



Deep Learning For Business

Seminar Presentation •

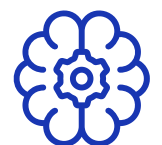
TRANSFORMER & PRE-TRAINED PHOBERT MODEL ANALYSIS OF VIETNAMESE FOOD REVIEWS

Presented by Nguyễn Thị Khánh Hà - 18520692

Instructors: Professor Đỗ Phúc & Ph.D. Trung Phan

TABLE OF CONTENTS

01	Problem
02	Transformers & PhoBERT Introduction
03	Dataset
04	Experiments
05	Challenges
06	References



1 PROBLEM

3

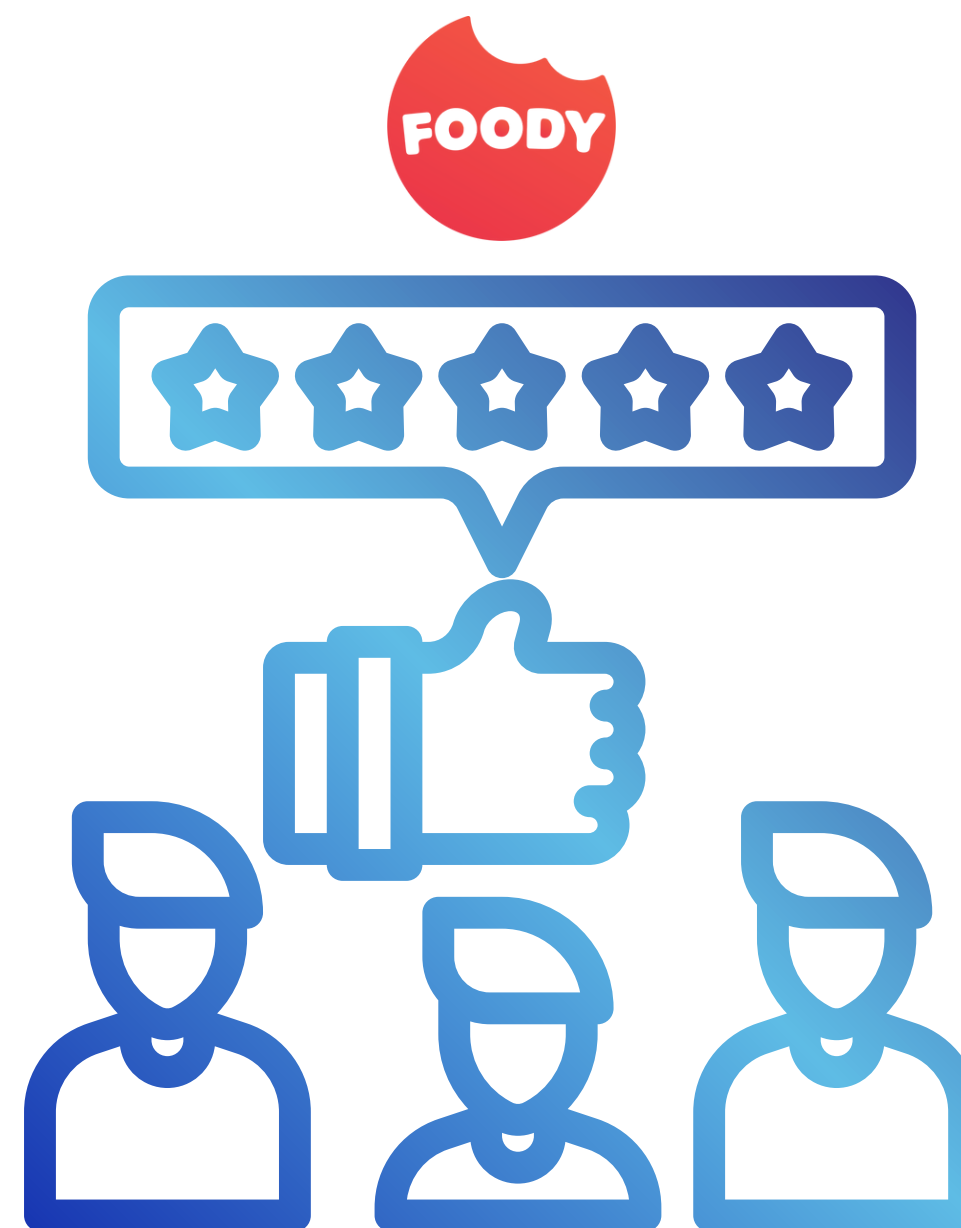
Problem: Analyzing the sentiment of Vietnamese Food Reviews on Foody.

