



BÁO CÁO KIỂM TRA TRÙNG LẶP

Thông tin tài liệu

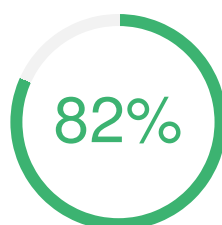
Tên tài liệu: DATN_Trần Tuấn Đạt_1851061821
Tác giả: Trương Xuân Nam
Điểm trùng lặp: 18
Thời gian tải lên: 17:56 06/07/2025
Thời gian sinh báo cáo: 18:01 06/07/2025
Các trang kiểm tra: 86/86 trang



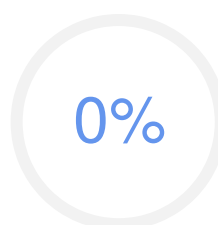
Kết quả kiểm tra trùng lặp



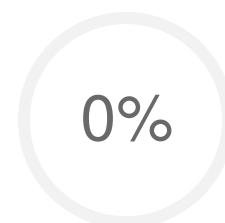
Có 18% nội dung trùng
lặp



Có 82% 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 machinelearningcoban.com tailieu.vn

Danh sách các câu trùng lặp

Câu 1. Trang 3: NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP Họ tên sinh viên Trần Tuấn Đạt Hệ đào tạo Đại học chính quy Lớp 60TH1 Ngành Công nghệ thông tin Khoa Công nghệ thông tin

Độ trùng lặp: **80%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP Họ tên sinh viên Hệ đào tạo Đại học chính quy Lớp DHCTTCK Z Ngành Công nghệ thông tin Khoa Công nghệ

Câu 2. Trang 4: 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

Câu 3. Trang 4: 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

Câu 4. Trang 5: hàng năm có hơn 400 000 sinh viên tốt nghiệp Tuy nhiên, có đến một nửa trong số đó ra trường không có việc làm hoặc làm việc trái với chuyên ngành

Độ trùng lặp: **62%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hàng năm có hơn 400 000 sinh viên đại học tốt nghiệp tuy nhiên hơn một nửa trong số sinh viên ra trường không tìm được việc làm hoặc làm các công việc ngoài chuyên

Câu 5. Trang 7: 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

Câu 6. Trang 7: các kết quả trong đồ án tốt nghiệp này là trung thực từ trong quá trình nghiên cứu giám sát và tiến hành thực hiện

Độ trùng lặp: **58%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đồ án tốt nghiệp này là công trình nghiên cứu, thực sự của cá nhân tôi, Các số liệu và kết quả nghiên cứu, nêu trong Đồ án này là trung thực do cá nhân tôi tiến hành thực

Câu 7. Trang 7: Việc tham khảo các nguồn tài liệu 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: Việc tham khảo các nguồn tài liệu đã được thực hiện trích dẫn và ghi nguồn tài

liệu tham khảo đúng quy định

Câu 8. Trang 8: Trong thời buổi hiện đại ngày nay, công nghệ thông tin cũng như những ứng dụng của nó không ngừng phát triển, các thuật ngữ như Big Data, Internet of Things, AI, Data mining đã trở nên quá quen thuộc

Độ trùng lặp: **54%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trong thời buổi hiện đại ngày nay, công nghệ thông tin cũng như những ứng dụng của nó không ngừng phát triển.

Câu 9. Trang 9: trình bày tập dữ liệu được sử dụng và cách thức xử lý dữ liệu đầu vào để phục vụ cho việc huấn luyện mô hình

Độ trùng lặp: **54%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: để phục vụ cho việc huấn

Câu 10. Trang 10: Trước tiên, em xin bày tỏ lòng biết ơn sâu sắc tới Ban giám hiệu trường. Đại học Thủy. Lợi cùng Ban chủ nhiệm khoa Công nghệ Thông tin đã luôn quan tâm, tạo điều kiện thuận lợi cho em trong suốt quá trình học tập và rèn luyện tại trường.

Độ trùng lặp: **61%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trường Ban Chủ nhiệm Khoa Công nghệ sinh học và các thầy cô giáo đã truyền đạt cho em những kiến thức quý báu trong suốt thời gian học tập và rèn luyện tại Trường Đại học Nông nghiệp Hà Nội em xin bày tỏ lòng kính trọng và biết ơn sâu sắc tới ThS Đặng Thị Thanh tâm, giảng viên bộ môn Công nghệ sinh học thực vật đã tận tình hướng dẫn và dạy dỗ em trong suốt quá trình học tập cũng như nghiên cứu em xin bày tỏ lòng biết ơn sâu sắc và

Câu 11. Trang 10: Trong suốt bốn năm học tập tại trường Đại học Thủy Lợi, em đã nhận được rất nhiều, sự hỗ trợ chỉ dẫn tận tình từ các thầy cô, giáo và sự giúp đỡ quý báu từ bạn bè

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: sự hỗ trợ, giúp đỡ dù ít hay nhiều dù gián tiếp hay trực tiếp từ người khác Trong suốt bốn năm học tập tại Trường Đại học Ngoại Ngữ Đại học Đà Nẵng, em đã nhận được rất nhiều sự giúp đỡ quan tâm của quý thầy cô gia đình, bạn bè

Câu 12. Trang 10: em xin gửi lời cảm ơn chân thành và sâu sắc nhất tới tập thể giảng viên Khoa Công nghệ thông tin, Trường Đại học Thủy Lợi, những người đã Truyền đạt cho chúng em những kiến thức quý báu, là nền tảng để em hoàn thành đồ án này

Độ trùng lặp: **63%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Khoa Công nghệ Thông tin Trường Đại học Công nghệ Thông tin và truyền Thông Đại học Thái Nguyên, những người đã tận tình giảng dạy, trang bị cho chúng Em những vốn kiến thức và kinh nghiệm quý báu, cung cấp cho chúng Em những điều kiện và môi Trường học tập tốt nhất Em xin bày tỏ lòng biết ơn sâu sắc nhất tới cô giáo Lưu Thị Liễu, người đã tận tình hướng dẫn, chỉ bảo Em trong suốt thời gian thực hiện đồ án Nhân dịp này Em xin gửi lời cảm ơn chân thành tới gia đình, bạn bè, những người thân đã cổ vũ, động viên tiếp thêm cho Em nghị lực để Em hoàn thành đồ án

Câu 13. Trang 10: Hoàng Quốc Dũng người đã tận tình hướng dẫn, định hướng và đồng hành cùng em trong suốt quá trình thực hiện đề tài, "Dự đoán cơ hội việc làm cho sinh viên sau tốt nghiệp"

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *đồng hành cùng em trong suốt quá trình thực hiện đề tài luôn tận tình hướng dẫn.*

Câu 14. Trang 10: Em cũng xin gửi lời cảm ơn đến lãnh đạo nhà Trường cùng các Phòng Ban chức năng đã trực tiếp và gián tiếp hỗ trợ Em trong quá trình học tập và thực hiện đề tài

Độ trùng lặp: 74%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *đến ban lãnh đạo trường Đại học cùng các Khoa, phòng ban chức năng đã trực tiếp và gián tiếp hỗ trợ Em trong suốt quá trình học tập và thực hiện đề tài*

Câu 15. Trang 10: Do hạn chế về thời gian và kinh nghiệm thực tiễn, bài báo cáo không tránh khỏi những thiếu sót

Độ trùng lặp: 89%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Do hạn chế về thời gian và kinh nghiệm thực tế nên báo cáo không tránh khỏi những thiếu sót Em rất mong được thầy cô và các bạn góp ý để bài báo cáo*

Câu 16. Trang 10: em kính mong nhận được sự chỉ bảo, góp ý quý báu từ các thầy, cô để em có thể hoàn thiện hơn về kiến thức cũng như kỹ năng, phục vụ tốt cho công việc sau này

Độ trùng lặp: 57%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *mong nhận được sự chỉ bảo, góp ý của các thầy cô giáo để Em có thể hoàn thiện được kiến thức cũng như có thể đáp ứng được cho công việc sau này*

Câu 17. Trang 17: DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ AI Trí tuệ nhân tạo (Artificial Intelligence)

Độ trùng lặp: 79%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH THUẬT NGỮ AI (Artificial Intelligence) Trí tuệ nhân tạo*

Câu 18. Trang 17: DFS một tổ chức phi lợi nhuận độc lập (Django Software Foundation)

Độ trùng lặp: 81%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Một tổ chức phi lợi nhuận độc lập)*

Câu 19. Trang 17: Python là một ngôn ngữ lập trình được sử dụng rộng rãi trong các ứng dụng web, phát triển phần mềm, khoa học dữ liệu và máy học (ML)

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: PYTHON là một ngôn ngữ lập trình được sử dụng rộng rãi trong các ứng dụng web, phát triển phần mềm, khoa học dữ liệu và máy học (ML)

Câu 20. Trang 17: ID3 là một thuật toán do Ross Quinlan phát minh, được sử dụng để tạo cây quyết định từ một tập dữ liệu

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ID3 là một thuật toán được John Australia) phát minh để tạo Decision tree từ một tập dữ liệu

Câu 21. Trang 17: C4 5 là một thuật toán được sử dụng để tạo cây quyết định do Ross Quinlan phát triển và là một phần mở rộng của thuật toán ID3

Độ trùng lặp: 73%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: 5 là một thuật toán được sử dụng để tạo cây quyết định được phát triển bởi Ross Quinlan C4 5 là phần mở rộng của thuật toán ID3 từ đó của Quinlan Các cây quyết định được tạo bởi C4 5 có thể được sử dụng để phân loại và vì lý do

Câu 22. Trang 17: RMSE Lỗi trung bình bình phương gốc (Root Mean Square Error)

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lỗi trung bình bình phương gốc (Root Mean Square Error RMSE)

Câu 23. Trang 17: GUI Giao diện đồ họa người dùng (Graphical User Interface)

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giao diện đồ họa người dùng

Câu 24. Trang 17: URL một loại Mã nhận dạng tài nguyên thống nhất (Uniform Resource Locator)

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: một loại mã nhận dạng tài nguyên thống nhất (Uniform Resource Identifier) Cấu trúc cơ bản của một liên kết URL

Câu 25. Trang 17: Data mining là hành động tự động tìm kiếm các kho thông tin lớn để tìm ra các xu hướng và các mẫu vượt ra ngoài các quy trình phân tích đơn giản

