

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI  
PHÂN HIỆU TẠI TP. HỒ CHÍ MINH  
BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐÒ ÁN TỐT NGHIỆP  
ĐỀ TÀI  
NGHIÊN CỨU THUẬT TOÁN RANDOM FOREST  
VÀ ỨNG DỤNG XÂY DỰNG  
TRANG WEB LUYỆN THI ĐẠI HỌC.**

Giảng viên hướng dẫn: ThS. TRẦN PHONG NHÃ

Sinh viên thực hiện: CHÂU QUẾ NHƠN

Lớp : CQ.61.CNTT

Khoa : 61

Tp. Hồ Chí Minh, 06 năm 2024

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI  
PHÂN HIỆU TẠI TP. HỒ CHÍ MINH  
BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP  
ĐỀ TÀI  
Nghiên cứu thuật toán RANDOM FOREST  
và ứng dụng xây dựng  
trang web luyện thi đại học.**

Giảng viên hướng dẫn: ThS. TRẦN PHONG NHÃ

Sinh viên thực hiện: CHÂU QUẾ NHƠN

Lớp : CQ.61.CNTT

Khoa : 61

Tp. Hồ Chí Minh, 06 năm 2024

**NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP**  
**BỘ MÔN: CÔNG NGHỆ THÔNG TIN**

-----\*\*\*-----

**Mã sinh viên: 6151071079 ..... Họ tên SV: Châu Quê Nhơn**

**Khóa: 61..... Lớp: CQ.61.CNTT**

**1. Tên đề tài**

Nghiên cứu thuật toán Random Forest và ứng dụng xây dựng trang web luyện thi đại học.

**2. Mục đích, yêu cầu**

- Xây dựng trang web luyện thi đại học với đầy đủ các tính năng để hỗ trợ người dùng ôn luyện, giải đề một cách hiệu quả.
- Huấn luyện và xây dựng mô hình máy học dự đoán điểm thi dựa trên lịch sử ôn luyện của người dùng.

**3. Nội dung và phạm vi đề tài**

- Đọc tài liệu mới của Laravel 11, NextJS 14.
- Nghiên cứu các tài liệu web khác như: caching, queue, OAuth2, websocket.
- Xây dựng trang web luyện thi với đầy đủ các tính năng đã đặt ra.
- Tìm hiểu về máy học, mô hình hồi quy trong máy học.
- Tìm hiểu về Decision Tree và thuật toán Random Forest.
- Nghiên cứu thư viện sklearn trong xây dựng mô hình máy học.
- Ứng dụng xây dựng mô hình dự đoán điểm thi.
- Nghiên cứu tích hợp mô hình vào trang web.

**4. Công nghệ, công cụ và ngôn ngữ lập trình**

- Công nghệ: Laravel 11, NextJS 14, Sckit Learn, Redis, WebSocket, MySQL, Flask.

- Công cụ: VSCode, DBeaver, Redis Insight.
- Ngôn ngữ lập trình: PHP, typescript, python.

## 5. Các kết quả chính dự kiến sẽ đạt được và ứng dụng

- Xây dựng trang web luyện thi với đầy đủ các tính năng đã đặt ra, đặc biệt là các tính năng thi thử, thi đấu realtime với những người dùng khác và thống kê học tập.
- Xây dựng được mô hình dự đoán điểm số dựa trên lịch sử ôn luyện trên web và tích hợp vào web.

## 6. Giáo viên và cán bộ hướng dẫn

Họ tên: Trần Phong Nhã

Đơn vị công tác: Bộ môn Công Nghệ Thông Tin, Trường Đại học Giao Thông Vận Tải phân hiệu Thành phố Hồ Chí Minh

Điện thoại: 0906761014

Email: tpnha@utc2.edu.vn

Ngày tháng 03 năm 2024

Trưởng BM Công nghệ Thông tin

Đã giao nhiệm vụ TKTN

Giáo viên hướng dẫn

ThS. Trần Phong Nhã

ThS. Trần Phong Nhã

Đã nhận nhiệm vụ TKTN

Sinh viên: Châu Quế Nhơn

Điện thoại: 0848611127

Ký tên:

Email: 6151071079@st.utc2.edu.vn

## LỜI CẢM ƠN

Lời đầu tiên, em xin gửi đến Quý thầy cô trong Bộ môn Công Nghệ Thông Tin, cùng Ban Giám Hiệu và tất cả Quý thầy cô trường Đại học Giao Thông Vận Tải phân hiệu tại Thành phố Hồ Chí Minh lời chúc sức khỏe và lòng biết ơn chân thành. Vậy là 4 năm Đại học đã đi đến năm cuối cùng. Nhờ vào sự quan tâm, dạy dỗ và chỉ bảo tận tình của thầy cô mà em đã hoàn thành tốt việc học trong 4 năm đại học và là tiền đề để em có thể hoàn thành đề tài tốt nghiệp này.

Trong quá trình thực hiện và hoàn thiện đề tài tốt nghiệp của mình, ngoài sự nỗ lực không ngừng học hỏi, tích lũy kiến thức trong 4 năm qua, đồng hành với chặn đường dài đó còn có sự quan tâm, giúp đỡ tận tình của các Quý thầy cô, bạn bè. Với tình cảm chân thành, cho phép em được một lần nữa bày tỏ lòng biết ơn sâu sắc đến tất cả mọi người đã tạo điều kiện và giúp đỡ em trong quá trình học tập, nghiên cứu.

Em muốn gửi lời cảm ơn đặc biệt đến thầy Trần Phong Nhã, là giảng viên hướng dẫn của em, người đã chỉ cho em hướng đi đúng đắn để có thể nghiên cứu một cách hiệu quả, truyền đạt cho em nhiều kiến thức bổ ích, những kinh nghiệm quý báu và cả những kỹ năng cần thiết để giúp em hoàn thành tốt đề tài này trong thời gian thực hiện đồ án.

Mặc dù em đã cố gắng hết sức mình để hoàn thành đề tài, nhưng sẽ không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến từ Quý thầy cô để em có thể bổ sung kiến thức thiếu sót của bản thân, rút ra nhiều bài học quý giá để giúp em phục vụ tốt hơn cho công việc thực tế sau này.

Lời nói sau cùng, một lần nữa em xin cảm ơn quý thầy cô trong thời gian 4 năm qua đã giúp đỡ em không những trong quá trình học tập mà còn những hoạt động khác, em cũng không biết nói gì hơn nữa ngoài những lời chúc gửi đến quý thầy cô, em xin kính chúc Quý thầy cô trong Bộ môn Công nghệ thông tin sẽ thật dồi dào sức khỏe, luôn vui vẻ và hạnh phúc, đạt được nhiều thành công trong cuộc sống cũng như sự nghiệp giảng dạy của mình.

Em xin chân thành cảm ơn!

## NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

*Tp. Hồ Chí Minh, ngày ..... tháng ..... năm 2024*

**Giáo viên hướng dẫn**

ThS. Trần Phong Nhã

## MỤC LỤC

LỜI CẢM ƠN.....	i
MỤC LỤC .....	iii
DANH MỤC CHỮ VIẾT TẮT .....	v
DANH MỤC HÌNH ẢNH.....	vi
DANH MỤC BẢNG BIÊU .....	viii
TỔNG QUAN ĐỀ TÀI .....	1
1. Giới thiệu về bối cảnh và cần thiết của đề tài.....	1
2. Mục tiêu của đề tài.....	2
3. Phạm vi và đối tượng nghiên cứu.....	3
4. Phương pháp nghiên cứu .....	4
CHƯƠNG 1. CƠ SỞ LÝ THUYẾT .....	5
1.1 Tổng quan về Laravel 11 .....	5
1.1.1 Giới thiệu.....	5
1.1.2 Vòng đời và kiến trúc của Laravel .....	6
1.1.3 Tính năng nổi bật .....	7
1.1.4 Ưu và nhược điểm của Laravel .....	11
1.2 NextJS 14 .....	12
1.2.1 Giới thiệu.....	12
1.2.2 Client Component .....	13
1.2.3 Ưu điểm, nhược điểm.....	14
1.3 Cơ sở dữ liệu .....	15
1.3.1 MySQL.....	15
1.3.2 Redis.....	15
1.4 Python và khoa học dữ liệu.....	16
1.4.1 Học máy trong Python .....	17
1.4.2 Thuật toán Random Forest .....	18
1.4.3 Thư viện Scikit learn .....	20
CHƯƠNG 2. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG .....	22
2.1 Thiết kế hệ thống .....	22
2.1.1 Sơ đồ Usecase .....	22
2.1.2 Sơ đồ Class.....	25

2.1.3 Sơ đồ tuần tự .....	25
2.2 Thiết kế cơ sở dữ liệu .....	30
2.2.1 Mô tả bài toán.....	30
2.2.2 Sơ đồ ER .....	32
2.2.3 Bảng thực thể cở sở dữ liệu.....	32
CHƯƠNG 3. XÂY DỰNG MÔ HÌNH DỰ ĐOÁN .....	36
3.1 Thu thập dữ liệu .....	36
3.1.1 Xác định các yếu tố ảnh hưởng đến điểm thi.....	36
3.1.2 Phân tích mối quan hệ giữa các yếu tố ảnh hưởng đến kết quả dự đoán .....	38
3.2 Tiền xử lý dữ liệu.....	40
3.3 Ứng dụng thuật toán Random Forest.....	43
3.3.1 Cây quyết định và rừng ngẫu nhiên .....	43
3.3.2 Ứng dụng xây dựng mô hình Random Forest.....	44
3.4 Xây dựng và đánh giá mô hình .....	48
3.4.1 Xây dựng và huấn luyện mô hình .....	48
3.4.2 Đánh giá hiệu suất của mô hình .....	50
CHƯƠNG 4. XÂY DỰNG CHƯƠNG TRÌNH.....	52
4.1 Thiết kế giao diện.....	52
4.1.1 Trang người dùng .....	52
4.1.2 Trang Admin .....	62
4.2 Tích hợp tính năng dự đoán .....	70
4.2.1 Triển khai API Prediction từ mô hình máy học .....	71
4.2.2 Thiết kế client.....	72
KẾT LUẬN & KIẾN NGHỊ.....	74
1. Kết quả đạt được .....	74
2. Hạn chế .....	74
3. Hướng phát triển .....	75
TÀI LIỆU THAM KHẢO .....	76

## **DANH MỤC CHỮ VIẾT TẮT**

<b>STT</b>	<b>Từ viết tắt</b>	<b>Điễn giải</b>	<b>Ghi chú</b>
1	THPT QG	Kỳ thi trung học phổ thông quốc gia	
2	BE	(Backend): Xử lý logic nghiệp vụ, truy xuất cơ sở dữ liệu và cung cấp API cho frontend.	
3	FE	(Frontend): Xử lý phía client	
4	ORM	Object-Relational Mapping	
5	PK	Primary Key	
6	FK	Foreign Key	
7	SSR	Server-Side Rendering	
8	CSR	Client-Side Rendering	
9	ML	Machine Learning	
10	API	Application Programming Interface	

## **DANH MỤC HÌNH ẢNH**

Hình 1. 1 Sơ đồ vòng đời của Laravel.....	6
Hình 2. 1 Sơ đồ Usecase tổng quát.....	22
Hình 2. 2 Sơ đồ phân rã Usecase Thảo luận.....	22
Hình 2. 3 Sơ đồ phân rã Usecase Tham gia thi đấu .....	23
Hình 2. 4 Sơ đồ phân rã Usecase Quản lý thông tin cá nhân .....	23
Hình 2. 5 Sơ đồ phân rã Usecase Quản lý câu hỏi .....	23
Hình 2. 6 Sơ đồ phân rã Usecase Quản lý bài thi.....	24
Hình 2. 7 Sơ đồ phân rã Usecase Quản lý phòng thi.....	24
Hình 2. 8 Sơ đồ phân rã Usecase Quản lý người dùng.....	24
Hình 2. 9 Sơ đồ Class .....	25
Hình 2. 10 Sơ đồ tuần tự thi thử .....	26
Hình 2. 11 Sơ đồ tuần tự đấu trường .....	27
Hình 2. 12 Sơ đồ tuần tự Thêm mới bản ghi .....	28
Hình 2. 13 Sơ đồ tuần tự Cập nhật bản ghi .....	29
Hình 2. 14 Sơ đồ tuần tự xóa bản ghi .....	30
Hình 2. 15 Sơ đồ ER.....	32
Hình 3. 1 Kiểm tra dữ liệu thiếu.....	41
Hình 3. 2 Mã hóa các biến phân loại dạng text sang number.....	42
Hình 3. 3 Chuẩn hóa và loại bỏ ngoại lệ .....	42
Hình 3. 4 Tách đặc trưng và biến mục tiêu.....	43
Hình 3. 5 Chia tập dữ liệu.....	43
Hình 3. 6 Ví dụ về cây quyết định và rừng ngẫu nhiên.....	44
Hình 3. 7 Ví dụ về một cây quyết định trong bài toán dự đoán điểm .....	47
Hình 3. 8 Huấn luyện mô hình với Random Forest trong Sklearn.....	49
Hình 3. 9 Tính toán giá trị để đánh giá mô hình .....	50
Hình 4. 1 Giao diện trang chủ.....	52
Hình 4. 2 Giao diện trang danh sách phòng thi .....	53
Hình 4. 3 Giao diện trang phòng chờ .....	53
Hình 4. 4 Giao diện trang thi đấu .....	54
Hình 4. 5 Demo chuyển câu hỏi realtime .....	54
Hình 4. 6 Giao diện trang kết quả thi đấu.....	54

Hình 4. 7 Hình ảnh demo về realtime notification nhắc nhở giờ thi .....	55
Hình 4. 8 Giao diện trang danh sách bài học.....	55
Hình 4. 9 Giao diện trang chi tiết bài học.....	56
Hình 4. 10 Giao diện trang thảo luận .....	56
Hình 4. 11 Giao diện trang chủ đề thảo luận.....	57
Hình 4. 12 Giao diện bình luận chủ đề thảo luận .....	57
Hình 4. 13 Giao diện trang bài tập .....	58
Hình 4. 14 Giao diện giải bài tập.....	58
Hình 4. 15 Giao diện trang đề thi .....	59
Hình 4. 16 Giao diện trang giải đề .....	59
Hình 4. 17 Giao diện thông tin người dùng.....	60
Hình 4. 18 Giao diện mục tiêu học tập.....	60
Hình 4. 19 Giao diện trang thống kê .....	61
Hình 4. 20 Giao diện thống kê bảng biểu.....	61
Hình 4. 21 Giao diện trang lịch sử làm bài.....	62
Hình 4. 22 Giao diện trang xác thực.....	62
Hình 4. 23 Giao diện trang Dashboard .....	63
Hình 4. 24 Giao diện trang Dashboard 2 .....	64
Hình 4. 25 Giao diện trang quản lý chương .....	64
Hình 4. 26 Giao diện trang quản lý chương .....	65
Hình 4. 27 Giao diện trang quản lý bài học .....	65
Hình 4. 28 Giao diện trang quản lý câu hỏi.....	66
Hình 4. 29 Giao diện trang quản lý đề thi .....	66
Hình 4. 30 Giao diện thêm/chỉnh sửa đề thi .....	67
Hình 4. 31 Giao diện trang quản lý bài tập.....	67
Hình 4. 32 Giao diện trang quản lý phòng thi đầu .....	68
Hình 4. 33 Giao diện thêm/sửa phòng thi.....	68
Hình 4. 34 Giao diện trang quản lý chủ đề thảo luận .....	69
Hình 4. 35 Giao diện trang quản lý người dùng.....	70
Hình 4. 36 Hình ảnh lưu mô hình vào pickle file .....	71
Hình 4. 37 Demo về api dự đoán từ mô hình .....	72
Hình 4. 38 Demo chức năng dự đoán .....	73

## **DANH MỤC BẢNG BIỂU**

Bảng 2. 1 Bảng thực thể User.....	32
Bảng 2. 2 Bảng thực thể Subject .....	33
Bảng 2. 3 Bảng thực thể Chapter.....	33
Bảng 2. 4 Bảng thực thể Practice .....	33
Bảng 2. 5 Bảng thực thể Lesson.....	33
Bảng 2. 6 Bảng thực thể Question.....	33
Bảng 2. 7 Bảng thực thể Exam.....	34
Bảng 2. 8 Bảng thực thể Topic .....	34
Bảng 2. 9 Bảng thực thể TopicComment .....	34
Bảng 2. 10 Bảng thực thể Arena .....	34
Bảng 2. 11 Bảng thực thể UserTarget .....	35
Bảng 2. 12 Bảng thực thể History .....	35
Bảng 2. 13 Bảng thực thể DayStatistics .....	35
Bảng 3. 1 Bảng các yếu tố cá nhân quyết định điểm thi .....	36
Bảng 3. 2 Bảng các yếu tố web quyết định điểm thi .....	37
Bảng 3. 3 Ví dụ về tập dữ liệu.....	45

## TỔNG QUAN ĐỀ TÀI

### 1. Giới thiệu về bối cảnh và cần thiết của đề tài.

Kỳ thi trung học phổ thông quốc gia có thể nói là cột mốc quan trọng, có tính quyết định cao nhất trong cuộc đời học sinh. Nó là kết quả của cả một quãng đường 12 năm gian nan, vất vả và đầy nỗ lực của học sinh trong hệ thống giáo dục phổ thông. Với mỗi học sinh, kỳ thi này không chỉ đánh dấu sự hoàn thành của giai đoạn trung học mà còn là cơ hội để họ chứng tỏ khả năng và kiến thức mà mình đã tích lũy được. Tầm quan trọng của kỳ thi trung học phổ thông quốc gia nằm ở việc nó không chỉ đánh giá hiệu suất học tập của học sinh mà còn định hình tương lai của họ. Kết quả của kỳ thi này có thể ảnh hưởng trực tiếp đến cơ hội tiếp cận giáo dục đại học và sự lựa chọn nghề nghiệp sau này của học sinh. Điều này làm cho áp lực và trách nhiệm của kỳ thi trở nên nặng nề và quan trọng hơn bao giờ hết. Ngoài ra, kỳ thi trung học phổ thông quốc gia cũng là công cụ đánh giá chất lượng của hệ thống giáo dục quốc gia. Kết quả của kỳ thi này phản ánh sự hiệu quả của quá trình dạy và học ở cấp trung học, từ chất lượng giáo viên đến nội dung chương trình học. Đồng thời, nó cũng là một phần trong việc đánh giá và cải thiện chất lượng giáo dục toàn diện của đất nước.

Việc ôn luyện có phương pháp và hiệu quả đóng vai trò quan trọng trong việc chuẩn bị cho kỳ thi trung học phổ thông quốc gia. Phương pháp ôn luyện hiệu quả giúp học sinh tối ưu hóa thời gian và năng lượng để nắm vững kiến thức cũng như phát triển kỹ năng cần thiết cho kỳ thi. Việc áp dụng phương pháp ôn luyện có hệ thống giúp học sinh xác định các mục tiêu cụ thể và lập kế hoạch hợp lý cho quá trình ôn tập. Bằng cách chia nhỏ kiến thức thành các đơn vị nhỏ hơn và ưu tiên làm việc trên những phần cần thiết nhất, học sinh có thể tập trung và hiệu quả hóa quá trình học tập. Đồng thời, việc sử dụng các phương tiện và tài liệu ôn luyện chất lượng cũng đóng vai trò quan trọng. Các sách giáo khoa, tài liệu ôn thi, bài tập mẫu và đề thi thử không chỉ giúp học sinh làm quen với định dạng của kỳ thi mà còn giúp họ củng cố kiến thức và rèn luyện kỹ năng làm bài thi một cách tự tin. Việc tìm kiếm sự hỗ trợ từ giáo viên, gia đình và bạn bè cũng là một phần quan trọng của quá trình ôn luyện. Sự hỗ trợ và động viên từ người thân và cộng đồng giúp học sinh vượt qua những khó khăn trong quá trình học tập và tạo ra một môi trường tích cực để phát triển.

Trong thời đại phát triển của công nghệ thông tin hiện nay, việc sử dụng các nền tảng công nghệ như web/app để học tập, ôn luyện mang lại nhiều hiệu quả cho mỗi

người. Sự tiện lợi và linh hoạt, người dùng có thể tự do học, luyện thi mọi lúc mọi nơi. Ngoài ra, việc truy cập vào nguồn tài nguyên phong phú từ internet giúp họ tiếp cận các kiến thức mới một cách dễ dàng và đa dạng hơn. Việc tham gia vào các hoạt động tương tác và hợp tác qua các nền tảng giáo dục từ xa cũng giúp họ rèn luyện kỹ năng giao tiếp và làm việc nhóm. Sử dụng công nghệ thông tin cũng giúp họ nhận được phản hồi và đánh giá nhanh chóng, từ đó điều chỉnh quá trình học tập một cách hiệu quả hơn. Cuối cùng, việc khuyến khích sự sáng tạo và khám phá qua các ứng dụng và công cụ giúp họ phát triển tư duy sáng tạo và thúc đẩy quá trình học tập. Từ đó, việc xây dựng trang web hỗ trợ học tập và luyện thi đại học có thể hỗ trợ người dùng dễ dàng tổng hợp kiến thức, thử sức với các dạng đề thi đại học là vô cùng cần thiết. Trang web này sẽ có thể giúp ích rất nhiều cho những sĩ tử sắp bước vào kì thi quan trọng nhất hoặc đối với những học sinh muốn ôn luyện sớm hơn để chuẩn bị kỹ cho kì thi này.

## **2. Mục tiêu của đề tài.**

Mục tiêu dự kiến của đề tài này sẽ là xây dựng được một trang web với các yêu cầu:

- Về giao diện: cấu trúc trang web chuẩn SEO, giao diện hiện đại đẹp mắt dễ sử dụng. Hỗ trợ responsive trên nhiều thiết bị khác nhau.
- Về chức năng:
  - + Xác thực: đăng nhập/đăng ký tài khoản.
  - + Phân quyền: Sẽ có 2 loại tài khoản: học sinh và giáo viên. Ngoài ra còn có vai trò quản trị viên/người kiểm duyệt.
    - + Tính năng hỏi – đáp: người dùng có thể tạo các chủ đề thảo luận, đặt các câu hỏi để mọi người có thể giúp đỡ, trao đổi, góp ý.
    - + Cung cấp lý thuyết, bài giải SGK: Lý thuyết và các bài tập (có đáp án) theo chương trình học chuẩn.
    - + Luyện tập giải đề: Học sinh có thể luyện tập giải các loại đề thi mẫu dưới hình thức trắc nghiệm trong thời gian quy định. Chức năng giải đề theo lộ trình học chuẩn để nâng cao hiệu quả ôn luyện.
      - + Tính năng đấu trường: người dùng có thể tạo phòng để những người tham gia khác có thể cùng tham gia thi cùng lúc với nhau (realtime).
      - + Thống kê tình hình ôn luyện của từng thí sinh, giúp thí sinh có thể nắm bắt được quá trình ôn luyện của bản thân.

Ngoài ra, mục tiêu quan trọng nhất là có thể xây dựng được một mô hình học máy đơn giản để dự đoán điểm số, hiệu suất học tập, hoặc khả năng làm bài thi của người dùng dựa trên dữ liệu lịch sử, đưa ra đề xuất những đề thi có thể cải thiện điểm yếu cho thí sinh, mang lại hiệu quả ôn luyện tốt nhất cho thí sinh.

### **3. Phạm vi và đối tượng nghiên cứu.**

#### **\* Phạm vi nghiên cứu:**

+ Yêu cầu của các sĩ tử đang ôn luyện: Đề tài này sẽ tập trung vào nghiên cứu và hiểu rõ những yêu cầu cần thiết của các sĩ tử đang ôn luyện để chuẩn bị cho kỳ thi trung học phổ thông quốc gia. Qua đó, trang web sẽ được thiết kế và phát triển để đáp ứng những yêu cầu này một cách hiệu quả nhất, bao gồm việc cung cấp các tài liệu ôn tập, bài kiểm tra mẫu, và đề thi thử.

+ Nghiên cứu về kỹ thuật lập trình: Đồng thời, nghiên cứu cũng sẽ tập trung vào các vấn đề kỹ thuật lập trình liên quan đến việc phát triển trang web. Điều này bao gồm việc tìm hiểu và áp dụng các công nghệ lập trình web như PHP Laravel và Next.js để xây dựng một trang web mạnh mẽ, linh hoạt và dễ bảo trì. Các nguyên tắc thiết kế và triển khai cũng sẽ được tìm hiểu để đảm bảo trải nghiệm người dùng tốt nhất.

+ Nghiên cứu về mô hình học máy có giám sát: Một phần quan trọng của đề tài sẽ là nghiên cứu về mô hình học máy có giám sát để xây dựng, huấn luyện và triển khai một hệ thống có khả năng dự đoán điểm số và hiệu suất học tập của các sĩ tử. Điều này đòi hỏi việc thu thập và tiền xử lý dữ liệu mẫu một cách kỹ lưỡng, sau đó áp dụng các mô hình học máy như hồi quy tuyến tính hoặc cây quyết định để tạo ra dự đoán chính xác và hữu ích cho người dùng.

#### **\* Đối tượng nghiên cứu:**

Đối tượng nghiên cứu chính của trang web luyện thi đại học chắc chắn sẽ tập trung vào các sĩ tử đang chuẩn thi thi trung học phổ thông quốc gia hoặc những học sinh đang học cấp trung học phổ thông muốn luyện thi đại học sớm. Bằng cách nghiên cứu các cộng đồng học sinh luyện thi đại học, các trang web luyện thi khác trên internet sẽ mang lại một cái nhìn toàn diện về nhu cầu và mong muốn của đối tượng này. Thông qua việc thu thập dữ liệu từ người dùng, bao gồm thông tin cá nhân, kết quả kiểm tra trước đó, hoạt động học tập và một số yếu tố khác, trang web có thể tạo ra những giải pháp tùy chỉnh và hữu ích nhằm nâng cao hiệu suất học tập của học sinh.

#### **4. Phương pháp nghiên cứu**

Các phương pháp nghiên cứu được sử dụng để thu thập và xử lý dữ liệu là:

- Khảo sát các yêu cầu của thí sinh chuẩn bị thi THPTQG: Khảo sát, thu thập thông tin về các yêu cầu, nhu cầu và mong muốn của thí sinh đang chuẩn bị thi Trung học Phổ thông Quốc gia (THPTQG). Điều này giúp định hình các tính năng và tài nguyên cần thiết trên trang web luyện thi.

- Tham khảo các trang web luyện thi khác: Tìm kiếm và tham khảo các trang web luyện thi khác trên internet để thu thập thông tin về các phương pháp và tài liệu ôn thi, cũng như để hiểu thêm về các nhu cầu và mong muốn của người dùng trong lĩnh vực này.

- Khảo sát các cộng đồng học sinh THPT: Tiến hành nghiên cứu và khảo sát các cộng đồng học sinh THPT trên các diễn đàn, mạng xã hội hoặc các nhóm chuyên ngành để hiểu rõ hơn về các thách thức và nhu cầu của họ trong quá trình luyện thi.

- Thu thập dữ liệu về các câu hỏi, đề thi chuẩn theo chương trình của Bộ Giáo Dục: Thu thập dữ liệu từ các nguồn đáng tin cậy về các câu hỏi và đề thi chuẩn theo chương trình của Bộ Giáo Dục, để xây dựng cơ sở dữ liệu đa dạng và chất lượng để phục vụ cho mục đích ôn thi.

- Nghiên cứu về các hàm trong thư viện sklearn của Python để xử lý dữ liệu cho mô hình máy học: Tiến hành nghiên cứu và tìm hiểu về các hàm và phương pháp trong thư viện scikit-learn của Python để tiền xử lý và chuẩn bị dữ liệu cho việc huấn luyện các mô hình máy học. Điều này bao gồm các phương pháp chuẩn hóa, mã hóa dữ liệu, và xử lý dữ liệu thiếu.

## CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

### 1.1 Tổng quan về Laravel 11

#### 1.1.1 Giới thiệu

*PHP là gì?*

PHP (Hypertext Preprocessor) [8] là một ngôn ngữ lập trình kịch bản phía máy chủ được thiết kế chủ yếu để phát triển web. Được tạo ra vào năm 1994 bởi Rasmus Lerdorf, PHP đã trở thành một trong những ngôn ngữ phổ biến nhất cho lập trình web nhờ vào sự dễ học, cú pháp linh hoạt và tích hợp mạnh mẽ với HTML. PHP cho phép các nhà phát triển tạo ra các trang web động và tương tác bằng cách xử lý dữ liệu từ biểu mẫu, quản lý phiên làm việc, và kết nối với các cơ sở dữ liệu như MySQL. Với cộng đồng phát triển lớn và sự hỗ trợ rộng rãi, PHP tiếp tục là lựa chọn ưa thích cho việc xây dựng các ứng dụng web từ nhỏ đến lớn.

*Laravel là gì?*

Laravel [17] là một framework phát triển ứng dụng web được viết bằng ngôn ngữ lập trình PHP. Nó cung cấp một cách tiếp cận đơn giản và linh hoạt cho việc xây dựng các ứng dụng web phức tạp. Laravel đi kèm với nhiều tính năng mạnh mẽ như hệ thống routing, ORM (Object-Relational Mapping), middleware, bảo mật, quản lý cơ sở dữ liệu và nhiều tính năng khác giúp người phát triển tạo ra các ứng dụng web chất lượng cao và dễ bảo trì. Điểm mạnh của Laravel nằm ở việc cung cấp cú pháp rõ ràng và cấu trúc dễ hiểu, làm cho việc phát triển ứng dụng trở nên nhanh chóng và hiệu quả.

