

# Requirements

## Graph Reduce Contrastive Learning ADM

---

### 1. Thông tin tổng quan dự án

- Mục tiêu của dự án: Xây dựng **code, figbug, support** đầy đủ **flow model 1, 2** về chủ đề **Graph Reduce Contrastive Learning ADM**
- Phạm vi công việc (Scope): **code model 1, 2**
- Đối tượng người dùng (User personas): **ManhVM, Chung nhận được code và hướng dẫn chạy, support cài đặt và chạy trên server**
- Các bên liên quan (Stakeholders): **Null**

### 2. Yêu cầu chức năng (Functional Requirements)

- Các tính năng chính và thêm:
  - **Chính:**
    - **Code** theo flow chính
    - Figbug
    - Support chạy
  - **Thêm:**
    - Chỉnh sửa các thông số
    - Thay đổi các thông số, mô hình
    - Config
- Luồng hoạt động của hệ thống (**Workflow**):

#### Main workflow: Apks

- trích xuất FCG
- tìm đồ thị bằng 2 phương pháp Local Degree và SCAN
- bổ sung thuộc tính đỉnh bằng 5 thuộc tính (in-degree, out-degree, closeness, Katz, clustering coefficient) concat với vector embedding của source code hàm dùng CodeBERT
- Sử dụng GraphSAGE tạo ra vector nhúng
- vector nhúng bởi GraphSAGE kết hợp theo cơ chế **attention** với **vector quyền** được nhúng bởi mô hình **SentenceTransformer minilm\_l6\_v2**
- > **phân loại: Fully Connected and Softmax, RF, DT**

#### Workflow 1:

- **Đo time training, tính trung bình cạnh, node**
- **fcg (cắt tỉa các loại) với phân loại FC + Softmax**
- **Tốt nhất hệ số -> Fully Connected and Softmax, RF, DT**
- **—**
- **Đo time training, tính trung bình cạnh, node**
- **Không sử dụng tỉa + Fully Connected and Softmax, RF, DT**

- Tỉa bằng API (thuật toán của thầy) -> FC + Softmax
- —
- Tỉa tốt nhất (tham số) + thay đổi GNN
- —
- Vector quyền (nhúng sentence bert) phân loại -> FC + Softmax
- Nhúng GraphSage -> FC + Softmax
- Concat 2 cái trên -> FC + Softmax
- —
- Thay đổi kịch bản chính thay đổi: vector quyền bằng 2 cách:
  - One hot
  - W2v

#### Workflow 2:

- Thêm contrastive learning + FC
- Không dùng vector quyền attention

- Các trường hợp sử dụng (Use Cases)
  - Sử dụng trong quá trình thí nghiệm
  - Thử nghiệm các loại thông số
- Quy trình nghiệp vụ liên quan: **Null**

### 3. Yêu cầu phi chức năng (Non-functional Requirements)

- Hiệu suất (Performance): **chạy được trên GPU 12G**
- Bảo mật (Security): **Null**
- Khả năng mở rộng (Scalability): **có thể mở rộng khi cơ sở hạ tầng lên**
- Độ tin cậy (Reliability): **Null**
- Trải nghiệm người dùng (UX/UI expectations): **Null**

### 4. Yêu cầu tích hợp

- Chạy trên **server**

### 5. Yêu cầu về dữ liệu

- Nguồn dữ liệu được cung cấp: **ManhVM cung cấp**

### 6. Yêu cầu về hạ tầng & triển khai

- Máy chủ (Cloud, On-premise, Hybrid): **Dùng server 12G GPU, 12G Ram**
- Hệ điều hành, công nghệ, framework ưu tiên: **Pytorch, DGL, Networkx, Networkit**
- Cách thức triển khai (CI/CD, Docker, Kubernetes, v.v.): **Null**

## 7. Yêu cầu bảo mật

- Cơ chế xác thực & phân quyền (OAuth, SAML, JWT): **Null**
- Quy tắc bảo vệ dữ liệu người dùng (GDPR, ISO 27001, v.v.): **Null**
- Chính sách sao lưu và phục hồi dữ liệu: **Null**

## 8. Yêu cầu bảo trì & hỗ trợ

- Chính sách cập nhật phần mềm: **Null**
- Hỗ trợ kỹ thuật sau triển khai: **Tuyên**

## 9. Yêu cầu về thời gian & ngân sách

- Deadline cho từng giai đoạn phát triển:
  - **Mong muốn: Thời gian 3 tuần (25/02/2025 -> 18/03/2025)**
  - **Muộn nhất:** Thời gian 1 tháng (25/02/2025 -> 25/03/2025)
- Ngân sách ước tính: **15tr**
- Nhân lực tham gia dự án: **Tuyên**

## 10. Yêu cầu đặc biệt khác

- Yêu cầu về **hỗ trợ hệ thống**
  - Thông báo checkpoint vào tối **T2 hàng tuần**
  - Các tiêu chuẩn hoặc quy định cần tuân thủ
-