



BÁO CÁO KIỂM TRA TRÙNG LẬP

Thông tin tài liệu

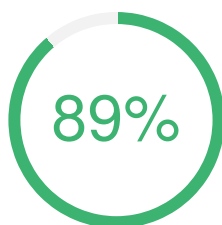
Tên tài liệu:	2051063432_LuongQuangTruong
Tác giả:	Trương Xuân Nam
Điểm trùng lặp:	11
Thời gian tải lên:	16:28 15/07/2024
Thời gian sinh báo cáo:	16:33 15/07/2024
Các trang kiểm tra:	60/60 trang



Kết quả kiểm tra trùng lặp



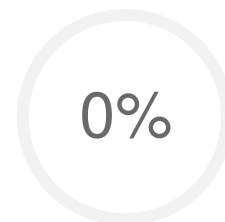
Có 11% nội dung trùng
lặp



Có 89% 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 xaydungso.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 5: Sinh viên đã hoàn thành và nộp bản Đồ án tốt nghiệp cho Hội đồng thi

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

3. Trang 6: Suy tim là một tình trạng y tế nghiêm trọng, trong đó tim không hoạt động hiệu quả để cung cấp đủ máu cho cơ thể

Độ trùng lặp: **73%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một tình trạng y tế nghiêm trọng, phức tạp, trong đó tim không thể hoạt động hiệu quả để

4. Trang 6: dù đã có sự tiến bộ trong chuẩn đoán và điều trị, nhưng suy tim vẫn là nguyên nhân hàng đầu gây ra tử vong và suy giảm chất lượng cuộc sống cho nhiều người

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Dù đã có nhiều tiến bộ trong chẩn đoán và điều trị, tuy nhiên di căn xa vẫn là nguyên nhân gây tàn phế, tử vong và giảm chất lượng cuộc sống cho

5. Trang 6: Ngoài ra, em có sử dụng Các phương pháp Phân Tích và Tiền xử lý dữ liệu Các phương pháp đánh giá mô hình

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tiền xử lý dữ liệu, Làm sạch dữ liệu, tích hợp và chuyển đổi dữ liệu, Rút gọn dữ liệu, phân cụm Định nghĩa Đặc điểm các ứng dụng của phân cụm dữ liệu, các phương pháp phân cụm phân lớp Định nghĩa Quá trình phân lớp dữ liệu, các vấn đề liên quan đến phân lớp dữ liệu, Một số phương pháp phân lớp các phương pháp đánh giá

6. Trang 8: Tác giả xin cam đoan đây là Đồ án tốt nghiệp của bản thân Tác giả

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tác giả xin cam đoan đây là Đồ án tốt nghiệp của bản thân Tác giả

7. Trang 8: Các kết quả trong Đồ án tốt nghiệp này là trung thực, và không sao chép từ bất kỳ một nguồn nào và dưới bất kỳ hình thức nào Việc tham khảo Các nguồn tài liệu (nếu có) đã được thực, hiện trích dẫn và ghi nguồn tài liệu tham khảo đúng quy định

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các kết quả trong Đồ án tốt nghiệp này là trung thực, và không sao chép từ bất kỳ một nguồn nào và dưới bất kỳ hình thức nào Việc tham khảo Các nguồn tài liệu (nếu có) đã được thực, hiện trích dẫn và ghi nguồn tài liệu tham khảo đúng quy định

8. Trang 9: em xin chân thành cảm ơn khoa Công nghệ thông tin, trường Đại học Thủy Lợi đã tạo điều kiện cho em thực hiện đồ án tốt nghiệp

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Em xin chân thành cảm ơn các thầy, cô trong Khoa Công Nghệ Thông Tin, trường Đại Học Thủy Lợi đã tạo điều kiện thuận Lợi cho Em trong quá trình Học tập tại trường cũng như trong thời gian thực hiện Đồ án tốt nghiệp

9. Trang 9: em xin gửi lời cảm ơn chân thành đến tất cả các Thầy Cô đã giảng dạy chúng em trong suốt thời gian qua

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Em xin gửi lời cảm ơn chân thành đến tất cả các thầy, cô đã giảng dạy chúng Em trong suốt thời gian qua

10. Trang 9: em cảm ơn thầy đã giúp đỡ, bổ sung cho em những kiến thức và cho em những lời khuyên, gợi ý để em có thể hoàn thành đồ án một cách nhanh chóng và hiệu quả nhất

Độ trùng lặp: 87%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cảm ơn thầy đã giúp đỡ, bổ sung cho Em những kiến thức cơ bản, và cho Em những lời khuyên, sự góp ý để Em có thể hoàn thành đồ án một cách nhanh chóng và có hiệu quả nhất

11. Trang 9: trong quá trình học tập và thực hiện đồ án em đã may mắn được sự chỉ bảo hướng dẫn tận tình của các thầy cô giáo và được gia đình bạn bè quan tâm, động viên luôn ở bên và tạo mọi điều kiện thuận Lợi để hoàn thành tốt đồ án này

