Nhập môn Khoa học Dữ liệu
Tài liệu ôn thi cuối kỳ

Đào Thành Mạnh
Lớp IT1-03 K66
(Tài liệu tự biên soạn)

---

40 Câu hỏi Khoa học Dữ liệu

1. Mục tiêu của phân tích dữ liệu khám phá (exploratory data analysis) là gì?

   Đáp án:

   - Tóm tắt dữ liệu, trực quan hóa và hiểu về dữ liệu
   - Trực quan hóa và làm sạch dữ liệu
   - Làm sạch dữ liệu, tối ưu hóa mô hình, tăng khả năng dự đoán
   - Hiểu về dữ liệu và biến đổi dữ liệu thành một số dạng

2. Những kết luận nào có thể rút ra từ biểu đồ hộp (box plot) trong phân tích dữ liệu khám phá?

   Đáp án:

   - Sự biến đổi có khác nhau giữa các nhóm con không?
   - Sự tập trung vị trí có khác nhau giữa các nhóm con không?
   - Có giá trị ngoại lai (outliers) không?
   - Có bất kỳ đặc trưng (biến) quan trọng nào không?

3. Phương pháp nào hiển thị dữ liệu phân cấp ở định dạng lồng nhau?

Đáp án:

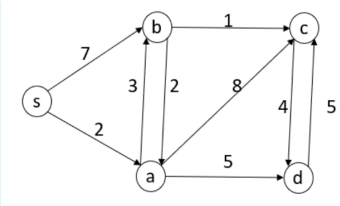
- Biểu đồ cột (Bar chart)
- Biểu đồ cây (Treemap)
- Biểu đồ dân số (Population pyramid)
- Không có phương pháp nào trong các lựa chọn trên

4. Hold-out có phải là một phương pháp để tiền xử lý và hiểu dữ liệu không?

Đáp án:

- Không, nó là một phương pháp để huấn luyện mô hình từ một tập dữ liệu cho trước.
- Không, nó là một chiến lược để đánh giá và lựa chọn mô hình.
- Có, tất nhiên.

5. Sử dụng thuật toán Dijkstra, độ dài đường đi ngắn nhất từ s đến c là bao nhiêu?



Đáp án:

- 8
- 6
- Không có đường đi từ s đến c
- 10

6. Scrapy có hỗ trợ chiến lược thu thập dữ liệu tăng dần (incremental crawling) không?

Đáp án:

- Có
- Không

7. Google Openrefine có thể nhập dữ liệu từ URL từ xa không?

Đáp án:

- Có
- Không

8. Sự khác biệt giữa học có giám sát (supervised learning) và học không giám sát (unsupervised learning) là gì?

Đáp án:

- Từ loại đầu ra, thường là một số thực trong học có giám sát
- Từ mục tiêu của thuật toán, học không giám sát thường không thực hiện dự đoán
- Từ dữ liệu huấn luyện, trong đó học có giám sát thường yêu cầu nhãn/phản hồi cho giai đoạn huấn luyện
- Từ cách chúng ta huấn luyện mô hình, học có giám sát có nghĩa là chúng ta phải cung cấp các bước chi tiết để máy học

9. Những kết luận nào có thể rút ra từ biểu đồ histogram trong phân tích dữ liệu khám phá?

Đáp án:

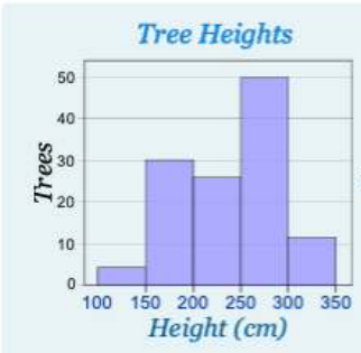
- Phân phối của dữ liệu có đối xứng hay lệch không?
- Sự phân tán của dữ liệu.
- Phân phối của tập các quan sát.
- Sự tập trung của dữ liệu.
- Có giá trị ngoại lai trong dữ liệu không?

10. Mô tả chính xác về XPath là gì?

Đáp án:

- XPath giống như một tệp XML.
- XPath là một ngôn ngữ truy vấn.
- XPath là một ngôn ngữ lập trình.
- XPath có thể được đọc bằng tài liệu Word.

11. Chỉ ra câu đúng.



Đáp án:

- Biểu đồ histogram của dữ liệu chiều cao cây
- Biểu đồ đếm số cây theo từng chiều cao trong dữ liệu

- Biểu đồ cột của dữ liệu chiều cao cây
- Biểu đồ đếm số cây trong dữ liệu

12. Kịch bản nào sau đây có thể không phù hợp với HDFS?

Đáp án:

- Lưu trữ một số lượng lớn các tệp nhỏ.
- Lưu trữ dữ liệu liên quan đến các ứng dụng yêu cầu truy cập dữ liệu độ trễ thấp.
- Các kịch bản yêu cầu ghi ngẫu nhiên vào cùng một tệp.
- Không có kịch bản nào được đề cập.

13. Thông tin nào bạn có thể thu được từ biểu đồ hộp (box plot)?

Đáp án:

- Độ lệch (Skewness)
- Phân phối xác suất
- Tứ phân vị dưới/trên
- Khoảng cách (Gap)

14. Điều gì không phải là vấn đề về chất lượng dữ liệu ở mức giá trị?

Đáp án:

- Từ đồng nghĩa (Synonym)
- Giá trị thiếu (Missing value)
- Vi phạm cú pháp (Syntax violation)

15. Sự khác biệt chính giữa Web-Scraper và Scrapy là gì?

Đáp án:

- Scrapy là một thư viện, trong khi Web-Scraper là một công cụ độc lập.
- Scrapy dựa vào XPath, trong khi Web-Scraper thì không.
- Scrapy là một thư viện, trong khi Web-Scraper là một tiện ích mở rộng của trình duyệt web.
- Web-Scraper tinh vi hơn Scrapy vì nó dựa vào hệ thống phân cấp bộ chọn.

16. Bạn đã tạo một hệ thống để dự đoán các cuộc tấn công mạng và bạn chắc chắn rằng nó có độ chính xác kiểm tra là 99%. Tuy nhiên, sếp của bạn nói rằng hệ thống của bạn vô dụng trong thực tế. Những lý do nào có thể là nguyên nhân?

Đáp án:

- Sếp của bạn không có đủ kiến thức để hiểu công việc khó khăn và hệ thống của bạn.
- Bạn không may mắn.
- Tập huấn luyện có thể có vấn đề.
- Độ chính xác có thể không phản ánh những gì sếp của bạn muốn trong lĩnh vực này.
- Việc đánh giá hệ thống của bạn có thể đã được thực hiện không chính xác.

17. Vai trò của hàm mất mát (loss function) là gì?

Đáp án:

- Đo lường mất mát/lỗi khi thực hiện dự đoán trong tương lai.
- Đo lường lỗi theo một cách nào đó và đóng vai trò là hàm mục tiêu để học từ dữ liệu.
- Không có vai trò trong quy trình khoa học dữ liệu.

18. Câu nào phù hợp nhất về việc lựa chọn mô hình?

Đáp án:

- Các câu khác đều sai.
- Lựa chọn mô hình liên quan đến việc thiết lập các tham số tốt nhất cho một mô hình khi học từ tập dữ liệu huấn luyện. Đôi khi nó đề cập đến việc chọn một mô hình từ nhiều mô hình.
- Lựa chọn mô hình chỉ liên quan đến việc chọn mô hình tốt nhất trong số các mô hình khác nhau khi làm việc với một vấn đề cụ thể.

19. Overfitting có thể đề cập đến tình huống nào?

Đáp án:

- Quá ít dữ liệu huấn luyện để máy học.
- Một phương pháp có thể dự đoán không chính xác hành vi của phương pháp khác.
- Một phương pháp có tỷ lệ lỗi nhỏ trên dữ liệu huấn luyện nhưng có tỷ lệ lỗi lớn hơn đáng kể trên dữ liệu tương lai.
- Quá nhiều dữ liệu huấn luyện nên máy có thể học dễ dàng.

20. Nhiệt độ thuộc loại dữ liệu nào?

Đáp án:

- Dữ liệu liên tục không có thứ tự.
- Dữ liệu rời rạc có thứ tự.
- Dữ liệu rời rạc không có thứ tự.
- Dữ liệu liên tục có thứ tự.

21. Ba lớp tạo nên kiến trúc của …… là backend, artist và scripting layers?

Đáp án:

- Matlab
- Matplotlib
- Pyplot
- Seaborn

22. Variety là một thách thức liên quan đến dữ liệu lớn, và nó đề cập đến điều gì?

Đáp án:

- Dữ liệu đến liên tục và nhanh chóng.
- Sức mạnh tính toán mà dữ liệu lớn yêu cầu.
- Dữ liệu có độ không chắc chắn cao do sự hiện diện của thông tin giả/nhiễu trong một số nguồn (đặc biệt là trên internet).
- Các loại dữ liệu khác nhau cần được xử lý: dữ liệu có cấu trúc/không có cấu trúc.

23. Đánh giá trong quy trình khoa học dữ liệu bao gồm những gì?

Đáp án:

- Đánh giá việc triển khai hệ thống trong thực tế.
- Phân tích, đánh giá, so sánh kết quả từ cả kịch bản ngoại tuyến và thực tế (nếu có).

24. Cho một hình ảnh grayscale không nén với 256 mức, cần bao nhiêu byte cho mỗi pixel?

Đáp án:

- 1

- 3
- 24
- 8

25. Mục đích của việc cân bằng histogram (histogram equalization) là gì?

Đáp án:

- Giảm nhiễu từ hình ảnh.
- Biểu diễn nội dung hình ảnh.
- Tăng độ sáng của hình ảnh.
- Tăng độ tương phản của hình ảnh.

