

TRƯỜNG ĐẠI HỌC TRÀ VINH
KHOA KỸ THUẬT VÀ CÔNG NGHỆ



ISO 9001:2015

MẠCH KHÁNH DUY

XÂY DỰNG HỆ THỐNG DỰ ĐOÁN ĐIỂM TÍCH LŨY
CHO SINH VIÊN TRƯỜNG ĐẠI HỌC TRÀ VINH

ĐỒ ÁN TỐT NGHIỆP
NGÀNH CÔNG NGHỆ THÔNG TIN

TRÀ VINH, NĂM 2024

TRƯỜNG ĐẠI HỌC TRÀ VINH
KHOA KỸ THUẬT VÀ CÔNG NGHỆ

XÂY DỰNG HỆ THỐNG DỰ ĐOÁN ĐIỂM TÍCH LŨY
CHO SINH VIÊN TRƯỜNG ĐẠI HỌC TRÀ VINH

ĐỒ ÁN TỐT NGHIỆP
NGÀNH CÔNG NGHỆ THÔNG TIN

Sinh viên: **Mạch Khánh Duy**

Lớp: **DA20TTA**

MSSV: **110120164**

GVHD: **ThS. Phạm Thị Trúc Mai**

TRÀ VINH, NĂM 2024

LỜI MỞ ĐẦU

Dù không quyết định tất cả nhưng điểm số vẫn luôn được quan tâm trên con đường học vấn của mỗi cá nhân. Bởi vì, đối với sinh viên nó đóng vai trò là công cụ đo lường năng lực hiểu biết về môn học đó, nếu điểm số quá thấp có thể dẫn đến những kết quả không mong muốn như: cảnh báo học vụ, rớt môn,... Bên cạnh đó, đối với cố vấn học tập, điểm số giúp đánh giá năng lực của từng cá nhân trong lớp thuộc quyền quản lý, từ đó giúp cố vấn học tập chủ động hơn trong việc quản lý lớp. Ngoài ra, điểm số còn có ý nghĩa quan trọng trong việc mở cánh cửa tương lai, dựa vào điểm số có thể đánh giá được điểm mạnh và điểm yếu của bản thân từ đó định hướng nghề nghiệp bản thân cần hướng đến. Đặc biệt hơn, điểm số tốt có thể mở ra vô số cơ hội về học bổng và tạo điều kiện thuận lợi cho các công việc sau này.

Dự đoán trước điểm số đóng vai trò hỗ trợ sinh viên trong quá trình học tập, kịp thời phát hiện điểm yếu kém của bản thân và khắc phục kịp thời tránh các trường hợp không mong muốn.

Ngoài việc giúp cho các sinh viên yếu kém, dự đoán điểm số còn có tác dụng giúp sinh viên thuộc dạng khá giỏi nâng cao điểm số hơn nữa hoặc hướng bản thân đến một tầm cao hơn nữa, một học bổng bậc cao hơn nữa.

Dự đoán điểm cũng còn giúp cố vấn học tập kiểm soát được tình hình học tập của các cá nhân trong lớp học thuộc quyền quản lý từ đó có những biện pháp học tập phù hợp để học sinh đạt được kết quả tốt hơn tránh các trường hợp rớt môn, không đủ điểm tích lũy,... dẫn đến việc ra trường trễ thời hạn.

Hệ thống dự đoán điểm tích lũy cho sinh viên sử dụng hai thuật toán Hồi quy tuyến tính và Rừng ngẫu nhiên. Dữ liệu yêu cầu của hệ thống bao gồm môn học và điểm số của từng môn để tiến hành tính toán dữ liệu dựa trên chương trình đào tạo của các ngành đào tạo.

LỜI CẢM ƠN

Đầu tiên, tôi xin gửi lời cảm ơn đến Ban Giám hiệu Trường Đại học Trà Vinh cùng Khoa Kỹ thuật và Công nghệ đã đồng ý và tạo điều kiện cho tôi thực hiện đồ án tốt nghiệp với đề tài mà tôi mong muốn. Bản thân tôi còn về mặt kiến thức còn rất nhiều hạn chế nhưng nhận được sự đồng ý của Ban Giám hiệu đã tiếp thêm động lực để tôi có thể hoàn thành tốt đề tài.

Thứ hai, tôi xin cảm ơn đến tất cả các cố vấn học tập của Bộ môn Công nghệ Thông tin đã giảng dạy tôi trong suốt 4 năm qua giúp tôi có đủ kiến thức nền tảng để thực hiện đề tài.

Cuối cùng, tôi xin chân thành gửi lời cảm ơn đến cô Phạm Thị Trúc Mai thuộc Khoa Kỹ thuật và Công nghệ đã nhiệt tình hướng dẫn tôi trong thời gian thực hiện đồ án tốt nghiệp. Do thời gian, tư duy, kiến thức còn rất hạn hẹp nên quá trình thực hiện đồ án tôi liên tục gặp rất nhiều trở ngại nhưng nhờ có sự động viên và hỗ trợ nhiệt tình từ cô đã giúp tôi vượt qua và có thể hoàn thành đồ án tốt nghiệp đúng thời hạn.

Trà Vinh, ngày tháng 7 năm 2024

Sinh viên thực hiện

Mạch Khánh Duy

[illegible]

Cố vấn học tập hướng dẫn
(ký và ghi rõ họ tên)

BẢN NHẬN XÉT ĐỒ ÁN, TỐT NGHIỆP
(Của cố vấn học tập hướng dẫn)

Họ và tên sinh viên: Mạch Khánh Duy

MSSV: 110120164

Ngành: Công nghệ Thông tin

Khóa: 2020

Tên đề tài: Xây dựng hệ thống dự đoán điểm tích lũy cho sinh viên Trường Đại học Trà Vinh

Họ và tên Giáo viên hướng dẫn: Phạm Thị Trúc Mai

Chức danh: Cố vấn học tập

Học vị: Thạc Sĩ

NHẬN XÉT

1. Nội dung đề tài:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2. Ưu điểm:

.....

.....

.....

3. Khuyết điểm:

.....

.....

.....

.....

4. Điểm mới đề tài:

.....

.....

.....

.....

.....

5. Giá trị thực trên đề tài:

.....

.....

.....

.....

.....

.....

.....

7. Đề nghị sửa chữa bổ sung:

.....

.....

.....

.....

.....

.....

.....

8. Đánh giá:

.....

.....

.....

.....

Trà Vinh, ngày tháng năm 2024
Cố vấn học tập hướng dẫn
(Ký & ghi rõ họ tên)

Phạm Thị Trúc Mai

MỤC LỤC

CHƯƠNG 1. ĐẶT VẤN ĐỀ	1
1.1. Lý do chọn đề tài.....	1
1.2. Mục tiêu	1
1.3. Nội dung	2
1.4. Đối tượng và phạm vi nghiên cứu.....	2
1.5. Phương pháp nghiên cứu	2
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT.....	3
2.1. Ngôn ngữ lập trình	3
2.1.1. Giới thiệu về Bootstrap 4.....	3
2.1.2. Giới thiệu về Python	4
2.1.3. Giới thiệu về Django	6
2.1.4. Giới thiệu về MySQL	8
2.2. Giới thiệu về thuật toán	9
2.2.1. Thuật toán Hồi quy tuyến tính.....	9
2.2.2. Thuật toán Rừng ngẫu nhiên.....	11
2.3. Giới thiệu về một số thư viện sử dụng	15
2.4. Mô hình hệ thống đề xuất.....	16
CHƯƠNG 3. HIỆN THỰC HÓA NGHIÊN CỨU	18
3.1. Mô tả đề tài	18
3.2. Thiết kế dữ liệu	19
3.2.1. Sơ đồ lớp	19
3.2.2. Sơ đồ Use case	22
3.3. Xây dựng mô hình đề xuất.....	25
3.3.1. Dữ liệu đầu vào	25
3.3.2. Tiền xử lý dữ liệu.....	26

3.3.3. Xây dựng mô hình.....	28
3.4. Cài đặt Django	34
3.5. Phác thảo giao diện	36
3.5.1. Phác thảo giao diện đăng nhập	36
3.5.2. Phác thảo giao diện trang chủ dùng chung.....	37
3.5.3. Phác thảo giao diện thêm, sửa, xóa dùng chung	37
3.5.4. Phác thảo giao diện cho sinh viên	39
3.6. Dữ liệu thử nghiệm.....	41
CHƯƠNG 4. KẾT QUẢ NGHIÊN CỨU	47
4.1. Giao diện của Admin.....	47
4.2. Giao diện của cố vấn học tập.....	56
4.3. Giao diện của sinh viên	57
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	59
5.1. Kết luận.....	59
5.2. Hướng phát triển	60
DANH MỤC TÀI LIỆU THAM KHẢO.....	61

DANH MỤC CÁC HÌNH

Hình 2.1 Ảnh minh họa cách hoạt động của Hồi quy tuyến tính	11
Hình 2.2 Ví dụ về Rừng ngẫu nhiên [9].....	15
Hình 2.3 Mô hình hệ thống đề xuất	17
Hình 3.1 Sơ đồ lớp của hệ thống	19
Hình 3.2 Sơ đồ tổng quát hệ thống	23
Hình 3.3 Sơ đồ tổng quan Cố vấn học tập	23
Hình 3.4 Sơ đồ tổng quát Sinh viên	24
Hình 3.5 Quản lý hệ thống của Admin	25
Hình 3.6 Dữ liệu mẫu được trích xuất từ Edusoft	26
Hình 3.7 Những dữ liệu không cần thiết.....	26
Hình 3.8 Dữ liệu mẫu bị lỗi.....	27
Hình 3.9 Cây thư mục sau khi tạo dự án.....	35
Hình 3.10 Trang chủ dự án Django	36
Hình 3.11 Phác thảo giao diện đăng nhập.....	36
Hình 3.12 Phác thảo giao diện trang chủ chung.....	37
Hình 3.13 Phác thảo giao diện thêm, sửa, xóa chung.....	38
Hình 3.14 Phác thảo giao diện thêm dữ liệu	38
Hình 3.15 Phác thảo giao diện thêm dữ liệu bằng tệp	39
Hình 3.16 Phác thảo giao diện sửa dữ liệu.....	39
Hình 3.17 Phác thảo giao diện xem điểm theo học kỳ	40
Hình 3.18 Phác thảo giao diện dự đoán điểm số cho sinh viên	40
Hình 4.1 Giao diện đăng nhập.....	47
Hình 4.2 Giao diện trang chủ Admin.....	48
Hình 4.3 Giao diện chức năng quản lý học kỳ niên khóa	48
Hình 4.4 Giao diện chức năng quản lý bộ môn.....	49

Hình 4.5 Chức năng quản lý lớp học	49
Hình 4.6 Chức năng quản lý khoa	50
Hình 4.7 Giao diện chức năng quản lý ngành	50
Hình 4.8 Giao diện chức năng chương trình đào tạo	50
Hình 4.9 Chức năng quản lý chương trình đào tạo.....	51
Hình 4.10 Chức năng quản lý môn học	52
Hình 4.11 Chức năng quản lý nhóm môn học.....	52
Hình 4.12 Chức năng quản lý loại học phần.....	53
Hình 4.13 Chức năng quản lý thuộc chương trình đào tạo	53
Hình 4.14 Chức năng quản lý sinh viên.....	54
Hình 4.15 Chức năng quản lý cổ vấn học tập	54
Hình 4.16 Chức năng quản lý tài khoản	55
Hình 4.17 Chức năng quản lý điểm số.....	55
Hình 4.18 Chức năng quản lý chi tiết bảng điểm.....	56
Hình 4.19 Chức năng quản lý thông tin sinh viên của cổ vấn học tập	56
Hình 4.20 Chức năng quản lý điểm số của cổ vấn học tập.....	57
Hình 4.21 Chức năng quản lý chi tiết bảng điểm của cổ vấn học tập	57
Hình 4.22 Chức năng dự đoán điểm số sinh viên.....	58
Hình 4.23 Kết quả dự đoán của sinh viên	58

DANH MỤC CÁC BẢNG, SƠ ĐỒ, HÌNH

Bảng 2.1. Một số thư viện sử dụng trong đề tài	5
Bảng 3.1. Bảng ví dụ về tiền xử lý dữ liệu môn học	27
Bảng 3.2. Bảng dự đoán một số môn.....	29
Bảng 3.3. Dữ liệu bảng Khoa	41
Bảng 3.4. Dữ liệu bảng Bộ môn	41
Bảng 3.5. Dữ liệu bảng Ngành	41
Bảng 3.6. Dữ liệu bảng chương trình đào tạo.....	41
Bảng 3.7. Dữ liệu bảng Học kỳ chương trình đào tạo (chương trình đào tạo năm 2020) .	42
Bảng 3.8. Dữ liệu bảng nhóm môn.....	42
Bảng 3.9. Dữ liệu bảng loại học phần.....	43
Bảng 3.10. Dữ liệu bảng môn học (một số môn học thuộc chương trình đào tạo năm 2020)	44
Bảng 3.11. Dữ liệu bảng thuộc chương trình đào tạo(một số môn học thuộc chương trình đào tạo năm 2020).....	44
Bảng 3.12. Dữ liệu bảng Cố vấn học tập (Dữ liệu một số Cố vấn học tập)	45
Bảng 3.13. Dữ liệu bảng Lớp (dữ liệu một số lớp)	45
Bảng 3.14. Dữ liệu bảng Sinh viên (Dữ liệu một số sinh viên)	46
Bảng 3.15. Dữ liệu bảng điểm(một số bảng điểm của một sinh viên)	46
Bảng 3.16. Dữ liệu chi tiết bảng điểm(một số chi tiết bảng điểm của một bảng điểm)....	46

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Ý nghĩa
3D	Three-Dimensional
CSS	Cascading Style Sheets
CART	Classification and Regression Trees
Edusoft	Phần mềm quản lý đào tạo tại Trường Đại học Trà Vinh. Hệ thống cung cấp dữ liệu điểm số làm tập mẫu đầu vào của đề tài đồ án tốt nghiệp này
GIL	Global Interpreter Lock
HTML	HyperText Markup Language
IE	Internet Explorer
MDI	Mean Decrease in Impurity
MSE	Mean Squared Error
MVC	Model-View-Controller
MVT	Model-View-Template
MYSQL	My Structured Query Language
OBB	Out of Bag
ORM	Object-Relational Mapping
SQL	Structured Query Language
URL	Uniform Resource Locator

CHƯƠNG 1. ĐẶT VẤN ĐỀ

1.1. Lý do chọn đề tài

Trong quá trình học tập, việc dự đoán điểm số của bản thân từ đó nhận biết được các môn yếu kém là rất cần thiết trong quá trình học tập và làm việc cho sau này. Việc phát hiện điểm yếu để có thể tìm biện pháp phù hợp và khắc phục nó càng sớm càng tốt là việc rất cấp bách. Song song đó, việc tìm ra điểm mạnh của bản thân để có thể tiếp tục hướng đến biến điểm mạnh thành lợi thế riêng của bản thân cũng là quan trọng không kém. Để có thể nhìn ra những điều đó kể từ lúc còn ngồi trên ghế nhà trường thì điểm số thể hiện rõ nhất.

Hệ thống dự đoán điểm cung cấp một cái nhìn nhanh chóng hơn về bản thân trong quá trình học tập. Việc phán đoán trước điểm số chỉ rõ những môn học yếu kém từ đó có thể đưa ra những biện pháp học tập. Đặc biệt đối với những môn học cần cải thiện kiến thức và điểm số để tránh những trường hợp không mong muốn như nợ môn, rớt môn thậm chí là cảnh báo học vụ. Hệ thống cũng giúp ích cho cô vấn học tập với vai trò là người cô vấn học tập có cái nhìn tổng quan về kỹ năng của các sinh viên trong lớp thuộc quyền quản lý, giúp nhận biết các sinh viên không theo kịp tiến độ của lớp từ đó có các biện pháp quan tâm đến từng cá nhân.

Chính vì vậy, xây dựng hệ thống dự đoán điểm tích lũy cho sinh viên Trường Đại học Trà Vinh là cần thiết, mang lại lợi ích trong việc học tập cho sinh viên và cung cấp một công cụ tiện ích trong việc theo dõi tình hình học tập của sinh viên dành cho cô vấn học tập.

1.2. Mục tiêu

Sử dụng thuật toán Hồi quy tuyến tính và Rừng ngẫu nhiên tiến hành tính toán dữ liệu dự đoán. Tìm hiểu và xây dựng hệ thống dự đoán điểm tích lũy bằng Django framework.

