|  |  |
| --- | --- |
| **Trường: ………………………****Tổ: ……………………………** | Họ và tên giáo viên:  |

## CHỦ ĐỀ F: GIẢI QUYẾT VẤN ĐỀ VỚI SỰ TRỢ GIÚP CỦA MÁY TÍNH

## FCS1. GIỚI THIỆU HỌC MÁY VÀ KHOA HỌC DỮ LIỆU

## BÀI 1. GIỚI THIỆU VỀ HỌC MÁY

Môn Tin Học; Lớp: 12

Thời gian thực hiện: 2 tiết

**I. MỤC TIÊU**

**1. Kiến thức**

Bài học cung cấp những kiến thức:

- Khái niệm Học máy.

- Hai loại mô hình gọc máy chính: học có giám sát và học không giám sát.

- Một số ứng dụng của Học máy.

**2. Năng lực**

Bài học góp phần củng cố và phát triển cho học sinh những năng lực với biểu hiện cụ thể như sau:

***Năng lực chung:***

*- Tự chủ và tự học:*HS biết tự tìm kiếm, chuẩn bị và lựa chọn tài liệu, phương tiện học tập trước giờ học, quá trình tự giác tham gia các và thực hiện các hoạt động học tập cá nhân trong giờ học ở trên lớp,…

- *Giao tiếp và hợp tác:*Khả năng phân công và phối hợp thực hiện nhiệm vụ học tập.

- *Giải quyết vấn đề và sáng tạo:*HS đưa ra các phương án trả lời cho câu hỏi, bài tập xử lý tình huống, vận dụng kiến thức, kĩ năng của bài học để giải quyết vấn đề thường gặp.

***Năng lực Tin học:***

- *Nla (Sử dụng và quản lí các phương tiện CNTT và truyền thông) và NLc (Giải quyết vấn đề với sự hỗ trợ của CNTT và truyền thông):*

+ Giải thích được sơ lược về khái niệm Học máy.

+ Nêu được vai trò của Học máy trong những công việc như: lọc thư rác, chẩn đoán bệnh, phân tích thị trường, nhận dạng tiếng nói và chữ viết, dịch tự động,…

**3. Phẩm chất**

- *Chăm chỉ:* Chủ động tìm hiểu và cập nhật những kiến thức mới, không ngừng học hỏi.

- *Trách nhiệm:* Hoàn thành các nhiệm vụ học tập, tích cực trong hoạt động nhóm.

**II. THIẾT BỊ DẠY HỌC VÀ HỌC LIỆU**

* GV: SGK, SBT Tin học 12 – Định hướng Khoa học máy tính – Cánh diều, bài trình chiếu (Slide), máy chiếu.
* HS: SGK, SBT Tin học 12 – Định hướng Khoa học máy tính – Cánh diều, vở ghi.

**III. TIẾN TRÌNH DẠY HỌC**

**A. HOẠT ĐỘNG KHỞI ĐỘNG**

**a) Mục tiêu:** Tạo hứng thú học tập cho HS, giúp HS thấy được lợi ích của máy tính trong việc tự động thực hiện nhiệm vụ thay con người.

**b) Nội dung:**HS nghiên cứu nội dung hoạt động **Khởi động** SGK trang 126, suy nghĩ trả lời câu hỏi theo kiến thức của mình.

**c) Sản phẩm:** Từ yêu cầu, HS vận dụng kiến thức đã học để trả lời câu hỏi **Khởi động** SGK trang 126.

**d) Tổ chức thực hiện:**

**Bước 1:** **GV chuyển giao nhiệm vụ:**

- GV chia lớp thành các nhóm 3 – 4 HS, sau đó nêu câu hỏi **Khởi động** tr.126 SGK cho các nhóm thảo luận:

*Cho Hình 1 và ba nhãn phân loại là “vuông”, “tròn”, “tam giác”. Cần gán nhãn phân loại cho từng đối tượng trong Hình 1. Em hãy trả lời các câu hỏi sau:*

*1) Nếu con người thực hiện thì nhiệm vụ trên là dễ hay khó?*

*2) Theo em, máy tính có thể tự động thực hiện nhiệm vụ trên thay con người hay không? Lập trình để máy tính làm công việc này là dễ hay khó?*

**

*Hình 1. Ví dụ một số đối tượng cần phân loại*

**Bước 2: HS thực hiện nhiệm vụ học tập:**

- Các nhóm HS thảo luận và trả lời câu hỏi.

- GV quan sát quá trình các nhóm thảo luận, giải đáp thắc mắc nếu HS chưa rõ.

**Bước 3:** **Báo cáo kết quả hoạt động, thảo luận:**

- GV mời một số nhóm báo cáo kết quả thảo luận.

***Gợi ý trả lời:***

*1) Nếu con người thực hiện thì nhiệm vụ trên là dễ vì:*

* *Hình ảnh có độ phân giải cao, các đối tượng được phân biệt rõ ràng.*
* *Các hình dạng cơ bản (vuông, tròn, tam giác) dễ nhận biết.*
* *Nhiệm vụ chỉ yêu cầu gán nhãn đúng cho từng đối tượng.*

*2) Máy tính có thể tự động thực hiện nhiệm vụ trên thay con người:*

* *Sử dụng các thuật toán học máy, cụ thể là phân loại hình ảnh, để tự động gán nhãn cho các đối tượng.*
* *Các mô hình học máy đã được huấn luyện trên tập dữ liệu ảnh lớn có thể nhận diện các hình dạng cơ bản với độ chính xác cao.*

*Việc lập trình để máy tính làm công việc này là khó vì:*

* *Yêu cầu kiến thức chuyên môn về học máy, xử lý ảnh và lập trình.*
* *Cần thu thập và chuẩn bị tập dữ liệu huấn luyện phù hợp.*
* *Lựa chọn và điều chỉnh mô hình học máy phù hợp với nhiệm vụ.*

*Tuy nhiên:*

* *Việc lập trình sẽ dễ dàng hơn nếu sử dụng các thư viện và công cụ học máy có sẵn.*
* *Có nhiều mô hình học máy được phát triển sẵn có thể áp dụng cho bài toán phân loại hình ảnh.*

*-*Các nhóm khác lắng nghe và góp ý.

