

## **ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP**

# **ĐỀ TÀI: XÂY DỰNG MÔ HÌNH LỘC ĐÁNH GIÁ NGƯỜI DÙNG NHẪM CẢI THIỆN CHẤT LƯỢNG CỦA HỆ THỐNG ĐỀ XUẤT**

Ngành: **Công nghệ thông tin**  
Chuyên ngành: **Công nghệ phần mềm**

Giảng viên hướng dẫn : Võ Đình Bảy  
Sinh viên thực hiện : Trương Văn Đạt  
MSSV: 2180608812 Lớp: 21DTHD6

TP. Hồ Chí Minh, 2025

[illegible]

**Giảng viên hướng dẫn**

## This image shows a full page of a document template designed for writing. It features a series of evenly spaced, horizontal grey lines that run across the entire width of the page. The lines are thin and light in color, providing a guide for text alignment without being distracting. There are no margins, headers, footers, or other markings present on the page.

**Hội đồng xét duyệt**

## **Mục lục**

I. LÍ DO CHỌN ĐỀ TÀI.....	4
II. MỤC TIÊU ĐỀ TÀI .....	4
III. TỔNG QUAN CƠ SỞ LÝ LUẬN .....	5
IV. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU .....	6
4.1. Nội dung nghiên cứu.....	6
4.2. Phương pháp nghiên cứu .....	7
4.3. Phạm vi nghiên cứu .....	8
V. DỰ KIẾN KẾT QUẢ .....	8
VI. TÀI LIỆU THAM KHẢO.....	8
VII. TIẾN ĐỘ THỰC HIỆN .....	9

## I. LÍ DO CHỌN ĐỀ TÀI

Trong bối cảnh hiện nay, các hệ thống đề xuất đã trở thành công cụ không thể thiếu trong nhiều lĩnh vực như thương mại điện tử, truyền thông xã hội, học trực tuyến và các nền tảng giải trí. Việc cải thiện chất lượng và độ chính xác của các hệ thống này có thể giúp nâng cao trải nghiệm người dùng và thúc đẩy sự phát triển của doanh nghiệp. Chính vì vậy, việc xây dựng một mô hình lọc đánh giá người dùng nhằm cải thiện hiệu quả của hệ thống đề xuất là một vấn đề quan trọng và có tính ứng dụng cao.

Các mô hình đề xuất hiện nay đôi khi vẫn chưa đáp ứng được yêu cầu về độ chính xác trong việc đưa ra các gợi ý phù hợp với nhu cầu và sở thích của người dùng. Việc ứng dụng các kỹ thuật lọc đánh giá người dùng sẽ giúp thu thập thông tin đánh giá một cách chính xác và khách quan hơn, từ đó tối ưu hóa việc đưa ra các đề xuất, giúp hệ thống hoạt động hiệu quả hơn.

Người dùng ngày càng yêu cầu các trải nghiệm cá nhân hóa cao hơn từ các dịch vụ trực tuyến. Một mô hình lọc đánh giá người dùng chính xác sẽ không chỉ giúp hệ thống đề xuất đưa ra những sản phẩm hoặc dịch vụ phù hợp với nhu cầu của người dùng mà còn giúp họ tiết kiệm thời gian và công sức khi tìm kiếm thông tin, sản phẩm, hoặc nội dung cần thiết.

Đề tài này không chỉ mang tính thực tiễn cao mà còn có giá trị học thuật sâu sắc. Việc nghiên cứu và xây dựng các mô hình lọc đánh giá người dùng sẽ giúp hiểu rõ hơn về hành vi người dùng, cải thiện các phương pháp phân tích dữ liệu và học máy, đồng thời đóng góp vào sự phát triển của các phương thức cải tiến hệ thống đề xuất trong khoa học máy tính và các ngành ứng dụng liên quan.

Một trong những thách thức lớn nhất khi xây dựng hệ thống đề xuất là cách thức xử lý dữ liệu người dùng một cách hiệu quả. Việc đánh giá và phân tích chính xác hành vi của người dùng thông qua các mô hình lọc đánh giá sẽ giúp cải thiện hệ thống đề xuất và tạo ra những khuyến nghị có giá trị, từ đó nâng cao chất lượng sản phẩm và dịch vụ.

