

# PHÂN LOẠI CÂU NHẬN XÉT TRÊN SHOPEE

Phạm Huỳnh Tấn Đạt - 230201004

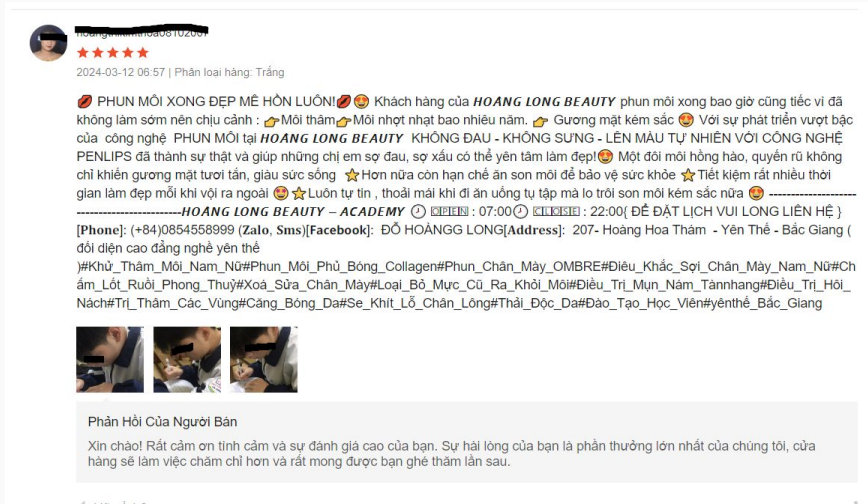
# Tóm tắt

- Link Github:  
<https://github.com/mynameuit/CS2205.APR2023/>
- Link YouTube video:  
<https://youtu.be/-33lfng3yW0>
- Tên: Phạm Huỳnh Tấn Đạt




# Giới thiệu


- Các cửa hàng trên shopee sử dụng các comment về sản phẩm của họ để nâng cao chất lượng cũng như dịch vụ.
- Một lượng lớn các câu comment “spam” làm tốn thời gian để lọc.

 **hoanglongbeauty10206**  
★★★★★  
2024-03-12 06:57 | Phản loại hàng: Trắng


🔥 PHUN MÔI XONG ĐẸP MÊ HỒN LUỒN! 🤩🤩 Khách hàng của **HOANG LONG BEAUTY** phun môi xong bao giờ cũng tiếc vì đã không làm sớm nên chịu cảnh : 🤔Môi thâm🤔Môi nhợt nhạt bao nhiêu năm, 🤔Gương mặt kém sắc 🤔 Với sự phát triển vượt bậc của công nghệ PHUN MÔI tại **HOANG LONG BEAUTY** KHÔNG ĐAU - KHÔNG SƯNG - LÊN MÀU TỰ NHIÊN VỚI CÔNG NGHỆ PENLIPS đã thành sự thật và giúp những chị em sợ đau, sợ xấu có thể yên tâm làm đẹp!🤩 Một đôi môi hồng hào, quyến rũ không chỉ khiến gương mặt tươi tắn, giàu sức sống 🌟 Hơn nữa còn hạn chế ăn son môi để bảo vệ sức khỏe 🌟 Tiết kiệm rất nhiều thời gian làm đẹp mỗi khi vội ra ngoài 🌟 Luôn tự tin , thoải mái khi đi ăn uống tụ tập mà lo trời son môi kém sắc nữa 🤩-----  
**HOANG LONG BEAUTY – ACADEMY** 📞 **OPEN** : 07:00🕒 **CLOSE** : 22:00( **ĐỂ ĐẶT LỊCH VUI LONG LIÊN HỆ** )  
[Phone]: (+84)0854558999 (Zalo, Sms)[Facebook]: ĐỒ HOÀNG LONG[Address]: 207- Hoàng Hoa Thám - Yên Thế - Bắc Giang ( đối diện cao đẳng nghệ yên thế )  
)#Khử\_Thâm\_Môi\_Nam\_Nữ#Phun\_Môi\_Phủ\_Bóng\_Collagen#Phun\_Chân\_Mày\_OMBRE#Điều\_Khắc\_Sợi\_Chân\_Mày\_Nam\_Nữ#Chăm\_Ấm\_Lối\_Ruột\_Phong\_Thủy#Xoa\_Sữa\_Chân\_Mày#Loại\_Bỏ\_Mực\_Cũ\_Ra\_Khỏi\_Môi#Điều\_Tri\_Mụn\_Nam\_Tàn nhang#Điều\_Tri\_Hội\_Nách#Tri\_Thâm\_Các\_Vùng#Căng\_Bông\_Dạ#Se\_Khít\_Lỗ\_Chân\_Lông#Thải\_Độc\_Dạ#Đào\_Tạo\_Học\_Viên#yênthế\_Bắc\_Giang




**Phản Hồi Của Người Bán**  
Xin chào! Rất cảm ơn tình cảm và sự đánh giá cao của bạn. Sự hài lòng của bạn là phần thưởng lớn nhất của chúng tôi, cửa hàng sẽ làm việc chăm chỉ hơn và rất mong được bạn ghé thăm lần sau.