**Bước 4: Đánh giá kết quả thực hiện:**

GV đánh giá kết quả của HS, dẫn dắt HS vào bài học mới: *Bài toán trong phần Khởi động có thể được giải quyết bằng thuật toán Học máy****.****Ở Chủ đề A, bài “Giới thiệu về Trí tuệ nhân tạo” cũng đã đề cập đến Học máy như một nhánh nghiên cứu trong ngành AI nhằm làm cho máy tính có khả năng học từ dữ liệu.**Vậy để giúp các em hiểu hơn về lĩnh vực này,**chúng ta sẽ cùng nhau đến với* ***Bài 1: Giới thiệu về Học máy.***

**B. HÌNH THÀNH KIẾN THỨC MỚI**

**Hoạt động 1. Tìm hiểu khái niệm học máy**

**a) Mục tiêu:** Giải thích được một cách sơ lược về khái niệm học máy.

**b) Nội dung:**

* GV giới thiệu sợ lược về khái niệm học máy.
* GV chia sẻ cho HS quá trình ”học” thông dụng của mô hình học máy.
* Sau khi nắm được sợ lược về học máy và quá trình ”học”, GV giới thiệu cho HS 2 phương pháp học máy cơ bản là học có giám sát và học không giám sát (tương ứng với hai loại dữ liệu là dữ liệu có nhãn và dữ liệu không có nhãn).
* Sau khi giới thiệu các nội dung về học máy, GV cho HS thực hiện hoạt động Làm trong SGK để củng cố lại kiến thức đã học.

**c) Sản phẩm:**

* HS hiểu được sơ lược về khái niệm học máy.
* HS hiểu được về quá trình ”học” thông dụng của mô hình học máy.
* HS nắm được hai phương pháp học máy cơ bản là học có giám sát và học không giám sát.

**d) Tổ chức thực hiện**

|  |  |
| --- | --- |
| **Sản phẩm dự kiến** | **Hoạt động của GV và HS** |
| **1. KHÁI NIỆM VỀ HỌC MÁY** *Học máy (Máy học) là một lĩnh vực của AI tập trung vào việc phát triển các thuật toán và mô hình cho phép máy tính tự học và cải thiện từ dữ liệu để đưa ra dự đoán hoặc quyết định dựa trên dữ liệu mà không cần lập trình rõ ràng.**A. Mô hình học máy:**Thuật toán học máy:* Thuật toán học máy rút ra các thông tin liên quan tới dưc liệu, các đặc điểm chung quan trọng,… từ dữ liệu đầu vào.*Mô hình học máy:*Các mô hình học máy được tạo ra từ các thuật toán học máy và trải qua quá trình huấn luyện bằng các dữ liệu huấn luyệnQua trình huấn luyện nhằm tạo ra mô hình học máy để giải quyết một bài toán cụ thể.*B, Quy trình học máy:** Thu thập dữ liệu
* Chuẩn bị dữ liệu
* Xây dựng mô hình
* Đánh giá mô hình
* Triển khai ứng dụng mô hình

**Câu hỏi củng cố kiến thức****Câu 1.** Chọn phương án trả lời đúng. Học máy là:A. Chương trình máy tính có khả năng đưa ra quyết định hay dự đoán dựa trên dữ liệu.B. Khả năng máy tính phân tích dữ liệu thu nhận được để đưa ra dự đoán hoặc quyết định dựa trên các quy tắc được xác định rõ ràng.C. Việc sử dụng các phương pháp và kĩ thuật cho phép máy tính học từ dữ liệu để đưa ra dự đoán hoặc quyết định mà không cần lập trình cụ thể.D. Chương trình máy tính có khả năng tự cải thiện hiệu suất thực hiện nhiệm vụ thông qua việc cập nhật các dữ liệu mới sau khi hoàn thành nhiệm vụ đó nhiều lần.**Câu 2.** Tại sao cần chia dữ liệu học máy thành hai phần: dữ liệu huấn luyện và dữ liệu kiểm tra?Trả lời:1. C. 2. Trong học máy, cần chia dữ liệu thành hai phần khác nhau: dữ liệu huấn luyện để huấn luyện mô hình (“dạy” máy tính học) và dữ liệu kiểm thử để đánh giá hiệu suất của mô hình, giúp xác định xem mô hình có hoạt động tốt với dữ liệu mới, chưa từng gặp trước đó không. Nếu sử dụng cùng một dữ liệu cho cả huấn luyện và kiểm thử, mô hình có thể chỉ học cách "ghi nhớ" dữ liệu chứ không thực sự "hiểu" hoặc tổng quát hoá từ dữ liệu. Chia dữ liệu giúp ngăn chặn vấn đề này (thuật ngữ chuyên môn trong Học máy được gọi là “quá khớp” – overfitting). Dữ liệu kiểm thử cung cấp một phương pháp đánh giá khách quan về khả năng tổng quát hoá và dự đoán của mô hình trên dữ liệu chưa biết. Như vậy, việc phân chia dữ liệu giúp đảm bảo mô hình học máy được xây dựng không chỉ hoạt động tốt trên dữ liệu mà nó đã gặp trong quá trình huấn luyện mà còn có khả năng áp dụng trong thực tế với dữ liệu mới. | **Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Đặt câu hỏi*?1.*Chắc hẳn mỗi bạn đều có một thư điện tử, các bạn đã từng thấy các thư được chuyển và mục có tên là Spam (thư rác), các em thấy các thư này có đặc điểm gì mà được chuyển vào Spam.?2. Thuật toán học máy có chức năng gì, thế nào là mô hình học máy?**HS:** Lắng nghe để thực hiện yêu cầu của GV.*?3* Em hãy tóm tắt các bước của quy trình học máy?**HS:** Lắng nghe để thực hiện yêu cầu của GV.**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm‌ vụ:****HS:‌** Tìm câu trả lời từ hiểu biết hoặc từ SGK. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:****GV:** Điều khiển hoạt động của của các HS, cho HS phát biểu, cho HS nhận xét nhau.**‌HS:‌** ‌ ‌Phát biểu trả lời câu hỏi.Các‌‌ HS ‌nhận‌ ‌xét nhau.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌ định:‌ ‌‌**GV‌ chính xác lại các câu trả lời và chính xác lại đáp án. |

