# Phần I. Câu hỏi trắc nghiệm nhiều lựa chọn A, B, C, D

**\* Mức độ biết**

**Câu 1.** Đường truyền mạng là gì?

1. Thiết bị dùng để phát sóng Wi-Fi
2. Thiết bị lưu trữ dữ liệu mạng
3. **Phương tiện truyền dữ liệu giữa các thiết bị mạng**
4. Phần mềm quản lí kết nối mạng

**Câu 2.** Thiết bị nào dưới đây thường được dùng để kết nối mạng có dây?

1. **Switch**
2. Wi-Fi
3. Bluetooth
4. NFC

**Câu 3.** Loại cáp nào sau đây được sử dụng phổ biến trong mạng LAN?

1. Cáp quang
2. Cáp đồng trục
3. **Cáp xoắn đôi (Twisted Pair)**
4. Cáp HDMI

**Câu 4.** Tên tiếng Anh của đường truyền không dây là:

1. **Wireless**
2. Fiber optic
3. Twisted pair
4. Wired

**Câu 5.** Một trong những đặc điểm nổi bật của đường truyền không dây là:

1. Tốc độ luôn cao hơn đường truyền có dây
2. **Không cần dùng cáp vật lí**
3. Không thể bị nhiễu sóng
4. Chỉ hoạt động trong mạng LAN

**Câu 6.** Cáp quang truyền dữ liệu bằng:

1. Dòng điện
2. Sóng vô tuyến
3. **Ánh sáng**
4. Từ trường

**Câu 7.** Trong mạng máy tính, thiết bị nào sau đây không đóng vai trò truyền dữ liệu?

1. Router
2. Switch
3. Cáp mạng
4. **Máy in**

**Câu 8.** Bluetooth là ví dụ của loại đường truyền:

1. Có dây
2. **Không dây**
3. Quang học
4. Đồng trục

# Mức độ hiểu

**Câu 1.** So với đường truyền có dây, đường truyền không dây có ưu điểm nào nổi bật nhất?

1. Ít bị nhiễu sóng
2. Tốc độ cao hơn
3. Không bị giới hạn khoảng cách
4. **Linh hoạt và tiện lợi trong kết nối**

**Câu 2.** Vì sao mạng LAN thường sử dụng cáp xoắn đôi thay vì cáp quang?

1. Vì cáp xoắn đôi có tốc độ cao hơn
2. Vì cáp quang dễ nhiễm từ trường
3. Vì cáp xoắn đôi rẻ và dễ lắp đặt hơn
4. Vì cáp xoắn đôi chỉ dùng được cho mạng không dây

**Câu 3.** Mục tiêu chính của việc lựa chọn đường truyền mạng phù hợp là gì?

1. Giảm chi phí bảo trì
2. Tăng tốc độ xử lý của thiết bị đầu cuối
3. **Tối ưu hiệu quả kết nối và truyền dữ liệu**
4. Hạn chế người dùng truy cập

**Câu 4.** Vì sao cáp quang thường được sử dụng cho đường truyền tốc độ cao?

1. Vì giá thành rẻ
2. Vì dễ bảo trì
3. **Vì không bị ảnh hưởng bởi nhiễu điện từ và tốc độ cao**
4. Vì dễ lắp đặt

# Mức độ vận dụng

**Câu 1.** Một công ty muốn thiết kế hệ thống mạng nội bộ ổn định, tiết kiệm chi phí và dễ triển khai. Nên chọn loại đường truyền nào?

1. Cáp quang
2. Cáp đồng trục
3. **Cáp xoắn đôi**
4. Wi-Fi

**Câu 2.** Trong một khu vực nhiều tầng, bạn muốn kết nối mạng cho nhiều thiết bị mà không thể đi dây mạng. Giải pháp phù hợp là:

1. Dùng mạng LAN có dây
2. **Sử dụng Wi-Fi với bộ phát sóng ở mỗi tầng**
3. Dùng cáp quang kéo dài
4. Không thể kết nối mạng

**Câu 3.** Một bạn học sinh muốn truyền file giữa hai điện thoại mà không có Internet. Cách truyền nhanh nhất là:

1. Gửi qua email
2. **Dùng Bluetooth**
3. Dùng cáp mạng
4. Chép vào USB

**Câu 4.** Nếu muốn kết nối máy tính ở tầng trệt với máy in ở tầng 3 mà đảm bảo tốc độ và ổn định, bạn nên chọn cách nào?

1. Kéo dây mạng LAN
2. Sử dụng sóng Bluetooth
3. Dùng cáp HDMI
4. **Sử dụng Wi-Fi kết hợp bộ khuếch đại sóng**

# Phần II. Câu hỏi trắc nghiệm Đúng/Sai a) b) c) d) Câu 1. Ứng dụng đường truyền trong trường học

Một trường THPT đang triển khai hệ thống mạng nội bộ để phục vụ việc dạy học trực tuyến và truy cập Internet. Trường có nhiều phòng học ở các tầng khác nhau, đồng thời cần đảm bảo tốc độ truy cập ổn định và chi phí hợp lí.

Xác định câu Đúng/Sai về việc lựa chọn và sử dụng đường truyền mạng:

* 1. Cáp quang là lựa chọn phù hợp nhất vì giá rẻ và dễ lắp đặt trong khuôn viên trường học.
  2. **Việc sử dụng Wi-Fi giúp học sinh kết nối linh hoạt, đặc biệt là với thiết bị di động như laptop, điện thoại.**
  3. **Có thể kết hợp đường truyền có dây cho phòng máy và đường truyền không dây cho khu vực học sinh sử dụng thiết bị cá nhân.**
  4. Chỉ cần một thiết bị phát Wi-Fi đặt tại phòng bảo vệ là đủ để phủ sóng toàn bộ khuôn viên trường.