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được sự chỉ bảo, và giúp đỡ của các thầy cô giáo và các bạn rất nhiều Trong lĩnh vực học tập và cuộc sống Đầu tiên, em xin chân thành cảm ơn các thầy cô giáo Trường Đại học Thủy Lợi và đặc biệt là các thầy cô giáo khoa Công nghệ thương tin đã dạy cho em có được những kiến thức để phục vụ cho việc thực hiện đồ án Đặc biệt, Trong tuần làm đồ án em đã được sự hướng dẫn nhiệt tình của Tiến sĩ, Giảng viên, BM Kỹ thuật Máy tính và Mạng Đồ Trường Xuân, cùng với các thầy cô giáo Khoa Công Nghệ thông tin và Tiến sĩ em xin gửi lời cảm ơn chân thành tới thầy cô đã giúp đỡ, bổ sung cho em những kiến thức, cho em những lời khuyên và sự góp ý để em có thể hồn thành đồ án một cách nhanh chóng và hiệu quả nhất Trong suốt thời gian học tập và hoàn thành đồ án em đã may mắn được thầy cơ chỉ bảo, dìu dắt và được gia đình, bạn bè quan tâm, động viên, luôn bên cạnh và tạo mọi điều kiện thuận lợi để cho em có thể hồn thành đồ án này Trong suốt q trình

12. Trang 9: Một lần nữa, em xin chân thành cảm ơn thầy cô giáo, bạn bè và gia đình, những người đã giúp đỡ, ủng hộ em trong thời gian vừa qua

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thầy cô giáo, gia đình, bạn bè và những người đã giúp đỡ, ủng hộ em trong thời gian vừa qua

13. Trang 10: x 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

14. Trang 13: 43 DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ TH trường hợp

Độ trùng lặp: **75%**

Nguồn: *Dữ liệu nội sinh*

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

15. Trang 15: Suy tim là một hội chứng lâm sàng phức tạp xảy ra khi tim không thể bơm đủ máu để đáp ứng nhu cầu của cơ thể

Độ trùng lặp: **88%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: lâm sàng phức tạp, xảy ra khi tim không thể bơm đủ máu để đáp ứng nhu cầu của cơ thể

16. Trang 15: Việc chuẩn đoán và điều trị suy tim sớm là rất quan trọng để cải thiện tiên lượng bệnh và giảm nguy cơ tử vong

Độ trùng lặp: **54%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là rất quan trọng để cải thiện tiên lượng bệnh và giảm tỉ lệ tử vong Người lớn tuổi là đối tượng có cơ địa đặc biệt (nhiều bệnh nền, nằm viện lâu, suy

17. Trang 16: Đề tài này hoàn toàn phù hợp với chuyên ngành được đào tạo

Độ trùng lặp: **91%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đề tài này hoàn toàn phù hợp với chuyên ngành mà tác giả đang được đào tạo

18. Trang 16: Nó giúp áp dụng những kiến thức, và kỹ năng đã được học vào việc giải quyết một vấn đề thực tiễn có ý nghĩa xã hội

Độ trùng lặp: **56%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dụng những kiến thức kỹ năng đã được học vào việc giải quyết những vấn đề thực tiễn >>

19. Trang 17: Kết quả của mô hình sẽ được đánh giá dựa trên các chỉ số hiệu suất như độ chính

xác độ nhảy,

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của mô hình được đánh giá dựa trên bốn chỉ số (phần 3 2), cụ thể là độ chính xác, (7), độ nhảy (8), PPV (9) và NPV (10) như Bảng 15 Bảng 15 Tóm tắt các chỉ số hiệu suất của thuật toán GBDT cho tất cả các hành vi các chỉ số độ chính xác, độ nhảy

20. Trang 19: học máy (machine learning) là một lĩnh vực của trí tuệ nhân tạo (AI Artificial Intelligence) tập trung vào việc phát triển các thuật toán và mô hình cho phép máy tính tự học từ dữ liệu và cải thiện hiệu suất của chúng trong việc thực hiện các nhiệm vụ cụ thể, mà không cần được lập trình rõ ràng

Độ trùng lặp: **66%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một lĩnh vực của trí tuệ nhân tạo (AI) tập trung vào việc phát triển các thuật toán và mô hình để giúp máy tính tự động Học hỏi từ dữ liệu mà không cần được lập trình trực tiếp Thay vì chỉ định rõ từng bước để giải quyết một vấn đề cụ thể machine learning cho phép máy tính tự động Học hỏi từ dữ liệu đầu vào và cải thiện hiệu suất của

21. Trang 19: học máy có ba loại chính học có giám sát (supervised learning), học không giám sát (unsupervised learning), và học tăng cường (reinforcement learning).

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ba loại chính Học có giám sát (supervised learning), Học không có sự giám sát (unsupervised learning), và Học tăng cường (reinforcement learning).

22. Trang 19: Trong Học có giám sát mô hình được huấn luyện trên một tập dữ liệu bao gồm các đầu vào và các đầu ra mong muốn từ đó Học cách ánh xạ các đầu vào đến các đầu ra một cách chính xác

Độ trùng lặp: **65%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dữ liệu huấn luyện được sử dụng Ba nhóm chính này là học có giám sát, (Supervised Learning) Trong học có giám sát, mô hình machine learning được huấn luyện trên một tập dữ liệu huấn luyện đã được gán nhãn, tức là có đầu ra mong muốn, cho mỗi điểm dữ liệu đầu vào Mục tiêu của học có giám sát, là xây dựng một mô hình dự đoán chính xác các đầu ra

23. Trang 19: các thuật toán phổ biến trong học có giám sát bao gồm hồi quy tuyến tính, hồi quy logistic, cây quyết định, rừng ngẫu nhiên

Độ trùng lặp: **66%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: học có giám sát bao gồm mạng thần kinh, bayes ngây thơ, hồi quy tuyến tính, hồi quy logistic, rừng ngẫu nhiên.

