ML project-Text summarization

The product design
What & Why

Background

Customer's goals and pains: Users need a way to quickly digest long pieces of text, such as research papers, legal documents, or news articles. Reading through these texts is time-consuming, and users want concise yet accurate summaries.

Value proposition

Value and pain alleviation: The text summarization model will allow users to save time and effort by providing summaries that capture the essence of long documents. It improves productivity, especially for professionals dealing with large amounts of written content.

objectives

What we will focus on:
-Develop an AI system that can summarize documents in multiple formats (PDFs, text, web articles)
.-Ensure high accuracy and readability in the summarization.

-Allow users to customize the length and format of the summary.

solution

Solution definition: Implement a text summarization product using LLMs like(transformers,groq) models.
Out-of-scope: Summarization for multimedia content like video or audio files, unless this is an explicit future goal.

Feasability

-Evaluate if there's sufficient labeled text summarization data.
-Identify if pre-trained models can be fine-tuned for this application.
-Ensure the resources (computation, data) and expertise required are available.
-Have the minimum cash for funding

data

Training data: Use datasets like CNN/Daily Mail or other news datasets for summarization.
Production data: User-uploaded documents in various formats (PDF, Word, HTML, plain text).
Labeling: For supervised learning, human-annotated summaries or third-party datasets will be required.

Modeling

-Fine-tune pre-trained models to summarization.
-Implement iterative improvements based on feedback and evaluation.
-Use extractive models like Groq or transformers for summarizing based on key sentence extraction

Metric

Key metrics: ROUGE scores (Recall-Oriented Understudy for Gisting Evaluation) to evaluate how well the generated summaries match reference summaries.
Additional metrics: User satisfaction based on human evaluation, readability scores, and time saved.

Evaluation

This ca done ,
Offline: Test on benchmark datasets, checking ROUGE scores and human readability.
Online: A/B test with users to assess performance in real-time, gather feedback on summary accuracy and usability.

Feedback

Gather user feedback on the clarity, accuracy, and usefulness of the generated summaries. Continuous model retraining and updating based on user feedback to improve performance over time.

Project

Define the team roles:
Data Scientist: Works on model selection and training.
Software Engineer: Integrates the model into a scalable, user-friendly application.
UX/UI Designer: Focuses on user interaction, ensuring easy-to-use interfaces.
Timeline: Set milestones for data collection, model development, and user testing.