['matter'] ’ t matter ['from']

t matter from ['where'] matter from where ['a'] from where a ['person']

where a person ['belongs'] a person belongs ['or'] person belongs or ['what'] belongs or what ['his'] or what his ['or']

what his or ['her']

his or her ['background'] or her background ['is']

her background is ['narendra'] background is narendra ['modi'] is narendra modi ['is']

modi is considered ['a'] is considered a ['master']

considered a master ['strategist'] a master strategist ['and'] master strategist and ['becomes'] strategist and becomes ['a']

and becomes a ['ray'] becomes a ray ['of'] a ray of ['hope'] ray of hope ['for']

of hope for ['billions', 'billions'] hope for billions ['of', 'of']

for billions of ['lives', 'indians'] billions of lives ['in']

of lives in ['india'] lives in india ['he'] in india he ['is'] india he is ['one'] he is one ['of']

is one of ['the']

one of the ['leaders', 'top', 'most'] of the leaders ['who']

the leaders who ['stay'] leaders who stay ['focused'] who stay focused ['on']

stay focused on ['developments'] focused on developments ['with'] on developments with ['him'] developments with him ['the'] with him the ['dignity']

him the dignity ['of']

the dignity of ['labor', 'labor'] dignity of labor ['is', 'and'] of labor is ['respected']

labor is respected ['and'] is respected and ['the'] respected and the ['working'] and the working ['class']

the working class ['is'] working class is ['supported'] class is supported ['greatly'] is supported greatly ['narendra']

supported greatly narendra ['modi'] greatly narendra modi ['is']

is the glorious ['son'] the glorious son ['of'] glorious son of ['late'] son of late ['damodardas']

of late damodardas ['mulchand'] late damodardas mulchand ['modi'] damodardas mulchand modi ['and'] mulchand modi and ['heeraben'] modi and heeraben ['damodardas'] and heeraben damodardas ['modi'] heeraben damodardas modi ['none'] damodardas modi none ['of']

modi none of ['the'] none of the ['prime']

of the prime ['ministers'] the prime ministers ['had'] prime ministers had ['taken'] ministers had taken ['office'] had taken office ['when'] taken office when ['their'] office when their ['mother'] when their mother ['was'] their mother was ['alive'] mother was alive ['it']

was alive it ['is'] alive it is ['mr'] it is mr ['modi'] is mr modi ['who']

mr modi who ['created'] modi who created ['history']

who created history ['eradicating'] created history eradicating ['black'] history eradicating black ['money'] eradicating black money ['from', 'from'] black money from ['our', 'india'] money from our ['country']

from our country ['dealing'] our country dealing ['with'] country dealing with ['strong'] dealing with strong ['hands'] with strong hands ['narendra'] strong hands narendra ['modi'] hands narendra modi ['has']

narendra modi has ['a', 'provided'] modi has a ['significant']

has a significant ['role'] a significant role ['in']

significant role in ['eradicating'] role in eradicating ['black']

in eradicating black ['money'] money from india ['he']

from india he ['demonetized'] india he demonetized ['the'] he demonetized the ['currency'] demonetized the currency ['notes'] the currency notes ['of'] currency notes of ['500']

notes of 500 ['and'] of 500 and ['1000'] 500 and 1000 ['rupees'] and 1000 rupees ['and'] 1000 rupees and ['later']

rupees and later ['introduced'] and later introduced ['a'] later introduced a ['complete'] introduced a complete ['new'] a complete new ['semblance'] complete new semblance ['of'] new semblance of ['indian'] semblance of indian ['currency'] of indian currency ['notes'] indian currency notes ['this'] currency notes this ['helped'] notes this helped ['a']

this helped a ['lot'] helped a lot ['in']

a lot in ['eliminating']

lot in eliminating ['corruption']

in eliminating corruption ['terrorism'] eliminating corruption terrorism ['and'] corruption terrorism and ['counterfeit'] terrorism and counterfeit ['currency'] and counterfeit currency ['from'] counterfeit currency from ['india'] currency from india ['our']

from india our ['15th'] india our 15th ['prime'] our 15th prime ['minister'] 15th prime minister ['is']

prime minister is ['considered'] minister is considered ['to'] is considered to ['be']

considered to be ['a'] to be a ['stern']

be a stern ['administrator'] a stern administrator ['and']

stern administrator and ['leader'] administrator and leader ['with'] and leader with ['strict'] leader with strict ['and']

with strict and ['protective'] strict and protective ['discipline'] and protective discipline ['these'] protective discipline these ['can'] discipline these can ['be']

