Name : Kolla Gayatri Shruti Roll No. : 2105279 Section : NLP-2

1. **Text Summarization**

**Objective**

# In this project, we aim scrape two websites **(CNBC-TV18 Edge article on Budget '24 — How it paints a picture of a resilient India and aljazeera’s article on Analysis: India’s 2024 interim budget shows a changing economy)** and pass it through two pre-trained **text summarization models from Hugging face repository** to get a concise overview on the 2024 Indian Budget.

**Methodology**

**Data Acquisition**

We obtain textual data through web scraping.

For web scraping we use BeautifulSoup.

The data from both the websites is joined.

**Text Segmentation**

The whole textual data is divided into chunks of 1000 length as the models don’t take input of length beyond that.

**Summarization**

The chunks are then passed through the model and the summary of all the individual chunks are aggregated.

The aggregate is again passed through the model to get the summary of the entire data

**We perform the same steps to pass the data through both the models**

**Result**

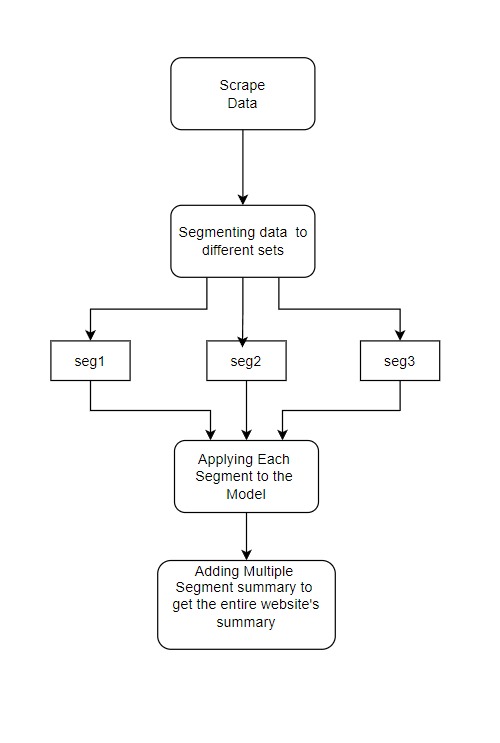
**Model 1 :**

"Indian's estimated contribution to global growth rising by 200 basis points in five years . The Budget's estimates for the fiscal year 2024–25 indicate a well-balanced approach . India's projected ascent to become the third-largest economy by 2027 puts India as a key player on the global economic stage . In her speech, Sitharaman said the average real income has risen by 50 percent, more than 250 million people have been lifted from poverty ."

**Model 2 :**

'The Interim Union Budget for the fiscal year 2024-25 was presented on February 6, 2024 at 9:36:25 AM IST. It proposes a well-balanced approach with total receipts other than borrowings estimated at ₹30.80 trillion. There is a substantial increase in the'

**NLP Diagram**

****

**Conclusion**

The 1st model works better than the 2nd one as it is providing a concise summary of the whole text, while the 2nd model is providing just a section of the whole text.

1. **Sentiment Analysis**

**Objective**

# In this project, we aim to get sentiment analysis on the 2024 Indian Budget by scraping two websites - **CNBC-TV18 Edge article on Budget '24 — How it paints a picture of a resilient India and aljazeera’s article on Analysis: India’s 2024 interim budget shows a changing economy** and passing it through two pre-trained **sentiment analysis models from Hugging face repository** and using different libraries like **pandas, matplotlib** to get the visual representation of the sentiment analysis.

**Methodology**

**We perform the same steps to pass the data through both the models**

**Data Acquisition**

We obtain textual data through web scraping.

For web scraping we use BeautifulSoup.

The data from both the websites is joined.

**Text Splitting**

The whole textual data is divided into an array of all sentences as the model gives positive negative and neutral scores for single sentences.

**Sentence Segmentation**

The 2nd model doesn’t take sentences larger than 512 length, therefore sentences exceeding that length are divided into chunks

**Information extraction**

All the sentences are passed through the model and the scores for positive, negative and neutral sentiments are counted and stored in an array

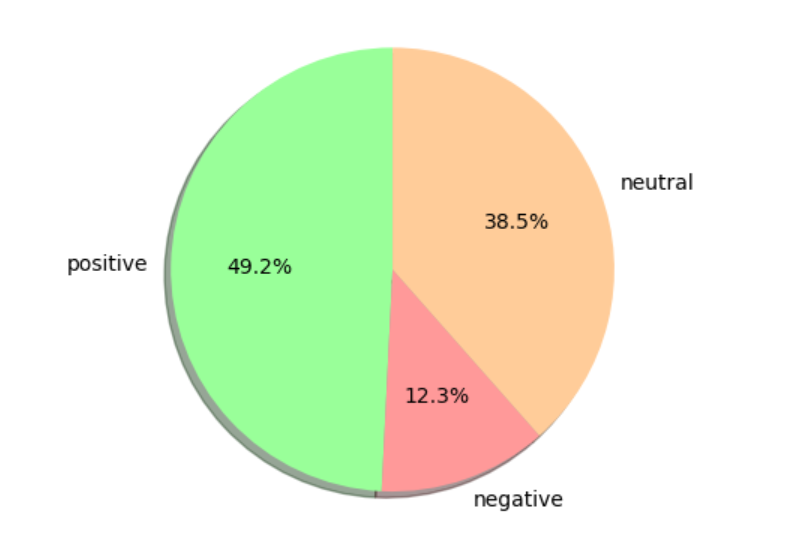
**Plotting the result**

For this step, **matplotlib** library is used to plot a pie chart for positive negative and neutral scores.