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: DATA MINING là hành động tự động tìm kiếm các kho thông tin lớn để tìm ra các xu hướng và các mẫu vượt ra ngoài các quy trình phân tích đơn giản

Câu 26. Trang 17: web còn gọi là trang website hoặc trang mạng, và nội dung liên quan được xác định bằng một tên miền chung và được xuất bản trên ít nhất một máy chủ web

Độ trùng lặp: **92%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: còn gọi là trang WEB hoặc trang mạng, và nội dung liên quan được xác định bằng một tên miền chung và được xuất bản trên ít nhất một máy chủ WEB (Theo Wikipedia) một website

Câu 27. Trang 18: lực lượng lao động, trẻ, được đào tạo từ các trường đại học, cao đẳng, là những người năng động, có trình độ và đóng vai trò quan trọng trong sự phát triển xã hội

Độ trùng lặp: **56%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: Lực lao động trong nền kinh tế nguồn Lực lao động quan trọng có trình độ, là những sinh viên được đào tạo trong các trường đại học, cao đẳng đây chính là Lực lượng lao động rất năng động có trình độ, và

Câu 28. Trang 18: Tuy nhiên, tình trạng thất nghiệp của sinh viên sau khi tốt nghiệp đã và đang tác động tiêu cực đến nền kinh tế xã hội

Độ trùng lặp: **53%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: đến tình trạng thất nghiệp của sinh viên sau khi tốt nghiệp Vấn đề xã hội mà gần như không có trong nền kinh tế

Câu 29. Trang 18: Qua nghiên cứu, có thể chỉ ra Một số Nguyên nhân chính như hoạt động sản xuất, kinh doanh của nhiều doanh nghiệp gặp khó khăn, dẫn đến việc cắt giảm nhu cầu tuyển dụng; các cơ quan, nhà nước, ngày càng có yêu cầu cao hơn Đối với chất lượng tuyển dụng công chức viên chức; sự gia tăng nhanh chóng số lượng cơ sở đào tạo khiến nguồn cung sinh viên vượt xa nhu cầu của thị trường lao động

Độ trùng lặp: **52%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: hoạt động sản xuất, kinh doanh của nhiều DN gặp khó khăn nhu cầu tuyển dụng lao động bị thu hẹp đối với các cơ quan tổ chức, nhà nước nhu cầu tuyển dụng công chức, viên chức, ngày càng có yêu cầu cao hơn về chất lượng. Bởi thế, không chỉ với các trường đại học ngoài công lập, ngay cả các trường đại học công lập có danh tiếng, không phải sinh viên nào tốt nghiệp ra trường cũng có ngay được công việc như mong muốn một nguyên nhân khác, việc có thêm nhiều cơ sở đào tạo (trường cao đẳng, đại học) ra đời dẫn đến số lượng sinh viên được đào tạo ở cùng các ngành, chuyên ngành ngày càng nhiều dẫn đến cung

Câu 30. Trang 18: Trên Thực tế mỗi năm, có rất nhiều sinh viên tốt nghiệp nhưng tỷ lệ đáp ứng được yêu cầu của nhà tuyển dụng vẫn còn thấp

Độ trùng lặp: **50%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: đáp ứng yêu cầu của thị trường, của nhà tuyển dụng Từ đó, số lượng người thất nghiệp ngày càng nhiều và càng có dấu hiệu ra tăng thực tế, cho thấy, trong hàng nghìn sinh viên mới ra trường mỗi năm chỉ có một bộ phận số ít sinh viên được lọt vào mắt những nhà tuyển dụng bởi lẽ, các sinh viên mới ra trường hầu như chưa có kỹ năng để làm việc Khi còn đi học, nhiều sinh viên cho rằng để có thể đáp ứng được yêu cầu

Câu 31. Trang 18: Trước những thách thức đó, các cơ sở đào tạo và bản thân sinh viên cần có cái nhìn mới chủ động hơn trong việc chuẩn bị cho cơ hội việc làm sau tốt nghiệp

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trước những thách thức nêu trên, đòi hỏi các cơ sở đào tạo và bản thân sinh viên phải có cách nhìn nhận mới, về vấn đề việc làm sau khi tốt nghiệp ra trường Do đó,

Câu 32. Trang 18: Nhận thức được vai trò then chốt của việc đào tạo nguồn nhân lực phù hợp với yêu cầu thực tiễn, trong những năm gần đây, các trường Đại học tại Việt Nam và trên thế giới đã thực hiện nhiều giải pháp nhằm nâng cao chất lượng đào tạo như Cập nhật chương trình, giáo trình, theo hướng hiện Đại đổi mới, phương pháp giảng, dạy tăng cường ứng dụng công nghệ thông tin chú trọng đào tạo kỹ năng mềm ngoại ngữ, tin học và khuyến khích sinh viên tham gia nghiên cứu khoa học

Độ trùng lặp: **53%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: việc đào tạo nguồn nhân lực đáp ứng với yêu cầu trong tình hình mới trong những năm qua, trường đại học công đoàn đã thực hiện nhiều giải pháp nhằm nâng cao chất lượng đào tạo như cập nhật chương trình, giáo trình, đào tạo theo hướng hiện đại tích cực đổi mới phương pháp dạy, học đẩy mạnh ứng dụng công nghệ thông tin, trong soạn giảng chú trọng giáo dục cho sinh viên những kỹ năng mềm,

Câu 33. Trang 19: kiểm tra độ chính xác của mô Hình bằng Phương pháp K Fold Cross Validation

Độ trùng lặp: **78%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: độ chính xác của mô hình phân lớp với phương pháp holdout phương pháp K Fold Cross Validation

Câu 34. Trang 19: một phiên bản nâng cao của phương pháp Hold Out chia dữ liệu thành K tập luân phiên sử dụng một tập để kiểm tra và K 1 tập còn lại để huấn luyện,

Độ trùng lặp: **52%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: để kiểm tra và K 1

Câu 35. Trang 19: đánh giá hiệu quả mô hình dựa trên các chỉ số như Precision, Recall, và F1 score,

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đánh giá các mô hình dự đoán Kết quả Bài nghiên cứu so sánh hiệu suất của các mô hình dựa trên các chỉ số như Accuracy, Precision, Recall F1 score và

Câu 36. Trang 20: đánh giá Mô Hình bằng Các chỉ số độ chính xác (Accuracy), Precision, Recall, F1 Score,

Độ trùng lặp: **73%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các chỉ số Đánh giá phổ biến bao gồm độ chính xác (Accuracy), Precision, Recall, F1 score và AUC 4 Điều Chính mô hình

Câu 37. Trang 21: Trong những năm gần đây, cùng VỚI sự bùng nổ CỦA cuộc cách mạng công nghiệp 4 0, các thuật ngữ như TRÍ TUỆ NHÂN TẠO (Artificial Intelligence AI), HỌC MÁY (Machine Learning) và HỌC SÂU (Deep Learning) đã dần trở nên phổ biến trở thành những khái niệm quen thuộc Trong thời đại công nghệ hiện nay

Độ trùng lặp: 72%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của học sâu với học máy và trí tuệ nhân tạo Trong những năm vừa qua, cùng với sự bùng nổ của cuộc Cách mạng Công nghiệp 4 0, các thuật ngữ như trí tuệ nhân tạo (AI), học máy (Machine Learning) và học sâu (Deep Learning) đang dần trở nên phổ biến, và trở thành những khái niệm

Câu 38. Trang 21: Machine learning là một công nghệ phát triển từ lĩnh vực trí tuệ nhân tạo (AI)

Độ trùng lặp: 91%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một công nghệ phát triển từ lĩnh vực trí tuệ nhân tạo Các thuật toán Machine learning là

Câu 39. Trang 21: Học Máy là một nhánh của trí tuệ nhân tạo (AI) và khoa Học Máy tính (Computer Science), phương pháp phân tích dữ liệu để tự động hóa việc xây dựng mô hình phân tích từ đó bắt chước cách con người Học dần dần cải thiện độ chính xác của nó mà không cần sự can thiệp hay trợ giúp của con người

Độ trùng lặp: 73%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: phương pháp phân tích dữ liệu để tự động hóa việc xây dựng mô hình phân tích Đây là một nhánh của trí tuệ nhân tạo (Artificial Intelligence AI) và khoa máy Học tính (Computer Science) tập trung vào việc sử dụng dữ liệu và thuật toán để bắt chước cách con người Học dần dần cải thiện độ chính xác của nó với sự can thiệp tối thiểu của con

Câu 40. Trang 21: Học máy là một thành phần quan trọng của lĩnh vực khoa Học dữ liệu đang phát triển

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Học máy là một thành phần quan trọng của lĩnh vực khoa Học dữ liệu đang phát triển

Câu 41. Trang 21: học máy sử dụng các phương pháp thống kê, và kỹ thuật xử lý dữ liệu để huấn luyện các mô hình phân tích giúp máy tính có thể đưa ra dự đoán hoặc phân loại dựa trên dữ liệu mẫu (training data) hoặc từ kinh nghiệm (những gì đã học được)

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: xử lý dữ liệu và đưa ra các quyết định dựa trên dữ liệu Dưới đây là những công việc phổ biến mà bạn có thể theo đuổi sau khi Học ngành này 1 Nhà khoa Học dữ liệu (Data Scientist) Vai trò Nhà khoa Học dữ liệu chịu trách nhiệm phân tích, dữ liệu thô để tìm ra những mẫu (pattern), xu hướng, và mối quan hệ ẩn trong dữ liệu Họ sử dụng các phương pháp thống kê thuật toán Học máy (machine learning) và công cụ dữ liệu để phát triển các mô hình dự đoán hoặc phân loại

Câu 42. Trang 22: Ngày nay, học máy đóng vai trò then chốt trong lĩnh vực Khoa học dữ liệu (Data Science)

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: khoa học dữ liệu (Data Science) Nếu Business Intelligence đóng vai trò then chốt trong

Câu 43. Trang 22: các thuật toán học máy (Machine Learning) được huấn luyện trên một tập dữ liệu, đầu vào gọi là dữ liệu, huấn luyện (training data), nhằm xây dựng một mô hình có khả năng phân tích và dự đoán

Độ trùng lặp: 67%

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 mới cho Các dữ liệu chưa từng được nhìn thấy trước đó Các thuật toán học

Câu 44. Trang 22: Khi tiếp nhận dữ liệu mới, Mô hình này sẽ áp dụng những gì đã học được từ dữ liệu quá khứ để đưa ra kết quả dự đoán hoặc phân loại

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: học được từ dữ liệu huấn luyện để đưa ra dự đoán hoặc phân loại

Câu 45. Trang 23: chuẩn bị dữ liệu sau khi Thu thập dữ liệu cần tiến hành các bước xử lý như làm sạch dữ liệu loại bỏ các thuộc tính dư thừa, mã hóa dữ liệu dạng chuỗi thành số gán nhãn trích xuất đặc trưng, chuẩn hóa / rút gọn dữ liệu để đảm bảo dữ liệu đầu vào phù hợp với yêu cầu của mô hình.

