



BÁO CÁO KIỂM TRA TRÙNG LẶP

Thông tin tài liệu

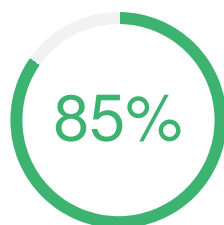
Tên tài liệu:	Báo cáo DATN TQChieu (v2)
Tác giả:	Tạ Quang Chiếu
Điểm trùng lặp:	15
Thời gian tải lên:	05:45 20/01/2024
Thời gian sinh báo cáo:	05:50 20/01/2024
Các trang kiểm tra:	62/62 trang



Kết quả kiểm tra trùng lặp



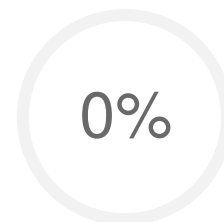
Có 15% nội dung trùng
lặp



Có 85% nội
dung không
trùng lặp



Có 0% nội dung
người dùng loại
trừ



Có 0% nội dung
hệ thống bỏ qua

Nguồn trùng lặp tiêu biểu

123docz.net tailieu.vn lib.uet.vnu.edu.vn

Danh sách các câu trùng lặp

1. Trang 5: Nhiệm vụ Đồ án tốt nghiệp đã được Hội đồng thi tốt nghiệp của Khoa thông qua

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Nhiệm vụ Đồ án tốt nghiệp đã được Hội đồng thi tốt nghiệp của Khoa thông qua

2. Trang 6: Sinh viên đã hoàn thành và nộp bản Đồ án tốt nghiệp cho Hội đồng thi ngày

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Sinh viên đã hoàn thành và nộp bản Đồ án tốt nghiệp cho Hội đồng thi ngày

3. Trang 7: Bài toán đề xuất phim là một phương tiện hữu hiệu để giải quyết vấn đề quá tải thông tin phim

Độ trùng lặp: **56%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một phương tiện có giá trị để giải quyết vấn đề quá tải thông tin

4. Trang 9: Tác giả xin cam đoan rằng Đồ án tốt nghiệp này là công trình nghiên cứu độc lập của bản thân Tác giả

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tác giả xin cam đoan đây là công trình nghiên cứu độc lập của bản thân Tác giả

5. Trang 9: Các kết quả nghiên cứu trong Đồ án được thực hiện một cách trung thực và không sao chép từ bất kỳ nguồn tài liệu nào

Độ trùng lặp: **73%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các kết quả nghiên cứu trong luận án là trung thực và không sao chép từ bất kỳ nguồn tài liệu khác dưới bất cứ hình thức Việc tham khảo Các nguồn tài liệu đã được thực hiện

6. Trang 9: Tất cả các nguồn tài liệu tham khảo đã được trích dẫn và ghi rõ nguồn gốc theo quy định

Độ trùng lặp: **84%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các nguồn tài liệu tham khảo đã được trích dẫn và ghi rõ nguồn gốc tài liệu tham khảo theo quy định

7. Trang 10: Nhờ sự chỉ bảo và giúp đỡ của thầy tôi đã học hỏi được nhiều kiến thức mới, và có thể hoàn thành đồ án tốt nghiệp kỳ này.

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tôi có thể hoàn thành đề án đúng thời hạn Qua đây tôi đã học hỏi được nhiều kiến

8. Trang 10: Với hạn chế về thời gian và kiến thức chuyên ngành, trong quá trình thực hiện đồ án, tôi không tránh khỏi việc mắc phải những sai sót

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thời gian và kiến thức còn hạn chế ở một số lĩnh vực nên không thể tránh khỏi việc mắc phải những thiếu sót trong quá trình thực hiện

9. Trang 10: do đó tôi rất mong nhận được sự đóng góp góp ý và đánh giá từ các thầy cô để có thể hoàn thiện đồ án của mình một cách tốt nhất

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Do đó, tôi rất mong nhận được sự đóng góp, ý kiến của các thầy cô trong khoa Đầu tư để có thể hoàn thiện chuyên đề tốt

10. Trang 14: Danh mục các từ VIẾT TẮT VÀ GIẢI THÍCH các THUẬT NGỮ

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: VIẾT TẮT VÀ GIẢI THÍCH THUẬT NGỮ x DANH MỤC CÁC TỪ VIẾT TẮT x GIẢI THÍCH CÁC THUẬT NGỮ

11. Trang 16: ngày nay, nhu cầu giải trí của con người ngày càng cao trong đó phim ảnh là một trong những hình thức giải trí phổ biến nhất

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của con người Ngày càng cao, và sử dụng dịch vụ karaoke là một trong những hình thức giải trí

12. Trang 17: bộ dữ liệu MovieLens MovieLens 100K là một bộ dữ liệu phổ biến được sử dụng trong nghiên cứu khoa học, bao gồm 100 000 xếp hạng từ 943 người dùng cho 1682 bộ phim với mỗi xếp hạng từ 1 đến 5

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 100 000 xếp hạng phim, từ 943 người dùng cho 1 682 Bộ phim, được phát hành vào tháng 04/1998 Bộ dữ liệu ml 1M là Bộ dữ liệu gồm 1 000 209 xếp hạng từ 6 040 người dùng cho 3 900 Bộ phim, được phát hành vào tháng 02/2003 Bộ dữ liệu ml 10M là Bộ dữ liệu gồm 10 000 054 xếp hạng từ 71 567 người dùng cho 10 681 Bộ phim, được phát hành vào tháng 02/2003 Bộ dữ liệu ml 20M là Bộ dữ liệu gồm 20 000 263 xếp hạng từ 138 493 người dùng cho 27 278 Bộ phim, được phát hành vào tháng 10/2016 Bộ dữ liệu ml 25M là Bộ dữ liệu gồm 25 000 095 xếp hạng từ 162 541 người dùng cho 62 423 Bộ phim, được phát hành vào tháng 10/2019 Bộ dữ liệu Movielens 1B Synthetic Dataset Đây là Bộ dữ liệu sử dụng tập dữ liệu ml 20M để sinh ra dữ liệu nhân tạo xấp xỉ gồm 1 223 962 043 xếp hạng cho dữ liệu huấn luyện và 12 709 557 xếp hạng cho dữ liệu kiểm tra Bên cạnh dữ liệu về xếp hạng phim, dự án Movielens còn cung cấp một