Xây dựng Website với các phân quyền và chức năng của từng phân quyền:

- Phân quyền quản trị: Có các chức năng cơ bản như quản lý ngành đào tạo, khoa, bộ môn,... Quản lý điểm số sinh viên, các nhóm môn, lĩnh vực,... để tạo ra dữ liệu mẫu cho việc dự đoán dữ liệu điểm số

- Phân quyền cô vấn học tập: Có các chức năng như quản lý thông tin sinh viên,

thống kê của lớp thuộc quyền quản lý, xem các loại dự đoán cho sinh viên, thống kê được các số liệu cơ bản của lớp

- Phân quyền sinh viên: Có các chức năng như quản lý được thông tin cá nhân, chọn loại dự đoán

1.3. Nội dung

- Tiến hành thu thập dữ liệu, xử lý, lọc và làm sạch các dữ liệu không cần thiết cho hệ thống

- Tìm hiểu về quy trình hoạt động của Hồi quy tuyến tính và Rừng ngẫu nhiên, lựa chọn dữ liệu đầu vào và dữ liệu đầu ra phù hợp với hệ thống

- Tìm hiểu về quy trình hoạt động của Django, tích hợp các thư viện cần thiết cho việc dự đoán bằng thuật toán Hồi quy tuyến tính và Rừng ngẫu nhiên

- Tiến hành xây dựng hệ thống dự đoán điểm cho sinh viên với các phân quyền và các chức năng khác nhau cho từng phân quyền

1.4. Đối tượng và phạm vi nghiên cứu

- Tìm hiểu về chương trình đào tạo của các ngành đào tạo
- Nghiên cứu về cách quản lý môn học và điểm số của từng môn
- Tìm hiểu về thuật toán Hồi quy tuyến tính và Rừng ngẫu nhiên cùng với các thư viện sử dụng trong hệ thống dành cho việc dự đoán dữ liệu

1.5. Phương pháp nghiên cứu

- Phương pháp lý thuyết: Nghiên cứu các tài liệu liên quan đến Hồi quy tuyến tính, Rừng ngẫu nhiên và các Framework Django.

- Phương pháp thu thập dữ liệu: Thu thập và xử lý dữ liệu điểm số của sinh viên.

- Phương pháp thực nghiệm: Xây dựng ứng dụng web bằng Framework Django đã lựa chọn để dự đoán điểm tích lũy sử dụng ít nhất 2 kỹ thuật học máy.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

2.1. Ngôn ngữ lập trình

2.1.1. Giới thiệu về Bootstrap 4

Bootstrap 4 là một framework front-end rất phổ biến, được sử dụng để phát triển giao diện web hiện đại và có khả năng phản hồi. Nó bao gồm các mẫu thiết kế dựa trên HTML và CSS. Nó cung cấp một bộ các lớp CSS và JavaScript mạnh mẽ để giúp xây dựng trang web thân thiện với người dùng và dễ bảo trì một cách nhanh chóng. [1]

Bootstrap 4 sử dụng hệ thống lưới linh hoạt và mạnh mẽ để dễ dàng chia cột trên các thiết bị khác nhau. Lưới được chia thành 12 cột và cung cấp các lớp cho mỗi cột, cho phép sắp xếp các phần tử vào các cột khác nhau một cách linh hoạt.

Bootstrap 4 cung cấp các thành phần và tiện ích đa dạng như nút, điều khiển biểu mẫu, dropdown, thanh điều hướng, hộp cảnh báo và nhiều hơn nữa, giúp thiết kế giao diện web nhanh chóng và dễ dàng. Ngoài ra, nó cũng cung cấp các lớp tiện ích để thực hiện các tác vụ như margin, padding, căn chỉnh, ẩn/hiện và các thao tác khác.

Bootstrap 4 được thiết kế để hỗ trợ nhiều thiết bị từ máy tính đến di động. Với hệ thống lưới linh hoạt và các lớp đáp ứng, có thể dễ dàng tạo ra các trang web thân thiện với người dùng trên mọi kích thước màn hình.

Bootstrap 4 cung cấp thanh điều hướng linh hoạt, cho phép tạo ra các thanh điều hướng có thể tùy chỉnh với menu thả xuống và biểu tượng điều hướng thân thiện với di động.

Bootstrap 4 cho phép dễ dàng tùy chỉnh và mở rộng bằng cách chỉnh sửa hoặc ghi đè các tệp CSS, hoặc sử dụng SASS để tạo biến và Mixin tùy chỉnh. Điều này giúp thích nghi và mở rộng Bootstrap theo yêu cầu cụ thể của dự án.

Bootstrap 4 được hỗ trợ trên hầu hết các trình duyệt hiện đại, bao gồm cả IE10+.

Nhược điểm:

Giao diện phổ biến và đơn điệu: Do Bootstrap được sử dụng rộng rãi, nhiều trang web có thể có thiết kế giống nhau, làm giảm sự độc đáo và sáng tạo của giao diện web.

Kích thước tệp lớn: Bootstrap bao gồm nhiều tính năng và thành phần, dẫn đến kích thước tệp lớn. Nếu không tối ưu hóa, điều này có thể làm tăng thời gian tải trang.

Đôi khi không linh hoạt trong thiết kế: Mặc dù Bootstrap cung cấp nhiều tùy chọn tùy chỉnh, nhưng đôi khi việc hoàn toàn tùy chỉnh theo ý muốn mà không cần phải viết lại một số mã Bootstrap có thể gặp khó khăn.

Cần học nhiều lớp CSS: Để sử dụng Bootstrap hiệu quả, cần học và quản lý nhiều lớp CSS của Bootstrap, điều này có thể làm phức tạp quá trình phát triển và bảo trì.

2.1.2. Giới thiệu về Python

Python là một ngôn ngữ lập trình thông dịch, mức cao, đa nền tảng được thiết kế để đọc dễ hiểu và viết dễ dàng. Dưới đây là một số điểm quan trọng và đặc điểm của ngôn ngữ lập trình Python: [2]

- Đọc dễ hiểu và cú pháp rõ ràng: Python chú trọng vào cú pháp đơn giản và rõ ràng, giúp người lập trình dễ đọc và hiểu mã nguồn. Điều này giúp tăng tính dễ bảo trì và phát triển mã nguồn một cách hiệu quả.

- Đa năng và mạnh mẽ: Python là một ngôn ngữ đa năng, có thể được sử dụng trong nhiều lĩnh vực như phát triển web, trí tuệ nhân tạo, khoa học dữ liệu, phân tích hệ thống, và nhiều lĩnh vực khác.

- Thư viện và Frameworks phong phú: Python có một cộng đồng lớn và năng động, cung cấp nhiều thư viện và frameworks như Django, Flask (phát triển web), TensorFlow (trí tuệ nhân tạo), NumPy (khoa học dữ liệu) và nhiều thư viện khác.

- Tương thích đa nền tảng: Mã nguồn Python có thể chạy trên nhiều hệ điều hành như Windows, macOS và Linux mà không cần sửa đổi đáng kể.

- Hỗ trợ cộng đồng rộng lớn: Cộng đồng Python rất lớn và chia sẻ kiến thức thông qua nhiều diễn đàn, tài liệu học tập, giúp người lập trình giải quyết vấn đề và học hỏi từ nhau.

- Thích hợp cho người mới bắt đầu: Python thường được chọn làm ngôn ngữ đầu tiên cho những người mới học lập trình vì cú pháp đơn giản và sức mạnh của nó.

- Đồng nhất với triển khai cấp doanh nghiệp: Python thường được sử dụng trong các môi trường doanh nghiệp lớn và được ưa chuộng cho các dự án lớn do tính nhất quán, dễ bảo trì.

- Tích hợp với các ngôn ngữ khác: Python có khả năng tích hợp với nhiều ngôn

ngữ khác nhau, giúp nó trở thành một phần của các hệ thống phức tạp.

Bảng 2.1. Một số thư viện sử dụng trong đề tài

Tên thư viện	Chức năng
Pandas	Pandas là một thư viện mạnh mẽ và phổ biến được thiết kế để thao tác và phân tích dữ liệu. Được xây dựng trên nền tảng của NumPy và cung cấp các cấu trúc dữ liệu và công cụ phân tích dữ liệu để sử dụng. [3]
Matplotlib	Là một thư viện mạnh mẽ và linh hoạt của Python được sử dụng để tạo ra các biểu đồ và đồ thị dữ liệu. Cung cấp nhiều công cụ và tùy chọn để trực quan hóa dữ liệu theo cách chuyên nghiệp và đẹp mắt. Matplotlib thường được sử dụng trong các lĩnh vực khoa học, kỹ thuật, kinh tế, và khoa học dữ liệu để phân tích và trình bày dữ liệu. [4]

Tuy Python là một ngôn ngữ lập trình mạnh mẽ và phổ biến, nhưng cũng có một số nhược điểm như:

- Chậm trong một số trường hợp: Python là ngôn ngữ thông dịch nên thường chậm hơn so với các ngôn ngữ biên dịch như C++ hoặc Java. Điều này đặc biệt rõ rệt trong các ứng dụng xử lý dữ liệu lớn hoặc ứng dụng yêu cầu thời gian thực.

- Quản lý bộ nhớ không hiệu quả: Python có tự động quản lý bộ nhớ nhưng đôi khi không hiệu quả trong một số tình huống. Điều này có thể gây vấn đề trong các ứng dụng đòi hỏi hiệu suất cao và tài nguyên lớn.

- Không phù hợp với ứng dụng yêu cầu thời gian thực: Python không phù hợp với các ứng dụng đòi hỏi xử lý thời gian thực hoặc yêu cầu xử lý rất nhanh như trò chơi 3D hay hệ thống nhúng ít trễ.

- Hạn chế về luồng (threading): CPython (trình thông dịch chính của Python) có Global Interpreter Lock (GIL) giới hạn việc sử dụng đa luồng thực sự. Điều này có thể làm giảm hiệu suất trong các ứng dụng yêu cầu hoạt động đa luồng.

- Phụ thuộc vào bên ngoài: Python có nhiều thư viện quan trọng, đôi khi vẫn cần

phụ thuộc vào các thư viện bên ngoài để có các tính năng mà Python không hỗ trợ mặc định.

Mặc dù có những nhược điểm này, Python vẫn là một trong những ngôn ngữ lập trình phổ biến nhất nhờ tính đơn giản, dễ học và được hỗ trợ mạnh mẽ từ cộng đồng. Những hạn chế của Python thường có thể được giải quyết thông qua các phiên bản Python khác nhau và các thư viện mở rộng.

2.1.3. Giới thiệu về Django

Django là một framework phát triển web cao cấp được viết bằng ngôn ngữ Python. Nó sử dụng mô hình kiến trúc MVT giúp các nhà phát triển tạo ra các trang web một cách nhanh chóng và hiệu quả. Django tập trung chủ yếu vào sự đơn giản, khả năng mở rộng và rất dễ dàng sử dụng đối với người mới tiếp cận.

Cách hoạt động của Django:

Django hoạt động theo mô hình kiến trúc MVC (Model-View-Controller), nhưng các thành phần này được gọi là MVT (Model-View-Template).

- Models đại diện cho cấu trúc dữ liệu và định nghĩa các bảng trong cơ sở dữ liệu. Mô hình được định nghĩa là các lớp Python trong tệp `models.py` của ứng dụng. Mô hình sử dụng ORM (Object-Relational Mapping) để thực hiện việc ánh xạ giữa các đối tượng Python và các bảng trong cơ sở dữ liệu.

- Views bao gồm logic nghiệp vụ và định nghĩa cách hiển thị dữ liệu cho người dùng. Mỗi giao diện là một hàm Python hoặc một lớp dựa trên hàm, nhận yêu cầu từ người dùng, xử lý và trả về phản hồi. Giao diện được định nghĩa trong tệp `views.py`.

- Templates định nghĩa cách dữ liệu được hiển thị dưới dạng HTML. Django sử dụng hệ thống mẫu riêng của mình để bao gồm các biến và chèn logic đơn giản vào HTML. Các mẫu thường được lưu trữ trong thư mục `templates` của ứng dụng.

URLconf là cấu hình để định nghĩa các mẫu URL và ánh xạ từng mẫu đó tới từng hàm view tương ứng. Điều này được định nghĩa trong tệp `urls.py` của mỗi ứng dụng Django. Bằng cách sử dụng biểu thức chính quy, có thể định nghĩa các mẫu URL và Django sẽ sử dụng chúng để định tuyến các yêu cầu từ người dùng tới các hàm view phù hợp.

Middleware là các lớp xử lý các yêu cầu trước khi chúng đến view và xử lý các phản hồi trước khi chúng được gửi lại cho client. Middleware có thể được sử dụng cho các nhiệm vụ như xử lý bảo mật, session, và xác thực người dùng.

Ngoài ra Django còn cung cấp các công cụ quản lý và thao tác với cơ sở dữ liệu mạnh mẽ như: Migration sẽ tự động tạo và áp dụng các thay đổi trong model vào cơ sở dữ liệu. ORM Tương tác với cơ sở dữ liệu thông qua các đối tượng Python mà không cần viết SQL thủ công.

Bên cạnh việc là một framework phát triển mạnh mẽ, Django cũng mang những hạn chế như:

- Khó khăn trong việc điều chỉnh và tùy chỉnh: Django cung cấp một cấu trúc mạnh mẽ và chuẩn mực cho các ứng dụng web, nhưng đôi khi điều này có thể làm cho việc điều chỉnh và tùy chỉnh đặc thù của ứng dụng trở nên phức tạp.

- Phù hợp với các ứng dụng lớn và phức tạp: Django là một framework mạnh mẽ nhưng có thể quá nặng và quá chuẩn mực đối với các ứng dụng nhỏ và đơn giản. Điều này có thể dẫn đến overhead không cần thiết khi chỉ cần xây dựng các trang web đơn giản.

- Phụ thuộc vào cấu trúc dự án chuẩn: Django yêu cầu các ứng dụng tuân thủ một cấu trúc dự án chuẩn, điều này có thể là hạn chế đối với các dự án có cấu trúc hoặc quy trình phát triển đặc biệt.

- Khả năng mở rộng của ứng dụng: Mặc dù Django hỗ trợ mở rộng tốt, nhưng đôi khi việc mở rộng ứng dụng có thể gặp phải một số rào cản, đặc biệt là khi cần tích hợp với các công nghệ khác hoặc mở rộng hệ thống lớn.

- Hiệu suất trong các ứng dụng yêu cầu thời gian thực: Django không phù hợp cho các ứng dụng yêu cầu xử lý thời gian thực cao, ví dụ như ứng dụng game hoặc các hệ thống nhúng với yêu cầu tương tác cao.

Mặc dù có những hạn chế này, Django vẫn rất mạnh mẽ và được ưa chuộng trong việc phát triển các ứng dụng web có quy mô lớn và cần tính bảo mật cao. Việc lựa chọn sử dụng Django hay không phụ thuộc vào yêu cầu cụ thể của dự án và kỹ năng của đội ngũ phát triển.

2.1.4. Giới thiệu về MySQL

Là ngôn ngữ cơ sở dữ liệu phổ biến chỉ xếp sau Oracle, nó đang được các ứng dụng có lượng truy cập nhiều nhất thế giới sử dụng. Ví dụ như Facebook, Netflix, Uber,... [5]

Cơ sở dữ liệu MySQL hoạt động theo hệ thống Client – Server, nó bao gồm một máy chủ SQL đa luồng để hỗ trợ các mặt khác nhau. Nó được cung cấp dưới dạng một thư viện đa luồng giúp nhà lập trình viên dễ dàng liên kết và ứng dụng của mình để tạo nên các sản phẩm độc lập và nhỏ lẻ nhanh hơn, dễ quản lý hơn. Bên cạnh đó MySQL có thể liên kết với nhiều thư viện bên phía Client khác nhau, thuận tiện cho việc giao tiếp giữa dữ liệu và giao diện người dùng.

MySQL có những điểm mạnh đáng chú ý như:

- Điều dễ nhận thấy nhất khi sử dụng MySQL là khả năng dễ dàng cài đặt và sử dụng, lập trình viên chỉ mất vài phút để có thể hoàn thành việc cài đặt sử dụng.

- Trong suốt quá trình từ lúc phát triển đến thời điểm hiện tại, MySQL đã trải qua quá trình thử nghiệm trên nhiều công ty và nhiều tình huống xử lý khác nhau, từ các công ty nhỏ lẻ đến các tập đoàn lớn nhất nhì thế giới cũng đang phụ thuộc vào nó để hoạt động các ứng dụng quan trọng vì độ tin cậy của nó

- Khả năng lưu trữ của MySQL được tích hợp khả năng sao chép gốc. Nghĩa là nó được tích hợp mở rộng quy mô cho phép số lượng lớn người dùng truy cập lớn. Ví dụ rõ ràng nhất là khả năng truy cập của Facebook với hàng tỷ lượt truy cập mỗi ngày cũng thông qua hệ thống quản lý dữ liệu của MySQL.

- Bên cạnh khả năng lưu trữ thì tốc độ phản hồi cũng là một ưu điểm. Thông qua nhiều điểm đánh giá tiêu chuẩn giữa MySQL đã chứng minh được tốc độ phản hồi của nó vượt trội hơn, ít tốn kém hơn so với các dịch vụ cơ sở dữ liệu khác.

- Khả năng tự khắc phục thảm họa cũng là một ưu điểm nổi trội, nó cũng cấp hoàn chỉnh một hệ thống tự động sao lưu dữ liệu khi có sự cố xảy ra, hệ thống khôi phục dữ liệu của MySQL cam kết không mất dữ liệu và thời gian khôi phục được tính bằng giây.