24. Trang 19: trong Học máy không giám sát mô hình được huấn luyện trên một tập dữ liệu chỉ bao gồm các đầu vào mà không có các đầu ra mong muốn

Độ trùng lặp: **64%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: học có giám sát, học có giám sát, là một phương pháp học máy Trong trí tuệ nhân tạo, Trong đó một thuật toán máy tính được huấn luyện trên một tập dữ liệu

25. Trang 19: Mục tiêu của học máy không giám sát là khám phá ra các cấu trúc hoặc mẫu ẩn trong dữ liệu

Độ trùng lặp: 65%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Mục tiêu của học không giám sát là khám phá cấu trúc ẩn trong dữ liệu và học các đặc trưng tự động từ tập dữ liệu mà không cần thông tin về đầu ra mong muốn trong quá trình huấn luyện, thuật toán học không giám sát sẽ tự động tìm kiếm các mẫu

26. Trang 19: học tăng cường là một phương pháp học mà trong đó một tác nhân (agent) học cách hành động trong một môi trường bằng cách thực hiện các hành động và nhận phần thưởng hoặc hình phạt

Độ trùng lặp: 52%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: thực hiện các hành động nào để cực đại hóa một khoản thưởng (reward) nào đó về lâu dài các thuật toán Học tăng cường cố gắng tìm một chiến lược ánh xạ các trạng thái của thế giới tới các hành động mà agent nên chọn trong các trạng thái đó [1] RL là một loại máy Học cho phép chúng tôi tạo ra các tác nhân AI Học hỏi từ môi trường bằng cách tương tác với nó để tối đa hóa phần thưởng tích lũy của nó Giống như cách chúng ta Học cách đi xe đạp, AI Học nó bằng cách thử và sai, các tác nhân trong

27. Trang 19: Mục tiêu của tác nhân là tối đa hóa tổng phần thưởng nhận được theo thời gian

Độ trùng lặp: 81%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Mục tiêu của tác nhân là tối đa hóa tổng phần thưởng mà nó nhận được trong một thời gian

28. Trang 19: Học tăng cường thường được áp dụng trong các lĩnh vực thực như chơi game điều khiển robot,

Độ trùng lặp: 72%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Học tăng cường thường được áp dụng trong các lĩnh vực như điều khiển robot, chơi game, và quyết định trong thời gian thực

29. Trang 20: Trong tài chính, học máy có thể được sử dụng để phát hiện các sai phạm, dự báo thị trường hoặc quản lý rủi ro

Độ trùng lặp: 58%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: có thể được sử dụng để dự

30. Trang 20: Với sự tiến bộ không ngừng của công nghệ, và khả năng tiếp cận dữ liệu, học máy hứa hẹn sẽ tiếp tục mang lại những phát minh và ứng dụng đột phá trong Tương lai

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Với sự tiến bộ không ngừng của công nghệ*

31. Trang 21: Học máy có giám sát, (Supervised Learning) là một kỹ thuật trong lĩnh vực trí tuệ nhân tạo nơi máy tính được huấn luyện để đưa ra dự đoán hoặc quyết định dựa trên dữ liệu đã được gắn nhãn.

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quyết định dựa trên các mẫu trong dữ liệu Học máy hoạt động như thế nào trong Học máy máy tính được cấp quyền truy cập một lượng lớn dữ liệu và được đào tạo để xác định các mẫu trong dữ liệu đó Những mẫu này có thể được dùng để đưa ra dự đoán hoặc quyết định về dữ liệu mới mà máy tính chưa từng thấy trước đây có ba loại Học máy chính Học có giám sát Học không giám sát và Học tăng cường trong Học có giám sát máy tính được cung cấp dữ liệu được dán nhãn trong đó mỗi điểm dữ liệu đã được phân loại máy tính sau đó được đào tạo để đưa ra dự đoán về dữ liệu mới dựa trên các mẫu mà nó đã Học được từ dữ liệu có dán nhãn trong Học không giám sát máy tính được cung cấp dữ liệu không được gắn nhãn

32. Trang 21: Sau khi mô hình được huấn luyện, nó sẽ được đưa vào quá trình kiểm thử dựa trên một tập dữ liệu chưa từng được đưa vào trong quá trình huấn luyện.