26. Thư viện nào trong Python nên được sử dụng cho phân tích dữ liệu khám phá?

Đáp án:

- SciPy và NumPy
- NLTK, Spacy
- Requests, Scrapy, BeautifulSoup
- Tensorflow, Keras, Scikit-learn
- Pandas
- Matplotlib

27. Hiểu biết về kinh doanh có phải là một bước quan trọng trong quy trình khoa học dữ liệu hướng sản phẩm không?

Đáp án:

- Không, nó không liên quan đến Khoa học Dữ liệu.

- Không, chúng ta có thể bỏ qua bước đó.
- Có, tất nhiên.

28. Giả sử bạn huấn luyện một bộ phân loại trên 10.000 điểm dữ liệu và đạt được độ chính xác huấn luyện là 99%. Tuy nhiên, khi bạn nộp lên Kaggle, độ chính xác chỉ là 67%. Điều nào sau đây có khả năng cải thiện hiệu suất của bạn trên Kaggle?

Đáp án:

- Huấn luyện trên nhiều dữ liệu hơn.
- Đặt hệ số điều chỉnh (nếu có) về 0.
- Sử dụng tập đánh giá (validation set) để điều chỉnh siêu tham số.
- Loại bỏ ngẫu nhiên một phần dữ liệu huấn luyện khi huấn luyện bộ phân loại.

29. Loại biểu đồ nào sẽ được tạo ra với đoạn code sau?

```
question.plot(kind='barh')
```

Chọn một:

Đáp án:

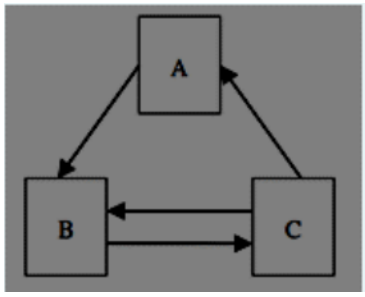
a. Biểu đồ cột ngang (Bar Graph)

b. Không có lựa chọn nào khác

c. Biểu đồ đường (Line graph)

d. Biểu đồ cột dọc (Column Graph)

30. Tính Pagerank của A với hệ số damping factor d = 0.7.



Đáp án:

- 0.3753
- 0.2245
- 0.3933
- 0.2314

31. Học cây quyết định bằng thuật toán ID3 sẽ dừng lại khi nào?

Đáp án:

- Cây đủ lớn
- Cây không thể phân loại chính xác tất cả dữ liệu huấn luyện
- Cây phân loại chính xác tất cả dữ liệu huấn luyện, hoặc tại bất kỳ đường dẫn nào tất cả các thuộc tính đã được sử dụng

32. Trong Scrapy, làm thế nào để lưu dữ liệu thu thập được vào cơ sở dữ liệu?

Đáp án:

- Viết một hook vào item pipelines
- Viết một hook vào downloader
- Viết một hook vào spider middleware

33. Chức năng nào chịu trách nhiệm hợp nhất kết quả được tạo ra bởi các hàm/tác vụ Map()?

Đáp án:

- Map
- Tất cả các lựa chọn trên
- Reducer
- Reduce

34. Câu nào sau đây là đúng?

Đáp án:

- Hive không phải là cơ sở dữ liệu quan hệ, mà là một công cụ truy vấn hỗ trợ các phần của SQL cụ thể cho việc truy vấn dữ liệu.
- Hbase không phải là cơ sở dữ liệu quan hệ nhưng hỗ trợ SQL.
- Pig là một cơ sở dữ liệu quan hệ với hỗ trợ SQL.
- Tất cả các lựa chọn trên.

35. Velocity là một thách thức của thời đại dữ liệu lớn, và nó đề cập đến điều gì?

Đáp án:

- Tốc độ phân tích
- Dữ liệu thay đổi mạnh mẽ
- Yêu cầu tính toán lớn
- Dữ liệu đến liên tục và nhanh chóng

36. Tệp robots.txt có thể thực sự ngăn chặn các trình thu thập thông tin không mong muốn không?

Đáp án:

- Có
- Không

37. Điều gì không phải là nguyên nhân gây ra nhiễu trong dữ liệu?

Đáp án:

- Sự khác biệt trong cách xem xét giữa thời điểm thu thập dữ liệu và thời điểm phân tích
- Thiết bị thu thập dữ liệu bị lỗi
- Lỗi con người khi nhập dữ liệu

38. Tại sao dữ liệu trong thế giới thực lại bẩn?

Đáp án:

- Không đầy đủ
- Được tích hợp
- Nhiễu
- Không nhất quán

39. Câu nào liên quan chặt chẽ nhất đến "Lời nguyền kích thước" (The curse of dimensionality)?

Đáp án:

- Kích thước cao có thể gây khó khăn cho việc lưu trữ và tính toán
- Khi kích thước tăng, thể tích không gian tăng nhanh đến mức dữ liệu trở nên thưa thớt. Sự thưa thớt này gây khó khăn cho bất kỳ phương pháp nào yêu cầu ý nghĩa thống kê.

- Khi kích thước tăng, độ khó của phân tích dữ liệu có thể không bị ảnh hưởng đáng kể

40. Thuật toán nổi tiếng nhất để xếp hạng trang web trong kết quả tìm kiếm là gì?

Đáp án:

- Textrank
- Webrank
- Pagerank

162 Câu hỏi Khoa học Dữ liệu

Câu 1

"Analytic approach" có phải là một bước quan trọng trong tiến trình xây dựng một sản phẩm khoa học dữ liệu?

Đáp án:

- Đương nhiên rồi
- Không, ta có thể bỏ qua bước này
- Không, nó không liên quan gì đến khoa học dữ liệu

Câu 2

Trong tiến trình xây dựng một sản phẩm khoa học dữ liệu, bước "Analytic approach" nói đến điều gì?

Đáp án:

- Việc biến đổi một bài toán thực tế về một bài toán khoa học dữ liệu
- Việc chọn một công cụ phân tích để giải quyết một bài toán khoa học dữ liệu
- Việc biến đổi một bài toán khoa học dữ liệu về một bài toán thực tế

Câu 3

Hiểu bài toán thực tế (Business understanding) có phải là một bước quan trọng trong quy trình Khoa học dữ liệu hướng sản phẩm?

Đáp án:

- Đúng, tất nhiên rồi
- Không, chúng ta có thể bỏ qua bước này
- Không, nó chẳng liên quan gì đến Khoa học dữ liệu

Câu 4

Trong khoa học dữ liệu, bước hiểu bài toán thực tế (Business understanding) là gì trong quá trình phát triển sản phẩm?

Đáp án:

- Là bước mà chúng ta cần hiểu rõ nhu cầu thực tế cần giải quyết
- Là bước để hiểu các nội dung kinh doanh chính của tổ chức
- Là bước để hiểu mối quan hệ giữa nhu cầu kinh doanh và Khoa học dữ liệu

Câu 5

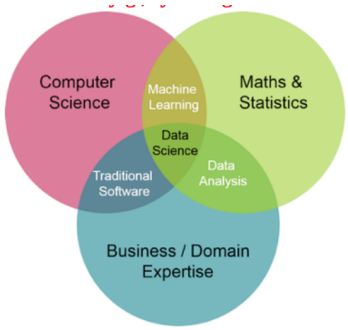
Trong quy trình Khoa học dữ liệu, giai đoạn hiểu/trực quan hoá dữ liệu nên được tiến hành sau bước mô hình hoá không?

Đáp án:

- Có
- Không, nó nên được thực hiện trước
- Các đáp án khác đều đúng

Câu 6

Hình sau đây gợi ý điều gì?



Đáp án:

- Khoa học dữ liệu là một lĩnh vực liên ngành, nó động đến rất nhiều lĩnh vực khác
- Khoa học dữ liệu là phần cốt lõi chung của Khoa học máy tính, Toán học, Thống kê, và Tri thức miền ứng dụng
- Khoa học dữ liệu là một ngành, phần giao chung giữa Khoa học máy tính, Toán học, Thống kê, và Tri thức miền ứng dụng

Câu 7

Đánh giá (Evaluation) có phải là một bước cốt lõi trong quy trình Khoa học dữ liệu, dù hướng sản phẩm hay hướng khám phá tri thức, hay không?

Đáp án:

- Đúng
- Không, nó chưa chắc cần thiết khi ta muốn khám phá tri thức mới từ dữ liệu
- Không

Câu 8

Đánh giá (Evaluation) trong quy trình Khoa học dữ liệu có thể bao gồm

Đáp án:

- Phân tích, kiểm định, so sánh các kết quả từ các kịch bản đã chọn (có thể gồm các kịch bản offline và real-life)
- Đánh giá việc triển khai một hệ thống trong thực tế

Câu 9

Trong Khoa học dữ liệu, điểm khác nhau chính giữa làm sạch và tiền xử lý dữ liệu là gì?

Đáp án:

- Làm sạch dữ liệu đương đầu phần lớn với dữ liệu nhiễu, trong khi tiền xử lý dữ liệu đương đầu phần lớn với dữ liệu thiếu
- Làm sạch dữ liệu đương đầu phần lớn với dữ liệu nhiễu, trong khi tiền xử lý dữ liệu đương đầu phần lớn với dữ liệu dư thừa
- Làm sạch dữ liệu thường thực hiện trước tiền xử lý dữ liệu, và nhắm đến việc phát hiện những dữ liệu bẩn
- Tiền xử lý dữ liệu bao gồm bước làm sạch dữ liệu

Câu 10

Phán đoán (Prediction) có phải là nhiệm vụ chính của Khoa học dữ liệu?

Đáp án:

- Đúng
- Không, nó chỉ là một trong những nhiệm vụ trong Khoa học dữ liệu
- Không, nó không thuộc lĩnh vực Khoa học dữ liệu

Câu 11

Phát biểu nào sau đây nói về ”Lời nguyền của số chiều” (The curse of dimensionality)?

