



ĐẠI HỌC QUỐC GIA TP. HCM  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Ngày nhận hồ sơ

(Do CQ quản lý ghi)

## THUYẾT MINH ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ CẤP SINH VIÊN 2022

### A. THÔNG TIN CHUNG

#### A1. Tên đề tài

- Tên tiếng Việt:

NHẬN DIỆN CÁC BÀI ĐÁNH GIÁ RÁC TRÊN CÁC TRANG THƯƠNG MẠI  
ĐIỆN TỬ TIẾNG VIỆT

- Tên tiếng Anh:

DETECTING SPAM REVIEWS ON VIETNAMESE E-COMMERCE WEBSITES

#### A2. Thời gian thực hiện

**06** tháng (kể từ khi được duyệt).

#### A3. Tổng kinh phí

Tổng kinh phí: **6** triệu đồng, gồm

- Kinh phí từ Trường Đại học Công nghệ Thông tin: **6** triệu đồng

#### A4. Chủ nhiệm

Họ và tên: **Đinh Văn Cơ**

Ngày, tháng, năm sinh: 15/04/2001

Giới tính (Nam/Nữ): Nam

Số CCCD: 08720103731; Ngày cấp: 14/02/2022; Nơi cấp: Cục cảnh sát quản lý  
hành chính về trật tự xã hội

Mã số sinh viên: 19521293

Số điện thoại liên lạc: 079.593.4432

Đơn vị (Khoa): Khoa Khoa học và Kỹ thuật Thông tin

Số tài khoản: 1015642141

Ngân hàng: Vietcombank

## A5. Thành viên đề tài

TT	Họ tên	MSSV	Khoa
1	Đinh Văn Cơ	1951293	KH&KTTT

## B. MÔ TẢ NGHIÊN CỨU

### B1. Giới thiệu về đề tài

#### B1.1. Lý do chọn đề tài

Trong những năm gần đây, thương mại điện tử ở Việt Nam đã có bước phát triển nhanh chóng. Thị trường thương mại điện tử ngày càng được mở rộng và hiện đại, thúc đẩy các doanh nghiệp trong việc sản xuất và kinh doanh, mọi người ngày càng biết đến và sử dụng các dịch vụ của thương mại điện tử như một hình thức mua sắm chủ yếu. Hiện nay, ở Việt Nam có nhiều sàn thương mại điện tử, trên mỗi nền tảng thì có những mặt hàng với nguồn gốc, xuất xứ, và chất lượng khác nhau. Người mua hàng thường sẽ cân nhắc để lựa chọn những sản phẩm phù hợp và chất lượng. Bên cạnh việc lựa chọn một số cửa hàng uy tín, thì một trong những yếu tố quan trọng để người dùng lựa chọn sản phẩm chính là đọc những bài đánh giá của người mua sản phẩm này trước đó. Trước đây, khi mới bắt đầu phát triển thì số lượng người dùng vẫn còn ít nên việc kiểm soát các bài đánh giá này cũng dễ dàng, có thể thực hiện theo cách thủ công. Tuy nhiên, số lượng người bán cũng như người dùng ngày càng nhiều thì việc kiểm soát được những bài đánh giá đó có đánh giá đúng hay không, hay chỉ là những bài đánh giá rác thì đối với các nền tảng thương mại điện tử này cũng là một điều hết sức khó khăn. Chính vì thế, việc lọc ra các bài đánh giá rác một cách tự động để người dùng có một cái nhìn đúng đắn về chất lượng sản phẩm họ sẽ mua là một điều rất cần thiết. Đó cũng chính là lý do mà bài toán nhận diện các bài đánh giá rác trên các trang thương mại điện tử ra đời. Đối tượng mà chúng tôi hướng đến trong bài toán này chính là các bài viết về các sản phẩm được người dùng đánh giá bằng tiếng Việt trên các trang thương mại điện tử hàng đầu ở Việt Nam cụ thể là Shopee và Tiki. Chúng tôi sẽ thu thập dữ liệu trên hai nền tảng này và đưa ra một quy trình gán nhãn nghiêm ngặt cho các bài đánh giá để có được một bộ dữ liệu tốt.

