



BÁO CÁO KIỂM TRA TRÙNG LẬP

Thông tin tài liệu

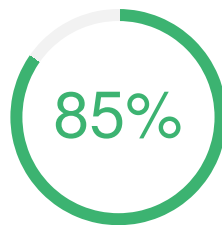
Tên tài liệu:	DATN-DinhNgocAnh
Tác giả:	Lê Nguyễn Tuấn Thành
Điểm trùng lặp:	15
Thời gian tải lên:	17:15 09/07/2024
Thời gian sinh báo cáo:	17:23 09/07/2024
Các trang kiểm tra:	78/78 trang



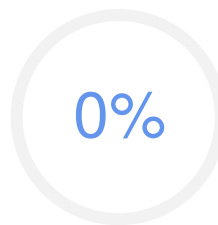
Kết quả kiểm tra trùng lặp



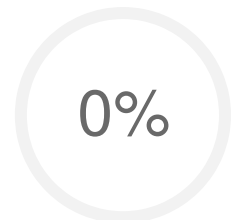
Có 15% nội dung trùng
lặp



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 bizflycloud.vn

Danh sách các câu trùng lặp

1. Trang 5: Nội dung các phần thuyết minh và tính toán Tỷ lệ % Chương 1 Cơ sở lý thuyết 30%

Độ trùng lặp: **76%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Nội dung các phần thuyết minh và tính toán Chương 1 Cơ sở lý thuyết

2. Trang 6: 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

3. 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

4. Trang 7: Đây là loại thường xuyên nhất trong tất cả ung thư và nguyên nhân chính gây tử vong ở phụ nữ trên toàn thế giới

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nguyên nhân chính gây tử vong đối với phụ nữ tại nhiều nước Theo GLOBOCAN , trên toàn thế giới

5. Trang 8: Tiền xử lý dữ liệu bằng cách xử lý các giá trị thiếu, chuẩn hóa dữ liệu và chia dữ liệu thành tập huấn luyện và tập thử nghiệm

Độ trùng lặp: **61%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tiền xử lý dữ liệu có thể bao gồm các bước như tách dữ liệu chia dữ liệu thành tập huấn luyện và tập kiểm tra, chuẩn hóa dữ liệu và

6. Trang 8: tối ưu hóa các tham số của các mô hình học máy bằng phương pháp điều chỉnh siêu tham số

Độ trùng lặp: **63%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: pháp Tối ưu hóa các mô hình học máy và các tham số của các mô

7. Trang 8: đánh giá hiệu suất mô hình đánh giá hiệu suất của các mô hình học máy trên tập thử nghiệm bằng các chỉ số như độ chính xác độ nhạy, độ đặc hiệu và F1 score

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: độ nhạy, độ đặc hiệu và độ chính xác, độ nhạy, được tính bằng tỉ số các cơn ho được xác, định chính xác, trên tổng số các cơn ho trong một tập hợp thử nghiệm độ đặc hiệu

8. Trang 8: Phân tích các yếu tố ảnh hưởng đến Hiệu suất của mô hình học máy

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phân tích các yếu tố đầu vào ảnh hưởng đến khả năng dự báo chính xác của mô hình học máy GB hiệu suất của mô

9. Trang 8: Đánh giá hiệu suất của từng mô hình bằng các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu

Độ trùng lặp: **65%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: hình bằng các chỉ số như độ chính xác, xác, suất

10. Trang 9: Em xin cam đoan đây là Đồ án tốt nghiệp của bản thân Em

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

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

11. Trang 9: 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

12. Trang 10: lời đầu tiên em xin gửi lời cảm ơn chân thành tới thầy cô trường Đại học Thủy lợi nói chung thầy cô khoa Công nghệ thông tin nói riêng đã tận tình giảng dạy truyền đạt những kiến thức, quý giá cho em cũng như các bạn sinh viên khác trong quá trình học tập tại trường từ đó làm cơ sở cho chúng em có kiến thức, để thực hiện đồ án này

Độ trùng lặp: **56%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: gửi Lời cảm ơn chân thành sâu sắc đến các thầy cô trường Đại học Thủy Lợi nói chung, và các thầy cô trong khoa Công nghệ thông tin nói riêng đã tận tình giảng dạy, cho em những kiến thức kinh nghiệm trong suốt thời gian học tập tại trường Đặc biệt em xin gửi Lời cảm ơn chân thành đến cô Trần Thị Ngân đã tận tình giúp đỡ chỉ bảo hướng dẫn tận tình cho em trong

13. Trang 10: Lê Nguyễn Tuấn Thành, giảng viên Khoa Công nghệ thông tin đã tận tình hướng dẫn, chỉ bảo, và góp ý cho Em trong suốt quá trình thực hiện đồ án

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: khoa Công nghệ thông tin đã tận tình chỉ dạy và quan tâm trong suốt quá trình học tập và rèn luyện tại trường em trân trọng biết ơn ThS Nguyễn Văn Hách đã tận tình hướng dẫn, chỉ bảo góp ý cho em trong suốt quá trình thực hiện

14. Trang 10: Với kinh nghiệm cũng như thời gian có hạn nên đồ án của em không thể không tránh khỏi những thiếu sót

Độ trùng lặp: 73%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cũng như thời gian có hạn nên bản đồ án của em không thể tránh khỏi những thiếu sót

15. Trang 10: Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của quý Thầy Cô để đồ án của Em được hoàn thiện tốt hơn

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của quý thầy cô và các bạn để bài đồ án của Em được hoàn thiện hơn

16. Trang 14: 59 DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ Kí hiệu Diễn GIẢI Ý nghĩa

Độ trùng lặp: 68%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ Tên TỪ VIẾT TẮT CSDL TK NCC vv Ý nghĩa

17. Trang 14: ANN Artificial Neural Network Mạng nơ ron nhân tạo SVM Support vector machine Máy vector hỗ trợ

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ANN Artificial Neural Network Mạng nơ ron nhân tạo SVM Support vector machine Máy vector hỗ trợ

18. Trang 16: Phương pháp nghiên cứu Phương pháp nghiên cứu sử dụng trong suốt quá trình thực hiện đồ án tốt nghiệp. Là Phương pháp nghiên cứu lý thuyết các mô hình để thực hiện với bài toán cụ thể

Độ trùng lặp: 57%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nghiên cứu lý thuyết + Phương pháp nghiên cứu tài liệu là Phương pháp được áp dụng xuyên suốt quá trình thực hiện đồ án tốt nghiệp

19. Trang 16: phương pháp thực nghiệm dựa trên tập dữ liệu thực tế để dự báo và có độ đo để xác định sự chính xác của mô hình