# Câu 2. Kết nối mạng tại gia đình

Một gia đình có 4 thành viên, sử dụng nhiều thiết bị như máy tính để bàn, laptop, điện thoại thông minh, và ti vi thông minh. Họ muốn kết nối Internet ổn định cho học tập, làm việc và giải trí.

Xác định câu Đúng/Sai về việc thiết lập hệ thống mạng gia đình:

1. **Sử dụng kết nối không dây Wi-Fi sẽ giúp các thiết bị di động truy cập mạng dễ dàng và linh hoạt hơn.**
2. **Cáp mạng có dây phù hợp để kết nối máy tính để bàn và Smart TV nhằm đảm bảo tốc độ truyền tải ổn định.**
3. Mỗi thiết bị trong gia đình cần có một đường truyền Internet riêng biệt để không bị gián đoạn khi truy cập.
4. **Nếu modem đặt ở tầng trệt, cần thêm bộ kích sóng (repeater) để mở rộng vùng phủ sóng Wi-Fi lên tầng trên.**

# Phần III. Tự luận

Em hãy trình bày ưu điểm và nhược điểm của ba loại đường truyền phổ biến: cáp xoắn đôi, cáp quang và kết nối không dây.

# Gợi ý đáp án:

* + Cáp xoắn đôi: giá thành rẻ, dễ lắp đặt, phù hợp với mạng LAN nhỏ; tuy nhiên, khoảng cách truyền ngắn, dễ nhiễu điện từ.
  + Cáp quang: tốc độ cao, truyền xa, không bị nhiễu điện từ; nhưng chi phí cao, đòi hỏi thiết bị chuyên dụng và khó lắp đặt.
  + Không dây (Wi-Fi, Bluetooth): linh hoạt, tiện lợi, không cần dây dẫn; nhưng dễ bị nhiễu sóng, tốc độ có thể không ổn định và phạm vi giới hạn.

# Phần I. Câu hỏi trắc nghiệm nhiều lựa chọn A, B, C, D

**\* Mức độ biết**

**Câu 1.** Thành phần nào dưới đây thường được sử dụng để kết nối các máy tính trong mạng nội bộ?

1. Repeater
2. **Switch**
3. Modem
4. Hub

**Câu 2.** Mô hình mạng phổ biến nhất trong thiết kế mạng nội bộ hiện nay là:

1. Hình lưới
2. Hình cây
3. **Hình sao**
4. Hình vòng

**Câu 3.** Trong thiết kế mạng, Router có vai trò chính là:

1. Khuếch đại tín hiệu
2. Chia sẻ máy in
3. **Kết nối nhiều mạng với nhau**
4. Ghi nhớ mật khẩu người dùng

**Câu 4.** Địa chỉ IP giúp xác định:

1. Dung lượng bộ nhớ máy tính
2. **Vị trí của thiết bị trong mạng**
3. Tên đăng nhập người dùng
4. Tốc độ xử lý của CPU

**Câu 5.** Trong mạng LAN, thiết bị kết nối trung tâm trong mô hình hình sao là:

1. Máy chủ
2. Router
3. **Switch**
4. Máy in

**Câu 6.** Việc đánh địa chỉ IP cho các thiết bị trong mạng nhằm:

1. Phân biệt tên người dùng
2. Truy cập trang web
3. **Định danh thiết bị trong mạng**
4. Mã hóa dữ liệu

**Câu 7.** Để đảm bảo an toàn khi thiết kế mạng, cần quan tâm đến:

1. Màu sắc thiết bị
2. Độ dài cáp mạng
3. **Biện pháp bảo mật và phân quyền**
4. Giá thành phần mềm

**Câu 8.** Trong một mạng nội bộ, thiết bị được gọi là máy trạm (client) là:

1. Máy chủ cung cấp dịch vụ
2. Thiết bị định tuyến
3. **Máy tính sử dụng tài nguyên từ máy chủ**
4. Máy in được chia sẻ

# Mức độ hiểu

**Câu 1.** Vì sao mô hình mạng hình sao thường được lựa chọn trong thiết kế mạng trường học hoặc văn phòng?

1. Tiết kiệm băng thông hơn hình vòng
2. **Dễ lắp đặt và dễ bảo trì khi có sự cố**
3. Có thể hoạt động không cần switch
4. Không cần cấu hình địa chỉ IP

**Câu 2.** Khi thiết kế mạng cho một khu vực có nhiều thiết bị di động, nên ưu tiên giải pháp nào?

1. Dùng cáp đồng trục
2. **Dùng Wi-Fi kết hợp với bộ mở rộng sóng**
3. Chỉ dùng mạng LAN có dây
4. Dùng mô hình mạng hình vòng

**Câu 3.** Điều gì xảy ra nếu hai thiết bị trong mạng được cấu hình cùng một địa chỉ IP?

1. Một thiết bị sẽ có tốc độ truy cập cao hơn
2. Hệ thống tự phân bổ địa chỉ mới
3. **Gây xung đột địa chỉ, dẫn đến mất kết nối**
4. Không ảnh hưởng đến hoạt động mạng

**Câu 4.** Mục đích của việc phân quyền truy cập trong thiết kế mạng là:

1. Giúp truy cập Internet nhanh hơn
2. **Bảo vệ thông tin và giới hạn quyền truy cập**
3. Dễ cấu hình thiết bị
4. Tiết kiệm điện năng

# Mức độ vận dụng

**Câu 1.** Một thư viện trường học có 10 máy tính học sinh và 2 máy tính quản lí. Nên chọn mô hình mạng nào và thiết bị nào để kết nối?