Đáp án:

- Khi số chiều dữ liệu tăng, kích cỡ của không gian dữ liệu sẽ tăng quá nhanh đến nỗi những tập dữ liệu chúng ta thu thập được sẽ quá thưa thớt (sparse). Việc thưa thớt này sẽ tạo ra thách thức lớn cho các phương pháp phân tích dữ liệu.
- Khi số chiều dữ liệu tăng, sự khó khăn trong phân tích dữ liệu sẽ không bị ảnh hưởng nhiều
- Số chiều cao có thể tạo ra nhiều khó khăn cho lưu trữ và tính toán

Câu 12

"Vagueness" là một thách thức trong kỷ nguyên của Dữ liệu lớn và nó đang nói về ...

Đáp án:

- Việc dữ liệu rất khó hiểu
- Vấn đề khó giao tiếp giữa nhà cung cấp và người sử dụng
- Khó khăn đối với một người không phải là chuyên gia để diễn giải kết quả phân tích
- Mức độ khó hiểu của những mẫu dữ liệu đến trong môi trường luồng
- Mức độ khó hiểu của các thuật toán phân tích dữ liệu

Câu 13

"Variability" là một thách thức trong kỷ nguyên của Dữ liệu lớn và nó đang nói về ...

Đáp án:

- Việc dữ liệu thay đổi nhiều
- Những thay đổi có thể xảy ra trong cấu trúc của nguồn dữ liệu
- Tốc độ mà dữ liệu đến trong môi trường luồng
- Các tốc độ khác nhau mà khi đó các nguồn dữ liệu được làm mới

Câu 14

Velocity là một thách thức của kỷ nguyên dữ liệu lớn, và nó nói tới

Đáp án:

- Đặc trưng thay đổi mạnh của dữ liệu
- Những tính toán lớn
- Đặc trưng đến liên tục và nhanh của dữ liệu
- Tốc độ phân tích dữ liệu

Câu 15

Veracity là một thách thức của kỷ nguyên dữ liệu lớn, và nó nói tới

Đáp án:

- Đặc trưng thay đổi mạnh của dữ liệu
- Những tính toán lớn
- Tốc độ đến liên tục của dữ liệu trong môi trường luồng
- Đặc trưng thiếu chắc chắn cao, do nhiễu, lỗi, mất mát, sai lệch, ...trong dữ liệu

Câu 16

Khoa học dữ liệu là một lĩnh vực liên ngành và vượt ra ngoài phạm vi của Khoa học máy tính

Đáp án:

- Đúng
- Sai

Câu 17

Đâu là ví dụ về một xpath đúng?

Đáp án:

- /node/text()
- //Parent[@id='1']/Children/child/@name
- span::text
- base::attr(href)

Câu 18

Đâu là ví dụ về một xpath đúng?

Đáp án:

- //a[contains(@href, "image")]/@href
- a[href*=image]::attr(href)
- //base/@href
- base::attr(href)

Câu 19

Đâu là ví dụ về một css selector đúng?

Đáp án:

- //a[contains(@href, "image")]/@href
- a[href*=image]::attr(href)
- a[href*=image]@attr(href)
- //a[contains(@href, "image")]::@href

Câu 20

Thuật toán Page Rank được sử dụng cho mục đích gì?

Đáp án:

- Để tìm kết quả phù hợp nhất với một truy vấn
- Để đo lường tầm quan trọng của một trang web
- Để xác định mức độ phổ biến của một trang web
- Để sắp xếp kết quả của công cụ tìm kiếm

Câu 21

Làm thế nào để có thể bóc tách được dữ liệu mong muốn khi viết bot thu thập dữ liệu trên scrapy?

Đáp án:

- Sử dụng bộ chọn (selector) expath và css để viết downloaders.
- Sử dụng bộ chọn (selector) expath và css để viết spiders.
- bộ chọn (selector) expath và css để viết item pipelines.

Câu 22

Đâu là ví dụ về một css selector đúng?

Đáp án:

- /node/text()
- //Parent[@id='1']/Children/child/@name
- span::text
- base::attr(href)

Câu 23

Scrapy bot có thể bỏ qua thông tin trong robots.txt hay không?

Đáp án:

- Có.
- Không.

Câu 24

Scrapy có hỗ trợ mặc định cơ chế thu thập dữ liệu tăng dần (incremental crawling strategy) hay không?

Đáp án:

- Không
- Có

Câu 25

Sử dụng robots.txt có chặn được các chương trình cào dữ liệu Internet hay không

Đáp án:

- Không
- Có

Câu 26

Theo cách của Scrapy, đâu là nơi có thể ép buộc các yêu cầu tải trang web phải sử dụng proxy?

Đáp án:

- Downloader middlewares
- Spider middlewares
- Downloader
- Spider

- Item pipelines

Câu 27

Trong Scrapy, đâu là nơi có thể tiến hành thay đổi thuộc tính user-agent trong quá trình thu thập dữ liệu?

Đáp án:

- Downloader middlewares
- Spider middlewares
- Downloader
- Spider

Câu 28

Trong Scrapy, làm thế nào để ghi dữ liệu thu thập được vào các cơ sở dữ liệu?

Đáp án:

- Viết mã lệnh thêm vào trong spider middleware.
- Viết mã lệnh thêm vào trong item pipelines.
- Viết mã lệnh thêm vào trong downloader.

Câu 29

Trong Scrapy, vai trò của thành phần downloader là gì?

Đáp án:

- Nhận các yêu cầu (requests) từ thành phần engine, đưa các yêu cầu này vào hàng đợi để xử lý sau.
- Tải về các trang web
- Bóc tách các trả lời (responses)

Câu 30

Trong Scrapy, vai trò của thành phần spider là gì

Đáp án:

- Bóc tách các trả lời (responses).
- Tải về nội dung các trang web.
- Điều phối luồng dữ liệu giữa tất cả các thành phần của Scrapy.

Câu 31

Đâu là giải thuật dùng để xếp thứ hạng các trang web trong kết quả trả về của máy tìm kiếm

Đáp án:

- Webrank
- Pagerank
- Textrank

Câu 32

Điều nào sau đây mô tả chính xác XPath?

Đáp án:

- XPath là một ngôn ngữ lập trình.
- XPath là một ngôn ngữ truy vấn.
- XPath là cấu trúc tệp tin XML.
- XPath có thể được đọc bởi Microsoft Word.

Câu 33

Bước nào trong phương pháp làm sạch dữ liệu sau đây không theo thứ tự thích hợp?

Đáp án:

- A. Trích xuất các trường dữ liệu có liên quan.
- B. Sửa chữa các vấn đề về chất lượng dữ liệu ở mức giá trị (value level).
- C. Chuẩn hóa giá trị dữ liệu.
- D. Khắc phục các vấn đề về chất lượng dữ liệu ở mức tập giá trị (value set level).
- E. Khắc phục các vấn đề về chất lượng dữ liệu ở cấp độ quan hệ
- F. Sửa chữa các vấn đề về chất lượng dữ liệu ở cấp độ đa quan hệ.
- G. Lấy phản hồi của người dùng

Câu 34

Google Openrefine có thể nhập dữ liệu từ Internet qua URL được không?

Đáp án:

- Có.
- Không.

Câu 35

Google Openrefine có thể được sử dụng để tự động phân nhóm dữ liệu không?

Đáp án:

- Có
- Không

Câu 36

Kỹ thuật faceting trong Google Openrefine là gì?

Đáp án:

- Cho phép nhìn thấy bức tranh toàn cảnh về dữ liệu.
- Cho phép lọc xuống chỉ tập hợp con các hàng mà bạn muốn thay đổi hàng loạt.
- Cho phép thực hiện các phán đoán xu thế từ dữu liệu.

Câu 37

Tại sao dữ liệu ngoài thực tiễn lại không sạch?

Đáp án:

- Không đầy đủ.
- Có nhiễu.
- Không nhất quán.

Câu 38

Đặc trưng mô hình hoá dữ liệu trong OLAP?

Đáp án:

- Lược đồ CSDL cần được chuẩn hoá, đảm boả dữ liệu được nhất quán.
- Thường sử dụng lược đồ CSDL phi chuẩn hoá.
- Thường sử dụng mô hình dữ liệu đa chiều.

Câu 39

Đâu không phải là nguyên nhân dẫn đến dữ liệu bị nhiễu?

Đáp án:

- Phương tiên, thiết bị thu thập dữ liệu bị lỗi.
- Lỗi do người nhập dữ liệu vào hệ thống.

- Do nhu cầu về dữ liệu khác nhau giữa thời điểm thu thập dữ liệu và thời điểm tiến hành phân tích dữ liệu.

Câu 40

Đâu không phải là vấn đề về chất lượng dữ liệu ở mức giá trị (value level)?

Đáp án:

- Giá trị bị thiếu.
- Vi phạm cú pháp.
- Các từ đồng nghĩa.

Câu 41

Đâu không phải là vấn đề về chất lượng dữ liệu ở mức tập giá trị (value set level)?

Đáp án:

- Tồn tại các từ đồng âm khác nghĩa.
- Vi phạm tính duy nhất.
- Vi phạm ràng buộc toàn vẹn.
- Vi phạm tập xác định.

Câu 42

Đâu là đặc trưng của OLAP?

Đáp án:

- Chủ yếu là các giao dịch thêm, sửa, xoá có thời gian thực hiện ngắn.
- Các truy vấn thường phức tạp và bao gồm các phép toán kết tập.
- Chủ yếu là các truy vấn ad-hoc.

- Thường truy cập tới nhiều bản ghi dữ liệu.
- Hỗ trợ ra quyết định.

Câu 43

Đâu là đặc trưng của OLTP?