- Tính bảo mật cao: MySQL tuân thủ đầy đủ các quy định của ngành và chính phủ, bao gồm cả việc bảo mật dữ liệu người dùng chung, nó sử dụng các tính năng bảo

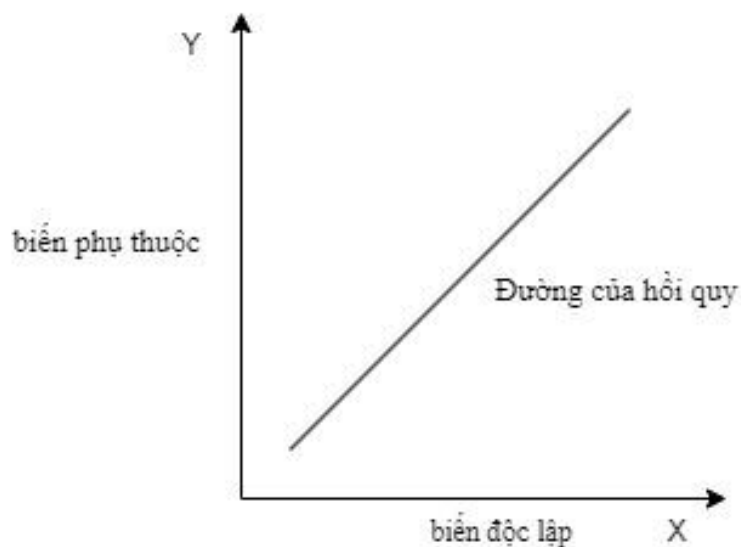
mật rất cao như tường lửa cơ sở dữ liệu, hệ thống mã hóa thông tin cá nhân, che giấu dữ liệu, xác thực/ủy quyền,...

2.2. Giới thiệu về thuật toán

2.2.1. Thuật toán Hồi quy tuyến tính

Là một trong những thuật toán dự đoán dữ liệu đơn giản, nó phân tích các dữ liệu đã biết để dự đoán các giá trị không xác định bằng các tuyến tính các dữ liệu liên quan. Hồi quy tuyến tính cung cấp một số công thức toán học dễ dàng giải thích để đưa ra các dự đoán. Nó đã được sử dụng lâu đời và áp dụng vào nhiều cho các phần mềm tính toán. [6]

Về các hoạt động của hồi quy tuyến tính chỉ đơn giản là nó vẽ một đồ thị bao gồm trục tung (y) và trục hoành (x). trong đó (x) là biến độc lập (biến giải thích hoặc biến dự báo), (y) là biến phụ thuộc. Để có thể dự đoán dữ liệu, mối quan hệ tuyến tính bắt buộc phải tồn tại giữa các biến phụ thuộc và độc lập.



Hình 2.1 Ảnh minh họa về biến phụ thuộc và biến giải thích

Các thông số của Hồi quy tuyến tính:

- Hệ số hồi quy (Regression Coefficients) bao gồm hai giá trị riêng biệt:

- + β_0 (Intercept): Hệ số này đại diện cho điểm giao cắt của đường hồi quy với trục y. Nó là giá trị của biến phụ thuộc khi tất cả các biến độc lập bằng 0.

+ β_i (Slope Coefficients): Các hệ số này đại diện cho độ dốc của đường hồi quy. Mỗi β_i cho biết sự thay đổi trong biến phụ thuộc khi biến độc lập tương ứng thay đổi một đơn vị, trong khi các biến khác được giữ nguyên.

- Phần dư độc lập: Phần dư là phần dữ liệu chênh lệch giữa dữ liệu đã biết và dữ liệu dự đoán, phần dư này thể hiện được mức độ chính xác của mô hình, phần dư càng lớn thì mô hình càng không chính xác và ngược lại phần dư càng gần 0 thì độ chính xác của mô hình càng đáng tin.

- Phương sai của phần dư (Residual Variance): Ký hiệu là σ^2 , phương sai của phần dư cho biết mức độ mà các giá trị thực tế của biến phụ thuộc (y) khác biệt so với các giá trị dự đoán \hat{Y} từ mô hình hồi quy. Phương sai của phần dư nhỏ cho thấy các giá trị dự đoán \hat{Y} rất gần với các giá trị thực tế y , tức là mô hình hồi quy có độ chính xác cao. Ngược lại, phương sai của phần dư lớn cho thấy sự khác biệt lớn giữa các giá trị dự đoán và giá trị thực tế, nghĩa là mô hình hồi quy không tốt trong việc dự đoán biến phụ thuộc.

- R^2 (Hệ số xác định - Coefficient of Determination): Hiểu đơn giản R^2 là thước đo mức độ biến thiên của dữ liệu mà các biến độc lập (biến giải thích hoặc biến dự báo) giải thích được cho biến phụ thuộc. Mức độ dao động của nó được biểu hiện từ 0 đến 1 hay có thể hiểu là từ 0% đến 100%, giá trị càng gần 1 thì mức độ giải thích được mô hình càng cao, theo đó mô hình càng đáng tin cậy và chính xác. [6]

- MSE (Mean Squared Error) là một chỉ số phổ biến dùng để đánh giá độ chính xác của các mô hình hồi quy trong học máy và thống kê. Nó đo lường mức độ chênh lệch giữa giá trị dự đoán của mô hình và giá trị thực tế. [7]

- Công thức: $MSE = (1/n) * \sum (y_i - \hat{y}_i)^2$
 - n : Số lượng điểm dữ liệu.
 - Y_i : Giá trị thực tế.
 - \hat{Y}_i : Giá trị dự đoán.
- MSE càng nhỏ, mô hình càng chính xác.

Hồi quy tuyến tính cũng bao gồm các loại khác nhau, tùy vào nhu cầu sử dụng của người dùng mà có thể lựa chọn mô hình hồi quy phù hợp:

- Hồi quy tuyến tính đơn giản: đây là mô hình đơn giản, dễ sử dụng và dễ giải thích nhất của Hồi quy tuyến tính, phương trình của nó có dạng như:

$$\hat{Y} = \beta_0 + \beta_1 X + \epsilon$$

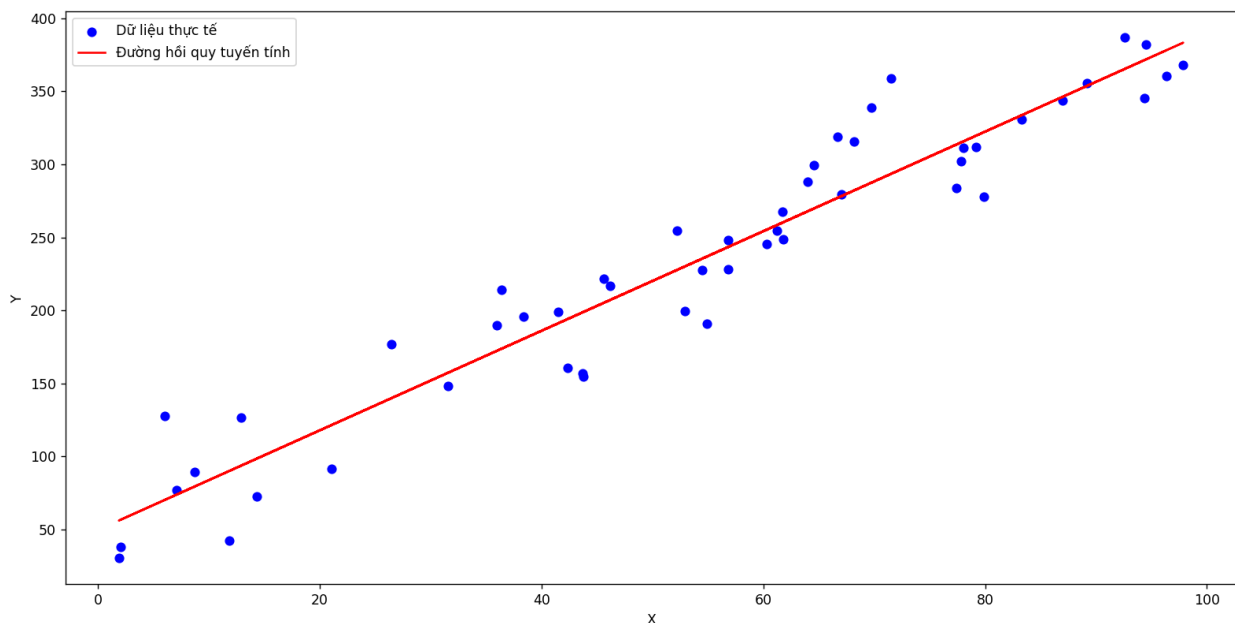
β_0 là hệ số chặn (intercept): đại diện cho giá trị của y khi x = 0.

β_1 là hệ số hồi quy (slope coefficient): cho biết mức độ thay đổi của y khi x thay đổi một đơn vị.

ϵ là phần dư (error term): đại diện cho sai số giữa giá trị thực tế và giá trị dự đoán.

Mô hình này thường được sử dụng phổ biến trong lĩnh vực phân tích và dự đoán giữa hai biến số, một số các ứng dụng như trong Y học, Kinh tế học, Giáo dục... Ví dụ đơn giản như tuổi và chiều cao ở học sinh, điểm số và học lực...

Dưới đây là một bài toán ví dụ về Hồi quy tuyến tính về dự đoán giá nhà với trục X là số M² và trục Y là giá nhà. Mỗi chấm xanh đại diện cho một dự đoán, chấm xanh càng gần với đường Hồi quy thì dự đoán đó càng chính xác với thực tế



Hình 2.1 Ảnh minh họa cách hoạt động của Hồi quy tuyến tính

2.2.2. Thuật toán Rừng ngẫu nhiên

2.2.2.1 Giới thiệu về Rừng ngẫu nhiên

Random Forest là một thuật toán học máy phổ biến được phát minh bởi Leo Breiman và Adele Cutler, sử dụng đầu ra của nhiều cây quyết định để đưa ra một kết quả

duy nhất. Do tính dễ sử dụng và linh hoạt, thuật toán này được chấp nhận rộng rãi và có thể áp dụng cho cả các bài toán phân loại và hồi quy. [8]

Vì mô hình Random Forest được cấu thành từ nhiều cây quyết định, nên trước hết việc giải thích ngắn gọn về thuật toán cây quyết định sẽ rất hữu ích. Cây quyết định cố gắng tìm ra phân tách tối ưu để chia dữ liệu, và thường được huấn luyện qua thuật toán cây phân loại và hồi quy (CART). Chất lượng của việc phân tách có thể được đánh giá bằng cách sử dụng các chỉ số như độ tạp Gini, lợi ích thông tin, hoặc sai số bình phương trung bình (MSE).

Cây quyết định là một thuật toán học có giám sát phổ biến, nhưng có thể gặp phải các vấn đề như thiên lệch hoặc quá khớp. Tuy nhiên, trong thuật toán Random Forest, nhiều cây quyết định tạo thành một tập hợp, và khi các cây này không liên quan lẫn nhau, có thể dự đoán kết quả chính xác hơn.

Thuật toán Random Forest mở rộng phương pháp Bagging bằng cách sử dụng cả Bagging và ngẫu nhiên hóa đặc trưng để tạo ra một tập hợp các cây quyết định không liên quan. Ngẫu nhiên hóa đặc trưng (hay còn gọi là Bagging đặc trưng hoặc phương pháp Random Subspace) tạo ra các tập con ngẫu nhiên của các đặc trưng, đảm bảo độ tương quan thấp giữa các cây quyết định. Đây là sự khác biệt chính giữa cây quyết định và Random Forest. Cây quyết định xem xét tất cả các phân tách đặc trưng có thể, trong khi Random Forest chỉ chọn một tập con của các đặc trưng. Bằng cách xem xét tất cả các biến động tiềm năng trong dữ liệu, Random Forest giảm thiểu rủi ro của quá khớp, thiên lệch và phương sai tổng thể, dẫn đến dự đoán chính xác hơn.

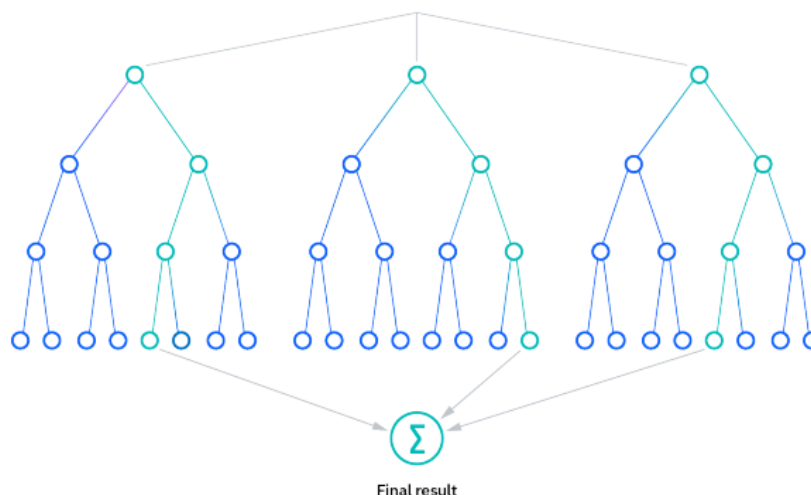
2.2.2.2 Cách thức hoạt động

Thuật toán rừng ngẫu nhiên có ba siêu tham số chính cần thiết lập trước khi huấn luyện. Đó là kích thước nút, số lượng cây và số lượng đặc trưng cần lấy mẫu. Với những điều này, rừng ngẫu nhiên có thể được sử dụng để giải quyết các vấn đề hồi quy hoặc phân loại.

Thuật toán rừng ngẫu nhiên bao gồm nhiều cây quyết định, mỗi cây được tạo từ các mẫu dữ liệu trích xuất từ một tập huấn luyện thay thế gọi là mẫu bootstrapping. Một phần ba của các mẫu huấn luyện này được giữ lại làm dữ liệu kiểm tra và gọi là mẫu out-

of-bag (oob). Sau đó, feature bagging được sử dụng để tạo thêm sự ngẫu nhiên, làm tăng sự đa dạng trong tập dữ liệu và giảm mối tương quan giữa các cây quyết định.

Tùy vào loại vấn đề, cách đưa ra dự đoán sẽ khác nhau. Đối với bài toán hồi quy, trung bình của các cây quyết định cá nhân sẽ được tính toán, còn đối với bài toán phân loại, phương pháp bỏ phiếu đa số sẽ được sử dụng, nghĩa là lớp được dự đoán là lớp xuất hiện nhiều nhất. Cuối cùng, các mẫu oob sẽ được sử dụng để kiểm tra chéo, hoàn thiện dự đoán.



Hình 2.2 Ảnh minh họa cách hoạt động của Rừng ngẫu nhiên

Các lợi ích khi sử dụng Rừng ngẫu nhiên:

- Giảm nguy cơ quá khớp: Cây quyết định có khả năng quá khớp với tất cả các mẫu dữ liệu huấn luyện. Tuy nhiên, trong Random Forest với số lượng cây quyết định đủ lớn, việc lấy trung bình các cây độc lập có thể giảm thiểu phương sai và lỗi dự đoán tổng thể, từ đó ngăn chặn việc mô hình bị quá khớp.

- Cung cấp tính linh hoạt: Random Forest được sử dụng rộng rãi trong lĩnh vực khoa học dữ liệu vì khả năng xử lý cả nhiệm vụ hồi quy và phân loại với độ chính xác cao. Bằng cách sử dụng phương pháp bagging đặc trưng (feature bagging), nó vẫn giữ được độ chính xác ngay cả khi một số dữ liệu bị thiếu, do đó trở thành công cụ hiệu quả để ước lượng giá trị thiếu.

- Dễ dàng xác định tầm quan trọng của đặc trưng: Random Forest giúp đánh giá đóng góp hay tầm quan trọng của các biến trong mô hình một cách dễ dàng. Có nhiều cách để đánh giá tầm quan trọng của đặc trưng. Gini importance và Mean Decrease in Impurity (MDI) thường được sử dụng để đo lường mức giảm độ chính xác khi loại bỏ

một biến cụ thể khỏi mô hình. Ngoài ra, còn có phương pháp Permutation Importance hay Mean Decrease Accuracy, trong đó giá trị của các đặc trưng trong mẫu oob được thay đổi ngẫu nhiên để xác định mức giảm độ chính xác trung bình.

Các thách thức khi sử dụng Rừng ngẫu nhiên:

- Quá trình tốn thời gian: Thuật toán Random Forest có khả năng xử lý các bộ dữ liệu lớn, cung cấp dự đoán chính xác hơn, nhưng do phải thực hiện tính toán trên từng cây quyết định riêng lẻ, quá trình xử lý dữ liệu có thể mất nhiều thời gian.

- Yêu cầu nhiều tài nguyên hơn: Random Forest xử lý các bộ dữ liệu lớn hơn, do đó cần nhiều tài nguyên hơn để lưu trữ dữ liệu đó.

- Phức tạp hơn: Dự đoán của một cây quyết định đơn lẻ dễ hiểu hơn so với dự đoán của Random Forest.