1. Hình vòng – Modem
2. **Hình sao – Switch**
3. Hình lưới – Hub
4. Hình cây – Repeater

**Câu 2.** Khi thiết kế mạng cho phòng máy tính với yêu cầu kết nối ổn định, chi phí thấp và dễ bảo trì, em nên:

1. Dùng mô hình lưới với router chuyên dụng
2. **Sử dụng mô hình hình sao với switch trung tâm**
3. Mỗi máy dùng một đường truyền riêng
4. Chỉ dùng kết nối không dây

**Câu 3.** Một doanh nghiệp muốn chia sẻ máy in giữa các máy tính trong cùng mạng LAN. Yêu cầu nào sau đây là đúng?

1. Máy in phải được gắn vào tất cả các máy tính
2. **Cần kết nối máy in với máy chủ hoặc thiết bị chia sẻ**
3. Phải cấu hình địa chỉ MAC cho máy in
4. Không cần nối mạng vẫn có thể in từ xa

**Câu 4.** Một bạn học sinh muốn thiết kế mạng trong gia đình để kết nối máy tính, điện thoại, TV. Giải pháp phù hợp là:

1. **Dùng mô hình mạng hình sao với router kết nối các thiết bị**
2. Chỉ cần modem, không cần thiết bị khác
3. Mỗi thiết bị sử dụng một địa chỉ IP giống nhau
4. Dùng cáp mạng nối từng thiết bị với máy chủ

# Phần II. Câu hỏi trắc nghiệm Đúng/Sai a) b) c) d) Câu 1. Thiết kế mạng cho thư viện trường học:

Một thư viện trường học chuẩn bị triển khai hệ thống mạng nội bộ để phục vụ tra cứu tài liệu, quản lý sách và truy cập Internet. Thư viện có 15 máy tính cho học sinh sử dụng và 5 máy dành cho nhân viên quản lý.

Xác định câu Đúng/Sai về những quyết định trong thiết kế mạng:

* 1. **Nên sử dụng mô hình mạng hình sao để dễ quản lý và khắc phục sự cố.**
  2. **Switch là thiết bị phù hợp để kết nối các máy tính trong mạng nội bộ của thư viện.**
  3. Tất cả các máy tính cần được nối trực tiếp với modem để truy cập Internet.
  4. **Việc đánh địa chỉ IP tĩnh cho các máy trong thư viện sẽ giúp quản lý mạng hiệu quả hơn**.

# Câu 2. Thiết kế mạng cho quán cà phê có Wi-Fi:

Một quán cà phê muốn cung cấp Wi-Fi miễn phí cho khách hàng. Không gian quán gồm hai tầng, tường dày, có nhiều thiết bị di động truy cập mạng cùng lúc. Chủ quán thuê một đường truyền cáp quang tốc độ cao.

Xác định câu Đúng/Sai về các quyết định thiết kế mạng phù hợp:

1. Chỉ cần một bộ phát Wi-Fi là đủ để phủ sóng toàn bộ không gian quán.
2. **Cần sử dụng router để chia sẻ kết nối Internet đến các thiết bị trong mạng.**
3. **Sử dụng thêm bộ mở rộng sóng (repeater) là cần thiết để tăng phạm vi phủ sóng.**
4. **Băng thông Internet thuê càng lớn thì số lượng thiết bị truy cập đồng thời càng nhiều** và ổn định.

# Phần III. Tự luận

Em hãy trình bày các bước cơ bản khi thiết kế một mạng LAN cho phòng học tin học gồm 20 máy tính. Nêu rõ vai trò của các thiết bị chính được sử dụng trong thiết kế mạng.

# Gợi ý đáp án:

* + Bước 1: Xác định nhu cầu sử dụng và số lượng máy cần kết nối.
  + Bước 2: Chọn mô hình mạng phù hợp (thường là mô hình hình sao).
  + Bước 3: Lựa chọn thiết bị mạng như switch, router, dây mạng.
  + Bước 4: Định cấu hình địa chỉ IP cho từng máy tính (IP tĩnh hoặc động).
  + Thiết bị chính:
    - Switch: Kết nối các máy tính với nhau.
    - Router: Kết nối mạng LAN với Internet.
    - Dây mạng: Truyền dữ liệu giữa các thiết bị.
    - Máy chủ (nếu có): Quản lý tài nguyên chung như máy in, phần mềm.

# Phần I. Câu hỏi trắc nghiệm nhiều lựa chọn A, B, C, D

**\* Mức độ biết**

**Câu 1.** Học máy là lĩnh vực của ngành nào?

1. Toán học
2. Mạng máy tính
3. **Trí tuệ nhân tạo**
4. Tin học văn phòng

**Câu 2.** Mục tiêu chính của học máy là gì?

1. **Dự đoán hoặc quyết định từ dữ liệu**
2. Lập trình các bước thủ công
3. Gắn nhãn dữ liệu
4. Gửi email tự động

**Câu 3.** Trong học máy, dữ liệu có nhãn là:

1. **Dữ liệu được gán nhãn hoặc giá trị đích**
2. Dữ liệu không có cấu trúc
3. Dữ liệu trống
4. Dữ liệu âm thanh

