**Giới thiệu về trí tuệ nhân tạo (Artificial Intelligence - AI)**

AI là một lĩnh vực đang phát triển nhanh chóng với mục tiêu giúp máy tính có thể thực hiện các tác vụ yêu cầu trí tuệ con người. Quá trình phát triển của AI hiện được chia thành 3 giai đoạn:

1. **Unit-AI**: AI đơn lẻ, chỉ có thể thực hiện những nhiệm vụ cụ thể.
2. **AGI (Artificial General Intelligence)**: AI có khả năng tương tự như con người, hiểu và thực hiện nhiều nhiệm vụ khác nhau.
3. **Super AI**: AI vượt xa trí tuệ con người, có khả năng tự học và cải tiến.

**Các bài toán phổ biến trong AI**

* **Phân lớp (Classification)**: Dự đoán nhãn hoặc phân loại dữ liệu.
* **Hồi quy (Regression)**: Dự đoán các giá trị liên tục.
* **Gom cụm (Clustering)**: Nhóm các dữ liệu tương tự lại với nhau.
* **Luật kết hợp (Association Rule Mining)**: Tìm mối quan hệ giữa các đối tượng trong dữ liệu.

**Quy trình giải quyết bài toán AI**

1. **Xử lý dữ liệu**: Làm sạch và chuẩn hóa dữ liệu.
2. **Huấn luyện mô hình**: Sử dụng dữ liệu để huấn luyện mô hình AI.
3. **Đánh giá mô hình**: Kiểm tra hiệu quả của mô hình với dữ liệu kiểm thử.
4. **Triển khai mô hình**: Đưa mô hình vào sử dụng thực tế.
5. **Nhận phản hồi**: Thu thập phản hồi và cải thiện mô hình.

**Một số bài toán phân lớp phổ biến và ứng dụng**

* Phân loại email thành **spam** hoặc **không spam**.
* Phân loại bệnh nhân dựa trên dữ liệu y tế.
* Nhận diện các **loài hoa** hoặc **động vật** dựa trên hình ảnh.
* Dự đoán khách hàng có **rủi ro tín dụng** cao hay không.
* Nhận diện các **giao dịch ảo** qua thẻ tín dụng.
* Phân biệt **tin thật** và **tin giả**.

**Các kỹ thuật phân lớp phổ biến**

**1. K-Nearest Neighbors (KNN)**

* **Ý tưởng**: Dự đoán nhãn cho một điểm mới dựa trên **K điểm gần nhất**.
* **Ưu điểm**: Đơn giản, dễ triển khai.
* **Nhược điểm**: Tốn thời gian khi dữ liệu lớn, nhạy cảm với nhiễu.

**2. Decision Tree (Cây quyết định)**

* **Ý tưởng**: Xây dựng mô hình phân loại dạng cây, với mỗi nút là một điều kiện.
* **Ưu điểm**: Dễ hiểu và giải thích kết quả.
* **Nhược điểm**: Có thể dẫn đến **overfitting** nếu cây quá phức tạp.

**3. Bayesian Classification (Phân loại Bayes)**

* **Ý tưởng**: Dựa vào **Định lý Bayes** để tính xác suất xảy ra của các nhãn dựa trên dữ liệu đầu vào.
* **Ưu điểm**: Hiệu quả với dữ liệu ít và nhiễu.
* **Nhược điểm**: Giả định các đặc trưng phải độc lập, điều này không phải lúc nào cũng đúng.

**4. Multiple Linear Regression (Hồi quy tuyến tính bội)**

* **Ý tưởng**: Dự đoán giá trị đầu ra dựa trên sự kết hợp tuyến tính của nhiều biến độc lập.
* **Ưu điểm**: Dễ hiểu và tính toán nhanh.
* **Nhược điểm**: Chỉ hiệu quả khi quan hệ giữa biến phụ thuộc và các biến độc lập là tuyến tính.

**Lựa chọn thuật toán học máy phù hợp**

* **Học có giám sát (Supervised Learning)**: Khi có cả input và output trong tập dữ liệu (ví dụ: Phân loại bệnh nhân).
* **Học không có giám sát (Unsupervised Learning)**: Khi chỉ có input mà không có output (ví dụ: Gom cụm khách hàng).