



**BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG**

NGHIÊN CỨU GIẢI PHÁP NÂNG CAO CHẤT LƯỢNG PHÂN TÍCH CẢM TÍNH CỦA KHÁCH HÀNG TRONG LĨNH VỰC DỊCH VỤ

Mã số đề tài: 2025-22



Thành viên thực hiện: Mai Nguyễn Xuân Thảo
Trần Phạm Song Nguyên
Hoàng Quốc Việt

Giảng viên hướng dẫn: TS. Đặng Đại Thọ

Đà Nẵng, ngày 23 tháng 9 năm 2025

Đánh giá trực tuyến

- **Phản ánh** cảm xúc, mức độ hài lòng và kỳ vọng của người tiêu dùng
- **Định hướng** hành vi tiêu dùng, **hình thành** uy tín thương hiệu
- **Tác động mạnh** đến chiến lược phát triển dịch vụ của doanh nghiệp
- **82% khách hàng** nghiên cứu sản phẩm trước khi mua hàng tại cửa hàng

Ngành dịch vụ Việt Nam

Động lực tăng trưởng **quan trọng** của nền kinh tế



Du lịch: **17,5 triệu lượt khách quốc tế** (2024)



Hàng không: **CAGR 17,4%** (2016–2021), dự kiến >20% (2016–2026)



Thương mại điện tử: **25 tỷ USD** (2024)