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thực nghiệm dựa trên dữ liệu huấn luyện mà chưa sử dụng dữ liệu kiểm tra (hay dữ liệu thực tế) để kiểm chứng mức độ chính xác của các Phương trình thực nghiệm; Chưa có nghiên cứu trong nước nào thực hiện nghiên cứu và so sánh hiệu quả của nhiều Phương pháp thực nghiệm để tìm ra Phương trình thực nghiệm tối ưu cho dự báo

20. Trang 17: Năm 1950 Nhà bác học Alan Turing đã tạo ra "Turing Test (phép thử Turing)" để xác định xem liệu một máy tính có trí thông minh thực sự hay không

Độ trùng lặp: 96%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 1950 Nhà bác học Alan Turing đã tạo ra "Turing Test (phép thử Turing)" để xác định xem liệu một máy tính có trí thông minh thực sự hay không

21. Trang 17: Để vượt qua bài kiểm tra đó, một máy tính phải có khả năng đánh lừa một con người tin rằng nó cũng là con người

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Để vượt qua bài kiểm tra đó, một máy tính phải có khả năng đánh lừa một con người tin rằng nó cũng là con người

22. Trang 17: Năm 1952 Arthur Samuel đã viết ra chương trình học máy (computer learning) đầu tiên

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Năm Arthur Samuel đã viết ra chương trình học máy (computer learning) đầu tiên

23. Trang 17: Chương trình này là trò chơi cờ đam, và hãng máy tính IBM đã cải tiến trò chơi này để nó có thể tự học và tổ chức những nước đi trong chiến lược để giành chiến thắng

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chương trình này là trò chơi cờ đam, và hãng máy tính IBM đã cải tiến trò chơi này để nó có thể tự học và tổ chức những nước đi trong chiến lược để giành chiến thắng

24. Trang 18: Năm 1957, Frank Rosenblatt đã thiết kế mạng nơron (neural network) đầu tiên cho máy tính, trong đó mô phỏng quá trình suy nghĩ của bộ não con người

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Năm 1957 Frank Rosenblatt đã thiết kế mạng nơron (neural network) đầu tiên cho máy tính, trong đó mô phỏng quá trình suy nghĩ của bộ não con người

25. Trang 18: Năm 1967 Thuật toán "nearest neighbor" đã được viết, cho phép các máy tính bắt đầu sử dụng những mẫu nhận dạng (pattern recognition) rất cơ bản

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thuật toán "nearest neighbor" đã được viết, cho phép các máy tính bắt đầu sử dụng những mẫu nhận dạng (pattern recognition) rất cơ bản

26. Trang 18: Nó được sử dụng để vẽ ra lộ trình cho một người bán hàng có thể bắt đầu đi từ một thành phố ngẫu nhiên nhưng đảm bảo anh ta sẽ đi qua tất cả các thành phố khác theo một quãng đường ngắn nhất

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Nó được sử dụng để vẽ ra lộ trình cho một người bán hàng có thể bắt đầu đi từ một thành phố ngẫu nhiên nhưng đảm bảo anh ta sẽ đi qua tất cả các thành phố khác theo một quãng đường ngắn nhất

27. Trang 18: Năm 1979 sinh viên tại trường đại học Stanford đã phát minh ra giỏ hàng "Stanford Cart" có thể điều hướng để tránh các chướng ngại vật trong một căn phòng

Độ trùng lặp: **94%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Sinh viên tại trường đại học Stanford đã phát minh ra giỏ hàng "Stanford Cart" có thể điều hướng để tránh các chướng ngại vật trong một căn phòng Năm

28. Trang 18: Năm 1981 Gerald Dejong giới thiệu về khái niệm Explanation Based Learning (EBL), trong đó một máy tính phân tích dữ liệu huấn luyện và tạo ra một quy tắc chung để nó có thể làm theo bằng cách loại bỏ đi những dữ liệu không quan trọng

Độ trùng lặp: **95%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Gerald Dejong giới thiệu về khái niệm Explanation Based Learning (EBL), trong đó một máy tính phân tích dữ liệu huấn luyện và tạo ra một quy tắc chung để nó có thể làm theo bằng cách loại bỏ đi những dữ liệu không quan trọng

29. Trang 18: Năm 1985 Terry Sejnowski đã phát minh ra NetTalk, nó có thể học cách phát âm các từ giống như cách một đứa trẻ tập nói

Độ trùng lặp: **92%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Terry Sejnowski đã phát minh ra NetTalk, nó có thể học cách phát âm các từ giống như cách một đứa trẻ tập nói

30. Trang 18: Năm 1990 machine learning đã dịch chuyển từ cách tiếp cận hướng kiến thức (knowledge driven) sang cách tiếp cận hướng dữ liệu (data driven)

Độ trùng lặp: **91%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Machine Learning đã dịch chuyển từ cách tiếp cận hướng kiến thức (knowledge driven) sang cách tiếp cận hướng dữ liệu (data driven)

31. Trang 18: Các nhà khoa học bắt đầu tạo ra Các chương trình cho máy tính để phân tích một lượng lớn dữ liệu và rút ra Các kết luận hay là "học" từ Các kết quả đó

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các nhà khoa học bắt đầu tạo ra Các chương trình cho máy tính để phân tích một lượng lớn dữ liệu và rút ra Các kết luận hay là học từ Các kết quả đó

32. Trang 18: Năm 1997 Deep Blue của hãng IBM đã đánh bại nhà vô địch cờ vua thế giới

Độ trùng lặp: **89%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Deep Blue của hãng IBM đã đánh bại nhà vô địch cờ vua thế giới Năm

33. Trang 18: [1] Năm 2006 Geoffrey Hinton đã đưa ra một thuật ngữ "deep learning" để giải thích các thuật toán mới cho phép máy tính "nhìn thấy" và phân biệt các đối tượng và văn bản trong các hình ảnh và video

Độ trùng lặp: **93%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Geoffrey Hinton đã đưa ra một thuật ngữ "deep learning" để giải thích các thuật toán mới cho phép máy tính "nhìn thấy" và phân biệt các đối tượng và văn bản trong các hình ảnh và video Năm

34. Trang 18: Năm 2010 Microsoft Kinect có thể theo dõi 20 hành vi của con người ở một tốc độ 30 lần mỗi giây, cho phép con người tương tác với máy tính thông qua các hành động và cử chỉ

Độ trùng lặp: **97%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 2010 Microsoft Kinect có thể theo dõi 20 hành vi của con người ở một tốc độ 30 lần mỗi giây, cho phép con người tương tác với máy tính thông qua các hành động và cử chỉ

35. Trang 18: Năm 2011 máy tính Watson của hãng IBM đã đánh bại các đối thủ là con người tại Jeopardy

Độ trùng lặp: **90%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Máy tính Watson của hãng IBM đã đánh bại các đối thủ là con người tại Jeopardy Năm