*Tại sao phải dùng Laravel?*

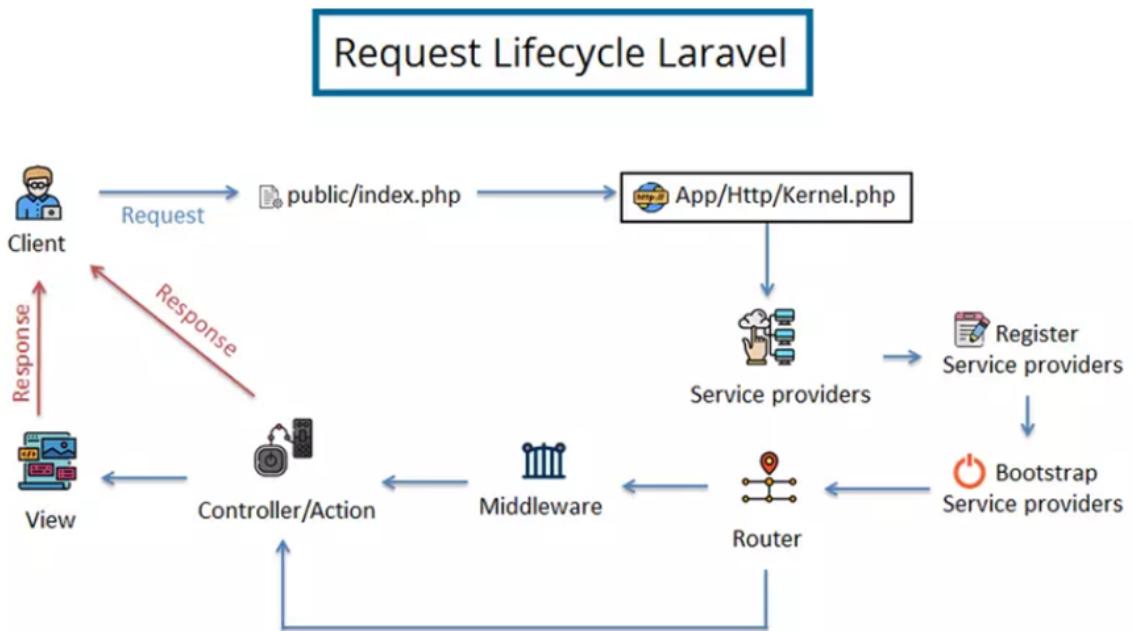
Laravel là một framework phát triển web PHP mạnh mẽ và phổ biến, được thiết kế để mang lại trải nghiệm phát triển dễ dàng và hiệu quả. Một số lý do mà nhiều nhà phát triển lựa chọn Laravel là do tính nhanh chóng và thuận tiện trong quá trình phát triển ứng dụng. Với Artisan Console, một công cụ dòng lệnh mạnh mẽ, nhà phát triển có thể tự động hóa nhiều công việc lặp lại, giảm thời gian và công sức đặc biệt là trong quá trình lặp lại. Điều này giúp tối ưu hóa quá trình phát triển và giảm bớt sự lặp lại không cần thiết.

Mặc dù PHP hiện nay đã dần trở nên lạc hậu và chậm cập nhật so với những công nghệ web phía server khác, nhưng với vị thế đã được xây dựng từ lâu, nó vẫn chiếm một phần lớn trong thị phần web hiện nay trên toàn thế giới. Trong cộng đồng lập trình viên PHP, có câu đùa rằng "PHP is dead, but Laravel forever alive". Điều này có phần đúng

trong khi PHP cập nhật quá chậm, thì tốc độ update và phát triển của Laravel vẫn có thể khiến các lập trình viên mới vào nghề “không kịp thích nghi”. Laravel hiện nay cung cấp hầu như mọi tính năng mà các framework hiện đại yêu cầu, như xử lý dữ liệu lớn, bảo mật, khả năng mở rộng và tích hợp tương thích cao với các dịch vụ bên thứ ba. Laravel sẽ release version mới vào quý 1 hàng năm, tính đến hiện tại đã là phiên bản Laravel 11.

Laravel 11 chạy trên phiên bản PHP mới ( $> 8.2$ ) đã như một cuộc cải cách lớn của Laravel Team để bắt kịp cuộc đua công nghệ với những Framework JS mạnh mẽ hiện nay như NestJS, NextJS, ... Trong phiên bản mới này, cấu trúc thư mục đã được rút gọn, tối ưu hơn so với những phiên bản trước đó. Ngoài ra, sự ra đời của Reverb đã giúp Laravel có một bước tiến lớn hơn khi lần đầu tiên, PHP có thể tự dùng để xây dựng một ứng dụng realtime mà không cần dùng đến bên thứ 3.

### 1.1.2 Vòng đời và kiến trúc của Laravel



Hình 1. 1 Sơ đồ vòng đời của Laravel

Quá trình vòng đời hoạt động của Laravel bắt đầu khi một yêu cầu (request) được gửi từ phía client. Đối tượng nhắm đến của mọi yêu cầu từ client chính là file public/index.php, một tập tin chính của ứng dụng. Màu xanh của mũi tên biểu thị hành động di chuyển từ client đến đích đến này.

Mặc dù số lượng mã nguồn trong file index.php không nhiều, nhưng nó đóng vai trò quan trọng như là khởi nguyên cho toàn bộ framework Laravel. File này thực hiện các công việc cơ bản như bao gồm file autoload.php để load các class cần thiết, sau đó

khởi tạo và chạy ứng dụng Laravel. Điều này bao gồm việc xử lý middleware, định tuyến (routing), và gọi controllers tương ứng với request.

Từ file index.php, quá trình xử lý được chuyển tiếp qua các giai đoạn tiếp theo của vòng đời hoạt động của Laravel, như chạy middleware, xử lý route, thực hiện controllers, và cuối cùng là gửi response về phía client. Dù đơn giản nhưng file index.php tạo nên một cơ sở hạ tầng vững chắc, làm nền tảng cho tính linh hoạt và hiệu suất của toàn bộ framework Laravel.

### **1.1.3 Tính năng nổi bật**

#### **1.1.3.1 Eloquent ORM**

Laravel cung cấp một bộ ánh xạ cơ sở dữ liệu hiện đại và vô cùng mạnh mẽ, được gọi là Eloquent ORM. Eloquent ORM (Object-Relational Mapping) là một trong những tính năng nổi bật nhất, cung cấp một cách thức đơn giản và thanh lịch để tương tác với cơ sở dữ liệu. Eloquent ORM cho phép ánh xạ các bảng trong cơ sở dữ liệu thành các đối tượng PHP (model), và các bản ghi (record) trong bảng thành các thể hiện (instance) của model đó. Điều này giúp cho việc thao tác với cơ sở dữ liệu trở nên trực quan và dễ dàng hơn, chỉ cần có hiểu biết về bản chất của truy vấn cơ sở dữ liệu thì có thể sử dụng ORM một cách dễ dàng, nhanh chóng, đồng thời tận dụng được các tính năng của lập trình hướng đối tượng trong PHP thông qua các class model.

Các tính năng chính được hỗ trợ mạnh mẽ bởi Eloquent ORM có thể kể đến như sau:

- Xem, thêm, sửa và xóa các bảng ghi (CRUD): Eloquent cung cấp các phương thức đơn giản và nhất quán để thực hiện các thao tác cơ bản với cơ sở dữ liệu, giúp giảm thiểu lượng code boilerplate và tăng tính dễ đọc của mã nguồn.
- Mối quan hệ (Relationships): Laravel là một trong những framework hỗ trợ relationship một cách đơn giản, dễ hiểu và dễ sử dụng nhất. Eloquent hỗ trợ định nghĩa và quản lý các mối quan hệ giữa các bảng (ví dụ: một-nhiều, nhiều-nhiều) một cách dễ dàng. Nhờ đó, có thể truy xuất và thao tác với dữ liệu liên quan giữa các bảng một cách thuận tiện.
- Kiểu collections: Kết quả truy vấn từ cơ sở dữ liệu được trả về dưới dạng các tập hợp Eloquent (collections). Collections cung cấp nhiều phương thức tiện lợi để lọc, sắp xếp, biến đổi và thao tác với dữ liệu một cách linh hoạt.

- Xử lý sự kiện (Events): Eloquent cho phép "lắng nghe" và phản ứng với các sự kiện xảy ra trong quá trình thao tác với model, chẳng hạn như khi một model được tạo, cập nhật hoặc xóa. Tính năng này giúp thực hiện các tác vụ phụ trợ một cách dễ dàng và tập trung.
- Bảo mật: Eloquent tự động bảo vệ ứng dụng khỏi các cuộc tấn công SQL Injection bằng cách sử dụng prepared statements. Prepared statements là một kỹ thuật giúp ngăn chặn việc chèn các đoạn mã độc vào câu truy vấn SQL, đảm bảo tính an toàn cho dữ liệu.

Ngoài ra, Eloquent ORM còn hỗ trợ phát triển web service với chuẩn Restful API một cách dễ dàng. API Resource là một tính năng của Laravel giúp cho việc xây dựng API trở nên dễ dàng và nhất quán hơn. API Resource cho phép chuyển đổi models và collections thành các cấu trúc JSON theo một cách định nghĩa trước, giúp cho việc trả về dữ liệu từ API trở nên dễ quản lý và bảo trì hơn.

#### \* Xử lý dữ liệu lớn

Một tính năng ưu Việt nữa, Laravel còn cung cấp một số phương thức và công cụ để xử lý dữ liệu lớn (big data), giúp cho quá trình làm việc với các tập dữ liệu có kích thước lớn trở nên dễ dàng và hiệu quả hơn. Các phương thức xử lý dữ liệu lớn được hỗ trợ như: Phương thức chunk cho phép chia nhỏ một truy vấn lớn thành nhiều phần nhỏ hơn để xử lý, tránh tình trạng tràn bộ nhớ khi làm việc với dữ liệu lớn. Phương thức cursor cũng cho phép lặp qua một tập kết quả lớn mà không cần tải toàn bộ dữ liệu vào bộ nhớ. Lazy Collections là một tính năng mới trong Laravel, cho phép làm việc với các tập dữ liệu lớn một cách hiệu quả hơn bằng cách chỉ tải dữ liệu khi cần thiết.

Ngoài ra, các công cụ xử lý dữ liệu lớn khác như: queue cho phép đưa các tác vụ xử lý dữ liệu nặng vào hàng đợi để xử lý nền, giúp giảm tải cho máy chủ web và tăng tốc độ xử lý. Laravel Jobs cung cấp một cách để định nghĩa các tác vụ xử lý dữ liệu một cách có cấu trúc và dễ quản lý. Bạn có thể đưa các jobs vào queue để xử lý nền. Events: Laravel Events cho phép bạn tạo ra các sự kiện khi có dữ liệu mới được thêm vào hoặc cập nhật, từ đó kích hoạt các tác vụ xử lý dữ liệu khác.

#### *1.1.3.2 Broastcasting*

##### \* Giới thiệu về Broastcasting

Broadcasting là một tính năng mạnh mẽ trong Laravel cho phép truyền tải các sự kiện trong thời gian thực từ phía máy chủ (backend) đến phía máy khách (frontend).

Điều này giúp tạo ra các ứng dụng web có khả năng cập nhật thông tin tức thời mà không cần phải tải lại trang hoặc gửi yêu cầu liên tục từ phía client [15].

**Cơ chế hoạt động:** Khi có một sự kiện xảy ra trong ứng dụng (ví dụ: người dùng đăng bài mới, có thông báo mới,...), một đối tượng Event sẽ được kích hoạt. Sự kiện này sẽ được gửi đến một hoặc nhiều kênh (channel) được định nghĩa trước. Các kênh này có thể là công khai (public) hoặc riêng tư (private), tùy thuộc vào yêu cầu của ứng dụng. Laravel sử dụng các trình điều khiển để gửi sự kiện đến các kênh. Các trình điều khiển phổ biến bao gồm Pusher, Redis, và Log. Phía máy khách sẽ lắng nghe các sự kiện trên các kênh mà họ đã đăng ký. Khi có sự kiện mới, trình duyệt sẽ nhận được thông báo và cập nhật giao diện tương ứng.

Broadcasting được sử dụng rộng rãi trong việc xây dựng các tính năng realtime cho ứng dụng web, chẳng hạn như:

- Thông báo: Gửi thông báo tức thời đến người dùng khi có sự kiện mới (ví dụ: tin nhắn mới, bình luận mới, yêu cầu kết bạn,...).
- Trò chuyện: Xây dựng các ứng dụng trò chuyện trực tuyến, nơi người dùng có thể gửi và nhận tin nhắn tức thời.
- Theo dõi trạng thái: Cập nhật trạng thái của một đối tượng trong thời gian thực (ví dụ: trạng thái đơn hàng, tiến độ công việc,...).
- Bảng tin trực tiếp: Hiển thị các cập nhật mới nhất về một chủ đề (ví dụ: tỉ số trận đấu, giá cổ phiếu,...) mà không cần phải tải lại trang.

#### \* Reverb trong Laravel 11

Reverb là một tính năng mới đáng chú ý được giới thiệu trong Laravel 11, cho phép xây dựng các ứng dụng realtime hoàn toàn bằng PHP mà không cần phụ thuộc vào các công nghệ bên thứ ba như Pusher hay Ably. Reverb hoạt động dựa trên giao thức WebSocket và Server-Sent Events (SSE), mang đến khả năng cập nhật dữ liệu tức thời cho ứng dụng web mà không cần yêu cầu từ phía client. Một số điểm mà nhà Laravel đã chính thức giới thiệu về Reverb với những ưu điểm như sau:

- Đơn giản và dễ sử dụng: Reverb được tích hợp trực tiếp vào Laravel, giúp việc cài đặt và sử dụng trở nên dễ dàng hơn so với việc tích hợp các dịch vụ bên ngoài.

- Hiệu suất cao: Reverb tận dụng sức mạnh của Swoole, một extension PHP hiệu suất cao, để xử lý các kết nối WebSocket và SSE, đảm bảo khả năng mở rộng và đáp ứng tốt với lượng lớn người dùng.
- Linh hoạt: Reverb cung cấp các API đơn giản và linh hoạt để gửi và nhận sự kiện, cho phép tùy chỉnh theo nhu cầu của ứng dụng.
- Bảo mật: Reverb hỗ trợ xác thực và ủy quyền người dùng, đảm bảo tính bảo mật cho các kết nối và dữ liệu truyền đi.
- Giảm sự phụ thuộc: Việc sử dụng Reverb giúp giảm sự phụ thuộc vào các dịch vụ bên thứ ba, giúp tiết kiệm chi phí và tăng tính kiểm soát đối với ứng dụng.

Với Reverb, việc xây dựng các ứng dụng realtime với Laravel trở nên dễ dàng và hiệu quả hơn bao giờ hết. Tính năng này hứa hẹn sẽ mở ra nhiều cơ hội mới cho các nhà phát triển Laravel trong việc tạo ra các ứng dụng web hiện đại và hấp dẫn. Tuy nhiên, với trình độ còn hạn chế cũng như thời gian làm đồ án có giới hạn, em vẫn chưa thể ứng dụng Reverb vào đồ án của mình mà vẫn đang phải sử dụng socket.io để tạo realtime cho trang web.

#### *1.1.3.3 Queue & Task Scheduling*

Queue (hàng đợi) và Task Scheduling (lập lịch tác vụ) là hai tính năng đặc biệt làm nên sự khác biệt của Laravel. Đây là 2 tính năng rất quan trọng trong Laravel, giúp cải thiện hiệu suất và trải nghiệm người dùng của ứng dụng web.

##### *\* Queue*

Queue là một cơ chế xử lý tác vụ nền (background jobs) trong Laravel, cho phép bạn trì hoãn việc thực thi các tác vụ trong thời gian hoặc cần xử lý bất đồng bộ. Thay vì thực hiện các tác vụ này ngay lập tức trong luồng xử lý chính của ứng dụng, bạn có thể đưa chúng vào một hàng đợi (queue) để xử lý sau đó.

Lợi ích của việc sử dụng queue:

- Cải thiện tốc độ phản hồi, các tác vụ nặng được xử lý nền, giúp giảm tải cho máy chủ web và tăng tốc độ phản hồi cho người dùng.
- Hỗ trợ xử lý bất đồng bộ, các tác vụ có thể được thực hiện độc lập với luồng xử lý chính, giúp ứng dụng hoạt động trơn tru hơn.
- Dễ dàng mở rộng khả năng xử lý của ứng dụng bằng cách tăng số lượng worker.
- Cung cấp các cơ chế để xử lý lỗi trong quá trình thực thi job, đảm bảo tính ổn định của ứng dụng.

- Hỗ trợ nhiều driver khác nhau như: database, redis, Amazon SQS, ...

\* *Task Scheduling*

Task Scheduling là một tính năng mạnh mẽ trong Laravel cho phép tự động hóa việc thực thi các tác vụ định kỳ hoặc theo lịch trình cụ thể. Bạn có thể sử dụng Task Scheduling để thực hiện nhiều loại tác vụ khác nhau, từ việc gửi email nhắc nhở, tạo báo cáo, sao lưu dữ liệu, đến việc thực hiện các tác vụ bảo trì hệ thống.

Laravel Task Scheduling hoạt động dựa trên cơ chế cron jobs của hệ điều hành. Cron job là một tiện ích cho phép lên lịch thực thi các lệnh hoặc script tại các thời điểm cụ thể. Laravel cung cấp một giao diện thân thiện để bạn định nghĩa các tác vụ và lịch trình của chúng trong file app\Console\Kernel.php. Khi đến thời điểm được lên lịch, Laravel sẽ tự động gọi cron job để thực thi tác vụ tương ứng.

Ưu điểm của Task Scheduling:

- Tự động hóa: Giúp tự động hóa các tác vụ định kỳ, giảm thiểu sự can thiệp thủ công.
- Cải thiện hiệu suất: Thực hiện các tác vụ nền, giúp giảm tải cho máy chủ web và tăng tốc độ phản hồi của ứng dụng.
- Linh hoạt: Hỗ trợ nhiều loại lịch trình khác nhau, từ đơn giản đến phức tạp.
- Đáng tin cậy: Sử dụng cơ chế cron job của hệ điều hành, đảm bảo tính ổn định và độ tin cậy cao.

Task Scheduling là một công cụ vô cùng hữu ích trong Laravel, giúp tự động hóa các tác vụ và cải thiện hiệu suất của ứng dụng. Với Task Scheduling, chúng ta có thể tập trung vào việc phát triển các tính năng chính của ứng dụng mà không cần lo lắng về việc thực hiện các tác vụ định kỳ một cách thủ công. Trong đề tài này, Task Scheduling được dùng để tạo tự động tạo thống kê theo ngày cho từng học sinh vào lúc 0h đêm mỗi ngày, giúp việc thống kê hiệu quả hơn mà không bị ảnh hưởng đến hiệu suất của trang web.

#### **1.1.4 Ưu và nhược điểm của Laravel**

\* Ưu điểm:

- Tài liệu chính thống dễ hiểu, có example code cho từng nội dung. Một trong số ít Framework có tài liệu từ trang chủ dễ hiểu nhất
- Có một cộng đồng rộng lớn với những chuyên gia hỗ trợ, những bài viết chia sẻ, hướng dẫn cho người dùng.

- Mã nguồn mở, cho phép lập trình viên phát triển các package để xây dựng ứng dụng web và tái sử dụng cũng như chia sẻ đến cộng đồng một cách dễ dàng thông qua composer (tương tự npm trong các framework js hiện đại).
- Tương tác với cơ sở dữ liệu một cách an toàn và dễ dàng thông qua Eloquent ORM. Eloquent cung cấp một bộ ORM mạnh mẽ, có thể dễ dàng truy vấn, đồng thời đã được thiết lập sẵn các biện pháp chống các lỗ hổng như injection. Ngoài ra Eloquent ORM còn cung cấp phương thức Chunk hỗ trợ xử lý dữ liệu lớn.
- Bảo mật cao với các phương thức bảo mật hiện đại như OAuth2, JWT,... thông qua các package mạnh mẽ như Passport, Sanctum, ... Và cơ chế phân quyền Spatie Permissions.
- Quản lý source code dễ dàng bằng các câu lệnh artisan, cho phép tạo file, cấu hình một cách tự động thông qua terminal.
- Khả năng tương thích với các thư viện frontend giúp dễ dàng xây dựng giao diện người dùng. Hiện tại, Laravel đang phát triển một frontend engine mới tên là Livewire với các tính năng như React nhưng chỉ cần trong 1 project và 100% PHP.

\* Nhược điểm:

- Laravel có nhiều tính năng và cấu hình phức tạp, vì thế nó đòi hỏi kiến thức lập trình PHP cao hơn so với các framework khác.
- Dù cung cấp nhiều tính năng và package mở rộng, nhưng khả năng mở rộng khi cần tích hợp hay phát triển các tính năng mới từ dự án cũ cũng sẽ gặp nhiều khó khăn.
- Với những tính năng và bộ package mạnh mẽ, Laravel sẽ tốn nhiều tài nguyên bộ nhớ để có thể chạy ứng dụng.

## 1.2 NextJS 14

### 1.2.1 Giới thiệu

NextJS[18] là framework mã nguồn mở được xây dựng trên nền tảng của React, cho phép chúng ta xây dựng các trang web tĩnh có tốc độ siêu nhanh và thân thiện với người dùng, cũng như xây dựng các ứng dụng web React. NextJS được ra đời vào năm 2016, thuộc sở hữu của Vercel. NextJS bắt đầu trở nên phổ biến vào năm 2018 và tiếp

tục tăng trưởng mạnh mẽ trong cộng đồng phát triển web vào những năm sau đó. Sự kết hợp của các tính năng như Server-side Rendering (SSR) với Static Site Generation (SSG) đã giúp NextJS trở thành sự lựa chọn hấp dẫn cho nhiều dự án phát triển ứng dụng web.

Next.js cung cấp một số tính năng mạnh mẽ giúp đơn giản hóa quá trình phát triển ứng dụng web như:

- Server-Side Rendering (SSR): Next.js hỗ trợ SSR, cho phép render trước các trang trên máy chủ.
- Static Site Generation (SSG): Next.js cũng hỗ trợ SSG, cho phép tạo ra các trang web tĩnh từ các thành phần React.
- Incremental Static Regeneration (ISR): Đây là một tính năng kết hợp giữa SSR và SSG, cho phép bạn tạo ra các trang tĩnh nhưng vẫn có thể cập nhật chúng một cách tăng dần theo thời gian.
- Routing: Next.js có hệ thống routing tích hợp, dựa trên cấu trúc thư mục, giúp dễ dàng định nghĩa các tuyến đường (route) cho ứng dụng.
- API Routes: Next.js cho phép bạn dễ dàng tạo các API route để xây dựng các ứng dụng backend.
- TypeScript Support: Next.js hỗ trợ TypeScript, một ngôn ngữ lập trình có kiểu dữ liệu mạnh, giúp bạn viết mã React rõ ràng, dễ bảo trì và ít lỗi hơn.

NextJS là một trong những web framework mới nhất và đang trở nên vô cùng phổ biến trong cộng đồng web hiện nay. Tốc độ cập nhật và phát triển của framework này chính là một điểm khiến framework này trở nên nổi bật và ưu dùng đối với những lập trình viên hiện đại. Tính đến nay, phiên bản mới nhất là Next.js 14, phiên bản mới nhất được phát hành vào tháng 10 năm 2023, mang đến nhiều cải tiến đáng kể như cơ chế App Router, Server Component, Streaming, ...

NextJS không chỉ đơn giản là một framework React, nó là một môi trường phát triển mạnh mẽ, mang lại hiệu suất tuyệt vời và cho trải nghiệm người dùng vượt trội. Bằng cách tận dụng các tính năng quan tích hợp sẵn, NextJS cho phép chúng ta xây dựng các ứng dụng React một cách nhanh chóng và hiệu quả.

### **1.2.2 Client Component**

Client Components là một tính năng mới được giới thiệu trong Next.js 14, mang đến sự linh hoạt và cải thiện hiệu suất cho các ứng dụng React. Với Client Components,

các nhà phát triển có thể lựa chọn render và thực thi các thành phần trực tiếp trên trình duyệt của người dùng (client-side rendering). Điều này đặc biệt hữu ích cho các thành phần có tính tương tác cao, yêu cầu cập nhật dữ liệu liên tục hoặc tận dụng các API của trình duyệt[16]. Client Components giúp giảm tải cho máy chủ và tăng tốc độ phản hồi của ứng dụng, vì chúng không cần phải được render lại từ đầu trên máy chủ mỗi khi có thay đổi.

Một điểm mạnh của Client Components là khả năng tương tác trực tiếp với DOM (Document Object Model) của trình duyệt. Điều này cho phép các nhà phát triển thực hiện các thao tác cập nhật giao diện người dùng một cách nhanh chóng và mượt mà, tạo ra trải nghiệm người dùng tốt hơn. Client Components cũng hỗ trợ việc sử dụng các thư viện JavaScript chỉ chạy trên client, mở rộng khả năng và tính linh hoạt của ứng dụng.

Tuy nhiên, Client Components cũng có một số hạn chế cần lưu ý. Do được thực thi trên client, chúng có thể làm tăng thời gian tải trang ban đầu nếu không được tối ưu hóa đúng cách. Ngoài ra, Client Components không phù hợp cho các thành phần chứa thông tin nhạy cảm hoặc cần được bảo vệ, vì mã nguồn của chúng có thể bị xem và sửa đổi bởi người dùng. Do đó, việc lựa chọn giữa Client Components và Server Components cần được cân nhắc kỹ lưỡng dựa trên yêu cầu cụ thể của từng thành phần và ứng dụng.

### **1.2.3 *Ưu điểm, nhược điểm***

**Ưu điểm:**

- **Hiệu suất cao:** Next.js tận dụng các kỹ thuật như SSR, SSG, ISR và code splitting để tối ưu hóa hiệu suất tải trang và trải nghiệm người dùng.
- **SEO thân thiện:** Nhờ SSR và SSG, các trang Next.js có thể được lập chỉ mục bởi các công cụ tìm kiếm một cách dễ dàng, giúp cải thiện thứ hạng SEO của ứng dụng.
- **Dễ phát triển:** Next.js cung cấp một cấu trúc dự án rõ ràng, hệ thống routing đơn giản và nhiều tính năng tiện ích khác, giúp giảm thiểu thời gian và công sức phát triển ứng dụng.
- **Cộng đồng lớn mạnh:** Next.js có một cộng đồng người dùng và nhà phát triển đông đảo, sẵn sàng hỗ trợ và chia sẻ kiến thức.

**Nhược điểm:**

- Khó làm chủ: Để tận dụng hết các tính năng của Next.js, bạn cần có kiến thức vững chắc về React và các khái niệm liên quan đến SSR, SSG, ISR.
- Hạn chế tùy biến: Next.js áp đặt một số quy ước về cấu trúc dự án và cách thức hoạt động, có thể gây khó khăn nếu bạn muốn tùy biến cao.
- Khó khăn trong việc tối ưu hóa: Việc tối ưu hóa hiệu suất của các ứng dụng Next.js phức tạp có thể đòi hỏi kiến thức chuyên sâu về các kỹ thuật tối ưu hóa.

### **1.3 Cơ sở dữ liệu**

#### **1.3.1 MySQL**

MySQL là một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở phổ biến được sử dụng rộng rãi trên toàn thế giới. Được phát triển bởi Oracle Corporation, MySQL cung cấp một cách tiếp cận đơn giản và mạnh mẽ cho việc lưu trữ và quản lý dữ liệu trong các ứng dụng web, từ các dự án cá nhân đến các hệ thống doanh nghiệp lớn.

Một trong những điểm mạnh của MySQL là tính nhất quán và độ ổn định cao. Nó cung cấp các tính năng cơ bản của một hệ quản trị cơ sở dữ liệu quan hệ như quản lý bảng, truy vấn dữ liệu, thêm, sửa, xóa dữ liệu và các chức năng phức tạp như giao dịch, khóa và xác thực người dùng..

Đối với những người mới bắt đầu, MySQL cung cấp một giao diện dễ sử dụng và một cộng đồng lớn với nhiều tài liệu và nguồn thông tin hữu ích. Nó cũng có sẵn trong các bản phân phối mã nguồn mở như MySQL Community Edition, giúp cho việc triển khai và phát triển các ứng dụng với chi phí thấp hoặc miễn phí.

Với tính linh hoạt và khả năng tương thích cao, MySQL có thể tích hợp với nhiều ngôn ngữ lập trình và framework phổ biến như PHP, Python, Java, Node.js và Ruby on Rails. Điều này giúp cho việc phát triển và triển khai các ứng dụng web trở nên linh hoạt và dễ dàng hơn [10].

=> MySQL là một hệ quản trị cơ sở dữ liệu mạnh mẽ và phổ biến với nhiều tính năng và ưu điểm. Tính đơn giản, tính linh hoạt và tính nhất quán của nó đã khiến cho MySQL trở thành một lựa chọn ưa thích cho việc lưu trữ và quản lý dữ liệu trong các ứng dụng web.