 **traip439**  
★★★★★  
2024-05-05 15:13 | Phản loại hàng: Kim loại đen + Silicone

**Phản Hồi Của Người Bán**  
Chúng tôi đã nhận được đánh giá tốt của bạn, chúng tôi rất vinh dự khi nhận được sự khẳng định của bạn và rất mong được phục vụ bạn một lần nữa.

 **dungtmth792**  
★★★★★  
2024-05-04 12:10 | Phản loại hàng: Silicone đen

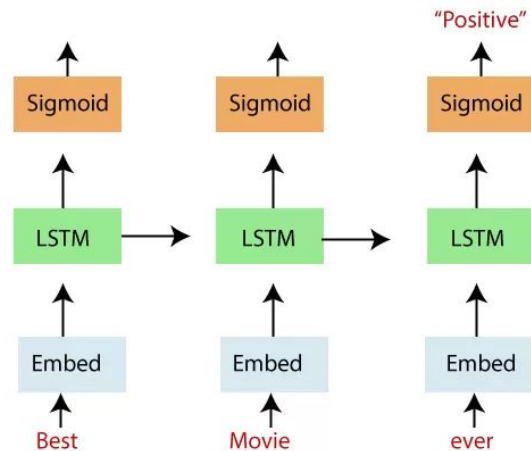
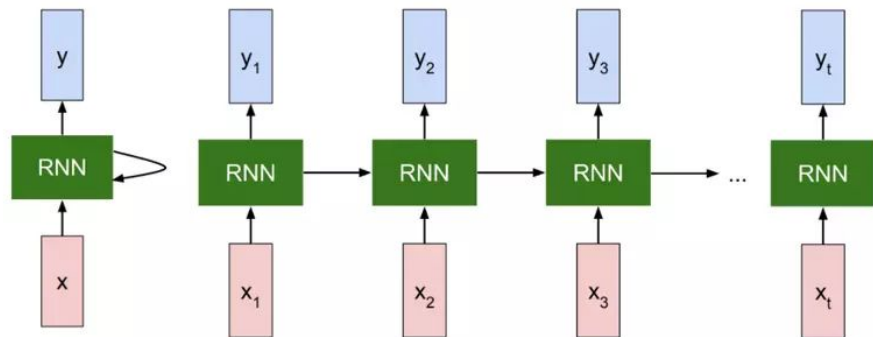
**Phản Hồi Của Người Bán**  
Chúng tôi đã nhận được đánh giá tốt của bạn, chúng tôi rất vinh dự khi nhận được sự khẳng định của bạn và rất mong được phục vụ bạn một lần nữa.

 **rubivoloc**  
★★★★★  
2024-04-29 21:02 | Phản loại hàng: Silicone đen

**Phản Hồi Của Người Bán**  
Quý khách thân mến, rất vui khi quý khách yêu thích sản phẩm của chúng tôi, đánh giá tốt của bạn là động lực lớn của shop. Chúc quý khách một ngày vui vẻ ạ.

# Giới thiệu

- Các thuật toán máy học và học sâu truyền thống không biểu diễn tốt về mặt ngữ nghĩa, gây tốn tài nguyên và thời gian.

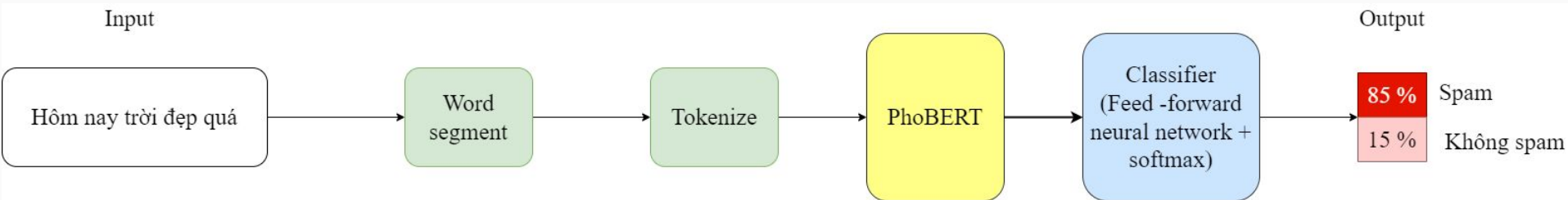


# Mục tiêu

- Áp dụng mô hình PhoBERT để xây dựng mô hình phân loại câu nhằm xây dựng mô hình phân loại với ứng dụng :
  - Hỗ trợ lọc và chặn câu comment spam.
  - Các tác vụ liên quan bài toán Sentiment Analysis
- Mô hình xử lý tốt ngôn ngữ tiếng Việt.
- Phân loại tốt với các câu comment về các sản phẩm điện tử trên shopee.