**Hoạt động 2. Tìm hiểu học có giám sát và học không giám sát**

**a) Mục tiêu:** Giúp HS biết được trong Học có giám sát, học không giám sát. Học có giám sát được ứng dụng trong các bài toán phân loại. Học không giám sát được áp dụng vào bài toán phân cụm.

**b) Nội dung:** HS tìm hiểu SGK, thảo luận tìm hiểu nội dung kiến thức theo yêu cầu của GV.

**c) Sản phẩm:** HS hoàn thành tìm hiểu kiến thức.

**d) Tổ chức thực hiện**

|  |  |
| --- | --- |
| **Sản phẩm dự kiến** | **Hoạt động của GV và HS** |
|

|  |  |  |  |
| --- | --- | --- | --- |
| **Phương pháp** | **Dữ liệu** | **Mục tiêu** | **Đặc điểm chính** |
| Học có giám sát | Dữ liệu có nhãn | Dự đoán kết quả dựa trên dữ liệu đầu vào có nhãn | Huấn luyện mô hình bằng dữ liệu đã được gắn nhãn |
| Học không giám sát | Dữ liệu không có nhãn | Phân loại hoặc phân nhóm dữ liệu mà không dựa trên nhãn trước | Khám phá cấu trúc ẩn trong dữ liệu |

 | **Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Đặt câu hỏi*? 1.* Các em hãy cho biết trong Học máy, tập dữ liệu được xây dựng như thế nào? Với mỗi loại dữ liệu, phương pháp là gì? Mục tiêu và đặc điểm của từng phương pháp?*? 2.* Em hãy tìm hiểu hai hình 25.3 và 25.4 trong SHS và giải thích cho hai hình đó?**HS:** Lắng nghe để thực hiện yêu cầu của GV.**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm ‌vụ:****HS:‌** Quan sát và lần lượt thực hiện các yêu cầu. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:‌****GV:** Điều khiển hoạt động của của các HS, cho HS trả lời và nhận xét câu trả lời.**‌HS:‌** ‌ ‌Lần lượt trả lời các yêu cầu.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌ định:‌ ‌‌**GV‌ nhận xét câu trả lời của HS và chính xác lại đáp án các câu hỏi. |

**Hoạt động 3: Tìm hiểu vai trò của Học máy trong một số công việc cụ thể**

**a) Mục tiêu:** Giúp HS biết được vai trò quan trọng của Học máy trong một số công việc cụ thể.

**b) Nội dung:** HS tìm hiểu SGK, thảo luận tìm hiểu nội dung kiến thức theo yêu cầu của GV.

**c) Sản phẩm:** HS hoàn thành tìm hiểu kiến thức.

**d) Tổ chức thực hiện**

|  |  |
| --- | --- |
| **Sản phẩm dự kiến** | **Hoạt động của GV và HS** |
| **3. MỘT SỐ ỨNG DỤNG CỦA HỌC MÁY** **Vai trò của Học máy** Nhờ khả năng tự học từ dữ liệu và tự hoàn thiện Học máy ngày càng có vai trò không thể thiếu trong các ứng dụng mà dữ liệu có quy mô và chủng loại đa dạng, không ngừng thay đổi theo thời gian.Một số ứng dụng thực tế của Học máy- Lọc thư rác: Giúp xây dựng mô hình có khả năng phân loại thư điện tử là thư rác hoặc thư thường, giúp cải thiện hiệu suất lọc thư theo thời gian bằng cách học hỏi từ dữ liệu và cập nhật mô hình.- Chẩn đoán bệnh: Xây dựng mô hình chẩn đoán bệnh, dự báo tình trạng sức khoẻ, giúp bác sĩ đưa ra phương án điều trị phù hợp và nhanh chóng cho bệnh nhân.- Phân tích thị trường: Xác định xu hướng, dự báo biến động giá cả, giúp người đầu tư và nhà kinh doanh hiểu rõ hơn về thị trường, tăng khả năng đưa ra quyết định đầu tư dựa trên thông tin và các phân tích kĩ thuật.- Nhận dạng tiếng nói: Xây dựng các mô hình âm thanh giúp máy tính có thể học, nhận dạng và tạo ra biểu diễn số hoá của âm thanh. - Nhận dạng chữ viết: Xây dựng mô hình hình học, phát triển mô hình học sâu cho phép học và trích xuất các đặc trưng phức tạp từ hình ảnh chữ viết tay, giúp cải thiện đáng kể khả năng nhận dạng chữ viết tay.Dịch tự động: Xây dựng mô hình dịch tự động.**Ghi nhớ:**Hai phương pháp học máy cơ bản là học có giám sát và học không giám sát, tuỳ theo tập dữ liệu cung cấp cho mô hình học máy là dữ liệu có nhãn hay không có nhãn. Học máy giúp xử lí lượng lớn dữ liệu một cách nhanh chóng và hiệu quả, bao gồm cả các dữ liệu không ngừng thay đổi theo thời gian, trợ giúp các quá trình ra quyết định cũng như tự động hoá các nhiệm vụ phức tạp.**Câu hỏi củng cố kiến thức**Vai trò quan trọng của Học máy trong các lĩnh vực khác nhau được thể hiện như thế nào?Trả lời:Đã tóm tắt ở trên bài học. | **Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Chia lớp thành 4 nhóm thảo luận về vai trò của Học máy qua những ví dụ cụ thể, yêu cầu mỗi nhóm sẽ tìm 2 ví dụ có ứng dụng Học máy (tham khảo trong sách giáo khoa và các nguồn khác) **HS:** Các nhóm HS lắng nghe để thực hiện yêu cầu của GV.**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm‌ vụ:****HS:‌** Các nhóm thảo luận về các ứng dụng của Học máy và sẵn sàng cho câu trả lời chung. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:‌****GV:** Điều khiển hoạt động của của các HS, cho đại diện các nhóm HS trả lời và nhận xét câu trả lời.**‌HS:‌** ‌ Các nhóm ‌lần lượt trả lời các yêu cầu.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌ định:‌‌‌**GV‌ nhận xét câu trả lời của các nhóm HS và chính xác lại đáp án các câu hỏi. |