#### **1.3.2 Redis**

Redis là một hệ thống cơ sở dữ liệu key-value mã nguồn mở được phát triển bởi Salvatore Sanfilippo. Nó là một trong những hệ thống cơ sở dữ liệu NoSQL phổ biến

nhất hiện nay, nổi tiếng với hiệu suất cao, khả năng mở rộng tốt và các tính năng linh hoạt.

Redis là một trong những công nghệ phổ biến được sử dụng trong việc cải thiện hiệu suất và tính năng của các ứng dụng Laravel. Laravel cung cấp tích hợp sẵn với Redis thông qua các package mạnh mẽ như predis/predis hoặc illuminate/redis.

- Caching: Một trong những cách phổ biến nhất để sử dụng Redis trong Laravel là để lưu trữ dữ liệu cache. Redis cung cấp hiệu suất cao và khả năng mở rộng tốt cho việc lưu trữ dữ liệu cache, giúp tăng tốc độ truy cập dữ liệu và giảm thiểu thời gian phản hồi của ứng dụng.

- Session Storage: Laravel cũng cho phép lưu trữ phiên người dùng trong Redis thay vì lưu trữ trên ổ đĩa, giúp tăng cường tính bảo mật và khả năng mở rộng của ứng dụng.

- Queue Backend: Redis cung cấp một backend mạnh mẽ cho hệ thống hàng đợi của Laravel. Bằng cách sử dụng Redis làm đối tượng hàng đợi, bạn có thể xử lý các công việc hàng đợi một cách hiệu quả và đồng bộ.

- Real-time Data: Redis cũng được sử dụng trong Laravel để xây dựng các tính năng real-time như chat, thông báo hoặc đồng bộ dữ liệu trực tiếp giữa các ứng dụng khác nhau.

- Rate Limiting: Redis cung cấp các công cụ mạnh mẽ để triển khai các giới hạn tốc độ, giúp kiểm soát lưu lượng của các yêu cầu từ các nguồn khác nhau.

Redis là một công nghệ mạnh mẽ được sử dụng trong Laravel để cải thiện hiệu suất và tính năng của các ứng dụng. Tích hợp Redis vào Laravel giúp tăng cường hiệu suất, đồng thời cung cấp các tính năng mở rộng và linh hoạt cho các ứng dụng web.

#### **1.4 Python và khoa học dữ liệu**

Python là một ngôn ngữ lập trình bậc cao, thông dịch, hướng đối tượng và có kiểu động. Python được tạo ra bởi Guido van Rossum vào cuối những năm 1980 và được phát hành lần đầu tiên vào năm 1991. Python được thiết kế với mục tiêu dễ đọc, dễ học và dễ sử dụng, với cú pháp đơn giản và rõ ràng.

Python đã nổi lên như một ngôn ngữ lập trình hàng đầu trong lĩnh vực máy học (Machine Learning), học sâu (Deep Learning) và trí tuệ nhân tạo (Artificial Intelligence - AI) nhờ vào những ưu điểm sau:

- Thư viện phong phú: Python sở hữu một hệ sinh thái không lồ các thư viện chuyên dụng cho máy học, học sâu và AI.
- Cộng đồng lớn mạnh: Python có một cộng đồng người dùng và nhà phát triển đông đảo trong lĩnh vực máy học, học sâu và AI.
- Tính đơn giản và dễ đọc: Cú pháp của Python đơn giản và rõ ràng, giúp giảm thiểu thời gian và công sức cần thiết để viết mã.
- Tính linh hoạt: Python là một ngôn ngữ đa năng, có thể được sử dụng cho nhiều mục đích khác nhau, từ phân tích dữ liệu, xây dựng mô hình, đến triển khai ứng dụng.
- Nền tảng độc lập: Python có thể chạy trên nhiều hệ điều hành khác nhau như Windows, macOS và Linux, giúp dễ dàng chia sẻ và cộng tác trong các dự án.

Nhờ những ưu điểm này, Python đã trở thành ngôn ngữ được lựa chọn hàng đầu cho việc phát triển các ứng dụng máy học, học sâu và AI. Các công ty công nghệ lớn như Google, Facebook, và Microsoft đều sử dụng Python rộng rãi trong các dự án nghiên cứu và phát triển sản phẩm của mình.

### **1.4.1 Học máy trong Python**

Python không chỉ là một ngôn ngữ lập trình phổ biến mà còn là một công cụ mạnh mẽ được sử dụng rộng rãi trong lĩnh vực máy học, học sâu và trí tuệ nhân tạo. Với cú pháp đơn giản, dễ đọc và khả năng tích hợp với nhiều thư viện chuyên dụng như NumPy, pandas, scikit-learn, TensorFlow, và PyTorch, Python đã trở thành sự lựa chọn hàng đầu của các nhà khoa học dữ liệu, kỹ sư máy học và các chuyên gia AI. Nhiều công ty công nghệ hàng đầu thế giới như Google, Facebook, Microsoft và OpenAI đều sử dụng Python làm ngôn ngữ chính để phát triển các ứng dụng và giải pháp AI tiên tiến. Sự phổ biến của Python trong cộng đồng khoa học và công nghệ đã tạo ra một hệ sinh thái phong phú với nhiều tài liệu, hướng dẫn và cộng đồng hỗ trợ, giúp việc học và áp dụng Python trở nên dễ dàng hơn bao giờ hết.

#### **1.4.1.1 Học máy là gì**

Học máy (Machine Learning - ML) là một lĩnh vực của trí tuệ nhân tạo (AI) tập trung vào việc phát triển các thuật toán và mô hình cho phép máy tính tự học từ dữ liệu và cải thiện hiệu suất của chúng trong việc thực hiện các nhiệm vụ cụ thể mà không cần lập trình rõ ràng. Mục tiêu của học máy là xây dựng các hệ thống có khả năng tự động phân tích và hiểu dữ liệu, từ đó đưa ra các quyết định hoặc dự đoán chính xác. Python,

với cú pháp đơn giản, dễ đọc và hệ sinh thái thư viện phong phú, đã trở thành một trong những ngôn ngữ lập trình hàng đầu trong lĩnh vực này [4].

Học máy được chia ra thành 2 loại chính thường thấy như sau: Học có giám sát (Supervised Learning) và học không giám sát (Unsupervised Learning), hay đơn giản hơn có thể hiểu là nhóm máy học có hướng dẫn và không có hướng dẫn. Supervised Learning là mô hình được huấn luyện trên một tập dữ liệu có nhãn (labeled data), trong đó mỗi mẫu dữ liệu được gắn với một kết quả mong muốn. Unsupervised Learning là mô hình huấn luyện trên một tập dữ liệu không có nhãn (unlabeled data).

#### *1.4.1.2 Nhóm học máy có giám sát*

Học máy giám sát là một nhánh của học máy, trong đó mô hình được huấn luyện trên một tập dữ liệu có nhãn (labeled data). Mỗi mẫu dữ liệu trong tập huấn luyện bao gồm một tập các đặc trưng (features) đầu vào và một nhãn (label) tương ứng, đại diện cho kết quả mong muốn. Mục tiêu của học máy giám sát là xây dựng một mô hình có khả năng học từ tập dữ liệu huấn luyện này để dự đoán nhãn cho các mẫu dữ liệu mới chưa biết trước nhãn.

Bản chất của học máy giám sát là tìm ra một hàm số hoặc một quy tắc ánh xạ từ không gian đặc trưng đầu vào sang không gian nhãn. Quá trình huấn luyện mô hình liên quan đến việc điều chỉnh các tham số của mô hình sao cho nó có thể dự đoán nhãn một cách chính xác nhất có thể trên tập dữ liệu huấn luyện. Sau khi được huấn luyện, mô hình có thể được sử dụng để dự đoán nhãn cho các mẫu dữ liệu mới.

#### **1.4.2 Thuật toán Random Forest**

##### *1.4.2.1 Cây quyết định*

Cây quyết định là một mô hình học máy được sử dụng rộng rãi để thực hiện các nhiệm vụ phân loại và hồi quy. Mô hình này hoạt động bằng cách phân chia dữ liệu thành các nhánh dựa trên các thuộc tính của dữ liệu. Mỗi nút trong cây quyết định đại diện cho một thuộc tính và mỗi nhánh đại diện cho một giá trị hoặc phạm vi giá trị của thuộc tính đó. Quá trình này tiếp tục cho đến khi đạt được một quyết định cuối cùng tại các nút lá, nơi xác định lớp hoặc giá trị dự đoán. Cây quyết định có lợi thế lớn ở tính trực quan và dễ hiểu, giúp người dùng dễ dàng theo dõi và lý giải quá trình ra quyết định.

Quá trình xây dựng cây quyết định bắt đầu bằng việc lựa chọn thuộc tính tốt nhất để chia tách dữ liệu tại mỗi nút. Các biện pháp như entropy và Gini impurity được sử

dụng để đánh giá chất lượng của các thuộc tính này, giúp xác định thuộc tính nào sẽ tạo ra sự phân chia dữ liệu tốt nhất. Sau khi chọn thuộc tính, dữ liệu được chia thành các nhánh dựa trên giá trị của thuộc tính đó. Quá trình này tiếp tục cho đến khi đạt điều kiện dừng, chẳng hạn như khi dữ liệu trong các nhánh đã đồng nhất hoặc cây đạt đến độ sâu tối đa.

Tuy nhiên, cây quyết định cũng gặp phải các vấn đề như overfitting và underfitting. Overfitting xảy ra khi cây quá phức tạp, học quá chi tiết các đặc điểm của dữ liệu huấn luyện, bao gồm cả nhiễu, dẫn đến hiệu suất kém trên dữ liệu kiểm tra. Ngược lại, underfitting xảy ra khi cây quá đơn giản, không thể nắm bắt được mối quan hệ quan trọng trong dữ liệu. Để giải quyết vấn đề này, có thể áp dụng các kỹ thuật như cắt tỉa (pruning) cây, giảm độ sâu tối đa của cây, hoặc yêu cầu một số lượng tối thiểu các mẫu tại mỗi nút lá.

Cây quyết định có nhiều ưu điểm, bao gồm tính dễ hiểu và khả năng xử lý dữ liệu phi tuyến tính. Tuy nhiên, nhược điểm của nó là dễ bị overfitting và nhạy cảm với những thay đổi nhỏ trong dữ liệu. Để tận dụng tối đa mô hình này, cần phải cẩn thận trong việc lựa chọn các tham số và kỹ thuật xử lý dữ liệu. Với những hiểu biết và kỹ thuật đúng đắn, cây quyết định có thể là một công cụ mạnh mẽ trong việc giải quyết các bài toán học máy và dự đoán dữ liệu [5].

#### *1.4.2.1 Thuật toán Random Forest và các phương pháp Ensemble*

Random Forest là một phương pháp học máy mạnh mẽ dựa trên việc kết hợp nhiều cây quyết định để cải thiện độ chính xác và ổn định của mô hình. Thay vì chỉ sử dụng một cây quyết định đơn lẻ, Random Forest xây dựng nhiều cây quyết định độc lập bằng cách sử dụng các mẫu ngẫu nhiên khác nhau từ tập dữ liệu gốc. Mỗi cây trong rừng sẽ đưa ra dự đoán và kết quả cuối cùng được xác định bằng cách lấy trung bình (đối với bài toán hồi quy) hoặc bỏ phiếu đa số (đối với bài toán phân loại) của các dự đoán từ tất cả các cây. Phương pháp này giúp giảm thiểu vấn đề overfitting, vì các sai số của các cây riêng lẻ có xu hướng bù trừ lẫn nhau.

Phương pháp ensemble là một kỹ thuật trong học máy nhằm cải thiện hiệu suất và độ chính xác của mô hình bằng cách kết hợp nhiều mô hình dự đoán khác nhau. Random Forest là một ví dụ điển hình của phương pháp ensemble, sử dụng kỹ thuật bagging (Bootstrap Aggregating). Bagging tạo ra nhiều mô hình bằng cách huấn luyện chúng trên các tập con khác nhau của dữ liệu huấn luyện, được lấy mẫu ngẫu nhiên với

hoàn nguyên. Kết quả cuối cùng được tổng hợp từ các mô hình này, giúp giảm phuơng sai và tăng độ ổn định của dự đoán.

Ngoài Random Forest, các phương pháp ensemble khác cũng được sử dụng rộng rãi, bao gồm boosting và stacking. Boosting là một kỹ thuật cải tiến hiệu suất bằng cách huấn luyện các mô hình tuần tự, mỗi mô hình mới cố gắng sửa các lỗi của mô hình trước đó. Các thuật toán như AdaBoost và Gradient Boosting là những ví dụ nổi bật của boosting. Stacking, mặt khác, kết hợp các mô hình khác nhau bằng cách huấn luyện một mô hình meta trên đầu ra của các mô hình cơ sở. Cả hai phương pháp này đều nhằm mục đích tận dụng sức mạnh của nhiều mô hình để đạt được hiệu suất tốt hơn so với việc sử dụng một mô hình đơn lẻ [5].

#### **1.4.3 Thư viện Scikit learn**

Scikit-learn là một thư viện mã nguồn mở hàng đầu trong lĩnh vực học máy (Machine Learning) được xây dựng trên nền tảng Python. Scikit-learn cung cấp một bộ công cụ toàn diện và dễ sử dụng để thực hiện các tác vụ học máy khác nhau, từ tiền xử lý dữ liệu, lựa chọn đặc trưng, huấn luyện mô hình, đánh giá mô hình đến lựa chọn mô hình tối ưu [6].

Các tính năng chính của Scikit-learn:

- Thuật toán học máy đa dạng: Scikit-learn cung cấp một bộ sưu tập phong phú các thuật toán học máy, bao gồm các thuật toán phân loại (classification), hồi quy (regression), phân cụm (clustering) và giảm chiều dữ liệu (dimensionality reduction).
- Tiền xử lý dữ liệu: Scikit-learn cung cấp nhiều công cụ để xử lý dữ liệu thô, chẳng hạn như chuẩn hóa (scaling), mã hóa (encoding), và xử lý dữ liệu thiếu (missing data imputation).
- Lựa chọn đặc trưng: Scikit-learn cung cấp các phương pháp để lựa chọn các đặc trưng quan trọng nhất trong tập dữ liệu, giúp cải thiện hiệu suất của mô hình và giảm thiểu overfitting.
- Đánh giá mô hình: Scikit-learn cung cấp nhiều độ đo để đánh giá hiệu suất của mô hình, chẳng hạn như độ chính xác (accuracy), độ đo F1 (F1 score), và diện tích dưới đường cong ROC (AUC-ROC).

- Lựa chọn mô hình: Scikit-learn cung cấp các kỹ thuật như tìm kiếm tham số ngẫu nhiên (randomized search) và tìm kiếm lưới (grid search) để tìm ra các siêu tham số tối ưu cho mô hình.

Ưu điểm của Scikit-learn:

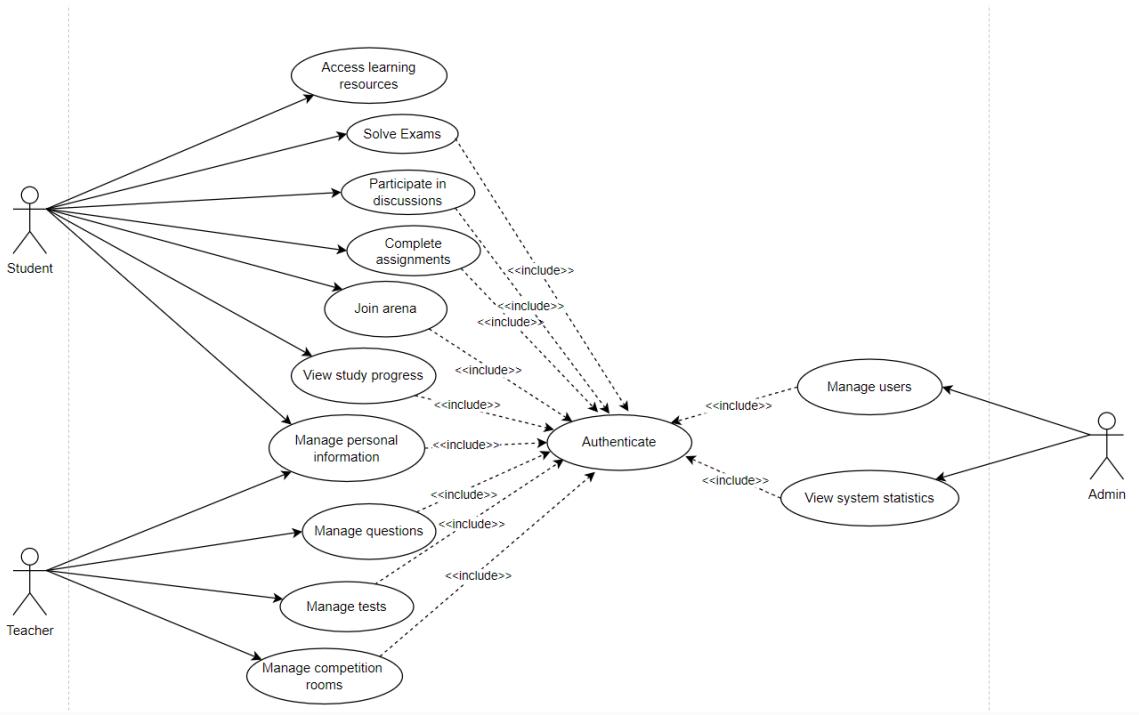
- Dễ sử dụng: Scikit-learn có một API (Application Programming Interface) đơn giản và nhất quán, giúp người dùng dễ dàng học và sử dụng thư viện.
- Hiệu quả: Scikit-learn được tối ưu hóa để hoạt động hiệu quả trên các tập dữ liệu lớn, giúp giảm thiểu thời gian và tài nguyên tính toán cần thiết.
- Linh hoạt: Scikit-learn cung cấp nhiều tùy chọn cấu hình cho các thuật toán, cho phép bạn điều chỉnh mô hình để phù hợp với bài toán cụ thể.
- Cộng đồng lớn mạnh: Scikit-learn có một cộng đồng người dùng và nhà phát triển lớn, sẵn sàng hỗ trợ và chia sẻ kiến thức.

## CHƯƠNG 2. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG

### 2.1 Thiết kế hệ thống

#### 2.1.1 Sơ đồ Usecase

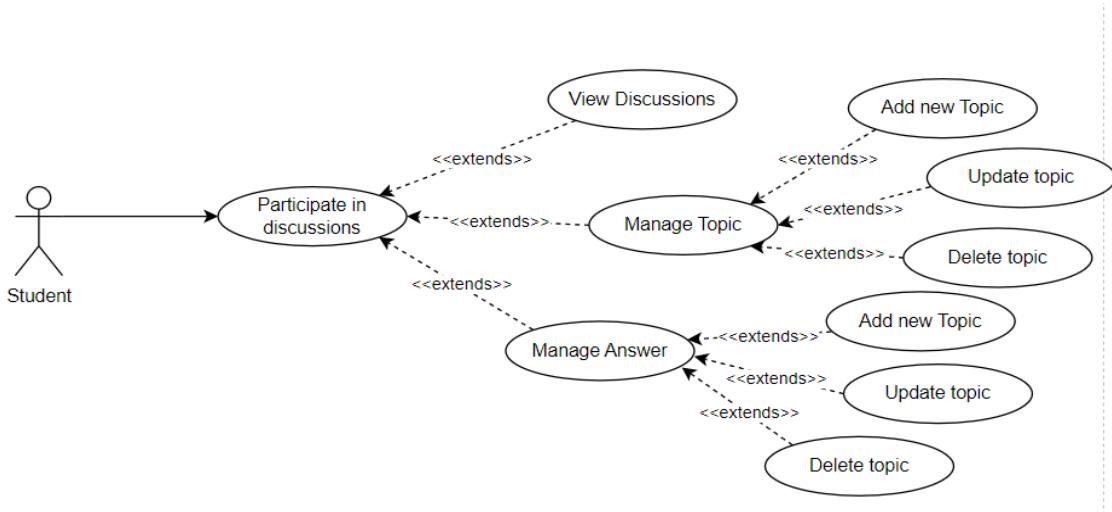
##### 2.1.1.1 Sơ đồ Usecase tổng quát



Hình 2. 1 Sơ đồ Usecase tổng quát

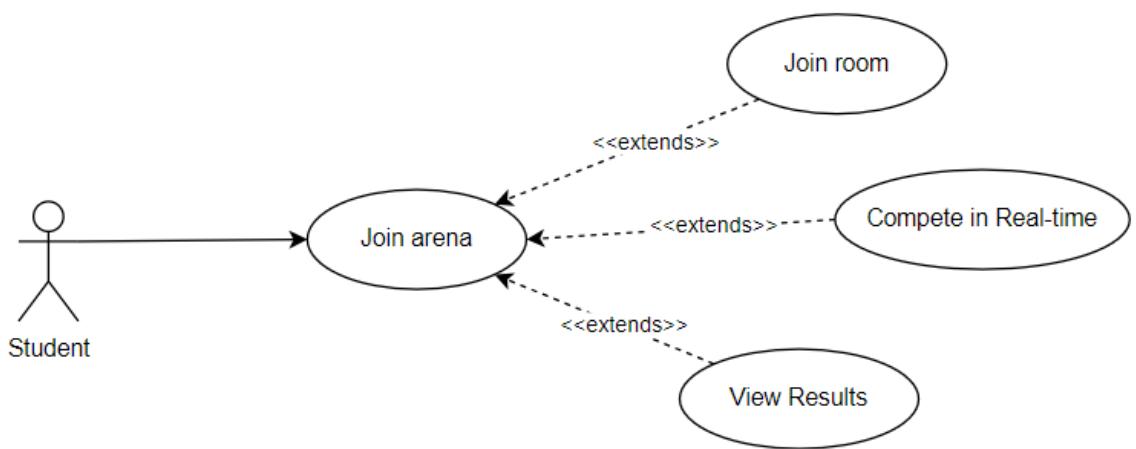
##### 2.1.1.2 Sơ đồ phân rã Usecase

###### \* Sơ đồ phân rã Usecase Thảo luận



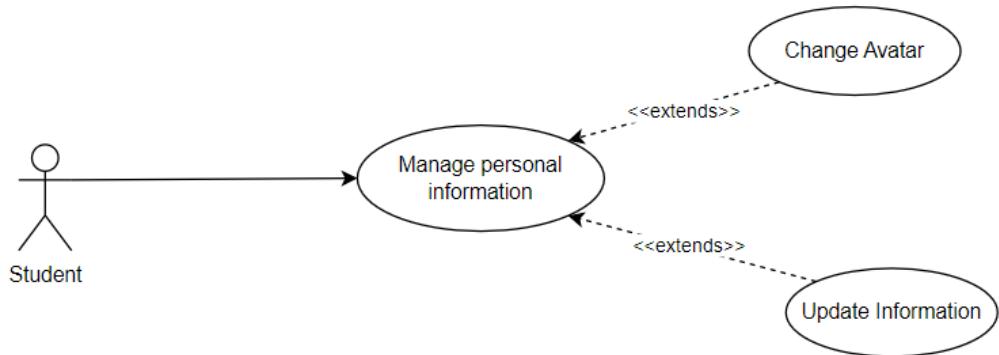
Hình 2. 2 Sơ đồ phân rã Usecase Thảo luận

\* Sơ đồ phân rã Usecase Tham gia thi đấu



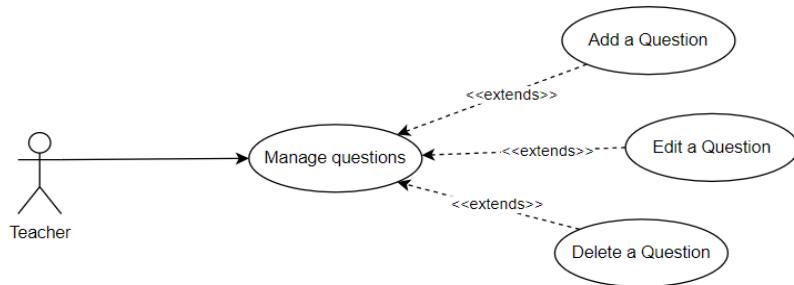
Hình 2. 3 Sơ đồ phân rã Usecase Tham gia thi đấu

\* Sơ đồ phân rã Usecase Quản lý thông tin cá nhân



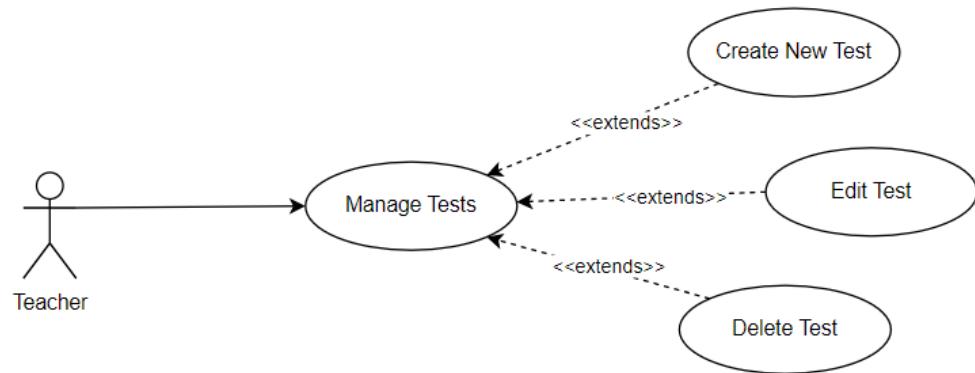
Hình 2. 4 Sơ đồ phân rã Usecase Quản lý thông tin cá nhân

\* Sơ đồ phân rã Usecase Quản lý câu hỏi



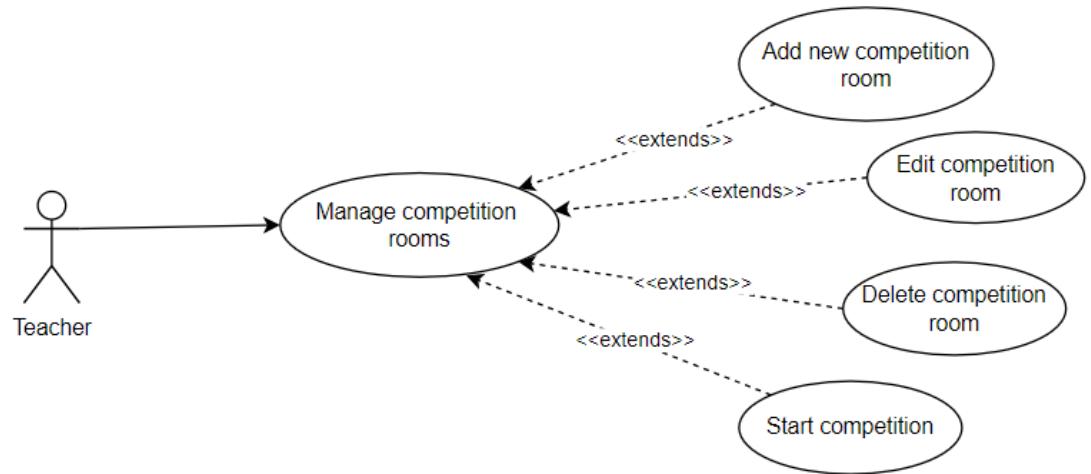
Hình 2. 5 Sơ đồ phân rã Usecase Quản lý câu hỏi

\* Sơ đồ phân rã Usecase Quản lý bài thi



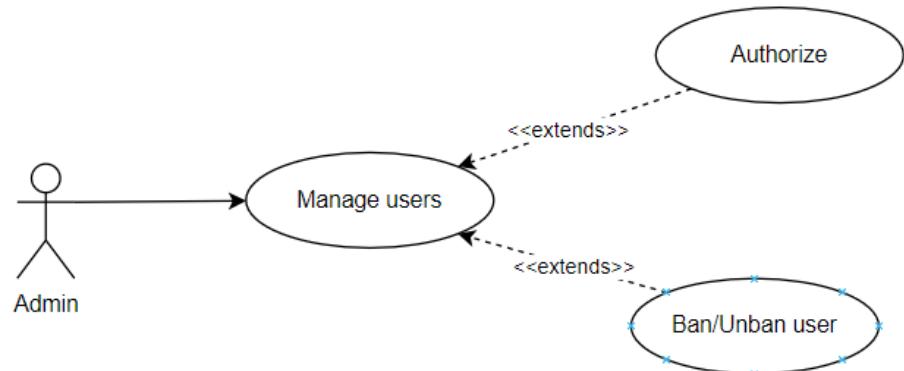
Hình 2. 6 Sơ đồ phân rã Usecase Quản lý bài thi

