



**MẠNG XÃ HỘI - IS353.P12**

# **DỰ ĐOÁN XẾP HẠNG TỐT NGHIỆP CỦA SINH VIÊN UIT DỰA VÀO THÀNH TÍCH HỌC TẬP VÀ CÁC YẾU TỐ KHÁC**

**GROUP 06**

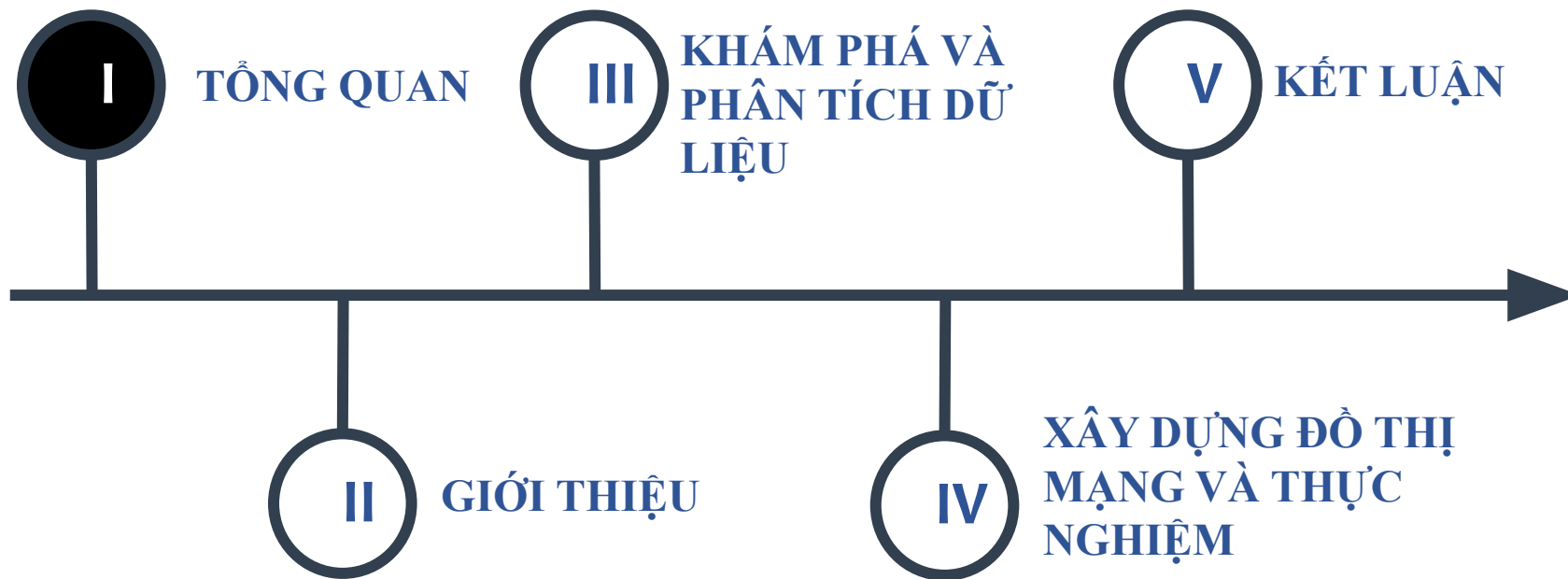
---

**GVHD: Th.S Nguyễn Thị Anh Thư**



## THÀNH VIÊN NHÓM

1	TRẦN NHẬT VĨ (nhóm trưởng)	22521659
2	BÙI ÁI ĐỨC	22520352
3	ĐẶNG ÁNH PHƯỚC	21520404
4	PHẠM NHẬT TÂN	22521311
5	BÙI TUẤN HUY	20520537
6	MAI HOÀNG VINH	22521673





# TỔNG QUAN

## 01

## GIỚI THIỆU ĐỀ TÀI

- **Tính cấp thiết trong quản lý và hỗ trợ sinh viên:** Giúp nhà trường phát hiện sớm những sinh viên có nguy cơ xếp hạng thấp để tư vấn, hỗ trợ, cũng như tối ưu hóa chất lượng giáo dục, đặc biệt trong các chương trình hỗ trợ kỹ năng và học thuật.
- **Khai thác tiềm năng của dữ liệu trong giáo dục:** Với sự đa dạng và toàn diện của loại hình dữ liệu này, hoàn toàn có thể từ đó trích xuất và nghiên cứu những mối quan hệ như sự tương tác giữa sinh viên qua các hoạt động và lớp học, ảnh hưởng của yếu tố thành tích và hoạt động khác đến xếp hạng tốt nghiệp của sinh viên.
- **Khả năng mở rộng:** Từ ý tưởng và cách thức thực hiện bài toán này, có thể mở rộng sang các bài toán khác như: Dự đoán tỷ lệ tốt nghiệp đúng hạn; Phân tích các yếu tố dẫn đến thành công trong nghề nghiệp sau tốt nghiệp,... đóng góp thiết thực vào việc cải thiện chất lượng giáo dục và tối ưu hóa nguồn lực trong quản lý sinh viên.



# TỔNG QUAN

01

## ĐỐI TƯỢNG VÀ PHẠM VI

**Đối tượng** của bài toán là các sinh viên thuộc khóa 8 - 14 (từ năm 2013 đến năm 2019) học tại trường Đại học Công nghệ thông tin

**Phạm vi** của bài toán bao gồm việc thu thập dữ liệu về các yếu tố cá nhân của sinh viên như tuổi, giới tính, quê quán, điểm trung bình học kỳ, điểm trung bình tích lũy, số tín chỉ tích lũy, trình độ ngoại ngữ, điểm rèn luyện, vv. cùng với xếp hạng tốt nghiệp và ngày cấp bằng của sinh viên từ năm 2017 - 2021.

## PHÁT BIỂU BÀI TOÁN

**Dữ liệu quá trình học tập:** Điểm trung bình các năm, điểm rèn luyện các năm, số tín chỉ tích lũy,...

**Dữ liệu khác:** Xử lý học vụ, bảo lưu quá trình học tập, thông tin xét tuyển (Điểm THPT),...



# TỔNG QUAN

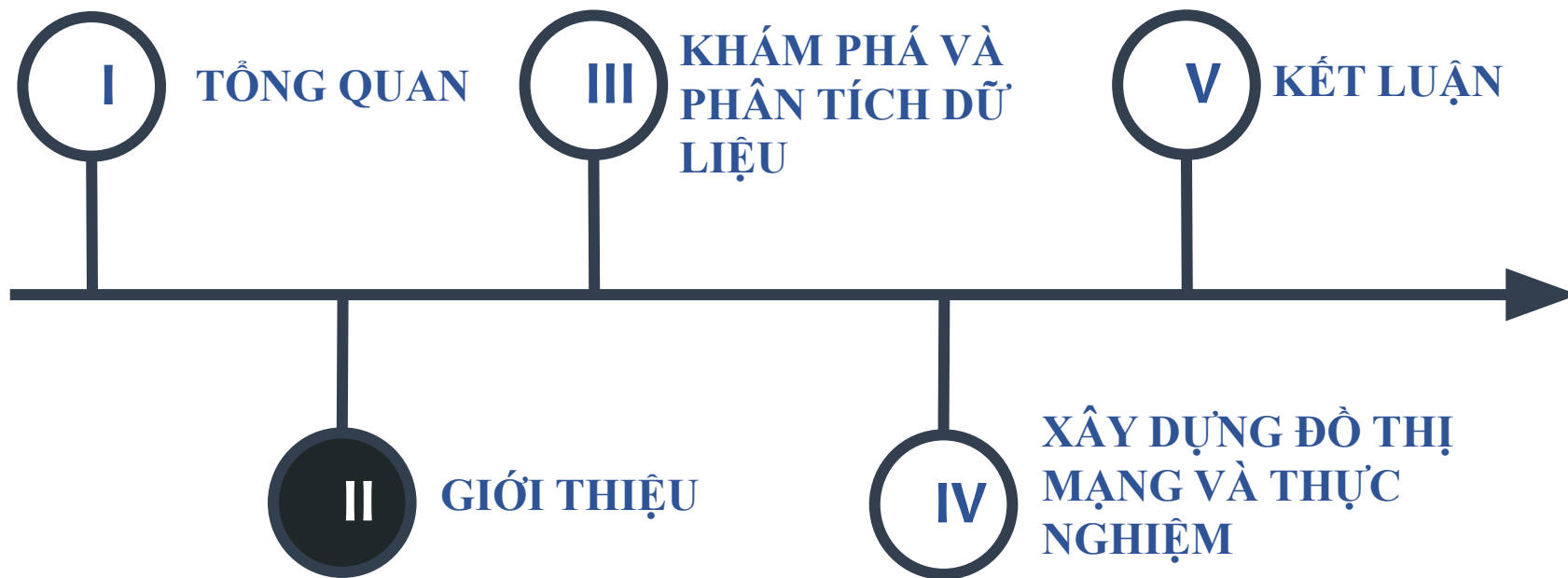
## 01

### MỤC TIÊU ĐỀ TÀI

- Ứng dụng đồ thị mạng xã hội để mô hình hóa mối quan hệ phức tạp trong dữ liệu sinh viên
- Kết hợp đồ thị mạng vào bài toán phân loại để tối ưu hóa dự đoán xếp loại tốt nghiệp

