

SVKM'S NARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES

**SCHOOL OF TECHNOLOGY AND
MANAGEMENT ENGINEERING**



**SUBJECT: NATURAL LANGUAGE PROCESSING
PROJECT REPORT/DOCUMENTATION**

SUBMITTED BY:

Name: Nishka Vijayvargiya
Class: B.Tech CE Sem VII
Batch: Batch 1
Roll no: T046
SAP Id: 70022200461

SUBMITTED TO:

Dr. Raj Gaurav Mishra
STME, NMIMS

1. Introduction

Natural Language Processing (NLP) enables machines to understand and process human language.

In today's information-heavy world, text summarization and keyword extraction are vital for quickly identifying the core meaning of large documents, news articles, or reports.

This project presents an NLP-based web application that automatically generates concise summaries and extracts relevant keywords using transformer models and semantic embeddings.

2. Objectives

1. Develop a user-friendly NLP web app for summarization and keyword extraction.
2. Use transformer-based models for high-quality summaries.
3. Use KeyBERT for semantically meaningful keyword extraction.
4. Enable user controls for summary and keyword tuning.
5. Visualize keyword importance through charts.

3. System Overview

System Architecture:

User Input → Preprocessing → Transformer Model (Summarization) → Summarized Text → KeyBERT Model → Keywords + Visualization

Tools and Technologies:

- Streamlit
- Transformers (BART-large-cnn)
- KeyBERT
- PyTorch (CPU)
- Matplotlib

4. Methodology

Step 1: User enters text into the Streamlit app.

Step 2: Text is summarized using facebook/bart-large-cnn (abstractive summarization).

Step 3: KeyBERT extracts semantically meaningful keyphrases.

Step 4: Results and relevance chart displayed, exportable as text file.

5. Results

Input: Long article

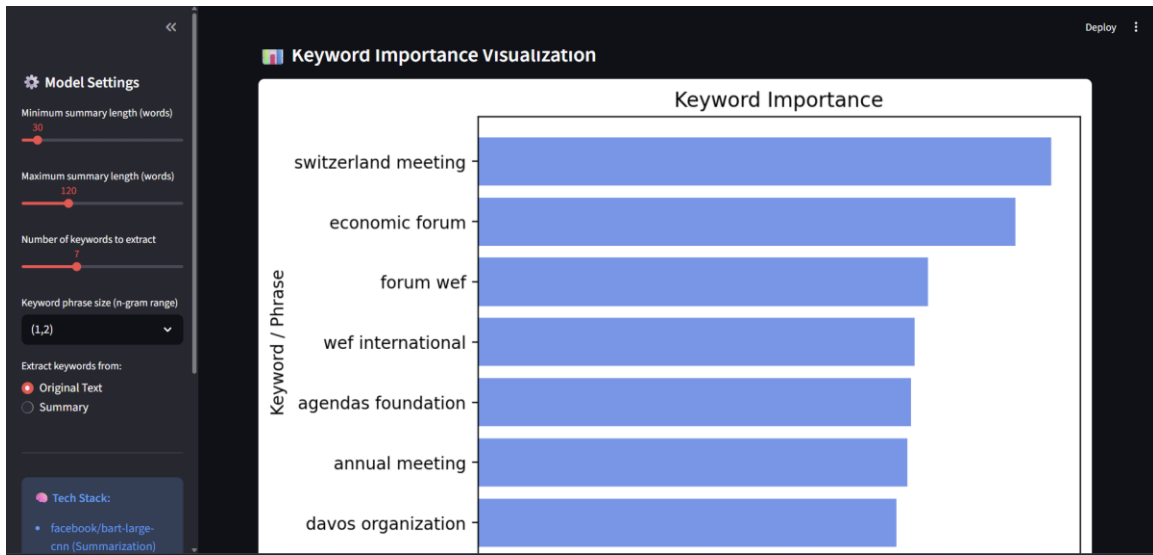
Output: Summarized text and top keywords

Visual Output: Bar chart showing keyword importance

File Output: summary_keywords.txt

The screenshot shows the 'Smart NLP Summarizer + Keyword Extractor' interface. On the left is a 'Model Settings' sidebar with sliders for 'Minimum summary length (words)' (set to 30), 'Maximum summary length (words)' (set to 120), and 'Number of keywords to extract' (set to 7). A dropdown for 'Keyword phrase size (n-gram range)' is set to '(1,2)'. Under 'Extract keywords from:', 'Original Text' is selected. A 'Generate Summary and Keywords' button is at the bottom of the sidebar. The main area has a title 'Smart NLP Summarizer + Keyword Extractor' and a description: 'This app uses advanced Transformer models for summarization and BERT embeddings for keyword extraction. Adjust parameters in the sidebar to experiment in real time.' Below this is a text input area with the placeholder 'Paste your text here:' containing a paragraph about the World Economic Forum (WEF). A 'Deploy' button is in the top right corner.

The screenshot shows the 'Generated Summary' and 'Keyword Extraction' results. The 'Generated Summary' section displays a green box with the summarized text: 'The World Economic Forum (WEF) is an international advocacy non-governmental organization and think tank, based in Cologne, Canton of Geneva, Switzerland. It was founded on 24 January 1971 by German engineer Klaus Schwab. Its stated mission is "improving the state of the world by engaging business, political, academic, and other leaders of society to shape global, regional, and industry agendas"'. The 'Keyword Extraction' section, titled 'Keyword Extraction', shows 'Extracted from: Original Text' and a list of keywords with their relevance scores: '• switzerland meeting — relevance: 0.59', '• economic forum — relevance: 0.56', '• forum wef — relevance: 0.46', '• wef international — relevance: 0.45', '• agendas foundation — relevance: 0.45', '• annual meeting — relevance: 0.44', and '• davos organization — relevance: 0.43'. The 'Model Settings' sidebar on the left is identical to the previous screenshot, but the 'Tech Stack' button is now visible at the bottom left.



6. Applications

- Research and news summarization
- Academic abstract generation
- SEO and content indexing
- Business intelligence text analysis

7. Conclusion

The Smart NLP Summarizer and Keyword Extractor demonstrates the potential of Transformer-based NLP applications for efficient and interpretable text understanding.