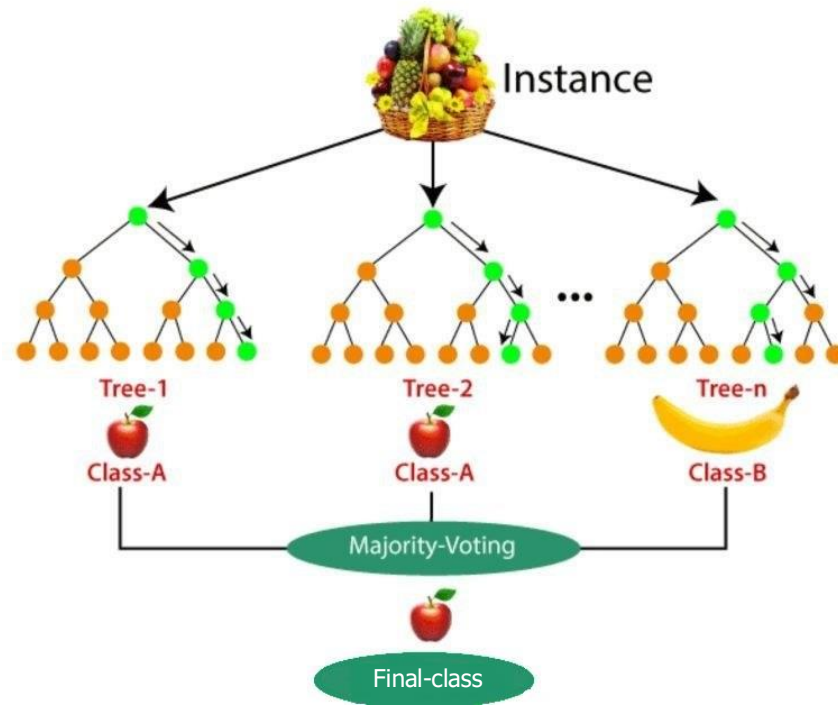
Một số lĩnh vực áp dụng Rừng ngẫu nhiên:

- Tài chính: Thuật toán này được ưa chuộng hơn các thuật toán khác vì nó giúp giảm thời gian dành cho quản lý và xử lý dữ liệu. Nó có thể được sử dụng để đánh giá khách hàng có tín dụng rủi ro cao, phát hiện gian lận và giải quyết các vấn đề về định giá quyền chọn.

- Y tế: Thuật toán Random Forest có thể được áp dụng trong sinh học tính toán, giúp các bác sĩ giải quyết các vấn đề như phân loại biểu hiện gen, phát hiện biomarker và chú thích trình tự. Do đó, bác sĩ có thể ước tính phản ứng thuốc đối với một loại thuốc cụ thể.

- Thương mại điện tử: Có thể sử dụng để xây dựng công cụ đề xuất sản phẩm bán hàng.

Dưới đây là một ví dụ cụ thể về dự đoán với mô hình Rừng ngẫu nhiên. Với tập dữ liệu mẫu gồm nhiều hình ảnh của các loại trái cây, tập dữ liệu này được chia cho bộ phận phân loại Rừng ngẫu nhiên. Tập dữ liệu được phân thành các tập con và được cấp cho cây quyết định. Trong giai đoạn huấn luyện mỗi cây quyết định sẽ cho ra một kết quả dự đoán. Khi đưa vào một ảnh trái cây mới thì dựa vào phần lớn kết quả dự đoán bộ phận loại Rừng ngẫu nhiên sẽ đưa ra kết quả cuối cùng.



Hình 2.2 Ví dụ về Rừng ngẫu nhiên [9]

2.3. Giới thiệu về một số thư viện sử dụng

Đối với thuật toán hồi quy tuyến tính, tôi có sử dụng thư viện Scikit-learn của Python với hàm LinearRegression từ module linear_model với cú pháp như sau:

```
from sklearn.linear_model import LinearRegression
```

LinearRegression trong scikit-learn là một lớp được sử dụng để triển khai mô hình hồi quy tuyến tính trong Python. Thuật toán này thích hợp để dự đoán giá trị của biến phụ thuộc dựa trên các biến độc lập, với giả định rằng mối quan hệ giữa các biến này là tuyến tính. [10]

Các đặc điểm chính của LinearRegression trong scikit-learn bao gồm:

- Dễ sử dụng và đơn giản: Cung cấp một phương pháp tiếp cận đơn giản để xây dựng và đánh giá mô hình hồi quy tuyến tính.
- Hỗ trợ nhiều ứng dụng thực tế: Phù hợp cho nhiều vấn đề như dự đoán, phân tích dữ liệu, và nghiên cứu mối quan hệ giữa các biến.
- Tối ưu hóa và hiệu suất: Được tối ưu hóa để đảm bảo hiệu suất cao khi xử lý các tập dữ liệu lớn.
- Tương thích với các công cụ và thư viện khác: Tương thích tốt với các thư viện chính như NumPy, Pandas, giúp trong việc xử lý và trực quan hóa dữ liệu.

LinearRegression là một công cụ quan trọng và cơ bản trong việc mô hình dữ liệu của scikit-learn, đóng vai trò quan trọng trong quá trình phân tích và mô hình hóa dữ liệu trong học máy.

Đối với thuật toán Rừng ngẫu nhiên, tôi sử dụng hàm RandomForestRegressor từ module ensemble của thư viện scikit-learn với cấu trúc như sau:

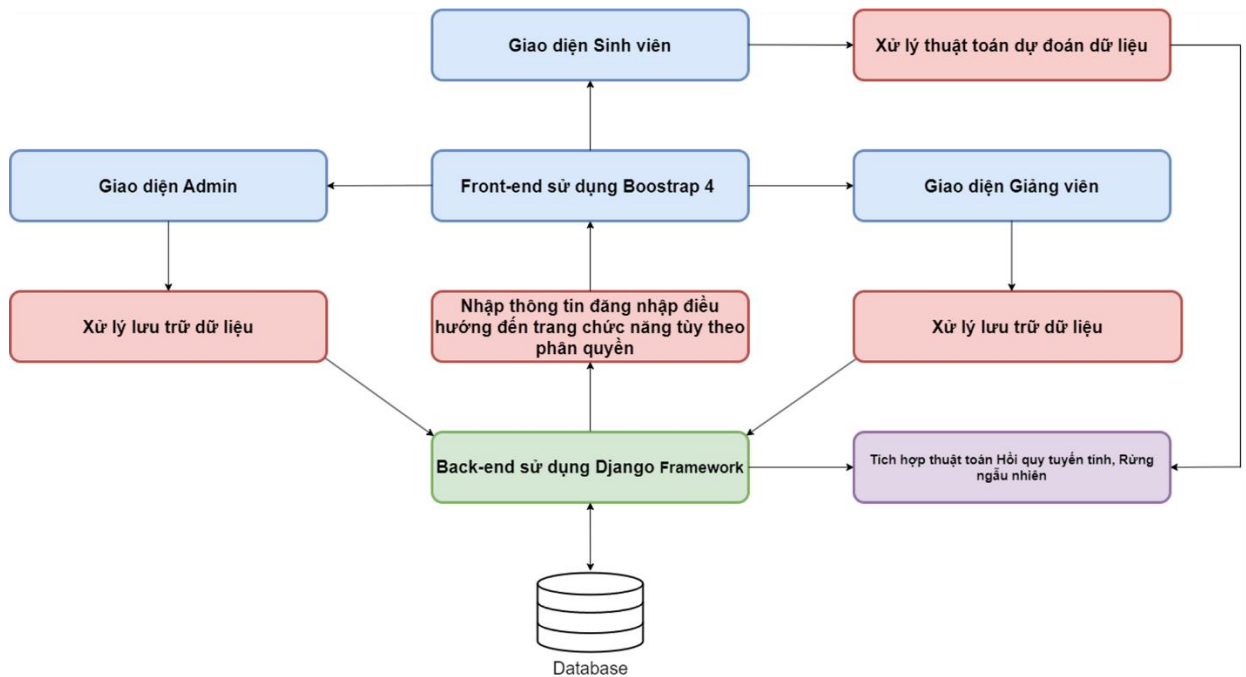
```
from sklearn.ensemble import RandomForestRegressor
```

RandomForestRegressor trong Python là một thuật toán thực hiện rừng ngẫu nhiên cho các bài toán hồi quy. Phương pháp này kết hợp nhiều cây quyết định để cải thiện độ chính xác của dự đoán và kiểm soát hiện tượng quá khớp. Dưới đây là những điểm chính về RandomForestRegressor: [10]

- Kết hợp cây quyết định: Mô hình này xây dựng nhiều cây quyết định trong quá trình học và lấy trung bình các dự đoán để cải thiện tính ổn định và độ chính xác.
- Giảm quá khớp: Bằng cách trung bình hóa dự đoán từ nhiều cây, RandomForestRegressor giúp giảm thiểu hiện tượng quá khớp so với việc sử dụng một cây quyết định đơn lẻ.
- Độ quan trọng của đặc trưng: Cung cấp thông tin về mức độ quan trọng của từng đặc trưng, giúp hiểu rõ hơn về các biến quan trọng đối với dự đoán.
- Xử lý mối quan hệ phi tuyến: Phù hợp để xử lý các mối quan hệ phức tạp và phi tuyến.
- Siêu tham số: Bằng cách điều chỉnh các siêu tham số như `n_estimators` (số lượng cây) và `max_depth` (độ sâu tối đa của cây), có thể tối ưu hóa hiệu suất của mô hình.

2.4. Mô hình hệ thống đề xuất

Từ các cơ sở lý thuyết đã nghiên cứu, tôi đề xuất mô hình hệ thống của đề tài Xây dựng hệ thống dự đoán điểm tích lũy cho sinh viên Trường đại học Trà Vinh như sau:



Hình 2.3 Mô hình hệ thống đề xuất

CHƯƠNG 3. HIỆN THỰC HÓA NGHIÊN CỨU

3.1. Mô tả đề tài

Đề tài tập trung nghiên cứu về việc quản lý điểm ở sinh viên ngành Công nghệ Thông tin, Khoa Kỹ thuật và Công nghệ, Trường Đại học Trà Vinh. Hệ thống có chức năng chính như quản lý dữ liệu điểm, danh sách sinh viên, dự đoán điểm tích lũy.

Hệ thống với ba phân quyền chính: Admin, cố vấn học tập, sinh viên. Mỗi phân quyền có các chức năng khác nhau. Người dùng bắt buộc đăng nhập trước khi sử dụng các chức năng của hệ thống.

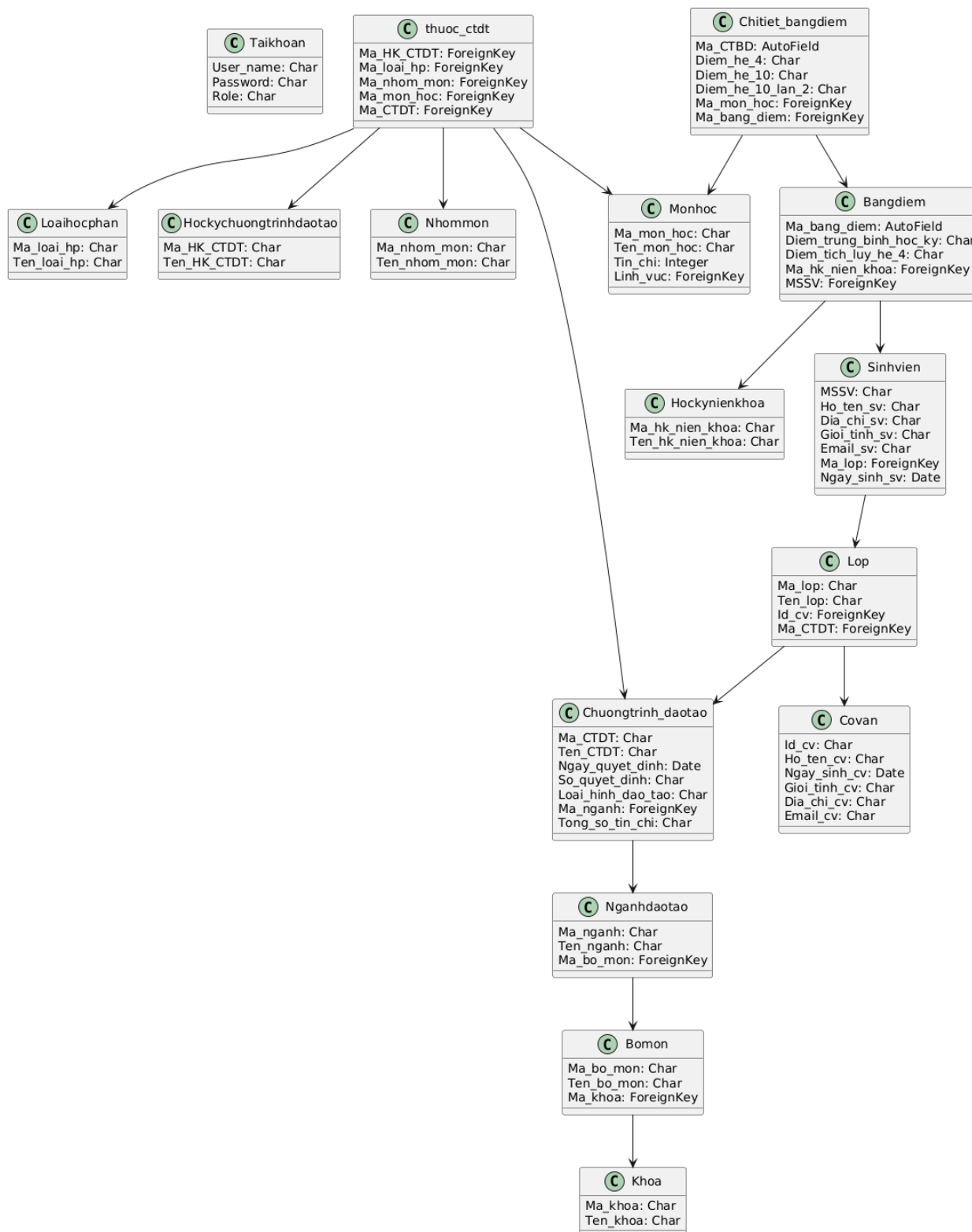
Đối với Admin: là cán bộ ở bộ môn Công nghệ Thông tin, có toàn quyền quản lý đối với các dữ liệu chung như khoa, bộ môn, ngành đào tạo, thông tin các chương trình đào tạo, môn học cho từng chương trình đào tạo, điểm số, danh sách sinh viên, danh sách cố vấn học tập,...

Đối với quyền cố vấn học tập: Là những cố vấn học tập được phân công quản lý và chịu trách nhiệm đối với một lớp học nhất định. Cố vấn học tập có các quyền như quản lý danh sách sinh viên trong lớp, danh sách điểm của từng sinh viên, nhập và chỉnh sửa điểm của sinh viên, thống kê xếp loại học lực, xem tiến độ học tập của cả lớp,...

Đối với quyền sinh viên: Sinh viên có những quyền trong hệ thống như xem và chỉnh sửa thông tin cá nhân, quản lý bảng điểm cá nhân, dự đoán điểm số, xem báo cáo thống kê điểm số, tiến độ học tập theo chương trình đào tạo.

3.2. Thiết kế dữ liệu

3.2.1. Sơ đồ lớp



Hình 3.1 Sơ đồ lớp của hệ thống

Chi tiết các bảng và mối quan hệ

Taikhoan: Đại diện cho thông tin tài khoản người dùng.

- User_name: Tên người dùng, là khóa chính.
- Password: Mật khẩu người dùng.
- Role: Vai trò của người dùng (Admin, Teacher, Student).

Covan: Đại diện cho thông tin cố vấn.

- Id_cv: ID cố vấn, là khóa chính.
- Ho_ten_cv, Ngay_sinh_cv, Gioi_tinh_cv, Dia_chi_cv, Email_cv: Các thông tin cá nhân của cố vấn.

Khoa: Đại diện cho các khoa trong trường.

- Ma_khoa: Mã khoa, là khóa chính.
- Ten_khoa: Tên khoa.

Bomon: Đại diện cho các bộ môn thuộc khoa.

- Ma_bo_mon: Mã bộ môn, là khóa chính.
- Ten_bo_mon: Tên bộ môn.
- Ma_khoa: Khóa ngoại trở tới Khoa.

Nganhdaotao: Đại diện cho các ngành đào tạo thuộc bộ môn.

- Ma_nganh: Mã ngành, là khóa chính.
- Ten_nganh: Tên ngành.
- Ma_bo_mon: Khóa ngoại trở tới Bomon.

Hockynienkhoa: Đại diện cho học kỳ niên khóa

- Ma_hk_nien_khoa: Mã học kỳ niên khóa (Khóa chính).
- Ten_hk_nien_khoa: Tên học kỳ niên khóa.

Hockychuongtrinhdaotao: Đại diện cho học kỳ chương trình đào tạo

- Ma_HK_CTDT: Mã học kỳ chương trình đào tạo (Khóa chính).
- Ten_HK_CTDT: Tên học kỳ chương trình đào tạo.

Loaihocphan: Đại diện cho loại học phần

- Ma_loai_hp: Mã loại học phần (Khóa chính).