## II. MỤC TIÊU ĐỀ TÀI

Mục tiêu chính của đề tài "Xây dựng mô hình lọc đánh giá người dùng nhằm nâng cao hệ thống đề xuất" là phát triển một mô hình lọc đánh giá hiệu quả, từ đó cải thiện chất lượng và độ chính xác của các hệ thống đề xuất. Cụ thể, các mục tiêu của đề tài bao gồm:

- Xây dựng mô hình lọc đánh giá người dùng:** Phát triển một mô hình lọc đánh giá người dùng dựa trên sự kết hợp giữa điểm số (rating) và nội dung bình luận (comment). Mô hình này sẽ giúp hệ thống phân tích và đánh giá mức độ tin cậy của các đánh giá từ người dùng, từ đó cung cấp thông tin chính xác hơn về sản phẩm, dịch vụ.
- Cải thiện chất lượng hệ thống đề xuất:** Sử dụng mô hình lọc đánh giá để tối ưu hóa quá trình đề xuất, giúp hệ thống đưa ra các gợi ý phù hợp và chính xác hơn với sở thích và nhu cầu thực tế của người dùng. Điều này sẽ góp phần nâng cao hiệu quả hoạt động của hệ thống đề xuất, từ đó nâng cao trải nghiệm người dùng.

3. **Nâng cao khả năng cá nhân hóa đề xuất:** Tạo ra một hệ thống đề xuất có khả năng cá nhân hóa cao, phù hợp với hành vi và sở thích của từng người dùng. Bằng cách phân tích chi tiết các đánh giá, hệ thống có thể đưa ra các đề xuất sản phẩm, dịch vụ một cách chính xác và kịp thời, giúp người dùng tiết kiệm thời gian tìm kiếm và lựa chọn.
4. **Tối ưu hóa dữ liệu người dùng để nâng cao độ chính xác:** Mô hình lọc đánh giá sẽ phân tích dữ liệu người dùng một cách chính xác, giúp nhận diện các xu hướng và thói quen của người dùng. Việc tối ưu hóa dữ liệu này sẽ giúp hệ thống đề xuất học hỏi và cải thiện liên tục từ những đánh giá và tương tác thực tế của người dùng.
5. **Đảm bảo tính khả thi và mở rộng của mô hình:** Mô hình lọc đánh giá được xây dựng sẽ có tính linh hoạt cao, có thể mở rộng và áp dụng trong nhiều lĩnh vực khác nhau như thương mại điện tử, giáo dục, giải trí và các nền tảng trực tuyến. Mục tiêu là tạo ra một mô hình có thể ứng dụng rộng rãi và dễ dàng điều chỉnh theo yêu cầu của từng hệ thống đề xuất.

### III. TỔNG QUAN CƠ SỞ LÝ LUẬN

Trong những năm gần đây, các hệ thống đề xuất đã trở thành một công cụ quan trọng trong nhiều lĩnh vực ứng dụng, đặc biệt là trong thương mại điện tử, nền tảng giải trí, học trực tuyến, và các mạng xã hội. Các hệ thống này giúp người dùng tìm kiếm thông tin, sản phẩm, và dịch vụ một cách nhanh chóng và chính xác, dựa trên sở thích và hành vi trước đó của họ. Tuy nhiên, để nâng cao hiệu quả của các hệ thống đề xuất, cần phải có những phương pháp tối ưu hóa, đặc biệt là trong việc lọc và đánh giá các thông tin người dùng cung cấp.