these can be ['seen'] can be seen ['through'] be seen through ['his'] seen through his ['works']

through his works ['policies'] his works policies ['speeches'] works policies speeches ['and']

policies speeches and ['initiation'] speeches and initiation ['of']

and initiation of ['various'] initiation of various ['schemes'] of various schemes ['he'] various schemes he ['maintains'] schemes he maintains ['a']

he maintains a ['great'] maintains a great ['image'] a great image ['when'] great image when ['it'] image when it ['comes'] when it comes ['to']

it comes to ['rising'] comes to rising ['from'] to rising from ['humble']

rising from humble ['beginnings'] from humble beginnings ['and'] humble beginnings and ['moving'] beginnings and moving ['to'] and moving to ['become']

moving to become ['the'] to become the ['prime']

become the prime ['minister'] of india campaigns ['led'] india campaigns led ['by'] campaigns led by ['modi'] led by modi ['poverty']

by modi poverty ['in'] modi poverty in ['a']

poverty in a ['farmers'] in a farmers ['life']

a farmers life ['has'] farmers life has ['been'] life has been ['reduced'] has been reduced ['to'] been reduced to ['a'] reduced to a ['great'] to a great ['extent']

a great extent ['thanks'] great extent thanks ['to'] extent thanks to ['the'] thanks to the ['helping'] to the helping ['hands'] the helping hands ['offered'] helping hands offered ['by'] hands offered by ['our'] offered by our ['prime']

by our prime ['minister']

our prime minister ['not', 'narendra'] prime minister not ['only']

minister not only ['poor'] not only poor ['farmers'] only poor farmers ['but'] poor farmers but ['he'] farmers but he ['also'] but he also ['helped'] he also helped ['reduce'] also helped reduce ['the']

helped reduce the ['poverty'] reduce the poverty ['level'] the poverty level ['from'] poverty level from ['other'] level from other ['sectors'] from other sectors ['he'] other sectors he ['has'] sectors he has ['eliminated'] he has eliminated ['the'] has eliminated the ['problem'] eliminated the problem ['of'] the problem of ['water'] problem of water ['from']

of water from ['india'] water from india ['carrying'] from india carrying ['the'] india carrying the ['work'] carrying the work ['to'] the work to ['the']

work to the ['next']

to the next ['stage'] the next stage ['after'] next stage after ['mr'] stage after mr ['atal'] after mr atal ['vihari']

mr atal vihari ['vajpayee'] atal vihari vajpayee ['modi'] vihari vajpayee modi ['showed'] vajpayee modi showed ['a'] modi showed a ['great']

showed a great ['interest'] a great interest ['in'] great interest in ['the']

interest in the ['construction'] in the construction ['of']

the construction of ['infrastructure'] construction of infrastructure ['in'] of infrastructure in ['india'] infrastructure in india ['a']

in india a ['generous'] india a generous ['and']

a generous and ['recognized'] generous and recognized ['campaign'] and recognized campaign ['make'] recognized campaign make ['in'] campaign make in ['india']

make in india ['was'] in india was ['started'] india was started ['by'] was started by ['mr'] started by mr ['narendra'] by mr narendra ['modi']

mr narendra modi ['in', 'he'] narendra modi in ['this', 'april'] modi in this ['campaign']

in this campaign ['he']

this campaign he ['conveyed'] campaign he conveyed ['the'] he conveyed the ['message'] conveyed the message ['to'] the message to ['manufacturers']

message to manufacturers ['that'] to manufacturers that ['it'] manufacturers that it ['is'] that it is ['best']

it is best ['to'] is best to ['use']

best to use ['indian']

to use indian ['materials']

use indian materials ['and'] indian materials and ['products'] materials and products ['rather'] and products rather ['than'] products rather than ['depending'] rather than depending ['on'] than depending on ['foreign'] depending on foreign ['goods'] on foreign goods ['this'] foreign goods this ['way']

goods this way ['our'] this way our ['money'] way our money ['will']

our money will ['circulate'] money will circulate ['within'] will circulate within ['the'] circulate within the ['country'] within the country ['and'] country and it ['will']

and it will ['help'] it will help ['to'] will help to ['reduce'] help to reduce ['the']

to reduce the ['inflation'] reduce the inflation ['rate'] the inflation rate ['to'] inflation rate to ['end'] rate to end ['with']

to end with ['india'] end with india ['has']

