

Span Detection and Explanation Generation for Aspect-based Compliment Analysis on Multi-domain Feedback

Phạm Tiến Dương, Đặng Thị Thúy Hồng, Nguyễn Văn Kiệt

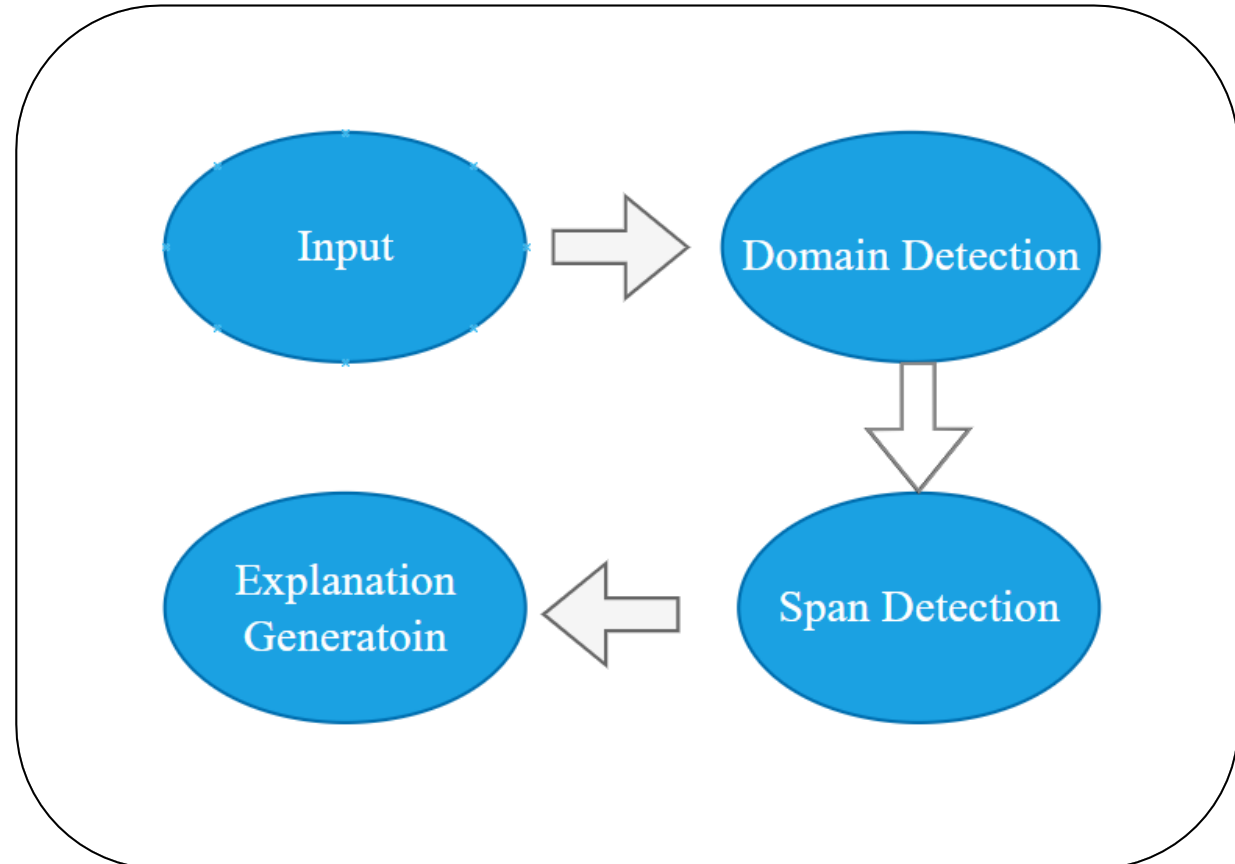
20521222@gm.uit.edu.vn, 20520523@gm.uit.edu.vn, kietnv@gm.uit.vn

Khoa Khoa học và Kỹ thuật Thông tin, Trường Đại học Công nghệ Thông tin,
Đại học Quốc gia Hồ Chí Minh

