

# PROJECT MÔN XỬ LÝ DỮ LIỆU LỚN

**Giảng viên:** Tiến sĩ Bùi Thanh Hùng  
Khoa Công nghệ thông tin  
Đại học Tôn Đức Thắng  
tg\_buithanhhung@tdtu.edu.vn  
**Website:** <https://sites.google.com/site/hungthanhbui1980>  
**Nộp bài: Classroom:** Nén thành 1 file gồm (code+dữ liệu+report)  
**Deadline:** 24h trước Session 15 hai ngày  
**Chấm bài:** Session 15

## Project

Thực hiện các yêu cầu sau và viết báo cáo theo mẫu gửi kèm:

1. Phân tích yêu cầu bài toán: Phân tích được yêu cầu của bài toán là gì (0.5 điểm)
2. Phương pháp giải quyết: Trình bày được các phương pháp giải quyết bài toán, Giải thích lý do tại sao chọn phương pháp này, Vẽ được sơ đồ tổng quát giải quyết bài toán (2 điểm)
3. Hiện thực - Viết code theo phương pháp giải quyết ở trên: Trình bày được cụ thể giải thuật sử dụng để giải quyết bài toán (Lưu đồ giải thuật), các tham số sử dụng, các thư viện sử dụng, code của bài toán ... (4.5 điểm)
4. Đánh giá kết quả đạt được: So sánh với ít nhất 1 phương pháp khác, Vẽ được biểu đồ so sánh giữa các phương pháp theo các độ đo ví dụ như: Accuracy, MSE, RMSE, MAP, .... (hãy lựa chọn ít nhất 2 độ đo trong các độ đo phổ biến để đánh giá bài toán trên) (1.25 điểm)
5. Hướng phát triển trong tương lai: Đưa ra được hướng phát triển trong tương lai và giải thích lý do tại sao lại đưa ra hướng phát triển đó (0.25 điểm).
6. Báo cáo bằng tiếng Việt theo mẫu: có đầy đủ các mục theo yêu cầu, phải tự căn chỉnh và soát lỗi chính tả, các công thức tự gõ, các hình, bảng đánh số, tài liệu tham khảo theo quy định và đánh chỉ mục vào trong bài từ nhỏ tới lớn (1 điểm)
7. Điểm làm việc nhóm (0.5 điểm)

## **ĐỀ TÀI**

### **HỆ THỐNG GỢI Ý KẾT BẠN**

Hãy xây dựng hệ thống gợi ý kết bạn (Recommender System) bằng phương pháp Graph-based  
Tham khảo:

Yêu cầu:

- Thu thập dữ liệu friend từ facebook (có thể lấy dữ liệu có sẵn)
- Xây dựng hệ thống gợi ý bạn (các bạn nào có cùng nhóm (clustering) thì gợi ý để kết bạn)

<https://arxiv.org/pdf/1707.07435.pdf>

<https://www.kaggle.com/georsara1/a-recommendation-engine-based-on-graph-theory>

<https://github.com/chandan-u/graph-based-recommendation-system>

<https://openreview.net/pdf?id=ByxxgCEYDS>