Động lực nghiên cứu

Hạn chế của các nghiên cứu trước đây

Chỉ sử dụng nội dung: **Có thể mất thông tin quan trọng**

Không sử dụng emoji vào việc phân tích: **Làm mất thông tin có ý nghĩa**

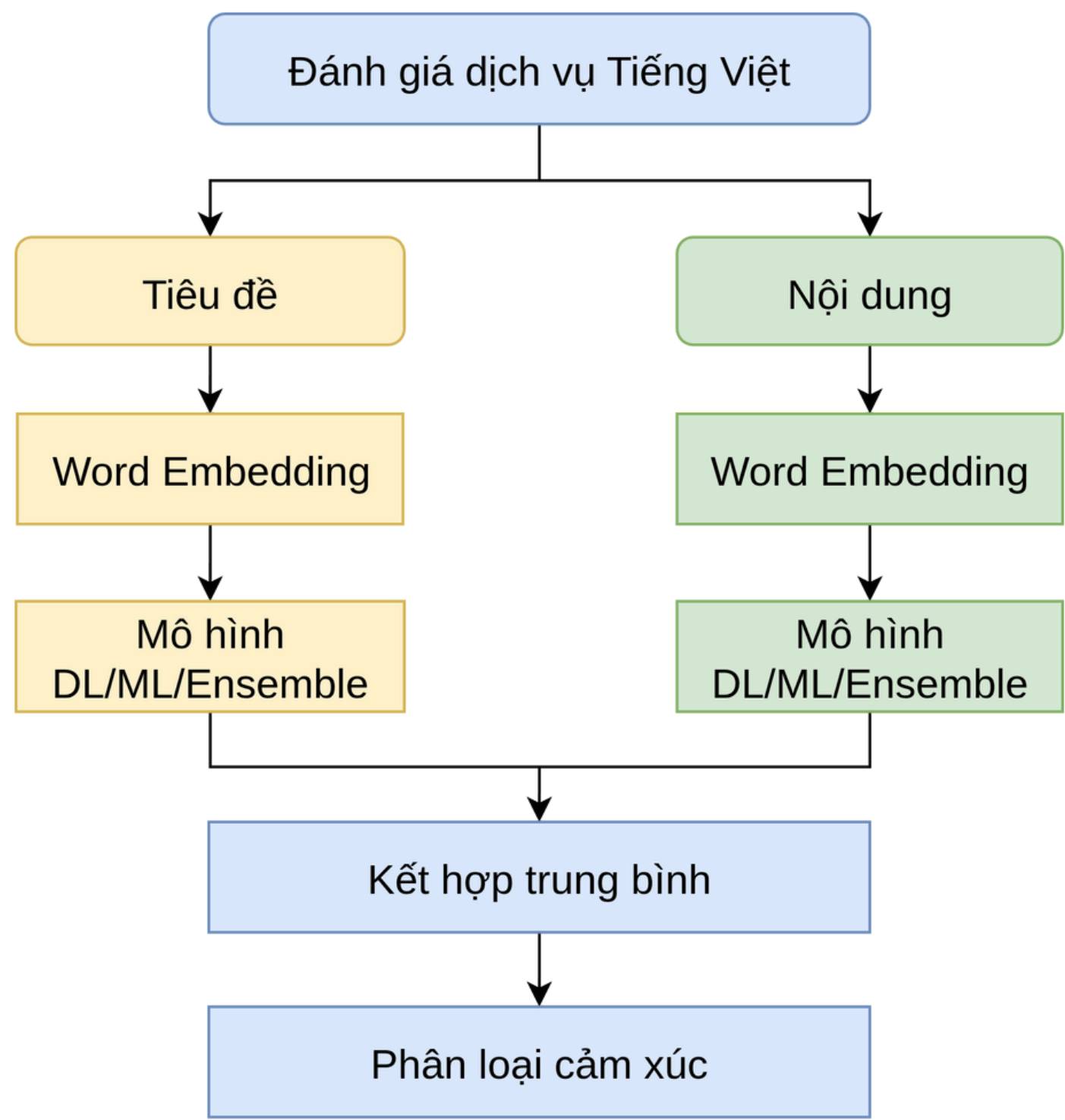


Giải pháp đề xuất

Kết hợp **Tiêu đề** và **Nội dung**

Sử dụng emoji vào phân tích

Phương pháp kết hợp Tiêu đề và Nội dung



STT	Tiêu đề	Nội dung
1	Bộ phận chăm sóc khách hàng kém	Bộ phận sales làm việc quá cứng nhắc, không quan tâm tới quyền lợi của khách hàng
2	Trải nghiệm về kỳ nghỉ dưỡng tại Amiana	Nhân viên nhiệt tình nhưng dịch vụ phòng quá kém so với giá trị nhận được. Khu mới ở quá xa so với khu hồ bơi và biển, mặc dù có xe di chuyển nhưng chờ đợi xe lâu nên ảnh hưởng đến thời gian của chuyến đi

Thực nghiệm

Mô hình

KNN, Decision Tree, Logistic
Regression, CNN, LSTM, BiLSTM,
Ensemble LSTM + CNN, Ensemble
BiLSTM + CNN