**C. LUYỆN TẬP** - **VẬN DỤNG**

**a) Mục tiêu:** Giúp HS được ôn tập lại những khái niệm trong Học máy và ứng dụng của Học máy và tầm quan trọng của Học máy.

**b) Nội dung:** GV giao nhiệm vụ cho HS, HS tìm hiểu xem lại và trả lời câu hỏi.

**c) Sản phẩm:** HS hoàn thiện hiểu biết về Học máy.

**d) Tổ chức thực hiện**

**Bước 1: Chuyển giao nhiệm vụ học tập**

**GV** đặt câu hỏi cho **HS**,

*Tại sao có thể nói Học máy có vai trò không thể thiếu trong các ứng dụng mà dữ liệu không ngừng thay đổi theo thời gian? Hãy chỉ ra một vài minh họa cụ thể?*

**Bước 2: HS thực hiện nhiệm vụ học tập**

**HS** tiếp nhận, thực hiện nhiệm vụ, thảo luận, đưa ra câu trả lời.

**GV** quan sát quá trình HS thảo luận, hỗ trợ khi **HS** cần.

**Bước 3: Báo cáo kết quả hoạt động và thảo luận**

**GV** cho **HS** trả lời

**Bước 4: Đánh giá kết quả, thực hiện nhiệm vụ học tập**

GV chính xác hoá lại các nội dung trả lời của HS.

Gợi ý câu trả lời:

Học máy có vai trò không thể thiếu trong các ứng dụng mà dữ liệu không ngừng thay đổi theo thời gian vì những lí do sau:

Học máy có thể tự động cập nhật mô hình dựa trên dữ liệu mới. Điều này rất quan trọng trong các ứng dụng mà dữ liệu thay đổi nhanh chóng, chẳng hạn như thị trường chứng khoán, chẩn đoán bệnh, hoặc các phòng chống các mối đe dọa mạng.

Học máy có thể phát hiện các mẫu và xu hướng mới trong dữ liệu. Điều này giúp các ứng dụng có thể thích ứng với những thay đổi của dữ liệu và cung cấp kết quả chính xác hơn.

🡪 Nói một cách ngắn gọn, Học máy đóng vai trò quan trọng trong các ứng dụng với dữ liệu không ngừng thay đổi do khả năng tự thích nghi và cập nhật liên tục của nó. Học máy có thể tự động hoá quá trình phân tích dữ liệu, nhận dạng xu hướng và mẫu dữ liệu mới mà không cần sự can thiệp thủ công liên tục. Ví dụ:

* Dự báo thời tiết: Mô hình học máy xử lí lượng lớn dữ liệu khí tượng và môi trường, tự cập nhật và điều chỉnh dự báo dựa trên dữ liệu mới nhất.
* Quản lí giao thông thông minh: Hệ thống học từ dữ liệu giao thông thực tế, nhận dạng mẫu và dự báo tắc nghẽn, giúp điều chỉnh đèn giao thông hoặc đưa ra khuyến nghị cho người lái. - Phát hiện gian lận tài chính: Học máy phân tích giao dịch liên tục, nhận biết hành vi bất thường và cập nhật mô hình nhận dạng để phòng chống gian lận hiệu quả hơn.
* Dự đoán giá cổ phiếu: Thị trường chứng khoán là một môi trường luôn thay đổi, với những biến động giá cả diễn ra liên tục. Học máy có thể được sử dụng để dự đoán giá cổ phiếu trong tương lai, giúp các nhà đầu tư đưa ra quyết định đầu tư chính xác hơn.
* Chẩn đoán bệnh: Các bệnh tật luôn có những biểu hiện mới, khiến cho việc chẩn đoán bệnh trở nên khó khăn hơn. Học máy có thể được sử dụng để phân tích dữ liệu y tế, phát hiện các mẫu và xu hướng mới, từ đó giúp các bác sĩ chẩn đoán bệnh chính xác hơn.
* Phát hiện mối đe dọa mạng: Các mối đe dọa mạng luôn xuất hiện và phát triển, đòi hỏi các giải pháp bảo mật phải được cập nhật thường xuyên. Học máy có thể được sử dụng để phân tích dữ liệu mạng, phát hiện các mối đe dọa mới, từ đó giúp các tổ chức bảo vệ hệ thống của mình khỏi các cuộc tấn công mạng…

|  |  |
| --- | --- |
| **Trường: ………………………****Tổ: ……………………………** | Họ và tên giáo viên:  |

## CHỦ ĐỀ F: GIẢI QUYẾT VẤN ĐỀ VỚI SỰ TRỢ GIÚP CỦA MÁY TÍNH

## FCS1. GIỚI THIỆU HỌC MÁY VÀ KHOA HỌC DỮ LIỆU

**BÀI 2. GIỚI THIỆU VỀ KHOA HỌC DỮ LIỆU**

## Môn Tin Học; Lớp: 12

Thời gian thực hiện: 2 tiết

**I. MỤC TIÊU**

**1. Kiến thức**

Bài học cung cấp những kiến thức:

- Khái niệm Khoa học dữ liệu.

- Các giai đoạn của một dự án Khoa học dữ liệu.