1. MỤC TIÊU

- Ngày nay, với sự phát triển bùng nổ của thương mại điện tử, người dùng thường có xu hướng tìm kiếm lời khuyên từ những người đi trước đã mua sản phẩm trước khi mua một mặt hàng. Chính vì thế, số lượng đánh giá ngày càng tăng và trở thành nguồn tài nguyên quý giá cho khách hàng và doanh nghiệp.
- Trong đề tài này, chúng tôi tiến hành xây dựng bộ dữ liệu UIT-ViSE4AM, cùng với đó áp dụng Machine Learning và Transfer Learning để giải quyết các tác vụ Domain Classification, Span detection và Explanation Generation.

Mô tả bài toán

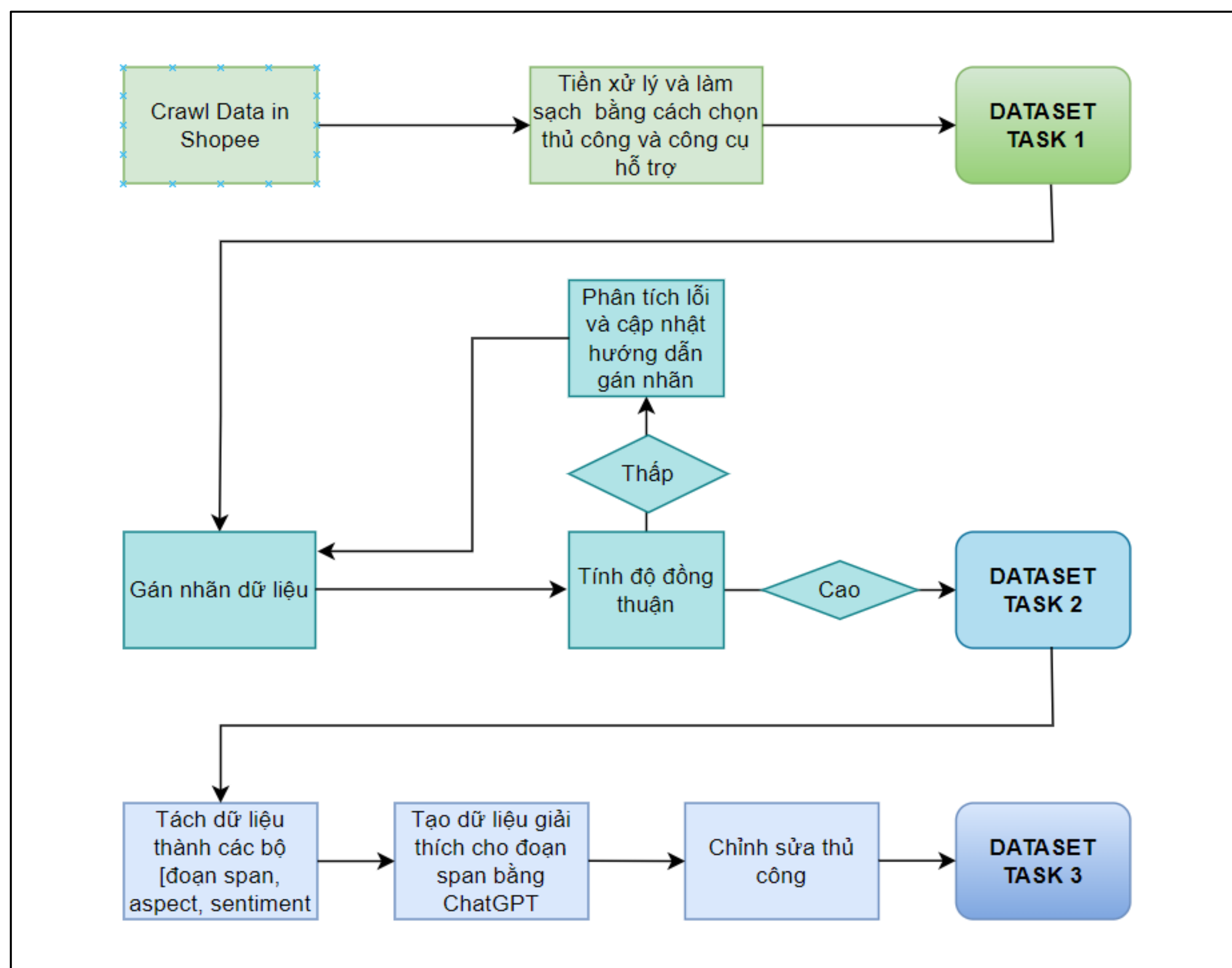


Hình 1: Đầu vào và đầu ra của bài toán

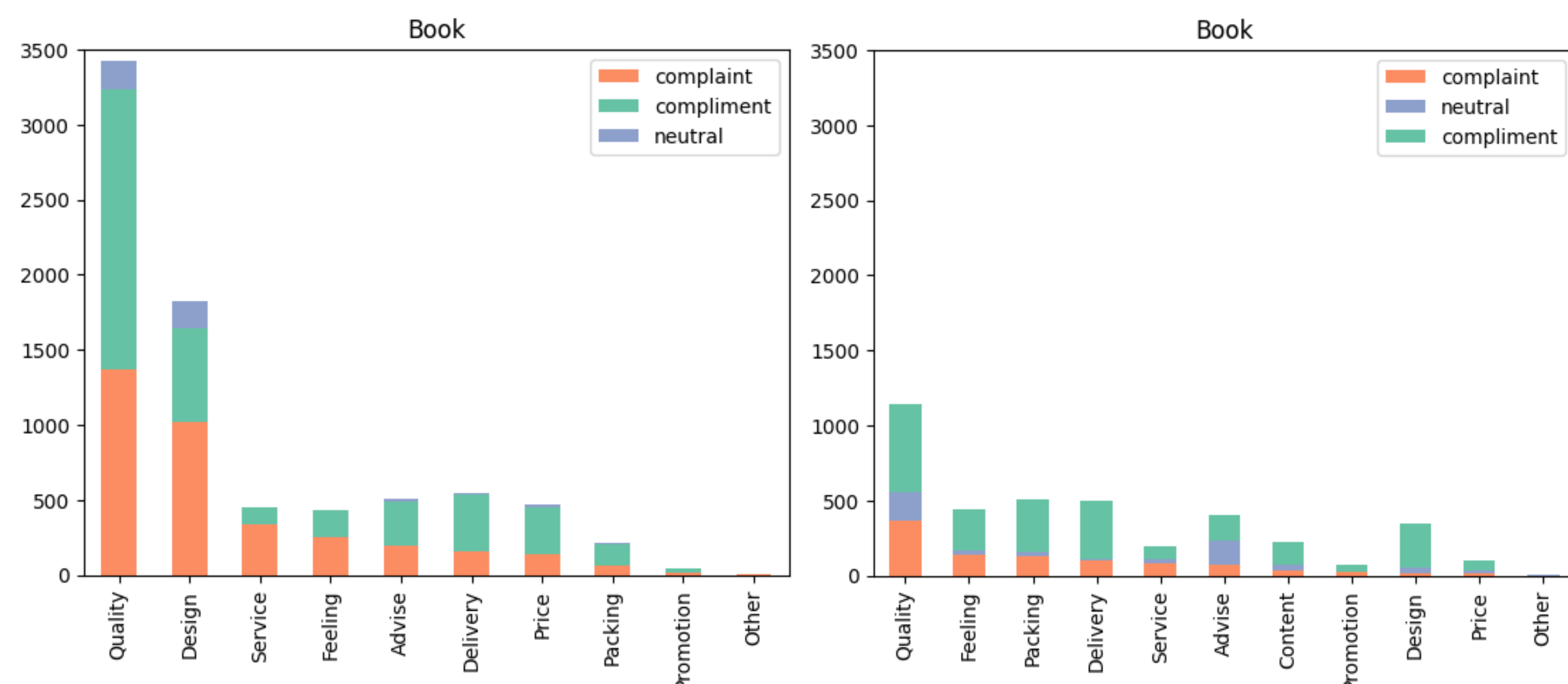
2. BỘ DỮ LIỆU

Chúng tôi xây dựng bộ dữ liệu UIT-ViSE4AM với 2 lĩnh vực Book và Fashion với quy trình như ở hình 2, kết quả thu được:

- 6000 câu bình luận cho tác vụ Domain Classification.
- 4483 câu bình luận gồm 11838 cặp span được gán nhãn với độ đồng thuận 81,25% cho tác vụ Span detection.
- 3953 cặp giải thích cho tác vụ Explanation Geneneration.