1. **Hệ thống đề xuất và vai trò của chúng:** Hệ thống đề xuất (recommender systems) là những công cụ giúp cung cấp các gợi ý sản phẩm, dịch vụ, hay nội dung cho người dùng dựa trên các tiêu chí cụ thể như sở thích, hành vi trước đó, hoặc thông tin từ các người dùng tương tự. Các phương pháp phổ biến được sử dụng trong hệ thống đề xuất bao gồm:
  - **Phương pháp lọc cộng tác (Collaborative Filtering):** Dựa trên hành vi của người dùng tương tự để đưa ra gợi ý. Đây là một trong những phương pháp phổ biến nhất và được sử dụng rộng rãi trong các nền tảng như Netflix, Amazon.
  - **Phương pháp lọc nội dung (Content-Based Filtering):** Dựa trên các đặc tính của sản phẩm hoặc dịch vụ để đưa ra đề xuất cho người dùng. Ví dụ, dựa trên các thông tin mô tả sản phẩm mà người dùng đã xem hoặc mua trước đó.
  - **Phương pháp kết hợp (Hybrid Methods):** Kết hợp nhiều phương pháp khác nhau để tối ưu hóa các đề xuất và giảm thiểu độ sai lệch.
2. **Lọc đánh giá người dùng trong hệ thống đề xuất:** Việc lọc và phân tích các đánh giá của người dùng có vai trò quan trọng trong việc nâng cao chất lượng các gợi ý của hệ thống đề xuất. Đánh giá người dùng thường được thể hiện dưới dạng điểm số (rating) và các bình luận (comment), và việc xử lý hiệu quả những dữ liệu này là điều kiện cần thiết để hệ thống đề xuất hoạt động chính xác.
  - **Lọc đánh giá dựa trên điểm số:** Phương pháp này sử dụng các điểm số đánh giá sản phẩm của người dùng để xác định mức độ phù hợp của sản

phẩm với người dùng trong tương lai. Tuy nhiên, điểm số đôi khi không phản ánh chính xác trải nghiệm của người dùng nếu không đi kèm với các bình luận giải thích chi tiết.

- **Lọc đánh giá dựa trên bình luận:** Các bình luận đi kèm với điểm số đánh giá có thể cung cấp thông tin quý giá về những khía cạnh cụ thể của sản phẩm mà người dùng đánh giá cao hoặc không hài lòng. Phân tích các bình luận này thông qua các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) sẽ giúp xác định được cảm xúc và các yếu tố quan trọng trong mỗi đánh giá.
- 3. **Kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) trong phân tích bình luận:** Các kỹ thuật xử lý ngôn ngữ tự nhiên đóng vai trò quan trọng trong việc phân tích nội dung bình luận của người dùng. Các phương pháp như phân tích cảm xúc (sentiment analysis), phân tích chủ đề (topic modeling), và phân loại văn bản (text classification) giúp hệ thống hiểu được cảm xúc (tích cực, tiêu cực, trung lập) và các yếu tố chủ yếu mà người dùng đề cập đến trong bình luận. Nhờ vào các kỹ thuật này, hệ thống có thể đưa ra các đề xuất chính xác và có giá trị hơn dựa trên sự phân tích chi tiết các đánh giá người dùng.
- 4. **Lọc đánh giá kết hợp giữa điểm số và bình luận:** Một phương pháp quan trọng trong việc lọc đánh giá là kết hợp cả điểm số và nội dung bình luận để nâng cao độ chính xác và sự khách quan của mỗi đánh giá. Điểm số sẽ cung cấp cái nhìn tổng quan về sự hài lòng của người dùng đối với sản phẩm, trong khi bình luận lại giải thích chi tiết hơn về lý do đằng sau điểm số đó. Bằng cách đối chiếu và phân tích sự khớp nhau giữa hai yếu tố này, hệ thống có thể phát hiện những đánh giá không chính xác hoặc không đáng tin cậy, từ đó cải thiện chất lượng của hệ thống đề xuất.
- 5. **Các thách thức trong việc lọc đánh giá người dùng:** Mặc dù phương pháp lọc đánh giá có thể giúp cải thiện chất lượng hệ thống đề xuất, nhưng cũng tồn tại nhiều thách thức. Một trong những vấn đề chính là sự không đồng nhất trong các đánh giá của người dùng. Điểm số có thể dễ dàng bị thao túng hoặc không phản ánh chính xác trải nghiệm của người dùng, trong khi bình luận có thể chứa đựng nhiều yếu tố cảm xúc hoặc không có đủ thông tin chi tiết. Do đó, việc kết hợp cả hai yếu tố này một cách hợp lý và hiệu quả là một vấn đề cần phải được giải quyết trong nghiên cứu này.