- Ten_loai_hp: Tên loại học phần.

Nhommon: Đại diện cho nhóm môn

- Ma_nhom_mon: Mã nhóm môn (Khóa chính).
- Ten_nhom_mon: Tên nhóm môn.

Chuongtrinh_daotao: Đại diện cho chương trình đào tạo

- Ma_CTDT: Mã chương trình đào tạo (Khóa chính).
- Ten_CTDT: Tên chương trình đào tạo.
- Ngay_quyet_dinh: Ngày quyết định.
- So_quyet_dinh: Số quyết định.
- Loai_hinh_dao_tao: Loại hình đào tạo.
- Ma_nganh: Mã ngành (liên kết tới Nganhdaotao).
- Tong_so_tin_chi: Tổng số tín chỉ.

Lop: Đại diện cho các lớp học thuộc chương trình đào tạo.

- Ma_lop: Mã lớp, là khóa chính.
- Ten_lop: Tên lớp.
- Id_cv: Khóa ngoại trở tới Covan.
- Ma_CTDT: Khóa ngoại trở tới Chuongtrinh_daotao.

Sinhvien: Đại diện cho sinh viên

- MSSV: Mã số sinh viên (Khóa chính).
- Ho_ten_sv: Họ tên sinh viên.
- Dia_chi_sv: Địa chỉ sinh viên.
- Gioi_tinh_sv: Giới tính sinh viên.
- Email_sv: Email sinh viên.
- Ma_lop: Mã lớp (liên kết tới Lop).
- Ngay_sinh_sv: Ngày sinh sinh viên.

Bangdiem: Đại diện cho bảng điểm

- Ma_bang_diem: Mã bảng điểm (Khóa chính, tự động tăng).
- Diem_tich_luy_he_4: Điểm tích lũy hệ 4.
- Ma_hk_nien_khoa: Mã học kỳ niên khóa (liên kết tới Hockynienkhoa).

- MSSV: Mã số sinh viên (liên kết tới Sinhvien).

Monhoc: Đại diện cho môn học

- Ma_mon_hoc: Mã môn học (Khóa chính).
- Ten_mon_hoc: Tên môn học.
- Tin_chi: Số tín chỉ.
- Linh_vuc: Lĩnh vực (liên kết tới chính nó).

Chitiet_bangdiem: Đại diện cho chi tiết của bảng điểm

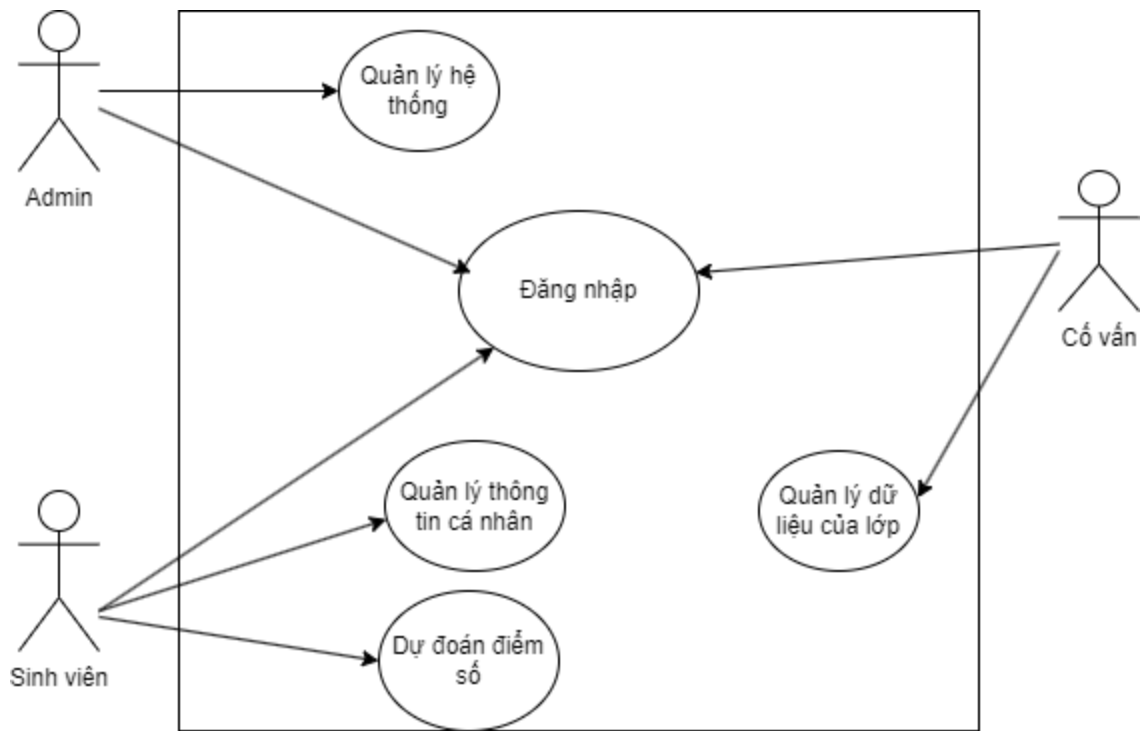
- Ma_CTBD: Mã chi tiết bảng điểm (Khóa chính, tự động tăng).
- Diem_he_4: Điểm hệ 4.
- Diem_he_10: Điểm hệ 10.
- Diem_he_10_lan_2: Điểm hệ 10 lần 2.
- Ma_mon_hoc: Mã môn học (liên kết tới Monhoc).
- Ma_bang_diem: Mã bảng điểm (liên kết tới Bangdiem).

thuoc_ctdt: Đại diện cho bảng chỉ các môn học thuộc chương trình đào tạo

- Ma_thuoc_CT_DT: Mã thuộc CTDT (Khóa chính).
- Ma_HK_CTDT: Mã học kỳ chương trình đào tạo (liên kết tới Hockychuongtrinhdaotao).
- Ma_loai_hp: Mã loại học phần (liên kết tới Loaihocphan).
- Ma_nhom_mon: Mã nhóm môn (liên kết tới Nhommon).
- Ma_CTDT: Mã chương trình đào tạo (liên kết tới Chuongtrinh_daotao).
- Ma_mon_hoc: Mã môn học (liên kết tới Monhoc)

3.2.2. Sơ đồ Use case

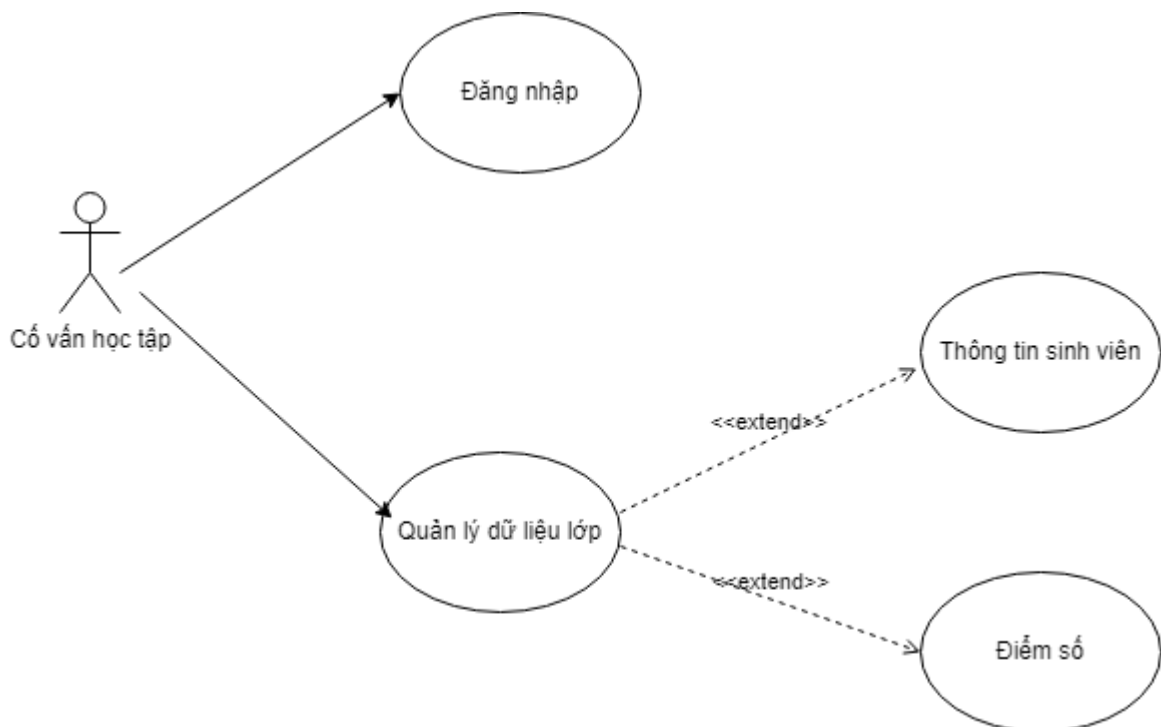
Use case tổng quát của hệ thống



Hình 3.2 Sơ đồ tổng quát hệ thống

Mô tả: Người dùng bắt buộc phải đăng nhập với tài khoản đã được cấp, mỗi tài khoản sẽ có những phân quyền riêng, tùy vào phân quyền sẽ được sử dụng các chức năng khác nhau.

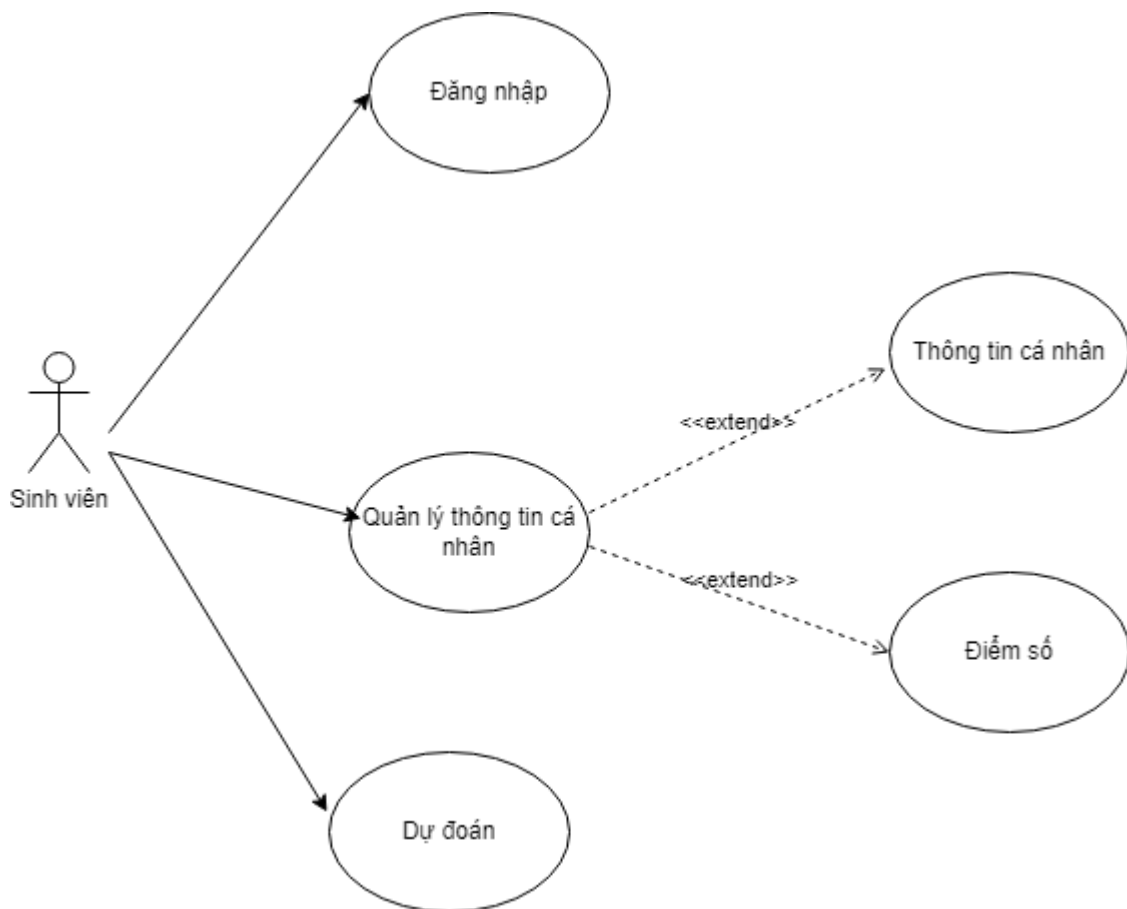
Use case phân rã chức năng Cố vấn học tập



Hình 3.3 Sơ đồ Use case phân rã chức năng Cố vấn học tập

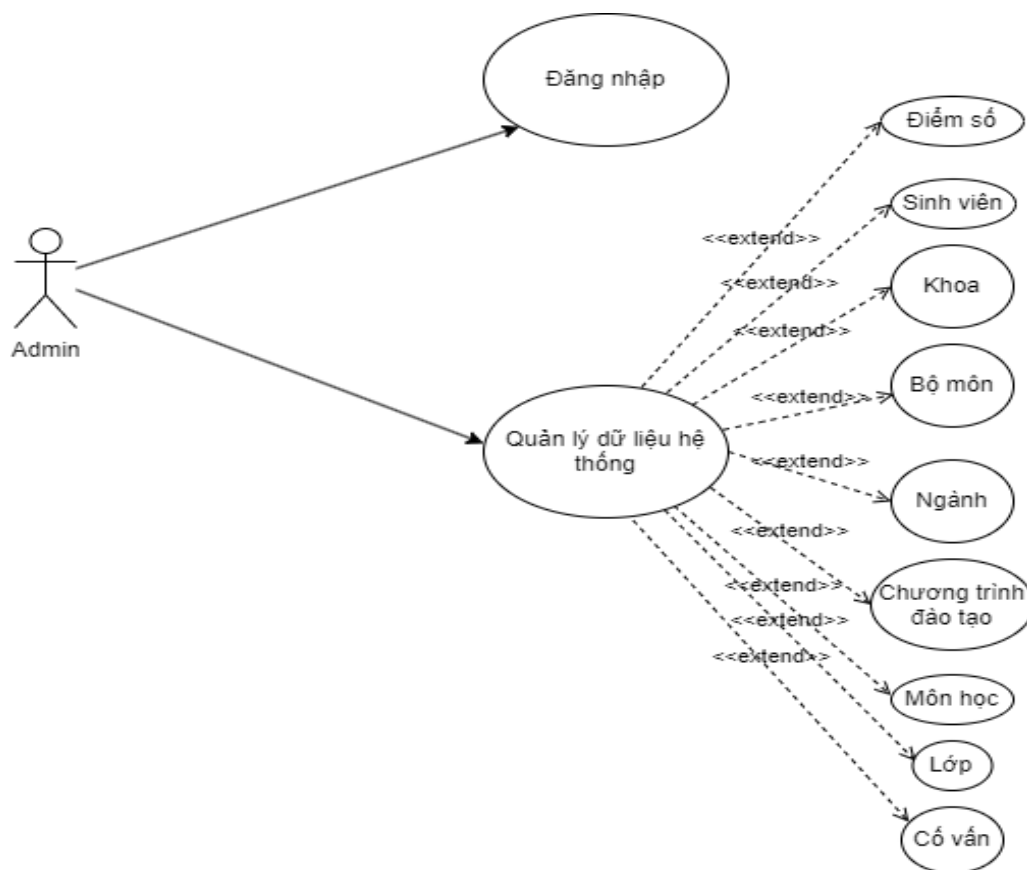
Mô tả: Cố vấn học tập sau khi đăng nhập vào hệ thống bằng tài khoản được cấp, cố vấn học tập có quyền quản lý thông tin lớp học, sinh viên, điểm số của sinh viên.

Use case phân rã chức năng Sinh viên



Hình 3.4 Sơ đồ Use case phân rã chức năng Sinh viên

Mô tả: Sinh viên sau khi đăng nhập vào hệ thống bằng tài khoản được cấp sẽ có quyền quản lý được thông tin cá nhân của sinh viên, điểm cá nhân và sử dụng chức năng dự đoán điểm số.



Hình 3.5 Sơ đồ Use case phân rã chức năng Admin

Tên use case: Quản lý hệ thống của Admin

Actor: Admin

Mô tả use case: Use case này cho phép Admin quản lý thêm, sửa, xóa dữ liệu của khoa, bộ môn, ngành, lớp học, sinh viên, môn học, điểm sinh viên, cổ vấn học tập, chương trình đào tạo.

3.3. Xây dựng mô hình đề xuất

3.3.1. Dữ liệu đầu vào

Hệ thống sử dụng điểm hệ 10 của toàn bộ sinh viên ngành Công nghệ Thông tin thuộc bộ môn Công nghệ Thông tin khoa Kỹ thuật và Công nghệ năm 2019 và 2020.

Điểm được lấy theo môn học và được chia theo cùng chương trình đào tạo, cùng nhóm môn, cùng loại học phần, cùng học kỳ chương trình đào tạo.

Dữ liệu mẫu được xử lý sắp xếp với mỗi cột là một môn và mỗi dòng là một sinh viên.