\* Sơ đồ phân rã Usecase Quản lý phòng thi



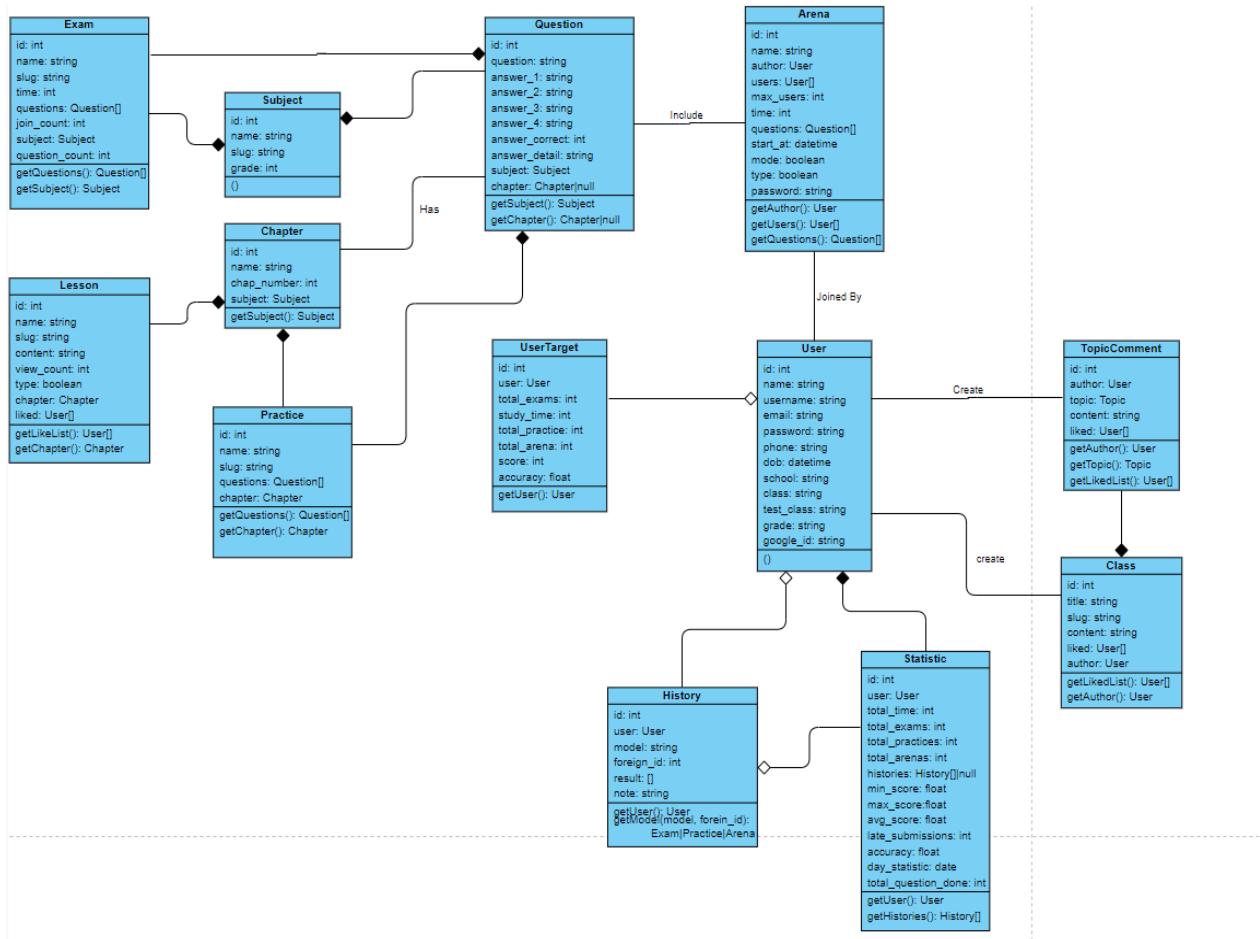
Hình 2. 7 Sơ đồ phân rã Usecase Quản lý phòng thi

\* Sơ đồ phân rã Usecase Quản lý người dùng



Hình 2. 8 Sơ đồ phân rã Usecase Quản lý người dùng

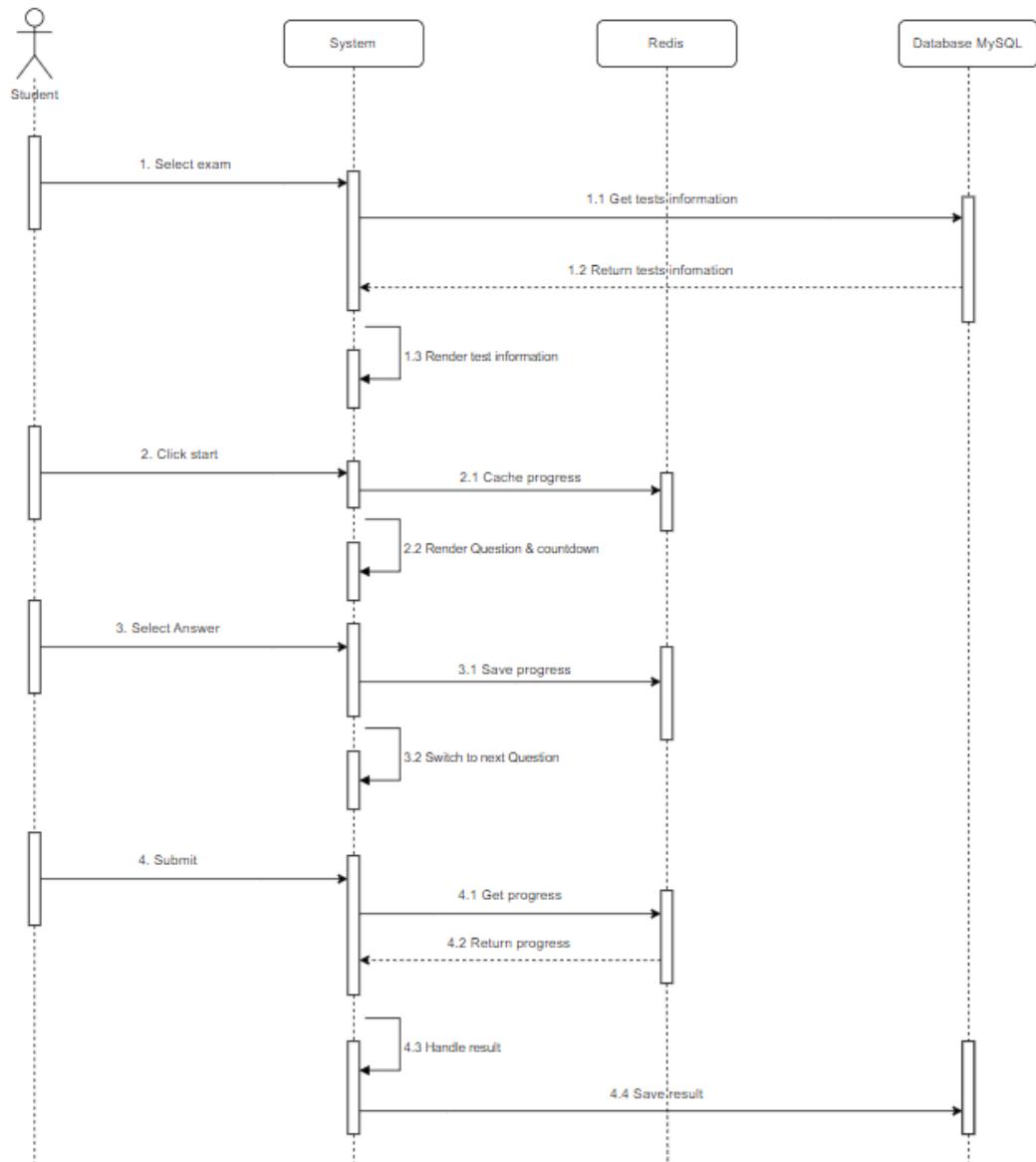
## 2.1.2 Sơ đồ Class



Hình 2. 9 Sơ đồ Class

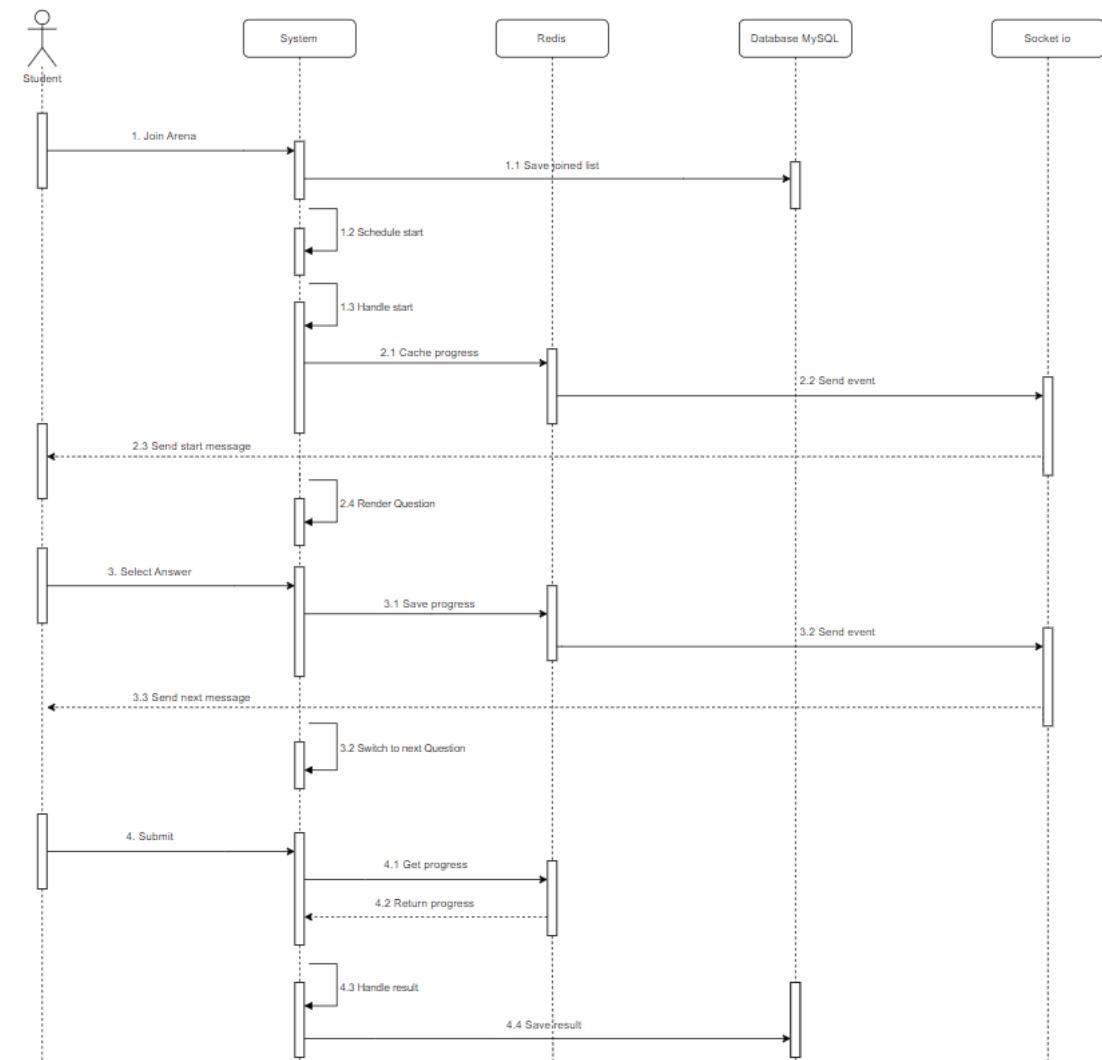
## 2.1.3 Sơ đồ tuần tự

### 2.1.3.1 Sơ đồ tuần tự thi thử



Hình 2. 10 Sơ đồ tuần tự thi thử

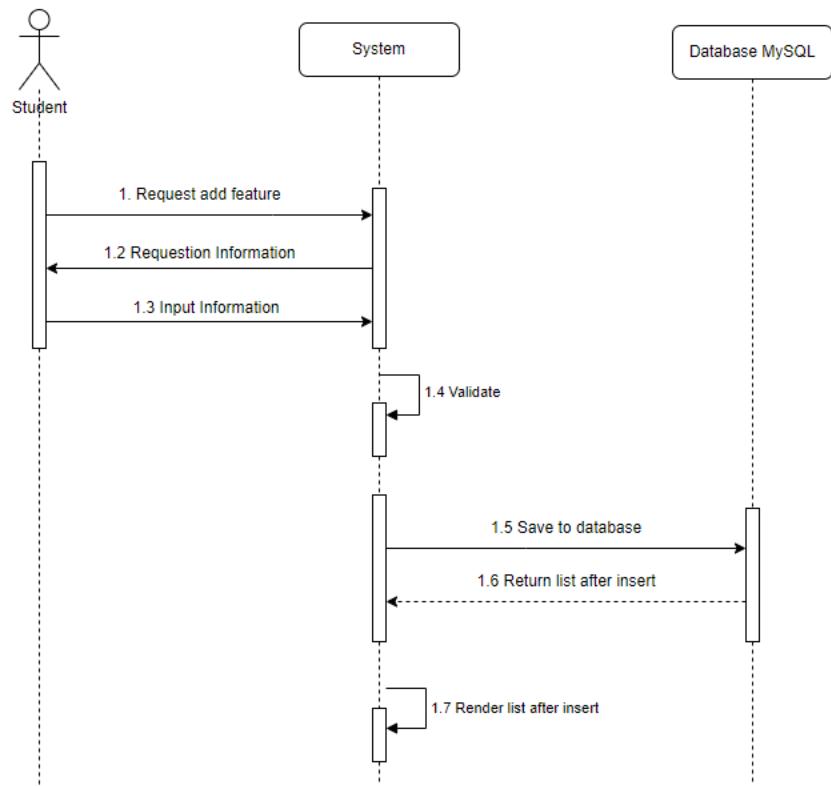
### 3.1.3.2 Sơ đồ tuần tự đấu trường



Hình 2. 11 Sơ đồ tuần tự đấu trường

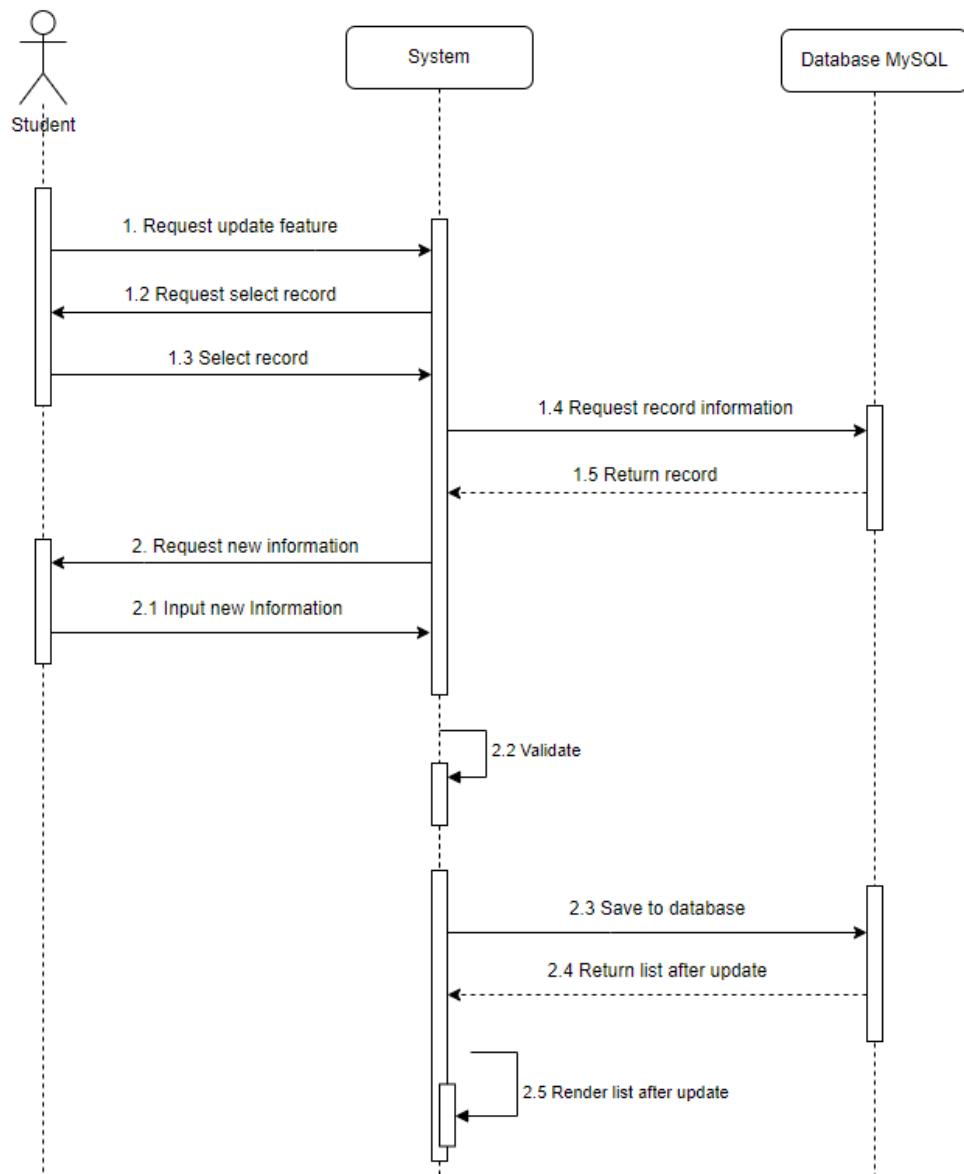
### 2.1.3.3 Sơ đồ tuần tự quản lý (thêm sửa xóa)

\* Thêm mới bản ghi



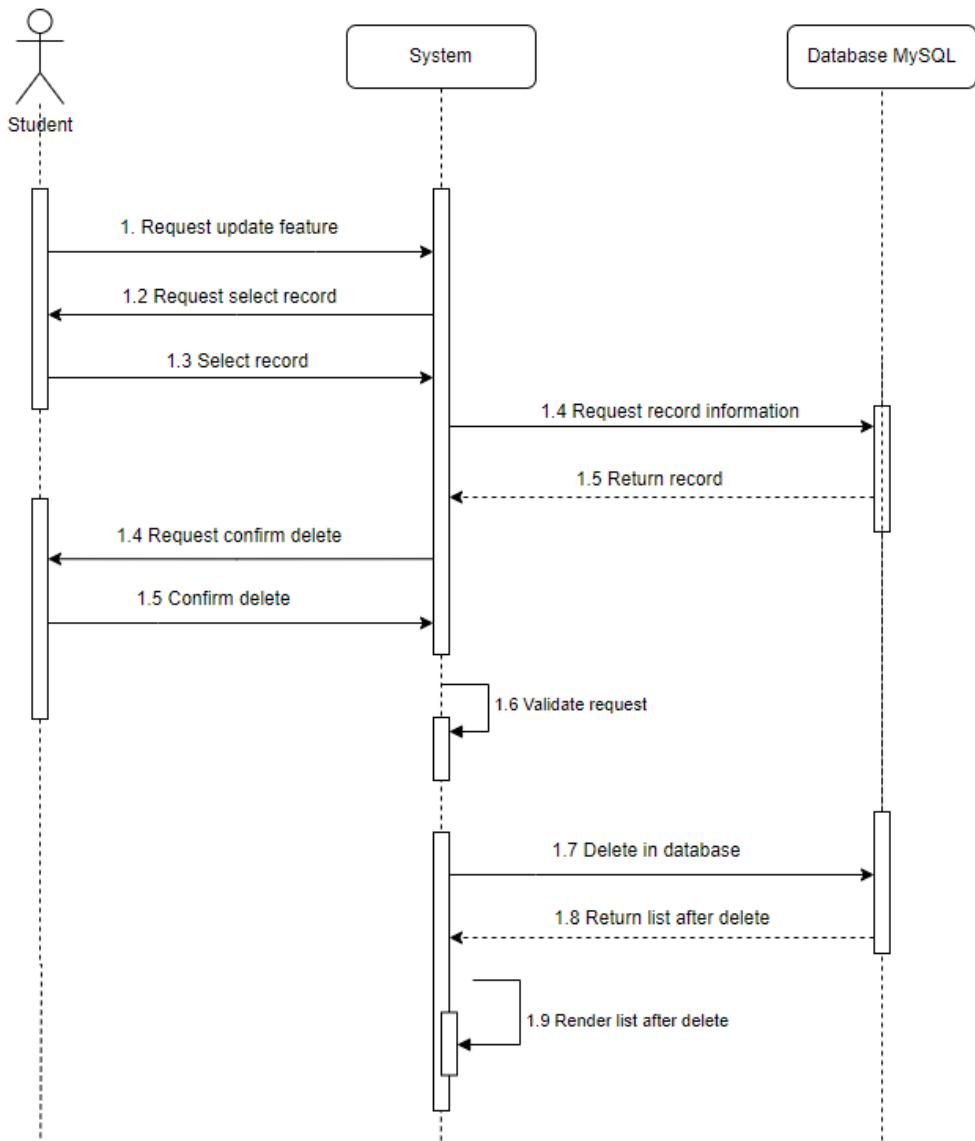
Hình 2. 12 Sơ đồ tuần tự Thêm mới bản ghi

\* Cập nhật bản ghi



Hình 2. 13 Sơ đồ tuần tự Cập nhật bản ghi

\* Xóa bản ghi



Hình 2. 14 Sơ đồ tuần tự xóa bản ghi

## 2.2 Thiết kế cơ sở dữ liệu

### 2.2.1 Mô tả bài toán

Trang web luyện thi Đại Học trực tuyến được thiết kế nhằm cung cấp một môi trường học tập toàn diện và hiệu quả, bao gồm các tính năng như hỏi – đáp, cung cấp lý thuyết và bài tập SGK, luyện tập giải đề, tính năng đấu trường và thống kê tình hình ôn luyện của người dùng. Cơ sở dữ liệu được xây dựng cần hỗ trợ việc quản lý người dùng, nội dung học tập, các kỳ thi và tương tác giữa người dùng một cách hiệu quả. Dữ liệu được lưu trữ:

1. Xác thực và Phân quyền: Nền tảng yêu cầu người dùng đăng ký và đăng nhập để sử dụng các tính năng. Người dùng có thể là học sinh, giáo viên hoặc quản trị viên/người kiểm duyệt. Mỗi loại tài khoản sẽ có quyền hạn khác nhau.

Bảng users: Lưu trữ thông tin của tất cả người dùng, bao gồm tên đăng nhập, email, mật khẩu và vai trò (role).

2. Tính năng hỏi – đáp: Người dùng có thể tạo các chủ đề thảo luận và đặt câu hỏi để nhận sự trợ giúp từ cộng đồng. Các bình luận sẽ được sử dụng để trao đổi và góp ý.

Bảng topics: Lưu trữ các chủ đề thảo luận được tạo bởi người dùng.

Bảng topic\_comments: Lưu trữ các bình luận trên từng chủ đề, giúp người dùng tương tác với nhau.

3. Cung cấp lý thuyết và bài tập SGK: Nền tảng cung cấp nội dung học tập theo chương trình chuẩn, bao gồm lý thuyết và các bài tập có đáp án.

Bảng subjects: Lưu trữ các môn học khác nhau.

Bảng chapters: Lưu trữ các chương trong từng môn học.

Bảng lessons: Lưu trữ các bài học trong mỗi chương, bao gồm nội dung lý thuyết và bài tập.

4. Luyện tập giải đề: Học sinh có thể luyện tập các loại đề thi mẫu dưới hình thức trắc nghiệm trong thời gian quy định để chuẩn bị cho kỳ thi thật.

Bảng exams: Lưu trữ thông tin về các đề thi, bao gồm thời gian làm bài và danh sách các câu hỏi.

Bảng questions: Lưu trữ các câu hỏi trắc nghiệm, bao gồm nội dung câu hỏi và các phương án trả lời.

5. Tính năng đấu trường: Người dùng có thể tạo phòng thi đấu để thi cùng lúc với nhau (realtime), giúp tăng cường tính cạnh tranh và tương tác.

Bảng arenas: Lưu trữ thông tin về các phòng thi đấu, bao gồm người tạo, thời gian bắt đầu và kết thúc, và danh sách người tham gia.

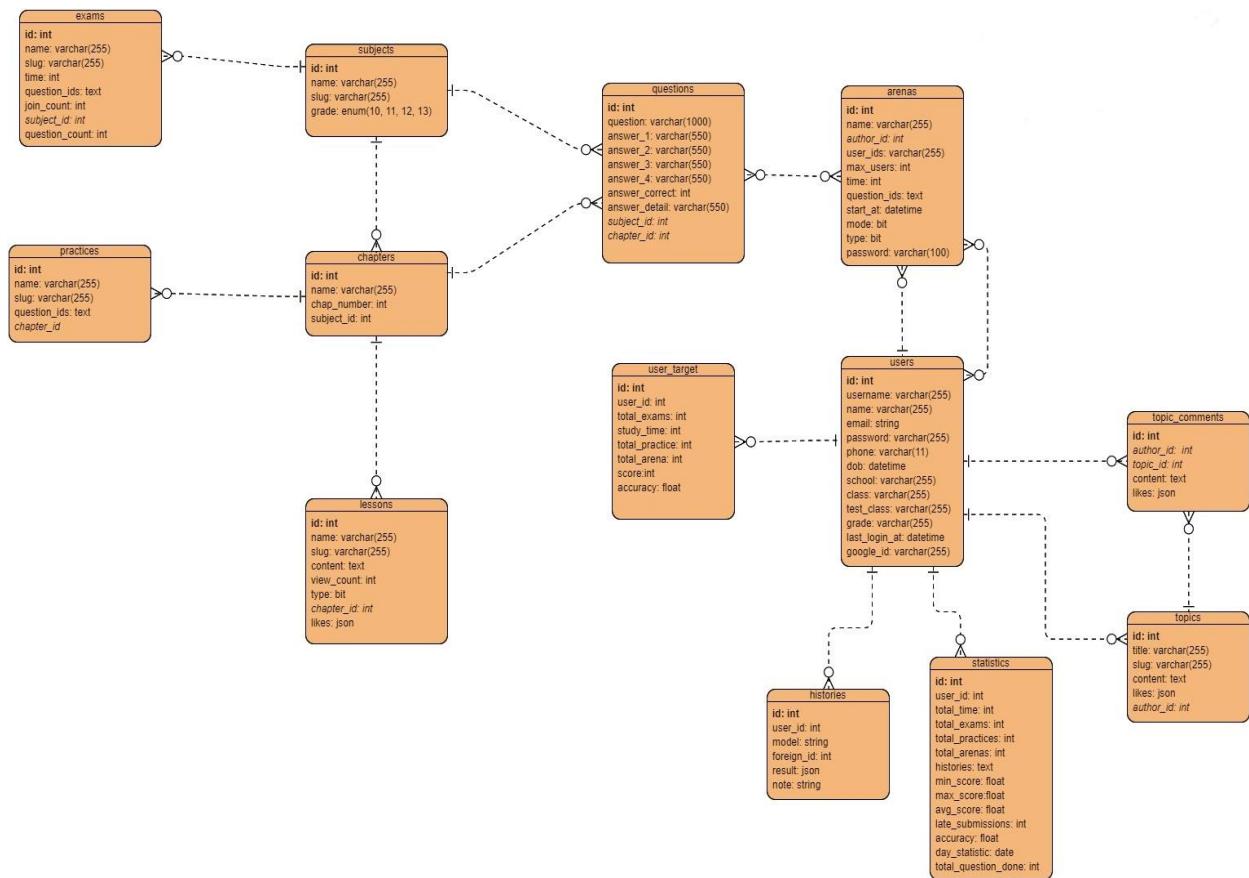
6. Thống kê tình hình ôn luyện của người dùng: Nền tảng theo dõi và cung cấp thông tin chi tiết về quá trình ôn luyện của từng người dùng, bao gồm thời gian học, số lượng bài thi và độ chính xác.

Bảng user\_stats: Lưu trữ các thống kê về hoạt động học tập của người dùng.

Bảng histories: Lưu trữ lịch sử ôn luyện và kết quả từng bài thi của người dùng.