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quá trình huấn luyện, dữ liệu XTestHoldout được đưa vào từng mô hình được huấn luyện, tốt nhất ở trên để thực thi bước kiểm tra Kết quả của quá trình kiểm tra sẽ trả về là dữ liệu dự đoán yPredHoldout Sau đó, nó

33. Trang 21: các chỉ số đánh giá như độ chính xác (accuracy), độ nhạy (recall), thường được sử dụng để đánh giá hiệu suất của mô hình

Độ trùng lặp: 72%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các chỉ số đánh giá hiệu suất của mô hình phân loại như độ chính xác (accuracy), độ nhạy (recall), độ chính xác dương tính (precision), và độ chính xác âm tính (specificity) Những chỉ số này giúp ta hiểu được

34. Trang 22: phân cụm, nhóm các điểm dữ liệu, thành các cụm, sao cho các điểm trong cụm, có các đặc điểm giống nhau hơn so với cụm, khác

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thành các cụm sao cho các điểm trong một cụm tương tự/tương đồng nhau và ít giống với các điểm trong các cụm khác

35. Trang 24: Quá trình này bao gồm các việc như làm sạch dữ liệu Xử lý giá trị thiếu, chuẩn hóa dữ liệu,

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trình làm sạch dữ liệu, xử lý giá trị thiếu, Trong trường hợp dữ liệu, có rất nhiều bản ghi có các giá trị thiếu, các phương pháp sau xử lý các giá trị thiếu,

36. Trang 24: Thông thường, khoảng 70 80% dữ liệu được đưa vào huấn luyện và phần còn lại được sử dụng để kiểm tra và xác thực

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: phần được sử dụng để huấn luyện và phần còn lại được sử dụng để kiểm

37. Trang 25: Sau khi huấn luyện, mô hình được đánh giá bằng cách sử dụng tập kiểm tra để xác định hiệu suất

Độ trùng lặp: 61%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được đánh giá bằng cách sử dụng tập kiểm tra

38. Trang 25: các chỉ số đánh giá ở đây có thể là độ chính xác độ nhạy F1 score, giúp chúng ta có thể có cái nhìn toàn diện về khả năng dự đoán của mô hình

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: về khả năng dự đoán của mô hình

39. Trang 25: trong học máy dữ liệu, là nền tảng và là yếu tố cốt lõi giúp các mô hình học máy có thể học và đưa ra dự đoán

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các mô hình học máy, có thể học

40. Trang 25: dữ liệu có thể được phân thành hai loại chính dữ liệu có cấu trúc và dữ liệu phi cấu trúc

Độ trùng lặp: 83%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thành hai loại chính Dữ liệu có cấu trúc (như bảng biểu, dự án) và Dữ liệu phi cấu trúc (

41. Trang 25: dữ liệu có cấu trúc là dữ liệu được tổ chức theo một cấu trúc rõ ràng và có thể lưu trữ và truy xuất trong các hệ thống cơ sở dữ liệu

Độ trùng lặp: 61%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: lưu trữ trong các hệ quản trị cơ sở Dữ liệu quan hệ (Relational Database Management System) và được truy xuất bằng ngôn ngữ truy vấn có cấu trúc (Structured Query Language) Dữ liệu phi cấu trúc (Unstructured data) Đây là dạng Dữ liệu không có cấu trúc rõ ràng hoặc không được tổ chức theo một mô hình truyền thống

42. Trang 25: Ngược lại, dữ liệu phi cấu trúc không có cấu trúc rõ ràng và không dễ dàng để lưu trữ trong các cơ sở dữ liệu truyền thống

Độ trùng lặp: 59%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: để lưu trữ dữ liệu đó Ngược lại, dữ liệu có cấu trúc có quy trình lưu trữ hợp lý hơn dữ liệu có cấu trúc và phi cấu trúc thường được lưu trữ trong các môi trường khác nhau, kho dữ liệu và hồ dữ liệu Kho dữ liệu dữ liệu có cấu trúc

43. Trang 27: Kiểu dữ liệu (Data Type) Các đặc trưng thường có một Kiểu dữ liệu xác định có thể là số nguyên, số thực, hoặc Kiểu rời rạc

Độ trùng lặp: 69%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Kiểu dữ liệu (data type) Các đặc trưng đều có dữ liệu mỗi dữ liệu đều có một Kiểu dữ liệu xác định. Chúng có thể là Kiểu số nguyên, số thực, hay cũng có thể là Kiểu rời rạc

44. Trang 27: với các tập dữ liệu nhỏ, chúng ta có thể sử dụng kỹ thuật như k fold để có thể đào tạo được mô hình tốt hơn

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Với các tập dữ liệu nhỏ, chúng ta có thể sử dụng các thuật toán so sánh để thiết kế

45. Trang 27: tập dữ liệu huấn luyện (Training dataset) tập dữ liệu huấn luyện là tập dữ liệu Được sử dụng với Mục tiêu đào tạo cho mô hình học máy

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dataset) Tập dữ liệu huấn luyện được sử dụng để huấn luyện mô hình học máy Tập dữ liệu

46. Trang 28: Chọn nguồn dữ liệu người dùng có thể chọn dữ liệu từ các nguồn khác nhau như các cơ sở dữ liệu các tập dữ liệu mở

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dữ liệu từ các nguồn khác nhau như tệp tin văn bản, Excel và các cơ sở dữ liệu quan hệ khác Data Editor cho phép Người dùng

47. Trang 28: Tập dữ liệu kiểm tra (Testing dataset) Testing dataset là Tập dữ liệu được sử dụng để đánh giá hiệu suất của mô hình học máy sau khi đã được huấn luyện bằng Tập dữ liệu dành cho việc đào tạo mô hình