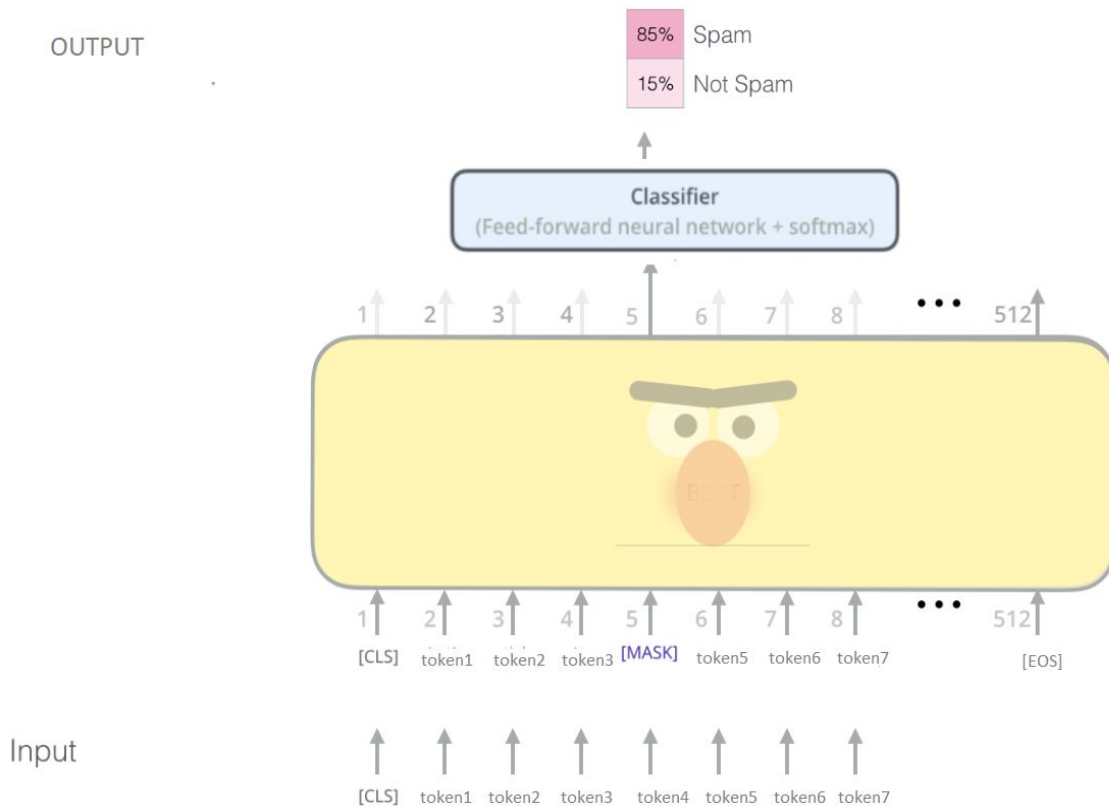
# Nội dung và Phương pháp

- Các câu comment sẽ được tiền xử lý: word segment và tokenize.
- Sau đó được đưa vào mô hình PhoBERT để tính toán và cuối cùng thông qua FFNN và softmax để ra xác suất từng lớp.



# Nội dung và Phương pháp

- Fine-tuning 2 phiên bản PhoBERT là base và large với bộ dữ liệu 10000 câu comment về các sản phẩm điện tử thu từ shopee.



# Nội dung và Phương pháp

- Bộ dữ liệu: 10000 câu comment về các sản phẩm điện tử thu được từ các cửa hàng trên shopee.
- Nhãn 0 cho câu không spam và nhãn 1 cho câu spam.

Input	Label
Hôm nay trời đẹp	1
Sản phẩm dùng tốt, 5 sao cho shop	0

- Độ đo Accuracy:

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{All predictions}}$$



# Kết quả dự kiến

- Thu được 2 mô hình phân loại câu comment ứng với 2 phiên bản PhoBERT là base và large.
- 2 mô hình có kết quả độ đo accuracy  $\geq 80\%$ .
- Một bộ dữ liệu 10000 câu comment về các sản phẩm điện tử từ shopee.
- Xây dựng ứng dụng và api minh họa.

# Tài liệu tham khảo

- [1]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR, abs/1810.04805.
- [2]. Dat Quoc Nguyen, & Anh Tuan Nguyen (2020). PhoBERT: Pre-trained language models for Vietnamese. CoRR, abs/2003.00744.