### THÁCH THỨC BÀI TOÁN

- Đầu vào dữ liệu
- Số lượng và tính đa dạng của các yếu tố ảnh hưởng đến xếp hạng tốt nghiệp
- Lựa chọn phương pháp giải quyết bài toán
- Giải thích kết quả và ứng dụng





# GIỚI THIỆU BỘ DỮ LIỆU

02

## GIỚI THIỆU TỔNG QUAN BỘ DỮ LIỆU

Bộ dữ liệu gốc được cung cấp gồm 15 bảng chứa thông tin của sinh viên thuộc khóa 8 - 14 (từ năm 2013 đến năm 2019)

ST T	Tên bảng	Ý nghĩa	Số mẫu
1	01.sinhvien	Thông tin của sinh viên đại học Công nghệ Thông tin. Gồm năm sinh, giới tính, nơi sinh, lớp, khoa, hệ đào tạo,... khóa 8 - 14	8316
2	02.diem	Điểm các môn học của từng sinh viên, năm 2012 - 2016.	99099
3	03.sinhvien_chungchi	Thông tin chứng chỉ Ngoại ngữ của sinh viên, năm 1999 - 2021.	3400
4	04.xeploaiav	Điểm kiểm tra Tiếng Anh đầu vào và thông tin xếp lớp anh văn của sinh viên.	6349
5	05.ThiSinh	Thông tin và điểm số của các thí sinh dự tuyển vào trường.	8234
6	06.giayxacnhan	Thông tin về các giấy xác nhận.	27259
7	08.XLHV	Thông tin xử lý học vụ sinh viên, năm 2017 - 2020.	3452

8	10.diemrl	Điểm rèn luyện của từng sinh viên, năm 2013 - 2020.	54057
9	12.baoluu	Thông tin bảo lưu của sinh viên, năm 2016 - 2020.	1880
10	14.totnghiep	Thông tin tốt nghiệp của sinh viên. Bao gồm xếp loại tốt nghiệp và ngày cấp bằng, năm 2017 - 2021.	1847
11	diemrl	Điểm rèn luyện của từng sinh viên, năm 2013 - 2022.	111978
12	diem_Thu	Điểm thành phần từng môn của sinh viên. năm 2006 - 2022	674273
13	sinhvien_dtb_hocky	Điểm trung bình toàn học kỳ của sinh viên, năm 2013 - 2022.	84952
14	sinhvien_dtb_toankhoa	Điểm trung bình toàn khóa, điểm trung bình tích lũy và số tính chỉ tích lũy của sinh viên.	13970
15	uit_hocphi_miengiam	Thông tin về mức miễn giảm học phí của sinh viên năm 2013-2019	5652





# GIỚI THIỆU BỘ DỮ LIỆU

02

## TIỀN XỬ LÝ DỮ LIỆU

### LÀM SẠCH DỮ LIỆU

1. Bảng ***sinhvien***: Thuộc tính **noisinh** và **diachi\_tinhtp** có nhiều giá trị khác nhau nhưng chung 1 ý nghĩa (ví dụ: Thành phố Hồ Chí Minh = Tp. Hồ Chí Minh) được đưa về 1 giá trị để xử lý dễ hơn. Xóa bỏ các thuộc tính có giá trị nhưng không có tên thuộc tính.
2. Bảng ***sinhvien\_chungchi***: điều chỉnh đúng tên của cột thuộc tính và xoá cột lydo vì dữ liệu cột này hầu hết là NULL. Điền bổ sung các giá trị thiếu của cột **tongdiem** bằng tổng điểm thành phần tương ứng của điểm dữ liệu. Lọc các mẫu dữ liệu bị trùng.
3. Bảng ***xeploaiav***: thêm giá trị xếp lớp cho một số mẫu dữ liệu bị thiếu.
4. Bảng ***thisinh***: đổi tên và chuyển kiểu dữ liệu của một cột để đồng bộ hóa và đảm bảo tính chính xác.
5. Bảng ***XLHV***: chỉnh sửa các đoạn dữ liệu bị lỗi và loại bỏ các dấu phẩy không cần thiết.
6. Bảng ***diemdrl*** đã được lọc trùng tạo cột khoa để đơn giản hóa quá trình phân tích và tổng hợp dữ liệu
7. Bảng ***baoluu*** đã được kiểm tra tính đúng đắn của các giá trị, thông qua quá trình so sánh thông tin ở cột **soqd**, cho thấy thông tin bảng dữ liệu này liên quan đến nhiều quyết định khác nhau như điều chỉnh, công nhận tốt nghiệp, xét anh văn, thu học phí.. Tên bảng dữ liệu này không tương ứng với thông tin trong bảng, nhóm quyết định không sử dụng trong bước phân tích.
8. Bảng ***totnghiep*** chỉ cần thống nhất lại dữ liệu trong cột **xeploai** như “Trung bình Khá” được đưa về cùng “TB Khá”.
9. Bảng ***sinhvien\_dtb\_hocky*** đã được đổi định dạng dữ liệu của một cột và kiểm tra lại để đảm bảo tính chính xác và đáng tin cậy của dữ liệu.
10. Bảng ***sinhvien\_dtb\_toankhoa*** đã được thêm giá trị NULL vào các điểm dữ liệu bị trống để dễ làm việc với dữ liệu.