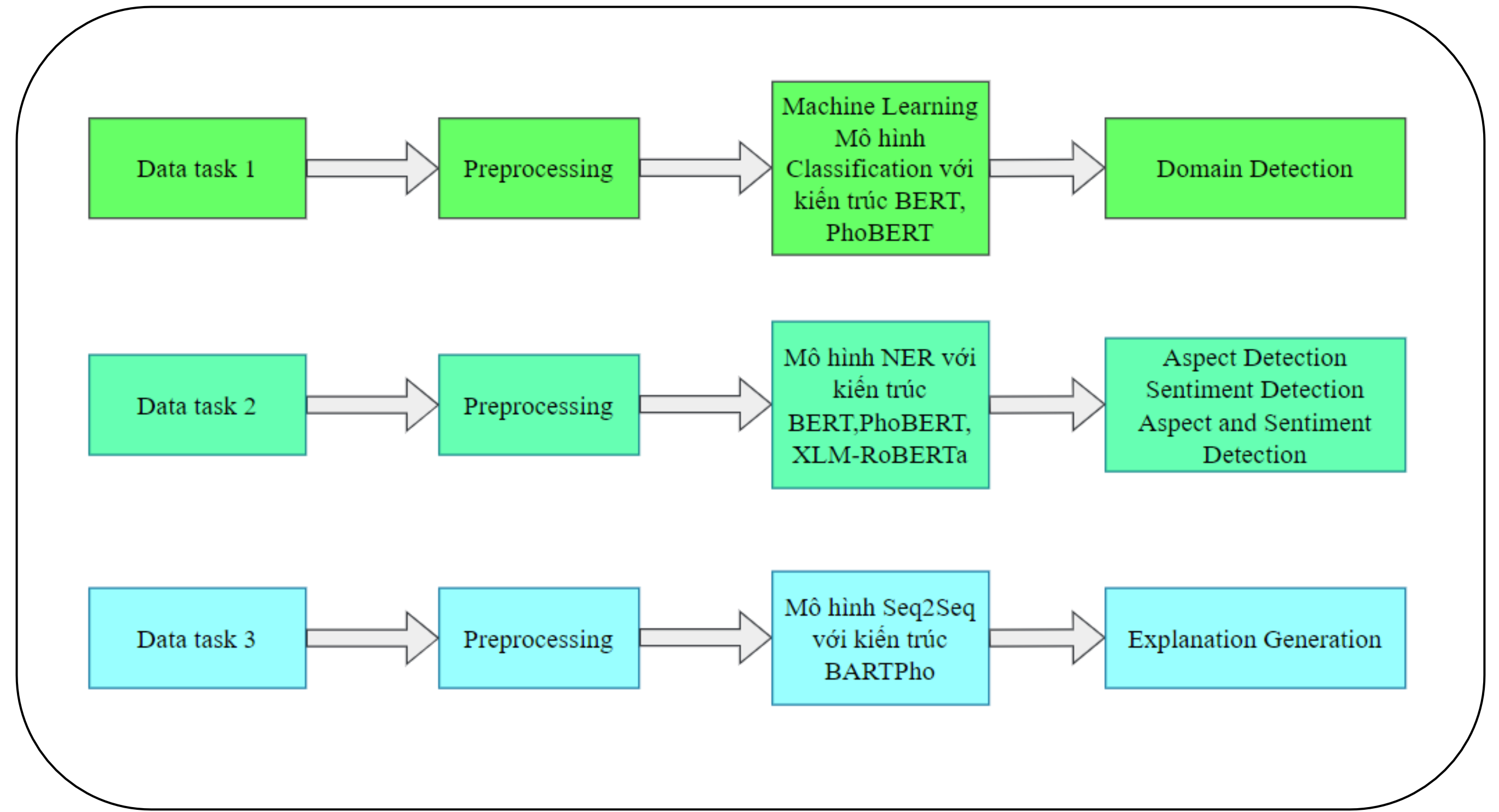
Hình 2: Quy trình xây dựng bộ dữ liệu UIT - ViSE4AM