with india has ['benefitted'] india has benefitted ['like'] has benefitted like ['never'] benefitted like never ['before'] like never before ['under'] never before under ['the'] before under the ['leadership'] under the leadership ['of'] the leadership of ['honourable']

leadership of honourable ['prime'] of honourable prime ['minister'] honourable prime minister ['mr'] prime minister mr ['narendra'] minister mr narendra ['modi'] narendra modi he ['has']

modi he has ['taken'] he has taken ['all'] has taken all ['the']

taken all the ['initiatives']

all the initiatives ['to'] the initiatives to ['make'] initiatives to make ['our'] to make our ['country'] make our country ['great'] our country great ['and']

country great and ['appreciable'] great and appreciable ['on']

and appreciable on ['the'] appreciable on the ['global'] on the global ['standard'] the global standard ['recent']

global standard recent ['endeavours'] standard recent endeavours ['of'] recent endeavours of ['narendra'] endeavours of narendra ['modi']

of narendra modi ['in'] modi in april ['2020']

in april 2020 ['narendra'] april 2020 narendra ['modi'] 2020 narendra modi ['an'] narendra modi an ['indian'] modi an indian ['politician'] an indian politician ['and'] indian politician and ['the'] politician and the ['current'] and the current ['prime'] the current prime ['minister'] current prime minister ['of'] of india overtook ['us']

india overtook us ['president'] overtook us president ['donald'] us president donald ['trump'] president donald trump ['as'] donald trump as ['the']

trump as the ['most'] as the most ['popular']

the most popular ['world', 'leaders'] most popular world ['leader'] popular world leader ['on']

world leader on ['facebook'] leader on facebook ['he'] on facebook he ['has'] facebook he has ['ranked'] he has ranked ['first'] has ranked first ['among'] ranked first among ['all']

first among all ['international'] among all international ['leaders']

all international leaders ['in'] international leaders in ['the'] leaders in the ['fight']

in the fight ['against'] the fight against ['the']

fight against the ['coronavirus'] against the coronavirus ['covid19'] the coronavirus covid19 ['pandemic']

coronavirus covid19 pandemic ['ensuring'] covid19 pandemic ensuring ['the'] pandemic ensuring the ['safety'] ensuring the safety ['and']

the safety and ['security'] safety and security ['of'] and security of ['indian']

security of indian ['citizens'] of indian citizens ['and'] indian citizens and ['offering'] citizens and offering ['all'] and offering all ['essential'] offering all essential ['help'] all essential help ['to'] essential help to ['other'] help to other ['countries']

to other countries ['in', 'article'] other countries in ['the'] countries in the ['aftermath']

in the aftermath ['of'] the aftermath of ['the']

aftermath of the ['pandemic'] of the pandemic ['prime']

the pandemic prime ['minister'] pandemic prime minister ['narendra'] decided to create ['a']

to create a ['distinct'] create a distinct ['ministry'] a distinct ministry ['of'] distinct ministry of ['ayush'] ministry of ayush ['now']

of ayush now ['selling']

ayush now selling ['medications'] now selling medications ['to'] selling medications to ['other'] medications to other ['countries'] other countries article ['370'] countries article 370 ['which'] article 370 which ['granted'] 370 which granted ['special'] which granted special ['status']

granted special status ['to'] special status to ['the'] status to the ['former']

to the former ['state'] the former state ['of'] former state of ['jammu'] state of jammu ['and'] of jammu and ['kashmir'] jammu and kashmir ['jk'] and kashmir jk ['was'] kashmir jk was ['repealed'] jk was repealed ['under'] was repealed under ['his']

repealed under his ['strong'] under his strong ['leadership'] his strong leadership ['narendra']

strong leadership narendra ['modis'] leadership narendra modis ['leadership'] narendra modis leadership ['has']

modis leadership has ['been'] leadership has been ['hailed'] has been hailed ['by']

been hailed by ['world'] hailed by world ['leaders']

by world leaders ['international'] world leaders international ['agencies']

leaders international agencies ['philanthropists'] international agencies philanthropists ['nobel'] agencies philanthropists nobel ['laureates'] philanthropists nobel laureates ['and']

nobel laureates and ['many'] laureates and many ['more'] and many more ['modiji'] many more modiji ['earned'] more modiji earned ['the'] modiji earned the ['uns'] earned the uns ['top']

the uns top ['environmental'] uns top environmental ['accolade'] top environmental accolade ['the']

environmental accolade the ['champion'] accolade the champion ['of']

the champion of ['the'] champion of the ['earth'] of the earth ['in', 'in']

the earth in ['october', 'october'] earth in october ['2018', '2018'] in october 2018 ['on', 'he'] october 2018 on ['february']