Đáp án:

- Chủ yếu là các giao dịch thêm, sửa, xoá có thời gian thực hiện ngắn.
- Hỗ trợ xử lý giao dịch, vận hành hàng ngày cho doanh nghiệp.
- Thường truy cập tới dữ liệu lịch sử, dữ liệu đa chiều.
- Thường là các truy vấn phức tạp.

Câu 44

Điều nào dưới đây là không đúng về OLAP?

Đáp án:

- Xử lý thông tin có tính lịch sử (được tạo ra trong quá khứ).
- Hỗ trợ phân tích nghiệp vụ.
- Khả mở, cho phép hàng triệu người sử dụng.
- Lưu trữ hàng triệu bản ghi dữ liệu.

Câu 45

Đâu là phát biểu đúng về Wrapper trong kiến trúc tích hợp dữ liệu ảo?

Đáp án:

- Là đoạn chương trình chuyển đổi dữ liệu từ định dạng ở nguồn qua định dạng chuẩn hoá của mediator

- Có thể cài đặt ở phía nguồn dữ liệu hoặc phía mediator
- Là thành phần không thể thiếu của mediator

Câu 46

Đâu là siêu dữ liệu có trong Danh mục nguồn dữ liệu (Data source catalog) trong kiến trúc tích hợp dữ liệu ảo?

Đáp án:

- Danh sách các bản dữ liệu của nguồn
- Khả năng truy vấn của nguồn (vd., Khả năng trả lời SQL)
- Tần suất cập nhật dữ liệu
- Kiểm soát truy cập, phân quyền

Câu 47

Đâu là phát biểu đúng về Apache Nifi?

Đáp án:

- Một công cụ ETL.
- Một nền tảng kho dữ liệu cho phép lưu trữ dữ liệu kích thước lớn.
- Một công cụ cho phép làm sạch và tiền xử lý dữ liệu.

Câu 48

Đâu là các khái niệm có trong Apache NiFi?

Đáp án:

- FlowFile
- FlowFile Processor

- Scheduler
- Process Group

Câu 49

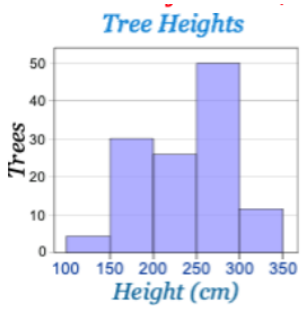
Đâu là các chiều thang đo khi nói về chất lượng dữ liệu?

Đáp án:

- Đầy đủ (Completeness), Xác thực (Validity), Toàn vẹn (Integrity)
- Tính thời điểm (Timeliness), Chính xác (Accuracy), Nhất quán (Consistency)
- Tính thời điểm (Timeliness), Tính cô lập (Isolation)
- Đầy đủ (Completeness), Toàn vẹn (Integrity), Giá trị (Value)

Câu 50

Biểu đồ này thể hiện điều gì?



Đáp án:

- Một biểu đồ cột thể hiện chiều cao của cây và số lượng cây tương ứng với chiều cao đó.
- Một biểu đồ histogram thể hiện phân bố chiều cao của cây.
- Một biểu đồ cột vẽ nhầm cần đổi tên trục trung và trục hoành.

Câu 51

Biểu đồ box plot cho phép rút ra kết luận gì trong phân tích thăm dò dữ liệu?

Đáp án:

- Có đặc trưng (biến) nào quan trọng ?
- Độ tập trung vị trí có khác nhau giữa các nhóm con không?
- Độ biến thiên có khác nhau giữa các nhóm con không?
- Có ngoại lệ không?

Câu 52

Biểu đồ histogram cho phép rút ra kết luận gì trong phân tích thăm dò dữ liệu?

Đáp án:

- Xem xét phân bố của tập các quan sát.
- Xem xét độ tập trung của dữ liệu.
- Xem xét sự phân tán của dữ liệu.
- Phân bố của dữ liệu là đối xứng hay lệch.
- Có ngoại lệ trong dữ liệu không?.

Câu 53

Biểu đồ scatter plot cho phép rút ra kết luận gì trong phân tích thăm dò dữ liệu?

Đáp án:

- Có mối quan hệ giữa biến X và Y hay không?
- Mối liên hệ có phải là tuyến tính hay không?
- Sự biến thiên của biến Y có phụ thuộc vào biến X hay không?
- Biến X, Y biến nào quan trọng hơn.

Câu 54

Phân tích thăm dò dữ liệu (EDA) là gì?

Đáp án:

- EDA không phải là một tập các kỹ thuật, mà là một triết lý về cách mà chúng ta nên làm khi muốn hiểu về dữ liệu
- EDA là tập các kỹ thuật cho phép chúng ta hiểu về dữ liệu bao gồm việc sử dụng các biểu đồ và các kỹ thuật thống kê.
- EDA là việc sử dụng các biểu đồ để hiểu dữ liệu

Câu 55

Thực hiện phân tích thăm dò dữ liệu như thế nào?

Đáp án:

- Xem xét các thuộc tính mô tả độ đo trung tâm và độ đo phân tán của dữ liệu
- Xem xét phân bố của dữ liệu
- Xem xét các mối liên hệ giữa các biến trong dữ liệu
- Xem xét đặc trưng cấu trúc của dữ liệu

Câu 56

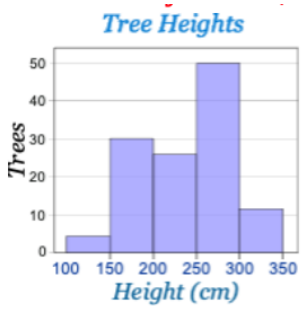
Trọng tâm của phân tích thăm dò dữ liệu EDA là gì?

Đáp án:

- EDA quan tâm tới cấu trúc, các ngoại lệ, và các mô hình từ dữ liệu
- EDA quan tâm tới tất cả các điểm dữ liệu trong tập dữ liệu
- Trực quan hoá và làm sạch dữ liệu
- EDA quan tâm tới các công cụ cho phép xem xét cấu trúc, các ngoại lệ từ dữ liệu

Câu 57

Với biểu đồ dưới đây thì phát biểu nào sai?



Đáp án:

- Số lượng các cây có chiều cao từ 250 tới 300 là nhiều nhất.
- Số lượng các cây có chiều cao từ 100 tới 150 là nhỏ nhất.
- Có 30 cây có chiều cao là 150.
- Có ít hơn hoặc bằng 50 cây có chiều cao là 300.

Câu 58

Với cùng 1 tập dữ liệu, kết quả của giải thuật phân cụm K-means phụ thuộc vào những yếu tố nào?

Đáp án:

- Cách tính độ đo khoảng cách.
- Cách gộp các cụm.
- Cấu hình số cụm ban đầu K.
- Một phép đoán khởi đầu cho các centroids.

Câu 59

Với cùng 1 tập dữ liệu, kết quả của giải thuật phân cụm phân cấp (Hierarchical clustering) phụ thuộc vào những yếu tố nào?

Đáp án:

- Cách tính độ đo khoảng cách.
- Cách gộp các cụm.
- Cấu hình số cụm ban đầu K.

Câu 60

Đâu là các thư viện và công cụ có thể sử dụng để thực hiện phân tích thăm dò dữ liệu?

Đáp án:

- NLTK, Spacy
- Requests, Scrapy, BeautifulSoup
- Tensorflow, Keras, Scikit-learn
- SciPy, NumPy, Matplotlib and Pandas

Câu 61

Khả năng tổng quát hoá (Generalization) và Quá khớp (overfitting) là hai mặt đối lập của các mô hình học máy

Đáp án:

- Đúng
- Không, chúng chưa chắc đã đối lập nhau
- Chúng là hai đặc trưng độc lập

Câu 62

Giả sử bạn muốn sử dụng một phương pháp học máy để phân tích tri thức ẩn bên trong một tập dữ liệu, nhưng không có ý niệm gì về những tri thức đó. Bạn có thể đưa về bài toán nào sau đây là phù hợp nhất?

Đáp án:

- Học không giám sát (Unsupervised learning)
- Học có giám sát (Supervised learning)
- Hồi quy (Regression)
- Phân loại nhiều lớp (Multiclass classification)

Câu 63

Quá trình học một cây quyết định bằng thuật toán ID3 sẽ dừng nếu

Đáp án:

- Cây đã phân loại chính xác hoàn toàn dữ liệu huấn luyện, hoặc tại bất kỳ một đường đi nào từ gốc đến lá, các thuộc tính đã được dùng hết
- Cây đã phân loại chính xác hoàn toàn dữ liệu huấn luyện
- Cây đã đủ lớn
- Cây không thể phân loại chính xác hoàn toàn dữ liệu huấn luyện

Câu 64

Vai trò của ”information gain” trong thuật toán ID3 khi học một cây quyết định là gì?

Đáp án:

- Đo đạc tính phân biệt của các thuộc tính để tìm một thuộc tính kiểm tra tại mỗi đỉnh
- Để xem độ tốt của một thuộc tính sau quá trình huấn luyện
- Để đo đạc lỗi tại mỗi đỉnh trong cây
- Không có vai trò gì

Câu 65

K-means là gì?

Đáp án:

- Một phương pháp phân cụm
- Một phương pháp phân loại
- Một phương pháp học có giám sát
- Một phương pháp để tính trung bình số học từ dữ liệu

Câu 66

Một phương pháp học máy có khả năng học được gì?

Đáp án:

- Một hàm mà có khả năng ánh xạ một điểm dữ liệu đầu vào đến một đầu ra
- Tri thức mới để phán đoán đầu ra
- Học để mô phỏng khả năng của con người
- Bất kỳ thứ gì

Câu 67

Sự khác nhau giữa học có giám sát và không giám sát nằm ở đâu?

