

Lập trình nâng cao

Cho tập [dữ liệu](#) về đánh giá game gồm những cột như sau

- [game](#): tên
- [platform](#): nền tảng
- [developer](#): nhà phát triển
- [genre](#): thể loại
- [rating](#): xếp hạng độ tuổi
- [release_date](#): ngày phát hành
- [metascore](#): điểm chuyên gia
- [user_score](#): điểm người chơi

Sinh viên thực hiện những yêu cầu sau

1. Đọc dữ liệu từ file [metacritic_games.csv](#)
2. Chuyển đổi các cột sau:
 - a. [release_date](#) thành kiểu ngày tháng.
 - b. [rating](#) thành kiểu phân loại.
3. Thống kê số lượng game phát hành theo [platform](#) và năm phát hành.
4. Lấy ra tất cả những game có phân loại và trên nền tảng là [PC](#), chúng ta sẽ sử dụng dữ liệu này cho các câu sau.
5. Vẽ biểu đồ tần suất cho điểm người chơi.

Hồi quy tuyến tính

6. Chuyển dữ liệu cột [rating](#) thành mã số của phân loại (cat.code).
7. Lấy ra các cột [release_date](#), [rating](#), [metascore](#), [userscore](#) rồi tách dữ liệu vừa lấy thành 2 phần:
 - a. [train](#): tất cả những game phát hành trước 2019.
 - b. [test](#): phần còn lại.
8. Chuẩn hóa bằng điểm Z cho các cột [rating](#), [metascore](#), [userscore](#) của cả hai phần. Tuy nhiên chỉ dùng trung bình và độ lệch chuẩn của phần [train](#) để chuẩn hóa cho cả hai phần.
9. Xây dựng mô hình hồi quy cho với X là [rating](#), [metascore](#) còn Y là [userscore](#).
10. Tính điểm của mô hình.

Tham khảo:

- [Kiểu phân loại.](#)
- [Chuẩn hóa dữ liệu.](#)
- [Phân tích hồi quy.](#)

Nộp bài: <https://forms.gle/6J2BTCsAA1EWs4PM8>