Project Progress Report: 09/17/2025
To: Basu Sir
From: Melina Singh
Date: 09/17/2025
Subject: Progress Update – Pipeline Adaptation for New Data Server & LLM Integration

Summary:
Following up on the Gemini LLM research and pipeline preparation completed yesterday, today's focus was on adapting the existing text processing pipeline to accommodate input from the new data server. The goal was to ensure seamless compatibility between the server-side data structure and the LLM-driven NLP module, especially for sentiment analysis and classification tasks.

I examined the data format and structure from the new server and updated the preprocessing logic accordingly to maintain consistency in token cleaning, sentence reconstruction, and prompt formatting. Special attention was given to handling metadata, encoding issues, and batch processing requirements for large-scale inference with the LLM.

What Was Done Today:

- Analyzed data schema and structure from the new data server.

- Refactored pipeline modules to support data ingestion from the new source.

- Ensured sentence-level formatting remains compatible with the LLM prompt expectations.

- Improved batch processing mechanisms to enable efficient prompt construction and LLM input handling.

- Implemented data validation checks to catch formatting or encoding inconsistencies early in the pipeline.

Current Status:

- The preprocessing pipeline is now fully compatible with the new data server.

- LLM-ready inputs are being generated from live server data, pending final Gemini API testing.

- Integration with Gemini remains blocked due to API access/billing constraints.

Next Steps:

- Continue resolving Gemini API access issues (billing or quota-related).

- Once API access is restored, begin sending new server data through the pipeline to:

    - Test end-to-end sentiment analysis

    - Parse and evaluate model responses on live Nepali data

- Monitor server integration for stability and refine error handling as needed

Note:
All recent code updates reflecting the server-side changes have been committed to the Company repository