36. Trang 19: Năm 2011 Google Brain đã được phát triển, và mạng deep nơon (deep neural network) của nó có thể học để phát hiện và phân loại nhiều đối tượng theo cách mà một con mèo thực hiện

Độ trùng lặp: **95%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Google Brain đã được phát triển, và mạng deep nơon (deep neural network) của nó có thể học để phát hiện và phân loại nhiều đối tượng theo cách mà một con mèo thực hiện Năm

37. Trang 19: Năm 2012 X Lab của Google phát triển một thuật toán machine learning có khả năng tự động duyệt qua các video trên YouTube để xác định xem video nào có chứa những con mèo

Độ trùng lặp: **97%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 2012 X Lab của Google phát triển một thuật toán machine learning có khả năng tự động duyệt qua các video trên YouTube để xác định xem video nào có chứa những con mèo

38. Trang 19: Năm 2014 Facebook phát triển DeepFace, một phần mềm thuật toán có thể nhận dạng hoặc xác minh các cá nhân dựa vào hình ảnh ở mức độ giống như con người có thể

Độ trùng lặp: **96%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 2014 Facebook phát triển DeepFace, một phần mềm thuật toán có thể nhận dạng hoặc xác minh các cá nhân dựa vào hình ảnh ở mức độ giống như con người có thể

39. Trang 19: Năm 2015 Amazon ra mắt nền tảng machine learning riêng của mình

Độ trùng lặp: **84%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Amazon ra mắt nền tảng machine learning riêng của mình 2015

40. Trang 19: Năm 2015 Microsoft tạo ra Distributed Machine Learning Toolkit, trong đó cho phép phân phối hiệu quả các vấn đề Machine Learning trên nhiều máy tính

Độ trùng lặp: **93%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Microsoft tạo ra Distributed Machine Learning Toolkit, trong đó cho phép phân phối hiệu quả các vấn đề Machine Learning trên nhiều máy tính Năm

41. Trang 19: Năm 2015 Hơn 3 000 nhà nghiên cứu AI và Robotics, được sự ủng hộ bởi những nhà khoa học nổi tiếng như Stephen Hawking, Elon Musk và Steve Wozniak (và nhiều người khác), đã ký vào một bức thư ngỏ để cảnh báo về sự nguy hiểm của vũ khí tự động trong việc lựa chọn và tham gia vào các mục tiêu mà không có sự can thiệp của con người

Độ trùng lặp: **93%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nhà nghiên cứu AI và Robotics, được sự ủng hộ bởi những nhà khoa học nổi tiếng như Stephen Hawking, Elon Musk và Steve Wozniak (và nhiều người khác), đã ký vào một bức thư ngỏ để cảnh báo về sự nguy hiểm của vũ khí tự động trong việc lựa chọn và tham gia vào các mục tiêu mà không có sự can thiệp của con người

42. Trang 19: Năm 2016 Thuật toán trí tuệ nhân tạo của Google đã đánh bại nhà vô địch trò chơi Cờ Vây, được cho là trò chơi phức tạp nhất thế giới (khó hơn trò chơi Cờ vua rất nhiều)

Độ trùng lặp: **95%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Năm Thuật toán trí tuệ nhân tạo của Google đã đánh bại nhà vô địch trò chơi Cờ Vây, được cho là trò chơi phức tạp nhất thế giới (khó hơn trò chơi Cờ vua rất nhiều)

43. Trang 20: Có nhiều lĩnh vực trong khoa học máy tính và trí tuệ nhân tạo,

Độ trùng lặp: **87%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: lĩnh vực trong khoa học máy tính và trí tuệ nhân tạo nó nghiên cứu và phát triển các thuật toán và mô hình máy tính để máy tính Có

44. Trang 21: mô hình dự đoán (Predictive Model) mô hình học máy học từ dữ liệu huấn luyện và được sử dụng để dự báo kết quả cho dữ liệu mới

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: sử dụng dữ liệu huấn luyện đã được gán nhãn trước đó Thuật toán học từ các cặp dữ liệu huấn luyện và nhãn tương ứng để xây dựng Mô hình dự đoán cho

45. Trang 22: Nói đơn giản đây là giai đoạn hệ thống tổ hợp các dữ liệu đầu vào, và sử dụng các thuật toán để hình thành, đánh giá và lựa chọn các phương án khả thi nhất

Độ trùng lặp: **64%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: hệ thống tổ hợp các dữ liệu đầu vào vận dụng thuật toán để hình thành, đánh giá, và chọn ra những phương án khả thi nhất

46. Trang 23: Tiếp đến, hệ thống sẽ cập nhật cho bộ dữ liệu ban đầu cách giải quyết mới nhằm tạo ra những giải pháp chính xác hơn theo thời gian

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cập nhật cho bộ dữ liệu ban đầu cách xử lý để có những phương án chuẩn xác hơn theo thời gian

47. Trang 23: dựa vào mục đích, cấu trúc và thuật toán tạo nên chúng thì Machine Learning Có thể chia thành bốn loại như sau học Có giám sát học không giám sát học bán giám sát và học tăng cường

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có giám sát, học không giám sát, học bán giám sát, và học tăng

48. Trang 23: Học máy có giám sát là một Kỹ thuật mà trong đó các thuật toán được huấn luyện trên các tập dữ liệu đã được gán nhãn, cho phép Chúng phân loại dữ liệu hoặc dự đoán kết quả chính xác

Độ trùng lặp: **59%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Học máy có giám sát (Supervised Machine Learning) là phương pháp mà trong đó máy tính được Học từ dữ liệu đã được đánh dấu trước, để phát triển các thuật toán có khả năng phân loại hoặc dự đoán kết quả một cách chính xác kỹ thuật này thường được áp dụng cho các bài toán phân lớp (Classification) Ví dụ chúng ta có một tập dữ liệu gồm các hình ảnh của chó và mèo trong giai đoạn đầu, máy tính được huấn luyện

49. Trang 24: Quá trình gán nhãn cho các bộ dữ liệu sử dụng trong học có giám sát thường được thực hiện thông qua các dịch vụ tuyển dụng tự do, như CrowdFlower (Figure Eight), Clickworker và Amazon Mechanical Turk

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trình gán nhãn cho các bộ dữ liệu dùng trong học giám sát thường được thực hiện thông qua các dịch vụ tuyển dụng tự do, như Amazon Mechanical Turk

50. Trang 25: được áp dụng rộng rãi trong nhiều lĩnh vực bao gồm phân khúc khách hàng, phát hiện gian lận và phân tích hình ảnh

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Được áp dụng rộng rãi trong nhiều lĩnh vực, bao gồm phân tích biểu hiện