**Câu 4.** Phương pháp học nào sử dụng dữ liệu có nhãn?

1. **Học có giám sát**
2. Học không giám sát
3. Học sâu
4. Học truyền thống

**Câu 5.** Phân chia dữ liệu thành nhóm là ứng dụng của:

1. Học có giám sát
2. **Học không giám sát**
3. Lập trình hướng đối tượng
4. Cơ sở dữ liệu

**Câu 6.** Chức năng của học máy trong phân tích thị trường là gì?

1. Quản lý kho
2. Tính toán thuế
3. **Dự báo biến động giá và hỗ trợ chiến lược**
4. Tạo bản đồ

**Câu 7.** Ví dụ nào sau đây thuộc ứng dụng của học có giám sát?

1. Phân chia khách hàng
2. **Nhận dạng chữ viết**
3. Xác định chủ đề
4. Khai thác dữ liệu chưa gán nhãn

**Câu 8.** Một trong những ứng dụng quan trọng của học máy trong chăm sóc sức khỏe là:

1. Đếm bước đi
2. **Chẩn đoán bệnh**
3. Quản lý đơn thuốc
4. Tư vấn tài chính

# Mức độ hiểu

**Câu 1.** Vì sao học máy không cần lập trình chi tiết từng bước?

1. Vì nó tự viết mã
2. Vì nó dùng AI
3. **Vì mô hình tự học từ dữ liệu**
4. Vì có bộ nhớ lớn

**Câu 2.** Vai trò của dữ liệu có nhãn trong học có giám sát là gì?

1. Không cần thiết
2. **Giúp mô hình học và dự đoán chính xác**
3. Làm mô hình đơn giản hơn
4. Giảm chi phí tính toán

**Câu 3.** Dữ liệu không có nhãn giúp học máy thực hiện điều gì?

1. **Phân tích thông tin ẩn và nhóm dữ liệu**
2. Tự động hóa gửi mail
3. Gán nhãn ngẫu nhiên
4. Tạo dữ liệu âm thanh

**Câu 4.** Lý do học không giám sát phù hợp để phân khúc khách hàng là:

1. Dữ liệu khách hàng luôn được gán nhãn
2. Không cần mô hình hóa
3. **Có thể phân nhóm từ các đặc điểm tương đồng**
4. Không cần xử lý dữ liệu

# Mức độ vận dụng

**Câu 1.** Một công ty muốn phân loại thư rác từ hộp thư đến. Họ nên dùng phương pháp học máy nào?

1. Học không giám sát
2. **Học có giám sát**
3. Học sâu
4. Tìm kiếm tuần tự

**Câu 2.** Để phát hiện gian lận trong giao dịch tài chính không có sẵn nhãn, nên dùng phương pháp nào?

1. **Học không giám sát**
2. Học có giám sát
3. Học tăng cường
4. Xử lý ảnh

**Câu 3.** Một nhóm nghiên cứu muốn mô hình hóa mối quan hệ giữa đặc điểm bệnh nhân và nguy cơ bệnh tật để đưa ra dự đoán. Phương pháp phù hợp là:

1. Học không giám sát
2. **Học có giám sát**
3. Phân loại không giám sát
4. Dò tìm chuỗi

**Câu 4.** Trong dịch tự động văn bản giữa các ngôn ngữ, học máy có thể:

1. Tìm từ điển tương ứng
2. Dùng mẫu câu cố định
3. **Học từ dữ liệu dịch để xây dựng mô hình**
4. Sử dụng ngữ pháp cơ bản

# Phần II. Câu hỏi trắc nghiệm Đúng/Sai a) b) c) d)

**Câu 1.** Trong một công ty tài chính A, quản lý muốn nâng cấp hiệu quả công việc phân tích của nhân viên công ty nên quyết định sử dụng Học máy để giải quyết. Hãy chọn ý đúng sai cho việc sử dụng Học máy trong phân tích thị trường.

* 1. **Đưa ra dự báo biến động giá cả, trợ giúp hình thành các chiến lược kinh doanh dựa trên các mô hình dự đoán.**
  2. **Giúp người đầu tư và nhà kinh doanh hiểu rõ hơn về thị trường, hỗ trợ khả năng đưa ra quyết định đầu tư dựa trên thông tin và các phân tích kĩ thuật.**
  3. Quyết định đầu tư dựa trên thông tin và các phân tích kĩ thuật.
  4. **Tự động hóá quá trình ra quyết định đầu tư.**

**Câu 2.** Một học sinh đang tìm hiểu về phân loại dữ liệu trong học máy. Bạn ấy được yêu cầu xác định vai trò của dữ liệu có nhãn và phương pháp học phù hợp để lọc thư rác.

Cho các phát biểu sau, hãy xác định đâu là đúng:

1. Dữ liệu có nhãn là dữ liệu đã được gán nhãn hoặc giá trị đầu ra.
2. **Học có giám sát sử dụng dữ liệu có nhãn để huấn luyện mô hình.**
3. Dữ liệu không có nhãn có thể dùng để dự đoán đầu ra cụ thể.
4. **Bài toán lọc thư rác là ứng dụng điển hình của học có giám sát.**

# Phần III. Tự luận

**Câu 1. Một công ty thương mại điện tử lớn** muốn **cá nhân hoá trải nghiệm mua sắm** cho từng người dùng. Hiện tại, trang web chỉ hiển thị sản phẩm theo thứ tự phổ biến hoặc giá bán, không dựa trên sở thích cá nhân. Điều này dẫn đến tỷ lệ thoát trang cao và doanh thu từ người dùng cũ không tăng trưởng.

Công ty có lượng lớn dữ liệu người dùng như:

* + Lịch sử mua hàng,
  + Sản phẩm đã xem,
  + Sản phẩm đã thêm vào giỏ nhưng chưa mua,
  + Thời gian tương tác trên từng loại sản phẩm.

Hãy đề xuất giải pháp xây dựng hệ thống học máy giải quyết vấn đề trên.

**Trả lời:**

1. **Thu thập và chuẩn hóa dữ liệu hành vi người dùng:** Lấy dữ liệu từ các lần truy cập, sản phẩm đã xem/mua và thời gian tương tác để tạo **hồ sơ người dùng**.
2. **Áp dụng học không giám sát để phân nhóm người dùng (clustering):** Dùng các thuật toán như K-means hoặc Hierarchical clustering để **phân khúc người dùng** dựa trên hành vi tương đồng.