Input

- Vietnamese food review dataset crawled from Foody
- Pre-trained PhoBERT model

Output

Sentiment analysis results of the Vietnamese food reviews, including whether the review is **Positive**, **Negative**, or **Neutral**.

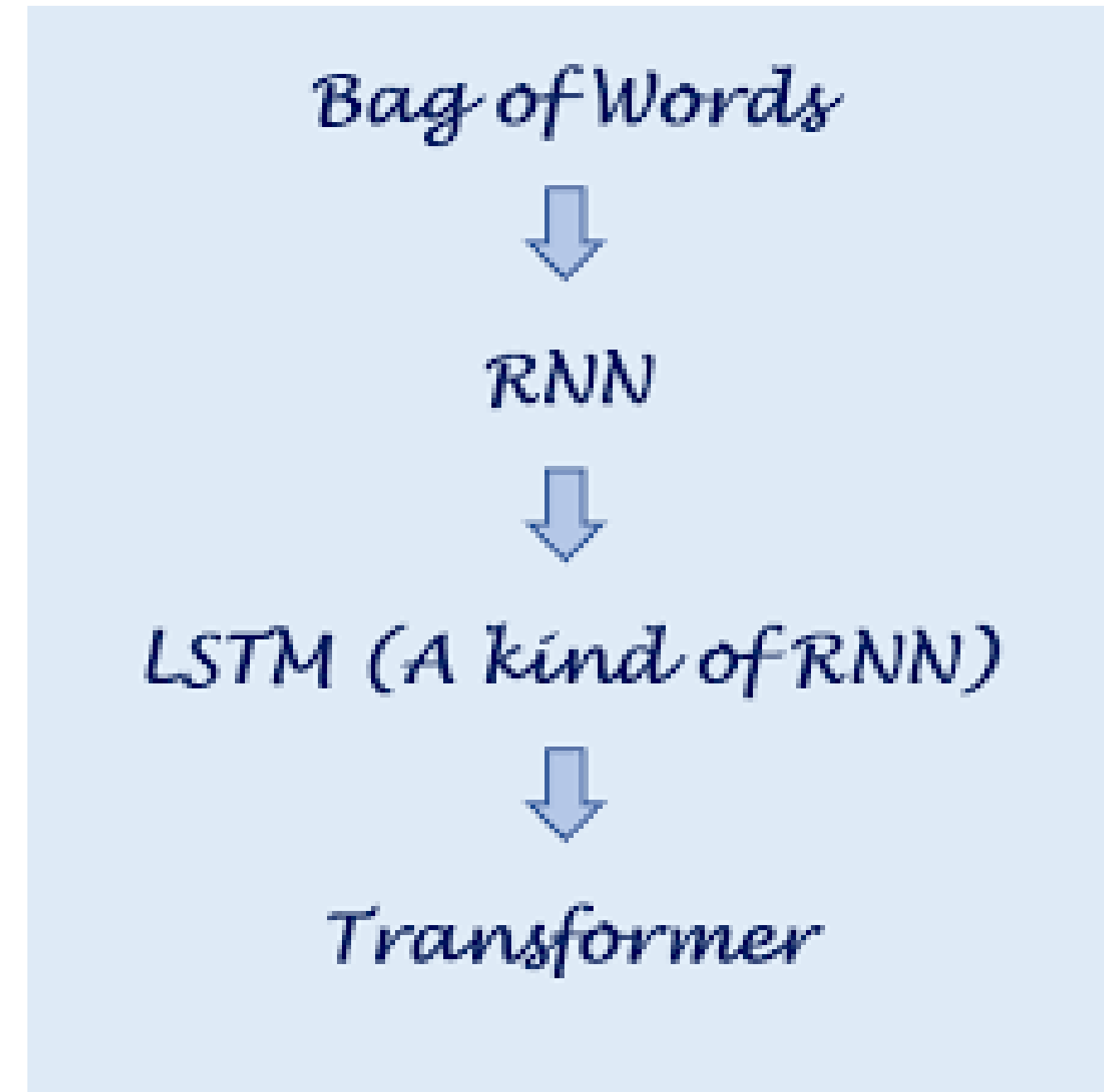


2 TRANSFORMERS & PHOBERT INTRODUCTION

4

WHY TRANSFORMER?