51. Trang 25: phân cụm là Một trong những kỹ thuật học không giám sát phổ biến nhất

Độ trùng lặp: 80%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: một trong những kỹ thuật học tập không giám sát phổ biến nhất là Phân cụm

52. Trang 25: Khai thác quy tắc kết hợp là một phương pháp dựa trên quy tắc để khám phá mối quan hệ thú vị giữa các điểm dữ liệu trong một tập dữ liệu lớn

Độ trùng lặp: 63%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một phương pháp học dựa trên quy tắc để học tập các mối quan hệ thú vị giữa các

53. Trang 26: Giảm kích thước là một kỹ thuật học máy không giám sát nhằm Giảm số lượng tính năng hoặc kích thước của tập dữ liệu

Độ trùng lặp: 59%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giảm kích thước là một kỹ thuật học không có giám sát giúp Giảm số lượng các tính năng trong tập dữ liệu

54. Trang 26: Các bài toán trong semi supervised learning, xảy ra khi có một lượng dữ liệu lớn, nhưng chỉ có một phần nhỏ trong số đó được gắn nhãn

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: chỉ có một phần nhỏ trong số đó được gắn nhãn Bằng cách sử dụng Semi Supervised Learning bạn có thể dùng phần dữ liệu

55. Trang 27: Học bán giám sát là một phương pháp nằm giữa Học có giám sát và Học không giám sát và nó giải quyết những thách thức chính của cả hai

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Học bán giám sát là một phương pháp Học trung hạ giữa Học có giám sát và Học

56. Trang 27: phương pháp này sử dụng một lượng nhỏ dữ liệu được gắn nhãn cùng với một lượng lớn dữ liệu chưa được gắn nhãn giúp giảm thời gian, và chi phí gắn nhãn thủ công so với Học có giám sát

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dữ liệu gắn nhãn với một lượng lớn dữ liệu chưa được gắn nhãn học có giám sát là Phương pháp sử dụng tập dữ liệu đã được gắn nhãn việc gắn nhãn thủ công này

57. Trang 27: một ví dụ điển hình khác trong nhóm này là khi chỉ một phần nhỏ ảnh hoặc văn bản

được gán nhãn (ví dụ như ảnh của con người, động vật hoặc các văn bản về khoa học, chính trị) trong khi phần lớn ảnh/ văn bản khác chưa có nhãn được thu thập từ internet.

Độ trùng lặp: **67%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chỉ Một phần trong chúng được gán nhãn Một ví dụ điển hình của nhóm này là chỉ có Một phần ảnh hoặc văn bản được gán nhãn (ví dụ bức ảnh về người, động vật, hoặc các văn bản khoa học, chính trị), và phần lớn các bức ảnh/văn bản khác chưa được gán nhãn được thu thập từ internet

58. Trang 27: Thực tế cho thấy nhiều bài toán Machine Learning thuộc nhóm này do việc thu thập dữ liệu có nhãn tốn rất nhiều thời gian và chi phí

Độ trùng lặp: **79%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thực tế cho thấy rất nhiều các bài toán Machine Learning thuộc vào nhóm này vì việc thu thập dữ liệu có nhãn tốn rất nhiều thời gian và có chi phí

59. Trang 27: Ngược lại, dữ liệu chưa có nhãn có thể được thu thập từ internet với chi phí thấp hơn

Độ trùng lặp: **83%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ngược lại, dữ liệu chưa có nhãn có thể được thu thập với chi phí thấp từ internet

60. Trang 28: Quá trình này căn cứ vào nguyên lý học từ phản hồi (feedback) và thưởng (reward) để tối ưu hóa một hàm phần thưởng đã được định nghĩa sẵn

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: học từ phản hồi (feedback) và thưởng (reward) để tối đa hóa một hàm phần thưởng được

61. Trang 28: Mục đích duy nhất của Học tăng cường là phát triển ra một mô hình hành vi hiệu quả nhằm tối đa hóa tổng phần thưởng tích lũy của đại lý

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ra một mô hình hành động phù hợp để tăng tổng phần thưởng tích lũy của đại lý

62. Trang 29: Trong học tăng cường, các nhà phát triển đã phát triển một phương pháp để khen thưởng các hành vi tốt và trừng phạt các hành vi xấu

Độ trùng lặp: **70%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trong học tăng cường, các nhà phát triển nghĩ ra một phương pháp khen thưởng các hành vi mong muốn và trừng phạt các hành

63. Trang 29: Phương pháp này thường gán các giá trị dương cho các hành động được khuyến khích để tác nhân có xu hướng thực hiện chúng, và gán các giá trị âm cho các hành vi xấu để giảm thiểu chúng

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phương pháp này gán các giá trị dương cho các hành động mong muốn để khuyến khích tác nhân và các giá trị âm cho các hành vi

64. Trang 29: Quá trình này giúp tác nhân tối ưu hóa tổng lượng phần thưởng để đạt được một giải pháp tối ưu

Độ trùng lặp: **66%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trình này giúp tác nhân tìm kiếm phần thưởng tổng thể dài hạn và tối đa để đạt được một giải pháp tối ưu

65. Trang 29: Theo thời gian, tác nhân học cách tránh các hành vi tiêu cực và tìm kiếm những hành vi tích cực hơn

Độ trùng lặp: **58%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Theo thời gian, tác nhân học cách tránh điều tiêu cực và tìm kiếm điều tích cực

66. Trang 29: Phương pháp học này đã được áp dụng trong trí tuệ nhân tạo (AI) như một Phương pháp Để hướng dẫn học máy không giám sát thông qua phần thưởng và hình phạt

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Phương pháp học này đã được áp dụng trong trí tuệ nhân tạo (AI) như một cách chỉ đạo việc học máy không giám sát thông qua phần thưởng và hình phạt để

67. Trang 29: Machine learning là một lĩnh vực quan trọng trong việc phát triển khoa học và công nghệ nhờ vào những ưu điểm sau

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quan trọng trong việc phát triển khoa học ở lĩnh vực này, mà còn đóng vai trò quan trọng thúc đẩy sự phát triển khoa học và công nghệ

68. Trang 30: có thể hoạt động tự động Sau khi thiết lập Ví dụ, như trong phần mềm an ninh mạng máy học có thể liên tục giám sát và phát hiện các hoạt động bất thường trong lưu lượng mạng mà không cần sự can thiệp của người quản trị

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: sau khi thiết lập, ban đầu, nó Có thể hoạt động mà không cần sự can thiệp của con người ví dụ máy học trong phần mềm an ninh mạng, Có thể liên tục theo dõi và xác định các điểm bất thường trong lưu lượng mạng, mà không cần sự can thiệp của quản trị

69. Trang 30: Các thuật toán đang cố gắng tìm hiểu sở thích của người dùng để hiển thị chính xác những nội dung trên màn hình