Ví dụ: nhóm thích đồ công nghệ, nhóm thích thời trang, nhóm mua vào khung giờ tối, v.v.

1. **Tùy biến giao diện theo từng nhóm:** Sau khi phân khúc, hiển thị các sản phẩm **ưu tiên phù hợp với từng nhóm**, giúp tăng tỷ lệ mua hàng.
2. **Liên tục cập nhật mô hình:** Do hành vi người dùng có thể thay đổi theo thời gian, mô hình học máy cần được **huấn luyện lại định kỳ** để đảm bảo độ chính xác cao.

**Phần I. Câu hỏi trắc nghiệm nhiều lựa chọn A, B, C, D**

**\* Mức độ biết**

**Câu 1. Mục tiêu chính của khoa học dữ liệu là gì?**

* 1. Thu thập dữ liệu
  2. Trực quan hóa dữ liệu
  3. **Phân tích, khai phá và sử dụng tri thức từ dữ liệu**
  4. Lưu trữ dữ liệu

# Câu 2. Toán học và thống kê hỗ trợ khoa học dữ liệu bằng cách nào?

1. Cung cấp các công cụ và kỹ thuật để xử lí, phân tích và khai phá dữ liệu
2. **Cung cấp cơ sở cho các phương pháp phân tích và khai phá dữ liệu**
3. Có vai trò quan trọng để hiểu ngữ cảnh và ý nghĩa của dữ liệu.
4. Trực quan hóa dữ liệu, tạo biểu đồ, thiết kế trang web

# Câu 3. Dữ liệu lớn thường được đặc trưng bởi “năm chữ V”. Đó là gì?

1. **Khối lượng (Volumn), vận tốc (Velocity), sự đa dạng (Variety), giá trị (Value), tính xác thực (Veracity)**
2. Âm lượng (Volumn), vận tốc (Velocity), sự đa dạng (Variety), giá cả (Value), tính xác thực (Veracity)
3. Khối lượng (Volumn), vận tốc (Velocity), phân loại (Variety), giá trị (Value), tính xác thực (Veracity)
4. Khối lượng (Volumn), vận tốc (Velocity), phân loại (Variety), giá cả (Value), tính xác thực (Veracity)

# Câu 4. Khối lượng dữ liệu lớn đề cập đến đặc trưng nào?

1. Sự đa dạng
2. Tốc độ
3. **Kích thước**
4. Tính xác thực

# Câu 5. Khoa học máy tính đóng vai trò gì trong khoa học dữ liệu?

1. **Cung cấp các công cụ và kỹ thuật để xử lí, phân tích và khai phá dữ liệu**
2. Cung cấp cơ sở cho các phương pháp phân tích và khai phá dữ liệu
3. Có vai trò quan trọng để hiểu ngữ cảnh và ý nghĩa của dữ liệu.
4. Trực quan hóa dữ liệu, tạo biểu đồ, thiết kế trang web

# Câu 6. "Veracity" trong dữ liệu lớn nói về điều gì?

1. Tốc độ dữ liệu được tạo ra
2. Nhiều loại dữ liệu khác nhau
3. Tính hữu ích của dữ liệu
4. **Sự cần thiết phải đảm bảo tính xác thực của dữ liệu**

# Câu 7. Trực quan hóa dữ liệu nhằm mục đích gì?

1. Hiểu rõ về nội dung, cấu trúc dữ liệu
2. **Biểu diễn dữ liệu trực quan, dễ hiểu**
3. Xây dựng, duy trì hệ thống tổ chức dữ liệu một cách khoa học
4. Quản lí hiệu quả các nguồn dữ liệu

# Câu 8. Để hiểu rõ giá trị sử dụng của dữ liệu trong thực tiễn, nhà khoa học dữ liệu cần kết hợp với yếu tố nào sau đây?

1. Thuật toán phân tích
2. Kỹ năng lập trình
3. **Tri thức chuyên ngành cụ thể**
4. Hệ điều hành phù hợp

# Câu 9. Phân tích dữ liệu nhằm mục đích gì?

1. Cải thiện quyết định dựa trên dữ liệu
2. Biểu diễn dữ liệu trực quan, dễ hiểu
3. **Xác định các đặc điểm quan trọng, nhận diện nhóm và xu hướng trong dữ liệu**
4. Tìm ra mối quan hệ ẩn chứa trong dữ liệu

# Câu 10. Y học cá nhân hóa ứng dụng Khoa học dữ liệu bằng cách nào?

1. Giới thiệu những sản phẩm hoặc nội dung mà khách hàng có thể quan tâm.
2. **Sử dụng bộ dữ liệu lớn về thông tin di truyền và y tế liên quan để đưa ra phát đồ điều trị phù hợp.**
3. Xem xét phản hồi khách hàng, có được những thông tin đầy đủ hơn về nhu cầu, sở thích và hành vi của họ giúp cải thiện trải nghiệm.
4. Sử dụng dữ liệu để cung cấp thông tin cho quá trình ra quyết định.