## 2.2.2 Sơ đồ ER

\* ERD Diagram



Hình 2. 15 Sơ đồ ER

## 2.2.3 Bảng thực thể cờ sở dữ liệu

Bảng 2. 1 Bảng thực thể User

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	username	STRING	
3	name	STRING	
4	email	STRING	
5	password	STRING	
6	phone	STRING	
7	invite_code	STRING	
8	avatar	STRING	
9	gender	ENUM	
10	dob	DATE	
11	address	STRING	
12	school	STRING	
13	class	STRING	
14	test_class	STRING	
15	grade	INT	

*Bảng 2. 2 Bảng thực thể Subject*

<b>STT</b>	<b>Trường</b>	<b>Kiểu</b>	<b>Ghi chú</b>
1	id	INT	PK
2	name	STRING	
3	slug	STRING	
4	grade	INT	

*Bảng 2. 3 Bảng thực thể Chapter*

<b>STT</b>	<b>Trường</b>	<b>Kiểu</b>	<b>Ghi chú</b>
1	id	INT	PK
2	name	STRING	
3	subject_id	STRING	

*Bảng 2. 4 Bảng thực thể Practice*

<b>STT</b>	<b>Trường</b>	<b>Kiểu</b>	<b>Ghi chú</b>
1	id	INT	PK
2	name	STRING	
3	questions	STRING	
4	slug	STRING	
5	subject_id	INT	FK

*Bảng 2. 5 Bảng thực thể Lesson*

<b>STT</b>	<b>Trường</b>	<b>Kiểu</b>	<b>Ghi chú</b>
1	id	INT	PK
2	name	STRING	
3	chap_id	INT	FK
4	content	STRING	
5	view_count	INT	
6	type	STRING	
7	subject_id	INT	
8	likes	STRING	

*Bảng 2. 6 Bảng thực thể Question*

<b>STT</b>	<b>Trường</b>	<b>Kiểu</b>	<b>Ghi chú</b>
1	id	INT	PK
2	question	STRING	
3	answer_1	STRING	
4	answer_2	STRING	
5	answer_3	STRING	

6	answer_4	STRING	
7	anwser_correct	INT	
8	anwser_detail	TEXT	
9	subject_id	INT	FK
10	chap_id	INT	FK
11	level	INT	

*Bảng 2. 7 Bảng thực thể Exam*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	name	STRING	
3	slug	STRING	
4	time	DATETIME	
5	questions	STRING	
6	join_count	INT	
7	complete_count	INT	
8	subject_id	INT	FK

*Bảng 2. 8 Bảng thực thể Topic*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	title	STRING	
3	slug	STRING	
4	content	STRING	
5	author	INT	FK

*Bảng 2. 9 Bảng thực thể TopicComment*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	topic	STRING	
3	author	INT	FK
4	content	STRING	
5	likes	STRING	

*Bảng 2. 10 Bảng thực thể Arena*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	name	STRING	
3	author	INT	FK
4	users	STRING	
5	max_users	INT	

6	time	INT	
7	questions	STRING	
8	start_at	DATETIME	
9	type	BIT	
10	password	STRING	

*Bảng 2. 11 Bảng thực thể UserTarget*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	user_id	INT	FK
3	total_time	INT	
4	total_exam	INT	
5	total_practice	INT	
6	total_arena	INT	
7	score	FLOAT	
8	late_submissions	INT	
9	accuracy	FLOAT	

*Bảng 2. 12 Bảng thực thể History*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	user_id	INT	FK
3	model	STRING	
4	foreign_id	INT	
5	result	STRING	
6	note	STRING	

*Bảng 2. 13 Bảng thực thể DayStatistics*

STT	Trường	Kiểu	Ghi chú
1	id	INT	PK
2	user_id	INT	FK
3	total_time	INT	
4	total_exam	INT	
5	total_practice	INT	
6	total_arena	INT	
7	histories	STRING	
8	min_score	FLOAT	
9	max_score	FLOAT	
10	avg_score	FLOAT	
11	late_submissions	INT	
12	accuracy	FLOAT	

## CHƯƠNG 3. XÂY DỰNG MÔ HÌNH DỰ ĐOÁN

### 3.1 Thu thập dữ liệu

#### 3.1.1 Xác định các yếu tố ảnh hưởng đến điểm thi

Để xây dựng mô hình dự đoán kết quả học tập, việc đầu tiên cần làm là phải thu thập dữ liệu để huấn luyện mô hình. Tập dữ liệu này cần xác định và phản ánh rõ những yếu tố có thể ảnh hưởng đến việc dự đoán điểm thi và hiệu suất của mô hình. Ngoài ra tập dữ liệu này cần đảm bảo đủ độ lớn và tính chính xác để mô hình có thể hoạt động đúng.

Dựa trên dữ liệu có thể thu thập để đánh giá trên web, khảo sát ở các diễn đàn học sinh như forum hocmai, thitot, các hội nhóm trên các trang mạng xã hội như Facebook, cùng với đó là các tài liệu nghiên cứu về phương pháp học tập hiệu quả (Hattie, 2009) và các báo cáo đánh giá kết quả học tập của các tổ chức giáo dục uy tín như trung tâm luyện thi hocmai, mclass. Từ đó xác định được các yếu tố chính sau đây có thể ảnh hưởng đến kết quả học tập và điểm thi của thí sinh.

Đầu tiên là những yếu tố chủ quan của thí sinh, bao gồm các thông tin cơ bản được thu thập từ chính người dùng dựa trên thực tế:

*Bảng 3. 1 Bảng các yếu tố cá nhân quyết định điểm thi*

STT	Thuộc tính	Giải thích
1	Age	Độ tuổi của người dùng
2	Grade	Khối lớp hiện tại của người dùng
3	StudyTime	Thời gian trung bình học mỗi ngày (giờ) (chỉ tính thời gian học chính khóa và bồi túc trên trường)
4	OnlineCourse	Có đăng ký khóa luyện thi trên mạng không
5	SchoolType	Loại trường THPT (non-public, public, specialized)
6	ClassType	Loại lớp (normal, gifted)
7	SelfStudy	Tần suất tự học (Từ 1 - rất ít đến 5 - rất nhiều)
8	GoingOut	Tần suất đi chơi với bạn bè (Từ 1 - rất ít đến 5 - rất nhiều)
9	Health	Tình trạng sức khỏe chung (Từ 1 - rất kém đến 5 - rất tốt)
10	LatestScore	Điểm trung bình năm học trước đó (thang 0 - 10)

Thứ hai là những yếu tố dữ liệu được thu thập từ chính trang web trong đề tài này, dựa trên kết quả ôn luyện thực tế của người dùng trên web:

*Bảng 3. 2 Bảng các yếu tố web quyết định điểm thi*

STT	Thuộc tính	Giải thích
1	DailyTimeSpent	Thời gian dành cho web mỗi ngày (giờ)
2	DayActive	Số ngày đã hoạt động trên web
3	ResourceAccessed	Số lượng tài nguyên học tập đã truy cập
4	ForumParticipationFrequency	Tần suất tham gia thảo luận trên diễn đàn (từ 1 - rất ít đến 5 - rất tích cực)
5	HelpfulAnswers	Số câu trả lời được đánh giá là hữu ích
6	ExercisesCompleted	Số bài tập đã làm
7	AverageExerciseScore	Điểm trung bình các bài tập
8	TestsCompleted	Số bài kiểm tra đã làm
9	AverageTestScore	Điểm trung bình các bài kiểm tra
10	AverageTestCompletionTime	Thời lượng làm bài trung bình 1 lần thi (%)
11	ArenaParticipation	Số lần tham gia đấu trường
12	ArenaQuestionsAnsweredPercent	Phần trăm câu hỏi trả lời được trong đấu trường
13	BestOfArenaCount	Số lần top 1 trong đấu trường
14	AverageArenaRanking	Vị trí trung bình trong đấu trường (từ 1 - rất thấp 5 - rất cao)
15	MostTime	Thời gian ôn luyện thường trên web
16	UsageTrend	Xu hướng hoạt động gần đây (increasing - tăng, decreasing - giảm)
17	PredictScore	Điểm thi được dự đoán

Trong đó thi trường PredictScore sẽ là dữ liệu mục tiêu của mô hình dự đoán này. Từ đó suy ra hiệu suất ôn luyện

### **3.1.2 Phân tích mối quan hệ giữa các yếu tố ảnh hưởng đến kết quả dự đoán**

#### **\* Mối Quan Hệ Tích Cực**

Thời gian học tập (StudyTime): Thời gian học tập càng nhiều, khả năng đạt điểm số cao càng lớn. Việc đầu tư nhiều thời gian vào việc học giúp học sinh nắm vững kiến thức và kỹ năng cần thiết để đạt kết quả tốt.

Luyện thi trực tuyến (OnlineCourse): Tham gia các khóa học trực tuyến cung cấp thêm kiến thức và kỹ năng, từ đó cải thiện điểm số. Khóa học trực tuyến thường có nội dung phong phú và đa dạng, hỗ trợ học sinh tiếp cận với nhiều nguồn tài liệu và phương pháp học tập hiện đại.

Loại Trường và Lớp Học (SchoolType & ClassType): Loại trường và lớp học có thể ảnh hưởng đến môi trường học tập và chất lượng giảng dạy. Các trường và lớp học chất lượng cao thường cung cấp môi trường học tập tốt hơn, giáo viên có trình độ cao hơn, và tài liệu học tập phong phú hơn.

Tần suất tự học (SelfStudy): Tần suất tự học cao thể hiện sự chăm chỉ và chủ động trong việc học, có thể dẫn đến điểm số tốt hơn. Tự học giúp học sinh rèn luyện kỹ năng tự quản lý và khả năng tự giải quyết vấn đề.

Tình trạng sức khỏe của học sinh (Health): Sức khỏe tốt giúp học sinh tập trung và học tập hiệu quả hơn. Một cơ thể khỏe mạnh tạo điều kiện tốt cho việc học tập và nâng cao khả năng tiếp thu kiến thức.

Điểm số trước đó (LatestScore): Điểm số năm trước thường có mối liên hệ với khả năng học tập và điểm số hiện tại. Học sinh có thành tích tốt trong quá khứ thường duy trì được phong độ học tập và đạt điểm cao hơn.

Truy cập tài nguyên học tập (ResourceAccessed): Truy cập nhiều tài nguyên học tập giúp mở rộng kiến thức và cải thiện điểm số. Tài nguyên học tập bao gồm sách vở, bài giảng, video hướng dẫn và các nguồn thông tin trực tuyến.

Tham gia diễn đàn (ForumParticipationFrequency): Tham gia tích cực vào diễn đàn thể hiện sự quan tâm đến việc học và có thể giúp hiểu bài sâu hơn. Thảo luận và chia sẻ kiến thức trên diễn đàn giúp học sinh học hỏi từ các bạn cùng lớp và giải quyết các thắc mắc một cách hiệu quả.

Câu trả lời hữu ích (HelpfulAnswers): Câu trả lời hữu ích chứng tỏ kiến thức vững vàng và khả năng áp dụng kiến thức tốt. Học sinh có khả năng trả lời các câu hỏi một cách chính xác và hiệu quả thường có kiến thức sâu rộng và kỹ năng phân tích tốt.

**Hoàn thành bài tập (ExercisesCompleted):** Làm nhiều bài tập giúp luyện tập và củng cố kiến thức. Bài tập là phương pháp rèn luyện kỹ năng và kiểm tra lại kiến thức đã học, giúp học sinh làm quen với các dạng bài tập và câu hỏi khác nhau.

**Điểm trung bình bài tập (AverageExerciseScore):** Điểm trung bình bài tập cao cho thấy khả năng làm bài tốt và hiểu bài sâu. Học sinh đạt điểm cao trong các bài tập thường có khả năng nắm vững kiến thức và áp dụng chúng một cách hiệu quả.

**Hoàn thành bài kiểm tra (TestsCompleted):** Làm nhiều bài kiểm tra giúp làm quen với dạng đề và áp lực thi cử. Kiểm tra là phương pháp đánh giá khả năng học tập và sự chuẩn bị cho các kỳ thi chính thức.

**Điểm trung bình bài kiểm tra (AverageTestScore):** Điểm trung bình bài kiểm tra phản ánh trực tiếp khả năng làm bài thi. Học sinh có điểm trung bình cao trong các bài kiểm tra thường có khả năng ứng dụng kiến thức và kỹ năng làm bài thi tốt.

**Thời gian hoàn thành bài kiểm tra (AverageTestCompletionTime):** Hoàn thành bài kiểm tra nhanh với điểm số cao cho thấy khả năng làm bài tốt và tốc độ xử lý thông tin nhanh. Học sinh có kỹ năng quản lý thời gian và làm bài hiệu quả thường đạt kết quả tốt hơn.

**Tham gia đấu trường (ArenaParticipation):** Tham gia đấu trường giúp luyện tập kỹ năng làm bài dưới áp lực thời gian. Đấu trường là môi trường học tập cạnh tranh, giúp học sinh rèn luyện kỹ năng làm bài nhanh và chính xác.

**Tỷ lệ trả lời đúng trong đấu trường (ArenaQuestionsAnsweredPercentage):** Tỷ lệ trả lời đúng cao trong đấu trường cho thấy khả năng xử lý câu hỏi nhanh và chính xác. Học sinh có tỷ lệ trả lời đúng cao thường có kiến thức vững vàng và kỹ năng làm bài tốt.

**Số lần đạt top 1 trong đấu trường (BestOfArenaCount):** Thường xuyên đạt top 1 trong đấu trường cho thấy khả năng vượt trội so với các bạn cùng học. Học sinh đạt top 1 thường có kỹ năng làm bài xuất sắc và kiến thức sâu rộng.

**Xếp hạng trung bình trong đấu trường (AverageArenaRanking):** Xếp hạng trung bình cao trong đấu trường thể hiện sự ổn định trong việc làm bài và kiến thức vững vàng. Học sinh có xếp hạng trung bình cao thường duy trì được phong độ học tập và đạt kết quả tốt.

Số ngày đã hoạt động trên web (DayActive): Số ngày hoạt động trên web nhiều có thể thể hiện sự chăm chỉ và quan tâm đến việc học. Học sinh dành nhiều thời gian trên web học tập thường có ý thức tự giác và nỗ lực trong học tập.

Xu hướng hoạt động gần đây (Increase): Hoạt động gần đây tăng lên có thể cho thấy sự nỗ lực và tiến bộ trong học tập. Học sinh có xu hướng hoạt động tăng thường có động lực và quyết tâm cao trong việc đạt kết quả tốt.

#### \* *Mối Quan Hệ Tiêu Cực*

Tần suất đi chơi (GoingOut): Dành nhiều thời gian đi chơi có thể ảnh hưởng đến thời gian học tập và kết quả học tập. Việc thiếu thời gian học tập do đi chơi nhiều có thể dẫn đến việc không nắm vững kiến thức và giảm điểm số.

Xu hướng hoạt động gần đây (Decrease): Hoạt động gần đây giảm xuống có thể là dấu hiệu của việc lơ là học tập. Học sinh có xu hướng hoạt động giảm thường thiếu sự chăm chỉ và quan tâm đến việc học, dẫn đến kết quả học tập kém.

#### \* *Có Thể Tăng Hoặc Giảm*

Độ tuổi (Age): Độ tuổi có thể ảnh hưởng đến sự trưởng thành và khả năng học tập, nhưng không phải yếu tố có sức ảnh hưởng quá lớn. Học sinh lớn tuổi hơn có thể có kỹ năng học tập tốt hơn, nhưng cũng có thể chịu áp lực từ nhiều yếu tố khác.

Khối lớp (Grade): Khối lớp có thể liên quan đến độ khó của kiến thức, nhưng không phải yếu tố có sức ảnh hưởng quá lớn. Học sinh ở các khối lớp khác nhau có thể gặp các thử thách khác nhau trong việc học tập.

Thời gian sử dụng trang Web mỗi ngày (DailyTimeSpent): Thời gian dành cho web mỗi ngày có thể có cả tác động tích cực (học tập) và tiêu cực (giải trí). Học sinh sử dụng web để học tập có thể cải thiện điểm số, nhưng nếu sử dụng quá nhiều cũng sẽ không tốt.

Thời gian học chủ yếu trong ngày (MostTime): Một số nghiên cứu cho thấy con người có thể tập trung và sáng tạo tốt hơn vào ban đêm, tuy nhiên nhịp sinh học là chu kỳ tự nhiên của cơ thể nên yếu tố này sẽ không ảnh hưởng quá nhiều đến điểm thi. Học sinh cần tìm thời gian học tập phù hợp với nhịp sinh học của mình để đạt hiệu quả tốt nhất.

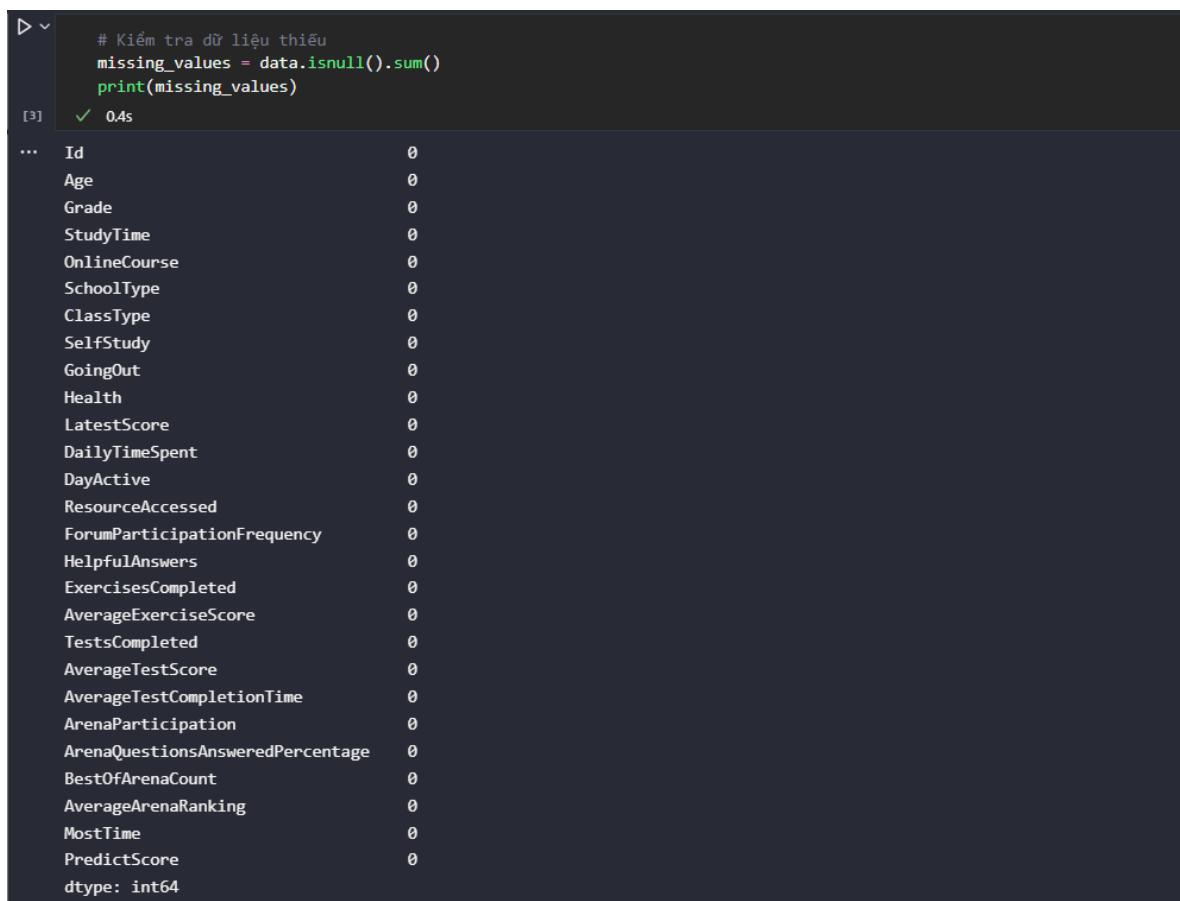
### **3.2 Tiền xử lý dữ liệu**

Sau khi hoàn thành quá trình thu thập dữ liệu mẫu, bước tiếp theo là tiến hành xử lý tập dữ liệu này. Các bước tiền xử lý dữ liệu thông thường bao gồm: xử lý dữ liệu

bị thiếu (missing values), mã hóa dữ liệu dạng chữ sang dạng số (categorical data encoding), chuẩn hóa dữ liệu (data normalization/standardization), loại bỏ các ngoại lệ (outlier removal), chia tách dữ liệu (data splitting), giảm số chiều của dữ liệu (dimensionality reduction), và áp dụng các kỹ thuật phân tích đặc trưng (feature engineering/feature scaling).

### Bước 1: Xử lý dữ liệu thiếu

Xử lý dữ liệu thiếu là một bước quan trọng trong quá trình tiền xử lý dữ liệu. Các phương pháp phổ biến để xử lý dữ liệu thiếu bao gồm loại bỏ các mẫu hoặc thuộc tính chứa giá trị thiếu, hoặc áp dụng các kỹ thuật thay thế giá trị thiếu bằng phương pháp trung bình, phương pháp trung vị, hoặc các kỹ thuật suy luận tiên tiến hơn. Điều này đảm bảo rằng dữ liệu đầu vào không bị sai lệch do thiếu thông tin.



```
# Kiểm tra dữ liệu thiếu
missing_values = data.isnull().sum()
print(missing_values)

[3]    ✓ 0.4s
...
...  Id          0
Age          0
Grade         0
StudyTime     0
OnlineCourse  0
SchoolType    0
ClassType     0
SelfStudy     0
GoingOut      0
Health         0
LatestScore   0
DailyTimeSpent 0
DayActive      0
ResourceAccessed 0
ForumParticipationFrequency 0
HelpfulAnswers 0
ExercisesCompleted 0
AverageExerciseScore 0
TestsCompleted 0
AverageTestScore 0
AverageTestCompletionTime 0
ArenaParticipation 0
ArenaQuestionsAnsweredPercentage 0
BestOfArenaCount 0
AverageArenaRanking 0
MostTime       0
PredictScore   0
dtype: int64
```

Hình 3. 1 Kiểm tra dữ liệu thiếu

Tập dữ liệu không chứa giá trị thiếu, bước này có thể được bỏ qua.

### Bước 2: Mã hóa các biến phân loại

Mã hóa dữ liệu phân loại sang dạng số là cần thiết để các mô hình học máy có thể xử lý. Các kỹ thuật mã hóa bao gồm mã hóa nhãn (label encoding) và mã hóa một-nóng (one-hot encoding). Việc sử dụng LabelEncoder từ thư viện sklearn.preprocessing để

mã hóa các biến phân loại thành các giá trị số giúp cho việc xử lý dữ liệu trở nên hiệu quả và phù hợp hơn với các mô hình học máy.

```

▷ ▾
    # Mã hóa các biến phân loại
    categorical_columns = ['OnlineCourse', 'SchoolType', 'ClassType', 'MostTime']
    label_encoders = {}
    for col in categorical_columns:
        le = LabelEncoder()
        data[col] = le.fit_transform(data[col])
        label_encoders[col] = le
    print("\nDữ liệu sau khi mã hóa các biến phân loại:")
    print(data[categorical_columns])

[32] ✓ 0.1s
...
Dữ liệu sau khi mã hóa các biến phân loại:
   OnlineCourse  SchoolType  ClassType  MostTime
0              1          1          0          1
1              0          2          1          1
2              1          0          0          1
3              0          2          0          1
4              0          2          0          0
...

```

*Hình 3. 2 Mã hóa các biến phân loại dạng text sang number*

### Bước 3: Chuẩn hóa và loại bỏ ngoại lệ

Chuẩn hóa dữ liệu (normalization/standardization) là quá trình đưa các thuộc tính về cùng một thang đo nhằm cải thiện hiệu quả và tốc độ hội tụ của mô hình. Loại bỏ các giá trị ngoại lệ bằng cách sử dụng điểm z-score giúp giảm thiểu tác động của các giá trị bất thường lên mô hình, từ đó nâng cao độ chính xác của dự đoán.

```

# Chuẩn hóa các đặc trưng và xóa bỏ ngoại lệ

numerical_columns = ['Age', 'Grade', 'StudyTime', 'SelfStudy', 'GoingOut', 'Health', 'AverageTestScore', 'AverageExerciseScore', 'LatestScore', 'TestsCompleted']
scaler = StandardScaler()
data[numerical_columns] = scaler.fit_transform(data[numerical_columns])

data = data[(np.abs(zscore(data[numerical_columns])) < 3).all(axis=1)]
print(data.head())

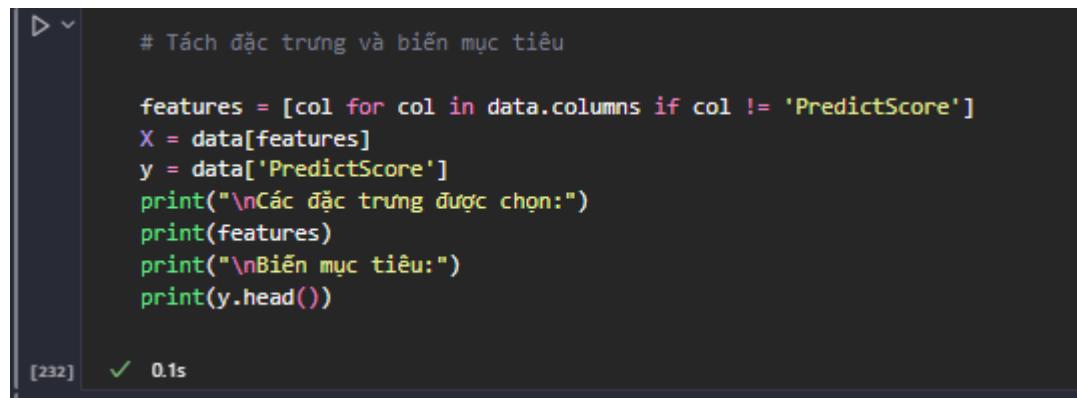
✓ 0.6s

```

*Hình 3. 3 Chuẩn hóa và loại bỏ ngoại lệ*

### Bước 4: Tách các đặc trưng và biến mục tiêu

Việc tách các đặc trưng (features) và biến mục tiêu (target) là bước cần thiết để xác định rõ ràng các biến độc lập và biến phụ thuộc trong mô hình học máy. Các đặc trưng được lựa chọn sẽ được đưa vào mô hình, trong khi biến mục tiêu là đối tượng cần dự đoán.



```
# Tách đặc trưng và biến mục tiêu

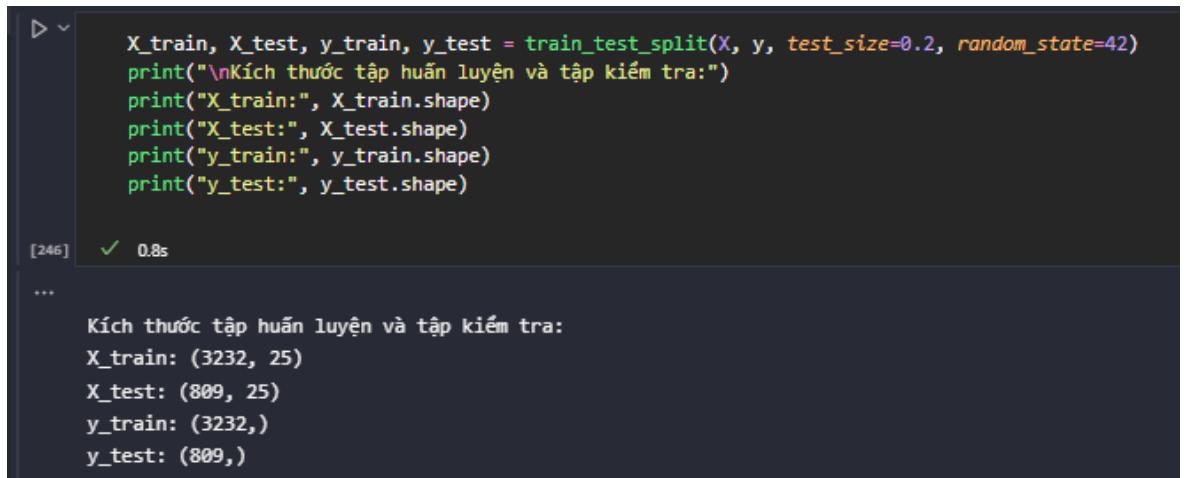
features = [col for col in data.columns if col != 'PredictScore']
X = data[features]
y = data['PredictScore']
print("\nCác đặc trưng được chọn:")
print(features)
print("\nBiến mục tiêu:")
print(y.head())

[232] ✓ 0.1s
```

Hình 3. 4 Tách đặc trưng và biến mục tiêu

#### Bước 5: Chia tách dữ liệu

Chia tách dữ liệu thành các tập huấn luyện và tập kiểm tra với tỷ lệ thông thường là 80-20 là một bước quan trọng để đánh giá hiệu suất của mô hình trên dữ liệu chưa từng thấy trước đó. Điều này giúp đảm bảo rằng mô hình không bị overfit vào tập huấn luyện và có thể tổng quát hóa tốt trên các dữ liệu mới.



```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print("\nKích thước tập huấn luyện và tập kiểm tra:")
print("X_train:", X_train.shape)
print("X_test:", X_test.shape)
print("y_train:", y_train.shape)
print("y_test:", y_test.shape)

[246] ✓ 0.8s
...
Kích thước tập huấn luyện và tập kiểm tra:
X_train: (3232, 25)
X_test: (809, 25)
y_train: (3232,)
y_test: (809,)
```

Hình 3. 5 Chia tập dữ liệu

Mỗi bước trong quy trình tiền xử lý dữ liệu đều đóng vai trò quan trọng trong việc chuẩn bị dữ liệu cho mô hình học máy. Việc xử lý dữ liệu thiếu, mã hóa dữ liệu phân loại, chuẩn hóa và loại bỏ ngoại lệ, chọn đặc trưng và chia tách dữ liệu đều góp phần làm cho dữ liệu phù hợp hơn với các thuật toán học máy, từ đó cải thiện hiệu suất và độ chính xác của mô hình dự đoán.