13. Trang 17: dữ liệu được thu thập thông qua trang web MovieLens (movielens umn edu) trong khoảng thời gian bảy tháng, từ ngày 19 tháng, 9 năm 1997 đến ngày 22 tháng, 4 năm 1998

Độ trùng lặp: 66%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được thu thập trong khoảng thời gian bảy tháng kể từ ngày 19 tháng 9 năm 1997 đến ngày 22 tháng 4 năm 1998

14. Trang 17: dữ liệu này đã được lọc để loại bỏ các người dùng có ít hơn 20 xếp hạng hoặc thiếu thông tin cá nhân đầy đủ

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: người dùng có ít hơn 20 xếp hạng hoặc thông tin nhân

15. Trang 17: U ITEM là file dữ liệu thông tin về các ITEM (phim) bao gồm các cột dữ liệu movie id (id phim), movie title (tiêu đề phim), release date (ngày phát hành), video release date (ngày phát hành video), URL IMDb, unknown, Action Adventure Animation Children's Comedy Crime Documentary Drama Fantasy Film Noir Horror Musical Mystery Romance Sci Fi Thriller War Western

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các thông tin về các

16. Trang 19: hệ đề xuất (Recommender System viết tắt là RS) là hệ thống sử dụng các kỹ thuật và công cụ phần mềm xử lý dữ liệu người dùng và dữ liệu về sản phẩm nhằm đáp ứng nhu cầu của người dùng về một sản phẩm hoặc dịch vụ nào đó [1]

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Recommender System viết tắt là RS) là Hệ thống sử dụng các kỹ thuật và công cụ phần mềm xử lý dữ liệu người dùng và dữ liệu về sản phẩm nhằm đáp ứng nhu cầu của người dùng về một sản phẩm hoặc dịch vụ nào đó

17. Trang 19: Dựa Trên cơ sở dữ liệu về hành vi tiêu dùng, trong quá khứ hệ thống thực hiện việc phân tích thông tin về sở thích, và thị hiếu của người dùng,

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trên cơ sở thông tin trong quá khứ, về hành vi của người dùng hệ thống thực hiện khai phá thông tin về sở thích thị hiếu tiêu dùng

18. Trang 19: Từ đó, hệ thống cung cấp các gợi ý để hỗ trợ người dùng trong việc lựa chọn sản phẩm và dịch vụ phù hợp với sở thích của họ

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: hệ thống cung cấp các gợi ý để giúp họ chọn lựa các sản phẩm hay dịch vụ phù hợp với nhu cầu và sở thích một cách nhanh chóng và dễ dàng hệ thống còn có thể cung cấp cho người dùng những gợi ý về các sản phẩm hay dịch vụ mới mà trước đây họ chưa từng

19. Trang 19: hệ thống đề xuất là một công cụ phần mềm, kỹ thuật cung cấp các đề xuất về các

đối tượng có thể hữu ích với người dùng như việc gợi ý bộ phim nên xem sản phẩm nên mua, bài hát nên nghe, hoặc tin tức nên đọc dựa trên quyết định của người dùng

Độ trùng lặp: **58%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: công cụ phần mềm kỹ thuật, cung cấp những đề xuất các đối tượng có thể hữu ích với người dùng. Những đề xuất liên quan đến quyết định của người dùng, như sản phẩm nào nên mua, cuốn sách nào nên đọc bài hát nào nên nghe, hay tin tức nào nên xem, Vào giữa thập niên s Hệ thống khuyến nghị được xem, như là một

20. Trang 20: Giáo dục Gợi ý nguồn tài nguyên học tập như sách, bài báo, khóa học trang Web, học tập cho người dùng thông qua các hệ thống của Foxtrot, InfoFinder

Độ trùng lặp: **62%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: Giáo dục gợi ý nguồn tài nguyên học tập nh sách, bài báo, khóa học địa chỉ web cho ng ời học Ví dụ hệ thống của Foxtrot, InfoFinder Giải trí gợi ý bài hát cho ng ời nghe (Ví dụ hệ thống của LastFM www last fm), gợi ý phim ảnh (Ví dụ hệ thống của Netflix, MovieLens, EachMovie), gợi ý các

21. Trang 20: người dùng là những cá nhân sử dụng hệ thống để thực hiện các thao tác mua bán, giao dịch, xem, đánh giá

Độ trùng lặp: **80%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: là Người sử dụng hệ thống để thực hiện các thao tác mua bán, giao dịch, xem, đánh giá

22. Trang 20: Bài toán được đặt ra như sau Cho tập hợp hữu hạn gồm N người dùng $U = \{u_1, u_2, \dots, u_N\}$, và M sản phẩm $P = \{p_1, p_2, \dots, p_M\}$

Độ trùng lặp: **72%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: như sau Cho tập hợp hữu hạn gồm N người dùng $U = \{u_1, u_2, \dots, u_N\}$ $P = \{p_1, p_2, \dots, p_M\}$ là tập hữu hạn gồm M sản phẩm Mỗi sản phẩm p_x

23. Trang 20: các đặc trưng tq \hat{T} thông thường là thông tin cá nhân của người dùng

Độ trùng lặp: **80%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: đặc trưng tq T thông thường là thông tin cá nhân của mỗi người dùng (Demographic Information) Ví dụ $i \in U$ là một người dùng thì Các đặc trưng

24. Trang 20: Ví dụ $u_i \in U$ là một người dùng thì các đặc trưng nội dung biểu diễn người dùng u_i có thể là $T \equiv \{\text{giới tính, tuổi, nghề nghiệp,}\}$