# \* Mức độ hiểu

**Câu 11. Vì sao cần có tri thức chuyên ngành trong khoa học dữ liệu?**

1. Để phân tích, khai phá dữ liệu và xây dựng các mô hình dự đoán.
2. **Để hiểu ngữ cảnh, ý nghĩa của dữ liệu và áp dụng kết quả phân tích chính xác**
3. Để kiểm tra giả thuyết, tính toán phân phối xác suất,…
4. Để xây dựng các thuật toán xử lí dữ liệu lớn một cách hiệu quả

# Câu 12. Mối liên hệ giữa học máy (machine learning) và khoa học dữ liệu được mô tả như thế nào?

1. Học máy thay thế toàn bộ khoa học dữ liệu
2. **Học máy là công cụ trong khoa học dữ liệu để trích xuất tri thức từ dữ liệu**
3. Không liên quan
4. Khoa học dữ liệu chỉ là một nhánh của học máy

# Câu 13. Vì sao việc tổ chức và quản lý dữ liệu quan trọng?

1. Tăng tính trực quan
2. Giảm lỗi người dùng
3. **Tạo hạ tầng vững chắc cho phân tích dữ liệu**
4. Giảm thời gian nhập dữ liệu

# Câu 14. Tại sao trực quan hóa lại hữu ích trong khoa học dữ liệu?

1. Tạo hình ảnh đẹp từ việc phân tích dữ liệu
2. Để chia sẻ dữ liệu hỗ trợ quá trình ra quyết định
3. **Giúp người dùng không chuyên dễ hiểu dữ liệu**
4. Tăng dung lượng tệp lưu trữ dữ liệu

# Câu 15. Khoa học dữ liệu giúp phát hiện gian lận như thế nào?

1. Tăng cường bảo mật trong các giao dịch tài chính
2. Gửi tin nhắn cảnh báo cho các đối tượng có hoạt động bất thường
3. **Phân tích mẫu giao dịch bất thường**
4. Khóa các tài khoản được cho là có các giao dịch bất thường.

# \* Mức độ vận dụng

**Câu 16. Một doanh nghiệp bán lẻ có thể sử dụng Khoa học dữ liệu như thế nào để tăng doanh thu?**

1. Điều chỉnh giá sản phẩm phù hợp với sự lên xuống của thị trường.
2. **Sử dụng hệ thống khuyến nghị dựa vào lịch sử mua hàng**
3. Mở thêm nhiều chi nhánh
4. Giảm bớt các nguyên liệu sản xuất tạo ra sản phẩm

# Câu 17. Trong lĩnh vực y tế, khoa học dữ liệu có thể giúp gì cho bác sĩ?

1. Phán đoán không cần thực hiện các loại xét nghiệm
2. **Dự đoán kết quả điều trị dựa trên dữ liệu bệnh nhân**
3. Giảm lượng bệnh nhân
4. Bỏ qua bước kiểm tra sức khỏe tổng quát

# Câu 18. Nếu dữ liệu của bạn thuộc nhiều dạng khác nhau như video, văn bản, và âm thanh, đặc trưng nào của dữ liệu lớn phù hợp nhất?

1. Vận tốc (Velocity)
2. Tính xác thức (Veracity)
3. **Sự đa dạng (Variety)**
4. Khối lượng (Volume)

# Câu 19. Một trong những thách thức lớn nhất trong việc xử lý dữ liệu lớn là gì?

1. Dễ bị mất điện
2. **Sự đa dạng và tốc độ sinh ra dữ liệu**
3. Giá trị của dữ liệu quá rõ ràng
4. Không thể lưu trữ

# Câu 20. Hệ thống khuyến nghị hoạt động hiệu quả nhất khi nào?

1. Khi có dữ liệu mô tả sản phẩm chính xác
2. **Khi người dùng để lại nhiều đánh giá và lịch sử hành vi**
3. Khi giảm số lượng sản phẩm
4. Khi có mạng xã hội

# Phần II. Câu hỏi trắc nghiệm Đúng/Sai a) b) c) d)

**Câu 1. Một công ty thương mại điện tử sử dụng Khoa học dữ liệu để phân tích lịch sử mua sắm và hành vi truy cập của khách hàng nhằm cá nhân hóa sản phẩm đề xuất. Câu hỏi đúng/sai:**

* 1. **Việc sử dụng dữ liệu lịch sử mua sắm giúp doanh nghiệp đề xuất sản phẩm phù hợp với khách hàng.**
  2. Khoa học dữ liệu nên sử dụng dữ liệu văn bản khi phân tích hành vi người dùng.
  3. **Trực quan hóa dữ liệu giúp doanh nghiệp dễ dàng nhìn nhận xu hướng và nhu cầu thị trường.**
  4. Sử dụng công cụ học máy là đủ để ra quyết định kinh doanh mà không cần tri thức chuyên ngành.

# Câu 2. Một nhà máy sản xuất sử dụng các thuật toán Khoa học dữ liệu để tối ưu hóa hiệu quả sản xuất và phân bổ tài nguyên.

1. **Khoa học dữ liệu có thể hỗ trợ tối ưu hóa dây chuyền sản xuất bằng cách phân tích dữ liệu vận hành.**
2. Dữ liệu về sản lượng và năng suất không liên quan đến phân tích khoa học dữ liệu.
3. **Các công cụ lập trình và thư viện là một phần của khoa học máy tính được dùng trong phân tích dữ liệu.**
4. Mục tiêu của Khoa học dữ liệu là xây dựng hệ thống phần mềm mà không đề cập đến khả năng ra quyết định.

# Phần III. Tự luận

**Câu 1. Giải pháp Phân Tích Dữ Liệu Lớn trong Kinh Doanh - Tối Ưu Hóa Chiến Lược Marketing**

Một công ty lớn trong lĩnh vực bán lẻ muốn tối ưu hóa chiến lược marketing của mình bằng cách phân tích dữ liệu hành vi khách hàng trên website và các nền tảng xã hội. Họ thu thập một lượng lớn dữ liệu từ các nguồn khác nhau, bao gồm thông tin khách hàng, thói quen mua sắm, lượt xem sản phẩm, tương tác trên mạng xã hội, v.v. Tuy nhiên, việc xử lý và phân tích dữ liệu này một cách nhanh chóng và hiệu quả trở thành một thách thức, vì khối lượng dữ liệu là quá lớn và có sự phân tán giữa các hệ thống khác nhau.