Độ trùng lặp: **55%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *đang cố gắng tìm hiểu sở thích của*

70. Trang 30: Phát hiện gian lận Machine Learning có thể phân tích các mẫu, chẳng hạn như cách ai đó thường chi tiêu hoặc nơi họ thường mua sắm, để xác định các giao dịch thẻ tín dụng có khả năng gian lận các nỗ lực đăng nhập hoặc spam email

Độ trùng lặp: 84%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Phát hiện gian lận (Fraud detection) Máy có thể phân tích các mẫu, như cách ai đó thường chi tiêu hoặc nơi họ thường mua sắm để xác định các giao dịch thẻ tín dụng có khả năng gian lận các nỗ lực đăng nhập hoặc email*

71. Trang 31: Xe ô tô tự lái công nghệ đằng sau ô tô tự lái chủ yếu dựa trên học, máy đặc biệt là học, sâu (Deep Learning)

Độ trùng lặp: 66%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Công nghệ đằng sau ô tô tự lái dựa trên máy, học đặc biệt là học sâu*

72. Trang 31: Chẩn đoán hình ảnh trong y tế Các chương trình học máy có thể được huấn luyện để kiểm tra và phân tích Các hình ảnh y tế hoặc Các thông tin khác nhằm hỗ trợ Chẩn đoán

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Chẩn đoán và hình ảnh y tế Các chương trình máy học có thể được đào tạo để kiểm tra Các hình ảnh y tế hoặc thông tin khác*

73. Trang 32: Nó thực hiện điều này bằng cách sử dụng một hàm sigmoid để ánh xạ mọi giá trị đầu vào thành một giá trị xác suất trong khoảng từ 0 đến 1

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *thành một giá trị trong khoảng từ 0 đến 1 trong học máy, ta sử dụng hàm sigmoid để ánh xạ*

74. Trang 32: Do giá trị của biến phụ thuộc trong hồi quy Logistic có thể không nằm trong khoảng [0,1], ta cần một hàm số để ánh xạ giá trị dự báo vào không gian xác suất [0,1]

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *logistic cần một hàm số có tác dụng chiếu giá trị dự báo lên không gian xác suất nằm trong khoảng [0,1]*

75. Trang 32: hồi quy Logistic Phân tích mối quan hệ giữa một hoặc nhiều biến độc lập và Phân loại

Độ trùng lặp: 78%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *phân tích mối quan hệ giữa một biến phụ thuộc với một hoặc nhiều biến độc lập*

khác Điều này không đòi hỏi giữa biến độc lập và

76. Trang 34: Trong bài toán phân loại nhị phân ta thường lựa chọn một ngưỡng xác suất để quyết định dự báo nhãn cho một quan sát

Độ trùng lặp: 76%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trong bài toán phân loại nhị phân chúng ta sẽ lựa chọn một ngưỡng threshold về xác suất để đưa ra dự báo nhãn cho một quan sát Giả định ta chọn ngưỡng xác suất

77. Trang 35: Như vậy ta có thể nhận ra những điểm thuộc về nhãn 1 sẽ nằm bên phải đường biên phân chia WT trong khi những điểm thuộc về nhãn 0 sẽ nằm bên trái

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Như vậy ta có thể nhận ra những điểm thuộc về nhãn 1 sẽ nằm bên phải đường biên phân chia $\setminus(\mathbf{w}^{\text{intercal}}\mathbf{x})$ trong khi những điểm thuộc về nhãn 0 sẽ nằm bên trái

78. Trang 35: Đồng thời đường biên phân chia hai nhãn 0 và 1 cũng là một phương trình tuyến tính

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đồng thời đường biên phân chia hai nhãn 0 và 1 cũng là một phương trình tuyến tính

79. Trang 35: sản xuất Hồi quy Logistic được áp dụng để dự đoán các sự cố máy móc, quản lý chất lượng, và tối ưu hóa quy trình sản xuất

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: quy trình Sản xuất để tăng hiệu suất, giảm lãng phí và tối ưu hóa sự sắp xếp của các hoạt động Việc tối ưu hóa này có thể được thực hiện thông qua việc áp dụng các phương pháp quản lý chất lượng công nghệ tiên tiến và tối ưu hóa quy trình

80. Trang 37: Ngoài việc thực hiện phân loại tuyến tính, SVM có thể thực hiện phân loại phi tuyến tính, một cách hiệu quả bằng cách sử dụng những gì được gọi là "thủ thuật hạt nhân"

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: việc thực hiện phân loại tuyến tính, SVM có thể thực hiện phân loại phi tuyến tính, một cách hiệu quả bằng cách sử dụng cái được gọi là thủ thuật hạt nhân,

81. Trang 37: khoảng cách từ siêu phẳng đến các điểm dữ liệu gần nhất của mỗi lớp được gọi là biên an toàn (margin)

Độ trùng lặp: 65%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của siêu phẳng được quyết định bởi Khoảng cách (được gọi là biên margin) của điểm dữ liệu gần nhất của mỗi lớp đến

82. Trang 38: Trong không gian hai chiều, khoảng cách từ một điểm có tọa độ $(0,)$ tới đường thẳng có phương trình $1 \pm 2 \pm = 0$ được xác định bởi công thức

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: không gian hai chiều, khoảng cách từ một điểm có tọa độ (x_0, y_0) tới đường thẳng có phương trình $w_1x + w_2y \pm b = 0$ được xác định bởi $1 \pm w_2^2$ Trong không gian

83. Trang 38: (2 7) Trong không gian ba chiều, khoảng cách từ một điểm có tọa độ $(0, , 0)$ tới đường thẳng có phương trình $1 \pm 2 \pm 3 \pm = 0$ được xác định bởi công thức

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: = được xác định bởi công thức Trong không gian ba chiều, khoảng cách từ một điểm có tọa độ

84. Trang 38: Những điểm làm cho biểu thức trong dấu giá trị tuyệt đối mang giá trị dương nằm về cùng một phía (gọi là phía dương của đường thẳng), Những điểm làm cho biểu thức mang giá trị âm nằm về phía ngược lại (gọi là phía âm)

Độ trùng lặp: 76%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Những điểm làm cho biểu thức trong dấu tri tuyệt đối mang dấu dương nằm về cùng một phía (tạm gọi là phía dương), Những điểm làm cho giá trị này mang dấu âm nằm về phía còn lại (gọi là phía âm)

85. Trang 38: Các điểm nằm trên đường thẳng hoặc mặt phẳng làm cho tử số bằng 0 tức là khoảng cách từ điểm đó đến đường thẳng/mặt phẳng là 0

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: điểm nằm trên đường thẳng/ mặt phẳng sẽ làm cho giá trị của tử số bằng 0, hay khoảng cách