Tập dữ liệu



Kết quả thực nghiệm

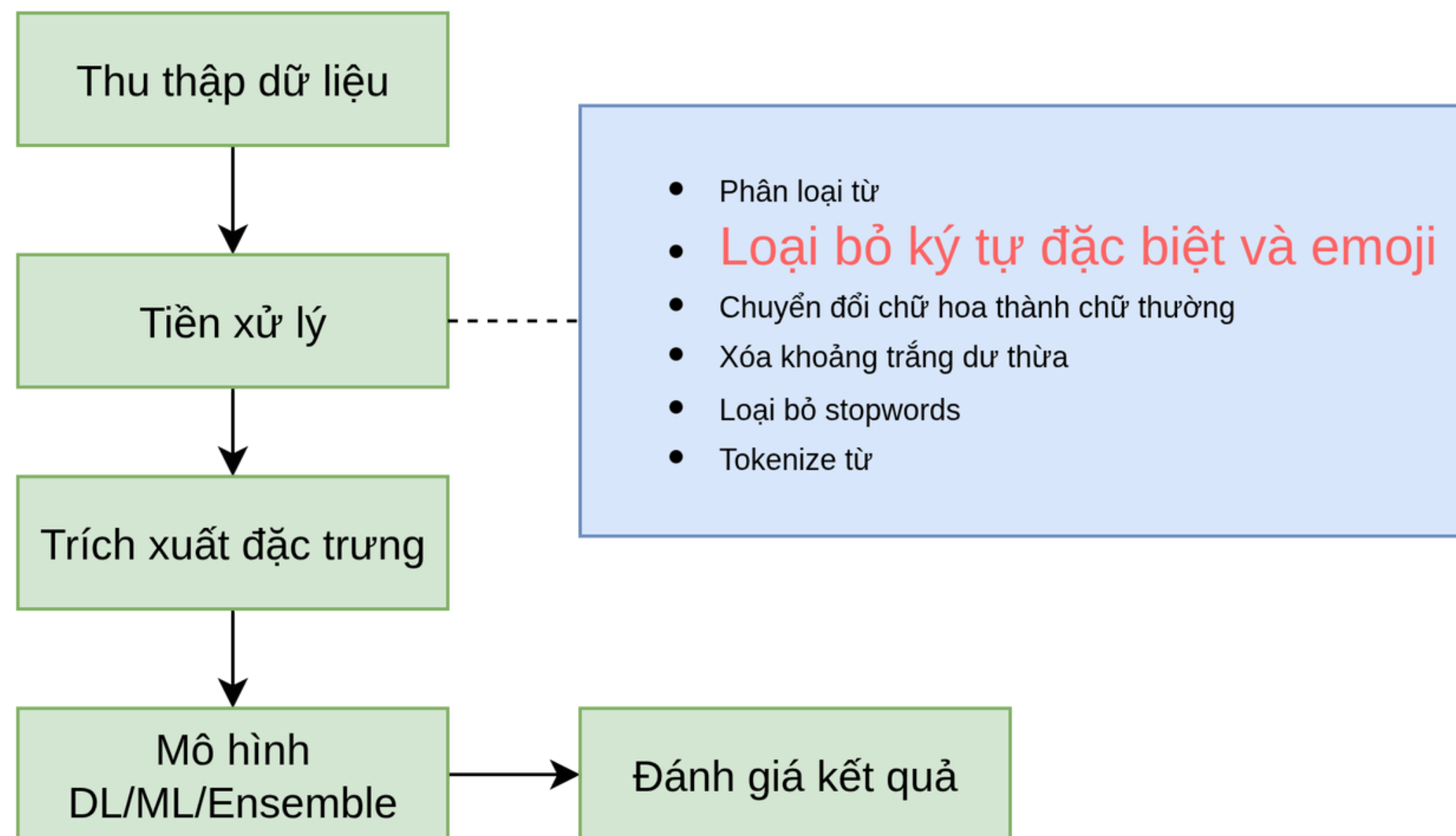
Mô hình		Kết hợp Tiêu đề và Nội dung				Chỉ dùng Nội dung			
		Accu	Prec	Reca	F1-sc	Accu	Prec	Reca	F1-sc
LSTM	Word2vec	87	86	87	87	73	69	72	72
	PhoBERT	87	76	87	81	78	75	76	75
CNN	Word2vec	88	87	88	87	78	75	78	77
	PhoBERT	87	76	87	81	78	75	76	75
BiLSTM	Word2vec	86	85	86	85	74	70	74	71
	PhoBERT	90	88	90	88	78	76	78	77
Sequential LSTM-CNN	Word2vec	86	84	86	85	77	75	77	77
	PhoBERT	89	85	89	87	78	76	78	77
Sequential BiLSTM-CNN	Word2vec	87	86	87	87	77	75	77	76
	PhoBERT	89	85	89	87	78	76	78	77
Logistic Regression	Word2vec	86	76	87	81	73	69	72	72
	PhoBERT	86	78	86	81	78	75	76	75