Đáp án:

- Tập huấn luyện, trong đó học có giám sát thường yêu cầu nhãn/đầu ra cho mỗi mẫu dữ liệu
- Kiểu đầu ra, trong đó học có giám sát thường có đầu ra là số thực
- Cách chúng ta huấn luyện một mô hình, học có giám sát thường yêu cầu chỉ dẫn chi tiết từng bước học ra sao
- Mục tiêu của thuật toán, học không giám sát thường không thực hiện phán đoán nào cả

Câu 68

Phương pháp bình phương tối thiểu học một hàm $f(x) = w_0 + w_1x_1 + \cdots + w_nx_n$ từ một tập học có cỡ M bằng cách tìm véctơ $\mathbf{w}^* = (w_0^*, w_1^*, \ldots, w_n^*)$, trong đó

Đáp án:

- $\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{M}(y_i - w_0 - w_1x_{i1} - \cdots - w_nx_{in})$
- $\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{M}(y_i - w_0 - w_1x_{i1} - \cdots - w_nx_{in})^2$
- $\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{M}(y_i - w_0 - w_1x_{i1} - \cdots - w_nx_{in})^2 + \lambda\|\mathbf{w}\|_2^2$

Câu 69

Vai trò của hàm lỗi thực nghiệm (empirical loss) là gì?

Đáp án:

- Để đo đạc lỗi phán đoán theo một nghĩa nào đó và hay được dùng làm hàm mục tiêu khi huấn luyện một mô hình
- Để đo đạc lỗi phán đoán trong tương lai
- Không có vai trò gì

Câu 70

Học máy (Machine Learning) cung cấp các phương pháp để phân tích dữ liệu và tạo các phán đoán cho dữ liệu trong tương lai

Đáp án:

- Đúng
- Sai, nó cung cấp các nền tảng để mở rộng tính toán
- Đúng, nó còn cung cấp các nền tảng để tăng tốc tính toán

Câu 71

Định lý "No-free-lunch" nói đến điều gì?

Đáp án:

- Không có thuật toán nào có thể đánh bại một thuật toán khác trên mọi miền bài toán
- Không có bữa trưa miễn phí nào cho ai
- Nếu không cố gắng nhất, một thuật toán không thể đánh bại các thuật toán khác

Câu 72

Overfitting nói đến tình huống nào?

Đáp án:

- Một phương pháp tạo ra tỉ lệ lỗi bé trên tập huấn luyện, nhưng có tỉ lệ lỗi lớn trên dữ liệu trong tương lai
- Một phương pháp có thể phán đoán không chính xác về hành vi của một phương pháp khác
- Quá ít dữ liệu huấn luyện
- Có quá nhiều dữ liệu huấn luyện đến nỗi một máy tính có thể học dễ dàng

Câu 73

Kỹ thuật nào sau đây có thể giúp giảm overfitting?

Đáp án:

- Sử dụng hiệu chỉnh (regularization), kỹ thuật thường giúp hạn chế không gian tìm kiếm khi huấn luyện một mô hình
- Dùng một phương pháp/mô hình mới
- Bỏ bớt dữ liệu nếu có quá nhiều

Câu 74

Học máy xuất hiện ở đâu trong một quy trình khoa học dữ liệu?

Đáp án:

- Bước mô hình hoá (modeling), nơi mà chúng ta dùng một phương pháp cụ thể để phân tích dữ liệu
- Bước hiểu dữ liệu
- Bước lựa chọn một cách tiếp cận để giải bài toán đang có

Câu 75

Veracity là một thách thức liên quan đến dữ liệu lớn và đề cập đến

Đáp án:

- Các loại dữ liệu khác nhau phải được xử lý: dữ liệu có cấu trúc / không có cấu trúc.
- Sức mạnh tính toán mà dữ liệu lớn yêu cầu.
- Dữ liệu đến liên tục và nhanh chóng.
- Dữ liệu có độ không chắc chắn cao do sự hiện diện của thông tin giả mạo / nhiễu trong một số nguồn (đặc biệt là trên internet).

Câu 76

Variety là một thách thức liên quan đến dữ liệu lớn và đề cập đến

Đáp án:

- Các loại dữ liệu khác nhau phải được xử lý: dữ liệu có cấu trúc / không có cấu trúc.
- Sức mạnh tính toán mà dữ liệu lớn yêu cầu.
- Dữ liệu đến liên tục và nhanh chóng.
- Dữ liệu có độ không chắc chắn cao do sự hiện diện của thông tin giả mạo / nhiễu trong một số nguồn (đặc biệt là trên internet).

Câu 77

Làm cách nào để phân tích dữ liệu khả mở cho dữ liệu lớn?

Đáp án:

- Song song hoá các giải thuật học máy
- Sử dụng kiến trúc xử lý thời gian thực
- Sử dụng phân tích thành phần chính (PCA)
- Sử dụng mô hình mạng thần kinh sâu

Câu 78

Chỉ ra phát biểu đúng:

Đáp án:

- Hadoop cần chạy trên phần cứng chuyên biệt, cấu hình cao để xử lý dữ liệu lớn.
- Hadoop 2.0 trở lên cho phép chạy các công việc không phải là các công việc MapReduce.
- Trong khung lập trình Hadoop, các tệp tin kết quả được phân chia thành các dòng hoặc bản ghi.
- Không đáp án nào đúng.

Câu 79

Chọn phát biểu đúng:

Đáp án:

- MapReduce thực hiện mang dữ liệu tới các nút tính toán.
- MapReduce mạng tính toán tới các nút chứa dữ liệu.
- Dữ liệu cho MapReduce bắt buộc nằm trên HDFS.
- Tất cả các đáp án.

Câu 80

Phát biểu nào sau đây không đúng về Apache Hadoop?

Đáp án:

- Xử lý dữ liệu phân tán với mô hình lập trình đơn giản, thân thiện hơn như MapReduce.
- Hadoop thiết kế để mở rộng thông qua ky thuật scale-out, tăng số lượng máy chủ
- Thiết kế để vận hành trên phần cứng phổ thông, có khả năng chống chịu lỗi phần cứng
- Thiết kế để vận hành trên siêu máy tính, cấu hình mạnh, độ tin cậy cao

Câu 81

Công cụ nào có thể sử dụng để hỗ trợ import, export dữ liệu vào ra hệ sinh thái Hadoop?

Đáp án:

- Oozie
- Flume
- Sqoop
- Hive

Câu 82

Vai trò của YARN?

Đáp án:

- Quản lý và phân phối tài nguyên trong cụm Hadoop
- Cung cấp giao diện người dùng mức cao, biến đổi truy vấn thành các job Mapreduce

- Cung cấp các chức năng phối hợp phân tán độ tin cậy cao như quản lý thành viên, bầu cử, giám sát trạng thái hệ thống

Câu 83

Hadoop là một hệ sinh thái bao gồm các thành phần nào:

Đáp án:

- MapReduce, YARN
- MapReduce, MySQL
- MapReduce, Skykeeper
- MapReduce, Heron

Câu 84

Hadoop đạt được độ tin cậy thông qua cơ chế nhân bản dữ liệu trên nhiều máy chủ, do đó không yêu cầu ……trên các nút máy chủ này.

Đáp án:

- RAID.
- Hệ thống tệp tin cục bộ (Local file system).
- Hệ điều hành.

Câu 85

Hàm ……chịu trách nhiệm tổng hợp kết quả từ các tác vụ Map().

Đáp án:

- Reduce.
- Map.

- Sort.
- Không có phương án nào.

Câu 86

Cơ chế tổ chức dữ liệu của Datanode trong HDFS?

Đáp án:

- Các chunk là các tệp tin trong hệ thống tệp tin cục bộ của máy chủ datanode.
- Các chunk là các vùng dữ liệu liên tục trên ổ cứng của máy chủ datanode.
- Các chunk được lưu trữ tin cậy trên datanode theo cơ chế RAID.

Câu 87

Cơ chế nhân bản dữ liệu trong HDFS?

Đáp án:

- Namenode quyết định vị trí các nhân bản của các chunk trên datanode.
- Datanode là primary quyết định vị trí các nhân bản của các chunk tại các secondary datanode.
- Client quyết định vị trí lưu trữ các nhân bản với từng chunk.

Câu 88

HDFS được lập trình bằng ngôn ngữ nào?

Đáp án:

- C++.
- Java.
- Scala.
- Không đáp án nào đúng.

Câu 89

Tác vụ .....có trách nhiệm xử lý một hoặc vài khối (chunk) dữ liệu và trả ra kết quả trung gian.

Đáp án:

- Map.
- TaskTracker.
- Tất cả các phương án.
- Reduce.

Câu 90

Thành phần .....có trách nhiệm thực thi các tác vụ (task) dược giao bởi JobTracker.

Đáp án:

- MapReduce
- Mapper
- TaskTracker
- JobTracker

Câu 91

Tình huống nào sau đây có thể không phù hợp với HDFS?

Đáp án:

- Đọc, ghi ngẫu nhiên vào tệp tin.
- Lưu trữ dữ liệu liên quan đến các ứng dụng yêu cầu quyền truy cập dữ liệu có độ trễ thấp.
- Lưu trữ các tệp tin kích thước nhỏ.
- Không có đáp án đúng

Câu 92

Đưa ra phát biểu đúng:

Đáp án:

- Một công việc MapReduce thường chia tập dữ liệu đầu vào thành các phần độc lập được các tác vụ map xử lý theo cách hoàn toàn song song
- MapReduce xem dữ liệu là các cặp khoá-giá trị.
- Các ứng dụng thường triển khai các giao diện Mapper và Reducer để cài đặt các phương thức map và reduce.
- MapReduce chỉ làm việc với dữ liệu trên Hadoop HDFS.