2018 on february ['22'] on february 22 ['2019']

february 22 2019 ['narendra'] 22 2019 narendra ['damodardas'] 2019 narendra damodardas ['modi'] modi was awarded ['the']

was awarded the ['prestigious'] awarded the prestigious ['seoul'] the prestigious seoul ['peace'] prestigious seoul peace ['prize'] seoul peace prize ['2018', 'he'] peace prize 2018 ['to']

prize 2018 to ['contribute'] 2018 to contribute ['to']

to contribute to ['international'] contribute to international ['collaboration'] to international collaboration ['and'] international collaboration and ['global'] collaboration and global ['economic']

and global economic ['prosperity'] global economic prosperity ['on'] economic prosperity on ['april'] prosperity on april ['12']

on april 12 ['2019'] april 12 2019 ['he'] 12 2019 he ['was'] 2019 he was ['also'] he was also ['awarded'] was also awarded ['the'] also awarded the ['order'] awarded the order ['of'] the order of ['st'] order of st ['andrew'] of st andrew ['russias']

st andrew russias ['highest'] andrew russias highest ['civilian']

russias highest civilian ['decoration'] highest civilian decoration ['for'] civilian decoration for ['his'] decoration for his ['second']

for his second ['term'] his second term ['as', 'as']

second term as ['prime', 'gujarats'] term as prime ['minister']

as prime minister ['of'] of india he ['ran'] india he ran ['on']

he ran on ['nationalism'] ran on nationalism ['in']

on nationalism in ['the'] nationalism in the ['2019'] in the 2019 ['general']

the 2019 general ['election'] 2019 general election ['and'] general election and ['earned'] election and earned ['a']

and earned a ['large'] earned a large ['mandate']

a large mandate ['benchmarks'] large mandate benchmarks ['of'] mandate benchmarks of ['modi'] benchmarks of modi ['’']

of modi ’ ['s'] modi ’ s ['success'] ’ s success ['after']

s success after ['being']

success after being ['elected'] after being elected ['to'] being elected to ['his'] elected to his ['second']

to his second ['term'] term as gujarats ['chief']

as gujarats chief ['minister'] gujarats chief minister ['in', 'he'] chief minister in ['2002'] minister in 2002 ['he']

in 2002 he ['focused'] 2002 he focused ['on'] he focused on ['the'] focused on the ['states'] on the states ['economic']

the states economic ['development']

states economic development ['and'] economic development and ['an'] development and an ['attractive'] and an attractive ['location']

an attractive location ['for'] attractive location for ['business'] location for business ['people'] for business people ['and']

business people and ['industrialists'] people and industrialists ['in']

and industrialists in ['2007'] industrialists in 2007 ['during'] in 2007 during ['his']

2007 during his ['third'] during his third ['term'] his third term ['as']

third term as ['cm'] term as cm ['he']

as cm he ['increased']

cm he increased ['agricultural'] he increased agricultural ['growth']

increased agricultural growth ['rates'] agricultural growth rates ['provided'] growth rates provided ['power']

rates provided power ['to'] provided power to ['all', 'every'] power to all ['villages']

to all villages ['and']

all villages and ['bolstered'] villages and bolstered ['the'] and bolstered the ['states'] bolstered the states ['rapid'] the states rapid ['development'] states rapid development ['when'] rapid development when ['he'] development when he ['was']

when he was ['gujarats'] he was gujarats ['chief']

was gujarats chief ['minister'] chief minister he ['launched'] minister he launched ['groundwater']

he launched groundwater ['conservation'] launched groundwater conservation ['initiatives'] groundwater conservation initiatives ['with'] conservation initiatives with ['the'] initiatives with the ['government']

with the government ['’'] the government ’ ['s'] government ’ s ['help'] ’ s help ['this']

s help this ['aided'] help this aided ['in'] this aided in ['the']

aided in the ['cultivation'] in the cultivation ['of'] the cultivation of ['bt'] cultivation of bt ['cotton'] of bt cotton ['by']

bt cotton by ['providing']

cotton by providing ['irrigation'] by providing irrigation ['through'] providing irrigation through ['tube'] irrigation through tube ['wells'] through tube wells ['gujarats'] tube wells gujarats ['governor']

wells gujarats governor ['narendra'] gujarats governor narendra ['modi'] governor narendra modi ['has']