86. Trang 38: Việc này có thể được tổng quát lên trên không gian nhiều chiều, khoảng cách từ một điểm có tọa độ x_0 tới một siêu mặt phẳng có phương trình $x_0 + = 0$ được xác định bởi

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tới một siêu mặt phẳng Trong không gian 2 chiều ta biết rằng khoảng cách từ một điểm có tọa độ tới đường thẳng có phương trình được xác định bởi Việc này có thể được tổng quát lên không gian nhiều chiều khoảng cách

87. Trang 38: Giả sử có hai lớp khác nhau được miêu tả bởi các điểm trong không gian nhiều chiều và hai lớp này là linearly separable, có nghĩa là tồn tại một siêu mặt phẳng có thể chia chính xác hai lớp đó

Độ trùng lặp: 66%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có hai lớp khác nhau được mô tả bởi các điểm trong không gian nhiều chiều hai lớp này phân tách tuyến tính, tức tồn tại một siêu phẳng phân chia chính xác hai lớp đó Hãy tìm

một siêu mặt phẳng

88. Trang 38: Để tìm một siêu mặt phẳng chia hai lớp này, ta cần tìm một mặt phẳng sao cho tất cả các điểm của một lớp nằm về cùng một phía của mặt phẳng đó, và tất cả các điểm của lớp còn lại nằm về phía ngược lại của mặt phẳng

Độ trùng lặp: 60%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: tìm một siêu mặt phẳng phân chia hai lớp đó tức tất cả các điểm thuộc một lớp nằm về cùng một phía của siêu mặt phẳng đó và ngược phía với toàn bộ các điểm thuộc lớp còn lại

89. Trang 39: để thể hiện đầu vào của một điểm dữ liệu D là số chiều của dữ liệu và N là số dữ liệu

Độ trùng lặp: 82%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: thể hiện đầu vào của một điểm dữ liệu, và ý là nhân của điểm dữ liệu, đó d là số chiều của dữ liệu, và N là số

90. Trang 39: Gọi các điểm vuông xanh thuộc class 1, các điểm tròn đỏ thuộc class 1, và mặt $+ = 1 \pm 2 \pm = 0$ là mặt phân chia giữa 2 classes

Độ trùng lặp: 52%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các điểm vuông xanh thuộc lớp 1, các điểm tròn đỏ thuộc lớp 1, và mặt $(\pm) \equiv 1 \pm 2 \pm$ là mặt phân c

91. Trang 40: 10) Do đó, Bài toán tối ưu của SVM chính là Bài toán tìm trọng số w và độ lệch b sao cho margin đạt giá trị lớn nhất

Độ trùng lặp: 60%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: SVM chính là bài toán tìm w và b sao cho margin này đạt giá trị lớn nhất

92. Trang 40: 11) trong đó sgn được định nghĩa là hàm xác định dấu, nhận giá trị 1 nếu đối số là không âm và nhận giá trị -1 nếu đối số là âm

Độ trùng lặp: 77%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Trong đó hàm sgn là hàm xác định dấu, nhận giá trị 1 nếu đối số là không âm và

93. Trang 41: Tài chính Phân tích rủi ro, tín dụng, phát hiện gian lận, dự đoán thị trường chứng

Độ trùng lặp: 81%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Tài chính Phân tích rủi ro dự đoán thị trường chứng khoán, phát triển các sản phẩm Tài chính mới Ngân hàng phát hiện gian lận, đánh giá tín dụng,

94. Trang 47: Mạng thần kinh nhân tạo (ANN), là một phương pháp học sâu xuất phát từ khái niệm Mạng thần kinh sinh học não người

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Mạng thần kinh nhân tạo (ANN) là một

95. Trang 48: Trong ANN, trừ lớp đầu vào thì tất cả các node thuộc các layer khác đều hoàn toàn được kết nối với các node thuộc layer trước nó

Độ trùng lặp: 76%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: layer thì tất cả các node thuộc các layer khác đều full connected với các node thuộc layer trước nó

96. Trang 48: Mỗi node thuộc lớp ẩn nhận vào ma trận đầu vào từ lớp trước và kết hợp với trọng số để ra được kết quả

Độ trùng lặp: 79%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nhận vào ma trận đầu vào từ layer trước và kết hợp với trọng số để ra được kết quả

97. Trang 50: Hàm kích hoạt (activation function) hoặc Hàm truyền được sử dụng nhằm thay đổi các tham số đầu vào sang một khoảng giá trị khác

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hàm truyền Hàm kích hoạt hay còn gọi là Hàm truyền có chức năng chuyển đổi thông số đầu vào sang một khoảng giá trị khác

98. Trang 51: Ban đầu, phương pháp này sẽ bắt đầu từ một điểm ngẫu nhiên trên hàm số và sau đó dịch chuyển điểm này theo hướng giảm dần của đạo hàm cho đến khi đến được điểm cực tiểu

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: một điểm ngẫu nhiên trên hàm số và sau đó điểm này sẽ được di chuyển theo chiều giảm của đạo hàm cho đến khi đạt ĐÀU KHÍ SỐ / THẨM DÒ KHAI THÁC ĐẦU KHÍ đến điểm cực tiểu

99. Trang 51: Nếu tốc độ học tập nhỏ, thì thuật toán sẽ phải thực hiện nhiều bước lặp để hàm số đến được điểm cực tiểu.

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nhỏ thuật toán sẽ phải thực hiện nhiều bước để

100. Trang 54: để tối ưu các trọng số ta thực hiện gradient descent để làm được điều này, ta cần tính được đạo hàm của trọng số và độ lệch

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Để làm được điều này ta cần tính được đạo hàm của hàm

101. Trang 59: False Positive (FP) đối tượng ở lớp Negative, mô hình dự đoán Positive

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: False Positive (FP) Đối tượng ở lớp Negative, mô hình phân Đối tượng vào lớp Positive (dự đoán

102. Trang 59: False Negative (FN) đối tượng ở lớp Positive, mô hình dự đoán Negative

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: False Negative (FN) Đối tượng ở lớp Positive, mô hình phân Đối tượng vào lớp Negative (dự đoán

103. Trang 59: Bốn loại trường hợp trên thường được biểu diễn dưới dạng ma trận hỗn loạn hay còn gọi là confusion matrix

Độ trùng lặp: 63%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Bốn trường hợp trên thường được biểu diễn dưới dạng ma trận hỗn loạn, (confusion matrix)

104. Trang 59: Trong thực tế có ba chỉ số để đánh giá một mô hình theo dạng dự đoán đó là Accuracy, Precision và Recall

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trong thực tế, có ba độ đo chủ yếu để đánh giá một mô hình phân loại là accuracy, precision and recall accuracy, được định nghĩa là tỷ lệ phần trăm dự đoán