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tập dữ liệu kiểm tra (Test Data) Đây là một phần khác của data set được sử dụng để kiểm tra hiệu suất của mô hình học máy sau khi nó đã được đào tạo dữ liệu kiểm tra thường không được sử dụng trong quá trình đào tạo và mô hình không biết trước kết quả của các mẫu này dữ liệu kiểm tra giúp đánh giá

48. Trang 28: tập dữ liệu kiểm tra được sử dụng để xác định xem mô hình có khả năng khái quát hóa tốt cho các dữ liệu mới hay không

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tập dữ liệu kiểm tra (test data set) Nếu độ chính xác là chấp nhận được (tức là cao), mô hình sẽ được sử dụng để xác định nhãn lớp cho các dữ liệu khác mới trong tương lai
Bước 2 (sử dụng mô hình) sử dụng mô hình đã được xây dựng ở bước 1 để phân lớp dữ liệu mới

49. Trang 29: Accuracy (độ chính xác) được tính bằng tỉ lệ số lượng dự đoán đúng so với tổng số lượng dự đoán

Độ trùng lặp: **60%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Độ nhạy Được tính bằng tỉ lệ giữa số lượng dự đoán đúng positive (true positive) và tổng số

50. Trang 30: Khi cần phân loại, một điểm dữ liệu mới KNN sẽ xem xét K điểm dữ liệu gần nhất
Trong tập huấn luyện và dựa vào đó để đưa ra quyết định.

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: phân loại một điểm dữ liệu mới, sẽ được phân loại trực tiếp từ K điểm dữ liệu gần nhất trong tập dữ liệu huấn luyện

51. Trang 30: Hơn nữa, việc tính toán khoảng cách từ tất cả các điểm dữ liệu trong tập huấn luyện có thể tốn kém về mặt tính toán nhất là khi số lượng điểm dữ liệu lớn hoặc số chiều của một điểm dữ liệu cao

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của một điểm dữ liệu mới bằng cách dựa vào thông tin của k điểm dữ liệu trong tập huấn luyện gần nó nhất không quan tâm đến việc các điểm dữ liệu

52. Trang 30: hồi quy tuyến tính là một mô hình học máy đơn giản dựa vào thống kê để hồi quy dữ liệu với các biến phụ thuộc với các giá trị liên tiếp trong khi các biến độc lập có thể là một trong hai giá trị liên tục hoặc đơn lẻ

Độ trùng lặp: **66%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thống kê để Hồi quy dữ liệu với biến phụ thuộc có giá trị liên tục trong khi các biến độc lập có thể có một trong hai giá trị liên tục hoặc là giá trị

53. Trang 31: Chọn K láng giềng gần nhất Sau khi Sắp xếp các điểm trong tập dữ liệu theo thứ tự tăng dần của khoảng cách đã
Tính thuật toán chọn ra K điểm dữ liệu gần nhất với điểm cần xét

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của K điểm dữ liệu gần nhất trong tập huấn luyện Nội dung thuật toán Mô tả thuật toán Xác định giá trị tham số K (số láng giềng gần nhất) tính, khoảng cách giữa điểm truy vấn phân lớp với tất cả các đối tượng trong tập dữ liệu huấn luyện sắp xếp khoảng cách theo thứ tự tăng dần và xác định K láng giềng gần nhất với điểm

54. Trang 32: Nó được sử dụng rộng rãi trong nhiều lĩnh vực, như y học, kinh tế sinh học, và khoa học, xã hội, để dự đoán khả năng xảy ra của một sự kiện

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được sử dụng rộng rãi trong nhiều lĩnh vực bao gồm y học, y tế, công cộng, kinh tế, học, khoa học, xã hội khoa học, môi trường, khoa học, dữ liệu và nghiên cứu thị trường Nó cung cấp một cách tiếp cận thống kê để đánh giá khả năng sống sót hoặc thời gian đến sự kiện trong một tập dữ liệu, và đưa ra các dự đoán về khả năng sống sót hoặc thời gian sống của một

55. Trang 33: hàm sigmoid đề cập đến một đường cong chữ S chuyển đổi bất kỳ giá trị thực nào thành phạm vi từ 0 đến 1

Độ trùng lặp: 62%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: chuyển đổi bất kỳ giá trị thực nào thành một phạm vi từ 0 đến 1 được biết đến với tên gọi là Hàm sigmoid

56. Trang 33: Hàm này đo lường sự khác biệt giữa giá trị dự đoán của mô hình và giá trị thực tế

Độ trùng lặp: 78%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đo lường sự khác biệt giữa giá trị dự đoán và giá trị thực tế Hiệu suất của mô hình

57. Trang 34: cây quyết định, (Decision Tree) là một trong những kỹ thuật học máy phổ biến được sử dụng trong cả phân loại và hồi quy

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: phân loại và hồi quy trong

58. Trang 35: Ý tưởng chính của cây quyết định là xây dựng một Mô hình dạng cây để dự đoán giá trị của các biến mục tiêu dựa trên các thuộc tính của dữ liệu đầu vào

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: xây dựng một cây quyết định để phân loại hoặc dự đoán giá trị đầu ra dựa trên các thuộc tính của dữ liệu đầu vào mô hình