Sau khi thực hiện thu thập và gán nhãn cho bộ dữ liệu, để có thể nhận diện được bài đánh giá rác chúng tôi sẽ sử dụng các phương pháp học sâu với các mô hình như Text-CNN, LSTM và GRU. Cùng với đó, chúng tôi sẽ thử nghiệm các mô hình học chuyển tiếp như Pho-BERT và BERT4New – đây là các mô hình đã cho kết quả cải thiện tốt trên những bài toán khác so với các mô hình học sâu trước đó. Với việc sử dụng các mô hình học chuyển

tiếp, chúng tôi hi vọng rằng bài toán sẽ cho khả năng nhận diện tốt hơn so với các mô hình học sâu truyền thống.

Mục tiêu đặt ra của bài toán này là xây dựng bộ dữ liệu các bài đánh giá của người dùng trên các trang thương mại điện tử và thử nghiệm các mô hình trong việc nhận diện các bài đánh giá rác. Từ đó đánh giá hiệu suất của các mô hình học sâu và chuyển tiếp trong việc phân loại văn bản nói chung và các bài đánh giá rác nói riêng.

### **B1.2. Các công trình liên quan**

Năm 2007, Jindan và Liu đã nghiên cứu về các bài đánh giá spam và chỉ ra các xây dựng mô hình phân loại để phát hiện xem liệu bài đánh giá của người dùng có phải là rác hay không [1]. Đến năm 2008, tác giả cũng đã chỉ ra những thách thức trong việc phát hiện bài đánh giá rác đến từ người dùng và đề xuất ba khía cạnh khác nhau của các bài đánh giá rác đó [2].

Việc phát hiện bài đánh giá rác có nhiều hướng tiếp cận, bên cạnh các phương pháp truyền thống như sử dụng quy luật hay và mô hình học máy thông thường, thì các mô hình học sâu cũng được ứng dụng trong việc phát hiện bài đánh giá rác. Năm 2017, Huayi Li và các cộng sự đã đề xuất phương pháp phát hiện các bài đánh giá rác với phương pháp sử dụng mẫu và quy luật [9]. Trong bài báo, tác giả đã chỉ ra rằng việc sử dụng mô hình Markov ẩn đã cho kết quả tốt hơn so với các baseline tại thời điểm đó. Ngoài ra, tác giả còn đề xuất một kiến trúc được gọi là mạng co-bursting dựa trên quan hệ co-bursting giúp phát hiện các nhóm người gửi bài đánh giá rác hiệu quả hơn các phương pháp hiện có. Năm 2018, Žabljak và Montenegro đã ứng dụng các phương pháp học máy trong việc phát hiện các bài đánh giá spam [14] với bộ dữ liệu được sử dụng bao gồm 5.18 triệu bài đánh giá từ 2.14 triệu người đánh giá trên nền tảng Amazon. Năm 2016, Yafeng Ren và Yue Zhang đã sử dụng các phương pháp học sâu như CNN, RNN, Bi-GRNN để phát hiện những ý kiến lừa đảo rác (deceptive opinion spam) [15].

Trên tiếng Việt cũng đã có một số bộ dữ liệu về bài đánh giá của người dùng từ các trang web bán hàng online [12, 13, 16] trong việc phân tích cảm xúc của người dùng nhưng chưa có bộ dữ liệu cho việc phát hiện các bài đánh giá rác. Vì thế, mục tiêu của chúng tôi là xây dựng bộ dữ liệu cho nhiệm vụ phát hiện các bài đánh giá rác của người dùng trên các trang thương mại điện tử.

