Project Progress Report: 09/21/2025

To: Basu Sir
From: Melina Singh
Date: 09/21/2025
Subject: Progress Update – Gemini API Integration for Sentiment Analysis

Following the earlier setup and validation of the data pipeline for compatibility with the new server infrastructure, today's work focused on advancing the **LLM integration phase**, specifically implementing **sentiment analysis** using the **Gemini 1.5 Flash model**. The integration marks a key step in enabling intelligent categorization and analysis of Nepali-language news content through large language models.

The updated pipeline now supports real-time inference of sentiment from article titles, classifying each as **Positive**, **Negative**, or **Neutral**. A dedicated sentiment classification function was implemented with built-in retry logic using exponential backoff and jitter. This ensures that temporary service interruptions from the Gemini API do not impact pipeline stability.

Prompt design was carefully refined to maximize clarity for the LLM while keeping it lightweight. The prompt is constructed dynamically using a fixed set of sentiment labels, and results are retrieved in a clean format for logging or further processing. The output was tested against a sample set of articles, and preliminary results indicate high alignment with expected sentiment tones.

Additional refinements included:

- Handling missing or incomplete data gracefully (e.g., skipping articles without valid titles).
- Supporting fallback to alternative fields (like content) where necessary.
- Isolating model-related logic for future reuse and easier testing.

Next Steps:

- Begin logging outputs to structured formats (CSV/JSON) for review and archival.
- Expand coverage to include article body/content where sentiment is more nuanced.

- Track response accuracy and begin compiling evaluation metrics for model feedback.
- Modularize current functions into reusable components within the broader NLP pipeline.