59. Trang 35: mỗi nút trong cây đại diện cho một thuộc tính của dữ liệu, mỗi nhánh đại diện cho một giá trị của thuộc tính đó, và mỗi lá đại diện cho một giá trị dự đoán của biến mục tiêu

Độ trùng lặp: 68%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cho một thuộc tính của dữ liệu, Mỗi nhánh thể hiện đầu ra của phép thử đó, còn nút lá đại diện cho giá trị dự đoán của biến mục tiêu

60. Trang 35: cây quyết định là một phương pháp thông dụng trong khai phá dữ liệu

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quyết định là một phương pháp thông dụng trong khai phá dữ liệu Khi đó, Cây quyết định

61. Trang 35: Nó quyết định mô tả cấu trúc của một cây, trong đó, các lá đại diện cho các phân loại còn cành đại diện cho các kết hợp của các thuộc tính dẫn tới phân loại đó.

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quyết định mô tả một cấu trúc cây, trong đó các lá đại diện cho các phân loại còn cành đại diện cho các kết hợp của các thuộc tính dẫn tới phân loại đó[1] một cây.

62. Trang 35: một quyết định có thể được học bằng cách chia tập hợp nguồn thành các tập con dựa theo một kiểm tra giá trị thuộc tính

Độ trùng lặp: 97%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quyết định có thể được học bằng cách chia tập hợp nguồn thành các tập con dựa theo Một kiểm tra giá trị thuộc tính [

63. Trang 35: thuật toán Cây quyết định là một trong những thuật toán học máy phổ biến và dễ hiểu nhất

Độ trùng lặp: 81%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cây quyết định là một trong những Thuật toán học máy phổ biến nhất nhờ tính dễ hiểu và

64. Trang 35: Cây được xây dựng đệ quy từ trên xuống và theo cách chia để trị

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Cây được xây dựng đệ quy từ trên xuống và theo cách chia để trị

65. Trang 36: Thông tin cần thiết để phân chia D theo thuộc tính A

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thông tin cần thiết để phân chia D theo thuộc tính A

66. Trang 36: $() = () ()^2 12$ Tiếp tục Áp dụng cho mỗi nút con của nút gốc cho đến khi đạt đến nút lá hoặc nút có entropy = 0 ta được Cây quyết định

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cho mỗi nút con của nút gốc này cho đến khi đạt đến nút lá hoặc nút có

67. Trang 37: Xử lý dữ liệu nhiễu loại bỏ hoặc giảm thiểu ảnh hưởng của nhiễu bằng các kỹ thuật làm sạch dữ liệu

Độ trùng lặp: 57%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: hoặc giảm thiểu ảnh hưởng của chúng đến tính chính xác của dữ liệu. Trong quá trình làm sạch dữ liệu các

68. Trang 37: Rừng ngẫu nhiên (Random Forest) là một trong những thuật toán học máy có giám sát được sử dụng rộng rãi

Độ trùng lặp: 75%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Random forest Rừng ngẫu nhiên là một thuật toán học máy có giám sát [8] Đây là một trong những thuật toán được sử dụng

69. Trang 37: Nó là một kỹ thuật Ensemble, tức là Nó kết hợp nhiều mô hình nhỏ để tạo ra một mô hình dự đoán mạnh mẽ hơn

Độ trùng lặp: 50%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: để tạo ra một mô hình dự đoán mạnh hơn

70. Trang 37: Sau đó Ở bước dự đoán với một dữ liệu mới, mỗi cây quyết định sẽ đưa ra kết quả của mình kết quả của thuật toán này sẽ được tổng kết dựa trên kết quả của từng cây quyết định

Độ trùng lặp: 55%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: mỗi cây quyết định sẽ khác nhau (có yếu tố random) Sau đó kết quả dự đoán được tổng hợp từ các cây quyết định ở bước huấn luyện thì mình, sẽ xây dựng nhiều cây quyết định các cây quyết định có thể khác nhau (phần Sau mình, sẽ nói mỗi cây được xây dựng như thế nào) Sau đó ở bước dự đoán với một dữ liệu mới.

71. Trang 38: Random forest được coi là một phương pháp mạnh mẽ và chính xác vì có nhiều cây quyết định tham gia vào quá trình này

Độ trùng lặp: 73%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Random Forest được coi là một phương pháp chính xác và mạnh mẽ vì số cây quyết định tham gia vào quá trình này

72. Trang 38: trước khi nó đưa ra quyết định cho bài toán thì phải chờ tất các cây trong rừng phải đưa ra dự đoán, cho cùng một đầu vào cho trước và thực hiện bỏ phiếu trên đó

Độ trùng lặp: 71%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: nó đưa ra dự đoán tất cả các cây trong rừng phải đưa ra dự đoán cho cùng một đầu vào cho Trước và sau đó thực hiện bỏ phiếu trên đó

73. Trang 39: Giả sử bộ dữ liệu của mình có n dữ liệu (sample) và mỗi dữ liệu có d thuộc tính

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Giả sử bộ dữ liệu của mình có n dữ liệu (sample) và mỗi dữ liệu có d thuộc tính (