- Một số thành tựu của Khoa học dữ liệu.

**2. Năng lực**

Bài học góp phần củng cố và phát triển cho học sinh những năng lực với biểu hiện cụ thể như sau:

***Năng lực chung:***

*- Tự chủ và tự học:*HS biết tự tìm kiếm, chuẩn bị và lựa chọn tài liệu, phương tiện học tập trước giờ học, quá trình tự giác tham gia các và thực hiện các hoạt động học tập cá nhân trong giờ học ở trên lớp,…

- *Giao tiếp và hợp tác:*Khả năng phân công và phối hợp thực hiện nhiệm vụ học tập.

- *Giải quyết vấn đề và sáng tạo:*HS đưa ra các phương án trả lời cho câu hỏi, bài tập xử lý tình huống, vận dụng kiến thức, kĩ năng của bài học để giải quyết vấn đề thường gặp.

***Năng lực Tin học:***

- *Nla (Sử dụng và quản lí các phương tiện CNTT và truyền thông) và NLc (Giải quyết vấn đề với sự hỗ trợ của CNTT và truyền thông):*

+ Nêu được sơ lược về khái niệm, mục tiêu của Khoa học dữ liệu.

+ Nêu được một số thành tựu của Khoa học dữ liệu và ví dụ minh họa.

**3. Phẩm chất**

- *Chăm chỉ:* Chủ động tìm hiểu và cập nhật những kiến thức mới, không ngừng học hỏi.

- *Trách nhiệm:* Hoàn thành các nhiệm vụ học tập, tích cực trong hoạt động nhóm.

**II. THIẾT BỊ DẠY HỌC VÀ HỌC LIỆU**

- GV: SGK, SBT Tin học 12 – Định hướng Khoa học máy tính – Cánh diều, bài trình chiếu (Slide), máy chiếu.

- HS: SGK, SBT Tin học 12 – Định hướng Khoa học máy tính – Cánh diều, vở ghi.

**III. TIẾN TRÌNH DẠY HỌC**

**A. HOẠT ĐỘNG KHỞI ĐỘNG**

**a) Mục tiêu:** Tạo hứng thú học tập cho HS.

**b) Nội dung:** HS nghiên cứu nội dung hoạt động **Khởi động** SGK trang 133 và đưa ra câu trả lời.

**c) Sản phẩm:** Từ yêu cầu, HS suy nghĩ trả lời câu hỏi **Khởi động** SGK trang 133.

**d) Tổ chức thực hiện:**

**Bước 1:** **GV chuyển giao nhiệm vụ:**

- GV yêu cầu HS hoạt động cá nhân, suy nghĩ trả lời câu hỏi **Khởi động**tr.133SGK:

*Có ý kiến cho rằng: Dữ liệu là tài sản quan trọng của tổ chức, doanh nghiệp. Theo em, nói như vậy là vì nguyên nhân nào sau đây:*

*1) Chi phí cao để thu thập, lưu trữ, bảo đảm an toàn dữ liệu.*

*2) Dữ liệu được sử dụng để tăng hiệu quả làm việc, tăng sức cạnh tranh của doanh nghiệp, tăng lợi ích kinh doanh.*

**Bước 2: HS thực hiện nhiệm vụ học tập:**

- HS vận dụng hiểu biết để trả lời câu hỏi.

- GV quan sát và hỗ trợ HS khi cần thiết.

**Bước 3:** **Báo cáo kết quả hoạt động, thảo luận:**

- GV mời một số HS xung phong trả lời câu hỏi **Khởi động**tr.133SGK.

***Gợi ý trả lời:***  *Dữ liệu là tài sản quan trọng của tổ chức, doanh nghiệp vì nó giúp tăng hiệu quả làm việc, tăng sức cạnh tranh và mang lại lợi ích kinh doanh cho doanh nghiệp. Doanh nghiệp cần đầu tư vào việc thu thập, lưu trữ và bảo đảm an toàn dữ liệu để khai thác tối đa tiềm năng của nó.*

- HS khác lắng nghe, nhận xét và bổ sung.

**Bước 4: Đánh giá kết quả thực hiện:**

GV đánh giá kết quả của HS, dẫn dắt HS vào bài học mới: *Dữ liệu là tài sản quan trọng của tổ chức, doanh nghiệp. Nhiều doanh nghiệp, bất kể quy mô, đều cần một chiến lược khoa học dữ liệu hiệu quả để thúc đẩy tăng trưởng và duy trì lợi thế cạnh tranh. Vậy để giúp các em hiểu rõ về khái niệm, mục tiêu và những thành tựu của Khoa học dữ liệu, chúng ta sẽ cùng nhau đến với* ***Bài 2: Giới thiệu về Khoa học dữ liệu.***

**B. HÌNH THÀNH KIẾN THỨC MỚI**

**Hoạt động 1. Tìm hiểu về Khoa học dữ liệu**

a) Mục tiêu: Xác định hiểu biết ban đầu về Khoa học dữ liệu của HS, đặc biệt là về phạm vi và các nội dung cụ thể của lĩnh vực này.

b) Nội dung: HS tìm hiểu SGK, thảo luận tìm hiểu nội dung kiến thức theo yêu cầu của GV.

c) Sản phẩm: HS hoàn thành tìm hiểu kiến thức.