### **B1.3. Thách thức của đề tài**

Thách thức lớn nhất của chúng tôi trong việc thực hiện đề tài là chưa có bộ dữ liệu cho tiếng Việt trên tác vụ nhận diện bài đánh giá của người dùng là bài đánh giá rác. Sau khi

nghiên cứu một số công trình liên quan, chúng tôi nhận thấy rằng có một số tác giả đã nghiên cứu về bài toán này nhưng đều thực hiện trên tiếng Anh hay một số ngôn ngữ khác và hiện tại bộ dữ liệu tiếng Việt vẫn chưa có. Điều này thúc đẩy chúng tôi cần phải xây dựng một bộ dữ liệu mà phải đáp ứng yêu cầu cho bài toán cũng như vấn đề chất lượng của bộ dữ liệu.

## **B2. Mục tiêu, nội dung, kế hoạch nghiên cứu**

### **B2.1 Mục tiêu**

Mục tiêu trong nghiên cứu của chúng tôi là xây dựng một bộ dữ liệu đủ tốt để đáp ứng yêu cầu cho bài toán. Sau khi xây dựng bộ dữ liệu, chúng tôi sẽ thử nghiệm bộ dữ liệu đã thu thập với các mô hình học sâu và học chuyển tiếp. Cuối cùng là nhận xét, so sánh và đánh giá kết quả đạt được giữa các mô hình. Chỉ ra kết quả tốt nhất mà mô hình đạt được cũng như những hạn chế trong kết quả, từ đó có thể tìm ra phương hướng cải thiện hiệu suất các mô hình sau này.

Do hiện tại vẫn chưa có bộ dữ liệu nào cho tiếng Việt về chủ đề nhận diện bài đánh giá rác, do đó việc xây dựng bộ dữ liệu là rất cần thiết, và chúng tôi mong muốn qua đề tài này sẽ có bộ dữ liệu có thể public cho cộng đồng và đóng góp kết quả đạt được của đề tài. Điều này sẽ góp phần thúc đẩy nghiên cứu và cho sự phát triển hơn nữa trong lĩnh vực Xử lý ngôn ngữ tự nhiên ở Việt Nam.

### **B2.2 Nội dung và phương pháp nghiên cứu**

**Nội dung 1:** Xây dựng bộ dữ liệu cho bài toán nhận diện các bài đánh giá rác trên các trang thương mại điện tử

Như đã đề cập trong phần thách thức của đề tài trước đó, do chưa có bộ dữ liệu trên tiếng Việt nên bước đầu chúng tôi cần phải xây dựng bộ dữ liệu. Sau khi khảo sát trên một số trang thương mại điện tử hàng đầu, chúng tôi đã quyết định chọn hai nền tảng là Shopee và Tiki để tiến hành thu thập dữ liệu. Để có thể có một bộ dữ liệu tổng quát, chúng tôi dự định sẽ thu thập một số bài đánh giá của người dùng trên tất cả các danh mục sản phẩm. Sau khi thu thập bộ dữ liệu, bước tiếp theo mà chúng tôi cần phải làm đó chính là xây dựng một hướng dẫn gán nhãn và thuê một số người gán nhãn, đánh giá kết quả gán nhãn, cuối cùng có được một bộ dữ liệu có thể đáp ứng yêu cầu cho bài toán.

**Nội dung 2:** Nghiên cứu các phương pháp phân loại cho bài toán nhận diện các bài đánh giá rác trên các trang thương mại điện tử

Trong bài toán của chúng tôi, nhiệm vụ mà chúng tôi thực hiện dựa trên tác vụ phân tích ý kiến (opinion analysis) của người dùng từ các câu bình luận và số sao đánh giá. Chúng tôi sẽ huấn luyện mô hình phân loại để phát hiện các bài đánh giá rác dựa trên các nhãn dữ liệu đã được gán trước đó. Trên thực tế, có nhiều hướng tiếp cận để giải quyết bài toán nhưng trong sự phát triển của deep learning trong những năm gần đây đã mang lại hiệu suất vượt trong nhiều bài toán. Do đó, hướng tiếp cận chính của chúng tôi trong bài toán này là ứng dụng các mô hình deep learning cho việc phân loại các bài đánh giá rác.

