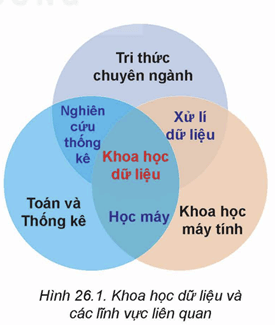
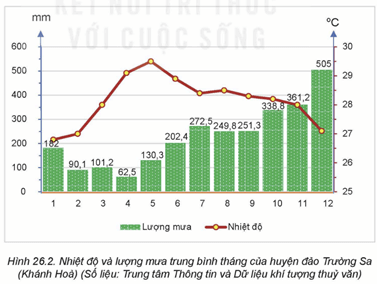
# Lý thuyết Bài 26: Làm quen với Khoa học dữ liệu

**Lý thuyết Tin học 12 Bài 26: Làm quen với Khoa học dữ liệu**  
**1. Khái niệm và mục tiêu của khoa học dữ liệu**  
Khoa học dữ liệu là lĩnh vực liên ngành kết hợp các phương pháp khoa học, quy trình, và thuật toán để khai thác tri thức từ dữ liệu. Nó tích hợp các kỹ thuật và công cụ từ:  
- Khoa học máy tính: Cung cấp công cụ và kỹ thuật để xử lý, phân tích và khai thác dữ liệu.  
-Toán học và thống kê: Cung cấp phương pháp phân tích dữ liệu và tạo mô hình dự đoán.  
- Tri thức chuyên ngành: Áp dụng các kiến thức đặc thù để hiểu và tận dụng dữ liệu nhằm hỗ trợ quyết định.  
Mục tiêu của khoa học dữ liệu là giúp tổ chức và cá nhân hiểu rõ dữ liệu và sử dụng tri thức từ dữ liệu để đưa ra quyết định phù hợp.  
  
b) Mục tiêu của khoa học dữ liệu:  
Mục tiêu chính của Khoa học dữ liệu là phân tích và khai phá dữ liệu để tạo ra tri thức và đưa ra các quyết định phù hợp. Các mục tiêu cụ thể bao gồm:  
1.Tổ chức và quản lý dữ liệu: Xây dựng và duy trì hệ thống tổ chức dữ liệu hiệu quả để đảm bảo tính toàn vẹn và sẵn sàng của dữ liệu, hỗ trợ phân tích và ra quyết định.  
2.Phân tích dữ liệu: Hiểu rõ nội dung và cấu trúc dữ liệu, xác định đặc điểm quan trọng, nhận diện nhóm và xu hướng để có cái nhìn toàn diện về dữ liệu.  
3. Trực quan hóa dữ liệu: Biểu diễn dữ liệu bằng sơ đồ, biểu đồ, hoặc hình ảnh để dễ dàng hiểu và phân tích dữ liệu.  
4. Tối ưu hóa quyết định: Sử dụng thuật toán tối ưu hóa để đưa ra quyết định tốt nhất dựa trên các ràng buộc và mục tiêu, chẳng hạn như tối ưu hóa quy trình sản xuất.  
5. Khám phá tri thức: Tìm ra các mối quan hệ ẩn chứa trong dữ liệu, xác định nguyên nhân và kết quả, và tạo ra tri thức mới để cải thiện hiểu biết và ứng dụng dữ liệu trong thực tiễn.  
Các mục tiêu này giúp tận dụng dữ liệu hiệu quả, cải thiện hoạt động và đưa ra quyết định thông minh.  
  