74. Trang 40: độ chính xác là Một trong những chỉ số đơn giản và phổ biến nhất được sử dụng để

đánh giá các mô hình hồi quy và phân loại

Độ trùng lặp: **64%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: một số chỉ số được sử dụng để đánh giá hiệu suất của mô hình như Accuracy (độ chính xác) là số liệu đơn giản và phổ biến nhất đo lường số dự đoán chính xác do mô hình đưa ra chia cho tổng số dự đoán Tuy nhiên, Độ chính xác thôi có thể không đủ để đánh giá các mô hình

75. Trang 41: Nó đo lường khả năng của mô hình phát hiện ra tất cả các trường hợp tích cực thực sự trong dữ liệu

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đo lường khả năng của mô hình trong việc phát hiện tất cả các trường hợp thực sự dương tính Nó

76. Trang 42: F1 score cung cấp một độ đo cân bằng giữa Recall, và Precision hai chỉ số quan trọng trong việc đánh giá các mô hình phân loại

Độ trùng lặp: **55%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trong việc đánh giá mô hình phân loại Tại sao F1 Score là một metric quan trọng để đánh giá hiệu suất của mô hình phân loại F1 Score là một metric quan trọng để đánh giá hiệu suất của mô hình phân loại vì nó kết hợp cả precision, và recall hai thành phần đánh giá quan trọng trong việc đo lường độ chính xác của mô hình precision, đo lường tỷ lệ các dự đoán đúng theo phân loại positive (đúng dương), trong khi recall đo lường tỷ lệ các trường hợp positive (đúng dương) mà mô hình dự đoán đúng Hiểu đơn giản, precision, đo lường khả năng của mô hình xác định đúng các trường hợp positive, trong khi recall đo lường khả năng của mô hình phát hiện được tất cả các trường hợp positive Tuy nhiên, precision, và recall đôi khi có mối quan hệ đảo ngược nhau Khi tăng precision, thường recall sẽ giảm và ngược lại Do đó, chỉ sử dụng precision, hoặc recall một mình không đủ để đánh giá hiệu suất của mô hình F1 Score giúp kết hợp cả precision, và recall thành một con số duy nhất, từ đó đánh giá tính chất cân bằng giữa

77. Trang 43: hơn nữa, mô hình này có thể giúp chúng ta hiểu rõ hơn về các yếu tố nguy cơ và phát triển các phương pháp điều trị mới

Độ trùng lặp: **54%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: này có thể giúp các nhà nghiên cứu hiểu rõ Hơn về các cơ chế cơ bản của bệnh và phát triển các phương pháp điều trị

78. Trang 43: Tuy nhiên, đây chỉ là một công cụ hỗ trợ và không thể thay thế hoàn toàn đánh giá lâm sàng của bác sĩ

Độ trùng lặp: **69%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tuy nhiên đây chỉ là một công cụ hỗ trợ và không thể thay thế hoàn toàn

79. Trang 45: làm sạch dữ liệu Sử dụng các phương pháp loại bỏ các bản ghi chứa giá trị thiếu hoặc điền các giá trị thiếu bằng giá trị trung bình trung vi hoặc giá trị gần nhất

Độ trùng lặp: **55%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: loại bỏ các giá trị trùng lặp trong dữ liệu Sử dụng các công cụ để thay thế các giá trị bị thiếu bằng giá trị trung bình, hoặc giá trị gần đúng khác Sử dụng các công cụ để chuẩn hóa dữ liệu đối với các biến có đơn vị khác nhau Sử dụng các công cụ để phát hiện và loại bỏ nhiễu trong dữ liệu ví dụ như bộ lọc số hoặc bộ lọc động Khi Làm sạch dữ liệu

80. Trang 45: huấn luyện mô hình mô hình được huấn luyện bằng cách đưa dữ liệu huấn luyện vào và điều chỉnh các thông số của mô hình để đạt được dự đoán chính xác nhất

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Huấn luyện mô hình này sẽ sử dụng dữ liệu đã được tiền xử lý để học các mẫu và quy tắc từ dữ liệu Huấn luyện Huấn luyện và điều chỉnh mô hình mô hình được Huấn luyện bằng cách thực hiện các vòng lặp trên dữ liệu Huấn luyện Trong quá trình này, mô hình cải thiện độ chính xác và khả năng dự đoán bằng cách điều chỉnh các tham số và trọng số Mục tiêu là tìm ra một mô hình tốt nhất có khả năng dự đoán chính xác

81. Trang 45: Đánh giá và tinh chỉnh mô hình được Đánh giá bằng cách sử dụng dữ liệu kiểm tra độc lập (chưa từng được sử dụng Trong quá trình training) để đo lường hiệu suất và độ chính xác của dự đoán

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đánh giá mô hình mô hình đã huấn luyện được Đánh giá trên dữ liệu kiểm tra để đo lường hiệu suất và

82. Trang 45: Python được thiết kế với tư tưởng giúp người học dễ đọc, dễ hiểu và dễ nhớ vì thế ngôn ngữ Python có hình thức rất sáng sủa, cấu trúc rõ ràng, thuận tiện cho người mới học

Độ trùng lặp: 69%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Python được thiết kế với ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ, Python là ngôn ngữ có hình thức rất sáng sủa, cấu trúc rõ ràng, thuận tiện cho người mới học lập trình và là ngôn ngữ lập trình dễ học cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu Python vì

83. Trang 46: Cấu trúc của Python cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu nói cách khác thì so với các ngôn ngữ lập trình khác chúng ta có thể sử dụng ít dòng code hơn để viết ra một chương trình trong Python

Độ trùng lặp: 51%

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, Python là một ngôn ngữ lập trình được sử dụng phổ biến ngày nay từ trong môi trường học đường cho tới các dự án lớn ngôn ngữ phát triển nhiều loại ứng dụng phần mềm khác nhau như các chương trình

84. Trang 46: Ban đầu, Python được phát triển để chạy trên nền Unix vì thế nó là mã nguồn mở

Độ trùng lặp: 72%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ban đầu, Python được phát triển để chạy trên nền Unix.