105. Trang 59: Accuracy được định nghĩa là tỉ lệ phần trăm dự đoán đúng trên tổng số lượng dữ liệu thử nghiệm

Độ trùng lặp: 63%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Accuracy được định nghĩa là tỷ lệ phần trăm dự đoán đúng cho dữ liệu thử nghiệm

106. Trang 59: Precision được định nghĩa là phần nhỏ của các ví dụ có liên quan trong số tất cả các ví dụ được dự đoán thuộc một lớp nhất định

Độ trùng lặp: 95%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Precision được định nghĩa là phần nhỏ của các ví dụ có liên quan (tích cực thực sự) trong số tất cả các ví dụ được dự đoán là thuộc một lớp nhất định. Recall được định nghĩa là phần nhỏ của các ví dụ được dự đoán thuộc

107. Trang 59: Recall được định nghĩa là phần nhỏ của các ví dụ được dự đoán thuộc một lớp so với tất cả ví dụ thực sự thuộc lớp đó

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: định Recall được định nghĩa là phần nhỏ của các ví dụ được dự đoán thuộc về một lớp so với tất cả các ví dụ thực sự thuộc

108. Trang 62: ngôn ngữ lập trình máy tính bậc cao Python thường được sử dụng để xây dựng trang web và phát triển phần mềm, thực hiện phân tích dữ liệu và tự động hóa các nhiệm vụ

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ngôn ngữ lập trình máy tính bậc cao thường được sử dụng để xây dựng trang web và phần mềm, tự động hóa các tác vụ và tiến hành phân tích dữ liệu Python.

109. Trang 62: Nó được Guido van Rossum phát triển vào cuối những năm 1980 tại Viện nghiên cứu quốc gia về toán học và khoa học máy tính ở Hà Lan

Độ trùng lặp: **89%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được Guido van Rossum phát triển vào cuối những năm tám mươi và đầu những năm chín mươi tại Viện Nghiên cứu Quốc gia về Toán học và Khoa học Máy tính ở Hà Lan

110. Trang 62: Python là ngôn ngữ có mục đích chung, có khả năng tạo ra nhiều chương trình khác nhau và không chuyên biệt cho bất kỳ vấn đề cụ thể nào

Độ trùng lặp: **81%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ngôn ngữ có mục đích chung, nghĩa là nó có thể được sử dụng để tạo nhiều chương trình khác nhau và không chuyên biệt cho bất kỳ vấn đề cụ thể nào Python là ngôn ngữ

111. Trang 63: NumPy (Numeric Python) là một thư viện Toán học rất phổ biến và mạnh mẽ trong Python

Độ trùng lặp: **89%**

Nguồn: *Dữ liệu nội sinh*

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

112. Trang 63: thư viện này được trang bị các hàm số đã được tối ưu, cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là xử lý dữ liệu ma trận và mảng, lớn với tốc độ nhanh hơn nhiều so với việc sử dụng Python đơn thuần

Độ trùng lặp: **82%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được trang bị các hàm số đã được tối ưu, 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 Python đơn thuần

113. Trang 63: numpy là một trong những thư viện quan trọng của python, đặc biệt là trong nghiên cứu về các phép toán số học

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *là một trong những thư viện hữu ích nhất của Python, đặc biệt là*

114. Trang 63: Năm 2005, Travis Oliphant đã tạo ra thư viện NumPy bằng cách hợp nhất các tính năng của Numarray và thư viện Numeric

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Năm 2005, Travis Oliphant đã tạo ra gói NumPy bằng cách kết hợp các tính năng của Numarray và gói Numeric*

115. Trang 63: Sử dụng NumPy, lập trình viên có thể thực hiện các thao tác sau

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Sử dụng NumPy, lập trình viên có thể thực hiện các thao tác sau*

116. Trang 63: Các biến đổi Fourier và Các quy trình để thao tác shape

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Các biến đổi Fourier và Các quy trình để thao tác shape*

117. Trang 63: NumPy tích hợp sẵn các hàm cho đại số tuyến tính và tạo số ngẫu nhiên

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *NumPy tích hợp sẵn các hàm cho đại số tuyến tính và tạo số ngẫu nhiên*

118. Trang 63: NumPy thường được sử dụng kết hợp với các gói như SciPy (Python Scientific) và Matplotlib (thư viện vẽ đồ thị)

Độ trùng lặp: **90%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *NumPy thường được sử dụng cùng với các gói như SciPy (Python Scientific) và Matplotlib (thư viện vẽ đồ thị) Sự kết hợp*

119. Trang 63: Sự kết hợp này rất phổ biến để thay thế cho MatLab, một nền tảng phổ biến trong tính toán kỹ thuật

Độ trùng lặp: **61%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *để thay thế cho MatLab, một nền tảng phổ biến cho tính toán kỹ thuật*

120. Trang 63: Điều quan trọng là Numpy là một thư viện mã nguồn mở, và miễn phí Trong khi MatLab là một thư viện mã nguồn đóng và yêu cầu phải trả phí

Độ trùng lặp: **72%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *là NumPy là một thư viện mã nguồn mở miễn phí, so với MatLab là một thư viện*

mã nguồn đóng và phải trả phí.

121. Trang 63: thư viện pandas trong python là một thư viện mã nguồn mở, mạnh mẽ được sử dụng rộng rãi để xử lý và phân tích dữ liệu

Độ trùng lặp: 74%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: viện pandas trong Python là một Thư viện mã nguồn mở hỗ trợ đắc lực trong thao tác dữ liệu pandas là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ, của ngôn ngữ lập trình Python Thư viện pandas được sử dụng rộng rãi

122. Trang 63: Đây là một công cụ thiết yếu trong ngôn ngữ lập trình python được áp dụng rộng rãi trong cả nghiên cứu và phát triển các ứng dụng về lĩnh vực khoa học dữ liệu

Độ trùng lặp: 62%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: dùng rộng rãi trong cả nghiên cứu lẫn phát triển các ứng dụng về khoa học dữ liệu

123. Trang 63: Thư viện này sử dụng cấu trúc dữ liệu Chính là Dataframe và cung cấp nhiều chức năng để xử lý và làm việc trên cấu trúc này

Độ trùng lặp: 70%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Thư viện này sử dụng một cấu trúc dữ liệu riêng là DataFrame Pandas cung cấp rất nhiều chức năng xử lý và làm việc trên cấu trúc dữ liệu này chính

124. Trang 64: TensorFlow là một thư viện mã nguồn mở end to end được thiết kế chủ yếu cho các ứng dụng machine learning

Độ trùng lặp: 71%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: TensorFlow là một thư viện mã nguồn mở end to end được tạo ra chủ yếu dành cho các ứng dụng