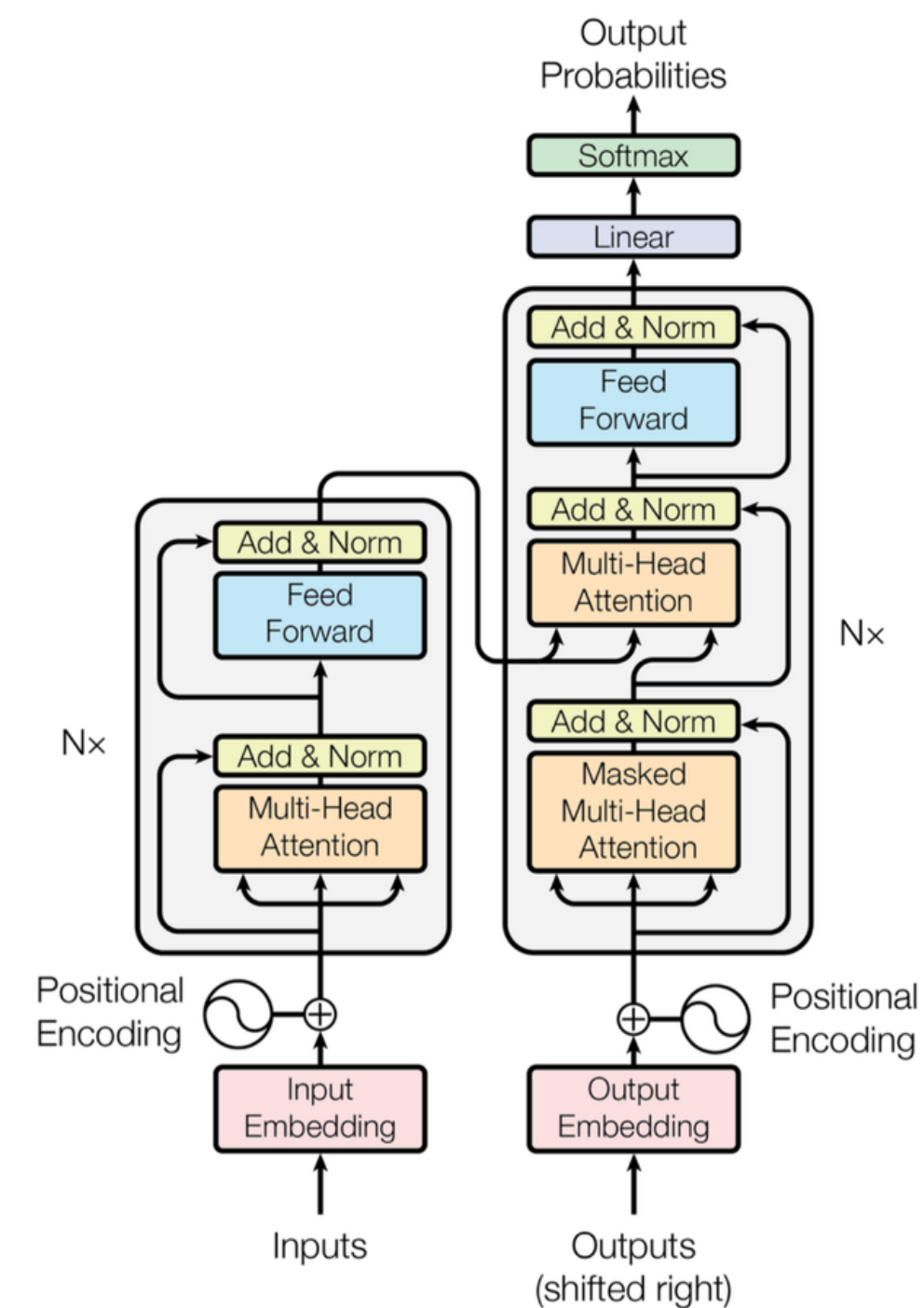
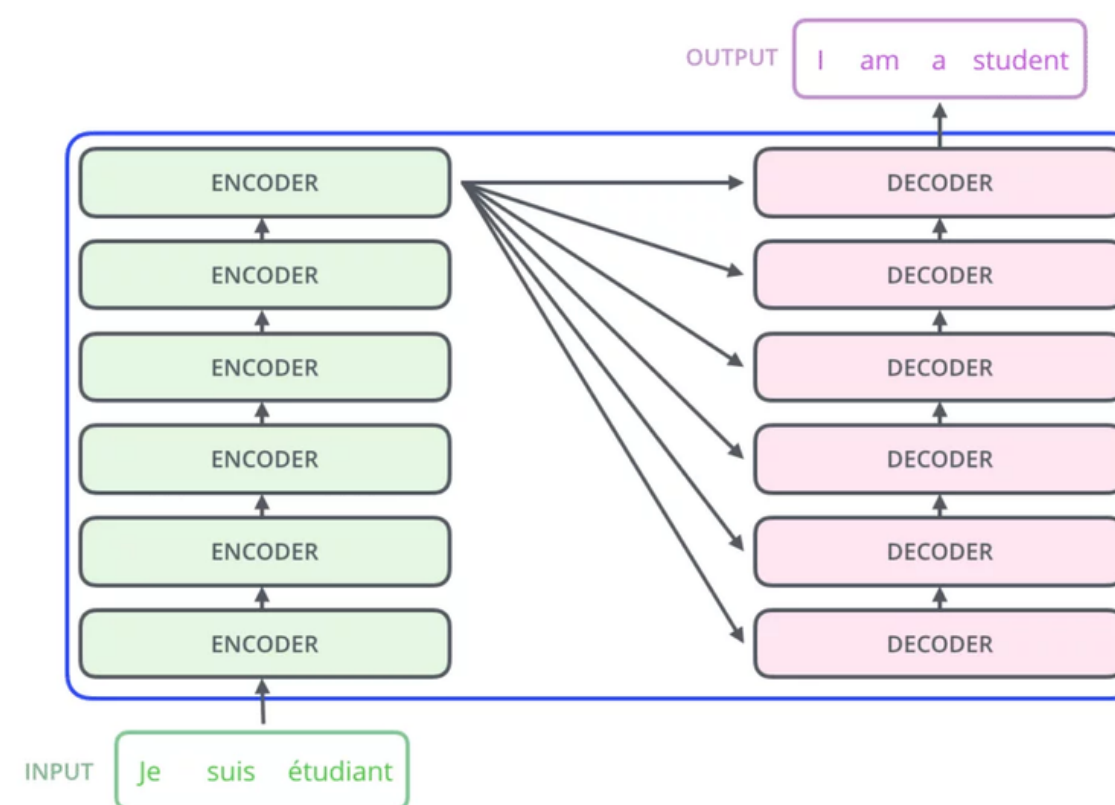