85. Trang 46: Python là một ngôn ngữ lập trình đa mẫu hình do nó hỗ trợ hoàn toàn lập trình hướng đối tượng và lập trình cấu trúc

Độ trùng lặp: 80%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Python là một ngôn ngữ lập trình đa mẫu hình, lập trình hướng đối tượng và lập trình cấu trúc được hỗ trợ hoàn

86. Trang 46: Nhờ vậy, Python có thể làm được rất nhiều việc và được sử dụng trong nhiều lĩnh vực khác nhau

Độ trùng lặp: 56%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: khác nhau và được sử dụng trong nhiều lĩnh vực khác nhau

87. Trang 47: Numpy (Numeric Python) là một thư viện toán học phổ biến và mạnh mẽ của Python

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Numpy (Numeric Python) là một thư viện toán học phổ biến và mạnh mẽ của Python

88. Trang 47: cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là dữ liệu ma trận và mảng, lớn với tốc độ xử lý nhanh hơn nhiều lần khi chỉ sử dụng core Python đơn thuần

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là dữ liệu ma trận và mảng, lớn với tốc độ xử lý nhanh hơn nhiều lần khi chỉ sử dụng core Python đơn thuần

89. Trang 47: Ma trận và phép toán Ma trận NumPy hỗ trợ Các phép toán Ma trận như nhân Ma trận chuyển vị, Ma trận tính Định thức nghịch đảo giải hệ phương trình tuyến tính v v

Độ trùng lặp: 65%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: giải hệ phương trình tuyến tính, \$ Ma trận + Khái niệm Ma trận + các phép toán Ma trận và tính chất cộng, trừ, nhân với số, nhân Ma trận lũy thừa + Ma trận nghịch đảo, và cách tìm bằng phương pháp Gauss Jordan + Ma trận chuyển vị Ma trận đối xứng, Ma trận hoán vị Bài tập \$ +\$ \$ định thức, +

90. Trang 47: Thư viện pandas trong python là một Thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Thư viện pandas trong python là một Thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu

91. Trang 47: Đây cũng là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình

python

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đây cũng là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình python

92. Trang 47: Thư viện này được sử dụng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thư viện này được sử dụng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu

93. Trang 48: thư viện này cho phép vẽ biểu đồ đường, biểu đồ cột, biểu đồ scatter, và nhiều loại biểu đồ khác

Độ trùng lặp: **70%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đường, biểu đồ cột, biểu đồ phân phối, biểu đồ hộp và nhiều loại biểu đồ khác
Bên cạnh đó, Thư viện này cũng cho phép

94. Trang 48: xử lý dữ liệu Seaborn cung cấp Các công cụ để xử lý và Trực quan hóa dữ liệu có cấu trúc, bao gồm việc xử lý dữ liệu thiếu, biến đổi dữ liệu và làm sạch dữ liệu để chuẩn bị cho Trực quan hóa

Độ trùng lặp: **55%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: liệu để chuẩn bị cho phân tích các bước tiền Xử lý dữ liệu có thể bao gồm loại bỏ dữ liệu trùng lặp, dữ liệu thiếu, dữ liệu nhiễu và dữ liệu không đúng định dạng Phân tích dữ liệu
Sau khi tiền Xử lý các tổ chức và doanh nghiệp sử dụng các công cụ phân tích dữ liệu để khai thác thông tin từ các tập dữ liệu lớn các kỹ thuật phân tích dữ liệu có thể bao gồm thống kê, khai phá dữ liệu học máy và trí tuệ nhân tạo trực quan hóa dữ liệu

95. Trang 49: thư viện Matplotlib là một trong những thư viện trực quan hóa dữ liệu phổ biến nhất trong Python

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: một Thư viện phổ biến nhất trong Python là Matplotlib giúp các bạn có thể trực quan hóa dữ liệu

96. Trang 49: bạn có thể tùy chỉnh các yếu tố như màu sắc, kích thước, chú thích và Tiêu đề, để tạo ra các biểu đồ tùy chỉnh

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Bạn có thể tùy chỉnh các yếu tố như tiêu đề nhãn trục, màu sắc, chú thích và các thuộc tính khác tùy chỉnh theo từng dạng biểu đồ

97. Trang 59: Xây dựng giao diện cho phép người dùng có thể nhập dữ liệu và trả kết quả cho người dùng

Độ trùng lặp: **59%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giao diện cho phép người dùng có thể nhập dữ liệu /*

--- Hết ---