

DATA SUMMARIZER using NLP
MINI PROJECT REPORT

18CSC305J - ARTIFICIAL INTELLIGENCE

Submitted by

Anukanksha Aashi (RA2011003010824)
Jigyasa Sharma (RA2011003010832)
Chinmoyee Gogoi (RA2011003010884)

Under the guidance of

M.Rajalakshmi

Assistant Professor, Department of Computer Science and Engineering

in partial fulfillment for the award of the degree
of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY



S.R.M. Nagar, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this project report “**DATA SUMMARIZER using NLP**” is the bona fide work of **Anukanksha Aashi (RA2011003010824)**, **Jigyasa Sharma (RA2011003010832)**, **Chinmoyee Gogoi (RA2011003010884)** of III Year/ VI Sem B.Tech (CSE) who carried out the miniproject work under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

M.Rajalakshmi

GUIDE

Assiatant Professor

Department of Computing Technologies

SIGNATURE

Dr. M. Pushpalatha

HEAD OF THE DEPARTMENT

Professor & Head

Department of Computing Technologies

ABSTRACT

The Data Summarizer using NLP project aims to develop a system that can automatically summarize large volumes of textual data using natural language processing techniques. The system will use advanced algorithms to analyze the input data and identify the most important and relevant information, which will then be summarized into a concise and coherent summary. This project has the potential to be of great value in various domains, including journalism, business, and research, where there is a need to quickly analyze and summarize large amounts of text-based data. By automating this process, the Data Summarizer using NLP project can significantly improve the efficiency and accuracy of data analysis and decision-making processes.

TABLE OF CONTENTS

| | |
|--|------------|
| ABSTRACT | iii |
| TABLE OF CONTENTS | iv |
| LIST OF FIGURES | v |
| ABBREVIATIONS | vi |
| 1 INTRODUCTION | 7 |
| 2 LITERATURE SURVEY | 8 |
| 3 SYSTEM ARCHITECTURE AND DESIGN | 9 |
| 4 METHODOLOGY | 10 |
| 5 CODING AND TESTING | 12 |
| 6 SREENSHOTS AND RESULTS | |
| 7 CONCLUSION AND FUTURE ENHANCEMENT | 17 |
| REFERENCES | 18 |

LIST OF FIGURES

Fig :1 (Proposed System)

Fig:2 (Working of NLP)

Fig 3: (Displayed Output)

Fig 4 (Working Output)

ABBREVIATIONS

- NLP: Natural Language Processing

CHAPTER 1

INTRODUCTION

Data summarization using Natural Language Processing (NLP) is the process of extracting important information from a large amount of data and presenting it in a concise and informative manner. NLP techniques are used to automatically analyze, process and extract meaning from text data.

Data summarization using NLP involves several steps including text pre-processing, feature extraction, and summarization. Text pre-processing involves cleaning and transforming raw text data into a suitable format for analysis. Feature extraction involves identifying important terms, phrases, or concepts from the text data that are relevant to the summarization task.

NLP techniques are widely used in various applications such as news summarization, social media analysis, and document summarization. With the increasing volume of text data being generated every day, data summarization using NLP is becoming increasingly important for efficient data analysis and decision-making. Summarization techniques can be categorized into two main types: extractive and abstractive.

Extractive summarization techniques are commonly used for NLP-based summarization tasks. These techniques involve identifying important sentences or phrases from the text based on certain criteria, such as sentence relevance, frequency of occurrence, or importance of keywords. The selected sentences or phrases are then combined to generate a summary that captures the most important information from the original text.

Abstractive summarization techniques, on the other hand, involve generating a summary that is not a verbatim extract of the text, but rather a summary that captures the essence of the original text. Abstractive summarization techniques are more complex and require advanced NLP models, such as neural networks and deep learning algorithms.

Data summarization using NLP has several applications, including:

- News summarization: NLP-based summarization can be used to generate a summary of news articles, enabling readers to quickly grasp the most important information without reading the entire article.
- Social media analysis: NLP-based summarization can be used to extract important insights from social media posts, such as sentiment analysis or topic modeling.

NLP techniques are widely used in various applications such as news summarization, social media analysis, and document summarization. With the increasing volume of text data being generated every day, data summarization using NLP is becoming increasingly important for efficient data analysis and decision-making.