# Trả lời: Để giải quyết vấn đề này, công ty có thể áp dụng một số giải pháp sau:

* + Sử dụng công nghệ xử lý dữ liệu lớn (Big Data): Áp dụng các công nghệ như Hadoop, Spark để xử lý dữ liệu phân tán và hỗ trợ phân tích dữ liệu lớn một cách nhanh chóng, giúp việc xử lý và lưu trữ dữ liệu trở nên đơn giản hơn.
  + Ứng dụng thuật toán học máy (Machine Learning): Sử dụng các thuật toán học máy để phân tích hành vi của khách hàng, dự đoán xu hướng tiêu dùng và tối ưu hóa chiến lược marketing, giúp công ty đưa ra các quyết định chính xác và kịp thời.
  + Trực quan hóa dữ liệu: Sử dụng các công cụ trực quan hóa dữ liệu như Tableau hoặc Power BI để trình bày dữ liệu một cách sinh động và dễ hiểu. Điều này giúp các nhà quản lý và nhân viên marketing dễ dàng nhận ra các mẫu và xu hướng trong dữ liệu.

**Như vậy:** Việc sử dụng công nghệ Big Data, thuật toán học máy và trực quan hóa dữ liệu sẽ giúp công ty xử lý và phân tích dữ liệu khách hàng một cách nhanh chóng và hiệu quả. Các công cụ và giải pháp này sẽ giúp công ty tối ưu hóa chiến lược marketing, nâng cao hiệu quả bán hàng và mang lại lợi nhuận cao hơn. Việc áp dụng các giải pháp máy tính và thuật toán trong quản lý dữ liệu lớn đã trở thành một yếu tố quan trọng trong chiến lược kinh doanh hiện đại.

**Phần I. Câu hỏi trắc nghiệm nhiều lựa chọn A, B, C, D**

**\* Mức độ biết**

**Câu 1. Vai trò chính của máy tính trong quy trình khoa học dữ liệu là gì?**

1. Triển khai các dự án khoa học dữ liệu
2. Chuyển đổi dữ liệu thành tri thức và thông tin hữu ích
3. Cung cấp sức mạnh tính toán, lưu trữ và tự động hóa
4. Xử lí dữ liệu có kích thước lớn

# Câu 2. Giai đoạn nào sau đây không nằm trong quy trình khoa học dữ liệu?

1. Thu thập dữ liệu
2. Lập trình trang web
3. Khám phá tri thức
4. Báo cáo kết quả

# Câu 3. Phát biểu nào sau đây đúng về vai trò của máy tính trong trực quan hóa dữ liệu?

1. Máy tính hỗ trợ tạo ra các bảng tính để thống kê dữ liệu một cách trực quan.
2. Máy tính không có vai trò lớn
3. Máy tính hỗ trợ tạo ra các biểu diễn dữ liệu trực quan
4. Máy tính dùng để lưu trữ dữ liệu trực quan

# Câu 4. Công nghệ nào cho phép xử lý dữ liệu mà không cần đầu tư phần cứng và những cơ sở hạ tầng đắt tiền?

1. Hệ điều hành
2. Phần mềm kế toán
3. Điện toán đám mây
4. Ổ đĩa di động

# Câu 5. Máy giải trình tự gene thường chỉ xác định được:

1. Toàn bộ hệ gene
2. Vị trí của từng gene
3. Các đoạn nucleotide ngắn
4. Vị trí của các đoạn nucleotide trên hệ gene

# Câu 6. Đâu là một ứng dụng của xử lý song song?

1. Làm sạch dữ liệu và huấn luyện mô hình.
2. Tăng tốc phân tích dữ liệu lớn
3. Cung cấp tài nguyên tính toán
4. Cài đặt phần mềm

# Câu 7. Lợi ích chính của việc tự động hóa trong khoa học dữ liệu là gì?

1. Giúp tăng số lượng nhân viên
2. Tạo ra nhiều báo cáo giấy
3. Giảm lỗi thủ công và tăng tốc xử lý
4. Giảm yêu cầu về dữ liệu