## IV. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 4.1. Nội dung nghiên cứu

Nội dung nghiên cứu của đề tài "Xây dựng mô hình lọc đánh giá người dùng nhằm nâng cao hệ thống đề xuất" sẽ bao gồm các bước chính sau:

- **Xây dựng mô hình lọc đánh giá người dùng:** Mô hình này sẽ được thiết kế để phân tích và đánh giá các yếu tố như điểm số (rating) và bình luận (comment) của người dùng trong việc đánh giá sản phẩm. Mục tiêu là cải thiện chất lượng đánh giá và giúp hệ thống đề xuất đưa ra những gợi ý phù hợp hơn với nhu cầu và sở thích người dùng.
- **Phân tích điểm số và bình luận của người dùng:** Sự kết hợp giữa điểm số đánh giá và bình luận là yếu tố quan trọng trong việc cải thiện độ chính xác của hệ thống đề xuất. Phần này sẽ nghiên cứu cách thức đối chiếu và phân tích giữa hai

yếu tố này để đảm bảo rằng các đánh giá của người dùng phản ánh chính xác chất lượng sản phẩm.

- **Xử lý ngôn ngữ tự nhiên (NLP) trong việc phân tích bình luận:** Việc sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên sẽ giúp hiểu rõ hơn về cảm xúc và nội dung trong các bình luận của người dùng. Phương pháp phân tích cảm xúc (sentiment analysis), phân tích chủ đề (topic modeling), và phân loại văn bản (text classification) sẽ được áp dụng để hiểu được các yếu tố chủ đạo trong các đánh giá.
- **Tối ưu hóa hệ thống đề xuất:** Dựa trên kết quả thu được từ mô hình lọc đánh giá, hệ thống đề xuất sẽ được tối ưu hóa để cung cấp các gợi ý chính xác và cá nhân hóa hơn cho người dùng. Phần này tập trung vào việc áp dụng các kỹ thuật học máy để cải thiện khả năng dự đoán và đưa ra đề xuất phù hợp với từng người dùng.
- **Đánh giá kết quả mô hình:** Cuối cùng, sẽ tiến hành đánh giá hiệu quả của mô hình bằng cách sử dụng các chỉ số như độ chính xác, độ phủ và độ hài lòng của người dùng. Các thử nghiệm sẽ được thực hiện trên tập dữ liệu thực tế để kiểm tra khả năng hoạt động của hệ thống trong môi trường thực tế.

## 4.2. Phương pháp nghiên cứu

Để đạt được các mục tiêu của đề tài, phương pháp nghiên cứu sẽ bao gồm các bước sau:

- **Nghiên cứu lý thuyết và tổng quan:** Đầu tiên, tiến hành nghiên cứu các lý thuyết cơ bản về hệ thống đề xuất, các phương pháp lọc đánh giá người dùng, và các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP). Nghiên cứu này sẽ giúp xây dựng nền tảng lý thuyết vững chắc cho việc phát triển mô hình.
- **Thu thập và xử lý dữ liệu:** Dữ liệu cần thiết sẽ được thu thập từ các nền tảng trực tuyến có hệ thống đánh giá người dùng, chẳng hạn như các trang web thương mại điện tử hoặc các nền tảng chia sẻ sản phẩm. Dữ liệu sẽ bao gồm điểm số đánh giá và bình luận của người dùng. Sau khi thu thập, dữ liệu sẽ được tiền xử lý để loại bỏ thông tin không cần thiết và chuẩn hóa cho phù hợp với mô hình nghiên cứu.
- **Phát triển mô hình lọc đánh giá:** Dựa trên dữ liệu đã thu thập, sẽ xây dựng một mô hình lọc đánh giá người dùng sử dụng phương pháp kết hợp điểm số và bình luận. Mô hình này sẽ áp dụng các kỹ thuật học máy, bao gồm các phương pháp phân tích cảm xúc và phân loại văn bản để phân tích bình luận người dùng và kết hợp với điểm số để đưa ra đánh giá tổng thể.
- **Áp dụng các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP):** Để xử lý các bình luận của người dùng, sẽ áp dụng các phương pháp như phân tích cảm xúc (sentiment analysis), phân tích chủ đề (topic modeling), và phân loại văn bản (text classification). Các mô hình NLP này giúp xác định các yếu tố quan trọng trong các bình luận và cải thiện sự chính xác của việc lọc đánh giá.
- **Tối ưu hóa và kiểm tra mô hình:** Sau khi xây dựng mô hình, sẽ áp dụng các thuật toán học máy như học sâu (deep learning), học có giám sát (supervised learning) hoặc học không giám sát (unsupervised learning) để tối ưu hóa mô hình. Mô hình sẽ được kiểm tra bằng các chỉ số độ chính xác (accuracy), độ phủ (coverage), và độ hài lòng của người dùng.