Dữ liệu đầu vào của hệ thống là bảng điểm toán khóa của sinh viên được trích xuất từ phần mềm quản lý Edusoft của Trường Đại học Trà Vinh có dạng như sau:

MaSV	HoLotSV	TenSV	Phai	NgaySinhC	NoiSinh	MaLop	MH006	MH007	MH008	MH009	MH010	MH011
Mã SV	Họ lót	Tên	Phái	Ngày sinh	Nơi sinh	Lớp	Ví tích phần A1 (110000)	Đại số tuyến tính (110001)	Ví tích phần A2 (110002)	Toán rời rạc (110003)	Xác suất thống kê (110006)	Ví tích phần A1 (110042)
110119060	Lý Thế	Vinh	0	12/08/2001	Tỉnh Trà Vinh	DA19TTA	3	2	2	2	2	3
110119080	Lý Kim	Lam	0	06/05/2001	Tỉnh Trà Vinh	DA19TTA	5.6	8.9			6.7	
110119036	Huỳnh Phi	Nhật	0	04/05/2001	Tỉnh Trà Vinh	DA19TTA	5.8	8.3	4.4	4.7	4.3	
110119146	Nguyễn Quốc	Dũng	0	11/08/2001	Tỉnh Thanh	DA19TTB	6.2	6.7	7.0	5.7	5.4	
110119061	Nguyễn Lâm Nguyên	Vũ	0	06/12/2001	Tỉnh Trà Vinh	DA19TTA	5.9	9.2	4.5	4.4	5.2	
110120038	Trần Thị Mỹ	Khánh	1	02/02/2002	Tỉnh Trà Vinh	DA20TTB		5.7	6.3	5.7		4.1
110119033	Ngô Thanh	Nguyễn	0	29/06/2001	Tỉnh Trà Vinh	DA19TTB	8.9	8.7	8.7	6.5	8.5	
110119147	Sơn Sô	Ni	0	01/01/2001	Tỉnh Trà Vinh	DA19TTB	5.1	4.3	5.3	4.5	7.4	
110119145	Ngô Ngọc Diệu	Hiền	1	30/11/2001	Tỉnh Trà Vinh	DA19TTB	6.0	6.5	8.4	6.1	7.1	
110120081	Nguyễn	Triển	0	17/02/2002	Tỉnh Trà Vinh	DA20TTA		9.3	7.5	8.0		8.7
110119144	Lê Tấn	Hỷ	0	13/04/2001	Tỉnh Trà Vinh	DA19TTB	6.9	6.5	6.5	5.9	7.1	
110120004	Trần Tiến	Anh	0	03/11/2002	Tỉnh Trà Vinh	DA20TTA		7.6	8.5	6.4		5.8
110120082	Trần Vỹ	Triết	0	26/08/2002	Tỉnh Trà Vinh	DA20TTA		5.3	5.9	5.3		6.8
110119038	Phan Minh	Nhựt	0	14/03/2001	Tỉnh Trà Vinh	DA19TTA	5.3	8.3	5.5	6.7	4.3	
110119058	Nguyễn Nhật	Trường	0	01/01/2001	Tỉnh Trà Vinh	DA19TTA	5.5	6.7	8.3	6.4	6.3	
113918001	Lê Anh	Hào	0	26/11/2000	Tỉnh Bạc Liêu	DA19TTA		6.0		6.6	8.3	6.3
110120077	Trần Quang	Tiến	0	04/09/2002	Tỉnh Trà Vinh	DA20TTA		8.2	7.8	7.4		7.4
110119006	Nguyễn Minh	Chiến	0	24/12/2001	Tỉnh Trà Vinh	DA19TTA	4.9	6.6	5.5	4.9	6.7	
110120108	Huỳnh Hữu	Nhân	0	20/06/2000	Tỉnh Trà Vinh	DA20TTB		6.5	5.8	VT		5.5
110119054	Trần Minh	Triết	0	07/11/2001	Tỉnh Trà Vinh	DA19TTA	VT					
110120060	Kim Thị Sô	Phi	1	11/01/2000	Tỉnh Trà Vinh	DA20TTA		6.6	7.3	6.9		6.6
110120132	Lê Tấn	Dạng	0	30/09/2002	Tỉnh Trà Vinh	DA20TTB		4.9		VT		VT
110119076	Nguyễn Mai	Thương	0	23/01/2001	Tỉnh Trà Vinh	DA19TTB	VT	6.4	VT	VT	VT	
110120011	Nguyễn Huỳnh	Cương	0	15/11/2002	Tỉnh Trà Vinh	DA20TTA						VT
110120170	Phạm Trọng	Nghĩa	0	13/12/2002	Tỉnh Trà Vinh	DA20TTA		5.6	5.5	4.1		6.0
110120006	Đặng Kim	Bắc	0	14/06/2002	Tỉnh Trà Vinh	DA20TTA		5.8	7.1	4.0		5.5

Hình 3.6 Dữ liệu mẫu được trích xuất từ Edusoft

3.3.2. Tiền xử lý dữ liệu

Sau khi có được dữ liệu mẫu được trích xuất từ Edusoft cần phải có những bước để loại bỏ những dữ liệu không cần thiết như Mã số sinh viên, tên môn,... Quá trình tiền xử lý bao gồm những bước sau:

Bước 1: Loại bỏ những cột dữ liệu không cần thiết như các cột chứa các thông tin cá nhân của sinh viên

MaSV	HoLotSV	TenSV	Phai	NgaySinhC	NoiSinh	MaLop
Mã SV	Họ lót	Tên	Phái	Ngày sinh	Nơi sinh	Lớp
110119060	Lý Thế	Vinh	0	12/08/2001	Tỉnh Trà Vinh	DA19TTA
110119080	Lý Kim	Lam	0	06/05/2001	Tỉnh Trà Vinh	DA19TTA
110119036	Huỳnh Phi	Nhật	0	04/05/2001	Tỉnh Trà Vinh	DA19TTA
110119146	Nguyễn Quốc	Dũng	0	11/08/2001	Tỉnh Thanh Hóa	DA19TTB
110119061	Nguyễn Lâm Nguyên	Vũ	0	06/12/2001	Tỉnh Trà Vinh	DA19TTA
110120038	Trần Thị Mỹ	Khánh	1	02/02/2002	Tỉnh Trà Vinh	DA20TTB
110119033	Ngô Thanh	Nguyễn	0	29/06/2001	Tỉnh Trà Vinh	DA19TTB
110119147	Sơn Sô	Ni	0	01/01/2001	Tỉnh Trà Vinh	DA19TTB
110119145	Ngô Ngọc Diệu	Hiền	1	30/11/2001	Tỉnh Trà Vinh	DA19TTB
110120081	Nguyễn	Triển	0	17/02/2002	Tỉnh Trà Vinh	DA20TTA
110119144	Lê Tấn	Hỷ	0	13/04/2001	Tỉnh Trà Vinh	DA19TTB
110120004	Trần Tiến	Anh	0	03/11/2002	Tỉnh Trà Vinh	DA20TTA
110120082	Trần Vỹ	Triết	0	26/08/2002	Tỉnh Trà Vinh	DA20TTA
110119038	Phan Minh	Nhựt	0	14/03/2001	Tỉnh Trà Vinh	DA19TTA
110119058	Nguyễn Nhật	Trường	0	01/01/2001	Tỉnh Trà Vinh	DA19TTB
113918001	Lê Anh	Hào	0	26/11/2000	Tỉnh Bạc Liêu	DA19TTA
110120077	Trần Quang	Tiến	0	04/09/2002	Tỉnh Trà Vinh	DA20TTA
110119006	Nguyễn Minh	Chiến	0	24/12/2001	Tỉnh Trà Vinh	DA19TTA
110120108	Huỳnh Hữu	Nhân	0	20/06/2000	Tỉnh Trà Vinh	DA20TTB

Hình 3.7 Những dữ liệu không cần thiết

Bước 2: Đối với dòng chứa tên môn học và mã môn học thì loại bỏ tên môn học chỉ giữ lại mã môn học, mã môn học này tương đồng với mã môn được lưu trong cơ sở dữ liệu của hệ thống. Ví dụ như:

Bảng 3.1. Bảng ví dụ về tiền xử lý dữ liệu môn học

Dữ liệu trước khi xử lý	Dữ liệu sau khi xử lý
MH008 Vi tích phân A2 (110002)	110002
MH009 Toán rời rạc (110003)	110003

Bước 3: Xử lý các giá trị chuỗi hoặc các giá trị bị trống. Trong dữ liệu mẫu được trích xuất từ Edusoft sẽ bao gồm những giá trị bằng kiểu chữ như: VT(vắng thi), MT(miễn thi),... và một số những liệu trống. Tiến hành thay thế các ô có dữ liệu bằng chữ bằng ô trống, sau khi thay thế tất cả dữ liệu chữ thì tiếp tục thay thế các ô trống (bao gồm cả ô trống thay thế cho dữ liệu chữ trước đó) bằng giá trị điểm trung bình của môn học đó.

192.08	192.09	192.11	192.13	193.13	193.14	193.15
1	1	1	1	1	1	1
7.7					6.5	
6					6.5	
7.2					7.7	
6.2					6	
		8.5				
		5.4				
6.5					8.7	
		6.8				
					5.3	
		7.7			6.3	
7.5						6.5
9.3						9
6.8					5.3	
		8.5			6.9	
	6.8				VT	
6.5				9.5		
7						
VT						VT
7.2					5.8	
		VT				VT

Hình 3.8 Dữ liệu mẫu bị lỗi

3.3.3. Xây dựng mô hình

Sau khi trải qua các bước tiền xử lý tiến hành xây dựng mô hình dự đoán điểm. Dữ liệu mẫu còn lại sẽ bao gồm dòng đầu tiên là mã môn học ứng với môn học được lưu trong cơ sở dữ liệu của hệ thống, mỗi dòng tiếp theo là dữ liệu điểm của tất cả môn của một sinh viên. Dữ liệu mẫu của sinh viên ngành Công nghệ Thông tin khóa 2019-2020 sau quá trình tiền xử lý với 395 dòng và 89 cột.

Chia tập dữ liệu thành hai phần: Tập huấn luyện với tỉ lệ 70% và tập kiểm tra với tỉ lệ là 30%. Tập huấn luyện sẽ được sử dụng để xây dựng mô hình, còn tập kiểm tra sẽ được sử dụng để đánh giá mô hình.

Khi nhận được thông tin môn cần dự đoán tiến hành truy vấn dữ liệu môn học đang ở học kỳ thứ mấy và lọc các thông tin về các kỳ học trước đó cùng với các môn học đã có điểm. Sau khi đã truy vấn được các môn học cần thiết tiến hành lấy điểm 9 môn tuyển tính nhất với điểm môn cần được dự đoán (sử dụng hàm Selectkbest với $K = 9$)

Sử dụng tập dữ liệu kiểm tra để đánh giá hiệu suất của mô hình Hồi quy tuyến tính và Rừng ngẫu nhiên trong việc dự đoán. Các chỉ số đánh giá độ chính xác và đáng tin cậy gồm R-squared (R^2), Mean Squared Error (MSE) cho cả 2 mô hình dự đoán.

Bảng 3.2. Bảng dự đoán một số môn

Biến độc lập (X)		Biến phụ thuộc (Y)		Hồi quy tuyến tính		Ràng ngẫu nhiên	
Mã môn	Tên môn	Mã môn	Tên môn	R ²	MSE	R ²	MSE
180001	Tư tưởng Hồ Chí Minh	220034	Chuyên đề Linux	0.69	0.06	0.71	0.06
190081	Học phần I: Đường lối QP và an ninh của ĐCSVN						
190082	Học phần II: Công tác quốc phòng và an ninh						
190083	Học phần III: Quân sự chung						
190084	Học phần IV: Kỹ thuật chiến đấu bộ binh và chiến thuật						
220018	Mạng máy tính						
220092	Nhập môn công nghệ thông tin						
220096	Cơ sở dữ liệu						

Biến độc lập (X)		Biến phụ thuộc (Y)		Hồi quy tuyến tính		Ràng ngẫu nhiên	
Mã môn	Tên môn	Mã môn	Tên môn	R ²	MSE	R ²	MSE
110042	Vi tích phân A1						
220239	Phân tích và thiết kế hệ thống thông tin	220071	Lập trình thiết bị di động	0.76	0.10	0.67	0.13
110003	Toán rời rạc						
180052	Chủ nghĩa xã hội khoa học						
190083	Học phần III: Quân sự chung						
190084	Học phần IV: Kỹ thuật chiến đấu bộ binh và chiến thuật						
220034	Chuyên đề Linux						
220055	Công nghệ phần mềm						
220060	Hệ quản trị cơ sở dữ liệu						

Biến độc lập (X)		Biến phụ thuộc (Y)		Hồi quy tuyến tính		Ràng ngẫu nhiên	
Mã môn	Tên môn	Mã môn	Tên môn	R ²	MSE	R ²	MSE
220086	Lập trình ứng dụng trên Windows						
110002	Vì tích phân A2	220109	Khai khoáng dữ liệu	0.60	0.51	0.67	0.43
110079	Kiến trúc máy tính						
170011	Tiếng Việt thực hành						
220065	Thương mại điện tử						
220086	Lập trình ứng dụng trên Windows						
220096	Cơ sở dữ liệu						
220099	Lập trình hướng đối tượng						
220233	Đại số đại cương						
420000	Kỹ thuật XD & ban hành văn bản						
110003	Toán rời rạc	220060	Hệ quản trị cơ sở dữ liệu	0.59	0.27	0.57	0.28

Biến độc lập (X)		Biến phụ thuộc (Y)		Hồi quy tuyến tính		Ràng ngẫu nhiên	
Mã môn	Tên môn	Mã môn	Tên môn	R ²	MSE	R ²	MSE
180001	Tư tưởng Hồ Chí Minh						
180053	Lịch sử Đảng Cộng sản Việt Nam						
220055	Công nghệ phần mềm						
220071	Lập trình thiết bị di động						
220086	Lập trình ứng dụng trên Windows						
220228	Kỹ thuật lập trình						
220239	Phân tích và thiết kế hệ thống thông tin						
320045	Thống kê và phân tích dữ liệu						
110042	Vị tích phân A1	220101	Hệ điều hành	0.67	0.51	0.64	0.56
110057	Quy hoạch tuyến tính						

Biến độc lập (X)		Biến phụ thuộc (Y)		Hồi quy tuyến tính		Ràng ngẫu nhiên	
Mã môn	Tên môn	Mã môn	Tên môn	R ²	MSE	R ²	MSE
180050	Triết học Mác – Lênin						
220099	Lập trình hướng đối tượng						
220100	Lý thuyết đồ thị						
290000	Phương pháp NC khoa học						
410291	Anh văn không chuyên 1						
410292	Anh văn không chuyên 2						
410294	Anh văn không chuyên 4						

3.4. Cài đặt Django

Cài đặt Django : Mở trình soạn thảo đang sử dụng để viết mã nguồn, mở terminal (hoặc Command Prompt trên Windows) và chạy lên sau để bắt đầu cài đặt Django:

```
pip install django
```

Tạo dự án: Sau khi cài đặt Django thành công, chạy câu lệnh sau để bắt đầu với một dự án Django mới:

```
django-admin startproject project_name
```

project_name là tên dự án người dùng tùy chọn, có thể thay đổi tùy ý của người dùng. Ví dụ muốn tạo một dự án có tên là myproject

```
django-admin startproject myproject
```

Các thư mục cần chú ý: Khi tạo thành công dự án mới, trong thư mục dự án có chứa một số thư mục con cần chú ý như:

init.py:c

- Đây là một tệp trống, nhưng nó làm cho thư mục được coi là một gói Python.
- Điều này cho phép nhập các mô-đun từ thư mục này trong các tệp Python khác.

settings.py:

- Tệp này chứa tất cả các cài đặt và cấu hình cho project Django.
- Chứa các cài đặt liên quan đến cơ sở dữ liệu, ứng dụng cài đặt, middleware, templates, tệp tĩnh, và nhiều thứ khác.

urls.py:

- Tệp này định nghĩa các URL patterns cho project.
- Sử dụng tệp này để ánh xạ các URL đến các view chức năng trong ứng dụng.

wsgi.py:

- Tệp này là một entry-point cho các server web để phục vụ project.