Kết quả của các mô hình với kết hợp Tiêu đề với Nội dung và chỉ dùng Nội dung

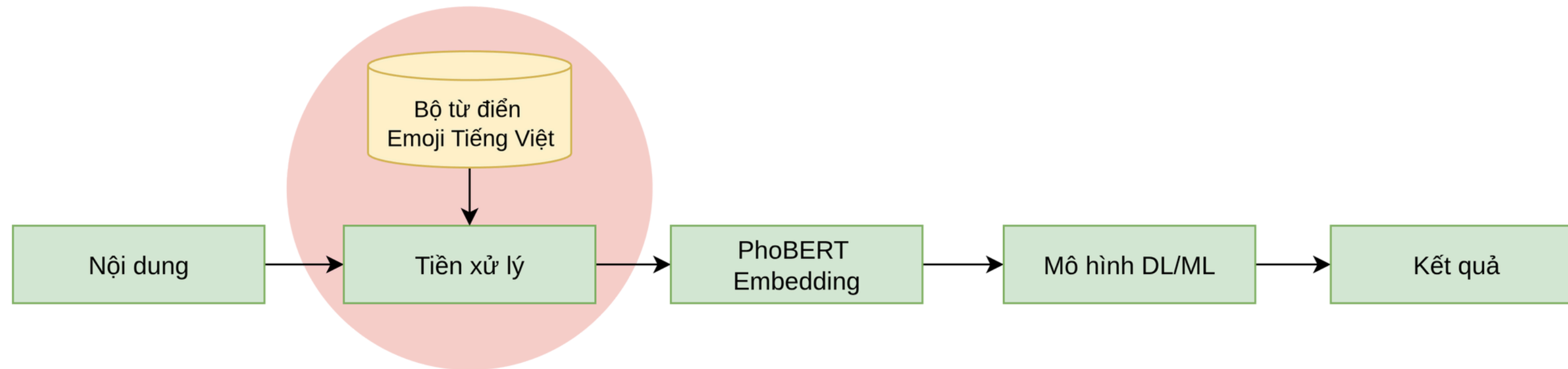
Phương pháp VED_PhoBERT

“Chúng ta cùng 🙌 cho sản phẩm này nào”

⇒ “Chúng ta cùng cho sản phẩm này nào”



















Phương pháp VED_PhoBERT



- Xây dựng từ điển 3.790 emoji tiếng Việt → mô tả dạng văn bản.
- Ví dụ về từ điển:

```
{  
  "Vietnamese_emoji": {  
    "0x1f603": "cười tươi",  
    "0x1f604": "cười rạng rỡ",  
    .....  
  }  
}
```

Code	Browser	Sample	GMail	SB	DCM	KDDI
U+1F603						
U+1F604					—	—
U+1F601						

Thực nghiệm

- **Mô hình ML/DL:** Logistic Regression và XGBoost, BiLSTM, BiLSTM-Attention, GRU, BiGRU, CNN, CNN-Trans-Enc, BiGRU+CNN-Trans-Enc
- **Dataset:** UIT-ViHSD và UIT-VSFC

Kết quả thực nghiệm

Mô hình	Bộ dữ liệu UIT-ViHSD			Bộ dữ liệu UIT-VSFC		
	VED_PhoBERT	ViSoBERT	PhoBERT	VED_PhoBERT	ViSoBERT	PhoBERT
BiLSTM	83	85	84	90	85	88
BiLSTM-Attention	83	83	84	92	83	89
GRU	79	83	82	93	86	86
BiGRU	84	82	84	93	88	89
CNN	86	83	81	91	84	87
Hồi quy Logistic	90	87	86	94	85	88
XGBoost	89	86	86	93	86	88
CNN-Trans-Enc	81	84	83	94	87	89
BiGRU-CNN-Trans-Enc	90	87	86	94	85	89

Kết quả thực nghiệm của VED_PhoBERT, ViSoBERT và PhoBERT trên hai tập dữ liệu chuẩn

Phương pháp E2V-PhoBERT

Tokenization:

→ ["Tôi", "yêu", "Việt_Nam", "❤️"]

Token IDs:

"Tôi" → 134

"yêu" → 3455;