Độ trùng lặp: **79%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: là một người dùng thì các đặc trưng nội dung biểu diễn người dùng

25. Trang 21: Mỗi sản phẩm $p_x \in P$ (với $x \equiv 1, 2, \dots, M$) có thể là phim, ảnh, sách, báo, hàng hóa,

hoặc bất kỳ thông tin nào mà người dùng cần

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: P có thể là phim, ảnh, tạp chí, tài liệu, sách, báo, hàng hóa dịch vụ hoặc bất kỳ dạng thông tin nào mà người dùng cần

26. Trang 21: Các đặc trưng cs \hat{C} nhận được từ Các phương pháp trích chọn đặc trưng trong lĩnh vực truy vấn thông tin

Độ trùng lặp: **86%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các đặc trưng cs C nhận được từ Các phương pháp trích chọn đặc trưng (feature selection) trong lĩnh vực truy vấn thông tin

27. Trang 21: Ví dụ, px \hat{P} là một phim thì các đặc trưng nội dung biểu diễn phim px có thể là $C = \{$ thể loại phim tiêu đề, nước sản xuất, diễn viên, đạo diễn, $\}$

Độ trùng lặp: **70%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: P là một phim thì các đặc trưng nội dung biểu diễn một phim có thể là $C = \{$ thể loại phim nước sản xuất, hãng phim diễn viên, đạo diễn $\}$

28. Trang 21: Mối quan hệ giữa tập sản phẩm P và tập người dùng U được biểu diễn thông qua trận đánh giá $R = [r_{ix}]$, với $i = 1, 2, \dots, N$; $x = 1, 2, \dots, M$

Độ trùng lặp: **88%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Mối quan hệ giữa tập người dùng U và tập sản phẩm P được biểu diễn thông qua ma trận đánh giá $R = [r_{ix}]$ với $i = 1, 2, \dots, N$; $x = 1, 2, \dots, M$

29. Trang 21: giá trị rix thể hiện đánh giá của người dùng ui \hat{U} cho một số sản phẩm px \hat{P}

Độ trùng lặp: **81%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trị rix thể hiện đánh Giá của người dùng ui U cho một số sản phẩm px P Thông thường Giá trị rix

30. Trang 21: Thông thường giá trị rix nhận một số giá trị thuộc miền $F = \{1, 2, \dots, g\}$ được thu thập trực tiếp bằng cách hỏi ý kiến người dùng hoặc thu thập gián tiếp Thông qua cơ chế phản hồi của người dùng [2]

Độ trùng lặp: **95%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thông thường giá trị rix nhận một giá trị thuộc miền $F = \{1, 2, \dots, g\}$ được thu thập trực tiếp bằng cách hỏi ý kiến người dùng hoặc thu thập gián tiếp Thông qua cơ chế phản hồi của người dùng

31. Trang 21: Những giá trị rix $= 0$ được hiểu là người dùng ui \hat{U} chưa biết đến hoặc không đánh giá sản phẩm px \hat{P} , Những ô điền ký tự là giá trị cần hệ đề xuất đưa ra dự đoán đánh giá

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: a người dùng Những giá trị = 0 được hiểu là người dùng chưa biết đến hoặc không đánh giá sản phẩm, Những ô điền ký tự là giá trị cần hệ tư vấn đư

32. Trang 22: Tiếp đến ta ký hiệu, P_i là tập các sản phẩm $p \in P$ được đánh giá bởi người dùng $u \in U$ và $u \in U$ được gọi là người dùng hiện tại, người dùng cần được tư vấn hay người dùng tích cực

Độ trùng lặp: 72%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tiếp đến, ta ký hiệu P_i là tập các sản phẩm $p \in P$ được đánh giá bởi người dùng $i \in U$ và U_x là tập các người dùng đã đánh giá sản phẩm $p \in P$. Với một người dùng cần được tư vấn $j \in U$ (được gọi là người dùng hiện thời, người dùng cần được tư vấn hay người dùng tích cực),

33. Trang 22: Nó thường được áp dụng trong hệ thống gợi ý để tìm các mục tương tự, với những mục mà người dùng đã thể hiện Sự quan tâm

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quan tâm với những mục tương tự hệ thống gợi ý tìm kiếm những mục tương tự với mục mà người dùng đã đánh giá để đưa ra gợi ý sự giống nhau giữa các mục được

34. Trang 23: biểu diễn nội dung của đối tượng khuyến nghị $p \in P$ [4], được ký hiệu là Content (p), thông qua tập $|C|$ đặc trưng nội dung của P

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được Biểu diễn thông qua tập K đặc trưng nội dung của p

35. Trang 23: Tập các đặc trưng của sản phẩm P được xây dựng bằng các kỹ thuật truy vấn thông tin

Độ trùng lặp: 90%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tập các đặc trưng sản phẩm p được xây dựng bằng các kỹ thuật truy vấn thông tin

36. Trang 23: Mô hình hóa sở thích người dùng $u \in U$ gọi là hồ sơ người dùng ký hiệu là UserProfile (u) [4]

Độ trùng lặp: 68%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Mô hình hóa sở thích người dùng $u \in U$ gọi tắt là hồ sơ người dùng (User's Profile), ký hiệu UserProfile(u)

37. Trang 23: hồ sơ của người dùng U thực chất là lịch sử truy cập hoặc đánh giá của người đó đối với các đặc trưng nội dung sản phẩm

Độ trùng lặp: 95%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đặc trưng nội dung của P Tập các đặc trưng sản phẩm p được xây dựng bằng các kỹ thuật truy vấn thông tin để thực hiện mục đích dự đoán những sản phẩm khác tương tự với p
 Cho $\underline{u} = \{u, \underline{u}, , \underline{u} M\}$ là tập gồm M người dùng với mỗi người dùng u gọi ContentBasedProfile(u) là Hồ sơ người dùng u Hồ sơ của người dùng u thực chất là lịch sử truy cập hoặc đánh giá của người đó đối với các