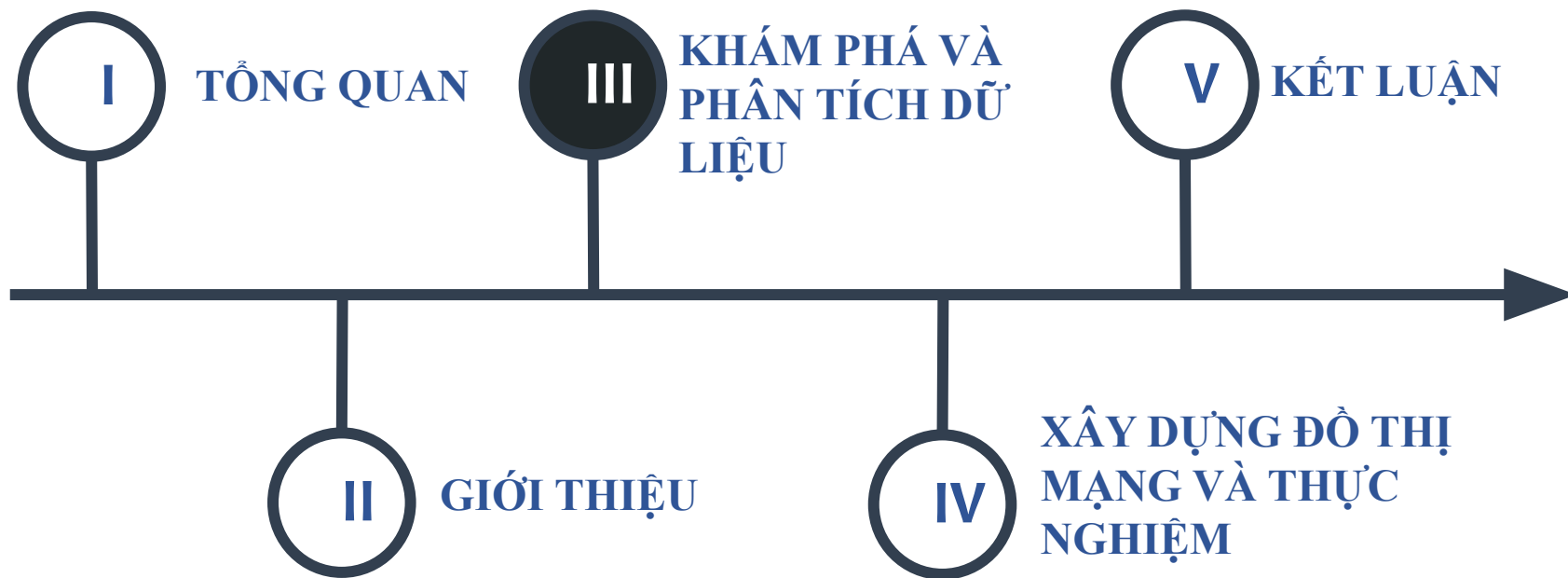
# GIỚI THIỆU BỘ DỮ LIỆU

02

## TIỀN XỬ LÝ DỮ LIỆU

### KẾT HỢP DỮ LIỆU

STT	Tên bảng	Ý nghĩa	Số mẫu
1	sv_totnghiep	Thông tin mẫu dữ liệu sinh viên tốt nghiệp.	1845
2	diem_hk_svt	Dữ liệu điểm theo học kỳ của các sinh viên tốt nghiệp.	17421
3	diem_nam_svt	Dữ liệu điểm của sinh viên tốt nghiệp theo từng năm.	8595
4	thisinh_svt	Dữ liệu thông tin xét tuyển đại học của sinh viên tốt nghiệp.	1811
5	drl_svt	Dữ liệu điểm rèn luyện của sinh viên tốt nghiệp.	9126
6	xlav_svt	Dữ liệu xếp loại anh văn của sinh viên tốt nghiệp.	313
7	chungchi_svt	Dữ liệu chứng chỉ ngoại ngữ của sinh viên tốt nghiệp.	967
8	diemtluy_svt	Dữ liệu điểm và tín chỉ tích lũy của sinh viên tốt nghiệp.	1831
9	XLHV_svt	Dữ liệu xử lý học vụ của sinh viên tốt nghiệp	187



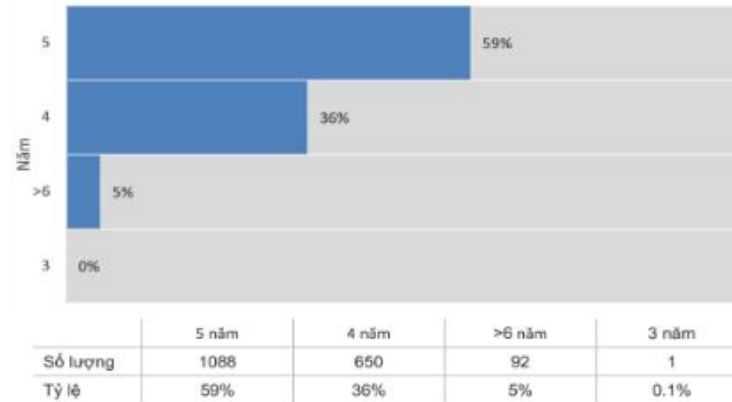


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM TRUNG BÌNH NĂM



*Hình 3-1: Thời gian đào tạo của các sinh viên đã tốt nghiệp giai đoạn 2017-2021.*

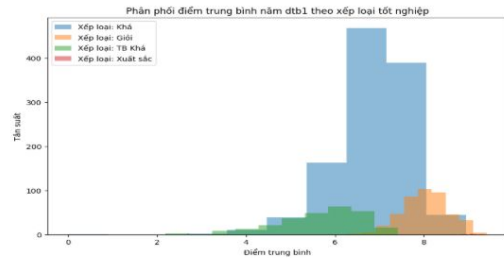


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

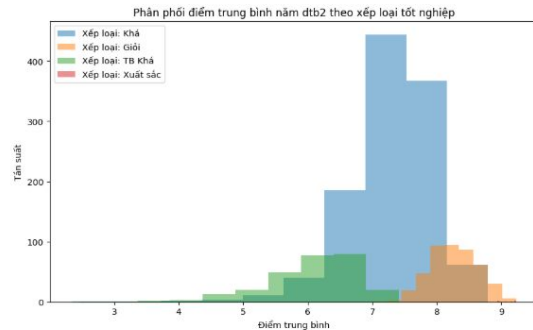
03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

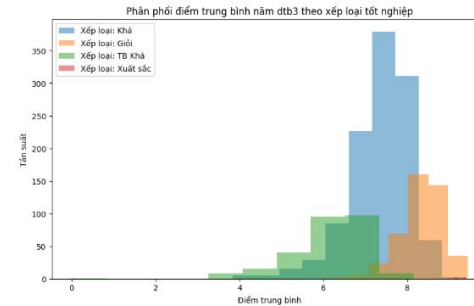
### ĐIỂM TRUNG BÌNH NĂM



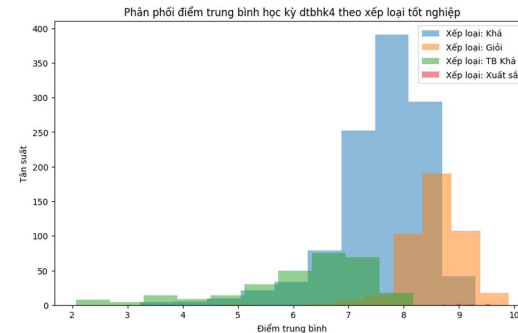
Hình 3-2: Phân phối điểm trung bình năm học thứ nhất theo xếp loại tốt nghiệp