Độ trùng lặp: 50%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thu thập dữ liệu máy tính chỉ có khả năng học tập khi được cung cấp đầy đủ các dữ liệu đầu vào (dataset), các dữ liệu đầu vào phải đảm bảo các yêu tố chính xác và sát với nhu cầu giải quyết vấn đề của mô hình không trộn lẫn với các dữ liệu sai, được gán nhãn, một cách cẩn thận B2 Tiền xử lý bước này dùng để Chuẩn hóa dữ liệu loại bỏ các thuộc tính không cần thiết, gán nhãn, dữ liệu mã hóa một số, đặc trưng, trích xuất đặc trưng, rút gọn dữ liệu nhưng

Câu 46. Trang 23: huấn luyện mô hình Tùy vào loại bài toán và đặc điểm của dữ liệu ta Lựa chọn thuật toán học máy phù hợp

Độ trùng lặp: 53%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lựa chọn thuật toán học máy phù hợp với tính chất dữ liệu, và yêu cầu bài toán Bước 4 Xây dựng và Huấn luyện mô hình

Câu 47. Trang 23: dữ liệu sau xử lý sẽ được chia thành hai phần tập huấn luyện (training set) và tập kiểm tra (test set)

Độ trùng lặp: 66%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được chia thành hai tập tập huấn luyện (training set) và tập kiểm tra (test set) tập huấn luyện chứa các điểm Dữ liệu

Câu 48. Trang 23: mô hình sẽ được huấn luyện trên tập huấn luyện từ đó có khả năng đưa ra dự đoán cho dữ liệu mới

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Mô hình sẽ được huấn luyện trên tập dữ liệu đó để học cách nhận diện và đưa ra dự đoán

Câu 49. Trang 23: đánh giá mô hình Sau khi huấn luyện, mô hình được kiểm tra bằng tập dữ liệu kiểm tra nhằm đánh giá hiệu quả

Độ trùng lặp: 78%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đánh giá mô hình Sau khi huấn luyện, mô hình được kiểm tra bằng cách đưa dữ liệu kiểm tra vào mô hình và Đánh giá hiệu

Câu 50. Trang 23: Các chỉ số đánh giá phổ biến gồm Accuracy (độ chính xác), Precision, Recall, F1 score.

Độ trùng lặp: 82%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các chỉ số đánh giá phổ biến bao gồm độ chính xác (accuracy), độ chính xác (precision), độ nhạy (recall), F1 score.

Câu 51. Trang 23: trong học máy (Machine Learning), các phương pháp được phân loại dựa trên cách mà mô hình học hỏi từ dữ liệu

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: học hỏi từ dữ liệu mà không cần được lập trình một cách cụ thể Cốt lõi của Machine Learning là việc tạo ra các mô hình dự đoán hoặc phân loại dựa trên mẫu dữ liệu đầu vào Machine Learning được áp dụng rộng rãi Trong

Câu 52. Trang 23: Hai nhóm phương pháp phổ biến nhất là học có giám sát (Supervised learning), và học không giám sát (Unsupervised learning).

Độ trùng lặp: 85%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Học không giám sát (Unsupervised Learning) Học bán giám sát (Semi Supervised Learning) và Học củng cố (Reinforcement Learning) Học có giám sát (Supervised Learning) Học có giám sát là nhóm phương pháp phổ biến nhất

Câu 53. Trang 24: học, có giám sát là nhóm thuật toán phổ biến nhất trong các thuật toán Machine Learning

Độ trùng lặp: 87%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: sắt là nhóm phổ biến nhất trong các thuật toán Machine learning Một cách toán Học Học có giám sát là

Câu 54. Trang 24: Trong Học có giám sát, mô hình được huấn luyện trên tập dữ liệu đã gán nhãn, nghĩa là mỗi dữ liệu đầu vào đi kèm với nhãn, cho đầu ra cụ thể để suy luận ra quan hệ giữa đầu vào và đầu ra

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: có giám sát, mô hình được huấn luyện bằng cách sử dụng một tập dữ liệu đã được gán nhãn, có nghĩa là mỗi mẫu dữ liệu đầu vào (input) đều có một nhãn, tương ứng (output) xác định mô hình sẽ học cách xác định mối quan hệ giữa đầu vào và đầu ra

Câu 55. Trang 24: Hồi quy logistic (Logistic Regression) Mạng nơ-ron (Neural Networks), máy vector hỗ trợ (SVM) k láng giềng gần nhất (KNN)

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Neural Networks) Máy vector hỗ trợ (Support vector Machines SVM), K láng giềng gần nhất (

Câu 56. Trang 26: Khác với học có giám sát, Phương pháp này sử dụng dữ liệu chưa được gán nhãn

Độ trùng lặp: 81%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Khác với học có giám sát, học không giám sát, sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận phương pháp này

Câu 57. Trang 27: Đây là phương pháp kết hợp giữa học có giám sát và không giám sát có ứng dụng tương tự như học có giám sát

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Đây là phương pháp kết hợp giữa học không giám sát và học có giám sát

Câu 58. Trang 27: học bán giám sát đặt nền tảng trung gian giữa hiệu suất của học có giám sát và hiệu quả của học không giám sát

Độ trùng lặp: 88%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: bán giám sát đặt nền tảng trung gian giữa hiệu suất của Học máy có giám sát và hiệu quả của Học máy không giám sát

Câu 59. Trang 27: Học tăng cường (RL) là một kỹ thuật Học máy đào tạo phần mềm đưa ra quyết định nhằm thu về kết quả tối ưu nhất kỹ thuật này bắt chước quy trình Học thử và sai mà con người sử dụng để đạt được mục tiêu đã đặt ra

Độ trùng lặp: 90%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Học tăng cường (RL) là kỹ thuật máy Học (ML) giúp đào tạo phần mềm đưa ra

quyết định nhằm thu về kết quả tối ưu nhất kỹ thuật này bắt chước quy trình Học thử và sai mà con người sử dụng để đạt được mục tiêu đã đặt ra

Câu 60. Trang 27: Học tăng cường giúp phần mềm tăng cường các hành động hướng tới mục tiêu, đồng thời bỏ qua các hành động làm xáo lãng mục tiêu, thuật toán Học tăng cường sử dụng mô hình khen thưởng và trừng phạt trong quy trình xử lý dữ liệu

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giúp phần mềm tăng cường các hành động hướng tới mục tiêu, đồng thời bỏ qua các hành động làm xáo lãng mục tiêu, sử dụng mô hình khen thưởng và trừng phạt trong quy trình xử lý dữ liệu các Thuật toán

Câu 61. Trang 28: Các thuật toán này tiếp thu ý kiến phản hồi của từng hành động và tự khám phá ra con đường xử lý tốt nhất để thu về kết quả cuối cùng thuật toán RL còn có khả năng trì hoãn khen thưởng

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Các thuật toán này tiếp thu ý kiến phản hồi của từng hành động và tự khám phá ra con đường xử lý tốt nhất để thu về kết quả cuối cùng thuật toán RL còn có khả năng trì hoãn khen thưởng

Câu 62. Trang 28: Chiến lược tổng thể tốt nhất có thể đòi hỏi phải đánh đổi một vài lợi ích trước mắt, vì vậy cách tiếp cận tốt nhất mà RL khám phá ra có thể bao gồm một số trừng phạt hoặc giai đoạn quay lui RL là phương thức hiệu quả giúp hệ thống trí tuệ nhân tạo đạt kết quả tối ưu trong môi trường chưa biết

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chiến lược tổng thể tốt nhất có thể đòi hỏi phải đánh đổi một vài lợi ích trước mắt, vì vậy cách tiếp cận tốt nhất mà RL khám phá ra có thể bao gồm một số trừng phạt hoặc giai đoạn quay lui RL là phương pháp hiệu suất cao giúp hệ thống trí tuệ nhân tạo (AI) đạt kết quả tối ưu trong môi trường chưa biết

Câu 63. Trang 29: một số Ứng dụng cụ thể o phát hiện và nhận diện hình ảnh đây là một trong những Ứng dụng phổ biến nhất của học máy và trí tuệ nhân tạo

Độ trùng lặp: 65%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trong hình ảnh ứng dụng của nhận diện hình ảnh nhận diện hình ảnh có nhiều ứng dụng quan trọng trong nhiều lĩnh vực khác nhau Dưới Đây là Một số ứng dụng của nhận diện hình ảnh 1 nhận diện khuôn mặt nhận diện khuôn mặt là Một trong những ứng dụng phổ biến nhất của nhận diện hình ảnh

Câu 64. Trang 29: kỹ thuật này cho phép hệ thống nhận diện và xác Định các đặc trưng của đối tượng trong hình ảnh số

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của các đối tượng trong các ảnh bất kì Kỹ thuật này nhận biết các đặc trưng của đối tượng

Câu 65. Trang 30: o Chẩn đoán y tế. Học máy hỗ trợ bác sĩ trong việc Chẩn đoán bệnh bằng cách phân tích dữ liệu từ triệu chứng, xét nghiệm, hoặc hình ảnh y tế như X quang, MRI.

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *hình ảnh y tế như X quang, MRI hỗ trợ bác sĩ trong việc*

Câu 66. Trang 30: o phân loại khách hàng và dự đoán sở thích, dựa trên các yếu tố như độ tuổi, giới tính, sở thích, hành vi tiêu dùng hay màu sắc, kích thước sản phẩm ưa thích.