CHAPTER 2

LITERATURE SURVEY

- Kim, J., J. K. Lee, and K. M. Lee. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks."
- *Proceedings of the IEEE® Conference on Computer Vision and Pattern Recognition*. 2016, pp. 1646-1654.
- Grubinger, M., P. Clough, H. Müller, and T. Deselaers. "The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information Systems."
- *Proceedings of the OntoImage 2006 Language Resources For Content-Based Image Retrieval*. Genoa, Italy. Vol. 5, May 2006, p. 10.
- He, K., X. Zhang, S. Ren, and J. Sun. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification."
- *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026-1034.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

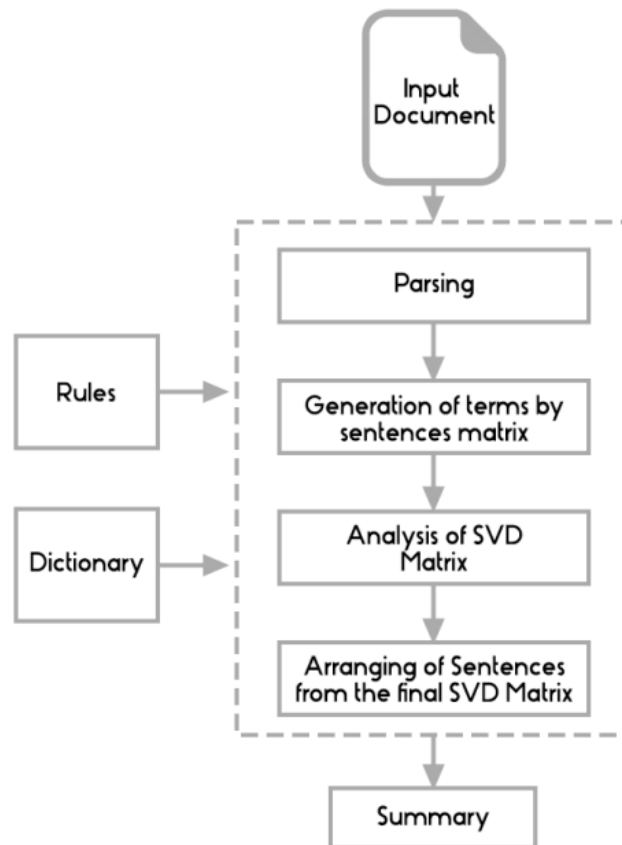


Fig :1 (Proposed System)

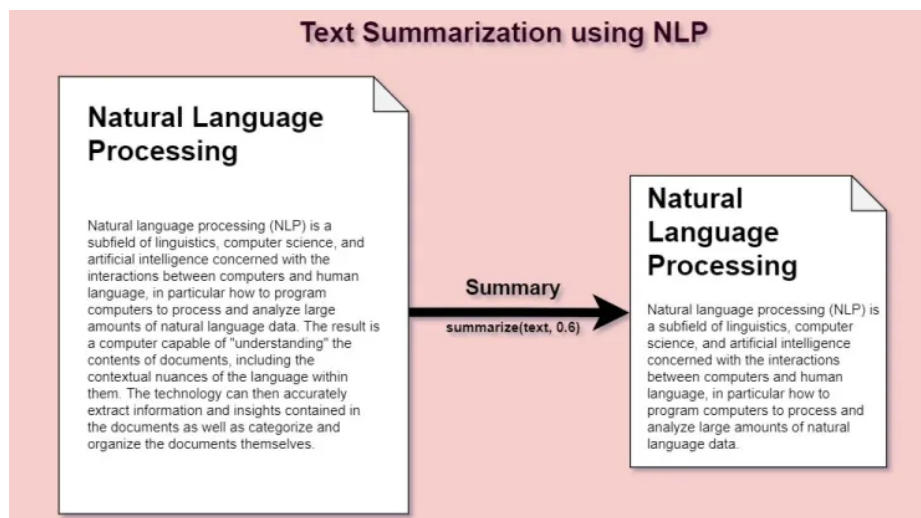


Fig:2 (Working of NPL)

CHAPTER 4

METHODOLOGY

Building a data summarizer using natural language processing (NLP) involves the following methodology:

- **Define the problem and determine the objective:** The first step in any NLP project is to determine the objective of the project and the problem that it aims to solve. In this case, the problem is to summarize large amounts of data and the objective is to develop an NLP-based solution that can accurately and efficiently summarize the data.