Các phương pháp tiếp cận cho bài toán của chúng tôi bao gồm:

- Các mô hình mạng nơ-ron học sâu:

**Text-CNN [8]:** Mạng nơ-ron tích chập (CNN) là mô hình được kết hợp từ nhiều lớp. Đầu tiên, dữ liệu sẽ được đi qua nhiều lớp với các bộ lọc khác nhau để trích xuất các thông tin có giá trị và sau đó các đặc trưng có giá trị này sẽ được phân loại vào các lớp khác nhau. Ngoài việc ứng dụng trong lĩnh vực thị giác máy tính để trích xuất các đặc trưng của hình ảnh, thì mô hình này còn được ứng dụng trong xử lý ngôn ngữ tự nhiên được gọi là Text-CNN, cũng có tác dụng trích xuất các thông tin có giá trị từ văn bản.

**LSTM [5]:** Long Short Term Memory là mô hình được cải tiến từ mạng nơ-ron hồi quy (Recurrent Neural Network - RNN). Mô hình LSTM học được các thông tin trước đó trong văn bản, đây là một vấn đề mà RNN gặp phải. Trong mô hình LSTM có 3 cổng đó là cổng input, cổng output, và cổng forget. Cổng input sẽ lựa chọn các thông tin đưa vào ngữ cảnh, cổng output có tác dụng quyết định các thông tin nào là cần thiết ở thời điểm hiện tại và cổng forget sẽ xóa các thông tin khỏi ngữ cảnh nếu thông tin này không có giá trị. Mô hình này giúp cho việc phân loại văn bản trở nên tốt hơn so với RNN vì mô hình có thể nắm được ngữ cảnh trong toàn bộ văn bản.

**GRU [2]:** Gated Recurrent Unit là một biến thể của LSTM. Mô hình này có độ phức tạp ít hơn so với LSTM vì chỉ có 2 cổng là cổng update và cổng reset so với 3 cổng của LSTM. Cổng update xác định các thông tin trước đó sẽ được giữ lại dùng cho tương lai, và cổng forget có tác dụng quyết định các thông tin nào được giữ lại và các thông tin nào sẽ được bỏ đi. Do chỉ có 2 cổng nên số lượng tham số trong mô hình GRU là ít hơn so với LSTM, giảm được bộ nhớ và thời gian huấn luyện mô hình.

- Mô hình học chuyển tiếp:

Transformer [1] là một kiến trúc được thiết kế để giải quyết nhiều bài toán trong xử lý ngôn ngữ tự nhiên. Nếu như các mô hình RNN thông thường xử lý dữ liệu đầu vào một cách tuần tự thì transformer có thể xử lý cùng lúc, điều này đã tạo nên sự khác biệt cho mô hình có thể tận dụng khả năng tính toán cùng lúc, giúp giảm thời gian xử lý. Transformer sử dụng cơ chế self-attention, đây là cơ chế giúp encoder nhìn vào các từ khác trong lúc mã hóa một từ cụ thể. Điều này giúp cho các từ có thể điều chỉnh trọng số của nó với các từ khác trong câu sao cho từ ở vị trí càng gần thì trọng số của nó càng lớn và càng xa thì trọng số càng nhỏ dần. Kiến trúc của mô hình transformer gồm 6 encoder và 6 decoder. Mỗi encoder gồm 2 lớp là self-attention và mạng truyền thẳng (FPN). Các decoder cũng có kiến trúc như vậy nhưng giữa chúng có một lớp attention để mô hình có thể tập trung vào các phần liên quan của đầu vào.