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *Phân loại khách hàng mục tiêu Dựa trên các yếu tố như độ tuổi, giới tính, vị trí địa lý, sở thích hành vi tiêu dùng và*

Câu 67. Trang 30: học máy có thể phân nhóm khách hàng và Dự đoán nhu cầu tiêu dùng để tối ưu chiến lược tiếp thị

Độ trùng lặp: 55%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *chiến lược tiếp thị*

Câu 68. Trang 31: o Trợ lý ảo cá nhân (Virtual Personal Assistants) Trợ lý ảo sử dụng học máy để hiểu và phản hồi các yêu cầu của người dùng thông qua văn bản giọng nói hoặc hình ảnh

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *thông qua giọng nói hoặc văn bản. Nó sử dụng công nghệ AI để hiểu và phản hồi lại các yêu cầu của người dùng một cách tự nhiên và linh hoạt Trợ lý ảo*

Câu 69. Trang 31: nhờ khả năng này, học máy được ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống

Độ trùng lặp: 76%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *được ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống và công nghiệp Nhờ khả năng*

Câu 70. Trang 32: phân loại nhị phân là một trong những bài toán cơ bản và phổ biến nhất trong học máy

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: *bài toán cơ bản trong học máy 1 2 1 bài toán Phân loại Phân loại là một trong những bài toán phổ biến và được nghiên cứu nhiều nhất trong học*

Câu 71. Trang 33: Phân cụm là một bài toán học không giám sát (unsupervised learning) trong đó Mục tiêu là chia tập dữ liệu thành các nhóm (cụm) sao cho các điểm dữ liệu trong cùng một nhóm có đặc điểm tương đồng với nhau và khác biệt so với các điểm dữ liệu ở nhóm khác

Độ trùng lặp: 61%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: là học không giám sát (unsupervised learning). [25] Hai bài toán điển hình của học không giám sát là Phân cụm (Clustering) và kết hợp (Association) Phân cụm (Clustering) là một kỹ thuật học máy không giám sát được sử dụng để nhóm các đối tượng (dữ liệu) lại với nhau sao cho các đối tượng trong cùng một nhóm (cụm) có tính chất tương tự nhau hơn so với các đối tượng ở các nhóm khác mục tiêu của Phân cụm là Phân chia dữ liệu thành các cụm sao cho

Câu 72. Trang 33: phân cụm, thường được sử dụng Trong khai phá dữ liệu, phân tích khách hàng, phát hiện bất thường và nhiều lĩnh vực khác

Độ trùng lặp: 62%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: trong khai phá dữ liệu, nhận dạng mẫu và học máy, Phân tích chuỗi thời gian được sử dụng để Phân cụm Phân loại, truy vấn theo nội dung, phát hiện bất thường và

Câu 73. Trang 34: Bài toán Hồi quy nhằm Dự đoán giá trị liên tục dựa trên dữ liệu đầu vào

Độ trùng lặp: 77%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: dự đoán giá trị liên tục dựa trên dữ liệu đầu vào

Câu 74. Trang 34: Từ tập dữ liệu quan sát được mô hình hồi quy tìm ra một hàm toán học mô tả mối quan hệ giữa các biến đầu vào và đầu ra

Độ trùng lặp: 57%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: mối quan hệ đầu vào và đầu ra ở đây thuần túy là quan sát thực nghiệm chứ không được, dựa trên một cơ sở lý thuyết Từ các quan sát thực nghiệm, mà tìm hiểu, ước đoán để giả thiết ra các dạng quan hệ có thể tồn tại giữa các biến đầu vào

Câu 75. Trang 34: độ chính xác của mô hình thường được đánh giá bằng sai số dự đoán trung bình (mean prediction error)

Độ trùng lặp: 61%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: sai số dự đoán trung bình (mean prediction error, MPE hay mean error, ME), trung bình tổng bình phương sai số dự đoán (mean squared error, MSE), căn bậc hai tổng bình phương sai số dự đoán (root mean square error, RMSE) Công thức tính các chỉ số này như sau Thẩm định mô hình Thẩm định mô hình (model evaluation, model qualification, model validation) chính là đánh giá tính dự đoán của mô hình

Câu 76. Trang 34: Cây quyết định (Decision Tree) là một thuật toán học có giám sát không tham số, được sử dụng cho cả bài toán phân loại và hồi quy Mô hình này có cấu trúc dạng Cây phân cấp, bao gồm nút gốc (root node) các nhánh (branches), nút bên trong (internal/decision nodes) và nút lá (leaf nodes)

Độ trùng lặp: 72%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Cây quyết định là một thuật toán học tập có giám sát, không tham số, được sử dụng cho cả nhiệm vụ phân loại và hồi quy có cấu trúc dạng Cây phân cấp, bao gồm nút gốc (root

node), các nhánh, các nút bên trong (internal node), và các nút lá (leaf nodes)

Câu 77. Trang 34: Quá trình phân chia tiếp tục cho đến khi các nút lá được hình thành

Độ trùng lặp: 78%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Quá trình phân chia tiếp tục để hình thành cây phân hoạch cho đến khi các nút lá

Câu 78. Trang 35: chính vì vậy, người ta cho rằng cây quyết định mô phỏng quá trình ra quyết định của con người

Độ trùng lặp: 56%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: mô phỏng quá trình ra quyết định của con người Nó dựa trên logic, Chính

Câu 79. Trang 35: Trong quá trình xây dựng cây, nó chia toàn bộ dữ liệu thành các tập dữ liệu con cho đến khi đưa ra quyết định

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Trong quá trình xây dựng cây, nó chia toàn bộ dữ liệu thành các tập dữ liệu con cho đến khi đưa ra quyết định

Câu 80. Trang 35: o xử lý linh hoạt cả dữ liệu định lượng và định danh cây quyết định có khả năng hoạt động hiệu quả với cả dữ liệu số và dữ liệu phân loại

Độ trùng lặp: 54%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Cây quyết định có khả năng Xử lý cả dữ liệu số và dữ liệu phân loại

Câu 81. Trang 36: Nó đại diện cho toàn bộ tập dữ liệu, và được phân chia thành hai hoặc nhiều tập con dựa trên các đặc trưng của dữ liệu.

Độ trùng lặp: 55%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Nó đại diện cho toàn bộ tập dữ liệu được chia thành hai hoặc nhiều tập

Câu 82. Trang 37: Cây quyết định hoạt động dựa trên chiến lược chia để trị (divide and conquer), thực hiện tìm kiếm tham lam (greedy search) để xác định các điểm phân tách tối ưu trong một Cây Quá trình này diễn ra theo hướng đệ quy từ trên xuống tiếp tục phân chia dữ liệu cho đến khi tất cả hoặc phần lớn các bản ghi được phân loại vào các nhãn lớp cụ thể

Độ trùng lặp: 57%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: chiến lược phân chia và chinh phục bằng cách thực hiện một tìm kiếm tham lam để xác định các điểm phân tách tối ưu trong một Cây Quá trình tách này sau đó được lặp lại theo cách thức đệ quy từ trên xuống, cho đến khi tất cả hoặc phần lớn các bản ghi đã được phân loại theo các nhãn lớp cụ thể tất cả các điểm dữ liệu

Câu 83. Trang 37: Tất cả các điểm dữ liệu của các tập con sau phân chia có được phân loại thành các tập đồng nhất hay không phụ thuộc đáng kể vào độ phức tạp của cây các cây có kích

thước, nhỏ thường dễ đạt được các nút lá thuần túy tức là Tất cả các điểm dữ liệu trong một nút lá thuộc về cùng một lớp

Độ trùng lặp: **63%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tất cả các điểm dữ liệu có được phân loại thành các tập đồng nhất hay không phần lớn phụ thuộc vào độ phức tạp của cây các cây nhỏ hơn có thể dễ dàng đạt được các nút lá thuần túy tức là các điểm dữ liệu trong một lớp duy nhất Tuy nhiên, khi một cây phát triển về kích thước

Câu 84. Trang 37: Tuy nhiên, khi cây phát triển quá lớn, việc duy trì độ thuần nhất trở nên khó khăn hơn và dẫn đến tình trạng một số cây con chứa quá ít dữ liệu

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tuy nhiên, khi một cây phát triển về kích thước, việc duy trì độ tinh khiết này ngày càng trở nên khó khăn và nó thường dẫn đến quá ít dữ liệu nằm trong một cây con nhất

Câu 85. Trang 37: Hiện tượng này được gọi là phân mảnh dữ liệu (data fragmentation) và thường dẫn đến quá khớp (overfitting)

Độ trùng lặp: **50%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hiện tượng này được gọi là phân mảnh dữ liệu

Câu 86. Trang 37: Do đó, cây quyết định thường ưu tiên các cây đơn giản hơn điều này phù hợp với nguyên lý của parsimony trong Occam sRazor các thực thể không nên được nhân lên quá mức cần thiết Nói cách khác, mô hình càng đơn giản càng tốt miễn là, vẫn đảm bảo được độ chính xác

Độ trùng lặp: **52%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cây quyết định ưu tiên những cây nhỏ, điều này phù hợp với nguyên tắc của parsimony trong Occam s Razor; nghĩa là các Razor; nghĩa là các thực thể không nên được nhân lên quá mức cần thiết Nói cách khác, cây quyết định chỉ nên tăng thêm độ phức tạp khi cần thiết vì lời giải thích đơn giản nhất thường là tốt.