Câu 93

Đưa ra đáp án đúng:

Đáp án:

- Hive không phải là một cơ sở dữ liệu quan hệ mà là một công cụ truy vấn hỗ trợ SQL để truy vấn dữ liệu
- HBase là một cơ sở dữ lớn có hỗ trợ SQL
- Pig là một cơ sở dữ liệu quan hệ có hỗ trợ SQL
- Tất cả các phương án.

Câu 94

Một trang tin cậy (authority page) về một chủ đề là gì?

Đáp án:

- Là trang được trỏ tới từ nhiều hub tốt
- Là trang được trỏ tới từ nhiều trang tin cậy
- Là trang trỏ đến nhiều hubs tốt

Câu 95

Xét ma trận $\hat{P}$ thu được bằng cách cộng số 0.1 vào tất cả các phần tử của ma trận xác suất chuyển P ở trên. Vậy $\hat{P}$ có tạo ra một chuỗi ergodic Markov không?

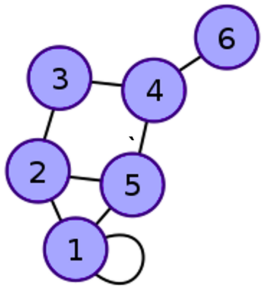
$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Đáp án:

- Có
- Không
- Chúng ta không thể nói gì về tính ergodic

Câu 96

Giá trị của ô [1,1] trong ma trận kề của đồ thị sau là bao nhiêu?



Đáp án:

- 0
- 1
- 2
- 4

Câu 97

Một chuỗi Markov ergodic là gì? (ergodic Markov chain)

Đáp án:

- Một chuỗi cho phép ta có thể đi dần dần từ bất kỳ trạng thái nào đến bất kỳ trạng thái khác với xác suất dương
- Một chuỗi cho phép ta có thể đi trực tiếp từ bất kỳ trạng thái nào đến bất kỳ trạng thái khác với xác suất dương
- Một chuỗi mà trong đó tồn tại một cặp trạng thái không thể đi đến nhau

Câu 98

Thuật toán PageRank xếp hạng các trang web như thế nào?

Đáp án:

- PageRank sử dụng tỉ lệ ghé thăm dài hạn (long-term visit rate) của mỗi trang web, và tỉ lệ đó được tính từ ma trận xác suất chuyển
- PageRank sử dụng số lượng kết nối vào mỗi trang web
- PageRank sử dụng số lượng kết nối ra từ mỗi trang web
- PageRank xếp hạng một cách ngẫu nhiên

Câu 99

Tác vụ nào không có trong phân tích liên kết?

Đáp án:

- Xếp hạng đồ thị
- Nhận diện cộng đồng
- Dự đoán liên kết
- Phân tích cảm xúc

Câu 100

Phương pháp mũ (Power method) có thể ...

Đáp án:

- tính tỉ lệ ghé thăm dài hạn cho mỗi trang web
- tính phân bố xác suất trạng thái dừng (steady-state probability distribution) cho một chuỗi Markov
- dùng một chuỗi Markov để dự đoán một chuỗi các trang sẽ ghé thăm
- tính một chuỗi các trang sẽ ghé thăm, khi cho trước điểm xuất phát

Câu 101

Cho ma trận xác suất chuyển P ở trên. Cho biết P có tạo ra một chuỗi ergodic Markov không?

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Đáp án:

- Có
- Không

Câu 102

Sự khác nhau giữa tập cơ sở (Base set S) và tập gốc (Root set W) trong thuật toán HITS là gì?

Đáp án:

- Tập cơ sở được xây dựng từ tập gốc
- Tập gốc được xây dựng từ tập cơ sở

- Tập cơ sở là cơ sở để đánh giá chất lượng của các trang tìm được bởi HITS, dựa trên tập gốc

Câu 103

Thuật toán HITS xếp hạng các trang web như thế nào?

Đáp án:

- HITS tìm ra một tập nhỏ các hubs và các trang tin cậy, sử dụng một thuật toán lặp để tính toán điểm số cho các trang
- HITS tìm ra một tập nhỏ các trang tin cậy, sử dụng một thuật toán lặp để tính toán phân bố xác suất trạng thái dừng
- HITS tìm ra một tập nhỏ các trang tin cậy, sử dụng một thuật toán lặp để tính toán tỉ lệ ghé thăm dài hạn

Câu 104

Trong các độ đo thứ hạng đỉnh dưới đây, độ đo nào chỉ dựa trên các đỉnh liền kề của đỉnh đang xét

Đáp án:

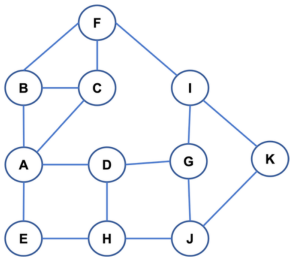
- Độ trung tâm lân cận (Closeness centrality)
- Độ trung tâm trung gian (Betweenness centrality)
- Độ quan trọng theo bậc (Degree prestige)
- Độ quan trọng lân cận (Proximity prestige)

Câu 105

Có bao nhiêu đường đi ngắn nhất từ A tới K trong đồ thị sau?

Đáp án:

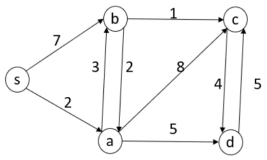
- 4



- 5
- 6
- 7

Câu 106

Sử dụng thuật toán Dijkstra, đường đi ngắn nhất từ s tới c có độ dài...



Đáp án:

- 8
- 10
- 6
- Không có đường đi từ s tới c

Câu 107

Cho một chuỗi Markov với 5 trạng thái và ma trận xác suất chuyển P ở trên. Giả sử chúng ta đang ở một trạng thái nào đó, được biểu diễn bởi véctơ xác suất $x = (0.1, 0, 0.2, 0.3, 0.4)$. Vậy ta sẽ di chuyển đến trạng thái 3 với xác suất là bao nhiêu nếu dùng một bước ngẫu nhiên?

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Đáp án:
0.350
0 0.1000000 3 0

Câu 108

Nếu độ sáng của một ảnh đa mức xám là 255, ảnh có đặc điểm nào dưới đây?

Đáp án:

- Ảnh trắng toàn bộ
- Ảnh đen toàn bộ
- Ảnh có một vài khối đen, một vài khối trắng
- Không có điểm gì quá đặc biệt, có các điểm ảnh có thể nhận giá trị đa dạng trong miền giá trị của nó.

Câu 109

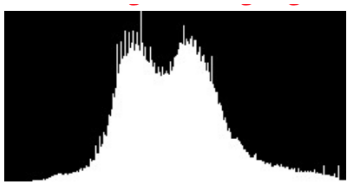
Nếu độ sáng của một ảnh đa mức xám là 0, ảnh có đặc điểm nào dưới đây?

Đáp án:

- Ảnh trắng toàn bộ
- Ảnh đen toàn bộ
- Ảnh có một vài khối đen, một vài khối trắng
- Không có điểm gì quá đặc biệt, có các điểm ảnh có thể nhận giá trị đa dạng trong miền giá trị của nó.

Câu 110

Cho 2 histogram tương ứng của 2 ảnh như hình dưới, nhận định nào dưới đây là đúng?



Histogram của ảnh I1

Histogram của ảnh I2

Đáp án:

- Độ sáng của ảnh I1 cao hơn độ sáng của ảnh I2.
- Độ sáng của ảnh I1 thấp hơn độ sáng của ảnh I2.
- Độ sáng của 2 ảnh tương tự nhau.
- Không so sánh được độ sáng của 2 ảnh này.

Câu 111

Cho ảnh đa mức xám 256 mức không nén, cần bao nhiêu bytes để lưu trữ mỗi điểm ảnh?

Đáp án:

- 1
- 3
- 8
- 24

Câu 112

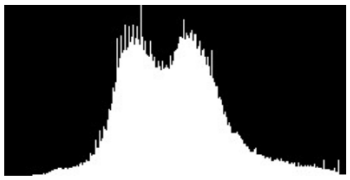
Trong không gian màu nào thành phần màu và độ sáng không được mã hóa tách biệt trong các kênh?

Đáp án:

- RGB
- HSV
- Lab
- YCbCr

Câu 113

Cho 2 histogram tương ứng của 2 ảnh như hình dưới, nhận định nào dưới đây là đúng?



Histogram của ảnh I1

Histogram của ảnh I2

Đáp án:

- Độ tương phản của ảnh I1 tốt hơn độ tương phản của ảnh I2.
- Độ tương phản của ảnh I2 tốt hơn độ tương phản của ảnh I1.
- Độ tương phản của ảnh I2 tương tự độ tương phản của ảnh I1.
- Không so sánh được độ tương phản của ảnh I1 và I2.

Câu 114

Nhận định nào dưới đây là đúng?

Đáp án:

- Histogram của hai ảnh khác nhau có thể giống nhau.
- Histogram của hai ảnh khác nhau luôn khác nhau.
- Histogram của ảnh luôn có 256 mức (256 bins).
- Nếu các đối tượng trong ảnh dịch sang trái 10 điểm ảnh, histogram của ảnh cũng được dịch sang trái.

Câu 115

Mục đích của cân bằng histogram là gì?

Đáp án:

- Tăng cường độ tương phản của ảnh.
- Tăng độ sáng của ảnh.
- Biểu diễn nội dung của ảnh.
- Giảm nhiễu.

Câu 116

Cho ảnh có đa mức xám 256 mức, cường độ sáng của điểm ảnh nhận giá trị trong khoảng nào?

Đáp án:

- Khoảng [0, 255]
- Khoảng [0, 100]
- Khoảng [0, 256]
- Khoảng [1, 256]

Câu 117

Nếu chúng ta chụp ảnh của cùng một đối tượng trong điều kiện chiếu sáng khác nhau, và biểu diễn chúng trong không gian màu Lab thì kênh màu nào có sự khác biệt lớn nhất giữa 2 ảnh?