- **Gather and preprocess the data:** The next step is to gather the data that needs to be summarized and preprocess it to make it ready for analysis. This may involve cleaning and normalizing the text data, removing stop words, and identifying key phrases and entities.

- **Develop the summarization model:** This is the heart of the project, where the NLP algorithms are used to develop a model that can accurately summarize the data. There are several approaches that can be used, including extractive summarization, where the most important sentences or phrases are extracted from the original text, and abstractive summarization, where a new summary is generated that captures the essence of the original text.

- **Evaluate the model:** Once the model has been developed, it needs to be evaluated to determine its accuracy and effectiveness. This may involve using metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and BLEU (Bilingual Evaluation Understudy) to compare the summary generated by the model to a reference summary or the original text.

- **Refine the model:** Based on the evaluation results, the model can be refined and optimized to improve its accuracy and effectiveness.

- **Deploy the model:** Once the model has been refined and optimized, it can be deployed and integrated into the target system or application.

- **Monitor and maintain the model:** Finally, it is important to monitor the performance of the model over time and make necessary updates and improvements to ensure that it continues to provide accurate and effective summaries.

CHAPTER 5

CODING AND TESTING

HTML CODE:

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1.0" />
    <title>Data Summarization</title>
    <link
      href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css"
      rel="stylesheet"
    />
    <link href="/static/css/jquery-ui.min.css" rel="stylesheet" />
  </head>

  <body>
    <header
      style="background-color: #004a56"
      class="fixed inset-0 w-full flex flex-col justify-center justify-items-center content-center h-20
rounded-b-lg"
    >
      
      <div class="self-center text-white">Data Summarization</div>
    </header>

    <section class="flex flex-wrap mt-20 w-full">
      <div class="w-full md:w-1/2">
        <textarea
```

```

id="dataBox"
class="w-11/12 md:h-3/4 m-2 p-2 border-black rounded-lg border self-center justify-center"
name="data"
id="data"
cols="30"
rows="10"
placeholder="Enter your Data"
required="required"
></textarea>
<div class="flex self-center">
  <h3>Summary Length</h3>
  <input type="range" class="m-2" min="20" max="1000" name="maxL" />
</div>
<div class="flex self-center">
  <button
    class="m-1 bg-green-500 hover:bg-blue-700 text-white font-semi-bold py-2 px-4 rounded-lg"
    onclick="Check()"
  >
    Submit
  </button>
  <button
    class="m-1 bg-red-500 hover:bg-blue-700 text-white font-semi-bold py-2 px-4 rounded-lg"
    type="reset"
  >
    Clear
  </button>
</div>
</div>
<div class="w-full md:w-1/2">
  <div class="mt-2 flex flex-col w-full md:h-screen">
    <textarea
      id='resultBox'
      style="background-color: #edffd2"
      class="w-11/12 border-green-600 rounded-lg p-2 border self-center justify-center md:h-3/4"
      rows="10"
      cols="30"
      name="result"
      readonly

```

```
        placeholder="Your Summary"
    >
</textarea>
    >
    <button
        class="m-2 bg-blue-400 hover:bg-blue-700 text-white font-semi-bold py-2 px-4 rounded-lg self-
center"
        onclick="myFunction()"
    >
        Copy text
    </button>
</div>
</div>
</section>

<script>
function Check(e) {
    console.log("Check");
    data = document.getElementById("dataBox").value;
    console.log(data);
    const xhttp = new XMLHttpRequest();
    xhttp.onload = function () {
        document.getElementById("resultBox").innerHTML = this.responseText;
    };
    xhttp.open("POST", "https://api-inference.huggingface.co/models/facebook/bart-large-cnn");
    xhttp.setRequestHeader("Authorization", "Bearer
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");
    xhttp.send(data)
}
</script>
</body>
</html>
```

PYTHON CODE:-

```
import requests

from flask import Flask,render_template,url_for
from flask import request as req


app = Flask(__name_)
@app.route("/",methods=["GET","POST"])
def Index():
    return render_template("index.html")


@app.route("/Summarize",methods=["GET","POST"])
def Summarize():
    if req.method == "POST":
        API_URL = "https://api-inference.huggingface.co/models/facebook/bart-large-cnn"
        headers = {"Authorization": "Bearer hf_ePsRrVQMJIgAiiqNdPfyBOELBQTLOkhaO"}

        data=req.form["data"]