### 3.3 Ứng dụng thuật toán Random Forest

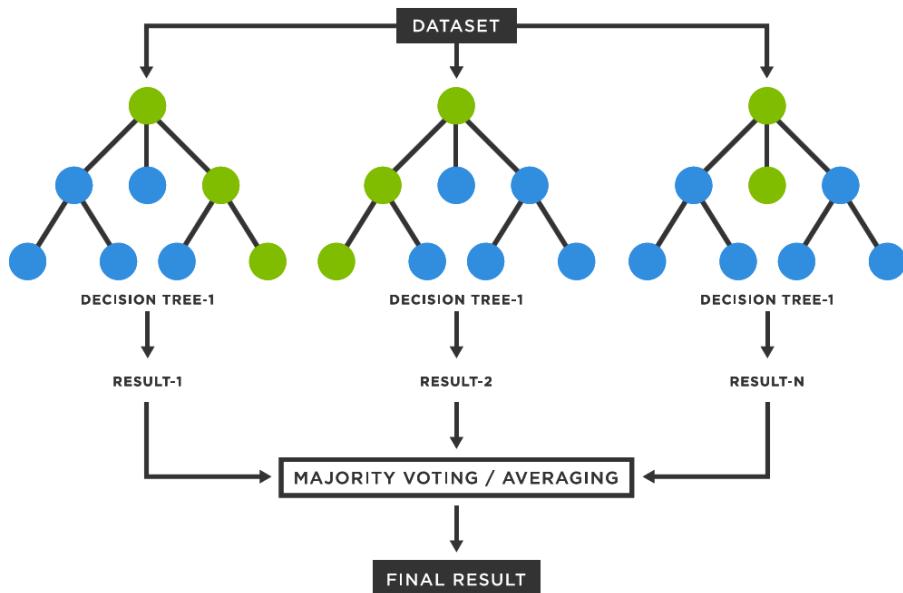
#### 3.3.1 Cây quyết định và rừng ngẫu nhiên

Random Forest là một thuật toán học máy sử dụng phương pháp Bagging để kết hợp nhiều cây quyết định nhằm cải thiện độ chính xác và giảm thiểu overfitting.

Trên thực tế, một khu “rừng” sẽ phải có thật nhiều “cây”, càng nhiều cây thì khu rừng càng lớn, càng phát triển mạnh. Cũng giống như vậy, về bản chất thì Random Forest là một tập hợp của nhiều cây quyết định, sau khi mỗi cây quyết định đưa ra một dự đoán riêng lẻ, Random Forest sẽ tổng hợp kết quả bằng cách lấy trung bình hoặc lấy kết quả được dự đoán nhiều nhất từ các cây quyết định thành phần.

Cây quyết định sẽ phân chia không gian biến số dựa trên các ngưỡng giá trị của các biến đầu vào sao cho mỗi nút lá (leaf node) của cây dự đoán một giá trị số. Khi tiến hành dự đoán, cây quyết định sẽ lan truyền giá trị từ gốc cây xuống đến lá tương ứng dựa trên các quy tắc phân chia đã được chỉ định từ trước.

Sau khi cây quyết định đưa ra kết quả dự đoán mà nó đã được học, đôi khi trong 1 mô hình cây với tập mẫu nào đó có thể dẫn đến dự đoán bị sai lệch. Vì thế, nếu ta xây dựng nhiều cây quyết định, mỗi cây sẽ đưa ra một dự đoán và tổng hợp kết quả lại, dự đoán cuối cùng sẽ là dự đoán chính xác nhất. Đó cũng chính là nguyên tắc của rừng ngẫu nhiên.



Hình 3. 6 Ví dụ về cây quyết định và rừng ngẫu nhiên

### 3.3.2 Ứng dụng xây dựng mô hình Random Forest

#### Bước 1: Bootstrap Sampling

Tập dữ liệu ban đầu có N mẫu với các đặc trưng (features) như sau: Age, Grade, StudyTime, OnlineCourse, SchoolType, LatestScore, ClassType, SelfStudy, GoingOut, Health, AverageTestScore, AverageExerciseScore, DailyTimeSpent,

ArenaQuestionsAnsweredPercentage, TestsCompleted, MostTime. Và biến mục tiêu là PredictScore.

Giả sử trong tập dữ liệu này có số lượng bản ghi là  $N=100$  và có số lượng đặc trưng  $M=16$ . Để tạo ra một tập bootstrap, chúng ta lấy mẫu ngẫu nhiên  $N$  lần với thay thế từ tập dữ liệu ban đầu. Ví dụ, giả sử sau khi bootstrap sampling, chúng ta có tập con sau:

Bảng 3. 3 Ví dụ về tập dữ liệu

Age	Grade	StudyTime	...	TestsCompleted	PredictScore
17	11	4	...	34	9.2
16	10	3	...	29	8.5
17	11	4	...	34	9.2
...	...	...	...	...	...
18	12	5	...	38	9.7

## Bước 2: Xây Dựng Cây Quyết Định từ Bootstrap Samples

Tại mỗi nút của cây quyết định, thay vì xem xét tất cả  $M$  đặc trưng, chúng ta chỉ xem xét một tập hợp con ngẫu nhiên của các đặc trưng (giả sử  $m = \sqrt{M}$ ).

Giả sử tại một nút, chúng ta chọn ngẫu nhiên 4 đặc trưng từ 16 đặc trưng. Để chọn đặc trưng tốt nhất để phân chia, ta tính toán Gini Impurity hoặc Mean Squared Error (MSE) cho từng giá trị ngưỡng.

\* *Gini Impurity*

Giả sử chúng ta đang phân chia dựa trên đặc trưng AverageTestScore (đặc trưng mạnh nhất trong tập dữ liệu), chúng ta tính Gini Impurity cho từng giá trị ngưỡng  $t$  theo công thức:

$$\text{Gini}(S) = 1 - \sum_{i=1}^e p_i^2$$

Trong đó  $p_i$  là tỷ lệ phần tử thuộc lớp  $i$  trong tập dữ liệu  $S$ .

Ví dụ, giả sử chúng ta có 2 giá trị ngưỡng để phân chia AverageTestScore: 5.0, và 8.0. Cần phải tính Gini Impurity cho từng giá trị này:

$$\text{Gini}(S | \text{AverageTestScore} < 8.0) = 1 - (0.6)^2 - (0.4)^2 = 0.48$$

$$\text{Gini}(S | \text{AverageTestScore} \geq 8.0) = 1 - (0.7)^2 - (0.3)^2 = 0.42$$

Tính tổng Gini Impurity:

$$\begin{aligned} Gini_{split} = & \left( \frac{n_{<8.0}}{n} \right) * Gini(S | AverageTestScore < 8.0) \\ & + \left( \frac{n_{\geq 8.0}}{n} \right) * Gini(S | AverageTestScore \geq 8.0) \end{aligned}$$

Tính tương tự với ngưỡng 5.0 và tất cả các đặc trưng được chọn khác. Chọn đặc trưng và ngưỡng phân chia mà có Gini Impurity nhỏ nhất.

Tại mỗi nút được phân chia, lặp lại quá trình lựa chọn đặc trưng và ngưỡng để phân chia dữ liệu con. Quá trình này tiếp tục cho đến khi đạt được các leaf node (nút lá), trong đó mỗi leaf node sẽ dự đoán giá trị của biến mục tiêu (ở đây là PredictScore).

#### \* Mean Squared Error (MSE)

Giả sử chúng ta đang phân chia dựa trên đặc trưng AverageTestScore, chúng ta tính MSE cho từng giá trị ngưỡng t.

$$MSE(S) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Trong đó:

- $y_i$  là giá trị thực của biến mục tiêu (PredictScore).
- $\bar{y}$  là giá trị trung bình của biến mục tiêu trong tập dữ liệu S.
- $n$  là số lượng mẫu trong tập dữ liệu S

Ví dụ, chúng ta có giá trị ngưỡng 8.0 để phân chia đặc trưng AverageTestScore.

Tập dữ liệu S sẽ được chia thành hai tập con:

- $S_{left}$ : Các mẫu có  $AverageTestScore < 8.0$
- $S_{right}$ : Các mẫu có  $AverageTestScore \geq 8.0$

Tính MSE cho từng tập con:

- Tính MSE cho  $S_{left}$ :

$$MSE(S_{left}) = \frac{1}{|S_{left}|} \sum_{i \in S_{left}} (y_i - \bar{y}_{left})^2$$

- Tính MSE cho  $S_{right}$ :

$$MSE(S_{right}) = \frac{1}{|S_{right}|} \sum_{i \in S_{right}} (y_i - \bar{y}_{right})^2$$

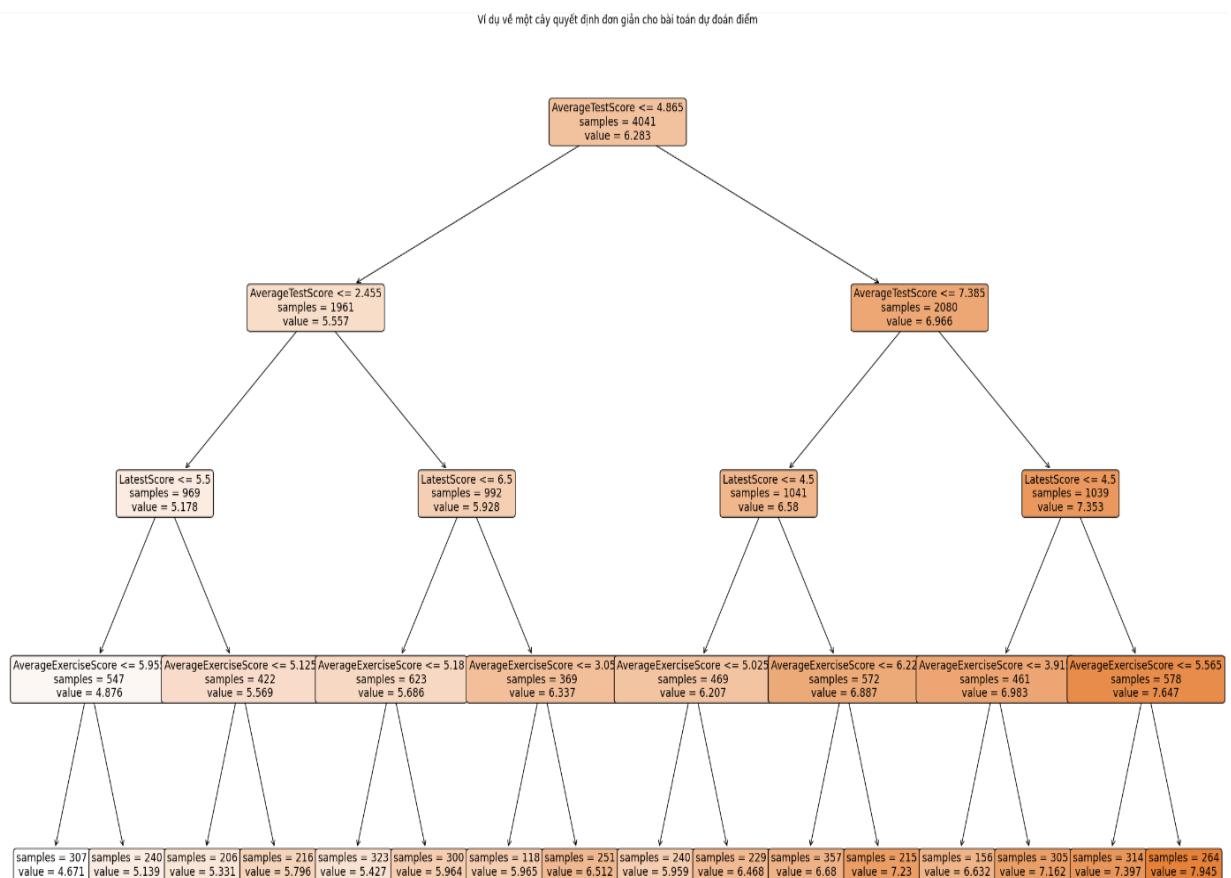
Tính tổng MSE sau khi phân chia ngưỡng:

$$MSE_{split} = \frac{n_{left}}{n} \cdot MSE(S_{left}) + \frac{n_{right}}{n} \cdot MSE(S_{right})$$

Chúng ta lặp lại các bước trên cho tất cả các giá trị ngưỡng của đặc trưng AverageTestScore, cũng như cho tất cả các đặc trưng khác được chọn ngẫu nhiên. Đặc trưng và ngưỡng phân chia tốt nhất sẽ là đặc trưng và ngưỡng có tổng MSE nhỏ nhất.

Tại mỗi nút, sau khi chọn được đặc trưng và ngưỡng phân chia tốt nhất, chúng ta phân chia tập dữ liệu tại nút hiện tại thành hai tập con:  $S_{left}$  và  $S_{right}$ . Sau đó, lặp lại quá trình lựa chọn đặc trưng và ngưỡng để phân chia dữ liệu con tại mỗi nút mới cho đến khi đạt được các nút lá. Các nút lá sẽ chứa các mẫu dữ liệu mà không thể phân chia thêm nữa. Tại mỗi nút lá, giá trị dự đoán của biến mục tiêu (PredictScore) sẽ là giá trị trung bình của biến mục tiêu trong tập dữ liệu con tại nút lá đó.

Dưới đây là một ví dụ đơn giản về cây quyết định đơn giản.



Hình 3. 7 Ví dụ về một cây quyết định trong bài toán dự đoán điểm

### Bước 3: Xây Dựng Nhiều Cây Quyết Định

Lặp lại bước 1 và bước 2 để xây dựng thêm nhiều cây quyết định. Ở đây chúng ta xây dựng 100 cây quyết định (tức là  $n\_estimators = 100$  trong scikit-learn).

#### Bước 4: Tổng Hợp Kết Quả từ Nhiều Cây Quyết Định

Sau khi xây dựng xong rừng ngẫu nhiên, chúng ta dự đoán bằng cách tổng hợp kết quả từ tất cả các cây quyết định. Với bài toán hồi quy, kết quả dự đoán là trung bình của tất cả các dự đoán từ các cây quyết định:

$$\hat{y} = \sum_{t=1}^T \hat{y}_t$$

Trong đó  $T$  là số lượng cây quyết định và  $\hat{y}_t$  là dự đoán từ cây quyết định thứ  $t$ .

Ví dụ, giả sử 100 cây quyết định đưa ra các dự đoán như sau:

$$\hat{y}^1 = 9.0, \hat{y}_2 = 8.8, \dots, \hat{y}_{100} = 9.2$$

Kết quả dự đoán cuối cùng sẽ là trung bình của tất cả các dự đoán này:

$$\hat{y} = \left(\frac{1}{100}\right) * \sum_{t=1}^{100} \hat{y}_t = 9.0$$

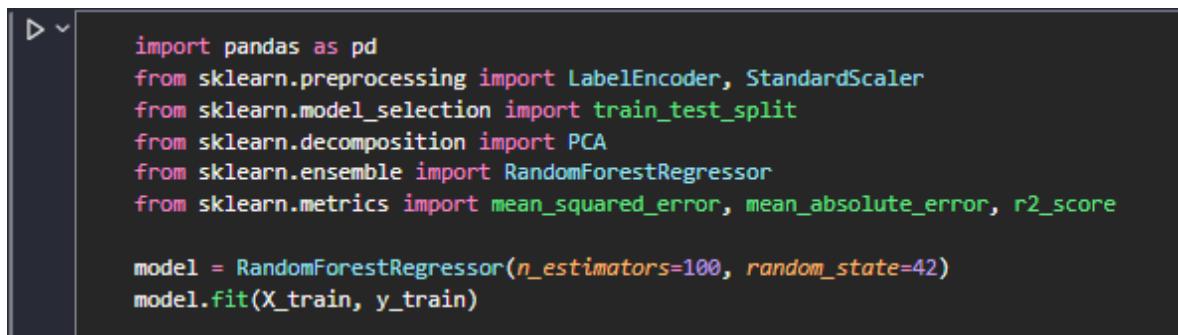
#### Tóm tắt các bước để xây dựng Random Forest:

- Bootstrap Sampling: Lấy mẫu ngẫu nhiên với thay thế từ tập dữ liệu ban đầu để tạo ra nhiều tập dữ liệu con.
- Xây Dựng Cây Quyết Định: Với mỗi tập dữ liệu con, xây dựng một cây quyết định bằng cách sử dụng một tập hợp con ngẫu nhiên của các đặc trưng tại mỗi nút để chọn đặc trưng tốt nhất để phân chia.
- Tổng Hợp Kết Quả: Dự đoán điểm thi bằng cách lấy trung bình các dự đoán từ tất cả các cây quyết định.

### 3.4 Xây dựng và đánh giá mô hình

#### 3.4.1 Xây dựng và huấn luyện mô hình

Scikit-learn là một công cụ mạnh mẽ và linh hoạt cho việc xây dựng mô hình học máy, đặc biệt là trong việc triển khai mô hình Random Forest. Việc sử dụng scikit-learn giúp tối ưu hóa quá trình xây dựng mô hình, từ chuẩn bị dữ liệu đến đào tạo mô hình, với tính dễ dàng và hiệu quả cao. Bên cạnh đó, việc tận dụng các tính năng linh hoạt và tài liệu phong phú của scikit-learn giúp người dùng tùy chỉnh và tinh chỉnh mô hình theo nhu cầu cụ thể nào đó.



```

import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

```

*Hình 3. 8 Huấn luyện mô hình với Random Forest trong Sklearn*

Đoạn mã trên là những bước cốt lõi trong quá trình xây dựng và huấn luyện mô hình Random Forest sử dụng thư viện scikit-learn, một trong những công cụ phổ biến cho việc phát triển các mô hình học máy. Mã này đóng vai trò quan trọng trong việc tạo ra một mô hình có khả năng dự đoán chính xác từ dữ liệu huấn luyện.

Đầu tiên, trong đoạn mã đầu tiên, chúng ta khởi tạo một mô hình Random Forest bằng cách sử dụng lớp RandomForestRegressor từ thư viện scikit-learn. Random Forest là một phương pháp học máy cơ bản, là một tập hợp của nhiều cây quyết định, mỗi cây đưa ra một dự đoán riêng và kết quả cuối cùng là kết hợp của các dự đoán này.

Thông qua tham số n\_estimators=100, chúng ta xác định số lượng cây quyết định sẽ được tạo ra trong mô hình. Điều này giúp tăng tính đa dạng và ổn định của mô hình, giảm thiểu hiện tượng quá mức phù hợp (overfitting) và nâng cao khả năng tổng quát hóa của mô hình.

Tham số random\_state=42 đảm bảo tính tái lập của quá trình huấn luyện mô hình. Bằng cách cố định giá trị random\_state, chúng ta có thể đảm bảo rằng mỗi lần chạy mã, chúng ta sẽ nhận được kết quả nhất quán, từ đó dễ dàng so sánh và đánh giá các mô hình khác nhau.

Tiếp theo, trong đoạn mã thứ hai, chúng ta tiến hành huấn luyện mô hình Random Forest trên tập dữ liệu huấn luyện, được đại diện bởi ma trận đặc trưng X\_train và vector mục tiêu y\_train. Phương thức fit() được gọi để bắt đầu quá trình huấn luyện.

Trong quá trình huấn luyện, scikit-learn tự động thực hiện các bước quan trọng như lấy mẫu Bootstrap và xây dựng các cây quyết định từ các tập dữ liệu con. Sau khi hoàn thành quá trình huấn luyện, mô hình Random Forest đã sẵn sàng để dự đoán giá trị mục tiêu cho các mẫu dữ liệu mới, đưa ra những dự đoán có tính chất tổng quát và chính xác.

### 3.4.2 Đánh giá hiệu suất của mô hình

Sau khi huấn luyện mô hình, chúng ta có thể tiến hành dự đoán trên tập kiểm tra  $y_{\text{test}}$ , sau đó sử dụng 1 số chỉ số để đánh giá hiệu suất và độ chính xác của mô hình như sai số toàn phương trung bình (MSE), sai số tuyệt đối trung bình và hệ số xác định  $R^2$ .

```

import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Dự đoán trên tập kiểm tra
y_pred = model.predict(X_test)

# Đánh giá mô hình
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'MSE: {mse}')
print(f'MAE: {mae}')
print(f'R²: {r2}')

```

[266] ✓ 0.9s

... MSE: 0.09005562945611864  
MAE: 0.2407400494437577  
R²: 0.9186146673801926

Hình 3. 9 Tính toán giá trị để đánh giá mô hình

\* MSE: đo lường độ chênh lệch trung bình bình phương giữa giá trị thực và giá trị dự đoán. Công thức tính MSE là:

$$MSE = \left(\frac{1}{n}\right) * \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Trong đó:

- $n$  là số lượng mẫu.
- $y_i$  là giá trị thực của mẫu thứ  $i$ .
- $\hat{y}_i$  là giá trị dự đoán của mẫu thứ  $i$ .

MSE càng nhỏ thì mô hình dự đoán càng chính xác. Trong trường hợp này, MSE = 0.09005562945611864, tức là sai số trung bình bình phương là rất nhỏ, cho thấy mô hình dự đoán khá chính xác.

\* MAE: đo lường độ chênh lệch trung bình tuyệt đối giữa giá trị thực và giá trị dự đoán. Công thức tính MAE là:

$$MAE = \left(\frac{1}{n}\right) * \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAE càng nhỏ thì mô hình dự đoán càng chính xác. Trong trường hợp này, MAE = 0.2407400494437577, tức là sai số trung bình tuyệt đối là khá nhỏ, cho thấy mô hình dự đoán khá chính xác.

\* R2 (R-squared): đo lường mức độ giải thích được của mô hình đối với biến mục tiêu. Công thức tính  $R^2$  là:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Trong đó:

- $\bar{y}$  là giá trị trung bình của biến mục tiêu.

$R^2$  dao động trong khoảng từ 0 đến 1. Giá trị  $R^2$  càng gần 1 thì mô hình dự đoán càng tốt. Trong trường hợp này,  $R^2 = 0.9186146673801926$ , tức là mô hình giải thích được 91.86% sự biến động của biến mục tiêu, cho thấy mô hình có độ chính xác cao.

## CHƯƠNG 4. XÂY DỰNG CHƯƠNG TRÌNH

### 4.1 Thiết kế giao diện

#### 4.1.1 Trang người dùng

The screenshot displays the GoUni platform's user interface. At the top, there is a navigation bar with links for 'Trang chủ' (Home), 'Hỏi đáp' (FAQ), 'Danh mục' (Categories), 'Đấu trường' (Arena), and 'Thống kê' (Statistics). A user profile for 'Châu Quế Nhơn @admin' is visible on the right.

**MỤC TIÊU HÔM NAY:**

- Thời gian ôn luyện: 178 / 240 phút (75%)
- Số đề đã giải: 0 / 5 đề (0%)
- Số bài tập đã làm: 0 / 5 bài (0%)
- Số phòng thi đã tham gia: 0 / 5 (0%)
- Điểm số tối thiểu: 0 / 8 (0%)
- Tỉ lệ làm đúng: 0 / 70% (0%)

**WEBSITE ĐÀNH RIÊNG CHO GIÁO VIÊN NEW:**

Lưu ý: Website chỉ sử dụng trên máy tính  
QUẢN LÝ TẤT CẢ LỚP HỌC  
TẠO PHÒNG THI ONLINE ĐỄ DÀNG  
THAM KHẢO HÀNG NGẦN TÀI LIỆU HAY  
[giaoVien.thitot.vn](http://giaoVien.thitot.vn)

**DANH MỤC:**

- Ngữ Văn
- Sinh Học
- Hóa Học
- Vật Lý
- GDCD
- Địa Lý
- Lịch Sử
- Tiếng Anh
- Toán Học

**PHÒNG THI ĐÀNH CHO BẠN:**

Xem thêm

**Quiz Section:**

Đề chính thức kỳ thi tốt nghiệp THPT năm 20...  
1

- Loại: đề thi
- Thời gian: 90 phút
- Số câu: 50 câu
- 0 người đã làm

**Lịch sử của tôi:**

- Bài tập tích phân
- Bài tập 19:46 15/06/2024
- Đề chinh phục 8+ Toán thầy Hồ Thức...
- Bài kiểm tra 18:3 15/06/2024
- CAO THỦ TOÁN HỌC
- Phòng thi đấu 13:9 12/06/2024
- Tên bài viết
- Phòng thi đấu 22:30 10/06/2024
- Tên bài viết

Hình 4. 1 Giao diện trang chủ

#### 4.1.1.1 Trang đấu trường

Tính năng đấu trường cho phép người dùng luyện tập thi thử với những người dùng khác, vừa mới mẻ gây hứng thú cho người học, vừa luyện tập khả năng giải đê nhanh và chính xác.

Với chức năng đấu trường, những thí sinh có thể tham gia và thi đấu realtime với nhau cùng một lúc. Khi bắt đầu làm bài, câu hỏi sẽ được phát cho tất cả thí sinh đã tham gia đấu trường. Các thí sinh sẽ phải giải câu đó đúng và thật nhanh để chọn đáp án chính xác. Nếu chọn đáp án sai, thí sinh đó sẽ bị mất lượt trả lời trong câu hỏi này. Nếu thí sinh chọn được đáp án chính xác và nhanh nhất sẽ được cộng điểm, đồng thời sẽ thông báo và chuyển tất cả thí sinh tham gia sang câu hỏi tiếp theo. Cứ như thế cho đến hết câu hỏi của bộ đề thi.

GoUni Trang chủ Hỏi đáp Danh mục Đầu trường Thông kê Châu Quế Nhơn

Trang Chủ / Kiểm tra / Thi online



### MỤC TIÊU HÔM NAY

Thời gian ôn luyện  
34 / 359 phút 10%

Số đề đã giải  
0 / 5 đề 0%

Số bài tập đã làm  
0 / 5 bài 0%

Số phòng thi đã tham gia  
0 / 5 0%

Điểm số tối thiểu  
0 / 0%

Tỉ lệ làm đúng  
0 % 0%

### Phòng thi

Bảng vàng Lịch sử

Tên phòng thi	Số câu	Thời gian	Loại phòng	Chế độ	Trạng thái
PEN - CUP BY HOCMAI	50 câu	1/25 thành viên 5 phút	Bắt đầu: 1 tuần sau	Loại phòng: Tự do	Đang chờ
Thi đấu Toán Học C-T-G MClass	47 câu	1/20 thành viên 15 phút	Bắt đầu: 3 ngày sau	Loại phòng: VIP	Đang chờ
THỦ SỨC CUỐI TUẦN	36 câu	3/10 thành viên 60 phút	Bắt đầu: 3 ngày sau	Loại phòng: VIP	Đang chờ
CAO THỦ TOÁN HỌC	4 câu	2/5 thành viên 1 phút	Bắt đầu: 2 ngày sau	Loại phòng: Tự do	Đang chờ
Tên bài viết	5 câu	4/5 thành viên 2 phút	Bắt đầu: 6 ngày trước	Loại phòng: Tự do	Đã thi xong
THỦ SỨC CUỐI TUẦN	39 câu	3/50 thành viên 10 phút	Bắt đầu: 1 tuần trước	Loại phòng: VIP	Đang diễn



Đề chinh phục 8+ Toán 2024 P2  
[Toán học]

Loại: đề thi  
Thời gian: 90 phút  
Số câu: 50 câu  
0 người đã làm

### Lịch sử của tôi

Đề chinh phục 8+ Toán 2024	18/3 15/06/2024
Bài kiểm tra	
CAO THỦ TOÁN HỌC	13:9 12/06/2024
Phòng thi đầu	
Tên bài viết	22:30 10/06/2024
Phòng thi đầu	
Tên bài viết	22:15 10/06/2024
Phòng thi đầu	

Hình 4. 2 Giao diện trang danh sách phòng thi

Trang Chủ / Kiểm tra / Thi online / PEN - CUP BY HOCMAI

### Thông tin phòng thi



**PEN - CUP BY HOCMAI**

4/25 + ❤️  
 Thi đấu ? 50 câu  5 phút/câu

Loại phòng: Công khai  
Trạng thái: Đã tham gia

**Bắt đầu ngay**

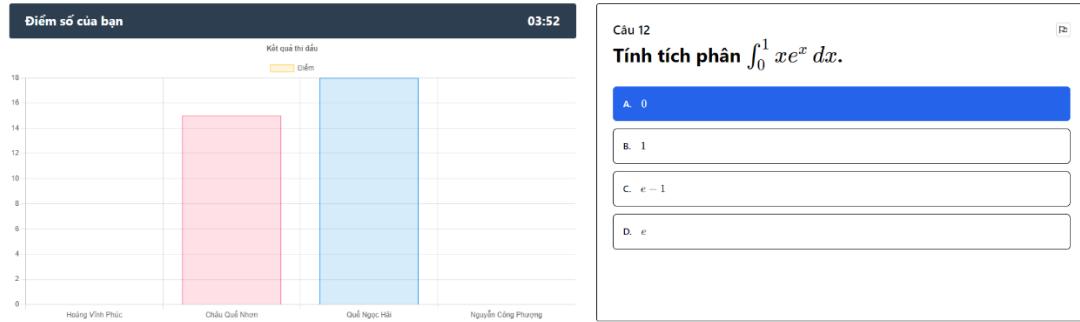
Rời khỏi phòng

### Danh sách thí sinh

Bắt đầu sau: 01:22:49:46

Hoàng Vĩnh Phúc @phuchv	Châu Quế Nhơn @admin	Quế Ngọc Hải @nhoncq2
Nguyễn Công Phương @nhoncq		

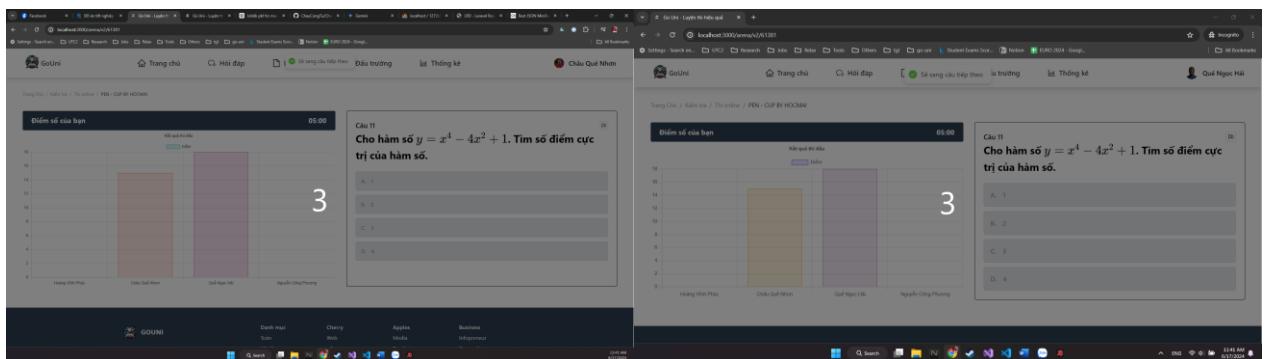
Hình 4. 3 Giao diện trang phòng chờ



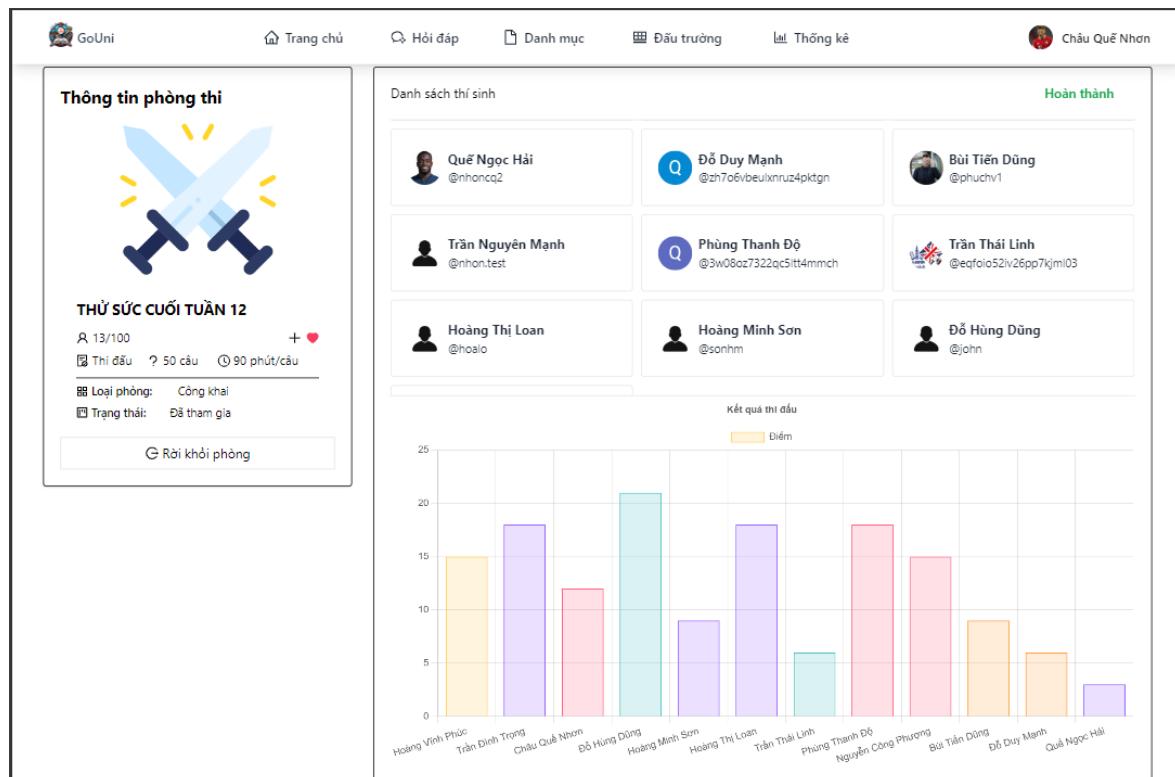
Câu 12  
Tính tích phân  $\int_0^1 xe^x dx$ .