38. Trang 23: UserProfile (u) được xây dựng bằng cách phân tích nội dung các sản phẩm mà người dùng đã từng truy nhập hoặc đánh giá dựa trên các kỹ thuật truy vấn thông tin

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: u) được xây dựng bằng cách phân tích nội dung các sản phẩm mà người dùng u đã từng truy nhập hoặc đánh giá dựa trên các kỹ thuật truy vấn thông tin

39. Trang 24: dự đoán đánh giá của người dùng u với sản phẩm p dựa trên độ tương tự nội dung của p với hồ sơ người dùng u

Độ trùng lặp: 69%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Dự đoán đánh giá của người dùng u với sản phẩm p dựa trên đánh giá trung (hoặc trọng số) của các sản phẩm tương tự Việc đo độ tương tự

40. Trang 24: Hệ thống sẽ ưu tiên tư vấn những đối tượng p có nội dung tương tự cao nhất với hồ sơ người dùng u

Độ trùng lặp: 79%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hệ thống sẽ ưu tiên khuyến nghị những đối tượng p có nội dung tương tự cao so với hồ sơ người dùng u

41. Trang 24: hệ thống không yêu cầu có quá nhiều dữ liệu từ những người dùng khác để đạt được độ chính xác đề xuất

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hệ thống không yêu cầu cần có quá nhiều dữ liệu từ những người dùng khác để đạt được độ chính xác

42. Trang 24: có thể đề xuất các nội dung mới phù hợp với sở thích của người dùng ngay cả khi người dùng chưa từng tương tác với nội dung đó, trước đây

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Có thể đề xuất nội dung từ những người dùng, mà bạn chưa từng thấy qua Do đó người dùng, Tik Tok luôn tìm thấy nội dung mới, Và, nhờ thuật toán của Tik Tok, hầu như toàn bộ nội dung mới, luôn phù hợp với sở thích của người dùng.

43. Trang 25: Hệ thống lọc công tác dựa trên việc thu thập và phân tích thông tin về hành vi, hoạt động hoặc sở thích của người dùng để dự đoán những gì họ sẽ thích dựa trên sự tương đồng với người dùng khác

Độ trùng lặp: **80%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trên việc thu thập và phân tích một lượng lớn thông tin về hành vi, hoạt động hoặc sở thích của người dùng và dự đoán những gì người dùng sẽ thích dựa trên sự tương đồng với người dùng khác Ưu điểm chính của phương pháp tiếp cận lọc cộng tác

44. Trang 26: 2 4 2 1 Lọc cộng tác dựa trên bộ nhớ (Memory based Collaborative Filtering)

Độ trùng lặp: **71%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lọc cộng tác dựa trên bộ nhớ (Memory based Collaborative Filtering)

45. Trang 26: Thuật toán Lọc cộng tác dựa trên bộ nhớ là dựa trên giá trị xếp hạng của người dùng trong ma trận người dùng Sản phẩm, hệ thống tính toán độ tương đồng giữa người dùng hiện tại với những người dùng tương tự để đưa ra đề xuất Sản phẩm, cho người dùng [1]

Độ trùng lặp: **77%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: lọc cộng tác dựa trên bộ nhớ (Memory based Collaborative Filtering) dựa trên giá trị xếp hạng của người dùng trong ma trận người dùng Sản phẩm, hệ thống tính toán độ tương đồng giữa người dùng hiện tại với những người dùng tương tự

46. Trang 26: lọc cộng tác theo bộ nhớ được chia thành hai loại chính lọc cộng tác dựa trên người dùng, và lọc cộng tác dựa trên nội dung

Độ trùng lặp: **70%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thành hai loại chính Lọc cộng tác dựa trên người dùng và Lọc cộng

47. Trang 27: Phương pháp lọc cộng tác theo bộ nhớ dựa vào người dùng sử dụng toàn bộ ma trận đánh giá để chọn ra một tập người dùng tương tự nhất với người dùng hiện tại, sau đó, kết hợp các đánh giá của tập những người dùng tương tự nhất này để dự đoán đánh giá cho người dùng hiện tại với sản phẩm chưa biết

Độ trùng lặp: **65%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: pháp lọc cộng tác theo bộ nhớ dựa vào người dùng sử dụng toàn bộ ma trận đánh giá để chọn ra một tập người dùng tương tự nhất với người dùng hiện tại Tiếp đó kết hợp các đánh giá của tập những người dùng tương tự nhất này để dự đoán đánh giá cho người dùng hiện tại với sản phẩm chưa biết

48. Trang 28: Xác định tập người dùng láng giềng gần bằng việc lựa chọn những người dùng có mức độ tương tự cao nhất với người dùng hiện tại

Độ trùng lặp: **52%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: chọn người dùng có mức độ tương tự cao nhất với

49. Trang 28: Đưa ra dự đoán của người dùng hiện tại với các sản phẩm chưa đánh giá bằng việc kết hợp tập đánh giá của các người dùng trong tập láng giềng

Độ trùng lặp: **52%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dự đoán đánh giá của bằng việc chọn ng ời với sản phẩm ch a đánh giá bằng việc kết hợp các đánh giá của các ng ời dùng trong tập láng giềng

50. Trang 28: phương pháp này thực hiện bằng cách tính toán mức độ tương tự giữa sản phẩm cần dự đoán đánh giá bởi người dùng với các sản phẩm đã được đánh giá thay vì tính mức độ tương tự giữa các người dùng trong hệ thống với người dùng hiện tại

Độ trùng lặp: **78%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phương pháp này được thực hiện bằng việc thay vì tính mức độ tương tự giữa users trong hệ thống với người dùng hiện thời , hệ tư vấn sẽ tính toán mức độ tương tự giữa sản phẩm cần dự đoán đánh giá bởi với các sản phẩm đã đ

