

### **Yêu cầu chung:**

- Với bài tập tìm hiểu cần có báo cáo tóm tắt nội dung đã tìm hiểu được.
- Với bài tập về ML yêu cầu tạo và update code lên github.

## **Bài 1:**

- Cho một ma trận có m hàng và n cột, sao cho trong 1 hàng thì các số có giá trị tăng dần từ trái sang phải, trong 1 cột thì các số cũng có giá trị tăng dần từ trên xuống dưới. Đếm xem có bao nhiêu số âm trong ma trận này với độ phức tạp là  $O(m+n)$ . Trình bày và code thuật toán.
- Có thuật toán nào đạt độ phức tạp tốt hơn không?

## **Bài 2:**

Xét quá trình đánh giá hiệu năng của một bài toán học máy thông thường:

- Có những chỉ số nào để đánh giá chất lượng một mô hình? Trình bày từng chỉ số.
- So sánh tương quan giữa prediction và recall. Khi nào prediction cần quan tâm hơn và khi nào thì recall cần được quan tâm hơn?
- So sánh chỉ số accuracy và f1-score? Khi nào thì đánh giá theo chỉ số accuracy và khi nào thì đánh giá theo chỉ số f1-score?
- So sánh khác biệt giữa f1-score micro và f1-score macro.

## **Bài 3:**

Bài toán phân loại với ML cơ bản

Dataset bài toán phân loại văn bản tự động:

<https://drive.google.com/file/d/1RtLXUXtWa4keC4-tTdWVzKve9H8z0u4W/view?usp=sharing>

Yêu cầu:

- Trình bày phương pháp xử lý dữ liệu. Tại sao xử lý như vậy?
- Trình bày mô hình lựa chọn và lý do lựa chọn mô hình.

- Trình bày hàm loss và cách hoạt động của hàm loss trong mô hình sử dụng trong mô hình?
- Đánh giá kết quả và trình bày lý do mô hình không đạt kết quả như mong muốn.
- Đề xuất ít nhất 1 phương pháp để cải tiến.

*Khuyến khích tìm hiểu thử nghiệm nhiều phương pháp xử lý dữ liệu, học máy/học sâu và phương pháp biểu diễn từ/văn bản trong bài toán này.*