Hình 3-3: Phân phối điểm trung bình năm học thứ hai theo xếp loại tốt nghiệp



Hình 3-4: Phân phối điểm trung bình năm học thứ ba theo xếp loại tốt nghiệp



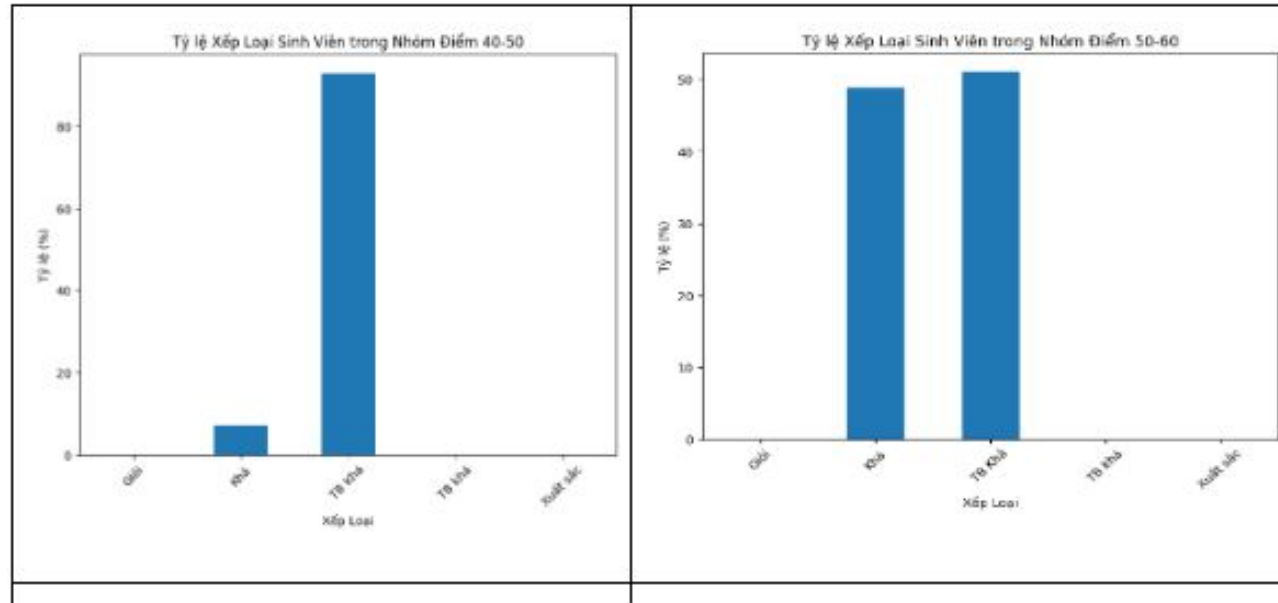


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM RÈN LUYỆN



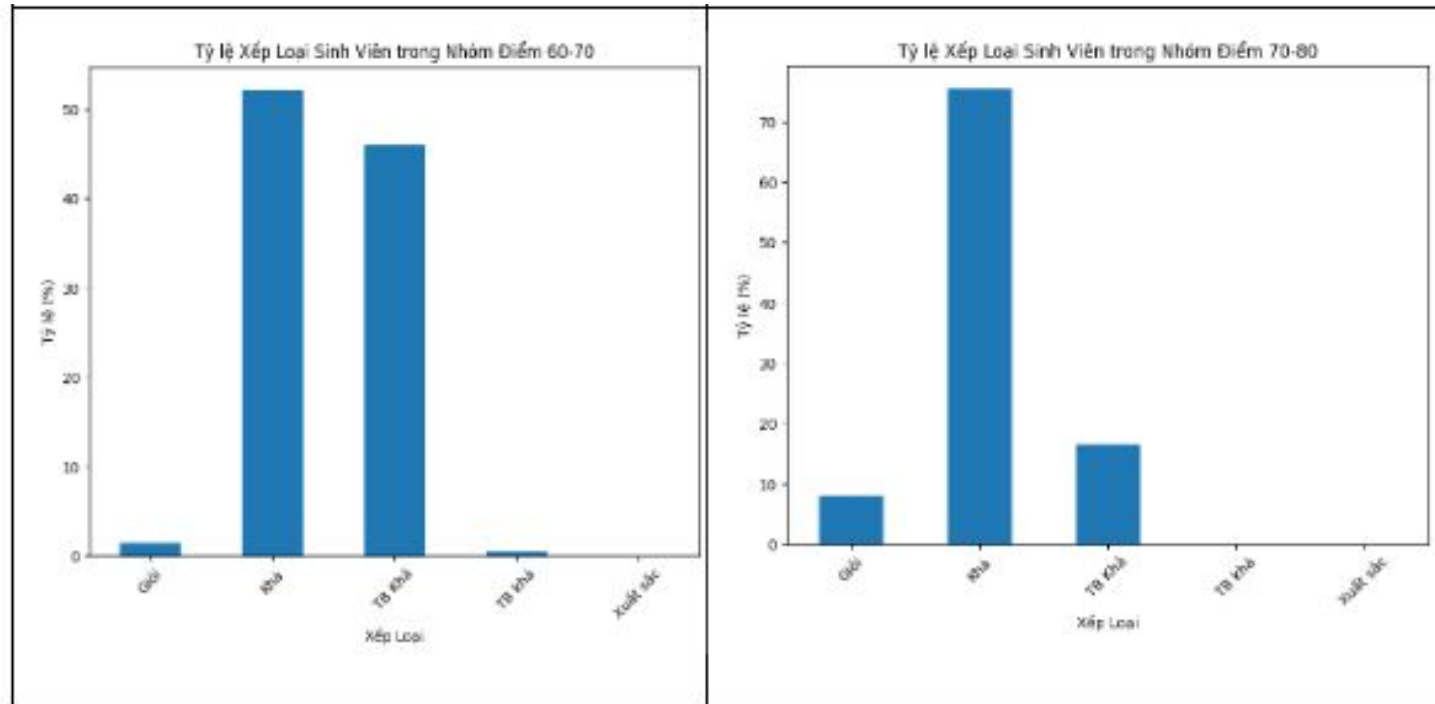


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM RÈN LUYỆN



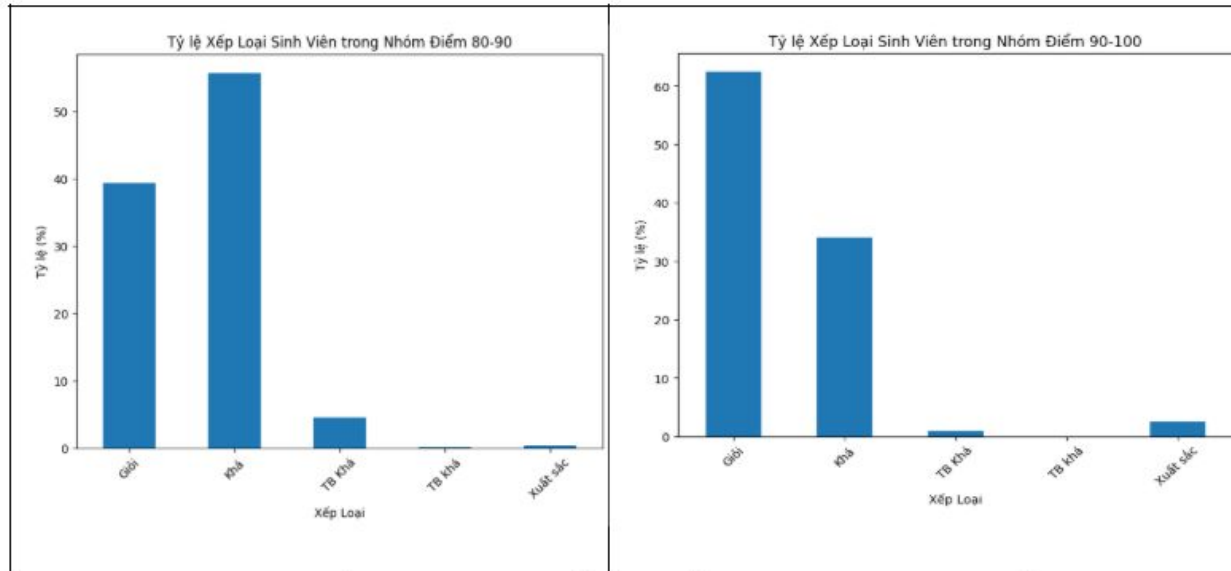


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM RÈN LUYỆN



**Bảng 3.1:** Bảng thể hiện tỉ lệ giữa xếp loại tốt nghiệp và các mức điểm rèn luyện





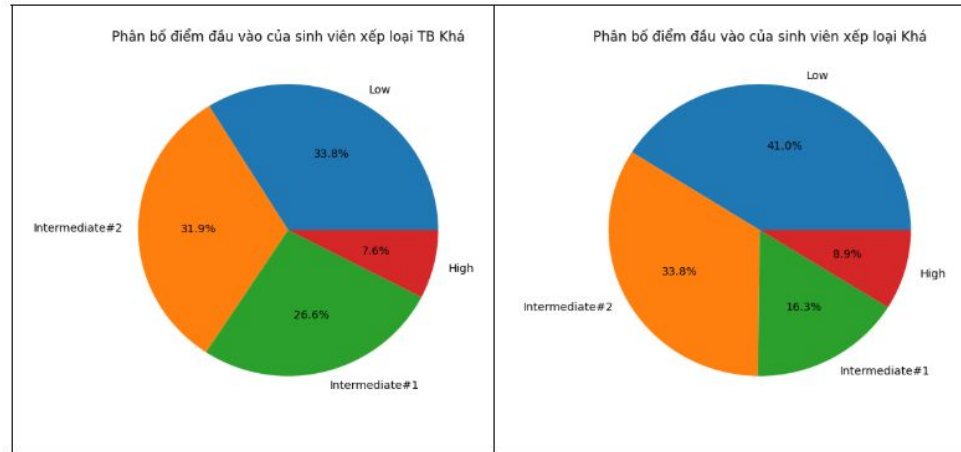
# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM THPT

- Low: nằm ở khoảng thấp hơn 24.5.
- Intermediate#1: nằm ở khoảng từ 24.5 đến nhỏ hơn 25.9.
- Intermediate#2: nằm ở khoảng từ 25.9 đến nhỏ hơn 29.5.
- High: trên 29.5.