51. Trang 28: Việc tính toán mức độ tương tự giữa hai sản phẩm dựa vào tập người dùng cùng đánh giá cả hai sản phẩm đó

Độ trùng lặp: **95%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Việc tính toán mức độ tương tự giữa hai sản phẩm được xem xét dựa vào tập người dùng cùng đánh giá cả hai sản phẩm

52. Trang 28: Sau đó chọn ra một tập sản phẩm láng giềng với sản phẩm cần dự đoán đánh giá

Độ trùng lặp: **94%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: sở đó, chọn ra một tập sản phẩm láng giềng với sản phẩm cần dự đoán đánh

53. Trang 28: Kết hợp các đánh giá của tập sản phẩm láng giềng này để đưa ra dự đoán đánh giá của sản phẩm cần dự đoán

Độ trùng lặp: **91%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: i Kết hợp các đánh giá của với tập sản phẩm láng giềng này để đưa ra dự đoán đánh giá của với sản phẩm c

54. Trang 28: 2 4 2 2 Lọc cộng tác dựa trên mô hình (Model based Collaborative Filtering)

Độ trùng lặp: **71%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lọc cộng tác dựa trên mô hình (Model based Collaborative Filtering)

55. Trang 28: Để khắc phục nhược điểm về thời gian tính toán và yêu cầu dung lượng bộ nhớ lớn, người ta đã phân biệt rõ ràng hai tác vụ chính trong phương pháp lọc cộng tác Một là tác vụ xây dựng mô hình từ dữ liệu huấn luyện, hai là tác vụ tính toán kết quả khuyến nghị từ mô hình nhận được

Độ trùng lặp: **98%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: khắc phục nhược điểm về thời gian tính toán và yêu cầu dung lượng bộ nhớ lớn, người ta đã phân biệt rõ ràng hai tác vụ chính trong phương pháp lọc công tác một là tác vụ xây dựng mô hình từ dữ liệu huấn luyện; hai là tác vụ tính toán kết quả khuyến nghị từ mô hình nhận được [

56. Trang 28: Cụ thể, hệ thống RS sử dụng các thuật toán học máy có giám sát hoặc không có giám sát để huấn luyện mô hình để rút gọn mô phỏng sở thích của người dùng

Độ trùng lặp: 95%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Cụ thể, hệ thống RS sử dụng các thuật toán học máy có giám sát hoặc không có giám sát để huấn luyện mô hình rút gọn mô phỏng sở thích của người dùng

57. Trang 28: Các thuật toán học máy có thể là cây quyết định, bộ phân loại Bayes, hồi quy, máy vector hỗ trợ (support vector machines), mạng nơron

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các thuật toán học máy có thể là cây quyết định, bộ phân loại Bayes, hồi quy, máy vector hỗ trợ (support vector machines), mạng nơron

58. Trang 28: Sau khi nhận được mô hình từ bước huấn luyện, hệ thống RS sử dụng trực tiếp mô hình này để đưa ra kết quả khuyến nghị

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Sau khi nhận được mô hình từ bước huấn luyện, hệ thống RS sử dụng trực tiếp mô hình này để đưa ra kết quả khuyến nghị

59. Trang 29: khả năng xử lý dữ liệu lớn và phức tạp. Lọc theo mô hình có thể xử lý các dữ liệu lớn và phức tạp Một cách hiệu quả

Độ trùng lặp: 61%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có Khả năng xử lý các dữ liệu lớn và phức tạp một

60. Trang 29: Lọc theo nội dung khai thác những khía cạnh liên quan đến các đặc trưng nội dung thông tin của những đối tượng cần Lọc

Độ trùng lặp: 83%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lọc theo nội dung khai thác những khía cạnh liên quan đến các đặc tr ng nội dung thông tin của những đối t ợng cần Lọc

61. Trang 29: Trái lại, lọc công tác khai thác những khía cạnh liên quan đến thói quen sử dụng các loại thông tin khác nhau của mỗi người dùng

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trái lại, lọc công tác khai thác những khía cạnh liên quan đến thói quen sử dụng các loại thông tin khác nhau của mỗi ng ời dùng

62. Trang 29: Mỗi phương pháp đều có những thế mạnh và hạn chế nhất định, do vậy để phát huy điểm mạnh và hạn chế những điểm yếu của từng kỹ thuật tư vấn riêng lẻ, các phương pháp lọc kết hợp được đưa ra nhằm cải thiện hiệu quả tư vấn sản phẩm mới phù hợp tới người dùng [2]

Độ trùng lặp: 77%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *pháp đều có những thế mạnh và hạn chế nhất định, do vậy để phát huy điểm mạnh và hạn chế những điểm yếu của từng kỹ thuật tư vấn riêng lẻ, các phương pháp lọc kết hợp được đưa ra nhằm cải thiện hiệu quả tư vấn sản phẩm mới phù hợp tới người dùng hiện thời [1][12]*

63. Trang 29: Nhiều nghiên cứu so sánh đã chứng minh rằng phương pháp kết hợp lọc cho kết quả dự đoán tốt hơn so với phương pháp lọc cộng tác và lọc nội dung thuần túy

Độ trùng lặp: 66%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Nhiều kết quả so sánh lọc kết hợp đã chứng tỏ phương pháp cho lại kết quả dự đoán tốt hơn so với các phương pháp lọc cộng tác và lọc nội dung thuần túy*

64. Trang 29: Đặc biệt, lọc kết hợp giảm thiểu tác động của vấn đề dữ liệu thưa và người dùng mới

Độ trùng lặp: 68%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Đặc biệt, lọc kết hợp hạn chế hiệu quả vấn đề dữ liệu thưa và người dùng mới*

65. Trang 29: Có trọng số (Weighted) các Điểm số của các thành phần đề xuất khác nhau được kết hợp dựa trên số lượng

Độ trùng lặp: 77%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Có trọng số (Weighted) điểm số của Các thành phần đề xuất khác nhau được kết hợp theo số lượng*

