**Tổng hợp các thuật toán trong học máy**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Thuật toán** | **Loại** | **Nguyên lí hoạt động** | **Ưu điểm** | **Nhược điểm** |
| ID3 | Phân loại | Sử dụng Entropy và Information Gain để xây dựng cây quyết định. | - Dễ hiểu, dễ triển khai.  - Không yêu cầu chuẩn hóa dữ liệu. | - Nhạy cảm với nhiễu.  - Dễ bị overfitting. |
| Decision Tree | Phân loại,  hồi quy | Xây dựng cây quyết định dựa trên tiêu chí chia giảm thiểu sự hỗn loạn (*Gini* hoặc *Entropy*). | - Dễ hiểu, trực quan.  - Không yêu cầu giả định phân phối dữ liệu. | - Overfitting dễ xảy ra.  - Hiệu suất giảm với dữ liệu lớn. |
| SVM | Phân loại,  hồi quy | Tìm siêu phẳng tối ưu để phân tách hai lớp, tối đa hóa khoảng cách tới *support vectors*. | - Hiệu quả với dữ liệu có nhiều đặc trưng.  - Làm việc tốt với dữ liệu nhỏ, rõ ràng. | - Khó hiệu chỉnh với dữ liệu lớn.  - Không tốt khi dữ liệu nhiễu hoặc không cân bằng. |
| Naïve Bayes | Phân loại | Dựa trên *Định lý Bayes*, giả định các thuộc tính là độc lập. | - Nhanh, hiệu quả với dữ liệu lớn.  - Dễ triển khai, tính toán. | - Giả định độc lập thường không đúng thực tế.  - Độ chính xác thấp nếu dữ liệu phụ thuộc lẫn nhau. |