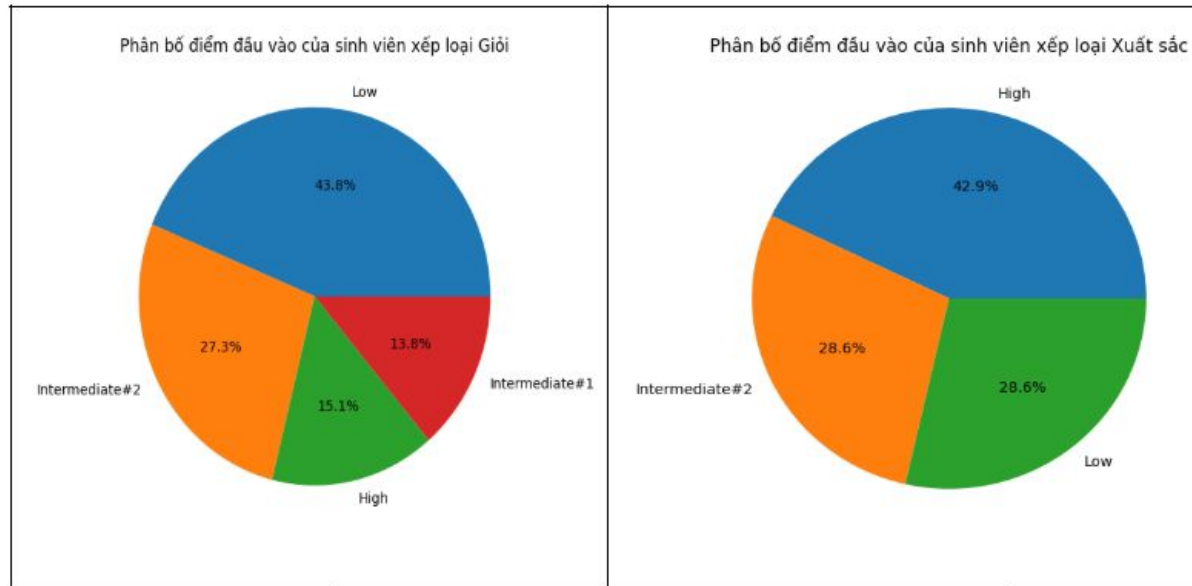


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

ĐIỂM THPT



*Bảng 3.2: Bảng thể hiện phân bố của xếp loại tốt nghiệp theo mức điểm đầu vào*

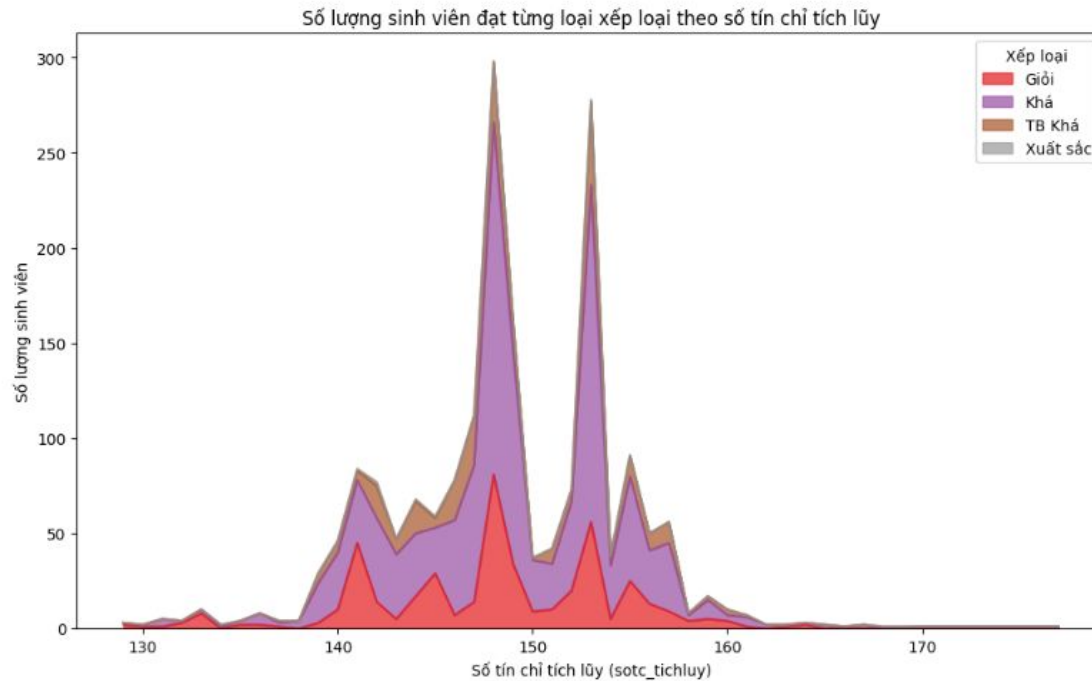


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### ĐIỂM TÍN CHỈ TÍCH LŨY



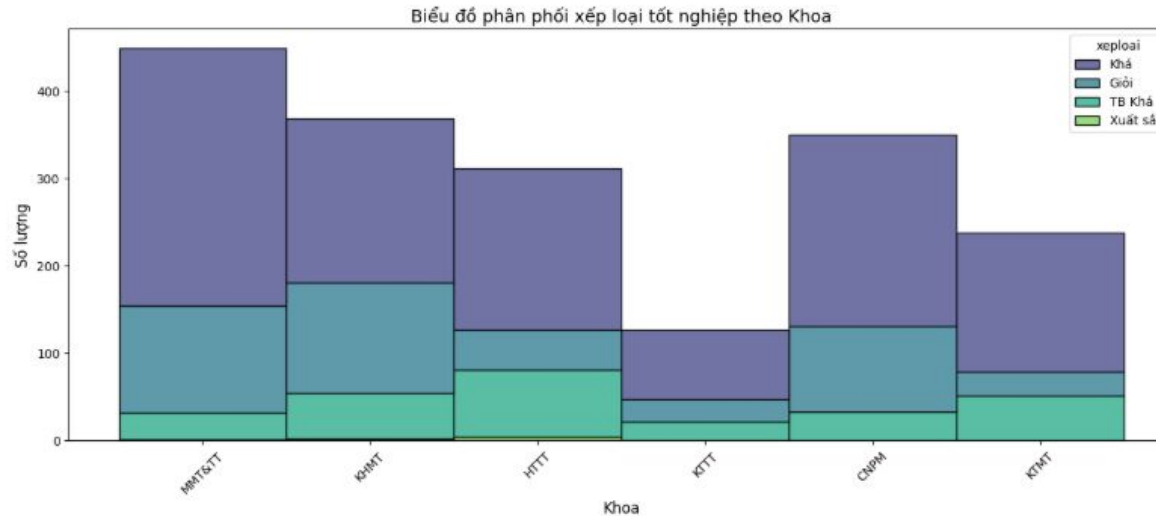


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

KHOA



*Hình 3-6: Biểu đồ phân phối xếp loại tốt nghiệp theo khoa*

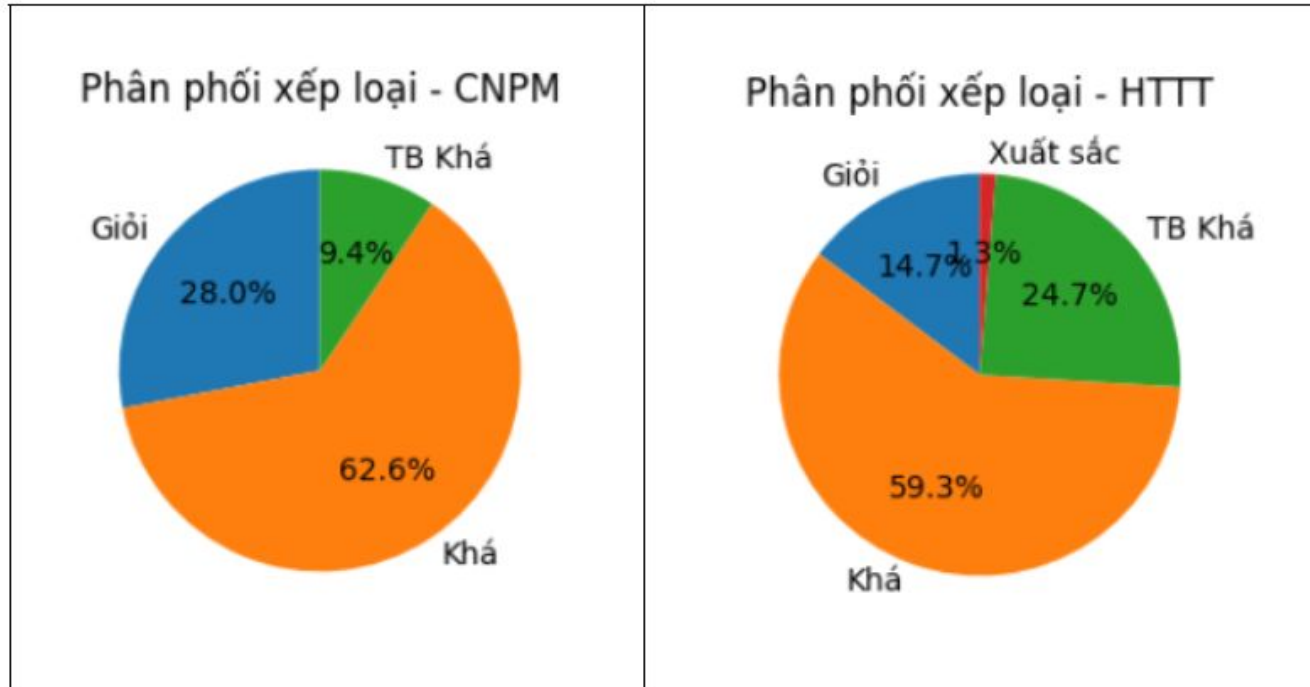


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

KHOA



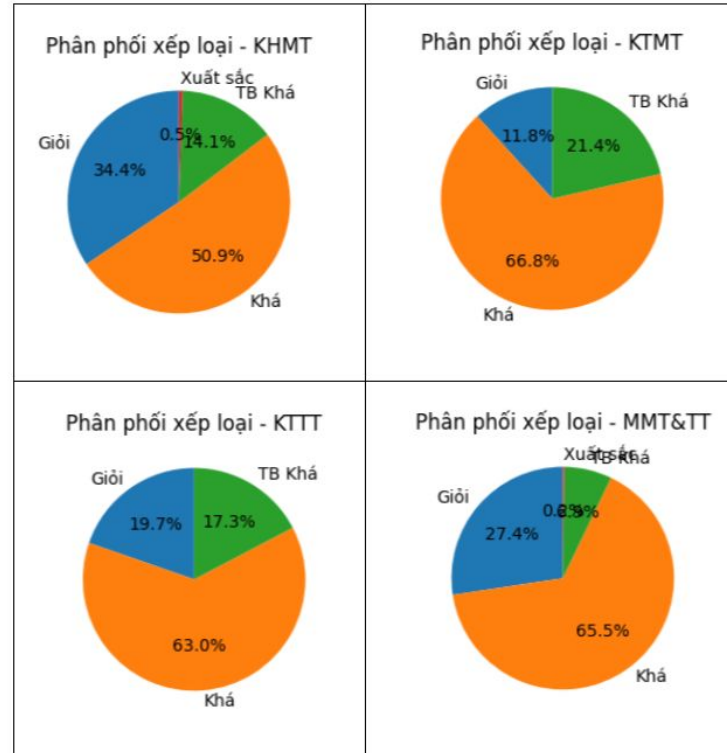


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

KHOA



Bảng 3.3: Bảng thể hiện phân phối chi tiết xếp loại tốt nghiệp của từng khoa cụ thể

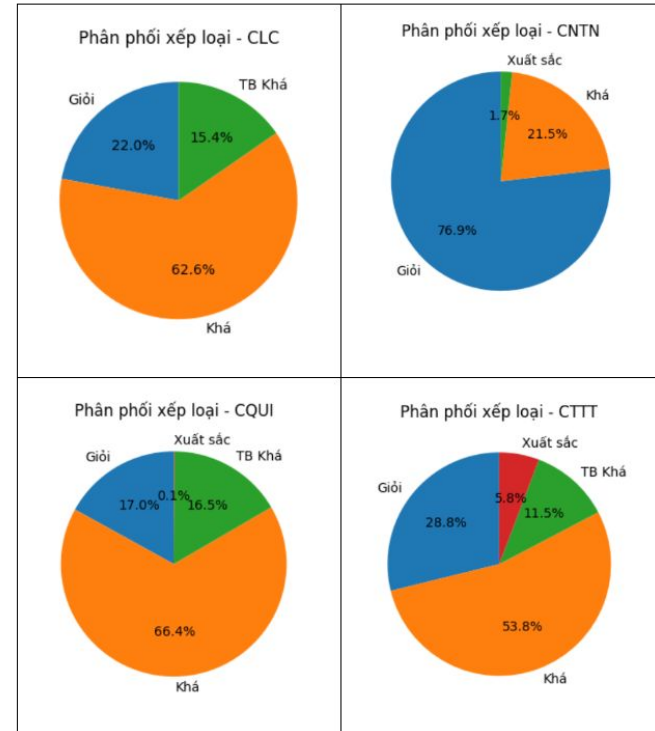
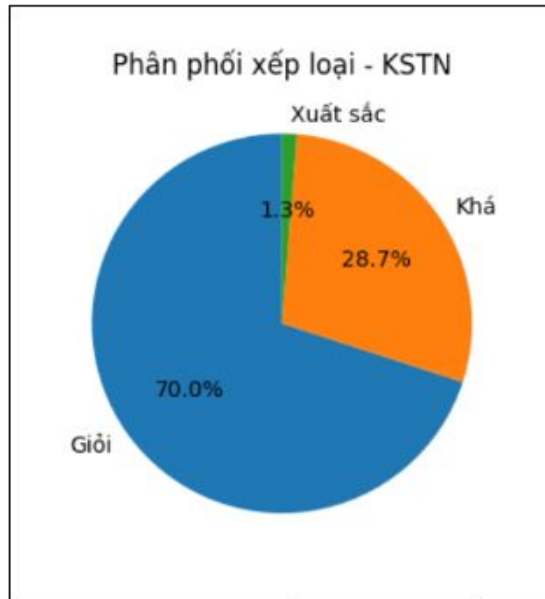


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

### HỆ ĐÀO TẠO

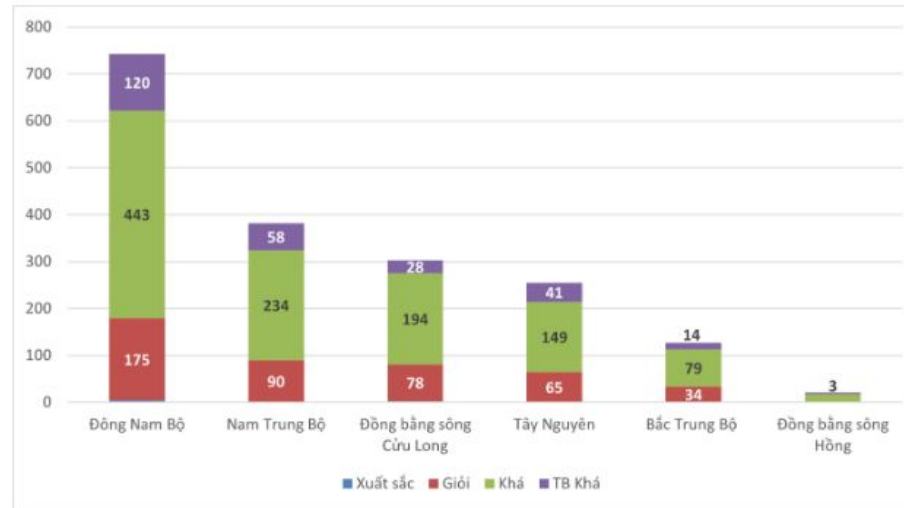




# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG VÙNG ĐỊA LÝ