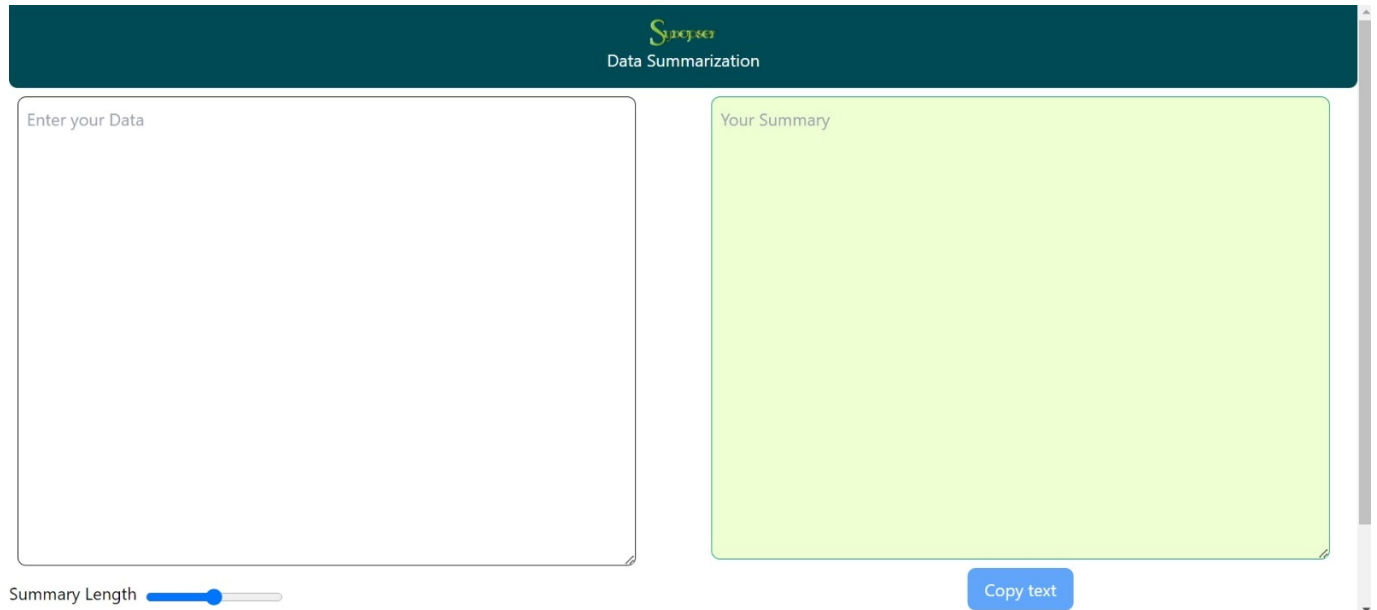
        #maxL=int(req.form["maxL"])
        #minL=maxL//4

        def query(payload):
            response = requests.post(API_URL, headers=headers, json=payload)
            return response.json()
        output = query({
            "inputs":data,
            "parameters":{"min_length":minL,"max_length":maxL},
        })[0]