d) Tổ chức thực hiện

|  |  |
| --- | --- |
| **Sản phẩm dự kiến** | **Hoạt động của GV và HS** |
| **1. KHOA HỌC DỮ LIỆU****Đáp án trắc nghiệm**- Khoa học dữ liệu tập trung vào việc nghiên cứu, phân tích và khai thác thông tin từ dữ liệu.- Khoa học dữ liệu thường bao gồm các công việc như thu thập, xử lí, phân tích và trực quan hoá dữ liệu để rút ra được các thông tin/tri thức hữu ích.- Nó liên quan đến việc áp dụng các phương pháp và kĩ thuật như Học máy, thống kê và phân tích dữ liệu để hiểu rõ hơn về dữ liệu và tìm kiếm các xu hướng hoặc mẫu hình của dữ liệu.- Hoạt động, A, B, D đều là những nội dung quan trọng của Khoa học dữ liệu. Phương án C không phải là một nội dung trực tiếp của Khoa học dữ liệu.**a) Khái niệm về Khoa học dữ liệu** *Khoa học dữ liệu là một lĩnh vực liênngành, sử dụng các phương pháp khoa học, quy trình, thuật toán 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.*Khoa học dữ liệu sử dụng các phương pháp và công cụ của: khoa học máy tính, toán học và thống kê kết hợp với tri thức chuyên ngành để giúp tổ chức, cá nhân hiểu rõ hơn về dữ liệu mình sởhữu và tận dụng tri thức này để đưa ra nhữngquyết định phù hợp.+) Khoa học máy tính cung cấp các ***công cụ*** và kĩ thuật để xử lí.+) Toán học và thống kê cung cấp cơ sở cho các ***phương pháp*** phân tích và khai phá dữ liệu.+) Tri thức chuyên ngành là ***tri thức của từng lĩnh vực***, ví dụ kinh doanh, y tế, khoa học xã hội,... có vai trò quan trọng để hiểu ngữ cảnh và ý nghĩa của dữ liệu. Có thể nhìn thấy mối quan hệ qua hình vẽ:**Mục tiêu của Khoa học dữ liệu**- Mục tiêu chính của Khoa học dữ liệu là phân tích và khai phá dữ liệu để có được tri thức, sau đó vận dụng tri thức đó để giải quyết vấn đề và đưa ra các quyết định phù hợp.- Các mục tiêu cụ thể bao gồm: * Phân tích và trực quan hoá dữ liệu: Biểu diễn dữ liệu một cách trực quan, dễ hiểu thông qua các sơ đồ, biểu đồ hay hình ảnh.
* Xây dựng mô hình dự đoán, dự báo: Sử dụng dữ liệu để xây dựng mô hình có khả năng dự đoán sự kiện trong tương lai:

Tối ưu hoá quyết định: Cải thiện quyết định dựa trên dữ liệu, bao gồm việc sử dụng các thuật toán tối ưu hoá để đưa ra quyết định tốt nhất dựa trên các ràng buộc và mục tiêu.* Phát hiện tri thức: Tìm ra các mối quan hệ ẩn chứa trong dữ liệu, xác định nguyên nhân và kết quả, tạo ra tri thức mới từ dữ liệu.

**b) Các giai đoạn của một dự án khoa học dữ liệu:**1. Xác định vấn đề:
2. Thu nhập dữ liệu
3. Chuẩn bị dữ liệu
4. Phân tích và khai phá dữ liệu
5. Đánh giá và giải thích
6. Ra quyết định và triển khai
 | **Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Cho HS xem clip ngắn nói về khoa học dữ liệu sau đó yêu câu HS chọn đáp án đúng khi nói về khoa học dữ liệu:[https://www.youtube.com/](https://www.youtube.com/watch?v=ppVZMy2G4AU)[watch?v=ppVZMy2G4AU](https://www.youtube.com/watch?v=ppVZMy2G4AU)?*1.* Có thể hiểu đơn giản Khoa học dữ liệu là lĩnh vực khoa học nghiên cứu về dữ liệu. Như vậy, đối tượng nghiên cứu của Khoa học dữ liệu chính là dữ liệu. Theo em, Khoa học dữ liệu không bao gồm công việc nào sau đây? Giải thích ngắn gọn.A. Nghiên cứu phát triển các phương pháp thu thập và quản lí dữ liệu.B. Khai phá các thông tin, trí thức từ dữ liệu thu để nâng cao hiệu quả kinh doanh, quản lí.C. Kinh doanh, phân phối dữ liệu thu thập được cho các cá nhân, tổ chức quan tâm.D. Phát triển và áp dụng các phương pháp và kĩ thuật để nhận biết các mẫu hình, các quan hệ và xu hướng có trong dữ liệu.*?2.* Khoa học dữ liệu là gì? Khoa học dữ liệu sử dụng các phương pháp và công cụ của các lĩnh vực nào?**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm‌ vụ:****HS:‌** Tìm câu trả lời từ hiểu biết hoặc từ SGK. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:****GV:** Điều khiển hoạt động của của các HS, cho HS phát biểu, cho HS nhận xét nhau.Cho các nhóm trả lời và nhận xét ở câu hỏi số 3.**‌HS:‌** ‌ ‌Phát biểu trả lời câu hỏi.Các‌‌ HS ‌nhận‌ ‌xét nhau.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌định:‌ ‌‌**GV chính xác lại các câu trả lời và chính xác lại đáp án.**Mục tiêu của Khoa học dữ liệu****Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Chia lớp thành 4 nhóm học tập. Dẫn dắt vào vấn đề thảo luận. GV cho mỗi nhóm chọn một phiếu học tập phiếu để thảo luận.*.* Mục tiêu chính của Khoa học dữ liệu là gì? Các mục tiêu cụ thể của Khoa học dữ liệu là gì? Mối quan hệ giữa các mục tiêu của Khoa học dữ liệu là gì? **Phiếu số 1**

|  |
| --- |
|  Các em hãy cho biết vai trò quan trọng của máy tính đối với sự phát triển của Khoa học dữ liệu từ góc độ sau bằng cách trả lời câu hỏi:*Về phân tích và khai phá dữ liệu*:- Làm thế nào máy tính giúp trong việc phân tích và khai phá dữ liệu phức tạp?- Em có thể đưa ra một ví dụ cụ thể về cách sử dụng máy tính để phân tích và dự đoán từ dữ liệu? |

