

KẾ HOẠCH & NỘI DUNG ĐỀ TÀI

Tên đề tài: Ứng dụng AI đánh giá phản hồi học phần theo cảm xúc và chủ đề

1. Phân tích bài toán (AI Thinking)

1. Mô tả bài toán:

- Bối cảnh: Giảng viên/Bộ phận khảo thí cần xử lý lượng lớn phản hồi học phần để cải thiện chất lượng dạy-học.
- Mục tiêu: Tự động phân loại cảm xúc (positive/neutral/negative) và rút trích chủ đề chính.
- Người dùng: Giảng viên, ban đào tạo, bộ phận khảo thí.

2. AI Thinking Steps:

- Mục tiêu AI:

- Đạt $\geq 80\%$ accuracy phân loại cảm xúc.
- Rút trích 5–10 chủ đề chính.

- Phương pháp AI:

- Sentiment Analysis: BERT fine-tuning + LDA-classifier (baseline hoặc tài nguyên hạn chế).
- Topic Modeling: LDA-topic modeling / BERTopic.

- Dữ liệu & rủi ro:

- Nguồn: Kaggle, Google Forms/LMS.
- Rủi ro: Bias (thiên lệch lớp/chủ đề), quyền riêng tư (ẩn danh).

- Công cụ / Mô hình: Python (pandas, scikit-learn, spaCy), Hugging Face Transformers, Gensim, BERTopic, Jupyter/Colab.

- Bối cảnh xã hội/đạo đức: Ẩn danh sinh viên (loại bỏ tên/ID), cân bằng dữ liệu giữa các lớp và chủ đề.

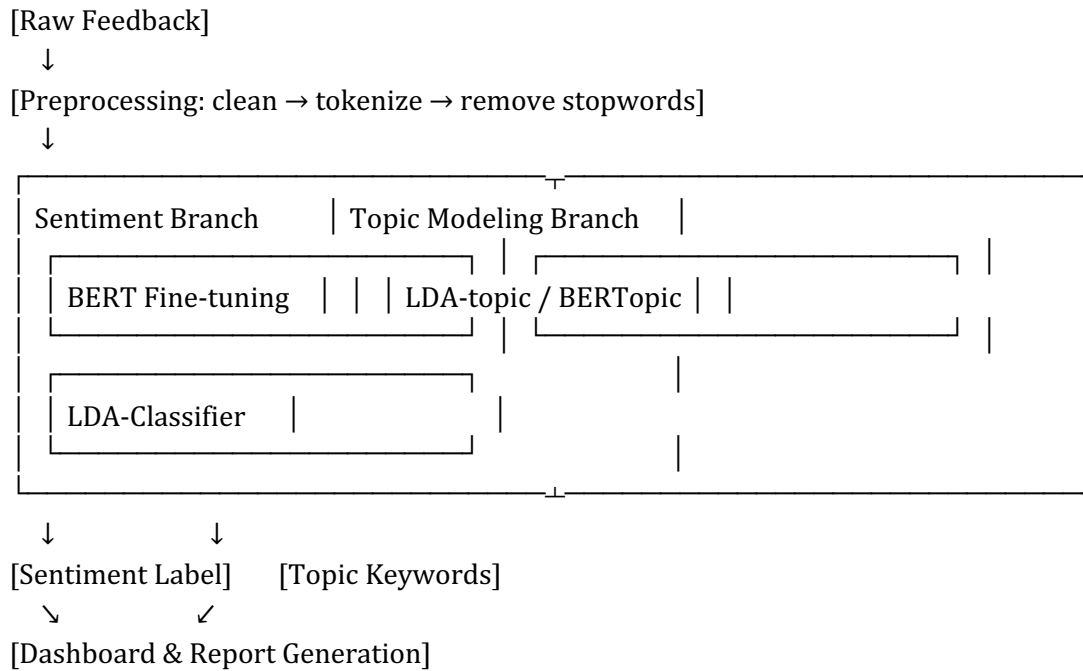
2. Thiết kế mô hình & pipeline

1. Input – Output:

- Input: File CSV/JSON chứa phản hồi văn bản.
- Output:

- Nhãn cảm xúc: {positive, neutral, negative}.
- Danh sách chủ đề chính (topic keywords).

2. Flowchart quy trình:



3. Lý do chọn kiến trúc:

- BERT: Xử lý ngữ cảnh tốt, accuracy cao.
- LDA-classifier: Baseline nhanh, dễ giải thích, nhẹ tài nguyên.
- LDA-topic/BERTopic: Rút trích chủ đề hiệu quả, trực quan.