Câu 87. Trang 38: Quy trình hoàn chỉnh có thể được hiểu rõ hơn bằng thuật toán sau o Bước 1 Bắt đầu với nút gốc (S), chứa toàn bộ tập dữ liệu huấn luyện

Độ trùng lặp: **63%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Quy trình hoàn chỉnh có thể được hiểu rõ hơn bằng cách sử dụng thuật toán dưới đây Bước 1 Bắt đầu cây với nút gốc (Đặt tên S), nút này chứa tập dữ liệu

Câu 88. Trang 38: o Bước 3 Dựa trên thuộc tính đã chọn Chia S thành các tập con, tương ứng với các giá trị có thể của thuộc tính đó

Độ trùng lặp: **64%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giá trị có thể của thuộc tính được chọn, chia tập ví dụ thành các tập con ứng với các giá trị có thể của thuộc tính được chọn, gán các tập con tương ứng các

Câu 89. Trang 38: o Bước 4 Tạo nút cây quyết định mới chứa thuộc tính tốt nhất o Bước 5 Tạo một cách đệ quy cây quyết định mới bằng cách sử dụng các tập con của tập dữ liệu đã Tạo ở Bước 3

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Bước 4 Tạo nút cây quyết định chứa thuộc tính tốt nhất Bước 5 Tạo một cách đệ quy cây quyết định mới Bước 5 Tạo một cách đệ quy cây quyết định mới bằng cách sử dụng các tập con của tập dữ liệu đã Tạo ở Bước 3

Câu 90. Trang 40: Thuật toán của Hunt, được phát triển vào những năm 1960 ban đầu nhằm mô hình hóa quá trình học tập của con người trong lĩnh vực Tâm lý học

Độ trùng lặp: 65%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Thuật toán của Hunt, được phát triển vào những năm 1960, để mô hình hóa việc học tập của con người trong Tâm lý học

Câu 91. Trang 40: ID3 (Iterative Dichotomiser 3) o được phát triển bởi Ross Quinlan ID3 là một trong những thuật toán xây dựng cây quyết định đầu tiên và phổ biến nhất

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một thuật toán quyết định cây (decision tree) Được phát triển bởi Ross Quinlan, vào những năm 1990 C4 5 là phiên bản cải tiến của thuật toán ID3 (Iterative Dichotomiser 3) và là một trong những thuật toán quyết định cây phổ biến nhất

Câu 92. Trang 40: o thuật toán sử dụng Entropy và information gain để lựa chọn thuộc tính phân tách tại mỗi nút

Độ trùng lặp: 57%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Information Gain để lựa chọn thuộc tính phân

Câu 93. Trang 41: Thường được sử dụng trong nghiên cứu thị trường và khoa học xã hội,

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được sử dụng trong nghiên cứu thị trường và khoa học xã hội

Câu 94. Trang 41: Trong quá trình xây dựng cây quyết định bằng thuật toán ID3 một bước quan trọng, là xác định thứ tự lựa chọn các thuộc tính tại mỗi nút của cây

Độ trùng lặp: 51%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: lựa chọn một thuộc tính cho phép thử là rất quan trọng nó quyết định sự thành công của việc xây dựng cây Thông thường việc chọn thuộc tính sử dụng cho thuật toán là sử dụng một độ đo gọi là Entropy và Information Gain của các thuộc tính để sắp xếp thứ tự chọn các thuộc tính ở mỗi bước Entropy của một thuộc tính được tính toán từ các thuộc tính phân lớp Đối với thuộc tính rời rạc, cần phải có các thông tin phân lớp của từng giá trị thuộc tính thuật toán ID3, xác định

sự phân lớp các đối tượng bằng cách kiểm tra các giá trị của các thuộc tính ID3, xây dựng cây quyết định

Câu 95. Trang 41: Trong các bài toán có nhiều thuộc tính và mỗi thuộc tính lại có nhiều giá trị việc tìm kiếm nghiệm tối ưu toàn cục cho toàn bộ cây là điều không khả thi về mặt tính toán

Độ trùng lặp: 51%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các bài toán có nhiều thuộc tính và mỗi thuộc tính có nhiều giá trị, khác nhau, việc tìm được nghiệm tối ưu thường là không khả thi

Câu 96. Trang 42: Sau khi chọn, được thuộc tính dữ liệu sẽ được chia thành các nút con (child nodes) tương ứng với các giá trị khác nhau của thuộc tính đó

Độ trùng lặp: 51%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các nút con (child node) tương ứng với các giá trị của thuộc tính, đó

Câu 97. Trang 42: Khi điều này xảy ra, nút con có thể được xem là nút lá (leaf node), không cần phân chia thêm nữa

Độ trùng lặp: 57%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: nút con này có thể được gọi là một nút lá (leaf node), tức ta không cần phân chia thêm nữa

Câu 98. Trang 42: một trong những hàm được sử dụng phổ biến nhất trong Lý thuyết thông tin để đo mức độ hỗn tạp là hàm entropy

Độ trùng lặp: 53%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: là Một độ đo trong lý thuyết thông tin được sử dụng để đo mức độ hỗn tạp của Một tập và Entropy=

Câu 99. Trang 42: Entropy là thuật ngữ thuộc Nhiệt động lực học, là thước đo của sự biến đổi, hỗn loạn hoặc ngẫu nhiên Năm 1948, Shannon đã mở rộng khái niệm sang lĩnh vực nghiên cứu, thống kê với công thức như sau

Độ trùng lặp: 97%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Entropy là thuật ngữ thuộc Nhiệt động lực học, là thước đo của sự biến đổi, hỗn loạn hoặc ngẫu nhiên Năm 1948, Shannon đã mở rộng khái niệm Entropy sang lĩnh vực nghiên cứu, thống kê với công thức như sau

Câu 100. Trang 43: Với một phân phối xác suất của một biến rời rạc x có thể nhận n giá trị khác nhau x_1, x_2, \dots, x_n

Độ trùng lặp: 100%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Với một phân phối xác suất của một biến rời rạc x có thể nhận n giá trị khác nhau x_1, x_2, \dots, x_n

Câu 101. Trang 43: Giả sử rằng xác suất để x nhận các giá trị này là $p_i = p(x=x_i)$

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giả sử rằng xác suất để x nhận các giá trị này là $p_i = p(x=x_i)$

Câu 102. Trang 43: $H(p) = -\sum_{i=1}^n p_i \log_2 p_i$ là logarit tự nhiên (Một số tài liệu dùng logarit cơ số 2, nhưng giá trị của $H(p)$ chỉ khác đi bằng cách nhân với một hằng số) và quy ước $0 \log(0) = 0$

Độ trùng lặp: 92%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là logarit tự nhiên (Một số tài liệu dùng logarit cơ số 2, nhưng giá trị của $H(p)$ chỉ khác đi bằng cách nhân với một hằng số) và quy ước $0 \log(0) = 0$

Câu 103. Trang 43: Giả sử bạn tung một đồng xu, entropy sẽ được tính như sau

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giả sử bạn tung một đồng xu, Entropy sẽ được tính như sau

Câu 104. Trang 43: Xét một ví dụ với $2n = 14$ được cho trên Hình 2.14. Trong trường hợp p là tính khiết nhất, tức một Trong hai giá trị p_i bằng 1, giá trị còn lại bằng 0, entropy của phân phối này là $H(p) = -\sum_{i=1}^n p_i \log_2 p_i$

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: được cho trên Hình 3. Trong trường hợp p là tính khiết nhất, tức một Trong hai giá trị p_i bằng 1, giá trị kia bằng 0, entropy của phân phối này là $H(p) = -\sum_{i=1}^n p_i \log_2 p_i$

Câu 105. Trang 43: Khi p là vẫn đục nhất, tức cả hai giá trị $p_i = 0.5$ hàm entropy đạt giá trị cao nhất.

Độ trùng lặp: 85%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Khi p là vẫn đục nhất, tức cả hai giá trị $p_i = 0.5$ hàm entropy đạt giá trị cao nhất.

Câu 106. Trang 44: Tổng quát lên với $2n$ khi có một giá trị $p_i = 1$ hàm entropy đạt giá trị nhỏ nhất ngược lại hàm đạt giá trị lớn nhất nếu tất cả các p_i bằng nhau (việc này có thể được chứng minh bằng phương pháp nhân tử Lagrange) ID3 còn được gọi là entropy based decision tree vì nó sử dụng hàm entropy trong việc đo độ vẫn đục của một phép phân chia

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: hàm entropy đạt giá trị nhỏ nhất, nếu có một giá trị $p_i = 1$, đạt giá trị lớn nhất nếu tất cả các p_i bằng nhau (việc này có thể được chứng minh bằng phương pháp nhân tử Lagrange), hàm entropy khiến nó được sử dụng trong việc đo độ vẫn đục của một phép phân chia của ID3, vì lý do này, ID3 còn được gọi là entropy based decision

Câu 107. Trang 44: Tiến hành phân chia dữ liệu thành các nhánh con ứng với từng giá trị của thuộc tính đã chọn

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: ứng với từng giá trị của thuộc tính đã chọn Bước Sắp xếp, phân chia dữ liệu với các

Câu 108. Trang 44: Chúng ta sẽ xây dựng phương pháp tính toán dựa trên mỗi node này

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Chúng ta sẽ xây dựng phương pháp tính toán dựa trên mỗi node này

Câu 109. Trang 45: Giả sử ta đang làm việc với một non leaf node với các điểm dữ liệu tạo thành một tập S với số phần tử là $|S|$ $N=S$

Độ trùng lặp: 95%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giả sử ta đang làm việc với một non leaf node với các điểm dữ liệu tạo thành một tập S với số phần tử là $|S|$

Câu 110. Trang 45: Giả sử thêm rằng trong số N điểm dữ liệu này, $1, 2, \dots, cN$ c \in điểm thuộc vào class c

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giả sử thêm rằng trong số $|N|$ điểm dữ liệu này, $|N_c|$ $C=1, 2, \dots, C$ $|N_c|$ điểm thuộc vào class $|c|$ \rightarrow Xác suất để mỗi điểm dữ liệu rơi vào một class $|c|$ được xấp xỉ bằng $\frac{|N_c|}{|N|}$

Câu 111. Trang 45: Xác suất để mỗi điểm dữ liệu rơi vào một class c được xấp xỉ bằng $\frac{|N_c|}{|N|}$

Độ trùng lặp: 95%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Xác suất để mỗi điểm dữ liệu rơi vào một class c được xấp xỉ bằng $\frac{|N_c|}{|N|}$

Câu 112. Trang 45: Dựa trên x , các điểm dữ liệu trong S được phân ra thành K child node S_1, S_2, \dots, S_K với số điểm trong mỗi child node lần lượt là n_1, n_2, \dots, n_K

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Dựa trên x các điểm dữ liệu trong S được phân ra thành K child node S_1, S_2, \dots, S_K với số điểm trong mỗi child node lần lượt là n_1, n_2, \dots, n_K

Câu 113. Trang 45: các node thường có số lượng điểm khác nhau nên Việc lấy trọng số này là cần thiết

Độ trùng lặp: 76%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: việc lấy trọng số này là quan trọng vì Các node thường có số lượng điểm khác nhau

