**Project Name:** Hindi Text Summarization Using Transformer Models

**Motivation (Abstract):** With the increasing amount of digital content in Hindi, reading and comprehending long articles is time-consuming. Hindi text summarization aims to convert lengthy texts into concise summaries while preserving the main ideas. This project focuses on building an automated summarization system using transformer-based models, enabling users to quickly grasp the key points of any Hindi text.

**Process (Pipeline / Steps):**

1. **Problem Understanding**
   * **Goal:** Convert long Hindi texts into short summaries in Hindi itself.
   * **Summarization type:**
     + **Abstractive:** Generate new sentences capturing the essence of the text.
     + **Extractive:** Select key sentences from the original text.
2. **Dataset Collection**
   * Collect Hindi text and their summaries from sources like:
     + XL-Sum (Hindi portion)
     + HindiSumm datasets
     + Hindi news articles with abstracts
3. **Data Preprocessing**
   * Clean text: remove HTML tags, special characters, and extra spaces.
   * Normalize Devanagari script.
   * Handle punctuation and sentence boundaries properly.
4. **Model Selection**
   * Use large pre-trained language models (~3B parameters), such as:
     + mT5-XL
     + LLaMA-3.2
5. **Text Formatting for Model**
   * **Input:** “summarize: ”
   * **Output:**
6. **Fine-Tuning the Model**
   * Use frameworks like Hugging Face Transformers.
   * Apply LoRA/PEFT for low compute environments.
   * Train with article–summary pairs.
7. **Evaluation**
   * Quantitative metrics: ROUGE, BLEU, BERTScore.
   * Qualitative assessment: Human evaluation for readability and factual accuracy.
8. **Inference (Prediction)**
   * Input Hindi text to the trained model.
   * Generate summaries using decoding methods such as beam search.
9. **Deployment**
   * Host the model on cloud platforms like AWS, GCP, or Hugging Face Spaces.
   * Provide a simple web or mobile interface for users to input text and get summaries.
10. **Use in Real World**
    * Summarize Hindi newspapers, government circulars, or educational content.
    * Enable faster understanding for students, professionals, and general readers.

**Learnings:** - Understanding Hindi NLP challenges like complex morphology and limited datasets. - Hands-on experience with transformer models and fine-tuning techniques. - Learning to evaluate summaries using ROUGE, BLEU, and human assessment. - Importance of preprocessing and proper tokenization in NLP pipelines.

**Conclusion:** The project demonstrates how transformer-based models can effectively summarize Hindi texts. It reduces reading time, enhances information accessibility, and can be extended to news aggregation, educational applications, and digital assistants. The work also provides valuable experience in Hindi NLP, model fine-tuning, and deployment.