

## **BÀI THỰC HÀNH 2: XÂY DỰNG GUIDELINE VÀ GÁN NHÃN DỮ LIỆU**

### **Hướng dẫn nộp bài:**

Nội dung nộp sẽ bao gồm:

- 01 file báo cáo (theo mẫu).
- 01 file zip chứa dữ liệu đã gán nhãn.

Đặt tên là **MSSV\_BaiThucHanh2.zip**, nhóm cử 1 bạn thành viên đại diện nộp bài lên course.

### **Yêu cầu:**

- Các bạn sinh viên chọn nhóm tối thiểu 2 bạn tối đa 5 bạn cho Bài Thực hành 1 và Bài Thực hành 2.
- Nếu có nhiều hơn một nhóm chọn cùng 1 bài toán thì cần đảm bảo dữ liệu thu thập của hai nhóm đôi một không có quá 10% samples trùng lặp.
- Hoàn thành các bài tập được giao.

**Bài 1:** Tìm hiểu các tiêu chuẩn của một benchmark.

**Bài 2:** Xây dựng guideline cho việc xây dựng dataset. Cần mô tả trước bài toán (xác định input và output).

**Bài 3:** Tiến hành gán nhãn cho 500 samples đã crawl và đánh giá độ đồng thuận cho dataset. Cần giải thích lý do chọn độ đo cho dataset.