=>Khoa học dữ liệu là một lĩnh vực liên ngành, sử dụng các công cụ của khoa học máy tính, toán học và thống kê để khám phá tri thức từ dữ liệu, kết hợp những tri thức đó với tri thức chuyên ngành làm cơ sở cho những quyết định phù hợp. Các mục tiêu cụ thể của Khoa học dữ liệu bao gồm thăm dò, khai thác, phân tích, khai phá và trực quan hoá dữ liệu, làm cơ sở xây dựng mô hình dự đoán, dự báo và tối ưu hoá quyết định, hướng tới mục tiêu cao nhất đó là khám phá tri thức từ dữ liệu.  
**2. Một số thành tựu của khoa học dữ liệu**  
Sự phát triển của Khoa học dữ liệu, AI và Học máy gắn liền với sự ra đời và phát triển của dữ liệu lớn. Dữ liệu lớn được đặc trưng bởi "năm chữ V" (5V):  
- Khối lượng (Volume): Kích thước rất lớn của tập dữ liệu.  
- Vận tốc (Velocity): Tốc độ tạo ra và cần phân tích dữ liệu.  
- Sự đa dạng (Variety): Nhiều loại dữ liệu khác nhau như văn bản, âm thanh, video.  
- Giá trị (Value): Tính hữu ích của dữ liệu.  
- Tính xác thực (Veracity): Đảm bảo dữ liệu chính xác, vì dữ liệu lớn thường có nhiều nhiễu và sai số.  
Mối quan hệ giữa AI, Học máy và Khoa học dữ liệu là gắn bó và tương hỗ, với nhiều thành tựu được coi là thành tựu chung của cả ba lĩnh vực. Một số thành tựu cụ thể của Khoa học dữ liệu bao gồm:  
- Đổi mới quá trình ra quyết định: Khoa học dữ liệu giúp tổ chức và cá nhân đưa ra quyết định chính xác hơn bằng cách phân tích và khai thác dữ liệu, dự báo xu hướng, phát hiện hoạt động gian lận và tối ưu hóa phân bổ tài nguyên.  
-Tự động hoá và đổi mới sáng tạo: Khoa học dữ liệu sử dụng mô hình học máy để tự động hoá công việc lặp đi lặp lại, giúp tiết kiệm thời gian và chi phí, đồng thời thúc đẩy sự đổi mới và sáng tạo trong các lĩnh vực như sản xuất, hậu cần, dịch vụ khách hàng, và quản lý tài chính.  
-Cá nhân hóa dịch vụ và cải thiện trải nghiệm khách hàng: Khoa học dữ liệu hỗ trợ cung cấp dịch vụ cá nhân hóa thông qua phân tích dữ liệu khách hàng, nâng cao trải nghiệm và sự hài lòng của khách hàng.  
**Trắc nghiệm Tin học 12 Bài 26: Làm quen với Khoa học dữ liệu**  
**PHẦN I.** **Câu trắc nghiệm nhiều phương án lựa chọn. Thí sinh trả lời từ câu 1 đến câu 10. Mỗi câu hỏi thí sinh chỉ lựa chọn một phương án.**  
**Câu 1:** Khoa học dữ liệu kết hợp các phương pháp và công cụ từ những lĩnh vực nào?  
A. Khoa học máy tính, toán học, và thống kê  
B. Y học, sinh học, và hóa học  
C. Kỹ thuật, vật lý, và hóa học  
D. Triết học, lịch sử, và ngôn ngữ học  
**Hiển thị đáp án**  
  
**Giải thích:** Khoa học dữ liệu kết hợp các phương pháp và công cụ từ khoa học máy tính, toán học, và thống kê để xử lý, phân tích và khai thác dữ liệu.  
  
  
**Câu 2:** Một trong những mục tiêu chính của khoa học dữ liệu là gì?  
A. Chỉ lưu trữ dữ liệu  
B. Tạo ra tri thức và đưa ra các quyết định phù hợp  
C. Phát triển phần cứng máy tính  
D. Dự đoán thời tiết  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Mục tiêu chính của khoa học dữ liệu là phân tích và khai phá dữ liệu để tạo ra tri thức và đưa ra các quyết định phù hợp.  
  
  
**Câu 3:** Khái niệm "dữ liệu lớn" được đặc trưng bởi bao nhiêu chữ V?  
A. 3  
B. 4  
C. 5  
D. 6  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Dữ liệu lớn được đặc trưng bởi "năm chữ V": Khối lượng (Volume), Vận tốc (Velocity), Sự đa dạng (Variety), Giá trị (Value), và Tính xác thực (Veracity).  
  