**Phiếu số 2**

|  |
| --- |
|  Các em hãy cho biết vai trò quan trọng của máy tính đối với sự phát triển của Khoa học dữ liệu từ góc độ sau bằng cách trả lời câu hỏi:*Về trực quan hoá dữ liệu*:- Tại sao trực quan hoá dữ liệu lại quan trọng và làm thế nào công nghệ máy tính hỗ trợ trong việc này?- Hãy mô tả một công cụ trực quan hoá dữ liệu mà em đã sử dụng. |

**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm‌ vụ:****HS:‌** Tìm câu trả lời từ hiểu biết hoặc từ SGK. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:****GV:** Điều khiển hoạt động của của các HS, cho HS phát biểu, cho HS nhận xét nhau.Cho các nhóm trả lời và nhận xét ở câu hỏi số 3.**‌HS:‌** ‌ ‌Phát biểu trả lời câu hỏi.Các‌‌ HS ‌nhận‌ ‌xét nhau.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌định:‌ ‌‌**GV chính xác lại các câu trả lời và chính xác lại đáp án.**b) Các giai đoạn của một dự án khoa học dữ liệu:****Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ‌****GV:** Chia lớp thành 4 nhóm học tập. Dẫn dắt vào vấn đề thảo luận. GV cho mỗi nhóm chọn một phiếu học tập phiếu để thảo luận.**Phiếu số 1:**

|  |
| --- |
|  Các em hãy cho biết vai trò quan trọng của máy tính đối với sự phát triển của Khoa học dữ liệu từ góc độ sau bằng cách trả lời câu hỏi:*Về phân tích và khai phá dữ liệu*:- Làm thế nào máy tính giúp trong việc phân tích và khai phá dữ liệu phức tạp?- Em có thể đưa ra một ví dụ cụ thể về cách sử dụng máy tính để phân tích và dự đoán từ dữ liệu? |

**Phiếu số 2:**

|  |
| --- |
|  Các em hãy cho biết vai trò quan trọng của máy tính đối với sự phát triển của Khoa học dữ liệu từ góc độ sau bằng cách trả lời câu hỏi:*Về tự động hoá quy trình Khoa học dữ liệu*:- Hãy nêu một ví dụ về cách máy tính giúp tự động hoá quy trình trong Khoa học dữ liệu.- Tự động hoá có thể gặp những thách thức gì trong Khoa học dữ liệu? |

**HS:** Lắng nghe để thực hiện yêu cầu của GV.**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm‌ vụ:****HS:‌** Tìm câu trả lời từ hiểu biết hoặc từ SGK. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:****GV:** Điều khiển hoạt động của của các HS, cho HS phát biểu, cho HS nhận xét nhau.Cho các nhóm trả lời và nhận xét ở câu hỏi số 3.**‌HS:‌** ‌ ‌Phát biểu trả lời câu hỏi.Các‌‌ HS ‌nhận‌ ‌xét nhau.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌định:‌ ‌‌**GV chính xác lại các câu trả lời và chính xác lại đáp án.B1: GV giao nhiệm vụ: Hướng dẫn tìm hiểu Quy trình của 1 dự án KHDLB2: HS thực hiện:B3: Đại diện nhóm trình bày, nhóm khác nhận xét, bổ sung.Kết luận, nhận định:**B4:** GV nhận xét kết quả, tổng hợp, chốt lại vấn đề ở hoạt động Làm.Sau khi thực hiện xong hoạt động Làm, GV chốt lại kiến thức cho HS ở hoạt động Ghi nhớ. |

**Hoạt động 2. Làm quen với dữ liệu lớn trong thực tế**

a) Mục tiêu:

- Giúp HS hiểu rõ hơn về khái niệm dữ liệu lớn (Big Data) không chỉ dừng lại ở kích thước lớn của dữ liệu mà còn mở rộng đến tính đa dạng và sự cập nhật liên tục, nhấn mạnh sự đa dạng và phức tạp của dữ liệu trong thực tế, bao gồm cả dữ liệu cấu trúc, phi cấu trúc và bán cấu trúc từ nhiều nguồn khác nhau.

- Thông qua việc đưa ra ví dụ cụ thể, câu hỏi giúp kết nối kiến thức lí thuyết với các ứng dụng thực tế, giúp HS dễ dàng hình dung và hiểu rõ cách thức sử dụng dữ liệu lớn trong cuộc sống và công việc.

b) Nội dung: HS tìm hiểu SGK, thảo luận tìm hiểu nội dung kiến thức theo yêu cầu của GV.

c) Sản phẩm: HS hoàn thành tìm hiểu kiến thức.