Hình 3: Thống kê số lượng Aspect và Sentiment của mỗi Domain

4. PHƯƠNG PHÁP THỰC HIỆN

Chúng tôi tiến hành thực hiện đề tài với 3 tác vụ theo quy trình sau:



Hình 4: Phương pháp thực hiện đề tài

4. THỰC NGHIỆM VÀ KẾT QUẢ

Chúng tôi tiến hành đánh giá các mô hình và đạt được kết quả như sau.

Bảng 1: Kết quả F1- Score trên mô hình máy học ở tác vụ Domain Classification

MODELS	TF IDF			COUNTER VECTO		
	Book	Fashion	Average	Book	Fashion	Average
SVM	90.78%	91.48%	91.13%	89.43%	90.42%	89.92%
NAIVE BAYES	75.77%	62.39%	69.08%	74.85%	57.50%	66.18%
DECISION TREE	87.62%	87.68%	87.65%	88.57%	88.61%	88.59%
RANDOM FOREST	90.34%	90.57%	90.46%	89.70%	90.03%	89.86%
LOGISTIC REGRESSION	90.75%	91.34%	91.05%	90.43%	91.15%	90.79%

Bảng 2: Đánh giá các mô hình ở tác vụ Domain Classification

MODELS		Precision	Recall	F1-Score
MACHINE LEARNING	SVM	91.40%	91.14%	91.13%
	NAIVE BAYES	75.26%	70.53%	69.08%
	DECISION TREE	88.59%	88.59%	88.59%
	RANDOM FOREST	90.34%	90.29%	90.29%
	LOGISTIC REGRESSION	91.24%	91.06%	91.05%
TRANSFER LEARNING	PHOBERT	91.59%	91.57%	91.57%
	BERT	90.75%	90.72%	90.72%

Bảng 3: Đánh giá F1-Score trên các mô hình ở tác vụ Span Detection

Mô hình	Aspect		Sentiment		Aspect#Sentiment	
	Fashion	Book	Fashion	Book	Fashion	Book
PhoBERT	46,61%	41,83%	50,88%	45,44%	42,72%	32,86%
XLM-ROBERTa	50,34%	48,08%	52,21%	48,78%	46,36%	40,46%
BERT	46,30%	42,59%	48,21%	42,20%	47,29%	34,43%

Bảng 4: Đánh giá các kết quả Bleu-Score ở tác vụ Explanation Generation

	Bleu 1	Bleu 2	Bleu 3	Bleu 4
BARTPho	54,78%	50,80%	47,53%	44,59%

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

- Kết luận:** Kết quả đạt được 91,48% trên F1-Score cho mô hình Support Vector Machine với phương pháp TF-IDF ở Domain Classification Task, 47,29% trên F1- Score ở Span Detection Task trên miền dữ liệu Fashion và 54,78% trên Bleu-1 cho tác vụ Explanation Generation.
- Hướng phát triển:** Chúng tôi đưa ra một số hướng phát triển như sau: (1) Mở rộng thêm bộ dữ liệu và tập khía cạnh trên nhiều miền dữ liệu khác. (2) Áp dụng thêm các pre-train model đa ngôn ngữ để tăng độ chính xác cho các tác vụ. (3) Cải thiện hiệu suất của nghiên cứu này để có thể áp dụng vào nhiều mảng nghiên cứu khác như phát triển bài toán Machine Comprehension Reading cho Tiếng Việt, khai thác ý kiến và phân tích khiếu nại người dùng.