66. Trang 29: Chuyển đổi (Switching) Hệ thống chọn giữa các thành phần đề xuất và áp dụng Hệ thống đã chọn

Độ trùng lặp: 90%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Chuyển đổi (Switching) Hệ thống chọn giữa các thành phần đề xuất và áp dụng Hệ thống*

67. Trang 30: Hỗn hợp (Mixed) Các đề xuất từ những người giới thiệu khác nhau được Kết hợp để đưa ra đề xuất

Độ trùng lặp: 85%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Hỗn hợp (Mixed) Các khuyến nghị từ những người giới thiệu khác nhau được trình bày cùng nhau để đưa ra đề xuất kết hợp tính năng (Feature Combination) Các tính năng được lấy từ Các nguồn tri thức khác nhau được kết hợp*

68. Trang 30: Kết hợp tính năng (Feature Combination) Các tính năng từ Các nguồn tri thức khác

nhau được Kết hợp và đưa ra cho một thuật toán gợi ý duy nhất

Độ trùng lặp: **83%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Kết hợp tính năng (Feature Combination) Các tính năng được lấy từ Các nguồn tri thức khác nhau được Kết hợp với nhau và được đưa ra cho một thuật toán gợi ý duy nhất

69. Trang 30: Tính năng tăng cường (Feature Augmentation) Một kỹ thuật gợi ý được sử dụng để Tính toán Một Tính năng hoặc tập hợp các Tính năng sau đó trở thành Một phần của đầu vào cho kỹ thuật tiếp theo

Độ trùng lặp: **87%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tính năng tăng cường (Feature Augmentation) Một kỹ thuật gợi ý được sử dụng để Tính toán Một Tính năng hoặc tập hợp các Tính năng sau đó là Một phần của đầu vào cho kỹ thuật tiếp theo

70. Trang 30: Cascade Các khuyến nghị được ưu tiên nghiêm ngặt, với những ưu tiên thấp hơn phá vỡ Các mối quan hệ trong việc tính điểm của những người cao hơn

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Cascade Các khuyến nghị được ưu tiên nghiêm ngặt, với những ưu tiên thấp hơn phá vỡ Các mối quan hệ trong việc tính điểm của những người cao hơn

71. Trang 30: Cấp độ meta (Meta level) Một kỹ thuật đề xuất được áp dụng và tạo ra Một số loại mô hình, sau đó trở thành đầu vào được sử dụng bởi kỹ thuật tiếp theo

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Cấp độ meta (Meta level) Một kỹ thuật đề xuất được áp dụng và tạo ra Một số loại mô hình, sau đó là đầu vào được sử dụng bởi kỹ thuật tiếp theo

72. Trang 30: Khi có một điểm dữ liệu mới cần phân loại, KNN sẽ tìm ra các điểm dữ liệu gần nhất trong tập dữ liệu huấn luyện và dự đoán nhãn của điểm dữ liệu mới dựa trên đa số nhãn của các điểm gần nhất đó

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: điểm dữ liệu gần nhất trong tập dữ liệu huấn luyện trong bài toán hồi quy, đầu ra của một điểm dữ liệu sẽ bằng chính đầu ra của điểm dữ liệu đã biết gần nhất trong trường hợp $K =$ trong trường hợp $K >$, đầu ra của điểm dữ liệu sẽ là trung bình có trọng số của đầu ra của những điểm gần nhất hoặc một mối quan hệ dựa trên khoảng cách tới các điểm gần nhất đó

73. Trang 30: Nguyên lý hoạt động của Thuật toán K Nearest Neighbors (KNN) là dựa trên giả định rằng các điểm dữ liệu có Tính chất tương đồng với nhau thường nằm gần nhau trong không gian đặc trưng

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đặc trưng trong dữ liệu mới dựa trên dữ liệu huấn luyện tính xác suất của lớp dữ

liệu cho dữ liệu mới Cuối cùng, chúng ta tính xác suất của từng lớp dữ liệu cho dữ liệu mới dựa trên xác suất của lớp dữ liệu và xác suất của các đặc trưng Naive Bayes là một thuật toán đơn giản và nhanh chóng, nó có thể hoạt động tốt với dữ liệu có kích thước lớn và có nhiều lớp dữ liệu thuật toán Naive Bayes thuật toán K Nearest Neighbors (KNN) K Nearest Neighbors (KNN) là

74. Trang 31: Giai đoạn này được coi là giai đoạn quan trọng trong các bài toán về Machine Learning, đây là input cho việc xây dựng mô hình bài toán

Độ trùng lặp: 56%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: đoạn quan trọng trong các bài toán về ML vì đây là input cho việc

75. Trang 31: Lựa chọn ra những đặc trưng tốt của bộ dữ liệu, và lược bỏ những đặc trưng không tốt của dữ liệu, gây nhiễu

Độ trùng lặp: 78%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: chọn ra những đặc trưng tốt của dữ liệu lược bỏ những đặc trưng không tốt của dữ liệu gây nhiễu

76. Trang 31: Các dữ liệu không cần thiết gây khó khăn cho việc tính toán thì có thể lược bớt nhưng vẫn giữ được độ chính xác của dữ liệu

Độ trùng lặp: 64%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: gây khó khăn cho việc tính toán thì phải giảm số chiều của dữ liệu nhưng vẫn giữ được độ chính xác của dữ liệu (reduce demension) Ở bước này cũng cần

77. Trang 31: Thông thường sẽ sử dụng cross validation để chia tập dataset thành 2 phần một phần dữ liệu phục vụ cho quá trình training và phần dữ liệu còn lại sử dụng cho mục đích test mô hình

Độ trùng lặp: 63%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Thông thường sẽ sử dụng cross validation (kiểm tra chéo) để chia tập dataset thành hai phần, một phần, phục vụ cho training và phần, còn lại phục vụ cho mục đích testing

78. Trang 32: cuối cùng chúng ta sẽ Đánh giá mô hình bằng cách kiểm tra mức độ lỗi của dữ liệu testing và dữ liệu training thông qua mô hình đã tìm được