Đáp án:

- L
- a
- b
- a và b

Câu 118

Có bao nhiêu kênh trong ảnh RGB?

Đáp án:

- 3
- 1
- 8
- 4

Câu 119

Mục đích của bộ phát hiện Canny?

Đáp án:

- Phát hiện biên
- Trích chọn đặc trưng cục bộ
- Trích chọn đặc trưng toàn cục
- Loại bỏ nhiễu

Câu 120

Bộ phát hiện Canny sử đụng đạo hàm bậc mấy trên ảnh?

Đáp án:

- Đạo hàm bậc nhất
- Đạo hàm bậc hai
- Cả đạo hàm bậc nhất và đạo hàm bậc 2
- Không sử dụng đạo hàm bậc 1 hay bậc 2.

Câu 121

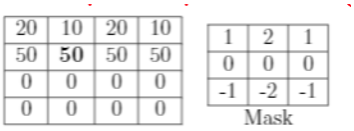
Cho ảnh gốc ở bên trái, bộ lọc nào đã được sử dụng để thu được ảnh kết quả ở bên phải??

Đáp án:

- Bộ lọc Sobel
- Bộ lọc trung vị
- Bộ lọc Gauss
- Bộ lọc trung bình

Câu 122

Cho một ma trận điểm ảnh (4x4) và một mặt nạ nhân chập, hãy cho biết giá trị của điểm ảnh (1,1) (điểm được in đậm) sau nhân chập?



Đáp án:

- -60
- 60
- 22
- 50

Câu 123

Nhận định nào dưới đây về nhân chập 2D KHÔNG chính xác?

Đáp án:

- Giá trị mới của điểm ảnh được tính bằng tổng có trọng số các giá trị điểm ảnh trong lân cận của nó..
- Cùng một hàm số được áp dụng lên tất cả các điểm ảnh.
- Nhân chập 2D có thể được dùng để loại nhiễu, tăng cường độ sắc nét của ảnh hoặc để phát hiện biên..
- Không đáp án nào trong số các đáp án được đề cập.

Câu 124

Mệnh đề nào sau đây là đúng?

Đáp án:

- Mặt nạ Laplace có thể dùng để tính đạo hàm bậc 2 của ảnh.
- Mặt nạ Laplace có thể dùng để tính đạo hàm bậc 1 của ảnh.
- Đạo hàm bậc 2 của ảnh không thể được sấp xỉ bởi phép nhân chập.

- Không đáp án nào trong số các đáp án được đề cập.

Câu 125

Điểm biên được xác định bằng cách nào ?

Đáp án:

- Tìm điểm đổi dấu (zero-crossing) trên đạo hàm bậc 2.
- Tìm cực trị địa phương trên đạo hàm bâc 1.
- Tìm điểm đổi dấu (zero-crossing) trên đạo hàm bậc 1.
- Tìm cực trị địa phương trên đạo hàm bâc 2.

Câu 126

Nhận định nào về đặc trưng ảnh sau đây là đúng?

Đáp án:

- Đặc trưng cục bộ mô tả nội dung một vùng nào đó trong ảnh.
- Đặc trưng cục bộ biểu diễn thông tin của toàn bộ bức ảnh.
- Histogram của ảnh là một đặc trưng cục bộ.
- Không đáp án nào trong số các đáp án được đề cập.

Câu 127

Nhận định nào về đặc trưng ảnh sau đây là đúng?

Đáp án:

- Đặc trưng toàn cục biểu diễn thông tin của toàn bộ bức ảnh.
- Đặc trưng cục bộ mô tả nội dung một vùng nào đó trong ảnh.
- SURF là một đặc trưng toàn cục.

- Không đáp án nào trong số các đáp án được đề cập.

Câu 128

Mặt nạ dưới đây dùng cho bộ lọc nào?

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Đáp án:

- Bộ lọc trung bình
- Bộ lọc trung vị
- Bộ lọc gauss
- Bộ lọc tăng cường độ sắc nét của cạnh

Câu 129

Mặt nạ dưới đây dùng cho bộ lọc nào?

1/16	2/16	1/16
2/16	4/16	2/16
1/16	2/16	1/16

Đáp án:

- Bộ lọc trung bình
- Bộ lọc trung vị
- Bộ lọc gauss
- Bộ lọc tăng cường độ sắc nét của cạnh

Câu 130

Các vùng để tính đặc trưng cục bộ được xác định bằng cách nào?

Đáp án:

- Sử dụng phương phân vùng ảnh
- Chia ảnh thành các mảnh sử dụng lưới chia định nghĩa trước
- Phát hiện các điểm đặc trưng và xác định vùng cục bộ xung quanh điểm đặc trưng đó.
- Tất cả các phương án được đề cập.

Câu 131

Hai ảnh dưới đây được là kết quả thu được khi áp dụng mặt nạ trung bình có kích thước khác nhau trên cùng một ảnh. Nếu ảnh bên trái là kết quả của bộ lọc có kích thước 9 x 9 thì ảnh bên phải là kết quả tương ứng của mặt nạ có kích thước nào?

Đáp án:

- 15 x 15
- 9 x 9
- 5 x 5
- 3 x 3

Câu 132

SIFT là gì?

Đáp án:

- Đặc trưng cục bộ
- Đặc trưng toàn cục
- Phương pháp tăng cường độ tương phản
- Bộ phát hiện biên

Câu 133

Giả sử bạn dùng K-means để phân tích dữ liệu từ Facebook để tìm các nhóm người dùng đặc biệt. Khi tăng số lượng nhóm K lên, lỗi phân cụm trên tập huấn luyện sẽ luôn giảm. Bạn có thể gặp khó khi chọn K để thu được kết quả phân cụm tốt nhất. Khi đó bạn nên làm gì?

Đáp án:

- Tìm một (hoặc vài) chuyên gia về lĩnh vực đó để đánh giá chất lượng của các nhóm/cụm tìm được
- Chọn K mà có lỗi phân cụm nhỏ nhất trên tập huấn luyện
- Chọn K mà có lỗi gần nhất với lỗi trung bình từ tất cả các thử nghiệm của bạn

Câu 134

Hold-out có phải là một phương pháp để tiền xử lý và hiểu dữ liệu?

Đáp án:

- Không, nó là một chiến lượng để đánh giá một mô hình
- Đúng, tất nhiên rồi
- Không, nó là một phương pháp để huấn luyện một mô hình từ một tập dữ liệu cho trước

Câu 135

Giả sử bạn đã xây dựng một hệ thống phát hiện các tấn công mạng và chắc chắn rằng hệ thống đó có độ chính xác (accuracy) trên tập kiểm thử là 99

Đáp án:

- Đánh giá của bạn về hệ thống đó có thể bị làm sai
- Độ chính xác có thể không phản ánh đúng mong muốn của sếp
- Sếp không có đủ tri thức để hiểu hệ thống đó và sự vất vả của bạn
- Tập huấn luyện có thể quá đơn giản.
- Bạn không may mắn

Câu 136

Đánh giá mô hình là

Đáp án:

- Việc đánh giá hiệu quả (chất lượng) của một mô hình hoặc phương pháp phân tích dữ liệu, bằng cách sử dụng một hoặc nhiều tập dữ liệu
- Việc đánh giá hiệu quả (chất lượng) của một mô hình hoặc phương pháp phân tích dữ liệu, chỉ bằng cách sử dụng các kịch bản thực tế
- Việc chúng ta khám phá một mô hình đã được học để tìm ra tri thức mới

Câu 137

Giả sử bạn huấn luyện một mô hình phân loại trên một tập huấn luyện gồm 10,000 điểm và thu được độ chính xác trên tập đó là 99

Đáp án:

- Đặt hệ số hiệu chỉnh (nếu có) bằng 0
- Huấn luyện trên nhiều dữ liệu hơn
- Dùng bước tối ưu tham số

- Bỏ bớt dữ liệu một cách ngẫu nhiên khi huấn luyện

Câu 138

Phát biểu nào sau đây là SAI?

Đáp án:

- Đánh giá mô hình và lựa chọn mô hình trong Học máy là hai thứ độc lập với nhau
- Đánh giá mô hình thường yêu cầu thực hiện bước lựa chọn mô hình
- Lựa chọn mô hình là một bước bắt buộc khi muốn so sánh nhiều mô hình (hoặc phương pháp) học máy khác nhau

Câu 139

Phát biểu nào sau đây là phù hợp nhất về lựa chọn mô hình (model selection)?

Đáp án:

- Lựa chọn mô hình quan tâm đến việc tìm thiết đặt tốt nhất về bộ (siêu) tham số trong một mô hình khi huấn luyện nó từ một tập dữ liệu. Đôi khi nó cũng nói đến việc lựa chọn một trong số các mô hình đang có.
- Lựa chọn mô hình chỉ quan tâm đến việc lựa chọn một mô hình tốt nhất từ một tập đang có.
- Các phát biểu khác đều sai.

Câu 140

Khi sử dụng một phương pháp để phân tích dữ liệu, hai lần chạy khác nhau có thể thu được hai kết quả khác nhau mặc dù sử dụng cùng thiết đặt cho bộ tham số. Lý do có thể từ đâu?

Đáp án:

- Do sự ngẫu nhiên khi chia tập dữ liệu đang có thành hai tập con dùng để huấn luyện và kiểm chứng
- Do việc sử dụng các thiết đặt khác nhau cho các tham số
- Do việc sử dụng các tập dữ liệu khác nhau
- Do dùng phương pháp đó sai cách
- Do tính ngẫu nhiên của thuật toán học/phân tích

Câu 141

Khi khám phá dữ liệu, bạn phát hiện ra rằng thuộc tính A có tương quan mạnh với nhãn lớp. Tuy nhiên, khi huấn luyện một mô hình học máy từ tập dữ liệu, A thường làm giảm đáng kể độ chính xác. Tại sao tình huống này có thể xảy ra?