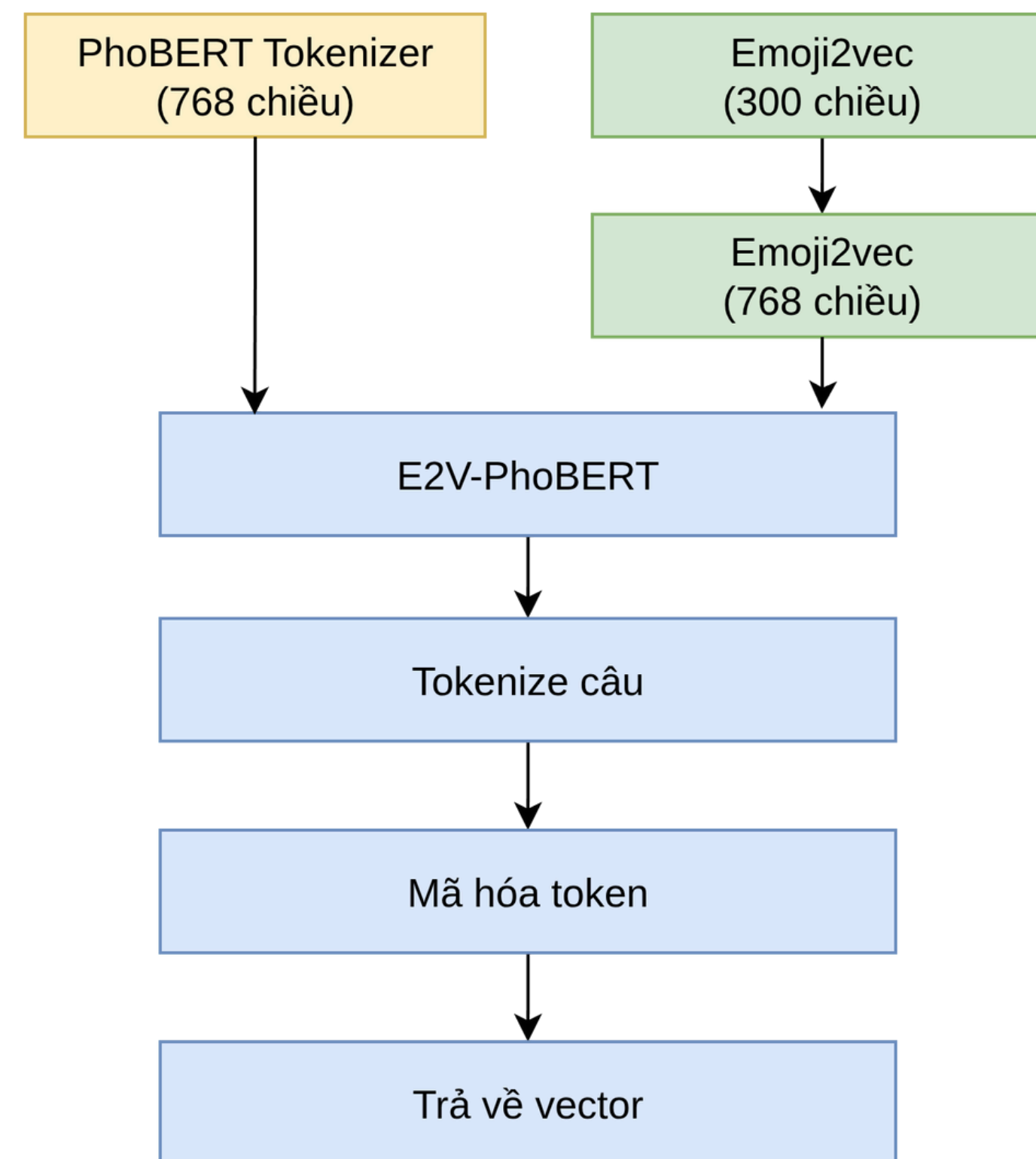
"Việt_Nam" → 44423;

"❤️" → 64321 (ID của emoji trong từ điển mở rộng)

mở rộng)

→ Tokenized: [134, 3455, 44423, 64321]

Embeddings: Kết hợp PhoBERT + Emoji2vec, sau đó lấy vector [CLS] từ 4 lớp cuối cùng cho kết quả đầu ra là [batch_size, 3072]