  
**Câu 4:** Trong khoa học dữ liệu, mục tiêu của việc "trực quan hóa dữ liệu" là gì?  
A. Lưu trữ dữ liệu an toàn  
B. Tăng tốc độ truy cập dữ liệu  
C. Biểu diễn dữ liệu bằng sơ đồ, biểu đồ, hoặc hình ảnh  
D. Phân loại dữ liệu  
**Hiển thị đáp án**  
  
**Đáp án:** C  
**Giải thích:** Trực quan hóa dữ liệu là việc biểu diễn dữ liệu bằng sơ đồ, biểu đồ, hoặc hình ảnh để dễ dàng hiểu và phân tích dữ liệu.  
  
  
**Câu 5:** Một trong những thành tựu của khoa học dữ liệu là gì?  
A. Phát minh ra máy tính cá nhân  
B. Phát hiện hoạt động gian lận  
C. Xây dựng các nhà máy sản xuất  
D. Khám phá vũ trụ  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Khoa học dữ liệu giúp phát hiện hoạt động gian lận bằng cách phân tích và khai thác dữ liệu để dự báo xu hướng và xác định các hành vi bất thường.  
  
  
**Câu 6:** Một ứng dụng của học máy trong khoa học dữ liệu là gì?  
A. Tạo ra năng lượng từ hạt nhân  
B. Tự động hoá công việc lặp đi lặp lại  
C. Thay đổi khí hậu  
D. Nghiên cứu lịch sử  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Học máy trong khoa học dữ liệu giúp tự động hóa công việc lặp đi lặp lại, tiết kiệm thời gian và chi phí.  
  
  
**Câu 7:** Khối lượng dữ liệu trong "dữ liệu lớn" đề cập đến điều gì?  
A. Tốc độ xử lý dữ liệu  
B. Kích thước của tập dữ liệu  
C. Độ chính xác của dữ liệu  
D. Đa dạng loại dữ liệu  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Khối lượng trong "dữ liệu lớn" đề cập đến kích thước rất lớn của tập dữ liệu.  
  
  
**Câu 8:** Mục tiêu của việc tối ưu hóa quyết định trong khoa học dữ liệu là gì?  
A. Giảm kích thước dữ liệu  
B. Sử dụng thuật toán tối ưu hóa để đưa ra quyết định tốt nhất  
C. Tăng độ phức tạp của dữ liệu  
D. Lưu trữ dữ liệu an toàn  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Tối ưu hóa quyết định trong khoa học dữ liệu sử dụng thuật toán tối ưu hóa để đưa ra quyết định tốt nhất dựa trên các ràng buộc và mục tiêu.  
  
  
**Câu 9:** Một ví dụ về việc cải thiện trải nghiệm khách hàng thông qua khoa học dữ liệu là gì?  
A. Tăng số lượng sản phẩm  
B. Cung cấp dịch vụ cá nhân hóa  
C. Tạo ra các sản phẩm mới  
D. Phát triển các chương trình truyền hình  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** Khoa học dữ liệu hỗ trợ cung cấp dịch vụ cá nhân hóa thông qua phân tích dữ liệu khách hàng, nâng cao trải nghiệm và sự hài lòng của khách hàng.  
  
  
**Câu 10:** Mối quan hệ giữa AI, Học máy và Khoa học dữ liệu là gì?  
A. Độc lập và không liên quan  
B. Gắn bó và tương hỗ  
C. Cạnh tranh và đối lập  
D. Không có mối quan hệ  
**Hiển thị đáp án**  
  
**Đáp án:** B  
**Giải thích:** AI, Học máy và Khoa học dữ liệu có mối quan hệ gắn bó và tương hỗ, với nhiều thành tựu được coi là thành tựu chung của cả ba lĩnh vực.  
  
  
**PHẦN II. Câu trắc nghiệm đúng sai.** **Thí sinh trả lời từ câu 1 đến câu 2. Trong mỗi ý a), b), c), d) ở mỗi câu, thí sinh chọn đúng hoặc sai**  
**Câu 1:** Khái niệm và mục tiêu của khoa học dữ liệu sau đây đúng hay sai?  
a) Khoa học dữ liệu là lĩnh vực liên ngành kết hợp các phương pháp khoa học, quy trình, và thuật toán để khai thác tri thức từ dữ liệu.  
b) Mục tiêu của khoa học dữ liệu là chỉ để lưu trữ dữ liệu mà không cần phân tích.  
c) Trực quan hóa dữ liệu không phải là một phần của khoa học dữ liệu.  
d) Tri thức chuyên ngành không liên quan đến khoa học dữ liệu.  
**Hiển thị đáp án**  
  