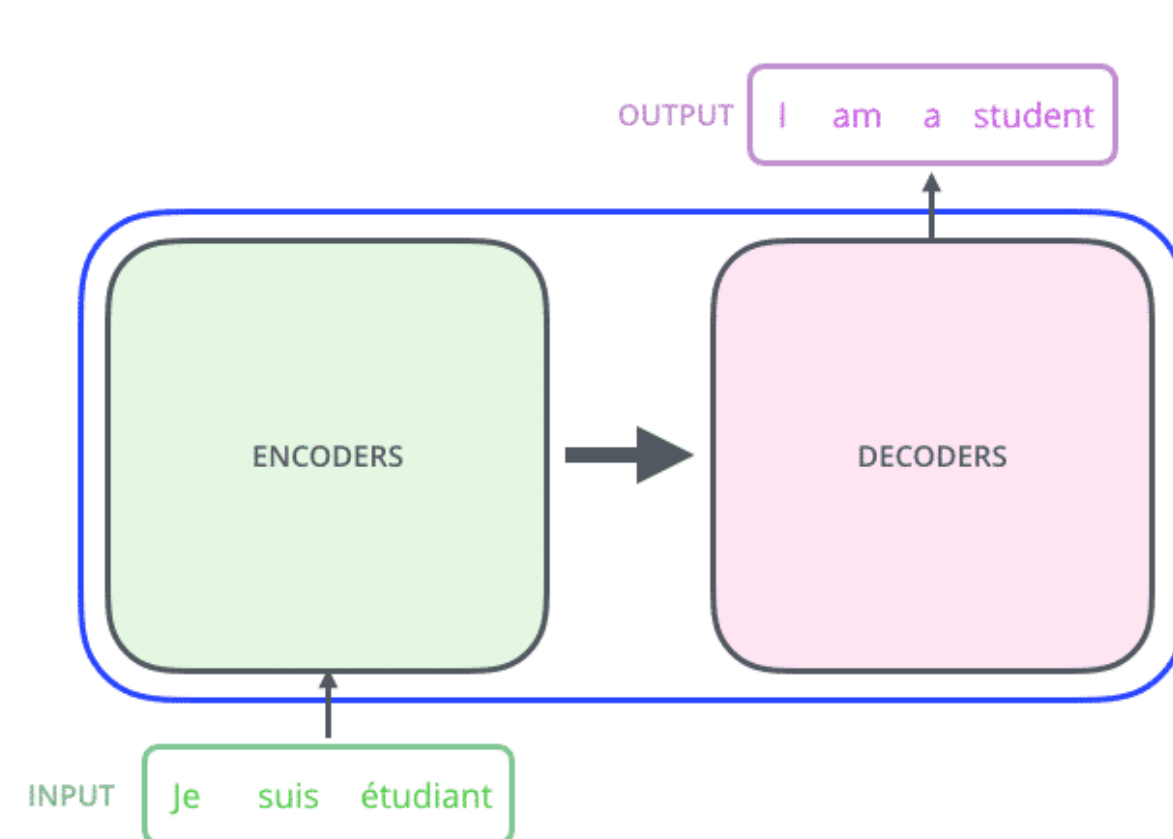
- Improving the limitations of RNNs and LSTM models.
 - Performance (CPU / GPU)
 - Vanishing Gradient & data loss issues of long sequences.