125. Trang 64: Đó là một thư viện toán học ký hiệu áp dụng luồng dữ liệu và lập trình có khả năng phân biệt để triển khai các nhiệm vụ đa dạng, chủ trọng vào đào tạo và suy luận các mạng nơ ron sâu (deep neural network)

Độ trùng lặp: 68%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: là một thư viện toán học ký hiệu sử dụng luồng dữ liệu và lập trình có thể phân biệt để thực hiện các nhiệm vụ khác nhau, tập trung vào đào tạo và suy luận các mạng nơ ron sâu (deep neural network)

126. Trang 64: các nhà phát triển có thể xây dựng các ứng dụng học máy bằng cách sử dụng các công cụ, thư viện và tài nguyên từ nhiều cộng đồng khác nhau

Độ trùng lặp: 73%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Các nhà phát triển tạo Các ứng dụng học máy bằng cách sử dụng Các công cụ.

thư viện và tài nguyên cộng đồng khác nhau

127. Trang 64: Google tích hợp công nghệ máy học này vào các sản phẩm của mình nhằm tối ưu hóa công cụ tìm kiếm, dịch, thuật, chú thích hình ảnh và các đề xuất khác

Độ trùng lặp: 50%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các sản phẩm của mình để cải thiện công cụ tìm kiếm, bản dịch chú thích hình ảnh

128. Trang 64: Khi nhập từ khóa vào thanh tìm kiếm, Google sẽ đề xuất các từ khóa liên quan và có thể tiếp theo để giúp người dùng tìm kiếm, hiệu quả hơn

Độ trùng lặp: 50%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: người dùng nhập từ khóa vào thanh tìm kiếm, Google sẽ đưa ra các đề xuất về từ tiếp theo có thể

129. Trang 64: Google muốn áp dụng công nghệ máy học để khai thác bộ dữ liệu khổng lồ của họ nhằm mang đến trải nghiệm tối ưu cho người dùng

Độ trùng lặp: 66%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: dùng công nghệ máy học để tận dụng bộ dữ liệu khổng lồ của họ nhằm mang đến cho người dùng trải nghiệm

130. Trang 64: Ba nhóm người sử dụng chủ yếu công nghệ này bao gồm Các nhà nghiên cứu nhà khoa học dữ liệu và Lập trình viên

Độ trùng lặp: 54%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các nhà nghiên cứu, các nhà khoa học dữ liệu lập trình viên

131. Trang 64: Tất cả đều có thể sử dụng cùng một bộ công cụ để hợp tác và tối đa hóa hiệu quả công việc của mình

Độ trùng lặp: 55%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: đều có thể sử dụng cùng một bộ công cụ để

132. Trang 64: TensorFlow là một thư viện do Nhóm Google Brain phát triển nhằm gia tăng tốc độ trong lĩnh vực học máy và nghiên cứu mạng nơ ron sâu

Độ trùng lặp: 70%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: TensorFlow là một thư viện do Nhóm Google Brain phát triển để tăng tốc công nghệ học máy và nghiên cứu mạng nơ ron sâu

133. Trang 65: Tensorflow được đặt tên là như vậy vì nó nhận đầu vào dưới dạng một mảng đa chiều, được gọi là tensor

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: TensorFlow vì nó nhận đầu vào là một mảng đa chiều, còn được gọi là Tensor

134. Trang 65: Bạn có thể xây dựng một lưu đồ hoạt động (hay còn gọi là đồ thị Graph) để thực hiện các hoạt động trên đầu vào đó

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Bạn có thể xây dựng một loại lưu đồ (flowchart lưu đồ hay còn gọi là sơ đồ quy trình) hoạt động (được gọi là Graph) mà Bạn muốn thực hiện trên đầu vào đó

135. Trang 65: Đây là lý do vì sao TensorFlow được đặt tên như thế, vì tensor đi qua nó và chảy qua một danh sách các hoạt động, trước khi ra phía bên kia

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nó chảy qua một danh sách các hoạt động và sau đó nó đi ra phía bên kia

136. Trang 65: Vào cuối năm 2010, các nhà nghiên cứu tại Stanford đã phát hiện rằng GPU rất giỏi trong các phép toán ma trận và đại số, giúp thực hiện các loại tính toán này một cách nhanh chóng

Độ trùng lặp: 80%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Vào cuối năm 2010, các nhà nghiên cứu tại Stanford đã phát hiện ra rằng GPU cũng rất giỏi trong các phép toán ma trận và đại số, do đó nó giúp chúng thực hiện các loại tính toán này rất nhanh

137. Trang 65: Ngoài ra, TensorFlow có thể được truy cập và điều khiển bởi nhiều ngôn ngữ khác, nhau, chủ yếu là Python

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ngoài ra, TensorFlow cũng có thể được truy cập và điều khiển bởi các ngôn ngữ khác chủ yếu là Python

138. Trang 65: Keras là một thư viện mạng nơ ron sâu, trong Python hỗ trợ đa dạng các kiểu mạng nơ ron và có khả năng xuất sắc trong xử lý dữ liệu, trực quan hóa và các tính năng khác

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Keras là một thư viện Python, dành cho mạng nơ ron chuyên sâu và nó cung cấp khả năng hỗ trợ tuyệt vời cho việc xử lý dữ liệu, trực quan hóa và nhiều tác vụ khác Keras được sử dụng phổ biến trong lĩnh vực học máy và thị giác máy tính thư viện này hỗ trợ một loạt các mạng nơ ron và

139. Trang 67: Hiển thị biểu đồ heat map thể hiện mức độ tương quan giữa các biến trong toàn bộ tập dữ liệu

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tương quan giữa các biến trong mô hình Từ *heat map* có thể thành lập biểu đồ thể hiện mức độ tương quan

140. Trang 70: Cụ thể thực hiện chuẩn hóa để cho giá trị trung bình gần bằng 0 và độ lệch chuẩn gần bằng 1

Độ trùng lặp: 68%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giá trị trung bình gần bằng 0, và độ lệch chuẩn gần bằng 1 (cụ thể,

141. Trang 73: 14 chênh lệch giữa giá trị thực tế và giá trị dự đoán của mô Hình SVM ANN

Độ trùng lặp: 72%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chênh lệch giữa giá trị dự đoán và giá trị thực tế

142. Trang 76: Tuy nhiên, mô hình mạng nơ ron nhân tạo đã cho thấy hiệu suất cao nhất, với độ chính xác cao nhất. Trong số ba mô hình

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nơ ron nhân tạo cho kết quả chính xác hơn so với phương pháp hồi quy trong ba mô hình dự đoán bằng mạng nơ ron đề xuất, mô hình 3 có độ chính xác cao nhất

--- Hết ---