a) Đúng: Khoa học dữ liệu tích hợp các kỹ thuật và công cụ từ khoa học máy tính, toán học, và thống kê để phân tích và khai thác dữ liệu.  
b) Sai: Mục tiêu chính của khoa học dữ liệu là phân tích và khai phá dữ liệu để tạo ra tri thức và đưa ra các quyết định phù hợp, không chỉ đơn thuần là lưu trữ.  
c) Sai: Trực quan hóa dữ liệu là một phần quan trọng của khoa học dữ liệu, giúp biểu diễn dữ liệu dưới dạng hình ảnh để dễ hiểu và phân tích.  
d) Sai: Tri thức chuyên ngành là một phần quan trọng trong khoa học dữ liệu, giúp hiểu và tận dụng dữ liệu trong các ngữ cảnh cụ thể.  
  
  
**Câu 2:** Phát biểu sau đây dứng hay sai về Một số thành tựu của khoa học dữ liệu?  
a) Khoa học dữ liệu giúp đổi mới quá trình ra quyết định bằng cách phân tích và khai thác dữ liệu.  
b) Khoa học dữ liệu chỉ áp dụng trong lĩnh vực tài chính.  
c) Tự động hóa là một trong những thành tựu của khoa học dữ liệu.  
d) Khoa học dữ liệu không liên quan đến việc cải thiện trải nghiệm khách hàng  
**Hiển thị đáp án**  
  
a) Đúng: Khoa học dữ liệu giúp tổ chức và cá nhân đưa ra quyết định chính xác hơn thông qua việc phân tích dữ liệu và dự báo xu hướng.  
b) Sai: Khoa học dữ liệu có ứng dụng rộng rãi trong nhiều lĩnh vực như sản xuất, y tế, marketing, và nhiều ngành khác.  
c) Đúng: Khoa học dữ liệu sử dụng mô hình học máy để tự động hoá các công việc lặp đi lặp lại, giúp tiết kiệm thời gian và chi phí.  
d) Sai: Khoa học dữ liệu giúp cải thiện trải nghiệm khách hàng thông qua phân tích dữ liệu để cung cấp dịch vụ cá nhân hóa và tối ưu hóa dịch vụ.  
  
  
**PHẦN III. Câu trả lời ngắn**. **Thí sinh trả lời từ câu 1 đến câu 3**  
**Câu 1:** Khoa học dữ liệu là lĩnh vực kết hợp các phương pháp khoa học, quy trình và thuật toán để làm gì?  
**Hiển thị đáp án**  
  
**Đáp án:** Để khai thác tri thức từ dữ liệu.  
**Giải thích:** Khoa học dữ liệu sử dụng các kỹ thuật từ khoa học máy tính, toán học và thống kê để xử lý và phân tích dữ liệu, nhằm trích xuất thông tin có giá trị và hỗ trợ ra quyết định.  
  
  
**Câu 2:** Khoa học dữ liệu tích hợp các kỹ thuật và công cụ từ những lĩnh vực nào?  
**Hiển thị đáp án**  
  
**Đáp án:** Khoa học máy tính, toán học và thống kê, tri thức chuyên ngành.  
**Giải thích:** Khoa học dữ liệu là một lĩnh vực liên ngành, kết hợp công cụ và kỹ thuật từ khoa học máy tính, phương pháp phân tích từ toán học và thống kê, và kiến thức chuyên ngành để khai thác và ứng dụng dữ liệu  
  
  
**Câu 3:** Một trong những thành tựu chính của khoa học dữ liệu là gì?  
**Hiển thị đáp án**  
  
**Đáp án:** Đổi mới quá trình ra quyết định.  
**Giải thích:** Khoa học dữ liệu giúp các tổ chức và cá nhân đưa ra quyết định chính xác hơn bằng cách phân tích và khai thác dữ liệu, dự báo xu hướng, và tối ưu hóa các quy trình, dẫn đến cải thiện hiệu quả hoạt động và quản lý.  
  
  
Xem thêm