- A. 0
- B. 1
- C.  $e - 1$
- D.  $e$

Hình 4. 4 Giao diện trang thi đấu

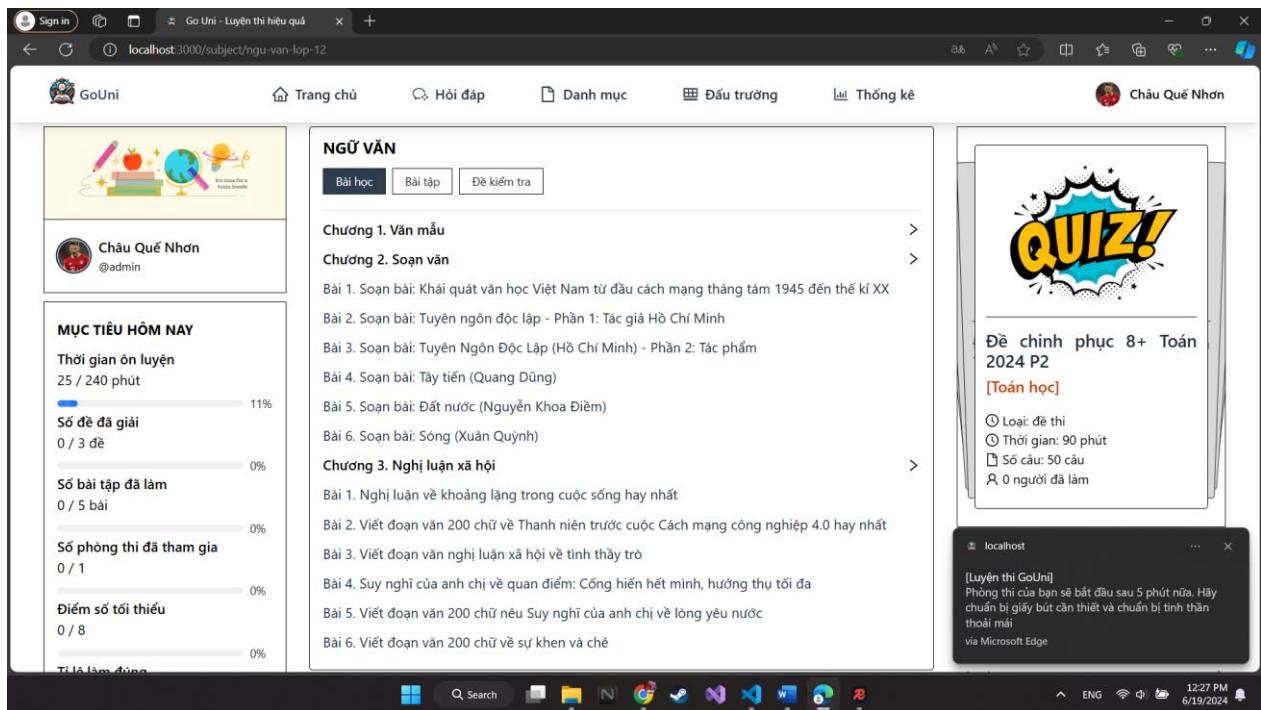


Hình 4. 5 Demo chuyển câu hỏi realtime



Hình 4. 6 Giao diện trang kết quả thi đấu

Trước thời gian bắt đầu phòng thi là 15 và 5 phút, hệ thống sẽ gửi các thông báo đầy đủ cho người dùng để nhắc nhở thi cho đúng thời gian. Người dùng chỉ cần click vào thông báo sẽ được đưa tới trang của phòng thi.



Hình 4. 7 Hình ảnh demo về realtime notification nhắc nhở giờ thi

#### 4.1.1.2 Trang tài liệu, bài học

The screenshot shows a Microsoft Edge browser window displaying the GoUni platform. The URL is Trang Chủ / Danh mục / Toán học. The main content area is titled 'TOÁN HỌC' and shows a list of chapters and exercises. On the right side, there is a large 'QUIZ!' box with the text 'Đề thi Vật Lý GoUni kì 1 - Mã đề 105 [Vật Lý]'. Below it, there are details about the exam: Type: thi, Time: 50 phút, Questions: 40 câu, and Participants: 0 người đã làm. At the bottom of the page, there is a section titled 'Môn học' with links for Ngữ Văn and Sinh Học.

Hình 4. 8 Giao diện trang danh sách bài học

**Ôn tập chương 1**

Cập nhật 17/06/2024 · 6 lượt xem · 0 hữu ích · [chia sẻ](#)

**1. Sự đồng biến, nghịch biến của hàm số**

Cho hàm số  $y=f(x)$ , khi đó:

- +  $f'(x)>0$  trên khoảng nào thì hàm số đồng biến trên khoảng đó.
- +  $f'(x)<0$  trên khoảng nào thì hàm số nghịch biến trên khoảng đó.

**Điều kiện để hàm số đồng biến, nghịch biến trên khoảng  $(a;b)$**

- + Để hàm số đồng biến trên khoảng  $(a,b)$  thì  $f'(x)\geq 0, \forall x\in(a,b)$ .
- + Để hàm số nghịch biến trên khoảng  $(a,b)$  thì  $f'(x)\leq 0, \forall x\in(a,b)$ .

**2. Cực trị của hàm số**

- [Chương 1. Ứng dụng đạo hàm để khảo sát và vẽ đồ thị của hàm số](#)
  - [Bài 1. Sự đồng biến, nghịch biến của hàm số](#)
  - [Bài 2. Cực trị của hàm số](#)
  - [Bài 3. Phương pháp giải bài toán cực trị có tham số đối với các hàm số cơ bản](#)
  - [Bài 4. Giá trị lớn nhất và giá trị nhỏ nhất của hàm số](#)
  - [Bài 5. Đồ thị của hàm số và phép tính tiến hệ tọa độ](#)
  - [Bài 6. Một số bài toán về khảo sát hàm số bậc ba, bậc bốn trùng phương](#)
  - [Bài 7. Phương pháp giải các bài toán tương giao đồ thi](#)
  - [Bài 8. Phương pháp giải các bài toán tiếp tuyến, sự tiếp xúc của hai đường cong](#)
  - [Bài 9. Ôn tập chương 1](#)
- [Chương 2. Hàm số lũy thừa, hàm số mũ và hàm số logarit](#)
- [Chương 3. Nguyên hàm, tích phân và ứng dụng](#)

**Hình 4. 9 Giao diện trang chi tiết bài học**

#### 4.1.1.3 Trang thảo luận

Học sinh tham gia ôn luyện có thể tạo các chủ đề, câu hỏi thảo luận để những thí sinh khác hoặc người hỗ trợ, quản trị viên sẽ cùng nhau trao đổi, hỗ trợ lẫn nhau, giúp quá trình ôn thi hiệu quả hơn.

**THẢO LUẬN**

Tạo chủ đề Đã like Hướng dẫn

**Trần Thái Linh** 17:53 16/06/2024 0

Quy tắc tính đạo hàm  
cho em hỏi mấy cái Quy tắc tính đạo hàm của hàm hợp chứng minh như thế nào vậy ạ?

**Hoàng Minh Sơn** 17:51 16/06/2024 0

Tuyển sinh đại học 2024: Thêm nhiều ngành mới  
[TẶNG BẠN] TRỌN BỘ Bí kíp học tốt 08 môn  
Chắc suất Đại học top - Giữ chỗ ngay!!  
...

**Trần Thị Linh** 17:48 16/06/2024 0

Lấy lại gốc môn Lý  
Chào mọi người, năm nay em học lớp 11 và theo học khối D. Từ lúc học c2 em đã xác định

**QUIZ!**

Đề thi lý thuyết Vật Lý THPT Quốc Gia  
**[Vật Lý]**

Loại: đề thi  
Thời gian: 60 phút  
Số câu: 45 câu  
0 người đã làm

**Lịch sử của tôi**

Đề chính phục 8+ Toán 2024  
Bài kiểm tra 18:3 15/06/2024

CAO THỦ TOÁN HỌC  
Phòng thi đầu 13:9 12/06/2024

THỦ SỨC CUỐI TUẦN 12

**Hình 4. 10 Giao diện trang thảo luận**

The screenshot shows a user profile for "Châu Quế Nhơn" (@admin) on the left. Below it is a section titled "MỤC TIÊU HÔM NAY" (Today's Goals) with various progress bars. The main content area displays a post titled "Mách học sinh Trung bình-khá cách đạt 8 điểm thi THPT quốc gia" by "Đỗ Hùng Dũng". The post includes a summary, a link to a file, and a link to a video. It also contains sections for "Xem thêm:" (See more), "Hà gục điểm 8 Vật lí trong 59 ngày", and "Điểm số tối thiểu". The right side features a sidebar with "Chủ đề liên quan" (Related topics) and several links.

Hình 4. 11 Giao diện trang chủ để thảo luận

A screenshot of a forum post titled "4 câu trả lời" (4 answers). The first answer is from "Quốc Ngọc Hải" at 17:24 16/06/2024, asking if anyone can help with learning Word from scratch. The second answer is from "Trần Nguyên Mạnh" at 17:24 16/06/2024, suggesting it's possible to learn it. The third answer is from "Nguyễn Công Phương" at 17:24 16/06/2024, stating that learning it from scratch is difficult. The fourth answer is from "Trần Thái Linh" at 17:24 16/06/2024, agreeing that it's very difficult. Below the answers is a "Thêm câu trả lời" (Add a reply) input field.

Hình 4. 12 Giao diện bình luận chủ đề thảo luận

#### 4.1.1.3 Trang bài tập

Tính năng giải bài tập cung cấp các đề ôn tập ngắn khoảng 10-30 câu cho học sinh để ôn luyện kiến thức cho từng chương học.

The screenshot shows a user profile for "Châu Quế Nhơn (@admin)" with a cartoon icon. Below it is a summary section titled "MỤC TIÊU HÔM NAY" (Today's Goals) showing completion percentages for various tasks:

- Thời gian ôn luyện: 203 / 359 phút (57%)
- Số đề đã giải: 0 / 5 đề (0%)
- Số bài tập đã làm: 0 / 5 bài (0%)
- Số phòng thi đã tham gia: 0 / 5 (0%)
- Điểm số tối thiểu: 0 / (0%)

The main menu on the right is titled "TOÁN HỌC" and includes tabs for "Bài học", "Bài tập" (selected), and "Đề kiểm tra". It lists chapters and sub-chapters:

- Chương 1. Ứng dụng đạo hàm để khảo sát và vẽ đồ thị của hàm số**
- Chương 2. Hàm số lũy thừa, hàm số mũ và hàm số logarit**
- Chương 3. Nguyên hàm, tích phân và ứng dụng**
- Chương 4. Số Phức**
- Chương 5. [Hình học] Khối đa diện và thể tích của chúng**

A separate box on the right is titled "Đề thi lý thuyết Vật Lý THPT Quốc Gia [Vật Lý]" with details: Loại: đề thi; Thời gian: 60 phút; Số câu: 45 câu; 0 người đã làm.

*Hình 4. 13 Giao diện trang bài tập*

Khi bắt đầu giải một tập bài, tiến trình sẽ được lưu lại cho đến khi người dùng nộp bài tập đó. Danh sách câu hỏi sẽ được phát ra dưới hình thức trắc nghiệm.

The worksheet is titled "Bài tập sóng cơ nâng cao" and includes the following details:

- Chí số: 29124
- Chế độ: Bài tập
- Đã làm: 2/10
- Thời gian làm bài: 05:10
- Buttons: "Làm lại" and "Nộp bài"

**Câu hỏi 1:**  
**Để tạo ra sóng dừng trên dây, điều kiện cần thiết là gì?**

A. Dây phải căng  
B. Dây phải mềm  
C. Hai đầu dây phải cố định (Selected)  
D. Hai đầu dây phải tự do

**Câu hỏi 2:**  
**Một sóng cơ có phương trình  $y = 0.01 \sin(2\pi t - 0.1x)$ . Tìm bước sóng của sóng.**

A. 10 m  
B. 20 m (Selected)  
C. 30 m  
D. 40 m

**Câu hỏi 3:**  
**Trong hiện tượng giao thoa sóng, hai nguồn sóng phải thỏa**

A sidebar on the right titled "Mục lục câu hỏi" contains numbered buttons from 1 to 10, where button 2 is highlighted in green.

*Hình 4. 14 Giao diện giải bài tập*

#### 4.1.1.4 Trang giải đề thi

Trang Chủ / Kiểm tra / Đề chinh phục 8+ Toán 2024

The screenshot shows a quiz interface. On the left, there's a large 'QUIZ!' logo. Below it, the title 'Đề chinh phục 8+ Toán 2024' is displayed, along with a progress bar showing 0 completed out of 50 questions, with a time of 84 phút | 8.8 điểm. Buttons for 'Bắt đầu thi' (Start test) and 'Tải về' (Download) are present. On the right, a 'Bảng xếp hạng' (Leaderboard) displays six users with their names, scores, and times:

#	Tên	Thời gian	Điểm
#1	Trần Nguyên Mạnh	84 phút	8.8 điểm
#2	Châu Quế Nhơn	82 phút	8.2 điểm (tối)
#3	Phùng Thành Độ	88 phút	8 điểm
#4	Đỗ Duy Mạnh	84 phút	7.8 điểm
#5	Đỗ Hùng Dũng	79 phút	7.8 điểm
#6	Nguyễn Công Phượng	79 phút	7.6 điểm

A message at the bottom of the leaderboard area says: 'Bạn đang ở chế độ xem trước, hãy bắt đầu ôn thi ngay nhé' (You are in preview mode, start studying now).

Hình 4. 15 Giao diện trang đề thi

Tính năng giải đề cung cấp các bài kiểm tra tuyển chọn theo chương trình học. Mỗi đề thi sẽ có số lượng câu hỏi và thời gian làm bài chuẩn theo quy chế thi trung học Phổ Thông Quốc Gia. Cơ chế đếm thời gian của tính năng bài tập và đề thi sẽ khác nhau, ở chức năng bài tập thì sẽ đếm xuôi thời gian và không giới hạn thời gian. Còn tính năng giải đề sẽ đếm ngược trong khoảng thời gian quy định.

Trang Chủ / Kiểm tra / Đề chinh phục 8+ Toán 2024

The screenshot shows a math problem: 'Câu hỏi 20: Tính tích phân:  $\int_0^1 e^x dx$ '. Below it are four answer options: A.  $e - 1$ , B.  $e$ , C. 1, and D.  $e + 1$ . Navigation buttons 'Trước' (Previous) and 'Sau' (Next) are at the bottom. To the right, a 'Mục lục câu hỏi' (List of questions) grid shows numbered boxes from 1 to 33, with box 20 highlighted in blue.

Hình 4. 16 Giao diện trang giải đề

#### 4.1.1.5 Trang cá nhân

##### a. Trang thông tin cá nhân

Trang Chủ / Châu Quế Nhơn

**Thông tin người dùng**

Mục tiêu ôn tập

Thống kê của tôi

**Tổng quan**

Họ tên Châu Quế Nhơn	Tên người dùng admin	Số điện thoại 0848611127	Địa chỉ Email admin@gouni.com.vn
Học lớp 12	Trường Trường THPT Số 1 An Nhơn	Khối theo học A	Khối ôn tập A A1
Ngày sinh 18/05/2002	Địa chỉ An Nhơn - Bình Định		

Student Admin Vip Teacher

Đổi ảnh đại diện

Vui lòng chọn ảnh nhỏ hơn 5MB  
Chọn hình ảnh phù hợp, không phản cảm

---

**Cá nhân**

Tên người dùng gouni@admin	Số điện thoại 0848611127
Giới tính Nam	Ngày sinh 18/05/2002

Hình 4. 17 Giao diện thông tin người dùng

### b. Trang mục tiêu cá nhân

**Thông tin người dùng**

Mục tiêu ôn tập

Thống kê của tôi

Châu Quế Nhơn  
@admin

**MỤC TIÊU HÔM NAY**

Thời gian ôn luyện hôm nay  
359 phút

Số đề sẽ giải  
5

Số bài tập sẽ làm  
5

Số phòng thi sẽ tham gia  
0

Điểm thấp nhất  
6.5

Tí lệ chính xác  
65 %

### MỤC TIÊU HÔM NAY

Thời gian ôn luyện  
253 / 359 phút

Số đề đã giải  
0 / 5 đề

Số bài tập đã làm  
0 / 5 bài

Số phòng thi đã tham gia  
0 /

Điểm số tối thiểu  
0 / 7

Tí lệ làm đúng  
0 / 65%

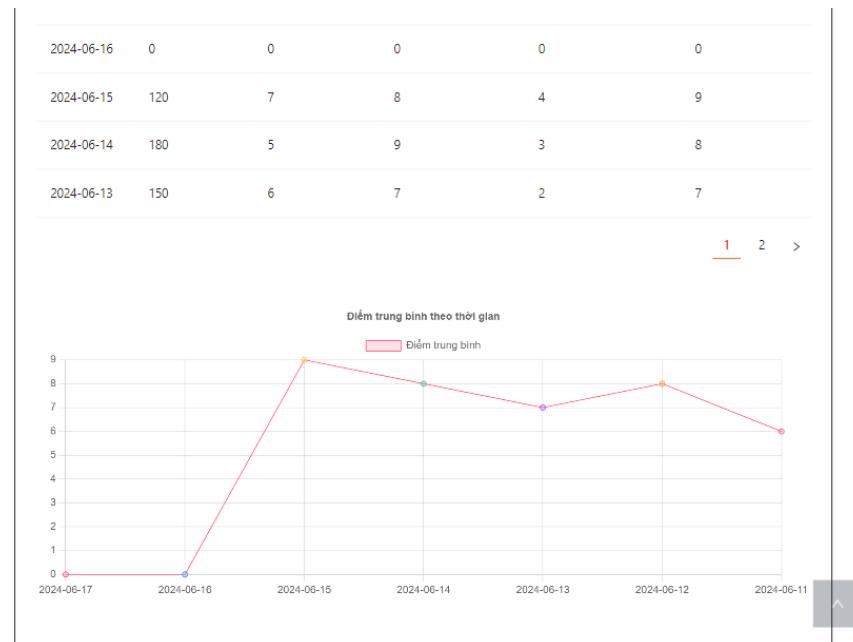
Hình 4. 18 Giao diện mục tiêu học tập

### c. Trang thống kê cá nhân

The screenshot shows a user profile for 'Châu Quế Nhơn' (born 18/05/2002, level A) and a statistics page. The sidebar on the left includes 'Thông tin người dùng', 'Mục tiêu ôn tập', and 'Thống kê của tôi'. The main area displays a summary of study activities over the last 7 days, followed by a detailed table and a line chart.

Ngày	Thời gian tổng	Số lượng bài thi	Số lượng luyện tập	Số lượng đấu trường	Điểm trung bình
2024-06-17	0	0	0	0	0
2024-06-16	0	0	0	0	0
2024-06-15	120	7	8	4	9

Hình 4. 19 Giao diện trang thống kê



Hình 4. 20 Giao diện thống kê bằng biểu

d. Trang lịch sử cá nhân

The screenshot shows a user profile at the top right with the name 'Châu Quế Nhơn'. The main content area displays a quiz history entry for 'Đề chinh phục 8+ Toán 2024' (Math Competition 8+ 2024). The entry includes the time '82:00', 'Số câu đúng' (Correct answers) '41', 'Điểm' (Score) '8.2', and a large green checkmark icon. Below this, there are three math questions with multiple-choice answers A, B, C, and D. To the right, there is a sidebar titled 'Lịch sử của tôi' (My History) listing past quizzes and exams.

Hình 4. 21 Giao diện trang lịch sử làm bài

#### 4.1.2 Trang Admin

##### 4.1.2.1 Giao diện xác thực

The three screenshots show the 'Đăng nhập' (Login), 'Đăng ký' (Registration), and 'Quên mật khẩu' (Forgot Password) pages. Each page has a red border and a header with the 'GoUni' logo. The 'Đăng nhập' page has fields for email and password, and social login buttons for Google and Facebook. The 'Đăng ký' page has similar fields and social login buttons. The 'Quên mật khẩu' page has a field for email and a 'Gửi yêu cầu' (Send request) button. All pages include links for 'Đăng nhập' and 'Quay về trang chủ' (Return to homepage).

Hình 4. 22 Giao diện trang xác thực

Chức năng xác thực người dùng bao gồm các tính năng quan trọng như đăng nhập tài khoản, đăng ký tài khoản và khôi phục mật khẩu. Đối với chức năng khôi phục mật khẩu, hệ thống sẽ gửi một liên kết đặt lại mật khẩu đến địa chỉ email của tài khoản người dùng. Mã token sẽ được lưu trữ trên Redis, và việc gửi email sẽ được đưa vào hàng đợi

(Queue), giúp quá trình xử lý khi người dùng nhấn nút "Gửi yêu cầu" không phải chờ đến khi email được gửi xong mới hoàn tất. Ngoài ra, hệ thống còn hỗ trợ tính năng đăng nhập thông qua các mạng xã hội phổ biến như Google, giúp người dùng đăng nhập dễ dàng hơn.

#### 4.1.2.2 Trang Dashboard

Trang Dashboard sẽ bao gồm các bảng, biểu đồ với các số liệu thống kê của cả hệ thống lẫn người dùng.

The screenshot shows the 'Dashboard' page with the following details:

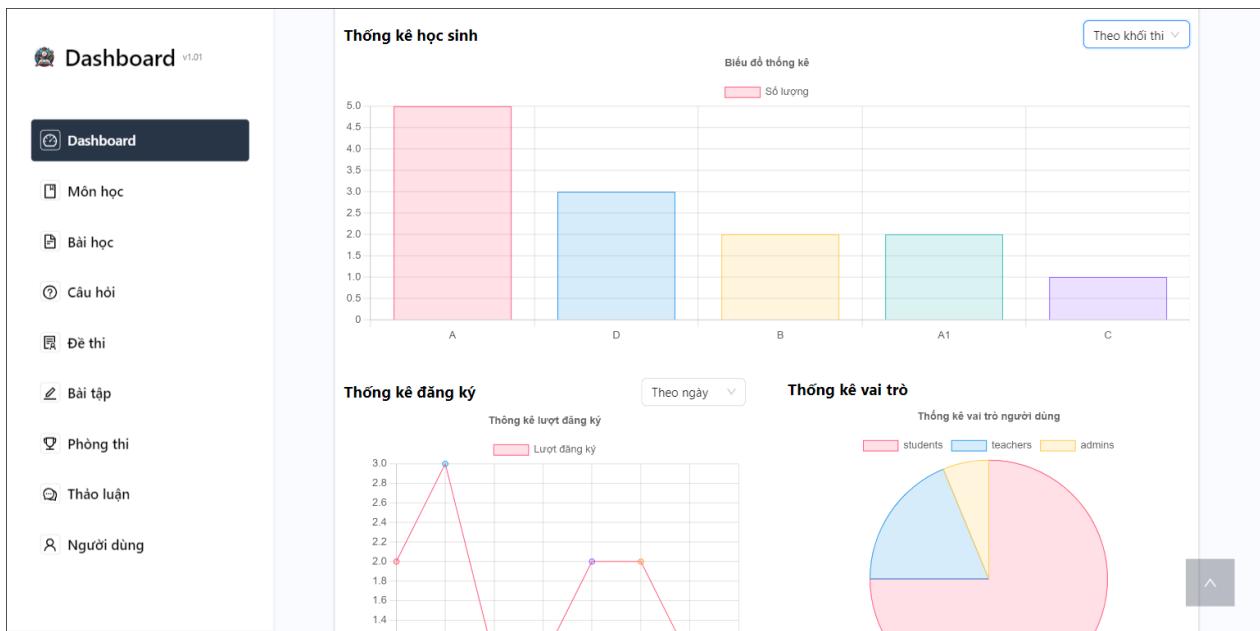
- Header:** Xin chào Châu Quế Nhơn, Quay lại trang web
- Sidebar:**
  - Dashboard** (highlighted)
  - Môn học
  - Bài học
  - Câu hỏi
  - Đề thi
  - Bài tập
  - Phông thi
  - Thảo luận
  - Người dùng
- Dashboard Summary:**

Số lượng thí sinh 12	Ngân hàng câu hỏi 604	Kho đề thi 6	Số lượt thi 30
-------------------------	--------------------------	-----------------	-------------------
- Table: Những học sinh sôi nổi nhất**

ID	Học sinh	Số bài tập	Số đề đã giải	Tham gia đấu trường	Tần suất thảo luận	Điểm sôi nổi
1	Quế Ngọc Hải	0	14	0	Rất Thấp	87
2	Châu Quế Nhơn	1	1	3	Thấp	36
3	Phùng Thành Đô	0	0	3	Rất Thấp	21
4	Nguyễn Công Phương	0	1	2	Rất Thấp	21
5	Rùi Tiến Dũng	0	0	0	Đã Tham Gia	15

Hình 4. 23 Giao diện trang Dashboard

Hệ thống sẽ thống kê các số liệu về bảng xếp hạng những học sinh ôn luyện sôi nổi nhất trên web. Quản trị viên có thể xuất danh sách ra file để theo dõi hoặc thống kê tùy mục đích.