- WSGI là một tiêu chuẩn giao diện giữa server web và các ứng dụng web được viết bằng Python.
- Tập này thường được sử dụng khi triển khai project lên server.

asgi.py:

- Tập này tương tự như wsgi.py, nhưng dành cho ASGI (Asynchronous Server Gateway Interface), một tiêu chuẩn mới hơn cho các ứng dụng web Python hỗ trợ cả giao diện đồng bộ và bất đồng bộ.
- Tập này hữu ích khi muốn triển khai các tính năng bất đồng bộ trong Django.

```
Project_name/
|- manage.py
|- project_name/
    |- _init_.py
    |- settings.py
    |- urls.py
    |- wsgi.py
    |- asgi.py
```

Hình 3.9 Cây thư mục sau khi tạo dự án

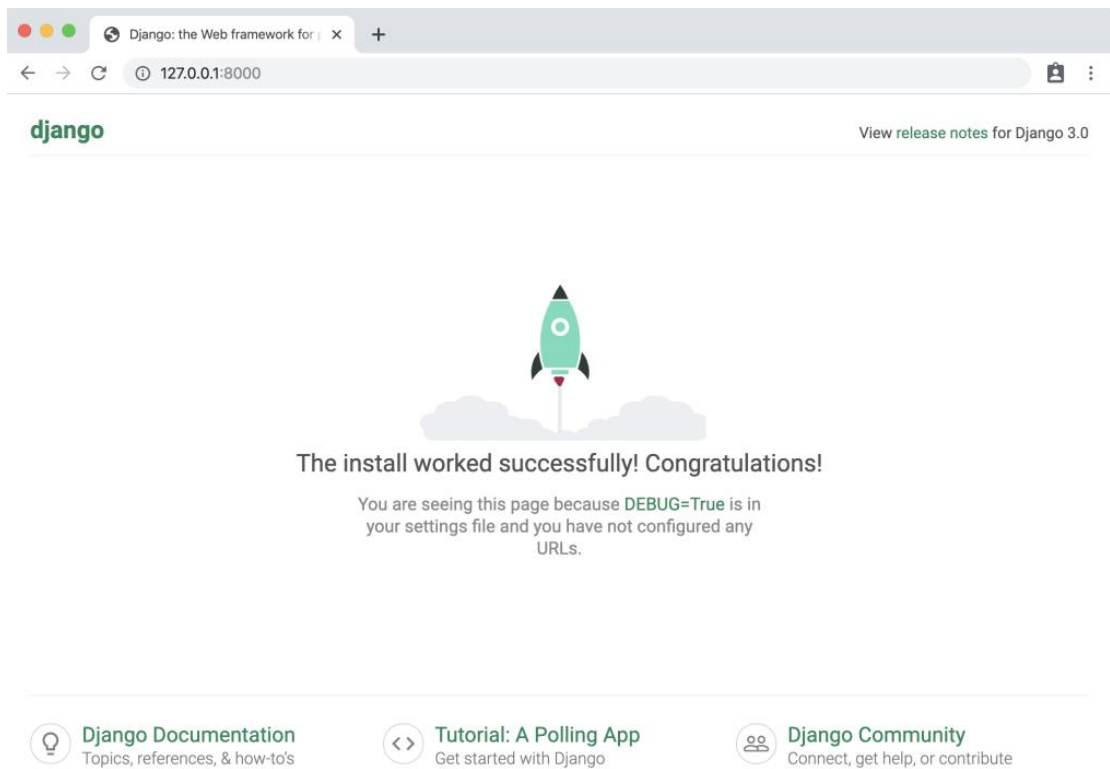
Di chuyển vào thư mục dự án: Sử dụng terminal để điều hướng vào thư mục của project Django và chạy câu lệnh sau để có thể thực thi trên dự án vừa tạo:

```
cd myproject
```

Chạy server phát triển: Sau khi di chuyển vào thư mục dự án vừa tạo, chạy câu lệnh sau để có thể khởi động server

```
Python manage.py runserver
```

Sau khi server khởi động thành công, mở trình duyệt và truy cập vào <http://127.0.0.1:8000/> để xem trang chủ của dự án Django. Nếu nhận được kết quả hiển thị như ảnh bên dưới thì Django đã được cài đặt và khởi động thành công.



Hình 3.10 Trang chủ dự án Django

3.5. Phác thảo giao diện

3.5.1. Phác thảo giao diện đăng nhập

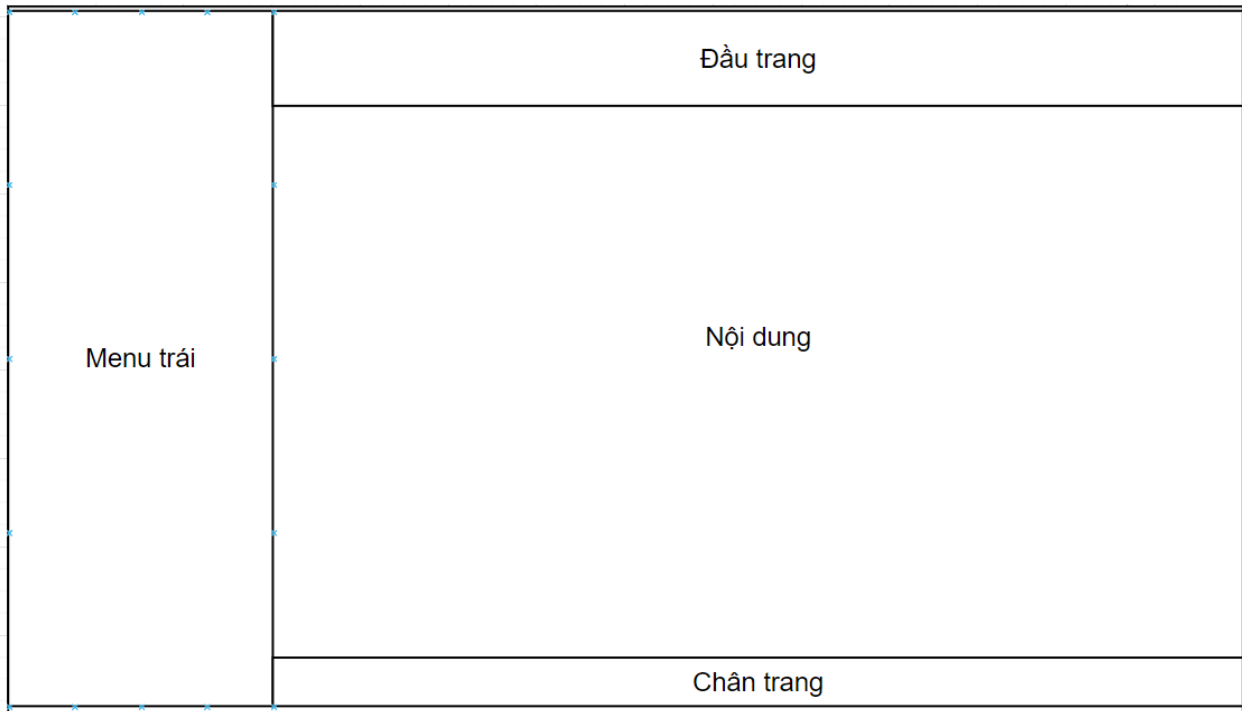
Người dùng bắt buộc thực hiện đăng nhập với tài khoản được cấp và được chuyển về giao diện tương ứng theo phân quyền của tài khoản

A wireframe diagram of a login interface. It features a central circle labeled 'Logo'. Below the logo, the text 'HỆ THỐNG DỰ ĐOÁN ĐIỂM TÍCH LŨY CHO SINH VIÊN TRƯỜNG ĐẠI HỌC TRÀ VINH' is displayed. Underneath this text are two input fields: the first is labeled 'Tên đăng nhập' and the second is labeled 'Mật khẩu'. Below the input fields is a button labeled 'Đăng nhập'.

Hình 3.11 Phác thảo giao diện đăng nhập

3.5.2. Phác thảo giao diện trang chủ dùng chung

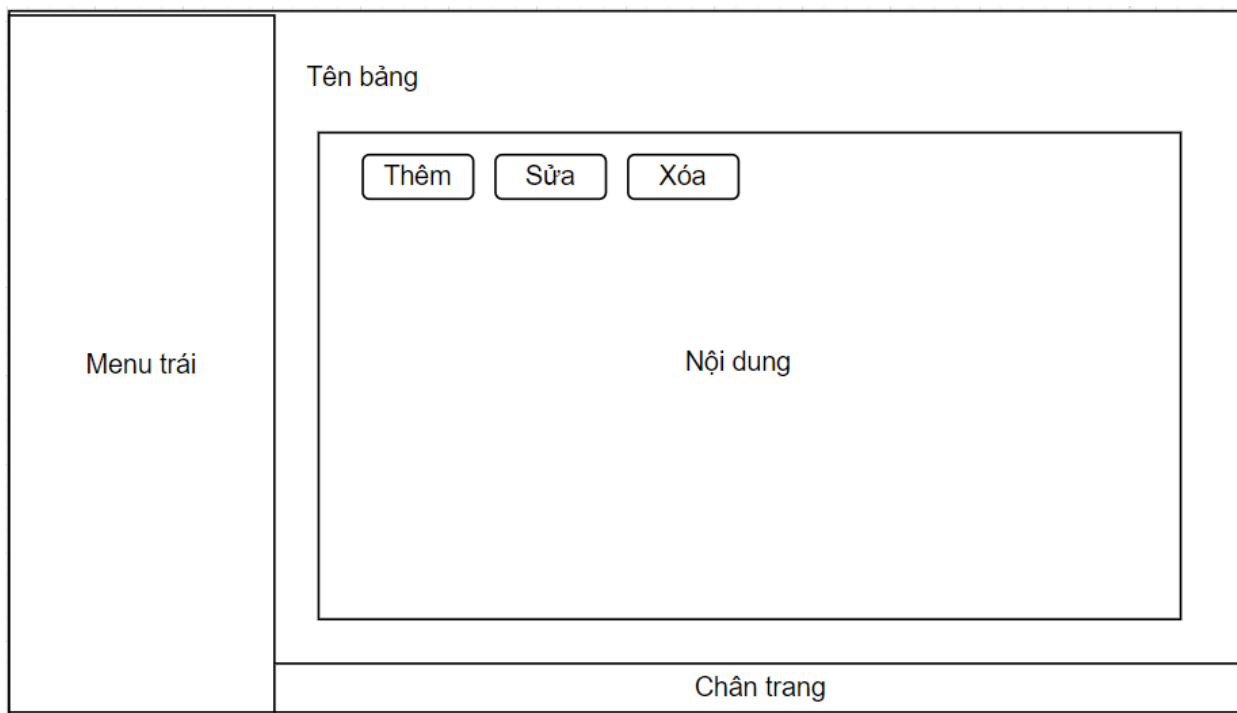
Sau khi đăng nhập hệ thống người dùng sẽ được chuyển đến trang chủ của hệ thống tùy theo phân quyền. Các phân quyền có chung giao diện trang chủ gồm các phần: Phần menu trái chứa các đường dẫn đến các chức năng chính của hệ thống, phần đầu trang chứa các thông kê cơ bản tùy theo mỗi phân quyền, phần nội dung chính của trang web và phần chân trang chứa một số thông tin khác.



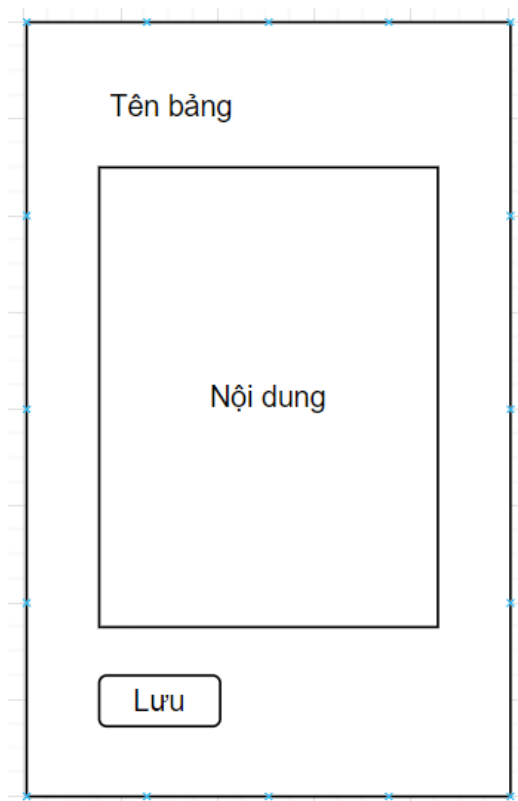
Hình 3.12 Phác thảo giao diện trang chủ chung

3.5.3. Phác thảo giao diện thêm, sửa, xóa dùng chung

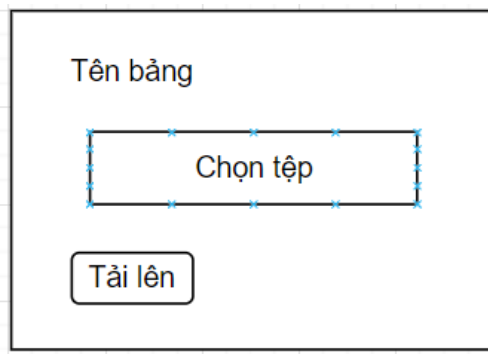
Với những phân quyền quản lý dữ liệu hệ thống sẽ có giao diện chức năng thêm, sửa, xóa chung. Đối với chức năng thêm có thể lựa chọn thêm bằng tệp để có thể thêm cùng lúc nhiều dữ liệu bằng tệp excel.



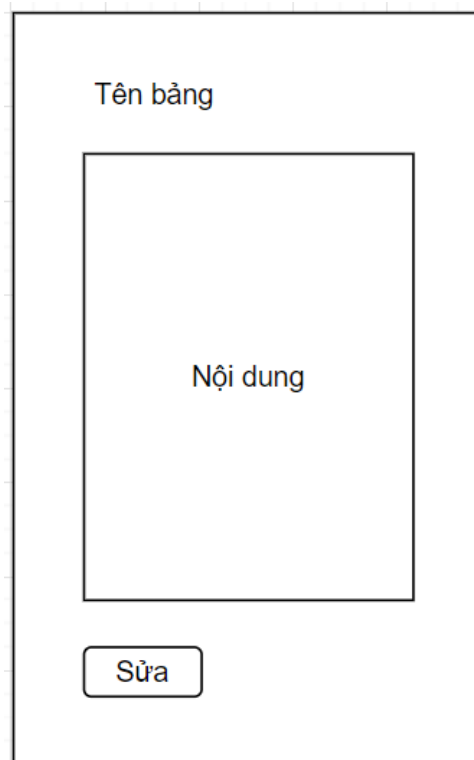
Hình 3.13 Phác thảo giao diện thêm, sửa, xóa chung



Hình 3.14 Phác thảo giao diện thêm dữ liệu



Hình 3.15 Phác thảo giao diện thêm dữ liệu bằng tệp



Hình 3.16 Phác thảo giao diện sửa dữ liệu

3.5.4. Phác thảo giao diện cho sinh viên

Chức năng xem điểm của hệ thống người dùng có thể lựa chọn xem tất cả các môn học có điểm hoặc lựa chọn học kỳ muốn xem. Sinh viên chỉ được xem điểm của cá nhân.

Tên sinh viên	Thanh lựa chọn học kỳ	Xem
	Nội dung	
Menu trái	Chân trang	

Hình 3.17 Phác thảo giao diện xem điểm theo học kỳ

Đối với chức năng dự đoán điểm cho sinh viên hệ thống sẽ hiển thị toàn bộ môn học thuộc chương trình đào tạo của sinh viên đó và điểm của môn học đó, sinh viên có quyền lựa chọn xem các môn học theo thanh lọc học kỳ. Sinh viên có thể lựa chọn dự đoán cho các môn học chưa có điểm.

Tên sinh viên	Thanh lựa chọn học kỳ	Xem					
	<table border="1"> <tr> <td>Môn học 1</td> <td>Điểm môn học 1</td> </tr> <tr> <td>Môn học 2</td> <td>Dự đoán</td> </tr> <tr> <td>Môn học 3</td> <td>Dự đoán</td> </tr> </table>		Môn học 1	Điểm môn học 1	Môn học 2	Dự đoán	Môn học 3
Môn học 1	Điểm môn học 1						
Môn học 2	Dự đoán						
Môn học 3	Dự đoán						
Menu trái	Chân trang						

Hình 3.18 Phác thảo giao diện dự đoán điểm số cho sinh viên

3.6. Dữ liệu thử nghiệm

Dữ liệu thử nghiệm được thu thập thực tế từ chương trình đào tạo của ngành Công nghệ Thông tin thuộc bộ môn Công nghệ Thông tin khoa Kỹ thuật Công nghệ Trường Đại học Trà Vinh.