        return render_template("index.html",result=output["summary_text"])
    else:
        return render_template("index.html")
if __name__=="main_":
    app.run(debug=True)
```

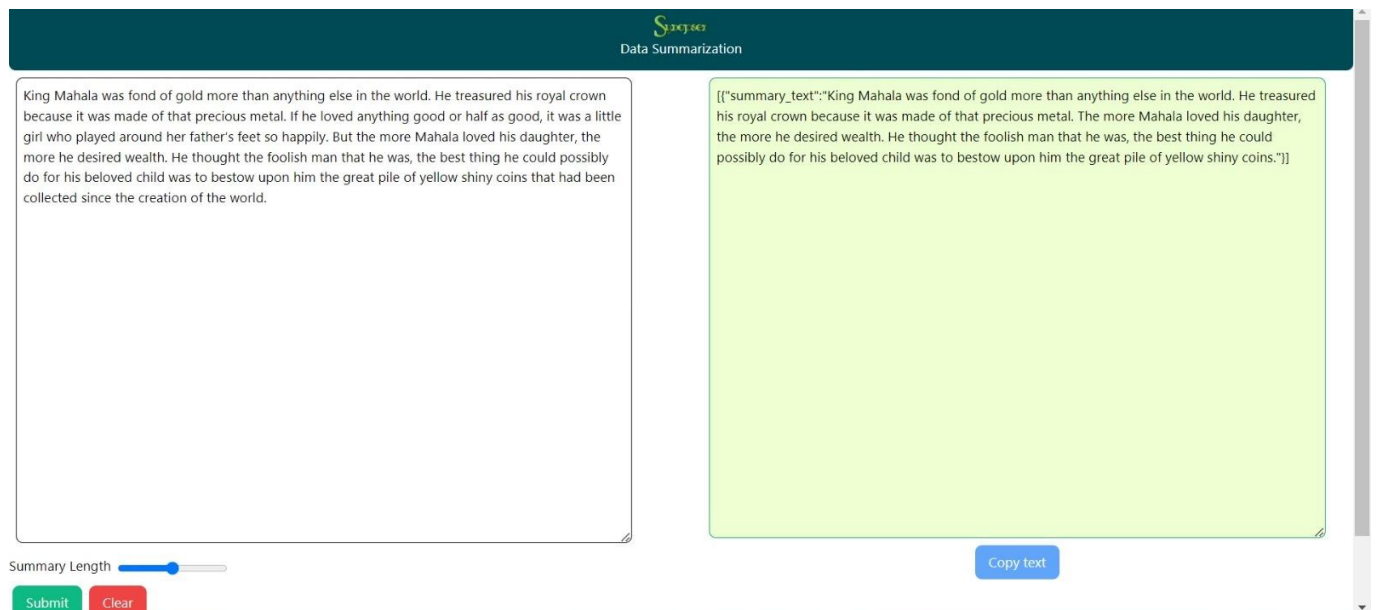
CHAPTER 6

SCREENSHOTS AND RESULTS



The screenshot shows a web application titled "Sumses Data Summarization". It features two main text areas: "Enter your Data" on the left and "Your Summary" on the right. Below the "Enter your Data" area is a "Summary Length" slider. Below the "Your Summary" area is a "Copy text" button. The interface is clean and ready for user input.

Fig 3: (Displayed Output)



This screenshot shows the same application with sample data. The "Enter your Data" field contains a paragraph about King Mahala. The "Your Summary" field displays a JSON object: `[{"summary_text": "King Mahala was fond of gold more than anything else in the world. He treasured his royal crown because it was made of that precious metal. The more Mahala loved his daughter, the more he desired wealth. He thought the foolish man that he was, the best thing he could possibly do for his beloved child was to bestow upon him the great pile of yellow shiny coins."}]`. At the bottom left, there are "Submit" and "Clear" buttons. The "Summary Length" slider is also visible.

Fig 4: (Working Output)

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, data summarization using NLP has the potential to be a highly effective tool for extracting key information from large datasets, reducing the time and effort required to analyze them. It has already shown promising results in various fields, including finance, healthcare, and education.

However, there are some limitations to consider, such as the accuracy of the summarization, which can be affected by factors like the quality and diversity of the dataset, the complexity of the language used, and the type of summarization technique used. Additionally, there is a risk of losing important context and nuances when summarizing large amounts of data, which can lead to misinterpretations and misunderstandings.

To enhance the effectiveness of data summarization using NLP, future research should focus on improving the accuracy and reliability of the summarization algorithms by incorporating more advanced machine learning techniques, such as deep learning and reinforcement learning. Another area of improvement is the development of more context-aware summarization models that can better capture the meaning and significance of the text.

Furthermore, it would be valuable to investigate how summarization can be combined with other NLP tasks, such as sentiment analysis and entity recognition, to provide more comprehensive insights into large datasets. Additionally, there is a need to evaluate the ethical and legal implications of data summarization, such as data privacy and bias, and develop guidelines and regulations accordingly.

REFERENCES

- <https://www.analyticsvidhya.com/blog/2021/05/deep-learning-for-image-super-resolution/>
- <https://blog.paperspace.com/image-super-resolution/>
- <https://in.mathworks.com/help/images/single-image-super-resolution-using-deep-learning.html>
- Image Super-Resolution Using Deep Convolutional Networks Chao Dong, Chen Change Loy, Member, IEEE, Kaiming He, Member, IEEE, and Xiaoou Tang, Fellow, IEEE.