Smart Lead Gen: A Generative AI for Hyper-Personalized Professional Outreach

- Milestone 5: Model Evaluation.
- Team: Group 7
- Course: DS and AI Lab
- Submission date: November 7th, 2025

This report details the evaluation setup, metrics, and findings for two key models: a personality classification model and an LLM-based email generation model.

1. Personality Classification Model (DistilBERT-LoRA)

This stage of the project aimed to predict an individual's Myers-Briggs Type Indicator (MBTI) personality type based on their written text.

Methodology and Setup

The model is a DistilBERT classifier fine-tuned using the LoRA (Low-Rank Adaptation) PEFT (Parameter-Efficient Fine-Tuning) method. The model was trained on a specialized MBTI dataset containing over 100,000 samples. Key metrics used for this task were **F1-score and accuracy**.

Performance Analysis and Findings

During the training process, the model demonstrated a gradual decrease in both training and validation losses, alongside a steady improvement in accuracy. Blackbox optimization was used to tune the LoRA hyperparameters, which successfully improved the model's convergence, stability, and generalization.

A key part of this evaluation was comparing the LoRA method against a traditional full fine-tuning approach:

- **Efficiency:** The LoRA method significantly reduced both training time and GPU usage. The full fine-tuning approach required **three times more training time** to achieve a performance level similar to the LoRA-based model.
- **Weight Updates:** While more time-consuming, the loss reduction per epoch was steeper in the full fine-tuning approach, indicating more substantial weight updates during training.
- **Conclusion:** Given the current limitations in GPU resources and time, LoRA is the preferred and more practical option.

Limitations and Future Work

Further training is required to improve the model's overall accuracy. This need stems from the inherent complexity of the MBTI dataset and the aforementioned resource constraints. Future work will focus on improving model accuracy once sufficient GPU resources become available.

---

2. LLM Email Generation Model

This evaluation (Milestone 5) assesses the performance of the 'generate_email' function, a prompt-based LLM developed in Milestone 4.

Objective

The model is designed to generate short, personalized, and ready-to-send B2B cold emails from a structured JSON lead profile.

The generated emails must adhere to several strict design rules:

- **Length:** Keep the email concise (target 60-120 words).
- **Call-to-Action (CTA):** Include **exactly one** clear CTA.
- **Links:** Contain **no links** in the first email.
- **Grounding:** Use *only* information from the provided lead and style profiles.
- **Structure:** Use a friendly greeting with the recipient's name and a proper closing with the sender's name.

Evaluation Setup

The model was tested on **7 unseen test leads** from diverse B2B contexts (e.g., fintech, SaaS, e-commerce, healthtech). Each lead profile was detailed, including data such as role, company history, recent activity, skills, and inferred MBTI.

Since there is no single "correct" email, performance was measured using **rule-based checks** rather than traditional accuracy. These checks included:

- Basic structure (subject/body present).
- Word count against the target band.
- Presence of correct greeting and closing names.
- Absence of URL-like text.
- A count of CTA verbs (e.g., "schedule," "chat") to verify the "exactly one" rule.
- Checks for personalization (lead's name/company) and objective grounding.
- A final **1-10 quality score** was assigned to quantify how well the email matched all rules.

Key Findings

The model's performance was mixed, with clear successes and specific areas needing improvement. Overall quality scores for the 7 emails fell in the middle of the 1-10 scale.

Successes:

- **Reliable Structure:** The model consistently produced a subject and body for all 7 test cases.
- **No Links:** The model successfully adhered to the "no links" rule.
- **Objective Grounding:** The model performed well at incorporating the lead's specific objective, using relevant terms like "activation," "churn," or "onboarding".

Areas for Improvement:

- **Length:** Emails were frequently longer than the 60-120 word target. They often read more like full examples or templates rather than short, concise cold emails.
- **Placeholders:** The model often used placeholders (e.g., [First Name]) instead of inserting the actual name from the profile. This caused the strict greeting/closing checks to fail, even when the email's tone was polite.
- **CTA Consistency:** While a CTA was generally present, the *number* of CTAs was inconsistent and did not reliably meet the "exactly one" rule.
- **Personalization:** The use of the lead's name and company was inconsistent, appearing in some emails but not all.

Overall Assessment

The current prompt-based model generates emails that are readable, polite, and relevant to the lead's business objective. However, it does not yet fully and consistently adhere to the specific design rules, particularly regarding length, placeholder usage, and CTA count.