d) Tổ chức thực hiện

|  |  |
| --- | --- |
| **Sản phẩm dự kiến** | **Hoạt động của GV và HS** |
| **2. MỘT SỐ THÀNH TỰU CỦA KHOA HỌC DỮ LIỆU**Khoa học dữ liệu có nhiều ứng dụng trong kinh tế - xã hội: Tài chính ngân hàng, chăm sóc sức khoẻ, trong kinh doanh, trong sản xuất…Khoa học dữ liệu đạt được 1 số thành tựu như sau:**a) Tính ưu việt của việc sử dụng máy tính và các thuật toán hiệu quả để xử lí dữ liệu lớn với Dự án Hệ gene người (Human Genome Project - HGP)**Dự án Hệ gene người (Human Genome Project - HGP) là dự án tìm cách khám phá những bí mật về cấu trúc di truyền của con người. HPG tạo ra một lượng dữ liệu khổng lồ và đòi hỏi nguồn lực tính toán hết sức to lớn 🡪 cần hệ thống máy tính mạnh, các thuật toán có độ chính xác cao và tốc độ nhanh để thực hiện.2) Độ phức tạp của dữ liệu- Kích thước dữ liệu: Chuỗi kí tự được nói ở trên của hệ gene người có độ dài khoảng 107,8 tỉ km. Việc giải trình tự toàn bộ hệ gene người tạo ra hàng trăm gigabyte dữ liệu thô.- Lưu trữ dữ liệu: Việc lưu trữ dữ liệu từ HGP là một thách thức đáng kể. Tổng dung lượng lưu trữ cho dữ liệu HGP được ước tính chiếm khoảng một trăm nghìn gigabyte.- Sức mạnh xử lí: Phân tích dữ liệu HPG đòi hỏi nguồn lực tính toán mạnh mẽ. Vào thời kì đỉnh cao, HGP dựa vào mạng lưới siêu máy tính trên khắp thế giới. Sức mạnh tính toán được sử dụng trong Dự án tương đương với hàng nghìn máy tính xách tay hiện đại hoạt động đồng thời.**b) Các dự án nghiên cứu và khám phá không gian vĩ trụ****c) Hệ thống giám sát đánh bắt cá toàn cầu****d) Các mô hình ngôn ngữ lớn****e) Mô hình phát hiện gian lận của American Express****Kết luận** Mỗi bài toán đều yêu cầu sức mạnh tính toán lớn và thuật toán phức tạp để xử lí và phân tích dữ liệu lớn một cách hiệu quả. Sự phát triển của máy tính hiện đại và thuật toán tiên tiến là chìa khoá để giải quyết các thách thức liên quan đến dữ liệu lớn trong các lĩnh vực. | **Bước‌ ‌1:‌ ‌Chuyển‌ ‌giao‌ ‌nhiệm‌ ‌vụ**Gọi 4 nhóm lên trình bày nội dung tìm hiểu trước ở nhà. (Giao việc cho 4 nhóm) với 3 cậu hỏi tìm hiểu:1) Dự án nhóm tìm hiểu là dự án gì?, mục đích của dự án là làm gì? Yêu cầu cần có để thực hiện dữ án là gì?2) Cho biết kích thước bài toán; quy mô dữ liệu; độ phức tạp tính toán…3) Kết quả, tính ưu việt của máy tính và các thuật toán hiệu quả được thể hiện như thế nào?**Nhóm 1: Dự án bộ Gen người HGP****Nhóm 2: Các dự án nghiên cứu và khám phá không gian vĩ trụ****Nhóm 3: Hệ thống giám sát đánh bắt cá toàn cầu****Nhóm 4: Các mô hình ngôn ngữ lớn****HS:** Lắng nghe để thực hiện yêu cầu của GV.**Bước‌ ‌2:‌ ‌Thực‌ ‌hiện‌ ‌nhiệm ‌vụ:****HS:‌** Quan sát và lần lượt thực hiện các yêu cầu. **‌GV:**‌ ‌Quan‌ ‌sát‌ và trợ giúp HS.**Bước‌ ‌3:‌ ‌Báo‌ ‌cáo,‌ ‌thảo‌ ‌luận:‌****GV:** Điều khiển hoạt động của của các HS, cho HS trả lời và nhận xét câu trả lời.GV cho các nhóm báo cáo cho câu số 3 và nhận xét.**‌HS:‌** ‌‌Lần lượt trả lời các yêu cầu.**Bước‌ ‌4:‌ ‌Kết‌ ‌luận,‌ ‌nhận‌ ‌ định:‌ ‌‌**GV‌ nhận xét câu trả lời của HS và chính xác lại đáp án các câu hỏi. |

**C. LUYỆN TẬP – VẬN DỤNG**

a) Mục tiêu: Giúp HS được ôn tập lại những kiến thức về Khoa học dữ liệu, Big Data.

b) Nội dung: GV giao nhiệm vụ cho HS, HS tìm hiểu xem lại và trả lời câu hỏi.

c) Sản phẩm: HS hoàn thiện hiểu biết về Học máy.

d) Tổ chức thực hiện

**Bước 1: Chuyển giao nhiệm vụ học tập**

**GV** đặt câu hỏi cho **HS**,

**1.** Nêu ngắn gọn vai trò của máy tính trong sự phát triển của Khoa học dữ liệu.

**2.** Trong trường hợp xấu nhất, để sắp xếp các đoạn nucleotide ngắn thành hệ gene người hoàn chỉnh, ước tính cần bao nhiêu phép thử?

**Bước 2: HS thực hiện nhiệm vụ học tập**

HS tiếp nhận, thực hiện nhiệm vụ, thảo luận, đưa ra câu trả lời.

GV quan sát quá trình HS thảo luận, hỗ trợ khi HS cần.

**Bước 3: Báo cáo kết quả hoạt động và thảo luận**

GV cho HS trả lời

**Bước 4: Đánh giá kết quả, thực hiện nhiệm vụ học tập**

GV chính xác hoá lại các nội dung trả lời của HS.

Gợi ý câu trả lời

**1.** Câu trả lời có thể trình bày dưới dạng tóm tắt các ý chính trong mục 1 của bài học (xử lí và lưu trữ dữ liệu; phân tích và khai phá dữ liệu; trực quan hoá dữ liệu; tự động hoá quy trình trong Khoa học dữ liệu; xử lí song song và điện toán đám mây; hợp tác và truyền thông).

**2.** Hệ gene người có thể được xem như là một chuỗi có độ dài khoảng 3 tỉ các kí tự A, C, G, T. Như vậy, trường hợp xấu nhất sẽ phải thực hiện khoảng 4^(3 tỉ) phép thử! Hãy nhớ lại bài toán xếp hạt thóc vào 64 ô của bàn cờ, ô đầu tiên xếp 1 hạt thóc, mỗi ô tiếp sau xếp số hạt thóc gấp đôi ô trước. Số lượng hạt thóc cần để xếp vào 64 ô bàn cờ tướng là 2^64 – 1, ước tính tương đương 536 tỷ tấn gạo! Tổ chức Nông Lương Liên Hiệp Quốc (FAO) ước tính tổng sản lượng ngũ cốc (gạo là 1 trong 5 loại ngũ cốc) thế giới năm 2018 khoảng 2,6 tỷ tấn. Như vậy số gạo nêu trên bằng số lương thực cả thế giới sản xuất trong hơn 2 thế kỉ! Từ đây có thể hình dung con số 4^(3 tỉ) lớn như thế nào!