**TRƯỜNG ĐẠI HỌC KINH TẾ TP.HỒ CHÍ MINH**

**PHÂN HIỆU VĨNH LONG**



**THUYẾT MINH**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN NĂM 2023**

**KHAI PHÁ DỮ LIỆU TRONG THƯƠNG MẠI ĐIỆN TỬ - XÂY DỰNG HỆ THỐNG KHUYẾN NGHỊ ANIME**

Thuộc nhóm chuyên ngành khoa học: Công nghệ thông tin

Giảng viên hướng dẫn: Lê Hoàng Son

Nhóm thực hiện:

Chủ nhiệm: Trần Thị Thùy Trang

Thành viên 1: Phạm Nguyễn Tấn Phát

Thành viên 2: Nguyễn Thị Huỳnh Như

Thành viên 3: Lê Hiếu Nghĩa

Thành viên 4: Phan Ngọc Huỳnh Như

**Vĩnh Long, 2023**

TRƯỜNG ĐẠI HỌC KINH TẾ TP.HCM

**PHÂN HIỆU VĨNH LONG**

KHOA CÔNG NGHỆ THÔNG TIN

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc lập – Tự do – Hạnh phúc**

Vĩnh Long, ngày 13 tháng 03 năm 2023

**THUYẾT MINH**

**Đề tài nghiên cứu khoa học sinh viên năm 2023**

1. **Tên đề tài:**

Khai phá dữ liệu trong thương mại điện tử - xây dựng hệ thống khuyến nghị anime.

1. **Tác giá/nhóm tác giả sinh viên đăng ký:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Lớp** | **Khóa** | **Khoa** |
| 1 | Chủ nhiệm:  Trần Thị Thùy Trang | 31211572267 | EC001 | 47 | Công nghệ thông tin |
| 2 | Phạm Nguyễn Tấn Phát | 31211570300 | EC001 | 47 | Công nghệ thông tin |
| 3 | Nguyễn Thị Huỳnh Như | 31211572259 | EC001 | 47 | Công nghệ thông tin |
| 4 | Lê Hiếu Nghĩa | 31211572254 | EC001 | 47 | Công nghệ thông tin |
| 5 | Phan Ngọc Huỳnh Như | 31211572260 | EC001 | 47 | Công nghệ thông tin |

1. **Thuyết minh**
   1. **Lý do chọn đề tài**

Dưới sự tác động, phát triển liên tục và nhanh chóng của Internet, làm cho các thông tin trên môi trường mạng ngày càng trở nên phong phú. Đi cùng với sự phát triển đó đã làm tăng đáng kể nhu cầu sử dụng các hệ thống khuyến nghị để lọc thông tin một cách hiệu quả hơn, tiết kiệm thời gian trong việc tìm kiếm các sản phẩm và dịch vụ. Hệ thống khuyến nghị là nơi đưa ra những gợi ý dựa trên các dữ liệu của người dùng như lọc những bài hát được nhiều người nghe, được nhiều lượt tìm kiếm nhất; các sản phẩm có nhiều lượt mua và được người tiêu dùng đánh giá tốt nhất; hay bảng xếp hạng các bộ phim có nhiều lượt xem,…

Anime là từ dùng để nói về các bộ phim hoạt hình sản xuất từ các truyện tranh Manga của Nhật Bản và Anime được cả thế giới công nhận rằng nó là một nét văn hóa rất đặc trưng của người Nhật. Với độ nổi tiếng toàn cầu đó thì cũng chứng minh được những bộ phim hoạt hình đến từ Nhật Bản có rất nhiều người hâm mộ trên khắp toàn thế giới. Nhưng để có một hệ thống chuyên biệt chuyên đưa ra những gợi ý liên quan đến những bộ phim mà khách hàng quan tâm thì vẫn còn quá ít.

Vì vậy, chúng em nghiên cứu và phát triển đề tài này với mong muốn có thể giúp những người yêu thích anime có thể dễ dàng tìm kiếm những bộ phim có thể loại giống nhau, dựa trên những tìm kiếm của khách hàng hiện tại để đưa ra những gợi ý.

* 1. **Mục tiêu nghiên cứu**

Trong báo cáo này, mục tiêu chính là đưa ra được một hệ thống khuyến nghị amine phù hợp nhất với nhu cầu của người dùng. Hệ thống có thể đưa vào ứng dụng được, nhằm mục tiêu gia tăng xác suất chọn được những bộ anime mà khách hàng yêu thích, làm tăng khả năng hứng thú cũng như trải nghiệm dịch vụ một cách tốt nhất khi sử dụng hệ thống.

Để làm được điều đó, trước hết chúng ta cần xây dựng được một hệ thống mô hình phục vụ cho việc dự đoán xu thế tìm kiếm anime của khách hàng, đâu là những bộ anime được người dùng yêu thích và đánh giá cao.