Mô hình này được sử dụng phổ biến trong những năm gần đây, đặc biệt là với sự xuất hiện của BERT [4] đã giúp cho các tác vụ phía sau (downstream task) đạt hiệu quả cao trên các tập dữ liệu có kích thước nhỏ và nó đã trở thành hướng tiếp cận cơ bản trong nhiều tác vụ xử lý ngôn ngữ tự nhiên. Trong tác vụ phân loại các bài đánh giá spam hay không spam cũng như loại spam của bài đánh giá là gì, chúng tôi huấn luyện 2 mô hình BERT cho tiếng Việt là PhoBERT [11] và BERT4News [12].

### **Nội dung 3: Thử nghiệm và so sánh các phương pháp, phân tích kết quả đạt được trên bộ dữ liệu**

Sau khi đã xây dựng được bộ dữ liệu và nghiên cứu các phương pháp để giải quyết bài toán, chúng tôi dự định chia bộ dữ liệu thành các tập train, tập test và tập dev với tỉ lệ 7:2:1. Sau khi có được các tập dữ liệu, chúng tôi sẽ huấn luyện các mô hình và đánh giá kết quả. Độ đo mà chúng tôi sử dụng để đánh giá sẽ phụ thuộc vào phân phối giữa các nhãn trong tập dữ liệu, nếu bộ dữ liệu có sự mất cân bằng quá lớn giữa các lớp, chúng tôi sẽ sử dụng F1-macro để đánh giá khả năng nhận diện của mô hình, nếu sự mất cân bằng là không cao, chúng tôi sử dụng cả hai độ đo đánh giá là Accuracy và F1-macro, còn nếu bộ dữ liệu hoàn toàn không bị mất cân bằng chúng tôi chỉ sử dụng Accuracy để đánh giá kết quả. Sau khi đánh giá kết quả chúng tôi thực hiện phân tích kết quả, tìm hiểu lý do tại sao mô hình vẫn còn nhận diện chưa tốt trên một số lớp từ đó tìm hướng giải quyết để cải thiện hiệu suất trong tương lai.

### **Nội dung 4: Báo cáo tổng kết**

Thực hiện báo cáo tổng kết cho đề tài.

B2.3 Kế hoạch nghiên cứu

Chúng tôi thực hiện nghiên cứu đề tài Nhận diện các bài đánh giá rác trên các trang thương mại điện tử tiếng Việt trong vòng 6 tháng kể từ, kế hoạch thực hiện chi tiết được trình bày như trong Bảng 1:

Bảng 1. Kế hoạch thực hiện đề tài

Công việc	T1	T2	T3	T4	T5	T6
Xây dựng bộ dữ liệu cho bài toán						
Tìm hiểu các phương pháp						
Cài đặt, thử nghiệm các mô hình						
Báo cáo tổng kết						

B3. Kết quả dự kiến

Với đề tài nhận diện các bài đánh giá rác trên các trang thương mại điện tử tiếng Việt, chúng tôi đặt ra những kết quả nghiên cứu được trình bày như trong Bảng 2 dưới đây:

Bảng 2. Kết quả nghiên cứu dự kiến của đề tài

Nội dung	Kết quả nghiên cứu
1	- Xây dựng bộ dữ liệu cho bài toán nhận diện các bình luận rác trên các trang thương mại điện tử. - Sản phẩm dự kiến: bộ dữ liệu các bài đánh giá của người dùng về các sản phẩm đã được gán nhãn.
2	- Nghiên cứu một số mô hình học sâu như Text-CNN, LSTM, GRU và các mô hình học chuyển tiếp như Pho-BERT và BERT4New cho bài toán phân loại văn bản. - Sản phẩm dự kiến: báo cáo tổng kết.
3	- Chia bộ dữ liệu thành các tập train, tập test, và tập test. - Cài đặt, thử nghiệm một số phương pháp phân loại trên bộ dữ liệu, so sánh và phân tích kết quả đạt được. - Sản phẩm dự kiến: mã nguồn (source code) và báo cáo tổng kết.
4	- Thực hiện báo cáo tổng kết cho đề tài. - Sản phẩm dự kiến: báo cáo tổng kết.