2 TRANSFORMERS ARCHITECTURE

5

Transformer Architecture





- BERT uses a Transformer-based architecture
- Variations: RoBERTa, ALBERT, DistilBERT, PhoBert
- Two available versions of PhoBERT: PhoBERT-base (150M parameters) & PhoBERT-large (350M parameters)
- Supported tools: RDRSegmenter, VnCoreNLP

3 DATA

7

comments_20_29_ho-chi-minh.csv

comments_1140_1149

comments_1160_1169

comments_1170_1179

comments_1200_1209

comments_1210_1219

comments_1220_1229

comments_1230_1239

comments_1240_1249

comments_1250_1259

comments_1260_1269

comments_1270_1279

comments_1280_1289

comments_1290_1299

comments_1300_1309

foodshops.csv

foodshops_an-giang.csv

foodshops_bac-giang.csv

foodshops_bac-kan.csv

foodshops_bac-lieu.csv

foodshops_bac-ninh.csv

foodshops_ben-tre.csv

File Origin

65001: Unicode (UTF-8)

Delimiter

Comma

Data Type Detection

Based on first 200 rows

commentor	content	rating	shop_name	
Cream Cheese	Lẩu 198k được tới 4 con bào ngư 4 con sò điệp kèm cái...	10	Lẩu Bào Ngư Thượng Hạng 198k - Phan Phú Tiên	https://www.foody.v
Anna Jullie	Lẩu chất lượng ok , nước dùng gà ngon , bào ngư tươi l...	8.8	Lẩu Bào Ngư Thượng Hạng 198k - Phan Phú Tiên	https://www.foody.v
Nguyễn Thanh	Ngon lắm giá lại rẻ , sẽ ủng hộ nữa	9.4	Lẩu Bào Ngư Thượng Hạng 198k - Phan Phú Tiên	https://www.foody.v
Mèo Ham Ăn	Ngon	6	Lẩu Bào Ngư Thượng Hạng 198k - Phan Phú Tiên	https://www.foody.v
Nhi Cẩm	Mỗi lần thèm là cứ order ở đây thôi , plemmm:)))	8	Lẩu Bào Ngư Thượng Hạng 198k - Phan Phú Tiên	https://www.foody.v
Canal1985	Không ngọt lắm. Ăn không bao giờ là ngán. Ngon.	9	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Linh My	Xôi ở đây ăn hoài vẫn thấy ngon , ghiền rau câu trứng c...	8.8	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Bảo Ngọc	Không ngon như mình nghĩ, xôi không vị gì hết, rau câu...	1	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	👍👍👍👍👍👍	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	Ngon lắm😊😊😊😊	8.8	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Lương Quang Dũng	Đặt mua 5 gói xôi thêm 3 miếng rau câu dừa + 3 miếng...	2.6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Nhi Cẩm	Xôi ngon vị cadé đặc trưng của người hoa , thích nhất r...	9	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Mèo Ham Ăn	Ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Mèo Ham Ăn	Ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	Xôi ngon , hương vị năm xưa !	10	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	Ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	ngon	6	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v
Anna Truong	Xôi ngon , hương vị năm xưa !	10	Xôi Cadé Trần Phú Chính Gốc - 咖椰糯米饭	https://www.foody.v

- The dataset consists of customer reviews about the food and services provided by restaurants in various cities and towns across Vietnam.
- The dataset is crawled from **Foody**, an online platform that offers reviews and information on restaurants, bars, cafes, and more.



4 EXPERIMENTS

8

```
✓ [5] # Just like PhoBERT: INPUT TEXT MUST BE ALREADY WORD-SEGMENTED!
0s import math
    sentence = 'Ngon lắm giá lại rẻ , sẽ ủng hộ nữa.'

    input_ids = torch.tensor([tokenizer.encode(sentence)])

    with torch.no_grad():
        out = model(input_ids)
        out = out.logits.softmax(dim=-1).tolist()
        # print(out)
        print(f'NEG: {round(out[0][0], 2) * 100}%, POS: {round(out[0][1], 2) * 100}%, NEU: {round(out[0][2], 2) * 100}%')
        # Output:

    NEG: 0.0%, POS: 99.0%, NEU: 1.0%
```

Result

5 CHALLENGES

9



- Understanding the inner principles of Transformer and BERT
- Collecting data, cleaning and pre-processing, and integrating it into the model.
- Dealing with limited computing resources for large models and datasets.



References

[1] Hugging Face. (n.d.). *wonrax/phobert-base-vietnamese-sentiment*. Retrieved from <https://huggingface.co/wonrax/phobert-base-vietnamese-sentiment>

[2] Hoang, T. H. (2020, May 17). *PhoBERT - Vietnamese Pre-trained Language Models [Video]*. YouTube. Retrieved from <https://www.youtube.com/watch?v=DVouDUMNFDY>

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is All You Need*. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).

[4] Pang, B., & Lee, L. (2005). *Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales*. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 115-124).

[5] Nguyen, T. L., & Nguyen, T. T. (2020). *PhoBERT: Pre-trained Language Models for Vietnamese*. arXiv preprint arXiv:2003.00744.

THANK YOU!