3. Kết quả thử nghiệm

1. Huấn luyện thử:

- Train/test split 80/20.
- BERT: 3 epochs, lr=2e-5.
- LDA-classifier: TF-IDF → LDA.
- LDA-topic: num_topics=7, passes=10.

Mô hình	Accuracy	Precision	Recall	Coherence
BERT-base	82 %	0.80	0.83	-
LDA-classifier (TF-IDF + LDA)	72 %	0.70	0.71	-
Logistic + TF-IDF	75 %	0.76	0.74	-
LDA-topic (7 topics)	-	-	-	0.42

Chú thích: Coherence đo lường tính nhất quán chủ đề.

2. Minh họa kết quả:

- Bảng: Accuracy & F1-score từng nhãn.
- Biểu đồ: Phân phối phần trăm cảm xúc; độ phổ biến chủ đề.
- Ví dụ: “Giảng viên nhiệt tình” → positive; topic: [“giảng viên”, “nhiệt tình”].

4. Đánh giá & Tư duy phản biện

1. Ưu điểm – Hạn chế:

- Ưu điểm:
 - BERT: độ chính xác cao, xử lý ngữ cảnh tốt.
 - LDA-classifier: triển khai nhanh, giải thích dễ, không cần GPU.
- Hạn chế:
 - BERT: cần GPU, thời gian huấn luyện lâu.
 - LDA-classifier: không bắt ngữ cảnh phức tạp.

2. Yếu tố xã hội/đạo đức/công bằng:

- Ẩn danh sinh viên.
- Cân bằng dữ liệu để giảm bias.
- Giải thích được quyết định của mô hình.

3. Gợi ý cải tiến & Triển khai thực tế:

- Thu thập phản hồi thực tế từ LMS.

- Mở rộng multi-label classification (nhiều khía cạnh).
- Xây dựng API + dashboard tương tác.
- Liên tục so sánh BERT & LDA-classifier, chọn mô hình phù hợp.

Danh sách Task cần thực hiện & Phân công (Có thứ tự ưu tiên)

Priority	Task	Phụ trách	Nội dung	Công cụ
1	Setup GitHub workspace	Đức	Tạo repo, khởi tạo thư mục, branch, .gitignore, PR template, badge CI, README	Git/GitHub
2	Phân tích bài toán & AI Thinking	Hồng, Đức	Mô tả bối cảnh, mục tiêu, phương pháp, dữ liệu, rủi ro	Google Docs, pandas, XMind
3	Thu thập & Tiền xử lý dữ liệu	Trọng, Nhật	Lấy dataset, làm sạch, tokenize, remove stopwords	pandas, spaCy/ NLTK
4	Thiết kế mô hình & pipeline	Hồng, Khang	Vẽ flowchart, mô tả input/output	draw.io/ Lucidchart
5	Triển khai Sentiment Analysis (BERT)	Đức, Khang	Fine-tune BERT, lưu checkpoints	Transformers, PyTorch, Colab
6	Triển khai baseline với LDA-classifier	Đức, Trọng	TF-IDF → LDA; so sánh metrics	scikit-learn
7	Triển khai Topic Modeling	Nhật, Long	LDA-topic/BERTopic; đánh giá coherence	Gensim, BERTopic, pyLDAvis
8	Đánh giá & So sánh mô hình	Hồng, Đức, Khang	Tính metrics; vẽ biểu đồ	pandas, matplotlib

9	Phân tích đạo đức & tác động xã hội	Trọng, Nhật	Đánh giá bias, privacy; đề xuất giải pháp	Google Docs, Canva
10	Soạn thảo báo cáo chi tiết	Hồng, Đức, Nhật	Viết 4 mục nội dung bắt buộc cần có; chèn sơ đồ, bảng, biểu đồ	Word/LaTeX
11	Review & Hoàn thiện báo cáo + self-assessment	Hồng, Đức	Kiểm tra chính tả, định dạng; hoàn thiện bảng đánh giá	Grammarly/VnCrate
12	Thiết kế slide thuyết trình	Trọng, Khang, Long	Tóm tắt nội dung, layout 5 bullet/slide	PowerPoint/Slides
13	Upload code & notebook lên GitHub	Khang, Long	Đẩy code; cập nhật README; tag version	Git/GitHub
14	Quay video demo	Hồng, Trọng	Ghi màn hình chạy mô hình; voice-over	OBS Studio/Loom
15	Diễn tập & Hoàn thiện thuyết trình	Đức, Nhật, Khang, Long	Chạy thử slide; điều chỉnh thời gian	Zoom/Google Meet
16	Kiểm tra cuối cùng & Nộp bài	Cả nhóm	Đánh giá toàn bộ; gửi link GitHub & Drive	Email, Drive