* 1. **Phương pháp nghiên cứu – mô hình dự kiến – nguồn số liệu dự kiến**
     1. **Phương pháp nghiên cứu**

Phương pháp thu thập dữ liệu: các dữ liệu về kho phim, nội dung, đánh giá của khán giả đối với các bộ phim anime

Phương pháp phân tích thiết kế: thiết kế mô hình khai phá luật kết hợp, mạng Nerual, cây quyết định, mạng Bayes.

Sử dụng các ngôn ngữ lập trình để xây dựng giải thuật.

Phương pháp thực nghiệm: dựa trên nguồn dữ liệu thu thập và tổng hợp được, tiến hành chuẩn hóa dữ liệu và ứng dụng xây dựng một hệ thống khuyến nghị các bộ phim anime khán giả nên xem.

* + 1. **Mô hình dự kiến**
       1. **Tổng quan nghiên cứu**

***Một số khái niệm:***

Vào đầu thập kỷ trước với sự phát triển của internet đã bao gồm một số loại ứng dụng web cơ bản liên quan đến thương mại điện tử, sự tương tác của con người, tin tức thực tế và chia sẻ thông tin. Do đó đã tạo nên nguồn dữ liệu khổng lồ trên internet, tạo điều kiện cho việc truy xuất thông tin, dữ liệu từ thực tế. Nhưng để chọn lựa ra những dữ liệu có giá trị cần thiết thì đó còn là một điều khó khăn. Do đó, hệ thống khuyến nghị đã trở thành một nhu cầu cần thiết cho tất cả mọi người. Khi thông tin thu thập càng nhiều thì tầm nhìn trên các ứng dụng, web càng mở rộng và có thể gợi ý cho người dùng những gì họ thực sự tìm kiếm trong thời gian ngắn nhất có thể. Do đó các trang web thương mại điện tử có thể bán nhiều sản phẩm, dịch vụ hơn những gì họ đã mong đợi, tất cả cũng nhờ hệ thống gợi ý sản phẩm. Không những thế người dùng còn có thể thấy được những sản phẩm, dịch vụ mới có liên quan đến thứ mà họ muốn mua. Các hệ thống khuyến nghị đã xuất hiện vào giữa những năm 1990 và chúng đề xuất các mục thực sự thú vị cho người dùng. Để giới thiệu một dịch vụ hoặc sản phẩm cho một người dùng cụ thể, các hệ thống khuyến nghị đã thu thập thông tin, dữ liệu về người dùng đó một cách âm thầm hoặc công khai. Quy trình thu thập công khai được thực hiện bởi sự nhận thức của người dùng, chẳng hạn như thực hiện các cuộc khảo sát. Còn các dữ liệu được thu thập một cách âm thầm khi người dùng nhấp vào liên kết; đưa ra nhận xét về một sản phẩm, dịch vụ đã từng sử dụng hoặc thu thập từ những tìm kiếm của khách hàng, hồ sơ người dùng trong đó có các tùy chọn và sở thích được lưu trữ. Có rất nhiều ứng dụng web giới thiệu các loại sản phẩm. Chẳng hạn, Amazon đề xuất sách cho khách hàng, YouTube là một ứng dụng web nổi tiếng đề xuất video…

Trong Khoa học máy tính (hay Computer Science), RS là một nhánh con của Khai phá dữ liệu (hay Data mining) và thuộc lĩnh vực lớn là Trí tuệ nhân tạo (Artificial Intelligence), mặc dù khá đa dạng về thuật toán nhưng chúng thường được chia thành ba nhóm thuật toán chính:

- Hệ thống khuyến nghị dựa trên nội dung (Content-based Recommendation System - CBRS).

- Hệ thống khuyến nghị dựa trên lịch sử hành vi người dùng – hay hệ thống khuyến nghị lọc cộng tác (Collaborative Filtering Recommendation System - CFRS).

- Hệ thống khuyến nghị Knowledge-Based Recommendation System - KBRS.

Thông tin về người sử dụng được dùng cho việc khuyến nghị tự động có thể chia thành hai loại: thông tin ẩn và thông tin hiện. Các thông tin ẩn bao gồm các thông tin như thời gian người mua xem sản phẩm, quá trình người dùng duyệt qua các sản phẩm trước khi chọn sản phẩm cuối để xem, nguồn mà người dùng truy cập đến trang hiện tại… Các nội dung này rất dễ thu thập, nhưng thường có độ nhiễu cao và không có tương tác trực tiếp với người dùng (hệ thống tự động ghi lại lịch sử). Các thông tin hiện là các kết quả nhận được khi người dùng trực tiếp đưa ra đánh giá về sản phẩm như thích/không thích, chấm điểm, bình luận… Các thông tin này thường khó thu thập hơn rất nhiều so với các thông tin ẩn, nhưng độ chính xác cao hơn và rất hữu ích cho việc khuyến nghị.