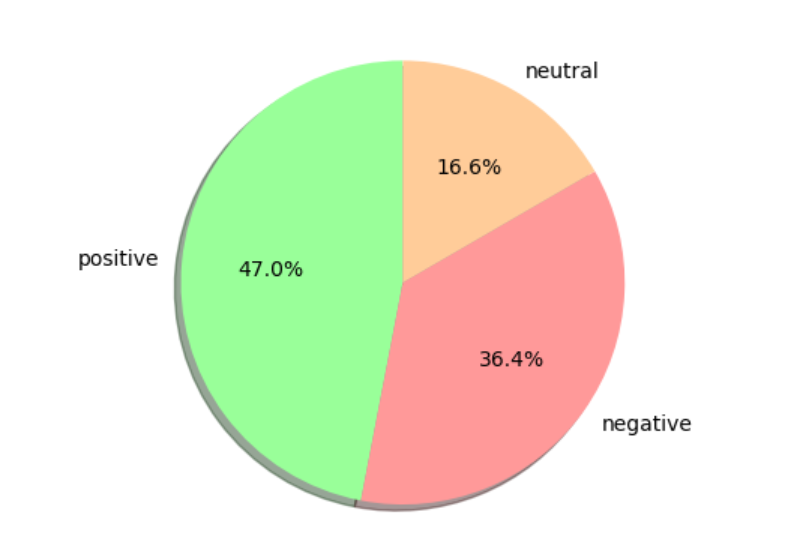
**We perform the same steps to pass the data through both the models**

**Result**

**Model 1 :**



**Model 2 :**



**Conclusion**

The sentiment analysis of both the models shows the dominance of positive sentiment by 47% - 50 % on 2024 Indian Budget.

The prediction of positive sentiments of both the models are pretty close.

1. **Information Extraction**

**Objective**

# In this project, we aim to do **named entity extraction(NER)** of 2024 Indian Budgetby scraping two websites **(CNBC-TV18 Edge article on Budget '24 — How it paints a picture of a resilient India and aljazeera’s article on Analysis: India’s 2024 interim budget shows a changing economy)** and passing it through two pre-trained **information extraction models from Hugging face repository.**

**Methodology**

**Data Acquisition**

We obtain textual data through web scraping.

For web scraping we use BeautifulSoup.

The data from both the websites is joined.

**Text Splitting**

The whole textual data is divided into an array of all sentences and then each sentence is passed through the NER model to get NER results.

**Color Coding Entities**

We use the library **termcolor** to color text and add **entity tags** according the **entity type** as obtained from the NER results.

**We perform the same steps to pass the data through both the models**

**Result**

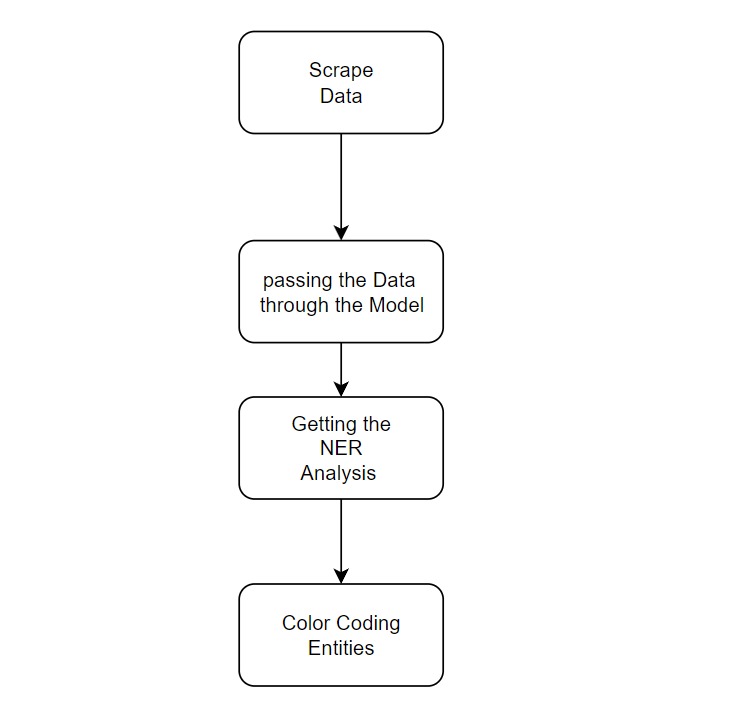
**Model 1 :**



**Model 2 :**



**NLP Diagram**

****

**Conclusion**

By identifying and labeling named entities present in the website's content, we gain valuable insights into the structure and context of the text.