Lab Activity 7: Climate Awareness NLP System

Name: Raghav Marwaha
Roll No.: E23CSEU1229
Batch: 41
Course: NLP Laboratory

---

1. Objective

The objective of this lab activity is to develop an intelligent Natural Language Processing system that can:

1. Detect whether climate-related statements are **factual** or **misinformation**
2. Generate **concise policy summaries** from lengthy environmental documents

This project contributes to public understanding and combating misinformation related to climate change.

The work aligns with **SDG 13 – Climate Action**, which emphasizes the need for awareness and responsible communication about environmental issues.

---

2. Dataset Information

Detail	Description
Dataset Used	**CLIMATE-FEVER (Hugging face)**
Domain	Climate change facts, policies, scientific claims
Task Type	Binary Classification
Labels	Factual vs. Misinformation
Use Case	Verifying climate claims and filtering biased information

Preprocessing Steps:

- Text cleaning (symbols, special characters)
- Tokenization using transformer tokenizer
- Train/Test/Validation split for fair evaluation

---

3. Methodology

Classification System

A transformer-based model is trained to categorize input statements.

Component	Purpose
DistilRoBERTa Transformer	Contextual understanding of scientific and political climate language
Dense Output Layer	Predict factual or misinformation label

Training Techniques:

- Adam optimizer
- Cross-entropy loss
- Checkpointing and early stopping for best performance retention

Summarization System

After classification, policy text is summarized using:

Model	Function
BART-CNN Abstractive Summarizer	Produces human-like, concise summaries of climate policy

This makes complex documents easier to interpret.

---

4. Results

Extracted from the executed Jupyter Notebook.

Metric	Score
Best Accuracy	75.27%
Best F1-Score	65.46%