Câu 114. Trang 45: * $\arg \max () \arg \min ()x \text{ xx } G \times H x = =S \text{ S}$ tức thuộc tính khiến cho chỉ số gain information đạt giá trị lớn nhất Trong ID3, tại mỗi node, thuộc tính được chọn được xác định dựa trên b) Ví dụ

Độ trùng lặp: 71%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: $\arg \max_{\{x\}} G(x, \mathcal{S}) = \arg \min_{\{x\}} H(x, \mathcal{S})$ \} tức thuộc tính khiến cho information Gain đạt giá trị lớn nhất , Trong ID3, tại mỗi node, thuộc tính được chọn được xác định dựa trên , \}

Câu 115. Trang 46: Tập dữ liệu này được trích từ cuốn Data Mining Practical Machine Learning Tools and Techniques, (trang 11) và là một ví dụ kinh điển thường được sử dụng trong các bài giảng về cây quyết định

Độ trùng lặp: 62%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Mining Practical Machine Learning Tools and Techniques trang 11 Đây là một bảng dữ liệu được sử dụng rất nhiều trong các bài giảng về cây quyết định

Câu 116. Trang 47: bài toán này có thể được xem là một bài toán dự đoán liệu đội bóng có ra sân thi đấu hay không dựa trên các yếu tố thời tiết quan sát được

Độ trùng lặp: 54%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một Bài toán dự đoán liệu đội bóng có chơi bóng không, dựa trên các quan sát thời tiết

Câu 117. Trang 47: Một số quy tắc ra quyết định đơn giản nhưng khá chính xác, dù chưa hẳn là tối ưu có thể được nêu như sau o Nếu outlook = sunny và humidity = high thì play = no o Nếu outlook = rainy và wind = strong, thì play = no o Nếu outlook = overcast thì play = yes o Ngoài ra Nếu humidity = normal thì play = yes o Trong các trường hợp còn lại, play = yes Ta thấy, Trong 14 giá trị đầu ra ở Bảng trên, có năm giá trị là no và chín giá trị là yes

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: đơn giản và khá chính xác có thể không phải là cách ra quyết định tốt nhất Nếu outlook = sunny và humidity = high, thì play = no Nếu outlook = rainy và windy = true thì play = no Nếu outlook = overcast, thì play = yes Ngoài ra Nếu humidity = normal, thì Trong 14 giá trị đầu ra ở Bảng trên, có năm giá trị bằng no và chín giá trị bằng yes

Câu 118. Trang 47: Tiếp theo, chúng ta sẽ tính tổng entropy có trọng số tại các node con (child node) nếu phân chia dữ liệu theo, từng thuộc tính outlook, temperature, humidity, wind, hoặc play Xét thuộc tính outlook, thuộc tính này có thể nhận một trong ba giá trị là sunny, overcast, và rainy

Độ trùng lặp: 64%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tiếp theo, chúng ta tính tổng có trọng số entropy của các child node nếu chọn một trong các thuộc tính Outlook, Temperature, Humidity, Wind Play để phân chia dữ liệu Xét thuộc tính Outlook, thuộc tính này có thể nhận một trong ba giá trị sunny, overcast rainy

Câu 119. Trang 49: với node con thứ nhất, outlook = sunny, khi áp dụng thuật toán ID3 ta chọn thuộc tính humidity để phân chia vì tổng entropy có trọng số sau bước này sẽ bằng 0 với output sẽ

là yes khi và chỉ khi humidity = normal Tương tự, node con Tương ứng với outlook = rainy sẽ tiếp tục được phân chia theo thuộc tính wind với kết quả là play = yes khi và chỉ khi wind = weak

Độ trùng lặp: **51%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: thứ nhất ứng Với Outlook = sunny, kết quả tính được bằng ID3, sẽ cho chúng ta thuộc tính Humidity vì tổng trọng số của entropy sau bước này sẽ bằng 0 Với output bằng yes khi và chỉ khi Humidity = normal Tương tự, child node ứng Với Outlook = Wind, sẽ được tiếp tục phân chia, Tương tự, child node ứng Với Outlook = Wind, sẽ được tiếp tục phân chia, bởi thuộc tính Wind, Với output bằng yes khi và chỉ khi Wind, = weak

Câu 120. Trang 50: Trong các thuật toán cây quyết định nói chung và ID3 nói riêng, nếu ta tiếp tục phân chia tất cả các node chưa thuần nhất đến cùng, ta có thể tạo ra một cây mà mọi điểm Trong tập huấn luyện đều được phân loại chính xác (giả định không có hai mẫu đầu vào giống nhau nhưng cho ra kết quả khác nhau)

Độ trùng lặp: **59%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: Trong các thuật toán cây quyết định nói chung, và ID3 nói riêng, nếu ta tiếp tục phân chia các đỉnh chưa tinh khiết, ta sẽ thu được một cây mà mọi điểm Trong tập huấn luyện đều được dự đoán đúng (giả sử rằng không có hai đầu vào giống nhau nào cho đầu ra khác nhau) Khi đó, cây có thể

Câu 121. Trang 51: Đầu tiên, ta chuẩn hoá information gain với tri thông tin phân tách (split information) Trong đó Split Info được tính như sau b) Điều kiện dừng

Độ trùng lặp: **84%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: Đầu tiên, ta chuẩn hoá information gain với tri thông tin phân tách (split information) Trong đó Split Info được tính như sau

Câu 122. Trang 52: Điều này dẫn đến khả năng overfitting cao tức là Mô hình hoạt động tốt trên tập huấn luyện nhưng kém hiệu quả trên dữ liệu mới

Độ trùng lặp: **64%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: mô hình hoạt động tốt trên tập huấn luyện nhưng lại kém hiệu quả trên dữ liệu mới

Câu 123. Trang 52: Python là một ngôn ngữ lập trình mã nguồn mở được sử dụng rộng rãi trong nhiều lĩnh vực như phát triển phần mềm, Khoa học dữ liệu, và học máy

Độ trùng lặp: **77%**

Nguồn: **Dữ liệu nội sinh**

Nội dung nguồn: phát triển phần mềm, phân tích dữ liệu, đến trí tuệ nhân tạo và học máy Điều này mở ra nhiều cơ hội nghề nghiệp cho người học Python được sử dụng trong nhiều lĩnh vực Python là một ngôn ngữ lập trình rất linh hoạt và được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau Dưới đây là một số ứng dụng phổ biến của Python 1 phát triển web Python được sử dụng để xây dựng các trang web và ứng dụng web thông qua các framework như Django và Flask Các trang web nổi tiếng như Instagram và Pinterest đều được xây dựng bằng Python 2 khoa học dữ liệu, và

Câu 124. Trang 53: với cấu trúc cú pháp rõ ràng, dễ đọc dễ viết, Python đặc biệt phù hợp với

người mới bắt đầu học lập trình

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cú pháp rõ ràng, dễ đọc, phù hợp Với người mới bắt đầu

Câu 125. Trang 53: Python là một ngôn ngữ đa nền tảng hỗ trợ nhiều mô hình lập trình khác nhau như lập trình mệnh lệnh (imperative programming), lập trình hướng đối tượng (object oriented programming), lập trình hàm (functional programming).

Độ trùng lặp: **67%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một ngôn ngữ đa hình (polymorphic), có nghĩa là nó hỗ trợ nhiều lập trình hướng đối tượng (object oriented programming), lập trình hàm (functional programming).

Câu 126. Trang 53: Python còn được Ứng dụng trong nhiều lĩnh vực như phát triển web phân tích và trực quan hóa dữ liệu thiết kế 3D CAD, tự động hóa tác vụ,

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dùng trong nhiều lĩnh vực khác nhau, từ Phát triển web, đến phân tích dữ liệu, và Phát triển trò chơi lĩnh vực ứng dụng cụ thể Phát triển web. Python là nền tảng cho nhiều trang web nổi tiếng như YouTube, Google, và Instagram Khoa học dữ liệu, Sử dụng Python để phân tích và trực quan hóa dữ liệu.

Câu 127. Trang 53: Ngôn ngữ Python hiện đang được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau. Nhờ vào tính linh hoạt và khả năng mở rộng mạnh mẽ.

Độ trùng lặp: **70%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: linh hoạt và khả năng mở rộng mạnh mẽ NoSQL ngày càng được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau

Câu 128. Trang 54: hỗ trợ kết nối và tương tác với nhiều hệ quản trị cơ sở dữ liệu

Độ trùng lặp: **88%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Hỗ trợ kết nối và tương tác với các hệ quản trị cơ sở dữ liệu như MySQL, PostgreSQL, SQLite và nhiều

Câu 129. Trang 55: Python là một ngôn ngữ dễ học và được xem là một trong những ngôn ngữ lập trình nổi bật nhất hiện nay

Độ trùng lặp: **65%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Python được xem là một trong những ngôn ngữ lập trình dễ học và dễ tiếp cận, phù hợp cho cả những lập trình viên mới bắt đầu 5 5 ngôn ngữ Javascript JavaScript là một ngôn ngữ

Câu 130. Trang 55: Python cung cấp rất nhiều framework GUI mạnh mẽ giúp việc thiết kế giao diện trở nên đơn giản và hiệu quả

Độ trùng lặp: **52%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thiết kế giao diện người dùng trở nên đơn giản và hiệu quả

Câu 131. Trang 56: Tuy nhiên, một số tính năng nâng cao chỉ có trong phiên bản thương mại

Độ trùng lặp: **78%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Tuy nhiên, một số tính năng bảo mật nâng cao chỉ có trong phiên bản

Câu 132. Trang 56: Tkinter là lựa chọn phổ biến cho những người mới bắt đầu, lập trình GUI với

Độ trùng lặp: **68%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là lựa chọn phổ biến cho những người mới bắt đầu vì nó giúp họ làm quen với

Câu 133. Trang 57: PySide là một dự án mã nguồn mở cung cấp ràng buộc Python cho Qt

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: PySide là một dự án phần mềm mã nguồn mở cung cấp các ràng buộc Python cho Qt

Câu 134. Trang 57: Qt là một bộ công cụ GUI đa nền tảng cho phép viết ứng dụng một lần và chạy trên nhiều hệ điều hành mà không cần chỉnh sửa lại mã nguồn Kết hợp sức mạnh của Qt và sự linh hoạt của Python PySide cho phép các lập trình viên phát triển ứng dụng GUI nhanh chóng và hiệu quả trên tất cả các nền tảng chính như Windows, Linux và macOS