modi has provided ['power'] has provided power ['to'] power to every ['village'] to every village ['in'] every village in ['addition'] village in addition ['he'] in addition he ['modified'] addition he modified ['the'] he modified the ['states'] modified the states ['power']

the states power ['distribution'] states power distribution ['system'] power distribution system ['by'] distribution system by ['dividing'] system by dividing ['agricultural'] by dividing agricultural ['and'] dividing agricultural and ['rural'] agricultural and rural ['electricity'] and rural electricity ['narendra'] rural electricity narendra ['modi'] electricity narendra modi ['introduced'] narendra modi introduced ['honouring'] modi introduced honouring ['the']

introduced honouring the ['interworldwidenational'] honouring the interworldwidenational ['day']

the interworldwidenational day ['of'] interworldwidenational day of ['yoga'] day of yoga ['during', 'is']

of yoga during ['his'] yoga during his ['speech'] during his speech ['to'] his speech to ['the'] speech to the ['united'] to the united ['nations']

the united nations ['general'] united nations general ['assembly'] nations general assembly ['thanks'] general assembly thanks ['to'] assembly thanks to ['his']

thanks to his ['efforts'] to his efforts ['the']

his efforts the ['international'] efforts the international ['day'] the international day ['of'] international day of ['yoga']

of yoga is ['observed']

yoga is observed ['on'] is observed on ['june'] observed on june ['21st'] on june 21st ['all'] june 21st all ['over'] 21st all over ['the'] all over the ['world'] over the world ['modis'] the world modis ['book'] world modis book ['aankh'] modis book aankh ['ka'] book aankh ka ['dhanya'] aankh ka dhanya ['che'] ka dhanya che ['has'] dhanya che has ['a']

che has a ['compilation'] has a compilation ['of'] a compilation of ['his']

compilation of his ['poems'] of his poems ['the']

his poems the ['madame'] poems the madame ['tussauds'] the madame tussauds ['wax'] madame tussauds wax ['museum'] tussauds wax museum ['in'] wax museum in ['london'] museum in london ['has']

in london has ['a'] london has a ['wax'] has a wax ['statue'] a wax statue ['of'] wax statue of ['modi'] statue of modi ['in'] of modi in ['2015'] modi in 2015 ['he'] in 2015 he ['was'] 2015 he was ['placed']

he was placed ['sixth'] was placed sixth ['on'] placed sixth on ['fortune'] sixth on fortune ['magazines'] on fortune magazines ['list'] fortune magazines list ['of'] magazines list of ['the'] list of the ['worlds']

of the worlds ['most']

the worlds most ['powerful'] worlds most powerful ['leaders'] most powerful leaders ['narendra'] powerful leaders narendra ['modi']

leaders narendra modi ['was'] modi was designated ['one'] was designated one ['of'] designated one of ['the'] of the top ['30']

the top 30 ['most']

top 30 most ['influential']

30 most influential ['individuals'] most influential individuals ['on'] influential individuals on ['the'] individuals on the ['internet']

on the internet ['and'] the internet and ['one'] internet and one ['of'] and one of ['forbes', 'the'] one of forbes ['top']

of forbes top ['ten'] forbes top ten ['most'] top ten most ['powerful'] ten most powerful ['people'] most powerful people ['on'] powerful people on ['the'] people on the ['planet'] on the planet ['he']

the planet he ['earned'] planet he earned ['the'] he earned the ['united'] earned the united ['nation'] the united nation ['’'] united nation ’ ['s'] nation ’ s ['highest']

’ s highest ['environmental']

s highest environmental ['honor'] highest environmental honor ['champions'] environmental honor champions ['of'] honor champions of ['the']

champions of the ['earth'] october 2018 he ['is'] 2018 he is ['the']

is the first ['indian'] the first indian ['to'] first indian to ['get'] indian to get ['the'] to get the ['2018'] get the 2018 ['seoul'] the 2018 seoul ['peace']

2018 seoul peace ['prize'] peace prize he ['is'] prize he is ['a']

is a beacon ['of'] a beacon of ['hope'] beacon of hope ['for']

billions of indians ['and'] of indians and ['one'] indians and one ['of']

of the most ['popular'] most popular leaders ['who'] popular leaders who ['focus'] leaders who focus ['on'] who focus on ['development'] focus on development ['even'] on development even ['our']

development even our ['prime'] even our prime ['minister'] minister narendra modis ['slogan'] narendra modis slogan ['main'] modis slogan main ['bhi']

slogan main bhi ['chowkidar', 'chowkidar'] main bhi chowkidar ['emphasizes', 'became'] bhi chowkidar emphasizes ['the'] chowkidar emphasizes the ['dignity'] emphasizes the dignity ['of']