Đáp án:

- A là một thuộc tính nhiễu
- A có tương quan âm với nhãn lớp
- Đánh giá của bạn có thể chưa kỹ lưỡng
- A có thể phổ biến và không có tính tách biệt
- Tình huống này không thể xảy ra

Câu 142

Khi khám phá dữ liệu, bạn phát hiện ra rằng thuộc tính A có tương quan rất bé với nhãn lớp. Tuy nhiên, khi huấn luyện một mô hình học máy từ tập dữ liệu, A thường làm tăng độ chính xác. Tại sao tình huống này có thể xảy ra?

Đáp án:

- A là một thuộc tính nhiễu
- A có tương quan âm với nhãn lớp
- Cách bạn đo độ tương quan có thể chưa mô tả đúng sự phụ thuộc ẩn giữa A và nhãn lớp

- A có thể cung cấp thêm tri thức cho mô hình
- A có thể phổ biến cho tất cả các nhãn lớp
- Tình huống này không thể xảy ra

Câu 143

3 tầng kiến trúc của ……là backend, artist, và scripting.

Đáp án:

- Seaborn
- Pyplot
- Matlab
- Matplotlib

Câu 144

Bạn sẽ làm gì khi muốn phân tích và khám phá dữ liệu?

Đáp án:

- Thống kê các thông số của dữ liệu (min, max, avg, std,…)
- Tính toán tần suất xuất hiện của các giá trị dữ liệu
- Vẽ biểu đồ histogram của dữ liệu
- Tất cả các đáp án khác đều đúng

Câu 145

Chỉ ra phát biểu đúng về biểu đồ scatter plot.

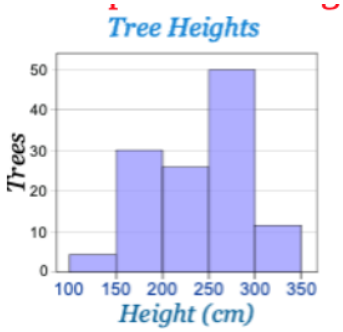
Đáp án:

- Tập hợp các điểm được vẽ theo cả hai chiều thẳng đứng và nằm ngang

- Tập hợp các điểm được vẽ ngẫu nhiên trong hệ trục toạ độ
- Tập hợp các điểm nằm tập trung quanh một đường thẳng
- Không phát biểu nào đúng

Câu 146

Chỉ ra phát biểu đúng về hình sau:



Đáp án:

- Biểu đồ cột về dữ liệu chiều cao của cây
- Histogram về dữ liệu chiều cao của cây
- Đồ thị hiển thị dữ liệu về số lượng cây
- Đồ thị hiển thị dữ liệu về chiều cao của cây

Câu 147

Khi phân tích histogram của dữ liệu, chúng ta muốn tìm kiếm những thông tin gì về dữ liệu?

Đáp án:

- Correlation
- Asymmetry

- Statistical information
- Outliers

Câu 148

Loại biểu đồ nào sau đây giúp biểu diễn trực quan dữ liệu dạng phân cấp tốt nhất?

Đáp án:

- Treemap
- Population pyramid
- Bar chart
- Các lựa chọn khác là sai

Câu 149

Loại biểu đồ nào thích hợp khi chúng ta muốn theo dõi sự thay đổi theo thời gian?

Đáp án:

- Line graph
- Column Graph
- Bar Graph
- Tất cả các lựa chọn khác

Câu 150

Loại công cụ trực quan hoá nào sẽ được sử dụng để biểu diễn độ phức tạp mã nguồn phần mềm?

Đáp án:

- Trực quan hoá khoa học

- Trực quan hoá toán học
- Trực quan hoá thông tin

Câu 151

Loại đồ thị nào ít nhập nhằng nhất và thường là lựa chọn tốt nhất để bắt đầu khám phá dữ liệu?

Đáp án:

- Table chart
- Pie Chart
- Radial column chart
- Bar chart

Câu 152

Một phiên bản của scatter plot cho phép hiển thị dữ liệu 3 chiều?

Đáp án:

- A heatmap
- A scatter map
- A bubble plot
- Các lựa chọn khác là sai

Câu 153

Một đối tượng cho phép giải thích cho các biểu tượng, màu sắc và các hình dạng mẫu được sử dụng trong các biểu đồ gọi là gì?

Đáp án:

- Legend

- Chart title
- Axis title
- Data label

Câu 154

Nhiệt độ thuộc loại dữ liệu nào trong các loại sau đây?

Đáp án:

- Dữ liệu rời rạc không sắp xếp
- Dữ liệu liên tục sắp xếp được
- Dữ liệu rời rạc sắp xếp được
- Dữ liệu liên tục không sắp xếp được

Câu 155

Những đặc trưng nào về dữ liệu có thể được trực quan hoá trong các biểu đồ scatterplot?

Đáp án:

- Correlation
- Associations
- Skewness
- Dispersion

Câu 156

Phát biểu nào đúng nhất về pie chart?

Đáp án:

- Pie chart được dùng khi chúng ta muốn thể hiện sự kết hợp của các thành phần khác nhau trong dữ liệu
- Pie chart là 1 đồ thị hình tròn được chia thành các mảng khác nhau, mỗi mảng biểu hiện sự thay đổi theo thời gian
- Pie chart được sử dụng khi muốn so sánh các hạng mục dữ liệu
- Các phát biểu khác là sai

Câu 157

Thông tin nào chúng ta có thể rút ra khi quan sát biểu đồ box plot?

Đáp án:

- Lower/upper quartile
- Gap
- Probability distribution
- Skewness

Câu 158

Thư viện nào cần được sử dụng nếu muốn trực quan hoá dữ liệu với Python?

Đáp án:

- Numpy
- Pandas
- Seaborn
- Pyplot, pandas, seaborn

Câu 159

Thư viện nào của Python thường được sử dụng để trực quan hoá dữ liệu?

Đáp án:

- NLTK, Spacy, ...
- Requests, Scrapy, BeautifulSoup, ...
- Tensorflow, Keras, scikit-learn, ...
- SciPy, NumPy, Matplotlib and Pandas, ...

Câu 160

Trong các phát biểu sau, đâu là phát biểu đúng nhất về việc lựa chọn kĩ thuật trực quan hoá phù hợp cho một loại dữ liệu?

Đáp án:

- Thu thập dữ liệu, Tổ chức dữ liệu và phân tích dữ liệu
- Sử dụng biểu đồ cột phù hợp cho tất cả các loại dữ liệu
- Tạo ra các câu hỏi từ một kĩ thuật trực quan hoá dữ liệu
- Tất cả các phát biểu khác đều đúng

Câu 161

Đâu là kết hợp đúng nhất về hàm (function) và tham số (parameter) để tạo ra 1 biểu đồ box plot trong Matplotlib?

Đáp án:

- Function = plot, and Parameter = type with value = ”box”
- Function = boxplot, and Parameter = type with value = ”plot”
- Function = plot and Parameter = kind with value = ”box”
- Function = plot and Parameter = kind with value = ”boxplot”

Câu 162

Đoạn code sau đây thể hiện đồ thị nào?

Đáp án:

- question.plot(kind='barh')
- Line graph
- Column Graph
- Bar Graph
- Các lựa chọn khác là sai

Phần tự luận

Câu 1: Mô tả phân tích về project của nhóm em

Project của nhóm em tập trung vào việc phân tích và đánh giá rủi ro tài chính của các doanh nghiệp sản xuất niêm yết trên sàn chứng khoán Việt Nam. Mục tiêu chính của project là xây dựng một mô hình dự đoán rủi ro tài chính dựa trên các chỉ số tài chính quan trọng như biên lợi nhuận, tỷ lệ nợ trên vốn chủ sở hữu (D/E), tỷ suất lợi nhuận trên vốn chủ sở hữu (ROE), và các chỉ tiêu tài chính khác.

Các bước thực hiện:

1. **Thu thập dữ liệu:** Dữ liệu được thu thập từ các báo cáo tài chính của các doanh nghiệp niêm yết trên sàn chứng khoán Việt Nam, bao gồm các chỉ số như doanh thu, lợi nhuận, nợ phải trả, và vốn chủ sở hữu.
2. **Tiền xử lý dữ liệu:** Dữ liệu được làm sạch, xử lý các giá trị thiếu, và chuẩn hóa để đảm bảo tính nhất quán.
3. **Phân tích dữ liệu:** Sử dụng các phương pháp phân tích thống kê và trực quan hóa dữ liệu để hiểu rõ hơn về tình hình tài chính của các doanh nghiệp.
4. **Xây dựng mô hình:** Áp dụng 13 mô hình, thuật toán học máy như Random Forest, Gradient Boosting, và Logistic Regression và các mô hình khác để xây dựng mô hình dự đoán rủi ro tài chính.
5. **Đánh giá mô hình:** Sử dụng các chỉ số như độ chính xác (Accuracy), Precision, Recall, và F1-Score để đánh giá hiệu quả của mô hình.
6. **Triển khai ứng dụng:** Phát triển một ứng dụng web để hiển thị kết quả dự đoán và cung cấp các biểu đồ trực quan hóa dữ liệu.

Kết quả:

- Mô hình đạt độ chính xác cao, giúp các doanh nghiệp nhận diện các yếu tố rủi ro tiềm ẩn và đưa ra các biện pháp điều chỉnh kịp thời.
- Ứng dụng web cung cấp giao diện thân thiện, dễ sử dụng, giúp người dùng dễ dàng theo dõi và đánh giá tình hình tài chính của doanh nghiệp.

Ý nghĩa:

Project không chỉ giúp các doanh nghiệp quản lý rủi ro tài chính hiệu quả hơn mà còn cung cấp công cụ hỗ trợ ra quyết định cho các nhà đầu tư và nhà quản lý.