- **Đánh giá kết quả:** Cuối cùng, kết quả mô hình sẽ được đánh giá dựa trên các chỉ tiêu đo lường như độ chính xác, độ tin cậy của các đề xuất, và mức độ cải thiện trải nghiệm người dùng. Các chỉ số này sẽ giúp xác định hiệu quả của mô hình trong việc nâng cao chất lượng hệ thống đề xuất.

#### 4.3. Phạm vi nghiên cứu

Nghiên cứu này sẽ tập trung vào việc xây dựng mô hình lọc đánh giá người dùng trong bối cảnh các hệ thống đề xuất cho thương mại điện tử. Dữ liệu đầu vào chủ yếu là các đánh giá sản phẩm từ người dùng, bao gồm điểm số và bình luận. Hệ thống đề xuất sau khi tối ưu hóa sẽ được kiểm tra và đánh giá qua các thử nghiệm thực tế để kiểm tra khả năng ứng dụng trong môi trường thực tiễn.

### V. DỰ KIẾN KẾT QUẢ

- Mô hình lọc đánh giá chính xác: Xây dựng mô hình lọc đánh giá người dùng có khả năng phân tích điểm số và bình luận để cung cấp đánh giá chính xác, đáng tin cậy về sản phẩm.
- Cải thiện hệ thống đề xuất: Hệ thống đề xuất sẽ trở nên chính xác và cá nhân hóa hơn, cung cấp gợi ý phù hợp với nhu cầu và sở thích người dùng.
- Nâng cao trải nghiệm người dùng: Người dùng sẽ nhận được các gợi ý chính xác, tiết kiệm thời gian và tăng sự hài lòng.

### VI. TÀI LIỆU THAM KHẢO

- [1] Nguyen, D. Q., & Nguyen, A. T., „PhoBERT: Pre-trained language models for Vietnamese,“ in *In Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.
- [2] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022)., „A survey on sentiment analysis methods, applications, and challenges,“ *Artificial Intelligence Review*, Bd. 55, Nr. 7, pp. 5731-5780, 2022.
- [3] N. T. Thinh, „Đánh giá model trong Machine Learning,“ 29 5 2022. [Online]. Available: <https://viblo.asia/p/danh-gia-model-trong-machine-learning-RnB5pAq7KPG>. [Zugriff am 18 05 2024].

## VII. TIẾN ĐỘ THỰC HIỆN

<b>Tuần</b> <b>Dự kiến nội dung thực hiện</b>	<b>1, 2</b>	<b>3, 4</b>	<b>5, 6</b>	<b>7, 8</b>	<b>9, 10</b>	<b>11, 12</b>	<b>13</b>
Nghiên cứu & Thu thập và phân tích dữ liệu							
Tiền xử lý dữ liệu & Gán nhãn dữ liệu							
Xây dựng mô hình lọc đánh giá							
Triển khai thử nghiệm trên tập dữ liệu & Kiểm tra mô hình							
Hoàn thiện mô hình và tối ưu hiệu suất hệ thống							
Viết tài liệu báo cáo & nội dung thuyết trình							
Chuẩn bị cho buổi bảo vệ đồ án & Tổng kết và đánh giá kết quả đạt được							