**Câu hỏi:**

Mỗi nhóm tìm dăm ba ví dụ về bài toán regression \*\*\***TRONG THỰC TẾ**\*\*\*  
Ghi rõ input, output và cách thu thập + xử lý data, commit vào github repository và dẫn link lên đây. Có thể ghi ra giấy rồi chụp hình hoặc viết ở dạng Markdown.

**Ví dụ 1:**

Dự đoán nhiệt độ và lượng mưa trung bình vào các ngày trong tháng tiếp theo ở Việt Nam

* Input: Là số nguyên: Tháng, năm cần dự đoán
* Output: Là số thực: Nhiệt độ trung bình, lượng mưa trung bình của tháng cần dự đoán
* Data: Thu thập dữ liệu từ các file dữ liệu được các đơn vị nghiên cứu khí hậu công bố như là <https://data.opendevelopmentmekong.net/vi/dataset/nhi-t-d-va-lu-ng-mua-trung-binh-hang-thang-t-i-vi-t-nam-t-1901-2015?type=dataset> của Đại học East Anglia (UEA).
* Xử lý dữ liệu: loại bỏ những dữ liệu không đầy đủ thông tin hoặc gán tự động cho các dữ liệu đó một con số.

**Ví dụ 2:**

Dự đoán lượng hàng cần sản xuất để đạt doanh thu cao nhất và doanh thu của công ty trong tháng/ năm tiếp theo

* Input: Là số nguyên: Tháng/ năm cần dự đoán
* Output: Là số nguyên: Lượng hàng cần sản xuất để đạt doanh thu cao nhất, doanh thu đạt được
* Data: Thu tập dữ liệu từ cơ sở dữ liệu của công ty hoặc mua dữ liệu từ công ty trên các trang web bán dữ liệu <https://www.vietdata.vn/> hoặc liên hệ trực tiếp công ty
* Xử lý dữ liệu: loại bỏ những dữ liệu không đầy đủ thông tin hoặc gán tự động cho các dữ liệu đó một con số.

**Ví dụ 3:**

Dự đoán tỉ lệ mắc bệnh tim mạch dựa trên cân nặng, tuổi tác, trạng thái bệnh tình, lượng mỡ, nồng độ các chất trong máu, ….

* Input: Là cả dạng số thực, số nguyên, chuỗi ký tự: cân nặng (số nguyên), tuổi tác (số nguyên), trạng thái bệnh tình (chuỗi hoặc số tùy theo cách lưu dữ liệu của cơ sở dữ liệu), lượng mỡ (số thực tính theo %), nồng độ các chất trong máu (số thực tính theo % hoặc tính theo đơn vị đo lường)
* Output: Là số thực: tỉ lệ mắc bệnh tim mạch của một bệnh nhân nào đó
* Dữ liệu: Thu thập bằng cách truy cập vào cơ sở dữ liệu của các bệnh viện hoặc mua dữ liệu từ bệnh viện
* Xử lý dữ liệu: loại bỏ những dữ liệu không đầy đủ thông tin hoặc gán tự động cho các dữ liệu đó một con số.