Thực nghiệm

- **Mô hình ML/DL:** Logistic Regression và XGBoost, BiLSTM, BiLSTM-Attention, GRU, BiGRU, CNN, CNN-Trans-Enc, BiGRU+CNN-Trans-Enc
- **Dataset:** UIT-ViHSD, UIT-VSFC và AIVIVN 2019

Kết quả thực nghiệm

Mô hình	UIT-ViHSD				UIT-VSFC				AIVIVN 2019			
	E2V-PhoBERT		ViSoBERT		E2V-PhoBERT		ViSoBERT		E2V-PhoBERT		ViSoBERT	
	Accu	F1-sc	Accu	F1-sc	Accu	F1-sc	Accu	F1-sc	Accu	F1-sc	Accu	F1-sc
BiLSTM	81	73	79	69	81	79	84	82	89	88	84	83
BiLSTM-Attention	84	77	79	70	85	81	84	82	87	88	84	83
GRU	86	79	83	76	84	82	88	86	85	84	86	86
BiGRU	82	74	76	66	81	78	82	80	91	90	87	85
CNN	85	78	79	70	82	80	82	79	89	87	90	90
CNN-Trans-	82	73	82	74	82	81	80	78	88	87	86	85
BiGRU+CNN-Trans-Enc	85	78	85	78	88	86	81	80	87	86	89	88
Logistic Regression	86	84	86	84	83	82	81	81	86	86	87	86
XGBoost	85	80	86	83	81	79	83	81	83	82	88	88

Kết quả thực nghiệm của E2V-PhoBERT và ViSoBERT trên ba tập dữ liệu chuẩn

Kết luận

Nghiên cứu đề xuất ba phương pháp kết hợp Tiêu đề, Nội dung và Emoji giúp nâng cao hiệu quả phân loại cảm xúc tiếng Việt:

- Mô hình DL với **Tiêu đề và Nội dung**: Trong tất cả các thực nghiệm nghiên cứu, phương pháp sử dụng kết hợp Tiêu đề luôn cho kết quả tốt hơn khi chỉ sử dụng Nội dung
- **VED_PhoBERT** đạt độ chính xác cao hơn ViSoBERT trong tất cả trường hợp thực nghiệm
- **E2V-PhoBERT** đạt độ chính xác cao hơn ViSoBERT trong hầu hết trường hợp thực nghiệm

=> **VED_PhoBERT đạt độ chính xác cao nhất so với các mô hình hiện tại**

Kết quả đề tài

01 sản phẩm phần mềm;

03 bài báo được chấp nhận tại hội nghị quốc tế:

- **Bài 1: Published**

- **Họ tên tác giả:** Dai Tho Dang, Quoc Viet Hoang, Nguyen Xuan Thao Mai, Ngoc Thanh Nguyen
- **Tên bài báo (Hội nghị):** E2V-PhoBERT: A Fine-Tuned PhoBERT Model with Enhanced Accuracy for High-Performance Vietnamese Sentiment Analysis (ACIIDS 2025)
- **Proceeding:** Communications in Computer and Information Science, vol 2493. Springer
- **Indexed:** Scopus, ISI

- **Bài 2: Accepted**

- **Họ tên tác giả:** Nguyen Xuan Thao Mai, Tran Song Nguyen Pham, Cong Phap Huynh, Dai Tho Dang
- **Tên bài báo (Hội nghị):** Sentiment Analysis of Hotel Customer Reviews (ICIIT 2025)
- **Indexed:** Scopus, ISI

- **Bài 3: Accepted**

- **Họ tên tác giả:** Cong Phap Huynh, Quoc Viet Hoang, Nguyen Xuan Thao Mai, Pham Song Nguyen Tran, Dai Tho Dang
- **Tên bài báo:** VED_PhBERT: Enhancing Vietnamese Sentiment Analysis with Emoji Descriptions Integration (CITA 2025)
- **Indexed:** Scopus



BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

CHÚNG EM



XIN CẢM ƠN!