**Hình 3-8:** Biểu đồ cột thể hiện số lượng sinh viên tốt nghiệp và phân bố xếp hạng sinh viên tốt nghiệp ở từng vùng địa lý tương ứng





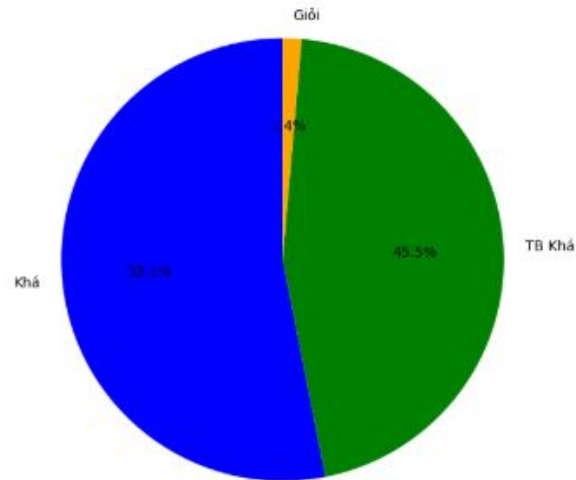
# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH CHI TIẾT CÁC YẾU TỐ ẢNH HƯỞNG

VI PHẠM HỌC VỤ

Tỷ lệ xếp loại tốt nghiệp của những người từng vi phạm (slvp > 0)





# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

03

## PHÂN TÍCH FEATURE IMPORTANCE CỦA RANDOM FOREST

### GIỚI THIỆU

- **Random Forest** là một thuật toán Machine Learning mạnh mẽ và linh hoạt được sử dụng rộng rãi trong cả bài toán phân loại (**classification**) và hồi quy (**regression**).
- **Feature Importance** trong Random Forest đo lường mức độ "đóng góp" của từng đặc trưng (feature) vào việc cải thiện chất lượng của mô hình.

Trong bài toán phân loại (**classification**), **impurity** thường được đo bằng:

- **Gini Impurity**: Đo độ hỗn tạp của các nhãn tại một nút trong cây.
- **Information Gain**: Sự tăng thông tin (giảm entropy) sau mỗi lần chia nhánh.