Kết quả của một hệ tư vấn là dự đoán “đánh giá” của người dùng cho một sản phẩm nhất định (ví dụ, khả năng 80% người dùng sẽ thích sản phẩm này) và danh sách những sản phẩm người dùng có thể sẽ thích.

* + - 1. **Mô hình hệ thống khuyến nghị**

Đưa ra khuyến nghị những bộ anime phù hợp yêu cầu người sử dụng.

Lọc cộng tác

Lọc dựa trên mô hình

Lọc dựa trên bộ nhớ

* Mạng Bayes
* Mạng Nerual
* Luật kết hợp
* Cây quyết định
* Lịch sử xem
* Đánh giá
* Bình luận
* Thích/không thích
* Bình luận
* Đánh giá.
* Số tập
* Thể loại, loại phim.
* Lượt xem

***Hệ thống khuyến nghị anime***

***Chức năng***

*Cải thiện trải nghiệm người dùng:* từ việc dự đoán và đưa ra những bộ phim anime đúng với sở thích của khách hàng sẽ làm tăng sự hài lòng của khách hàng.

*Tăng hiệu năng hoạt động bằng tự động hóa:* việc khuyến nghị sản phẩm truyền thống thường được làm thủ công, hiệu quả không cao và bị hạn chế về hiệu năng.

*Biến khách hàng tiềm năng thành khách hàng thật:* đúng với mục đích của bài toán, hệ thống sẽ tư vấn được những sản phẩm tiềm năng mà ngay cả khách hàng cũng chưa nghĩ đến.

* 1. **Kết quả kỳ vọng**

Về kiến thức: Thông qua nội dung và kết quả của bài nghiên cứu sẽ góp phần cung cấp cho người đọc về những kiến thức tổng quan về hệ thống khuyến nghị (Recommendation System – RS). Hệ thống khuyến nghị giúp tự động gợi ý, đưa ra các nội dung liên quan cho người dùng dựa trên thông tin mà họ đã hoặc đang xem mà không cần họ phải cung cấp thêm bất kỳ một dữ liệu nào khác.

Về thực tiễn: Việc ứng dụng hệ thống khuyến nghị có thể cải thiện trải nghiệm người dùng từ việc dự đoán và đưa ra danh sách thể loại anime đúng với sở thích, cũng như hệ thống sẽ gợi ý được những thể loại khác (sản phẩm tiềm năng) mà ngay cả khách hàng cũng chưa biết đến.

* 1. **Kết cấu báo cáo dự kiến**

*Chương 1:* Tổng quan.

* 1. Mục tiêu nghiên cứu.
  2. Phương pháp nghiên cứu.
  3. Đối tượng và phạm vi nghiên cứu.
  4. Kết cấu đề tài.

*Chương* 2: Cơ sở lý thuyết.

1. Tổng quan về khai phá dữ liệu.
2. Các thuật toán máy học.
3. Hệ thống khuyến nghị.

*Chương 3*: Xây dựng hệ thống khuyến nghị anime.

1. Ứng dụng, triển khai hệ thống khuyến nghị anime.
2. Bài toán.
3. Xây dựng hệ thống.

*Chương 4:* Kết luận.

1. Kết quả đạt được.
2. Hướng phát triển.

Tài liệu tham khảo.

* 1. **Kinh phí đề xuất: 3 triệu**
  2. **Thời gian thực hiện: 6 tháng**

**Tài liệu tham khảo**

1. Vũ Đức Thi (1997), *Cơ sở dữ liệu-Kiến thức và thực hành*, Nhà xuất bản Thống kê, Hà Nội.

2. Vũ Đức Thi (1999), *Thuật toán trong tin học*, Nhà xuất bản Khoa học và kỹ thuật, Hà Nội.

3. Lê Tiến Vương (1996), *Nhập môn cơ sở dữ liệu quan hệ*, Nhà xuất bản Khoa học và kỹ thuật, Hà Nội.

(\*) Zhu, Bo Bobadilla, Jesus Ortega, Fernando. (2018). *Reliability quality measures for recommender systems*. Information Sciences.

4. Isinkaye, F.O., Y.O. Folajimi, and B.A. Ojokoh (2015), Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*.16(3): p. 261-273.

5. Guo, G. (2012). Resolving Data Sparsity and Cold Start in Recommender Systems. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (eds) User Modeling, Adaptation, and Personalization. UMAP 2012. Lecture Notes in Computer Science, vol 7379. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31454-4\_36

# 6. Emrah Inan, Fatih Tekbacak, Cemalettin Oztuek (2018), Moreopt: A goal programming based movie recommender system. *Journal of Computational Science*, page 43 – 50.

**Ý kiến của giảng viên hướng dẫn**

Đề tài khả thi và phù hợp với chuyên ngành học cũng như khả năng nghiên cứu của sinh viên.