**DANH SÁCH ĐỀ TÀI MÔN**

**KHAI THÁC DỮ LIỆU VÀ KHAI THÁC TRI THỨC**

**Nhóm sinh viên chọn 1 trong 10 thuật toán sau (hoặc thuật toán khác được sự đồng ý của giảng viên) để tìm hiểu và thực hành.**

1. C4.5
2. K-Means
3. SVM: Support Vector Machine
4. Apriori
5. EM
6. PageRank
7. AdaBoost
8. kNN: k-Nearest Neighbors
9. Naïve Bayes
10. CART: Classification and Regression Trees

**Yêu cầu:**

* Slide:
  + Giới thiệu bài toán/ ứng dụng data mining
  + Tìm hiểu/ phân tích thuật toán
  + So sánh vs thuật toán tương tự/ cập nhật thuật toán mới nhất (Cuối kỳ)
* Demo/ code
  + Tiền xử lý dữ liệu
  + Phân tích dữ liệu
  + Khai thác dữ liệu và Khai phá tri thức
  + Phân tích kết quả đạt được

**Báo cáo:**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Nội dung** | **Điểm** | **Ghi chú** |
| **1** | Bảng phân công công việc   * Ai/ làm gì? * Deadline | 1 | Vai trò Leader |
| **2** | Tìm hiểu/ phân tích vấn đề   * Liệt kê các câu hỏi | 1 | Ứng dụng thực tế/ Mindset |
| **3** | Thu thập dữ liệu   * Databases | 1 | Nguồn: Kaggle, UCI, Scikit-Learning, Google,…  (hoặc tự thu thập) |
| **4** | Tiền xử lý dữ liệu   * Data Warehouse * Data Integration | 2 |  |
| **5** | Phân tích/ tìm hiểu thuật toán phù hợp | 2 | Hiểu thuật toán |
| **6** | Phân tích/ đánh giá kết quả | 1 | Hiểu/ trình bày kết quả |
| **7** | Sử dụng tri thức | 2 | Sự dụng thông tin và tri thức khai thác được từ dữ liệu hổ trợ ra quyết định |
|  | **Tổng điểm** | **10** |  |

**Demo/ Code:**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Nội dung** | **Điểm** | **Ghi chú** |
| 1 | Các thông số mô tả dữ liệu | 1 | Mô tả thống kê |
| 2 | Sử dụng công cụ Tiền xử lý dữ liệu  - Trực quan hóa dữ liệu | 2 | Làm sạch dữ liệu, xử lý missing data |
| 3 | Huấn luyện mô hình | 2 | Thư viện |
| 4 | Trình bày/ trực quan hóa kết quả đạt được   * Lựa chọn chỉ số/ độ đo đánh giá | 2 | Trình bày các chỉ số đo được |
| 5 | Trình bày/ phân tích/ Báo cáo kết quả | 2 | Thảo luận |
| 6 | Hướng ứng dụng kết quả đạt được | 1 | Câu trả lời cho các câu hỏi từ đầu |
|  | **Tổng** | **10** |  |