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Qt là một ứng dụng đa nền tảng và Python, GUI Framework, cho phép các lập trình viên viết ứng dụng một lần và triển khai chúng trên nhiều hệ điều hành mà không cần viết lại mã nguồn Kết hợp sức mạnh của Qt và Python, PySide cung cấp Qt Framework phong phú cho các lập trình viên sử dụng Python, phát triển ứng dụng GUI nhanh chóng trên tất cả các hệ điều hành chính Link tải framework PySide PyGUI PyGUI là một Python, GUI Framework đa nền tảng

Câu 135. Trang 58: Trong quá trình học lập trình với ngôn ngữ Python, rất nhiều người quan tâm đến việc xây dựng các ứng dụng có giao diện người dùng đồ họa (GUI) tương tự như Windows Form

Độ trùng lặp: **63%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Trong quá trình học lập trình với ngôn ngữ Python, rất nhiều người quan tâm tới việc tạo các ứng dụng có giao diện như Windows Form

Câu 136. Trang 58: Qt Designer là một công cụ hỗ trợ thiết kế giao diện người dùng đồ họa (GUI) một cách nhanh chóng và trực quan, sử dụng các widget có sẵn từ bộ công cụ Qt GUI

Độ trùng lặp: **55%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Qt Designer là một công cụ hỗ trợ đắc lực cho việc thiết kế giao diện người dùng đồ họa (GUI) cho các ứng dụng Qt. Nó cung cấp một giao diện kéo thả trực quan, giúp bạn dễ dàng sắp xếp các widget (

Câu 137. Trang 59: Qt là một Application framework đa nền tảng được viết bằng ngôn ngữ C++ dùng để phát triển các ứng dụng trên desktop, hệ thống nhúng và mobile

Độ trùng lặp: **87%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Qt là một application framework đa nền tảng, viết trên ngôn ngữ C++, , được dùng để phát triển các ứng dụng trên desktop, hệ thống nhúng và mobile

Câu 138. Trang 59: Nó Hỗ trợ nhiều nền tảng bao gồm Linux, OS X, Windows, VxWorks, QNX, Android, iOS, BlackBerry, Sailfish OS và một số nền tảng khác

Độ trùng lặp: **87%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: nền tảng bao gồm Linux, OS X, Windows, VxWorks, QNX, Android, iOS, BlackBerry, Sailfish OS và một số nền tảng khác

Câu 139. Trang 59: PyQt là giao diện Python của Qt, là sự kết hợp giữa ngôn ngữ lập trình Python và thư viện Qt.

Độ trùng lặp: **65%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: của Qt, kết hợp của ngôn ngữ lập trình Python và thư viện Qt, là

Câu 140. Trang 59: Đây là một thư viện bao gồm các thành phần giao diện điều khiển (widgets graphical control elements) PyQt API bao gồm nhiều module với số lượng lớn các lớp (classes) và hàm (functions), hỗ trợ việc thiết kế giao diện người dùng cho các phần mềm chức năng

Độ trùng lặp: **71%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là một thư viện bao gồm các thành phần giao diện điều khiển (widgets , graphical control elements) PyQt API bao gồm các module bao gồm số lượng lớn với các classes và functions hỗ trợ cho việc thiết kế ra các giao diện giao tiếp với người dùng của các phần mềm chức năng

Câu 141. Trang 60: QtCore gồm các thành phần cốt lõi không thuộc GUI, ví dụ làm việc với thời gian, tệp và thư mục, kiểu dữ liệu, streams, URLs, mime types, luồng

Độ trùng lặp: **57%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Gồm phần lõi không thuộc chức năng GUI, ví dụ dùng để làm việc với thời gian, file và thư mục, các loại dữ liệu, streams, URLs, MIME

Câu 142. Trang 61: Giả sử bạn đã lưu tệp của mình từ Qt Designer dưới dạng dialog ui Sau đó, bạn có thể tạo một tệp khác, chẳng hạn như main.py, với nội dung Sau o Cách 1 from PyQt5 import uic from PyQt5.QtWidgets import QApplication

Độ trùng lặp: **96%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giả sử bạn đã lưu tệp của mình từ Qt Designer dưới dạng dialog ui Sau đó, bạn có thể tạo một tệp khác, chẳng hạn như main.py, với nội dung Sau Cách 1 from PyQt5 import uic from PyQt5 QtWidgets import QApplication

Câu 143. Trang 62: Sau đó, khi bạn gọi python main.py trên dòng lệnh, hộp thoại của bạn sẽ mở ra

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Sau đó, khi bạn gọi python main.py trên dòng lệnh, hộp thoại của bạn sẽ mở ra

Câu 144. Trang 62: Ngoài ra các bạn cũng có thể chuyển đổi từ tệp ui (tệp giao diện người dùng) sang tệp.py (tệp Python) bằng câu lệnh trên Terminal như sau

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Ngoài ra, các bạn cũng có thể chuyển đổi từ tệp ui (tệp giao diện người dùng) sang tệp.py (tệp Python) bằng câu lệnh trên Terminal như sau #

Câu 145. Trang 62: Qt được sử dụng để phát triển giao diện người dùng đồ họa (GUI) và các ứng dụng đa nền tảng có thể chạy trên tất cả các nền tảng máy tính để bàn chính và hầu hết các nền tảng di động hoặc những hầu hết các chương trình GUI được xây dựng bằng Qt đều có giao diện tự nhiên, vì vậy Qt được phân loại như một widget toolkit

Độ trùng lặp: 79%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Qt được sử dụng để phát triển giao diện người dùng đồ họa (GUI) và các ứng dụng đa nền tảng, chạy trên tất cả các nền tảng, máy tính để bàn lớn và hầu hết các nền tảng, di động hoặc những hầu hết các chương trình GUI được tạo bằng Qt đều có giao diện tự nhiên, trong trường hợp này Qt được phân loại là widget toolkit

Câu 146. Trang 63: Qt hỗ trợ nhiều trình biên dịch khác nhau, bao gồm GCC C++ Visual Studio và có hỗ trợ quốc tế hóa toàn diện Qt cũng cung cấp Qt Quick, bao gồm một ngôn ngữ kịch bản gọi là QML cho phép sử dụng JavaScript để xử lý logic

Độ trùng lặp: 78%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trình biên dịch khác nhau, bao gồm trình biên dịch GCC C++, và bộ Visual Studio, và có hỗ trợ quốc tế hóa Qt cũng cung cấp Qt Quick, bao gồm một ngôn ngữ kịch bản lệnh được gọi là QML, cho phép sử dụng JavaScript để cung cấp logic

Câu 147. Trang 63: Với Qt Quick, việc phát triển ứng dụng nhanh chóng cho thiết bị di động trở nên dễ dàng hơn, trong khi phần logic có thể được viết bằng mã gốc để đảm bảo hiệu năng tối ưu các tính năng nổi bật khác của Qt bao gồm o truy cập cơ sở dữ liệu SQL, o phân tích cú pháp XML, và JSON, o quản lý luồng xử lý (multithreading) o hỗ trợ lập trình mạng (networking)

Độ trùng lặp: 56%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Với Qt Quick, việc phát triển ứng dụng nhanh chóng cho Các thiết bị di động trở nên khả thi, trong khi logic vẫn có thể được viết bằng mã gốc để Các tính năng khác bao gồm Truy

cập cơ sở dữ liệu SQL Phân tích cú pháp XML Phân tích cú pháp JSON Quản lý luồng và Hỗ trợ mạng PyQt là một thư viện của

Câu 148. Trang 63: Một số thống kê cho thấy Qt không chỉ được sử dụng trong các ứng dụng máy tính mà còn xuất hiện rộng rãi trong các thiết bị nhúng và đồ điện gia dụng

Độ trùng lặp: 60%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cho thấy Qt không chỉ xuất hiện trong máy tính, mà còn trong các thiết bị nhúng và đồ điện gia dụng

Câu 149. Trang 64: PyCharm là một nền tảng hybrid được JetBrains phát triển như một IDE dành cho Python

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: PyCharm là một nền tảng hybrid được JetBrains phát triển như một IDE cho Python

Câu 150. Trang 64: Nó thường được sử dụng để phát triển các ứng dụng Python

Độ trùng lặp: 93%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Nó thường được sử dụng để phát triển các ứng dụng doanh nghiệp Python

Câu 151. Trang 64: Một số công ty lớn như Twitter, Facebook, Amazon và Pinterest cũng sử dụng PyCharm làm môi trường phát triển chính cho Python của họ

Độ trùng lặp: 65%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Twitter, Facebook, Amazon và Pinterest cũng sử dụng PyCharm làm

Câu 152. Trang 64: Pycharm có thể chạy trên Windows, Linux, hoặc Mac OS

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: PyCharm có thể chạy trên Windows, Linux hoặc Mac OS

Câu 153. Trang 64: Ngoài ra, nó tích hợp các module và package hỗ trợ giúp lập trình viên tiết kiệm thời gian và công sức khi phát triển phần mềm bằng Python

Độ trùng lặp: 63%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: các module và các package giúp các lập trình viên phát triển phần mềm bằng Python tiết kiệm thời gian và công sức

Câu 154. Trang 64: ngôn ngữ và giao diện tối giản, thân thiện, giúp lập trình viên dễ dàng tập trung và định hình nội dung

Độ trùng lặp: 69%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giao diện tối giản, thân thiện, giúp các lập trình viên dễ dàng định hình nội dung Đa dạng Ngôn ngữ lập trình như HTML, CSS, JavaScript, C++, giúp người dùng thỏa sức sáng tạo và

Câu 155. Trang 64: Tích hợp các tính năng quan trọng như bảo mật (Git), tăng tốc xử lý vòng lặp

Độ trùng lặp: 87%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Tích hợp các tính năng quan trọng như tính năng bảo mật (Git), khả năng tăng tốc xử lý vòng lặp (

Câu 156. Trang 65: o giúp viết mã chất lượng cao hơn o bao gồm các lược đồ màu cho từ khóa, lớp và hàm tăng khả năng đọc

Độ trùng lặp: 79%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: viết mã chất lượng cao hơn Nó Bao gồm các lược đồ màu (color schemes) cho từ khóa, lớp và hàm, Điều này Giúp tăng khả năng đọc

Câu 157. Trang 67: Scikit learn (sklearn) là một trong những thư viện mạnh mẽ và phổ biến nhất cho các thuật toán học máy cổ điển, được viết bằng Python.