of labor and ['seeks'] labor and seeks ['the'] and seeks the ['support'] seeks the support ['of'] the support of ['the'] support of the ['working'] of the working ['people'] the working people ['he'] working people he ['used'] people he used ['this'] he used this ['term'] used this term ['because'] this term because ['he'] term because he ['believes'] because he believes ['he'] he believes he ['too'] believes he too ['is']

he too is ['standing']

too is standing ['steadfast'] is standing steadfast ['and'] standing steadfast and ['doing'] steadfast and doing ['his'] and doing his ['job']

doing his job ['as'] his job as ['the'] job as the ['nations']

as the nations ['chowkidar'] the nations chowkidar ['he'] nations chowkidar he ['further'] chowkidar he further ['stated'] he further stated ['that'] further stated that ['any'] stated that any ['indian']

that any indian ['fighting'] any indian fighting ['against']

indian fighting against ['corruption'] fighting against corruption ['filth'] against corruption filth ['social'] corruption filth social ['evils'] filth social evils ['and']

social evils and ['other'] evils and other ['issues'] and other issues ['for'] other issues for ['indias']

issues for indias ['prosperity'] for indias prosperity ['is'] indias prosperity is ['a'] prosperity is a ['chowkidar'] is a chowkidar ['the']

a chowkidar the ['slogan'] chowkidar the slogan ['main'] the slogan main ['bhi']

bhi chowkidar became ['famous'] chowkidar became famous ['as'] became famous as ['a']

famous as a ['result'] as a result ['of']

a result of ['this']

for m in range (5):

number= random.randint(1,len(words\_tokens)-words) curr\_sequence = ' '.join(words\_tokens[number:number+words]) output = curr\_sequence

for i in range(100):

if curr\_sequence not in ngrams.keys(): break

possible\_words = ngrams[curr\_sequence] next\_word =

possible\_words[random.randrange(len(possible\_words))] output += ' ' + next\_word

seq\_words = nltk.word\_tokenize(output) curr\_sequence = ' '.join(seq\_words[len(seq\_words)-

words:len(seq\_words)]) print(output,"\n")

politician and the current prime minister of india after breaking the bar of a povertystricken teaselling boy he has seamlessly become a developmentoriented leader narendra damodardas modi was born on 17th september 1950 he is a prominent figure who showed us success is not related to the caste system it doesn ’ t matter from where a person belongs or what his or her background is narendra modi is a motivation for every indian he became the prime minister of india overtook us president donald trump as the most popular world leader on facebook he has ranked first among all international leaders in

civilian decoration for his second term as prime minister of india after breaking the bar of a povertystricken teaselling boy he has seamlessly become a developmentoriented leader narendra damodardas modi was born in a lowermiddleclass family at vadnagar gujarat he had a keen interest in politics since the early days of his childhood after completing his higher education in his hometown he decided to join rashtriya swayamsevak sangh this is popularly known as rss in our country during his earlier ages of life he was headstrong and was not that keen on the concept of marriage since then he has dedicated his entire

and kashmir jk was repealed under his strong leadership narendra modis leadership has been hailed by world leaders international agencies philanthropists nobel laureates and many more modiji earned the uns top environmental accolade the champion of the earth in october 2018 on february 22 2019 narendra damodardas modi is the glorious son of late damodardas mulchand modi and heeraben damodardas modi none of the prime ministers had taken office when their mother was alive it is mr modi who created history eradicating black money from our country dealing with strong hands narendra modi has provided power to every village in addition he modified

demonetized the currency notes of 500 and 1000 rupees and later introduced a complete new semblance of indian currency notes this helped a lot in eliminating corruption terrorism and counterfeit currency from india our 15th prime minister is considered to be a stern administrator and leader with strict and protective discipline these can be seen through his works policies speeches and initiation of various schemes he maintains a great image when it comes to rising from humble beginnings and moving to become the prime minister of india he ran on nationalism in the 2019 general election and earned a large mandate benchmarks of

has dedicated his entire life to his motherland at the age of 17 narendra modi decided to create a distinct ministry of ayush now selling medications to other countries article 370 which granted special status to the former state of jammu and kashmir jk was repealed under his strong leadership narendra modis leadership has been hailed by world leaders international agencies philanthropists nobel laureates and many more modiji earned the uns top environmental

accolade the champion of the earth in october 2018 he is the first indian to get the 2018 seoul peace prize he is a member of the parliament mp who