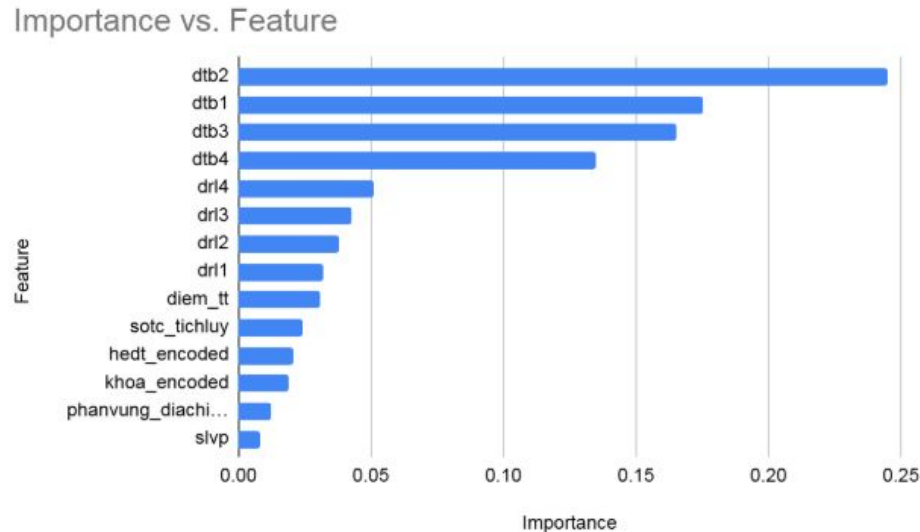


# KHÁM PHÁ VÀ PHÂN TÍCH DỮ LIỆU

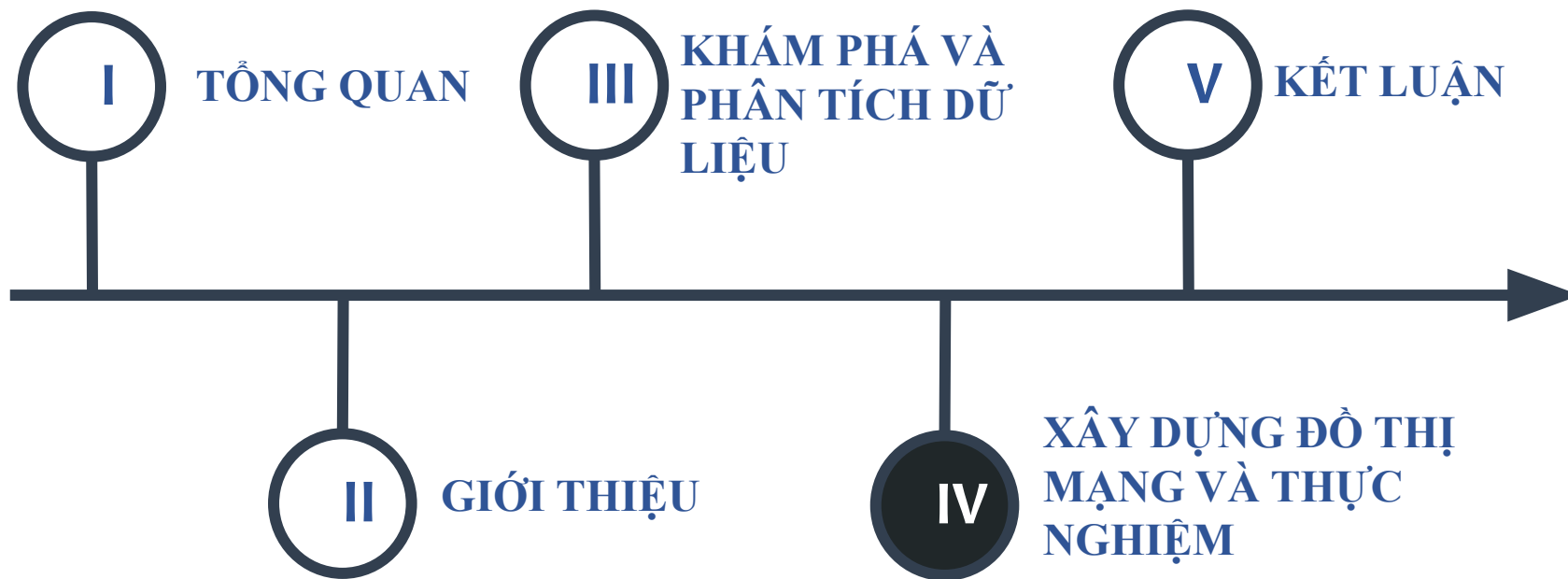
03

## PHÂN TÍCH FEATURE IMPORTANCE CỦA RANDOM FOREST

QUÁ TRÌNH TÍNH TOÁN VÀ KẾT QUẢ ĐẠT ĐƯỢC



*Hình 3-10: Biểu đồ cột thể hiện mức độ ảnh hưởng của các đặc trưng*





# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## PHƯƠNG PHÁP SỬ DỤNG

### THUẬT TOÁN LOUVAIN

- Thuật toán Louvain là một thuật toán đồ thị được sử dụng để **phân cụm đồ thị** dựa trên việc tối ưu **độ đo modularity**.
- Thuật toán Louvain hoạt động bằng cách lần lượt thực hiện hai giai đoạn: phân cụm và tái phân cụm.



# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## PHƯƠNG PHÁP SỬ DỤNG

### THUẬT TOÁN SUPPORT VECTOR CLASSIFICATION

- Support Vector Classification (SVC) là một thuật toán máy học dựa trên lý thuyết Support Vector Machines (SVM), được thiết kế để giải các bài toán phân loại. Mục tiêu chính của SVC là tìm một **siêu phẳng tối ưu** (optimal hyperplane) để phân tách các lớp trong không gian đặc trưng.
  - **SVC hoạt động qua hai bước chính:**
- Tìm siêu phẳng tối ưu
- Giải bài toán tối ưu hóa
  - vector trọng số  $w$ , độ dời  $b$ , khoảng cách lỗi  $\xi$ .

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i$$



# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## PHƯƠNG PHÁP SỬ DỤNG

### THUẬT TOÁN RANDOM FOREST

- Random Forest là một thuật toán máy học thuộc nhóm **ensemble learning**, kết hợp nhiều cây quyết định (decision trees) để đưa ra kết quả dự đoán mạnh mẽ và chính xác hơn.
- Mục tiêu của Random Forest là cải thiện độ chính xác của mô hình và giảm hiện tượng **overfitting** mà một cây quyết định đơn lẻ dễ gặp phải.
- Random Forest thực hiện các bước sau:
  - Lấy mẫu dữ liệu
  - Xây dựng cây quyết định
  - Dự đoán



# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

## 04

## XÂY DỰNG ĐỒ THỊ MẠNG

### CÁC ĐẶC TRƯNG

- Dựa vào lần áp dụng Feature Importance gần nhất, các đặc trưng được xếp theo thứ tự mức độ ảnh hưởng (Importance), và chúng tôi đã lựa chọn **3 đặc trưng quan trọng nhất phù hợp để xây dựng đồ thị mạng xã hội sinh viên**, gồm:
  - ❖ Điểm trung bình năm 2 (dtb2)
  - ❖ Điểm chênh lệch đầu vào (diem\_tt2)
  - ❖ Số lần vi phạm nội quy (slvp)





# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

## 04

## XÂY DỰNG ĐỒ THỊ MẠNG

### XÂY DỰNG MẠNG XÃ HỘI

#### Cấu trúc mạng xây dựng:

- ❖ Đỉnh (nodes): Đại diện cho các sinh viên đã tốt nghiệp (tổng cộng 1831 đỉnh).
- ❖ Cạnh (edges): Kết nối giữa các sinh viên dựa trên các tiêu chí tương đồng về một số đặc trưng đã được chọn lựa dưới đây:
  - Điểm trung bình năm 2
  - Điểm chênh lệch đầu vào
  - Số lần vi phạm nội quy

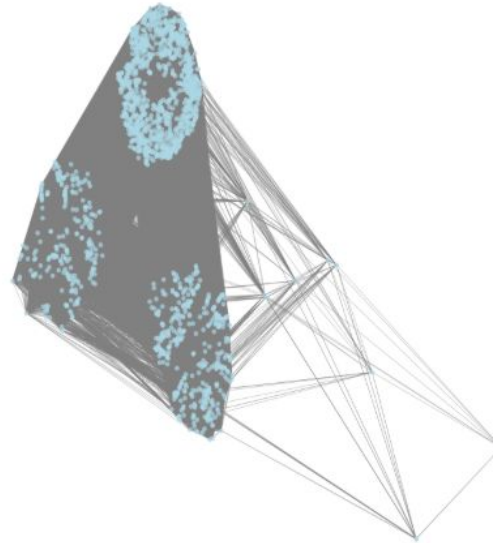


# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## XÂY DỰNG ĐỒ THỊ MẠNG

### XÂY DỰNG MẠNG XÃ HỘI



*Hình 4.2: Phân bố các đỉnh trên đồ thị mạng xã hội*



# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## KỊCH BẢN THỰC NGHIỆM

- **Thực nghiệm 1:** Tiến hành huấn luyện mô hình và đánh giá chỉ trên dữ liệu các đặc trưng “dtb” và “drl” của 4 năm học, ngoài ra còn có các đặc trưng như điểm thpt, số tín chỉ, khoa, hệ đào tạo với mô hình SVC và Random Forest.
- **Thực nghiệm 2:** Tiến hành huấn luyện mô hình và đánh giá chỉ trên dữ liệu các đặc trưng “dtb” và “drl” của 4 năm học, ngoài ra còn có các đặc trưng như điểm thpt, số tín chỉ, khoa, hệ đào tạo và kết hợp nhãn đã được phân cụm từ đồ thị mạng xã hội sử dụng thuật toán Louvain, dự đoán bằng mô hình SVC và Random Forest, Đồng thời sử dụng lại tính năng Feature Importance để đánh giá chi tiết hơn mức độ ảnh hưởng của việc áp dụng mạng xã hội.



# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

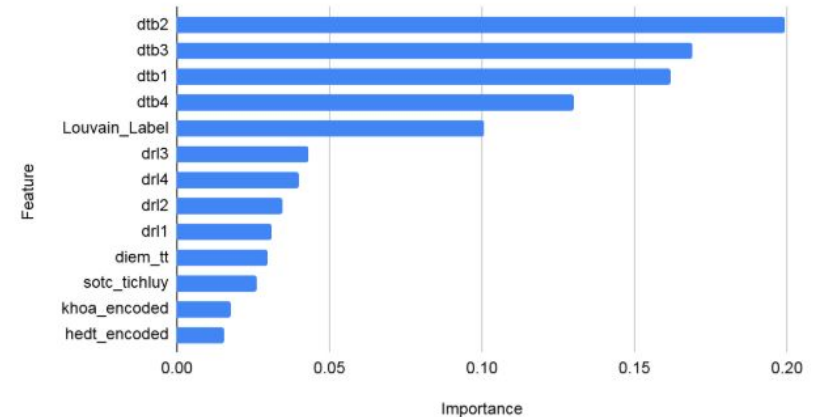
04

## KẾT QUẢ THỰC NGHIỆM

Nội dung	macro precision	macro recall	macro f1-score	weighted precision	weighted recall	weighted f1-score	accuracy
RFC_1	66.62%	65.68%	66.11%	90.12%	90.46%	90.27%	<b>90.46%</b>
RFC_2	68.06%	67.24%	67.63%	91.79%	92.10%	91.93%	<b>92.10%</b>
SVC_1	65.87%	65.23%	65.51%	85.96%	89.91%	89.72%	<b>89.91%</b>
SVC_2	64.80%	64.89%	64.82%	88.79%	89.10%	88.93%	<b>89.10%</b>

Bảng 4.2 Bảng so sánh kết quả của 2 thực nghiệm chính

Importance vs. Feature





# XÂY DỰNG ĐỒ THỊ MẠNG VÀ THỰC NGHIỆM

04

## KẾT QUẢ THỰC NGHIỆM

Thực nghiệm	Accuracy
RFC_1(offmxh)	90.46%
RFC_off_khoa_hedt_1(offmxh)	91.28%
RFC_2	92.10%
RFC_off_khoa_hedt_2	92.37%

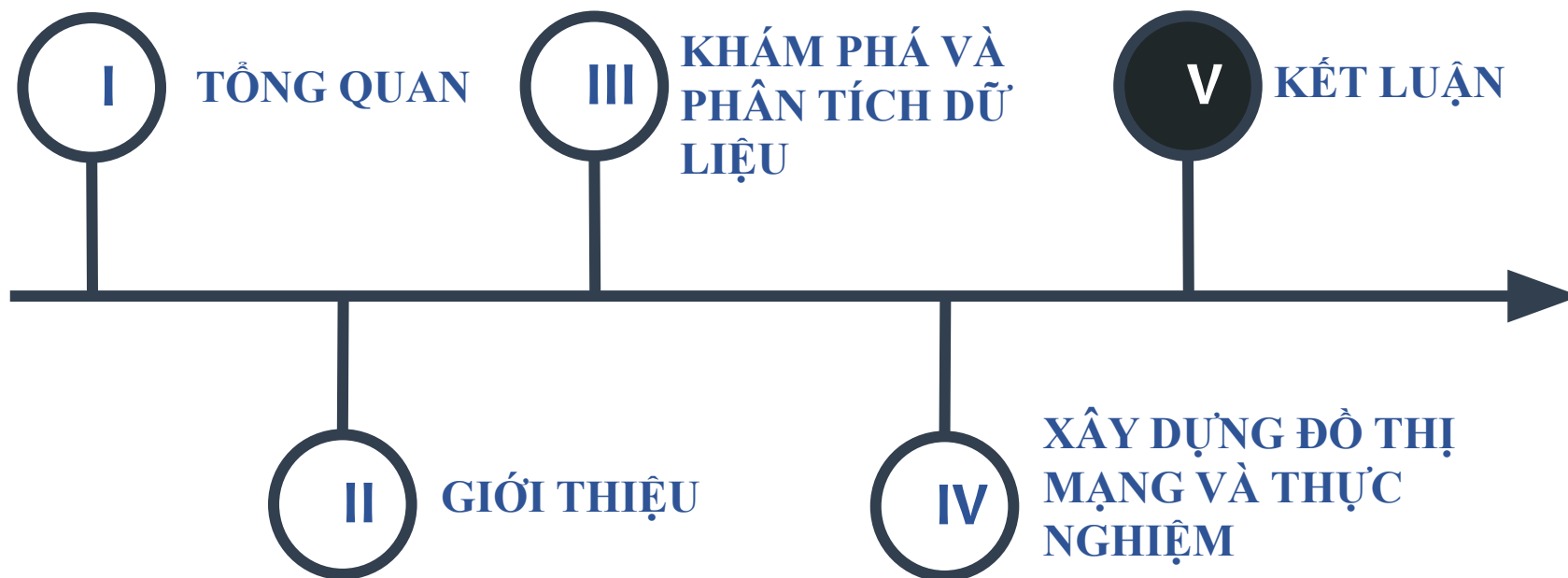
Bảng 4.3 So sánh kết quả của 2 thực nghiệm chính sau khi đã loại đặc trưng khoa và hệ đào tạo với mô hình Random Forest

Thực nghiệm	Accurac
SVC_1(offmxh)	89.91
SVC_off_khoa_hedt_1(offmxh)	89.65
SVC_2	89.10
SVC_off_khoa_hedt_2	90.46

Bảng 4.4 So sánh kết quả của 2 thực nghiệm chính sau khi đã loại đặc trưng khoa và hệ đào tạo với mô hình SVC

Nội dung	macro precision	macro recall	macro f1-score	weighted precision	weighted recall	weighted f1-score	accuracy
3nam_offmxh	63.63%	63.19%	63.40%	86.92%	87.19%	87.05%	<b>87.19%</b>
3nam_mhx	64.69%	66.01%	65.32%	88.73%	88.83%	88.76%	<b>88.83%</b>

Bảng 4.5 Bảng kết quả trên phương pháp Random Forest với 3 năm đầu tiên trong 2 trường hợp chưa và đã áp dụng đồ thị mạng xã hội





# KẾT LUẬN

05

**Đề tài:** Dự đoán xếp hạng tốt nghiệp của sinh viên UIT

- **Mục tiêu:** Dự đoán xếp loại tốt nghiệp dựa vào thành tích học tập và yếu tố hỗ trợ.
- **Phương pháp:**
  - Phân tích dữ liệu: Điểm trung bình (GPA), điểm rèn luyện (drl), yếu tố ngoại cảnh (vi phạm học vụ, điểm thi THPT).
  - Mô hình: Random Forest Classifier đạt **92.10% độ chính xác**, kết hợp với nhãn từ đồ thị mạng xã hội, tối đa lên tới **92.37%** khi loại bỏ yếu tố khoa và hệ đào tạo.
- **Kết quả:**
  - GPA và drl là yếu tố quan trọng nhất.
  - Đồ thị mạng xã hội cải thiện hiệu quả dự đoán, mở ra hướng phát triển mới.

**Hạn chế và Hướng phát triển:**

- Mở rộng dữ liệu, tích hợp yếu tố phi học thuật (ngoại khóa).
- Áp dụng thuật toán Deep Learning để tăng hiệu quả.
- Dự đoán sự tiến bộ qua các năm học để hỗ trợ sinh viên phát triển toàn diện.

**Ý nghĩa:**

- Ứng dụng công nghệ dữ liệu trong giáo dục.
- Nền tảng cho nghiên cứu nâng cao chất lượng giáo dục và hỗ trợ sinh viên.