Độ trùng lặp: 60%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: Scikit learn là một trong những thư viện học máy mạnh mẽ và phổ biến nhất trong Python cung cấp một bộ công cụ phong phú cho việc khai thác và phân tích dữ liệu thư viện này cung cấp các thuật toán học máy

Câu 158. Trang 67: Thư viện này cung cấp công cụ xử lý các bài toán machine learning và thống kê

Độ trùng lặp: 75%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các công cụ xử lý các bài toán machine learning và

Câu 159. Trang 67: Nó hỗ trợ hầu hết các thuật toán học có giám sát và không giám sát và có thể chạy được trên nhiều nền tảng khác nhau

Độ trùng lặp: 62%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: hỗ trợ hầu hết các thuật toán học có giám sát và không giám sát Scikit learning cũng có thể được

Câu 160. Trang 68: Nó cung cấp một tập hợp các hàm toán học cấp cao hữu ích cho tính toán khoa học đại số tuyến tính biến đổi Fourier và xử lý số ngẫu nhiên

Độ trùng lặp: 64%

Nguồn: Dữ liệu nội sinh

Nội dung nguồn: các hàm toán học cấp cao Nó rất hữu ích cho các tính toán khoa học cơ bản trong ML Nó đặc biệt hữu ích cho các khả năng đại số tuyến tính biến đổi Fourier, và số ngẫu nhiên

Câu 161. Trang 68: Matplotlib là thư viện phổ biến để trực quan hóa dữ liệu trong Python

Độ trùng lặp: **84%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: để trực quan hóa dữ liệu trong Python Matplotlib là thư viện vẽ đồ thị phổ biến

Câu 162. Trang 68: Nó được sử dụng để Hiển thị dữ liệu dưới dạng các loại Biểu đồ như Biểu đồ đường Biểu đồ cột, và đặc biệt hữu ích Khi cần phân tích trực quan các mẫu dữ liệu

Độ trùng lặp: **51%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dữ liệu dưới dạng các

Câu 163. Trang 69: Nó cung cấp cấu trúc dữ liệu tối ưu và công cụ mạnh mẽ để xử lý dữ liệu chuỗi thời gian và dữ liệu có cấu trúc

Độ trùng lặp: **61%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cung cấp cấu trúc dữ liệu được tối ưu hóa và linh hoạt mà bạn có thể sử dụng để thao tác với dữ liệu chuỗi thời gian và dữ liệu có cấu trúc

Câu 164. Trang 69: Pandas thường được sử dụng trong khoa học dữ liệu, học máy và phân tích dữ liệu.

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: thường được sử dụng trong khoa học dữ liệu, học máy và trực quan hóa dữ liệu, để phân tích và ra quyết định dữ liệu, thô sau khi được làm sạch và phân tích

Câu 165. Trang 70: kết quả xét nghiệm cũng có thể dương tính hoặc âm tính

Độ trùng lặp: **100%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Kết quả xét nghiệm cũng có thể dương tính hoặc âm tính

Câu 166. Trang 70: o precision tỉ lệ người thật sự mắc bệnh trên tổng số người được chẩn đoán là dương tính Ví dụ, Nếu precision = 0.9, thì cứ 100 người được chẩn đoán là dương tính sẽ có 90 người thật sự mắc bệnh

Độ trùng lặp: **85%**

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: số người được chẩn đoán là dương tính nếu Precision = 0.9, thì cứ 100 người được chẩn đoán là dương tính thì sẽ thật sự có 90 người mắc bệnh Precision càng cao thì xác suất người được chẩn đoán là dương tính có khả năng mắc bệnh càng cao Recall là Tỉ lệ người được chẩn đoán là dương tính thật sự mắc bệnh trên tổng số người thật sự nếu Precision = 0.9, thì cứ 100 người được chẩn đoán là dương tính thì sẽ thật sự có 90 người mắc bệnh Precision càng cao thì xác suất người được chẩn đoán là dương tính có khả năng mắc bệnh càng cao Recall là Tỉ lệ người được chẩn đoán là dương tính thật sự mắc bệnh trên tổng số người thật sự

Câu 167. Trang 70: o Recall tỉ lệ người được chẩn đoán là mắc bệnh trên tổng số người thật sự

mắc bệnh Nếu Recall = 0.9, thì trong 100 người thật sự mắc bệnh mô hình sẽ chẩn đoán đúng 90 người

Độ trùng lặp: 74%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cao Recall là Tỷ lệ người được chẩn đoán là dương tính thật sự mắc bệnh trên tổng số người thật sự người thật sự mắc bệnh Nếu Recall = 0.9, thì cứ 100 người mắc bệnh thì sẽ chẩn đoán 90 người

Câu 168. Trang 71: F1 Score là trung bình điều hòa của Precision và Recall

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trung bình điều hòa của Precision và Recall

Câu 169. Trang 71: Precision = số điểm Positive dự đoán đúng / tổng số điểm dự đoán là Positive

Độ trùng lặp: 71%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: dự đoán là Positive

Câu 170. Trang 71: hệ số xác định R^2 (R squared) là chỉ số Đánh giá mức độ mô hình giải thích được sự biến thiên của dữ liệu

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là sự biến thiên của

Câu 171. Trang 71: trong đó là giá trị dự đoán, là giá trị thực

Độ trùng lặp: 61%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: là giá trị dự đoán,

Câu 172. Trang 72: MSE tính trung bình của bình phương sai số giữa giá trị thực tế và giá trị dự đoán

Độ trùng lặp: 94%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: tính trung bình của bình phương sai số giữa giá trị thực tế và giá trị dự đoán

Câu 173. Trang 72: giá trị thực tế

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giá trị RMSE,

Câu 174. Trang 72: sai số tuyệt đối trung bình (Mean Absolute Error MAE) là một metric đánh giá mô hình bằng cách tính trung bình của giá trị tuyệt đối sai số giữa giá trị thực tế và giá trị dự đoán

Độ trùng lặp: 81%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: trung bình của bình phương Sai số giữa giá trị thực tế và giá trị dự đoán. Giả sử ta có một bài toán bài toán dự đoán giá nhà. Coi giá trị thực tế của nhà thứ i là y_i , còn giá trị dự đoán của căn nhà đó là \hat{y}_i . Vậy, MSE có thể được tính như sau: MAE (Mean Absolute Error) là 1 metric đánh giá mô hình bằng cách tính trung bình giá trị tuyệt đối Sai số

Câu 175. Trang 72: = cách đo sự khác biệt giữa giá trị dự đoán và giá trị thực tế

Độ trùng lặp: 86%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cách đo sự khác biệt giữa các giá trị dự đoán và giá trị thực tế

Câu 176. Trang 72: giá trị thực tế

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giá trị RMSE,

Câu 177. Trang 72: Lỗi trung bình bình phương gốc (Root Mean Square Error RMSE) là độ lệch chuẩn của phần dư (lỗi dự đoán)

Độ trùng lặp: 70%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Lỗi trung bình bình phương (RMSE) là độ lệch chuẩn của phần dư (Lỗi dự đoán)

Câu 178. Trang 73: giá trị thực tế

Độ trùng lặp: 75%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: giá trị RMSE,

Câu 179. Trang 75: xây dựng mô hình dự đoán cơ hội việc làm dành cho sinh viên sau tốt nghiệp với các thông số sau

Độ trùng lặp: 58%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: cơ hội việc làm dành cho sinh viên sau tốt nghiệp cực kỳ tốt Công việc sau tốt nghiệp ngành Công nghệ thông tin Chuyên viên thiết kế, Xây dựng và quản lý các

Câu 180. Trang 76: Dữ liệu được chia thành tập huấn luyện và kiểm tra (train/test) theo tỷ lệ K

Độ trùng lặp: 77%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Dữ liệu được chia thành tập huấn luyện và kiểm tra theo tỷ lệ

Câu 181. Trang 76: Ở mỗi lần lặp, 1 phần dữ liệu sẽ được sử dụng để kiểm tra, (test), trong khi k 1 phần còn lại được sử dụng để huấn luyện (train)

Độ trùng lặp: 65%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: phần sẽ được sử dụng làm tập kiểm tra trong khi (k 1) phần còn lại dùng để huấn luyện mô hình Quá trình này được lặp lại K lần

Câu 182. Trang 76: giao diện phải đảm bảo thân thiện, trực quan và dễ sử dụng cho người dùng

Độ trùng lặp: 73%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Giao diện thân thiện, trực quan và dễ sử dụng cho người dùng </

Câu 183. Trang 82: Sau khi thực hiện quá trình huấn luyện và đánh giá mô hình qua 7 lần (K = 7), kết quả các độ đo được tổng hợp trong bảng Sau

Độ trùng lặp: 52%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Sau khi thực hiện quá trình huấn luyện và đánh giá mô hình trên với bộ dữ liệu huấn luyện MNIST, có được 5 độ chính xác cho 5 quá trình huấn luyện trong quá trình sử dụng k fold để đánh giá Có thể thấy đây là 1 kết quả

Câu 184. Trang 84: em rất mong nhận được những góp ý, phản hồi từ thầy cô để hoàn thiện hơn trong các dự án tiếp theo

Độ trùng lặp: 62%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: Em rất mong nhận được những sự góp ý, phản hồi từ thầy cô để Em có thể phát triển và hoàn thiện kiến thức hơn nữa Sau cùng, Em xin kính chúc Quý thầy cô trong

Câu 185. Trang 85: Danh mục tài liệu tham khảo <https://machinelearningcoban.com/>

Độ trùng lặp: 100%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: DANH MỤC TÀI LIỆU THAM KHẢO <https://machinelearningcoban.com/> / / /

Câu 186. Trang 85: Tài liệu sử dụng Qt Designer <https://doc.qt.io/qt-6/qtdesigner-manual.html>

Độ trùng lặp: 71%

Nguồn: *Dữ liệu nội sinh*

Nội dung nguồn: [//doc Qt io/qt 6/qtdesigner manual html](https://doc.qt.io/qt-6/qtdesigner-manual.html)

--- Hết ---