#### **B4. Tài liệu tham khảo**

- [1]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is all you need (Năm 2017).
- [2]. Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder–decoder approaches. Trong: Proceedings of SSST- 8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. trang 103 - 111. Association for Computational Linguistics, Doha, Qatar (Tháng 10, 2014).
- [3]. Cho, K., van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder–decoder approaches. Trong: Proceedings of SSST- 8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. trang 103 - 111. Association for Computational Linguistics, Doha, Qatar (Tháng 10, 2014).
- [4]. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. Trong: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). trang 4171 - 4186. Association for Computational Linguistics, Minneapolis, Minnesota (Năm 2019).
- [5]. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), trang 1735 - 1780 (Năm 1997).
- [6]. Jindal, N., Liu, B.: Review spam detection. In: Proceedings of the 16th international conference on World Wide Web. trang 1189 - 1190 (Năm 2007).
- [7]. Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the 2008 international conference on web search and data mining. trang 219 - 230 (Năm 2008).
- [8]. Kim, Y.: Convolutional neural networks for sentence classification. Trong: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). trang 1746 - 1751. Association for Computational Linguistics, Doha, Qatar (Tháng 10, 2014).
- [9]. Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., Shao, J.: Bimodal distribution and co-bursting in review spam detection. Trong: Proceedings of the 26th International Conference on World Wide Web. trang 1063 - 1072 (Năm 2017).
- [10]. Luc Phan, L., Huynh Pham, P., Thi-Thanh Nguyen, K., Khai Huynh, S., Thi Nguyen, T., Thanh Nguyen, L., Van Huynh, T., Van Nguyen, K.: Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. Trong: Qiu, H.,



- Zhang, C., Fei, Z., Qiu, M., Kung, S.Y. (eds.) Knowledge Science, Engineering and Management. trang 647 - 658. Springer International Publishing, Cham (Năm 2021).
- [11]. Nguyen, D.Q., Tuan Nguyen, A.: PhoBERT: Pre-trained language models for Vietnamese. Trong: Findings of the Association for Computational Linguistics: EMNLP 2020. trang 1037 - 1042. Association for Computational Linguistics, Online (Tháng 11, 2020).
- [12]. Nguyen, T.C., Nguyen, V.N.: Nlpbk at vlsp-2020 shared task: Compose transformer pretrained models for reliable intelligence identification on social network. arXiv preprint arXiv:2101.12672 (Năm 2021).
- [13]. Nguyen, H.T., Nguyen, H.V., Ngo, Q.T., Vu, L.X., Tran, V.M., Ngo, B.X., Le, C.A.: Vlsip shared task: sentiment analysis. Journal of Computer Science and Cybernetics 34(4), trang 295 - 310 (Năm 2018).
- [14]. Radovanović, D., Krstajić, B.: Review spam detection using machine learning. Trong: 2018 23rd International Scientific-Professional Conference on Information Technology (IT). trang 1 - 4 (Năm 2018).
- [15]. Ren, Y., Zhang, Y.: Deceptive opinion spam detection using neural network. Trong: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. trang 140 - 150 (Năm 2016).
- [16]. Van Thin, D., Nguyen, N.L.T., Truong, T.M., Le, L.S., Vo, D.T.: Two new large corpora for vietnamese aspect-based sentiment analysis at sentence level. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20(4) (Tháng 5, 2021).

*Ngày 23 tháng 04 năm 2022*  
**Giảng viên hướng dẫn**  
(Ký và ghi rõ họ tên)

Lưu Thanh Sơn

*Ngày 23 tháng 04 năm 2022*  
**Chủ nhiệm đề tài**  
(Ký và ghi rõ họ tên)

Đinh Văn Cơ