Hình 4. 24 Giao diện trang Dashboard 2

#### 4.1.2.3 Trang quản lý

Hệ thống cung cấp bộ quản lý đầy đủ các tính năng của trang web như môn học, tài liệu học tập, bài tập, câu hỏi, đề thi, đấu trường và cả quản lý người dùng.

##### a. Quản lý môn học

The table lists subjects for Grade 12:

Lớp	Môn học	Khối lớp	Actions
Lớp 12			+ Thêm mới
39	Ngữ Văn	12	
38	Sinh Học	12	
37	Hóa Học	12	
36	Vật Lý	12	
35	GDCD	12	
8	Địa lý	12	
7	Lịch sử	12	
5	Tiếng anh	12	
1	Toán học	12	

GolUni Admin © 2024

Hình 4. 25 Giao diện trang quản lý chương

Mỗi khối lớp sẽ có nhiều môn học khác nhau, quản trị viên có thể quản lý thêm, sửa xóa môn học đối với từng khối lớp. Ngoài ra, ứng với mỗi môn học sẽ phân ra thành nhiều chương.

ID	Số chương	Tên chương	Môn	Khối lớp	
33	I	Hóa học và văn đề phát triển kinh tế, xã hội, môi trường	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
32	II	Phân biệt một số chất vô cơ. Chuẩn độ dung dịch	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
31	III	Crom - Sắt - Đồng	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
30	IV	Kim loại kiềm - Kim loại kiềm thổ - Nhóm	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
29	V	Đại cương về kim loại	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
28	VI	Polime và vật liệu polime	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
27	VII	Amin - Amino Axit - Protein	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
26	VIII	Cacbonhidrat	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>
25	IX	Este - Lipit	Hóa Học	12	<a href="#">Sửa</a> <a href="#">Xóa</a>

GoUni Admin © 2024

*Hình 4. 26 Giao diện trang quản lý chương**b. Quản lý tài liệu bài học*

ID	Tên bài học	
41	Viết đoạn văn 200 chữ về sự khen và chê	<a href="#">Sửa</a> <a href="#">Xóa</a>
40	Viết đoạn văn 200 chữ nêu Suy nghĩ của anh chị về lòng yêu nước	<a href="#">Sửa</a> <a href="#">Xóa</a>
39	Suy nghĩ của anh chị về quan điểm: Công hiến hết mình, hưởng thụ tối đa	<a href="#">Sửa</a> <a href="#">Xóa</a>
38	Viết đoạn văn nghĩ luận xã hội về tình thầy trò	<a href="#">Sửa</a> <a href="#">Xóa</a>
37	Viết đoạn văn 200 chữ về Thanh niên trước cuộc Cách mạng công nghiệp 4.0 hay nhất	<a href="#">Sửa</a> <a href="#">Xóa</a>
36	Nghị luân về khoảng lảng trong cuộc sống hay nhất	<a href="#">Sửa</a> <a href="#">Xóa</a>

GoUni Admin © 2024

*Hình 4. 27 Giao diện trang quản lý bài học**c. Quản lý câu hỏi*

ID	Câu hỏi	Môn học	Khối lớp	
292	Đồ thị của hàm số $y = \log(x)$ có tiệm cận đứng là:	Toán học	12	Sửa Xóa
293	Phương trình mặt phẳng đi qua ba điểm A(1,2,3), B(4,5,6), C(7,8,9) là:	Toán học	12	Sửa Xóa
294	Cho hàm số $y = x^4 - 4x^2 + 4$ . Tính giá trị cực đại của hàm số.	Toán học	12	Sửa Xóa
295	Tính diện tích hình phẳng giới hạn bởi đồ thị hàm số $y = x^2$ và trục hoành từ $x = -2$ đến $x = 2$ .	Toán học	12	Sửa Xóa
296	Cho số phức $z = 2 + 2i$ . Tính mô-đun của số phức $z$ .	Toán học	12	Sửa Xóa
297	Cho hàm số $y = \frac{x^2-1}{x-1}$ . Đường tiệm cận ngang của đồ thị hàm số là:	Toán học	12	Sửa Xóa
298	Khảo sát sự biến thiên của hàm số $y = \sin x$ trên khoảng $[0, 2\pi]$ .	Toán học	12	Sửa Xóa
299	Giá trị lớn nhất của hàm số $f(x) = 2x - x^2$ trên đoạn $[0, 2]$ là:	Toán học	12	Sửa Xóa
278	Cho hàm số $y = x^2 - 4x + 4$ . Giá trị cực tiểu của hàm số là:	Toán học	12	Sửa Xóa

Hình 4. 28 Giao diện trang quản lý câu hỏi

Ngân hàng câu hỏi của trang web. Tất cả câu hỏi sẽ được lưu tại đây và phân theo khối lớp, môn học, chương, ... Người dùng có role giáo viên có thể quản lý các câu hỏi này bao gồm thêm sửa xóa các câu hỏi trắc nghiệm.

#### d. Quản lý đề thi

ID	Đề thi	Thời gian	Câu hỏi	Lượt thi	
16	Đề chính phục 8+ Toán 2024 P2	90	50	0	Sửa Xóa
15	Đề chính phục 8+ Toán Khóa C-T-G MClass	90	50	0	Sửa Xóa
14	Đề ôn tập Toán THPT QG #1	90	45	0	Sửa Xóa
13	Đề chính phục 8+ Toán 2024	90	50	1	Sửa Xóa
12	Đề toán 15 buổi sáng #1	15	10	1	Sửa Xóa
11	Đề chính thức kỳ thi tốt nghiệp THPT năm 2023 môn Toán #105	90	50	0	Sửa Xóa

Hình 4. 29 Giao diện trang quản lý đề thi

**Thêm đề thi mới**

**Thông tin**

**Câu hỏi**

**Xác nhận**

**STT** Câu hỏi Hành động

1	Cho hàm số $f(x) = x^2 - 2x + 5$ . Khẳng định nào dưới đây đúng?	Xóa
2	Điểm M trong hình bên là điểm biểu diễn của số phức nào dưới đây?	Xóa
3	Từ một nhóm học sinh gồm 5 nam và 8 nữ, chọn ngẫu nhiên 4 học sinh. Xác suất để 4 học sinh được chọn có cả nam và nữ bằng	Xóa
4	Tính tích phân: $\int_0^1 \sqrt{x} dx$	Xóa
5	Tính đạo hàm của hàm số ( $y = x^2$ )	Xóa
6	Tính đạo hàm của hàm số ( $y = \ln(x)$ )	Xóa
7	Tính đạo hàm của hàm số $y = e^x$	Xóa

Hình 4. 30 Giao diện thêm/chỉnh sửa đề thi

e. Quản lý bài tập

**Dashboard** v1.01

Xin chào Châu Quế Nhơn

Quay lại trang web

**Quản lý bài tập**

+ Thêm mới

12 Toán học Hàm số lũy thừa, hàm số mũ Xác nhận Clear

ID	Đề thi	Câu hỏi	Lượt thi	Sửa	Xóa
12	Bài tập hàm số lũy thừa, hàm số mũ #4	42	0	Sửa	Xóa
11	Bài tập hàm số lũy thừa, hàm số mũ #3	10	0	Sửa	Xóa
10	Bài tập hàm số lũy thừa, hàm số mũ #2	30	0	Sửa	Xóa
9	Bài tập hàm số lũy thừa, hàm số mũ #1	25	0	Sửa	Xóa

GoUni Admin © 2024

Hình 4. 31 Giao diện trang quản lý bài tập

### f. Quản lý đấu trường

Hình 4. 32 Giao diện trang quản lý phòng thi đấu

Người dùng có role “Teacher” sẽ được cung cấp quyền quản lý phòng thi đấu như tạo mới, chỉnh sửa và xóa phòng thi đấu. Chỉ những phòng thi ở chế độ đang chờ mới được phép chỉnh sửa.

Hình 4. 33 Giao diện thêm/sửa phòng thi

Với chức năng thêm và sửa phòng thi sẽ chia thành 3 bước: nhập thông tin phòng, thông tin đề thi và cuối cùng sẽ là xác nhận lại thông tin trước khi lưu lại. Sẽ có 2 chế độ phòng thi, bao gồm tạo phòng thi cổ điển và chế độ thi đấu đối kháng giữa những người dùng. Quá trình thêm đề thi, người thêm có thể tự thêm thủ công các câu hỏi hoặc có thể thêm nhanh (sẽ tự lấy câu hỏi tương ứng trong ngân hàng trắc nghiệm).

### g. Quản lý chủ đề thảo luận

<b>Quản lý chủ đề thảo luận</b>				
ID	Tiêu đề	Nội dung	Người đăng	
18	Quy tắc tính đạo hàm	cho em hỏi mấy cái Quy tắc tính đạo hàm của hàm hợp chung minh như thế nào vậy ạ?	Trần Thái Linh	Xóa Đến ➔
17	Tuyển sinh đại học 2024: Thêm nhiều ngành mới	 [TẶNG BẠN] TRỌN BỘ Bí kíp học tốt 08 môn...	Hoàng Minh Sơn	Xóa Đến ➔
16	Lấy lại gốc môn Lý	Chào mọi người, năm nay em học lớp 11 và theo học khối D. Từ lúc học c2 em đã xác định học khối D và bỏ bê mấy môn tự...	Trần Thái Linh	Xóa Đến ➔
15	Mách học sinh Trung bình-khá cách đạt 8 điểm thi THPT quốc gia	<i>Teen 2k6 ơi, Trong hai tháng cuối cùng trước kỳ thi THPT quốc gia 2024 bạn không đủ thời gian để ôn luyện toàn diện...</i>	Đỗ Hùng Dũng	Xóa Đến ➔
14	Bí kíp để "xử lý gọn" 50 câu Tiếng Anh trong 60 phút	Có chiến lược làm bài thi rõ ràng, hiệu quả sẽ giúp sức rất nhiều cho bạn bén cạnh kiến thức và kỹ năng tốt. Để trả l...	Hoàng Thị Loan	Xóa Đến ➔
13	mọi người giúp em với ạ hứu em k hiú	tìm GTLN và GTNN của $(\sin(x))^4 + (\cos(x))^4$	Trần Nguyên Mạnh	Xóa Đến ➔
12	Cảm nhận triết ngắn "Mùi của mây trời" của Nguyễn Ngọc Tư	mùi long não trên người bà già sôc lên mũi tôi trong lúc giúp bà cài dây an toàn bà hỏi mở cửa sổ được không một câu hỏi...	Bùi Tiến Dũng	Xóa Đến ➔
11	giúp mình với	giúp mình với...	Quế Ngọc Hải	Xóa Đến ➔

Hình 4. 34 Giao diện trang quản lý chủ đề thảo luận

### h. Quản lý người dùng

Quản trị viên trong hệ thống có hai nhiệm vụ chính là phân quyền và kiểm soát truy cập. Họ có quyền thiết lập và quản lý các quyền hạn cho người dùng khác, đảm bảo rằng mỗi người dùng chỉ có quyền truy cập và hoạt động trong phạm vi được cấp phép. Đồng thời, quản trị viên cũng có thể thực hiện việc khóa tài khoản của những người dùng vi phạm để bảo vệ tính toàn vẹn và an toàn của hệ thống. Điều này giúp duy trì một môi trường làm việc an toàn và hiệu quả, đồng thời đảm bảo sự tuân thủ các chính sách và quy định của hệ thống.

## Quản lý người dùng

ID	Username	Họ tên	Email	SĐT	Role	Ban	Role
39	nhoncq	Nguyễn Công Phương	quenhon2002@gmail.com	0848611127	<span>STUDENT</span>		
13	john	Đỗ Hùng Dũng	johndoe@example.com		<span>STUDENT</span>		
12	sonhm	Hoàng Minh Sơn	son.tottenham@gmail.com		<span>STUDENT</span>		
11	hoalo	Hoàng Thị Loan	hoangloanvff@gmail.com		<span>STUDENT</span>		
10	eqfoio52iv26pp7kjm03	Trần Thái Linh	skyland.0410@gmail.com		<span>STUDENT</span>		
9	3w08oz7322qc5ltt4mmch	Phùng Thanh Đô	nhon.q.chau@tgi-sol.com		<span>STUDENT</span>		
8	nhon,test	Trần Nguyên Mạnh	nhontest@gmail.com		<span>STUDENT</span> <span>TEACHER</span>		
7	phuchv1	Bùi Tiến Dũng	cnhon5496@gmail.com	0848611127	<span>STUDENT</span>		
6	zh7o6vbeulxnruz4pktgn	Đỗ Duy Mạnh	cqn1805@gmail.com		<span>STUDENT</span>		
5	nhoncq2	Quế Ngọc Hải	quenhon.dev@gmail.com	0848611127	<span>STUDENT</span>		
4	admin	Châu Quế Nhơn	admin@gouni.com.vn	0848611127	<span>STUDENT</span> <span>ADMIN</span> <span>VIP</span> <span>TEACHER</span>		
3	ezxq8qyu5jzefk0mp1iow	Trần Đình Trọng	6151071079@st.utc2.edu.vn		<span>STUDENT</span> <span>TEACHER</span>		

1 2 >

Hình 4. 35 Giao diện trang quản lý người dùng

## 4.2 Tích hợp tính năng dự đoán

Mô hình dự đoán điểm thi trên trang web được xây dựng dựa trên thuật toán Random Forest. Thuật toán này sử dụng một tập hợp các cây quyết định để phân tích dữ liệu đầu vào và đưa ra các dự đoán. Trang web sẽ sử dụng mô hình này để dự đoán điểm thi của người dùng dựa trên các thông tin về quá trình ôn luyện, bao gồm số lượng bài kiểm tra đã hoàn thành, điểm số của các bài kiểm tra đó, và các yếu tố khác liên quan đến việc học tập.

Kết quả dự đoán được cung cấp bởi mô hình này giúp người dùng hiểu rõ hơn về tình trạng ôn luyện hiện tại của mình. Người dùng có thể thấy được điểm mạnh và điểm yếu của mình, từ đó có thể điều chỉnh cách ôn luyện để phù hợp hơn cho kỳ thi đại học. Ví dụ, nếu dự đoán cho thấy điểm số trong một môn học cụ thể đang ở mức thấp, người dùng có thể dành nhiều thời gian hơn để ôn tập môn đó hoặc thay đổi phương pháp học tập. Ngược lại, nếu điểm số trong một môn học khác đang ở mức cao, người dùng có thể tập trung vào các môn khác mà họ cảm thấy còn yếu.

Nhờ vào khả năng dự đoán chính xác của thuật toán Random Forest, trang web cung cấp một công cụ hữu ích giúp người dùng tối ưu hóa quá trình ôn luyện và nâng cao cơ hội đạt được kết quả tốt trong kỳ thi đại học.

#### **4.2.1 Triển khai API Prediction từ mô hình máy học**

Ngôn ngữ lập trình Python cung cấp rất nhiều thư viện hỗ trợ phát triển và tích hợp API như Django, FastAPI, Flask.

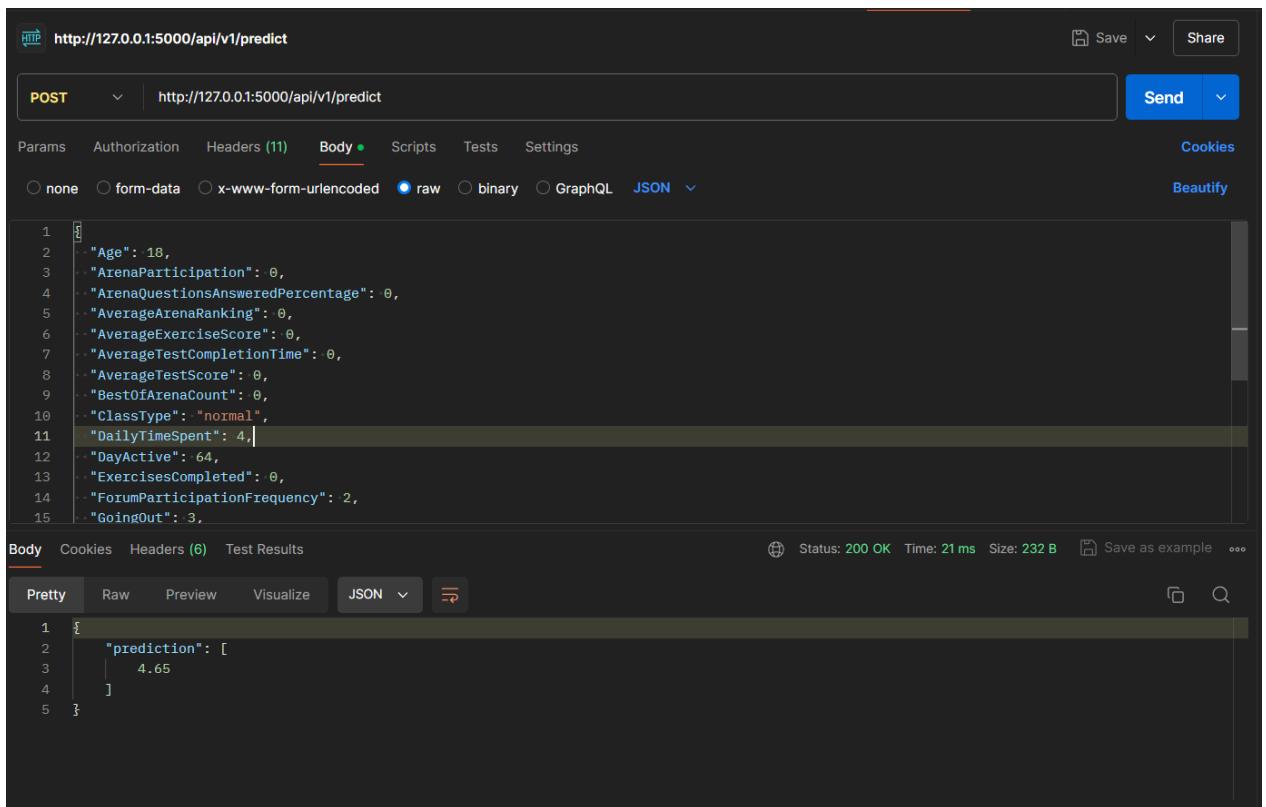
Đầu tiên, cần phải lưu tất cả đối tượng cần thiết như mô hình dự đoán đã được train và các công cụ tiền xử lý dữ liệu cần thiết vào file pickle để lưu trạng thái của đối tượng và tái sử dụng lại sau này mà không cần phải khởi tạo lại từ đầu.

```
# Tạo thư mục 'pkl' nếu chưa tồn tại
if not os.path.exists('pkl'):
    os.makedirs('pkl')

# Lưu các đối tượng vào thư mục 'pkl'
joblib.dump(model, 'pkl/model.pkl')
joblib.dump(scaler, 'pkl/scaler.pkl')
joblib.dump(label_encoders, 'pkl/label_encoders.pkl')
joblib.dump(categorical_columns, 'pkl/categorical_columns.pkl')
joblib.dump(numerical_columns, 'pkl/numerical_columns.pkl')
joblib.dump(features, 'pkl/features.pkl')
```

*Hình 4. 36 Hình ảnh lưu mô hình vào pickle file*

Sau đó cần phải cài Flask vào file python để phát triển API để sử dụng mô hình cho dự đoán từ phía client. Trong file app.py để phát triển API sẽ load lại các pickle file để tiến hành tiền xử lý lại dữ liệu được gửi đến trong request. Cuối cùng sử dụng mô hình đã được huấn luyện để dự đoán điểm dựa trên các dữ liệu mới này sau đó trả về kết quả cho người dùng.



Hình 4. 37 Demo về api dự đoán từ mô hình

#### 4.2.2 Thiết kế client

Để phục vụ dự đoán phía client, phía backend sẽ cần có thêm 3 API để hỗ trợ việc thu thập dữ liệu.

Đầu tiên, sẽ có một API để lấy danh sách các bài kiểm tra và các môn học thuộc khối thi của người dùng đang truy cập. API này giúp xác định những môn học mà người dùng cần dự đoán điểm thi. Thứ hai, sẽ có một API để thu thập các thông tin cơ bản của người dùng, bao gồm tuổi, lớp, loại trường học, và các thông tin khác có thể ảnh hưởng đến dự đoán điểm thi. Những dữ liệu này là yếu tố quan trọng để mô hình dự đoán có thể đưa ra kết quả chính xác. Cuối cùng, một API sẽ tính toán các số liệu cần thiết được thu thập từ lịch sử hoạt động của người dùng trên trang web. API này sẽ phân tích dữ liệu như số lượng bài kiểm tra đã làm, điểm số của các bài kiểm tra đó, và các hoạt động học tập khác.

Sau khi đã thu thập xong các dữ liệu cần thiết, hệ thống sẽ gọi API dự đoán cho từng môn học với bộ dữ liệu tương ứng. Quá trình này đảm bảo rằng mỗi môn học đều có dự đoán điểm thi riêng biệt, giúp người dùng có cái nhìn tổng quan và chi tiết về tình trạng ôn luyện của mình.

**1**

Châu Quế Nhơn  
18/05/2002  
Khối A  
Student Admin Vip Teacher

GounAI  
Xem dự đoán ôn luyện từ GounAI

**2**

Hiệu suất ôn tập  
B-  
Tổng điểm dự đoán: 14.81  
Xem dự đoán chi tiết C Dự đoán lại

**3**

Chi tiết dự đoán	
Hiệu suất ôn tập	
Toán học 5.51	Dự đoán điểm từng môn học Điểm dự đoán: C-
Vật Lý 4.65	Điểm dự đoán: C-
Hóa Học 4.65	Điểm dự đoán: C-
<input type="checkbox"/> Dự đoán lại	
Đóng	

Hình 4. 38 Demo chức năng dự đoán

## **KẾT LUẬN & KIẾN NGHỊ**

### **1. Kết quả đạt được**

Sau quãng thời gian khoảng 13 tuần nghiên cứu và thực hiện đồ án, em cũng đã học hỏi, tiếp thu thêm được nhiều kiến thức hay và bổ ích về mặt kiến thức chuyên môn và cả các kỹ năng mềm. Về mặt kỹ năng chuyên môn, em đã nâng cao kiến thức về lập trình web như caching, microservice, realtime, docker, deploy. Năm được cách xây dựng trang web với restful API. Nâng cao trình độ về PHP Laravel như Queue, Cache và cách sử dụng Redis để cải thiện hiệu suất ứng dụng. Mở rộng chuyên môn fullstack với các kỹ năng Frontend như NextJS, Tailwind và các thư viện hỗ trợ giao diện mạnh mẽ của npm. Về mặt kỹ năng mềm, em cũng đã nâng cao được các kỹ năng nghiên cứu, kỹ năng đọc sách, giải quyết vấn đề và kỹ năng quản lý thời gian.

Từ những kết quả đạt được về mặt kiến thức và kỹ năng chuyên môn sau 13 tuần nghiên cứu, cùng với toàn bộ kiến thức đã được học suốt 4 năm đại học, em cũng đã hoàn thành được nhiệm vụ thiết kế tốt nghiệp về mặt sản phẩm, đó là xây dựng xong trang web luyện thi với đầy đủ các tính năng đặt ra ban đầu:

- Xác thực người dùng, có social login. Phân quyền cho người dùng. Tính năng tạo chủ đề thảo luận, tương tác câu trả lời như stackoverflow. Cung cấp bài học lý thuyết, các tài liệu học tập. Luyện tập giải đề theo bài tập từng chương hoặc làm các đề kiểm tra trong thời gian quy định.

- Tính năng đấu trường với 2 chế độ: Tạo phòng thi để tất cả thí sinh có thể tham gia làm bài cùng lúc trong thời gian quy định và tạo phòng thi đấu người tham gia có thể thi đấu đối kháng với nhau theo thời gian thực.

- Trang Admin với đầy đủ các tính năng quản lý CRUD. Thống kê kết quả ôn luyện bằng bảng, biểu đồ.

- Ứng dụng thuật toán Random Forest cho bài toán hồi quy vào xây dựng mô hình dự đoán điểm thi để đánh giá hiệu suất ôn luyện của người dùng tại web.

### **2. Hạn chế**

Do thời gian nghiên cứu và thực hiện đồ án có hạn cũng như trình độ chuyên môn còn nhiều hạn chế, trang web sẽ không tránh khỏi còn tồn tại 1 số bug ở cả client lẫn server. Chưa thực hiện caching được các static assets cũng như api khiến mỗi lần vào web đều phải tải lại tài nguyên, gây tốn bộ nhớ trình duyệt. Các đoạn xử lý logic,

database phía server vẫn chưa được tối ưu, làm chậm tốc độ tải trang. Dữ liệu mẫu còn khá ít, chưa test đầy đủ được các testcase cũng như chưa xử lý được với dữ liệu lớn.

Mô hình được xây dựng còn khá đơn giản, chưa có tập dữ liệu thực từ người dùng để huấn luyện chính xác hơn so với thực tế, chưa có những bước xử lý dữ liệu hiệu quả hơn. Các chỉ số MSE và MAE vẫn còn thể hiện độ sai lệch nhiều (~ 0.1 trên thang điểm 10).

### **3. Hướng phát triển**

Nếu có thêm thời gian để nghiên cứu sâu hơn về phát triển Web và các tính năng AI tích hợp cho trang web. Em muốn tập trung vào ba hướng phát triển chính sau để nâng cao chất lượng và hiệu quả của trang web:

- Khắc phục những vấn đề còn tồn đọng: việc đầu tiên cần làm là phải kiểm thử sâu và fix những lỗi còn sót lại. Tối ưu hóa hệ thống api backend để cải thiện, tối ưu hóa logic xử lý, truy vấn cơ sở dữ liệu, và caching các file assets, static API nhằm tăng tốc độ tải trang và cải thiện trải nghiệm người dùng.
- Cải thiện UI và phát triển thêm tính năng mới:
  - + Cải thiện giao diện người dùng cho sinh động đẹp mắt hơn, tối ưu UX/UI để nâng cao trải nghiệm người dùng.
  - + Phát triển thêm tính năng lớp học. Cung cấp bộ quản lý lớp học cho giáo viên, giáo viên có thể tạo lớp học miễn phí và có phí. Giáo viên có thể tạo những bộ đề kiểm tra độc quyền cho lớp học là tạo thống kê cho học sinh trong lớp học. Tích hợp streaming để giáo viên có thể tạo meeting dạy học, giải đề online trực tuyến ngay trên lớp trong lớp học của mình.
- Nâng cao và mở rộng mô hình máy học:
  - + Nếu có được một tập dữ liệu chính xác hoàn toàn từ người dùng sẽ có thể xây dựng một rồng ngẫu nhiên với những cây quyết định đa dạng hơn, đưa ra kết quả dự đoán với hiệu suất tốt hơn, có thể sử dụng trong thực tế.
  - + Nghiên cứu phát triển một mô hình máy học LLM (Large Language Model) để có thể hỗ trợ học sinh luyện tập các bài dưới hình thức tự luận, gợi ý giải những bài tập khó một cách chi tiết.

## TÀI LIỆU THAM KHẢO

- [1]. Christopher M. Bishop, “*Pattern Recognition and Machine Learning (Information Science and Statistics)*”, NXB Springer (2016).
- [2] Jiawei Han, Micheline Kamber & Jian Pei, “*Data Mining Concepts And Techniques 4th Edition*”, NXB Morgan Kaufmann (2022).
- [3] Frank E. Harrell Jr. “*Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*”, NXB Springer Series in Statistics (2015).
- [4] Andreas C. Müller và Sarah Guido. “*Introduction to Machine Learning with Python*”, NXB O'Reilly Media (2017).
- [5] David R. Jackson. “*Decision Trees and Random Forests: A Visual Introduction for Beginners*”, NXB CreateSpace Independent Publishing Platform (2017).
- [6] Aurélien Géron. “*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*”, NXB O'Reilly Media (2019).
- [7] John D. Kelleher, Brian Mac Namee và Aoife D'Arcy. “*Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*”, NXB The MIT Press (2015).
- [8] Matthew Setter. “*Modern PHP: New Features and Good Practices*”, NXB Packt Publishing (2015).
- [9] Matt Zandstra. “*PHP Objects, Patterns, and Practice*”, NXB Apress (2016).
- [10] Paul DuBois. “*MySQL Cookbook*”, NXB O'Reilly Media (2014).
- [11] Jon Duckett. “*PHP & MySQL: Server-side Web Development*”, NXB Wiley (2009).
- [12] Joel Murach & Ray Harris. “*Murach's PHP and MySQL*”, NXB Mike Murach & Associates Inc. (2018).
- [13] Robin Nixon. “*Learning PHP, MySQL & JavaScript: With jQuery, CSS & HTML5*”. NXB O'Reilly Media (2015).
- [14] Amundsen M. & Richardson L. “*RESTful Web APIs: Services for a Changing World*”. NXB O'Reilly Media (2013).
- [15] Lengstorf J. & Leggeter P. “*Realtime Web Apps: With HTML5 WebSocket, PHP, and jQuery*”. NXB Apress (2013).

- [16] Emmitt Scott. “*SPA Design and Architecture: Understanding Single Page Web Applications*”. NXB Manning (2015).
- [17]. “Laravel's official documents”. <https://laravel.com/docs/11.x> , truy cập 03/2024.
- [18]. “Vercel's official Nextjs document”. <https://nextjs.org/docs> , truy cập 05/2024.