Độ trùng lặp: 86%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Cuối cùng, chúng ta sẽ đánh giá mô hình bằng cách đánh giá mức độ lỗi của dữ liệu testing và dữ liệu training thông qua mô hình tìm được

79. Trang 32: Nếu không đạt kết quả mong muốn chúng ta cần thay đổi các tham số của thuật toán học để tìm ra các mô hình tốt hơn sau đó kiểm tra, và đánh giá lại mô hình phân lớp và cuối cùng chọn ra mô hình phân lớp tốt nhất cho bài toán của chúng ta

Độ trùng lặp: 84%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: kết quả mong muốn, của chúng ta thì phải thay đổi các tham số của thuật toán học để tìm ra các mô hình tốt hơn, và kiểm tra đánh giá lại mô hình phân lớp và cuối cùng chọn ra mô hình phân lớp tốt nhất cho bài toán của chúng

80. Trang 34: singular value Decomposition (SVD) một phương pháp từ đại số tuyến tính thường được sử dụng như một kỹ thuật giảm kích thước trong học máy, là một trong những phương pháp thuộc nhóm matrix factorization được phát triển lần đầu bởi những nhà hình học vi phân

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một trong những phương pháp thuộc nhóm matrix factorization được phát triển lần đầu bởi những nhà hình học vi phân

81. Trang 34: Ban đầu mục đích của phương pháp này là tìm ra một phép xoay không gian sao cho tích vô hướng của các vector không thay đổi

Độ trùng lặp: 85%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ban đầu mục đích của phương pháp này là tìm ra một phép xoay không gian sao cho tích vô hướng của các vector không thay đổi

82. Trang 34: Từ mối liên hệ này khái niệm về ma trận trực giao đã hình thành để tạo ra các phép xoay đặc biệt

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Từ mối liên hệ này, khái niệm về ma trận trực giao đã hình thành để tạo ra các phép xoay đặc biệt

83. Trang 34: Phương pháp SVD đã được phát triển dựa trên những tính chất của ma trận trực giao và ma trận đường chéo để tìm ra một ma trận xấp xỉ với ma trận gốc

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phương pháp SVD đã được phát triển dựa trên những tính chất của ma trận trực giao và ma trận đường chéo để tìm ra một ma trận xấp xỉ với ma trận gốc

84. Trang 34: Phương pháp này sau đó đã được ứng dụng rộng rãi trong các lĩnh vực như hình học vi phân, hồi quy tuyến tính, xử lý hình ảnh, clustering, các thuật toán nén và giảm chiều dữ liệu, và đặc biệt hiệu quả trong các bài toán recommendation

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phương pháp này sau đó đã được ứng dụng rộng rãi trong các lĩnh vực như hình học vi phân, hồi quy tuyến tính, xử lý hình ảnh, clustering, các thuật toán nén và giảm chiều dữ liệu, và đặc biệt hiệu quả trong các bài toán recommendation

85. Trang 34: SVD là phương pháp giúp phân rã bất cứ ma trận nào thành tích của 3 ma trận với tính chất đặc biệt [7]

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: SVD là phương pháp giúp phân rã bất cứ ma trận nào thành tích của ma trận với tính chất đặc biệt

86. Trang 34: S là ma trận đường chéo $n \times m$ chứa các giá trị riêng của ma trận A

Độ trùng lặp: 90%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ma trận đường chéo $n \times m$ chứa các giá trị riêng của ma trận m V là ma trận

87. Trang 34: V là ma trận $n \times n$ vec tơ riêng trái và V cũng là ma trận trực giao

Độ trùng lặp: 90%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: vec tơ riêng trái và V cũng là ma trận trực giao

88. Trang 35: Trường hợp $m = n$ có thể xếp vào một trong hai Trường hợp trên

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có thể xếp vào một trong hai Trường hợp trên

89. Trang 35: là một ma trận đường chéo với các phần tử trên đó giảm dần và không âm

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một ma trận đường chéo với các phần tử trên đó giảm dần và không âm

90. Trang 35: sử dụng, SVD trên ma trận xếp hạng với các hàng đại diện cho người dùng và các cột đại diện cho phim

Độ trùng lặp: 62%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các cột đại diện cho các mục, các hàng đại diện cho người Sử dụng

91. Trang 36: SVD có thể được áp dụng trong nhiều lĩnh vực khác nhau như xử lý ảnh, xử lý ngôn ngữ tự nhiên và hệ thống gợi ý

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dùng trong nhiều lĩnh vực khác nhau như xử lý ảnh, xử lý ngôn ngữ tự nhiên đồng thời cũng được

92. Trang 37: độ tương tự Cosine similarity là một cách đo độ tương tự (measure of similarity) giữa hai vectơ khác không của một không gian tích vô hướng

Độ trùng lặp: 89%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một cách đo Độ tương tự (measure of similarity) giữa hai vectơ khác không của một không gian tích vô hướng

93. Trang 37: Độ tương tự này được định nghĩa bằng giá trị cosine của góc giữa hai vectơ, và cũng là tích vô hướng của cùng các vectơ, đơn vị để cả hai đều có chiều dài 1

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tương tự này được định nghĩa bằng giá trị cosine của góc giữa hai vectơ, và cũng là tích vô hướng của cùng các vectơ, đơn vị để cả hai đều có chiều dài 1 giá trị cosine của 0° là 1 và bé hơn 1 với bất kỳ góc nào trong khoảng các radian

94. Trang 37: Giá trị cosine của 0° là 1, và bé hơn 1, với bất kỳ góc nào trong khoảng các radian $(0, \pi]$

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giá trị cosine của 0° là 1, và bé hơn 1, với bất kỳ góc nào trong khoảng các radian

