**THỂ LỆ CHI TIẾT - NHIỆM VỤ 2: XỬ LÝ DỮ LIỆU CẢM BIẾN MÀU**

**1. THỂ LỆ**

Ban tổ chức sẽ cung cấp cho mỗi đội thi hai tập dữ liệu JSON:

1. Tập dữ liệu có nhãn: Chứa các dòng dữ liệu cảm biến đã được gán nhãn màu sắc. Tập dữ liệu này giúp thí sinh phân tích, tìm quy luật giữa các giá trị RGB và màu sắc tương ứng.
2. Tập dữ liệu không có nhãn: Chứa các dòng dữ liệu có nhãn "**UNKNOWN**". Nhiệm vụ của thí sinh là xây dựng thuật toán để dự đoán màu sắc chính xác cho những dòng dữ liệu này.

Mỗi đội cần viết chương trình Python để đọc dữ liệu, trực quan hóa, xây dựng mô hình phân loại, và áp dụng mô hình đó để gán nhãn màu sắc cho dữ liệu chưa biết.

**2. CHI TIẾT VỀ DỮ LIỆU**

**2.1. Dữ liệu đầu vào**

* Dữ liệu được cung cấp trong tệp json.
* Mỗi dòng dữ liệu chứa 4 thông tin chính:
  + objRed: Giá trị cảm biến màu đỏ.
  + objGreen: Giá trị cảm biến màu xanh lá cây.
  + objBlue: Giá trị cảm biến màu xanh dương.
  + objColor: Màu sắc của vật thể (có thể là một trong các màu: RED, GREEN, BLUE, YELLOW, hoặc UNKNOWN).
  + id: Mã định danh của dữ liệu.

Ví dụ về dữ liệu có nhãn:

|  |
| --- |
| **{**  "objColor"**:**"RED"**,**  "objRed"**:**19**,**  "objGreen"**:**1**,**  "objBlue"**:**1**,**  "id"**:**"935950420m19"  **},**  **{**  "objColor"**:**"YELLOW"**,**  "objRed"**:**11**,**  "objGreen"**:**7**,**  "objBlue"**:**2**,**  "id"**:**"838442491m21"  **}** |

Ví dụ về dữ liệu chưa có nhãn (cần phân loại):

|  |
| --- |
| **{**  "objColor"**:**"UNKNOWN"**,**  "objRed"**:**41**,**  "objGreen"**:**27**,**  "objBlue"**:**6**,**  "id"**:**"166521374m40"  **},**  **{**  "objColor"**:**"UNKNOWN"**,**  "objRed"**:**14**,**  "objGreen"**:**33**,**  "objBlue"**:**9**,**  "id"**:**"592106652m41"  **}** |

**2.2. Yêu cầu đối với thí sinh**

1. Đọc và trực quan hóa dữ liệu:
   * Sử dụng Python để đọc tệp JSON vào DataFrame bằng thư viện pandas.
   * Hiển thị dữ liệu dưới dạng bảng.
   * Vẽ biểu đồ thể hiện sự phân bố giá trị RGB để quan sát quy luật màu sắc.
2. Phát triển thuật toán phân loại màu sắc:
   * Xây dựng một thuật toán để xác định màu sắc dựa trên giá trị objRed, objGreen và objBlue.
   * Hiện thực thuật toán bằng Python.
3. Phân loại dữ liệu thi đấu:
   * Chạy thuật toán để gán nhãn màu sắc cho các dòng có objColor = "UNKNOWN".
   * Ghi lại dữ liệu đã phân loại vào một tệp JSON mới để nộp bài.

**3. PLATFORM THỰC HIỆN : Pnote**

Giới thiệu về PNote

PNote là một nền tảng lập trình trực tuyến giúp thí sinh viết, chạy và kiểm thử mã Python một cách thuận tiện. Nó hỗ trợ việc đọc và xử lý dữ liệu, vẽ biểu đồ trực quan, cũng như thực hiện các thuật toán phân loại màu sắc trong cuộc thi.

Các tính năng chính của PNote bao gồm:

* Chạy mã Python trực tuyến mà không cần cài đặt phần mềm.
* Hỗ trợ thư viện khoa học dữ liệu như Pandas, Matplotlib, NumPy.
* Dễ dàng nhập và xuất dữ liệu từ tệp JSON.

Thí sinh sẽ sử dụng PNote để viết chương trình xử lý dữ liệu và phát triển thuật toán phân loại màu sắc.

**4. CÁCH CHẤM ĐIỂM**

* Mỗi mẫu "UNKNOWN" được phân loại chính xác sẽ được cộng 1 điểm.
* Ví dụ, nếu có 1000 mẫu UNKNOWN, đội nào phân loại đúng 200 mẫu sẽ nhận được 200 điểm.
* Nếu hai đội có số điểm bằng nhau, đội nào có thuật toán tối ưu hơn (tốc độ xử lý nhanh hơn, mã nguồn gọn gàng hơn) sẽ được đánh giá cao hơn.

**5. QUY ĐỊNH VÀ HƯỚNG DẪN NỘP BÀI**

* Mỗi đội sẽ nộp 2 tệp:
  1. Tệp JSON kết quả (output.json) chứa dữ liệu đã được phân loại đầy đủ.
  2. Tệp Python (solution.py hoặc solution.ipynb) chứa mã nguồn xử lý dữ liệu và phân loại màu.
* Bài làm sẽ được kiểm tra bằng bộ test độc lập để đảm bảo tính chính xác.