# Câu 8. Lưu trữ dữ liệu của HGP gặp thách thức vì:

1. Dữ liệu dễ mất
2. Thiếu điện
3. Khối lượng dữ liệu quá lớn
4. Không có người lưu

# Câu 9. Tính năng nào của máy tính giúp phân tích dữ liệu theo thời gian thực?

1. Giao diện người dùng đẹp
2. Khả năng xử lý nhanh và mạnh
3. Phần mềm gõ văn bản
4. Màn hình độ phân giải cao

# Câu 10. Một lợi ích của phân tích thời gian thực trong dự án khoa học là gì?

1. Có thêm thời gian nghỉ
2. Giảm độ chính xác
3. Ra quyết định nhanh chóng
4. Không cần kiểm tra kết quả

# \* Mức độ hiểu

**Câu 11. Nếu không có máy tính, bước nào sau đây trong phân tích dữ liệu sẽ bị ảnh hưởng nhiều nhất?**

1. Thu thập dữ liệu từ người
2. Vẽ hình minh họa
3. Huấn luyện mô hình học máy
4. Tạo thư mục

# Câu 12. Nếu không có điện toán đám mây, hạn chế lớn nhất với các nhà nghiên cứu nhỏ lẻ là gì?

1. Không thể gặp nhau
2. Không có phần mềm
3. Thiếu tài nguyên phần cứng để xử lý dữ liệu
4. Không hiểu thuật toán

# Câu 13. Lưu trữ hàng trăm gigabyte dữ liệu đòi hỏi:

1. Bộ nhớ RAM lớn
2. Ổ cứng di động cá nhân
3. Cơ sở hạ tầng lưu trữ quy mô lớn
4. Sách vở

# Câu 14. Việc huấn luyện mô hình học máy yêu cầu nhất điều gì dưới đây?

1. Hình ảnh đẹp
2. Tài nguyên tính toán lớn và dữ liệu chất lượng
3. Thiết kế dễ nhìn
4. File PDF

# Câu 15. Một nhà khoa học dữ liệu muốn đảm bảo kết quả có thể lặp lại, cần?

1. Ghi nhớ trong đầu
2. Ghi vào giấy
3. Dùng quy trình tự động và công cụ kiểm thử
4. Tạo ảnh minh họa

# \* Mức độ vận dụng

**Câu 16. Khi xây dựng một mô hình học máy trên dữ liệu lớn, yếu tố nào sau đây cần ưu tiên?**

1. Thiết kế website đẹp
2. Khả năng tính toán và lưu trữ dữ liệu
3. Chia sẻ trên mạng xã hội
4. Thiết kế đồ họa

# Câu 17. Muốn tăng tốc quá trình huấn luyện mô hình AI, bạn nên áp dụng?

1. Viết tay dữ liệu
2. Dùng bộ xử lý song song hoặc GPU
3. Thêm hình ảnh nền
4. Sử dụng máy in

# Câu 18. Trong dự án HGP, việc nào sau đây thể hiện rõ tính “tích hợp dữ liệu”?

1. Một nhóm tự làm tất cả
2. Nhiều tổ chức cùng chia sẻ dữ liệu gene
3. Dùng bảng tính đơn giản
4. Sử dụng máy ảnh

# Câu 19. Vì sao cần sử dụng các thuật toán có khả năng mở rộng khi xử lý dữ liệu di truyền?

1. Dễ học hơn
2. Dùng cho mọi môn học
3. Đáp ứng quy mô dữ liệu ngày càng lớn
4. Không có lý do

# Câu 20. Nếu bạn cần trực quan hóa xu hướng tăng trưởng qua nhiều năm, bạn nên dùng công cụ nào?

1. Bộ xử lý văn bản
2. Thư viện trực quan hóa như Matplotlib, Tableau
3. Công cụ chat
4. Máy tính bỏ túi

# Phần II. Câu hỏi trắc nghiệm Đúng/Sai a) b) c) d)

**Câu 1.** Một nhóm các nhà khoa học từ nhiều quốc gia cùng nhau tham gia HGP. Mỗi nhóm phụ trách giải mã một phần gene khác nhau, dữ liệu được chia sẻ qua nền tảng số và hợp nhất để hoàn chỉnh bản đồ gene người.

* 1. HGP là dự án chỉ do Mỹ thực hiện và không cần sự hợp tác quốc tế.
  2. Việc tích hợp dữ liệu từ nhiều nhóm giúp hoàn chỉnh bản đồ gene người nhanh và hiệu quả hơn.
  3. Việc dùng nền tảng điện toán đám mây đã giúp các nhà khoa học dễ dàng chia sẻ và truy cập dữ liệu di truyền mọi lúc, mọi nơi.
  4. Nếu không có công nghệ lưu trữ và xử lý dữ liệu hiện đại, khối lượng dữ liệu hàng trăm gigabyte của HGP sẽ gây quá tải và gián đoạn quá trình nghiên cứu.

**Câu 2. Phân tích dữ liệu thời gian thực trong dịch tễ học:** Cô Lan là nhà khoa học dữ liệu trong lĩnh vực y tế công cộng. Cô theo dõi sự lây lan của một bệnh truyền nhiễm qua các khu vực và dùng mô hình để dự đoán điểm nóng trong tương lai nhằm đưa ra biện pháp kịp thời.

1. Cô Lan có thể sử dụng biểu đồ và bản đồ nhiệt để trực quan hóa dữ liệu lây nhiễm theo thời gian thực.
2. Dữ liệu thời gian thực luôn sẵn sàng và không bao giờ bị thiếu sót.
3. Cô Lan xây dựng một dashboard (bảng điều khiển) tương tác để theo dõi sự lan truyền dịch bệnh theo từng khu vực cụ thể.
4. Nếu không có thuật toán có khả năng mở rộng, hệ thống phân tích sẽ bị chậm trễ khi số lượng điểm dữ liệu tăng đột biến theo thời gian.

# Phần III. Tự luận

**Câu 1.** Nêu giải pháp để ứng phó với dịch bệnh nhờ dữ liệu thời gian thực, giúp cơ quan y tế phản ứng nhanh, chính xác và có cơ sở khoa học trong việc ra quyết định phòng chống dịch.

**Trả lời:**

* **Xây dựng hệ thống thu thập dữ liệu tự động theo thời gian thực** (Thiết lập biểu mẫu số hóa tại các trạm y tế, kết nối về một trung tâm quản lý dữ liệu. Dữ liệu được cập nhật liên tục (thời gian thực hoặc theo đợt) từ các điểm giám sát.)
* **Ứng dụng công cụ trực quan hóa dữ liệu** (Dùng bản đồ nhiệt (heatmaps), biểu đồ, đồ thị để hiển thị mức độ lây lan dịch bệnh theo khu vực. Tạo dashboard tương tác giúp lãnh đạo y tế theo dõi biến động ca bệnh ngay lập tức.)
* **Tích hợp mô hình dự báo dịch tễ học** (Áp dụng các mô hình học máy hoặc toán học (ví dụ: SIR, SEIR) để dự đoán xu hướng dịch bệnh. Cảnh báo sớm các “điểm nóng” giúp chủ động phân bổ nguồn lực và triển khai phòng chống.)
* **Tối ưu hiệu năng hệ thống (**Đảm bảo nền tảng xử lý dữ liệu có khả năng mở rộng, không bị nghẽn khi dữ liệu tăng đột biến. Sử dụng hạ tầng mạnh như HPC hoặc cloud để giữ ổn định khi dịch lan rộng.)