95. Trang 38: Cấu trúc của Python cho phép người sử dụng viết mã lệnh với số lần nhập phím tối thiểu

Độ trùng lặp: 83%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu

96. Trang 38: vì vậy so với các ngôn ngữ lập trình khác, chúng ta có thể sử dụng ít dòng mã hơn để viết ra một chương trình

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có thể sử dụng ít dòng mã hơn để viết một chương trình

97. Trang 38: Python có sẵn trên nhiều nền tảng, bao gồm Windows, macOS, Linux, và các hệ thống UNIX khác nhau

Độ trùng lặp: 57%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có sẵn trên nhiều nền tảng, bao gồm Windows, MacOS, Linux Android và

98. Trang 39: Nó Cung cấp các cấu trúc dữ liệu và công cụ phân tích dữ liệu mạnh mẽ làm cho nó trở thành một công cụ có giá trị để làm việc với dữ liệu có cấu trúc

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: công cụ để phân tích dữ liệu và xử lý các tập dữ liệu lớn Pandas cung cấp các cấu trúc dữ liệu linh hoạt để xử lý và phân tích dữ liệu đặc biệt là dữ liệu có cấu trúc như bảng có tính năng xử lý dữ liệu bị thiếu, hỗ trợ các phép toán theo nhóm và phân tích thống kê dữ liệu Được tích hợp tốt với NumPy và các thư viện phân tích dữ liệu khác như Matplotlib và Scikit learn Matplotlib Là một thư viện trực quan hóa dữ liệu mạnh mẽ.

99. Trang 40: trực quan hóa dữ liệu Mặc dù Pandas chủ yếu Là một thư viện thao tác dữ liệu nhưng nó tích hợp tốt với các thư viện khác như Matplotlib và Seaborn

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tích dữ liệu khác như Matplotlib và Scikit learn Matplotlib là một thư viện Trực quan hóa dữ liệu

100. Trang 40: SciPy gồm các gói con (submodule) cho đại số tuyến tính, tối ưu hóa, tích hợp và thống kê

Độ trùng lặp: **61%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cho đại số tuyến tính, tích hợp và thống kê

101. Trang 42: thư viện tkinter là một thư viện trong Python được sử dụng để tạo giao diện người dùng đồ họa (GUI)

Độ trùng lặp: **60%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được sử dụng để tạo giao diện người dùng

102. Trang 43: Visual Studio Code hỗ trợ đa dạng các chức năng debug, đi kèm với Git, có syntax highlighting.

Độ trùng lặp: **76%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Visual Studio Code hỗ trợ chức năng Debug, đi kèm với Git, có Syntax Highlighting

103. Trang 44: bộ dữ liệu MovieLens 100k là một bộ dữ liệu phổ biến được sử dụng trong nghiên cứu khoa học, bao gồm 100 000 xếp hạng từ 943 người dùng cho 1682 bộ phim với mỗi xếp hạng từ 1 đến 5

Độ trùng lặp: **54%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 100K là Bộ dữ liệu gồm 100 000 xếp hạng phim, từ 943 người dùng cho 1 682 Bộ phim, được phát hành vào tháng 04/1998 Bộ dữ liệu ml 1M là Bộ dữ liệu gồm 1 000 209 xếp hạng từ 6 040 người dùng cho 3 900 Bộ phim, được phát hành vào tháng 02/2003 Bộ dữ liệu ml 10M là Bộ dữ liệu gồm 10 000 054 xếp hạng từ 71 567 người dùng cho 10 681 Bộ phim, được phát hành vào tháng 02/2003 Bộ dữ liệu ml 20M là Bộ dữ liệu gồm 20 000 263 xếp hạng từ 138 493 người dùng cho 27 278 Bộ phim, được phát hành vào tháng 10/2016 Bộ dữ liệu ml 25M là Bộ dữ liệu gồm 25 000 095 xếp hạng từ 162 541 người dùng cho 62 423 Bộ phim, được phát hành vào tháng 10/2019 Bộ dữ liệu Movielens 1B Synthetic Dataset Đây là Bộ dữ liệu sử dụng

104. Trang 44: dữ liệu được thu thập thông qua trang web MovieLens (movielens umn.edu) trong khoảng thời gian bảy tháng, từ ngày 19 tháng. 9 năm 1997 đến ngày 22 tháng. 4 năm 1998

Độ trùng lặp: **66%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được thu thập trong khoảng thời gian bảy tháng kể từ ngày 19 tháng 9 năm 1997

đến ngày 22 tháng 4 năm 1998

105. Trang 44: dữ liệu này đã được lọc để loại bỏ các người dùng có ít hơn 20 xếp hạng hoặc thiếu thông tin cá nhân đầy đủ

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: người dùng có ít hơn 20 xếp hạng hoặc thông tin nhân

106. Trang 62: [3] N T Nghe "Chương 2 Hệ thống gợi ý Kỹ thuật và ứng dụng Cần Thơ Khoa Công nghệ thông tin và Truyền thông Đại học Cần Thơ," [Online]

Độ trùng lặp: 73%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chương Hệ thống gợi ý Kỹ thuật và ứng dụng, Nguyễn Thái Nghe, Khoa công nghệ thông tin và truyền thông Trường Đại học Cần Thơ Email ntngh@

107. Trang 62: TÍN "Phát TRIỂN MỘT SỐ PHƯƠNG PHÁP KHUYẾN NGHỊ HỖ TRỢ TÌM KIẾM THÔNG TIN HỌC THUẬT DỰA TRÊN TIẾP CẬN PHÂN TÍCH MẠNG XÃ HỘI," 2016

Độ trùng lặp: 96%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: triển một số phương pháp khuyến nghị hỗ trợ tìm kiếm thông tin học thuật dựa trên tiếp cận phân tích mạng xã HỘI

108. Trang 62: TRƯƠNG "Nghiên CỨU VÀ ỨNG DỤNG KỸ THUẬT SVD VÀO HỆ THỐNG GỢI Ý," 2018

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cứu và ứng dụng kỹ thuật SVD vào hệ thống gợi Ý

--- Hết ---