Bảng 3.3. Dữ liệu bảng Khoa

Ma_khoa	Ten_khoa
KTCN	Kỹ thuật Công nghệ
NNTS	Nông nghiệp Thủy sản

Hệ thống lưu trữ hiện tại có 2 bộ môn thuộc khoa Kỹ thuật Công nghệ:

Bảng 3.4. Dữ liệu bảng Bộ môn

Ma_bo_mon	Ten_bo_mon	Ma_khoa
XD	Xây dựng	KTCN
BM_CNTT	Công nghệ Thông tin	KTCN

Hệ thống lưu trữ hiện tại có 2 ngành thuộc bộ môn Công nghệ Thông tin:

Bảng 3.5. Dữ liệu bảng Ngành

Ma_nganh	Ten_nganh	Ma_bo_mon
N_TTNT	Trí tuệ nhân tạo	BM_CNTT
N_CNTT	Công nghệ Thông tin	BM_CNTT

Hệ thống lưu trữ hiện tại có 2 chương trình đào tạo thuộc ngành Công nghệ Thông tin với các quyết định chính thức của trường Đại học Trà Vinh:

Bảng 3.6. Dữ liệu bảng chương trình đào tạo

Ma_CTDT	Ten_CTDT	Ngay_QD	So_QD	Loai_hinh_DT	Ma_nganh	Tong_so_TC
CNTT-	Chương trình đào	07/7/2019	373/QĐ-	Đào tạo hệ	N_TTNT	138

2019	tạo ngành công nghệ thông tin năm 2019		ĐHTV	chính quy		
CNTT- 2020	Chương trình đào tạo ngành công nghệ thông tin khóa 2020	07/9/2020	Số 3455/QĐ - ĐHTV	Đào tạo hệ chính quy	N_TTNT	150

Với mỗi chương trình đào tạo sẽ được chia ra các học kỳ, với mỗi chương trình đào tạo sẽ có số lượng học kỳ khác nhau. Với học chương trình đào tạo Công nghệ Thông tin năm 2019 sẽ có 7 học kỳ và chương trình đào tạo Công nghệ Thông tin năm 2020 sẽ có 8 học kỳ.

Bảng 3.7. Dữ liệu bảng Học kỳ chương trình đào tạo (chương trình đào tạo năm 2020)

Ma_HK_CTTD	Ten_HK_CTTD
HKI - CNTT	Học kỳ 1 - Công nghệ thông tin
HKII - CNTT	Học kỳ 2 - Công nghệ thông tin
HKIII - CNTT	Học kỳ 3 - Công nghệ thông tin
HKIV - CNTT	Học kỳ 4 - Công nghệ thông tin
HKV - CNTT	Học kỳ 5 - Công nghệ thông tin
HKVI - CNTT	Học kỳ 6 - Công nghệ thông tin
HKVII - CNTT	Học kỳ 7 - Công nghệ thông tin
HKVIII - CNTT	Học kỳ 8 - Công nghệ thông tin

Trong mỗi chương trình đào tạo sẽ có những nhóm môn khác nhau, một nhóm môn sẽ bao gồm nhiều môn khác nhau.

Bảng 3.8. Dữ liệu bảng nhóm môn

Ma_nhom_mon	Ten_nhom_mon
CN	Kiến thức chuyên ngành
CSN	Kiến thức cơ sở ngành

GDTC-QPAN	Giáo dục thể chất và giáo dục Quốc phòng - An ninh
KH-XH-NV	Kiến thức Khoa học Xã hội và Nhân văn
LLTT	Kiến thức lý luận chính trị
NG-NG	Kiến thức ngoại ngữ
NK	Ngoại khóa
T-KH-TN	Kiến thức Toán - Tin học - Khoa học tự nhiên
TN	Tốt nghiệp

Bên cạnh việc phân chia nhóm môn, các môn học còn được chia theo loại học phần.

Bảng 3.9. Dữ liệu bảng loại học phần

Ma_loai_hp	Ten_loai_hp
BB	Học phần bắt buộc
TC	Học phần tự chọn

Với mỗi chương trình đào tạo sẽ có những môn học ứng với số tín chỉ của chương trình đào tạo đó. Một số môn học sẽ có trong chuỗi tiên quyết của chương trình đào tạo. Bên cạnh đó một số môn sẽ không nằm trong bất cứ chuỗi tiên quyết nào.

Ví dụ: Môn vi tích phân A1 sẽ là môn tiên quyết của môn vi tích phân A2 thì sẽ được tính là nằm trong chuỗi tiên quyết. Môn Thực tập đồ án chuyên ngành sẽ không có tiên quyết cho môn nào và cũng không có môn nào tiên quyết thì sẽ không nằm trong chuỗi tiên quyết.

Bảng 3.10. Dữ liệu bảng môn học (một số môn học thuộc chương trình đào tạo năm 2020)

Ma_mon_hoc	Ten_mon_hoc	Tin_chi	Mon_Tiep_theo
110001	Đại số tuyến tính	2	Null
110042	Vi tích phân A1	3	Vi tích phân A2
110002	Vi tích phân A2	2	Null
110003	Toán rời rạc	2	Cơ sở dữ liệu
180001	Tư tưởng Hồ Chí Minh	2	Lịch sử Đảng Cộng sản Việt Nam
170011	Tiếng Việt thực hành	2	Kỹ thuật XD & ban hành văn bản
180050	Triết học Mác - Lênin	2	Kinh tế chính trị Mác - Lênin
180051	Kinh tế chính trị Mác - Lênin	2	Chủ nghĩa xã hội khoa học
220114	Thực tập tốt nghiệp (CNTT)	3	Null

Với mỗi môn học sẽ phải thuộc một loại học phần, một nhóm môn học, ở một học kỳ chương trình đào tạo và thuộc một chương trình đào tạo.

Bảng 3.11. Dữ liệu bảng thuộc chương trình đào tạo(một số môn học thuộc chương trình đào tạo năm 2020)

Ma_mon_hoc	Ma_HK_CTD	Ma_loai_hp	Ma_nhom_mon	Ma_CTD
110001	HKI - CNTT	BB	T-KH-TN	CNTT-2020
110042	HKI - CNTT	BB	T-KH-TN	CNTT-2020
180050	HKI - CNTT	BB	LLTT	CNTT-

				2020
190081	HKI - CNTT	BB	GDTC-QPAN	CNTT-2020
220092	HKI - CNTT	BB	T-KH-TN	CNTT-2020
220228	HKI - CNTT	BB	CSN	CNTT-2020
410291	HKI - CNTT	BB	NG-NG	CNTT-2020
450015	HKI - CNTT	BB	KH-XH-NV	CNTT-2020

Dữ liệu điểm được quản lý với bảng điểm lưu điểm trung bình của mỗi học kỳ cho mỗi sinh viên, mỗi bảng điểm sẽ có chi tiết bảng điểm lưu điểm của từng môn học.

Bảng 3.12. Dữ liệu bảng Cố vấn học tập (Dữ liệu một số Cố vấn học tập)

Ma_cv	Ho_ten_cv	Dia_chi_cv	Gioi_tinh_cv	Email_cv
GV01	Đoàn Phước Miền	Trà Vinh	Nam	DPM@gmail.com
GV02	Nguyễn Ngọc Đan Thanh	Trà Vinh	Nữ	NNDT@gmail.com
GV03	Nguyễn Mộng Hiền	Trà Vinh	Nam	NMH@gmail.com

Bảng 3.13. Dữ liệu bảng Lớp (dữ liệu một số lớp)

Ma_lop	Ten_lop	Id_cv	Ma_CTDT
DA19TTA	Công nghệ thông tin lớp A khóa 2019	GV01	CNTT_2019
DA19TTB	Công nghệ thông tin lớp B khóa 2019	GV03	CNTT_2019
DA20TTA	Công nghệ thông tin lớp A khóa 2020	GV02	CNTT_2020

Bảng 3.14. Dữ liệu bảng Sinh viên (Dữ liệu một số sinh viên)

MSSV	Ho_ten_sv	Dia_chi_sv	Gioi_tinh_sv	Ngay_sinh_sv	Ma_lop
110118135	Nguyễn Trung Tuấn	Trà Vinh	Nam	01/01/2002	DA20TTA
110120002	Huỳnh Trần Tuấn Anh	Trà Vinh	Demo	01/01/2002	DA20TTA
110120008	Nguyễn Thị Ngọc Chăm	Trà Vinh	Demo	01/01/2002	DA20TTA

Bảng 3.15. Dữ liệu bảng điểm(một số bảng điểm của một sinh viên)

Diem_trung_binh_hoc_ky	MSSV	Ma_hk_nien_khoa
3.94	110120002	HKI 2020-2021
3.66	110120002	HKII 2020-2021
3.64	110120002	HKI 2021-2022

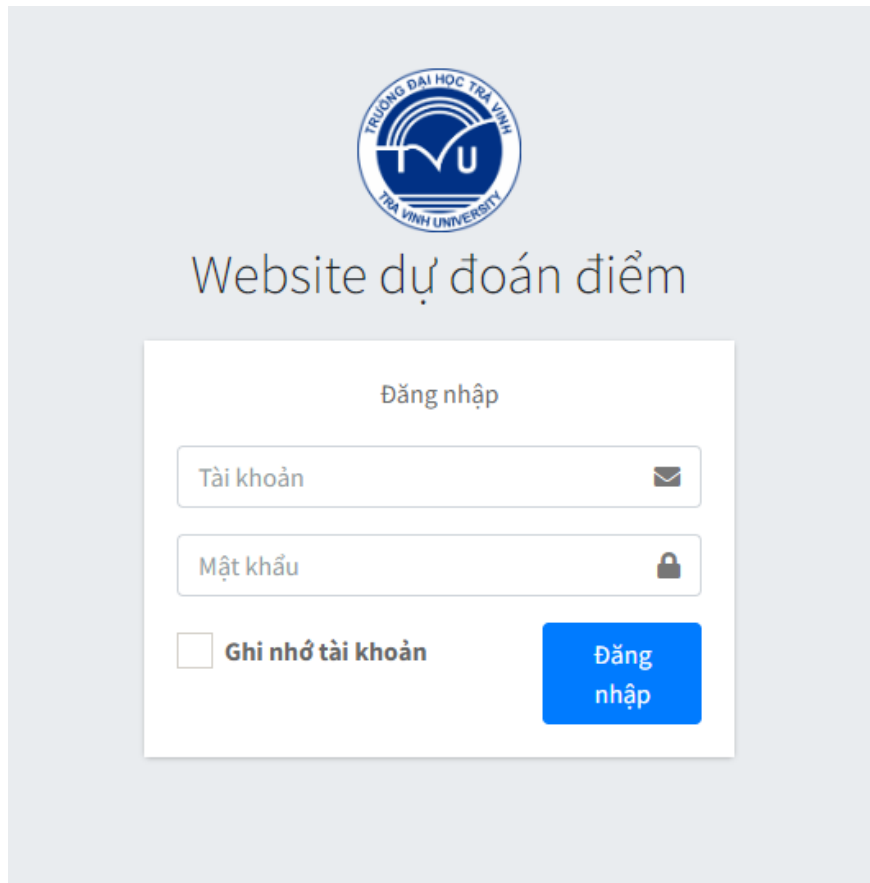
Bảng 3.16. Dữ liệu chi tiết bảng điểm(một số chi tiết bảng điểm của một bảng điểm)

Ma_mon_hoc	Diem_he_4	Diem_he_10	Diem_he_10_lan_2
110001	4	9.5	Null
190082	3.5	8.8	Null
190084	3	7.4	Null

CHƯƠNG 4. KẾT QUẢ NGHIÊN CỨU

4.1. Giao diện của Admin

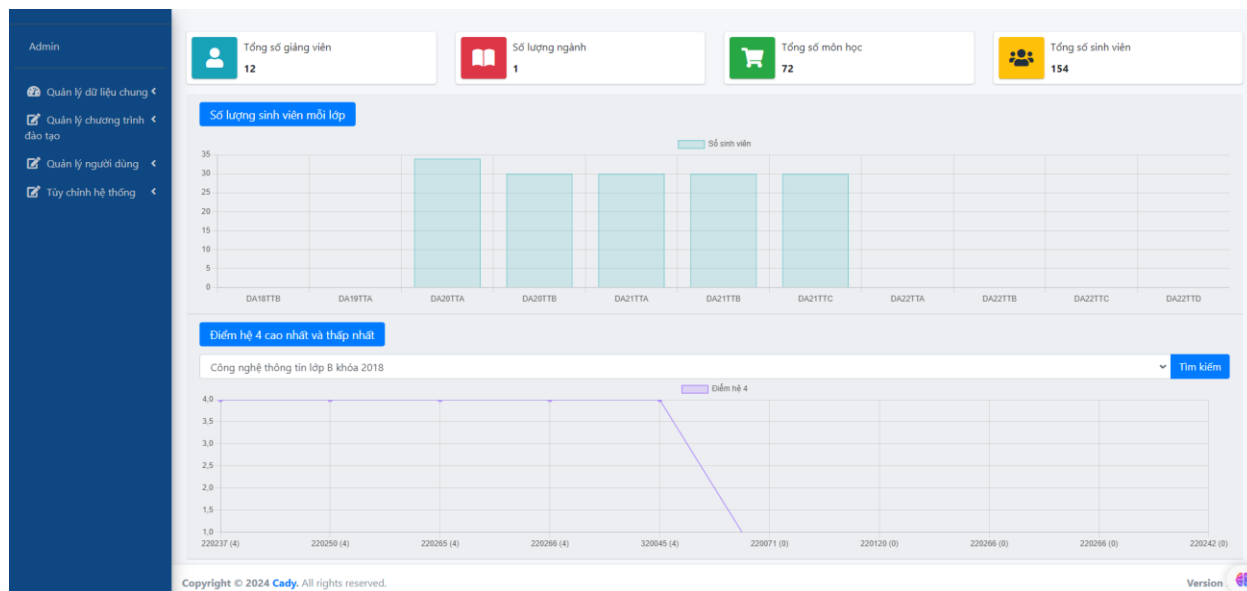
Người dùng đăng nhập bằng tài khoản và mật khẩu được cấp. Đối với tài khoản của admin mặc định là admin, còn đối với cố vấn học tập tài khoản mặc định sẽ là email của cố vấn học tập, đối với sinh viên tài khoản mặc định sẽ là mã số sinh viên, mật khẩu chung mặc định của tất cả tài khoản là 123. Tùy vào phân quyền của tài khoản mà người dùng sẽ được đưa đến các trang tương ứng.



The image shows a login page for Tra Vinh University. At the top center is the university's logo, a circular emblem with 'TRƯỜNG ĐẠI HỌC TRÀ VINH' and 'TRA VINH UNIVERSITY' around a central 'TVU' monogram. Below the logo, the text 'Website dự đoán điểm' is displayed. The main content is a white login box with a light gray border. Inside the box, the title 'Đăng nhập' is centered. There are two input fields: 'Tài khoản' (Account) with an envelope icon and 'Mật khẩu' (Password) with a lock icon. Below these is a checkbox labeled 'Ghi nhớ tài khoản' (Remember account). To the right of the checkbox is a blue button with white text that says 'Đăng nhập' (Login).

Hình 4.1 Giao diện đăng nhập

Giao diện trang chủ: Trang chủ hệ thống hiển thị thống kê sơ bộ của các dữ liệu quản lý, hiển thị thanh điều hướng tới những chức năng quản lý của hệ thống. Sơ đồ thứ nhất hiển thị tổng số sinh viên của mỗi lớp, sơ đồ thứ hai hiển thị điểm hệ 4 cao nhất và thấp nhất của lớp được chọn. Có thể xem thông tin chi tiết bằng cách nhấp chuột.



Hình 4.2 Giao diện trang chủ Admin

Giao diện chức năng quản lý học kỳ niên khóa: Chức năng thêm, sửa, xóa học kỳ niên khóa. Tùy vào năm học mà mã học kỳ niên khóa sẽ được quy định ứng với học kỳ và năm học. Học kỳ niên khóa được thêm mới vào sau mỗi học kỳ.

Danh sách học kỳ			
Thêm học kỳ mới			
<input type="checkbox"/>	Mã học kỳ	Tên học kỳ	Hành động
<input type="checkbox"/>	HKI 2019-2020	Học kỳ I năm 2019-2020	Sửa
<input type="checkbox"/>	HKI 2020-2021	Học kỳ I năm 2020-2021	Sửa
<input type="checkbox"/>	HKI 2021-2022	Học kỳ I năm 2021-2022	Sửa
<input type="checkbox"/>	HKI 2022-2023	Học kỳ I năm 2022-2023	Sửa
<input type="checkbox"/>	HKI 2023-2024	Học kỳ I năm 2023-2024	Sửa
<input type="checkbox"/>	HKII 2019-2020	Học kỳ II năm 2019-2020	Sửa
<input type="checkbox"/>	HKII 2020-2021	Học kỳ II năm 2020-2021	Sửa
<input type="checkbox"/>	HKII 2021-2022	Học kỳ II năm 2021-2022	Sửa
<input type="checkbox"/>	HKII 2022-2023	Học kỳ II năm 2022-2023	Sửa

Hình 4.3 Giao diện chức năng quản lý học kỳ niên khóa

Chức năng quản lý bộ môn: Nhấn vào nút thêm bộ môn và tiến hành thêm bộ môn mới, mỗi bộ môn sẽ phải thuộc một Khoa riêng biệt, CART có sai sót có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của bộ môn. Chọn vào ô vuông để đánh dấu một hoặc nhiều bộ môn muốn xóa và nhấn nút xóa.

Danh sách bộ môn

Thêm bộ môn

<input type="checkbox"/>	Mã bộ môn	Tên bộ môn	Mã khoa	Hành động
<input type="checkbox"/>	CNTT	Bộ môn công nghệ thông tin	Khoa kỹ thuật và công nghệ	Sửa

Xóa

Hình 4.4 Giao diện chức năng quản lý bộ môn

Chức năng quản lý lớp học: Nhấn vào nút thêm lớp và tiến hành thêm lớp mới, mỗi lớp sẽ phải học theo một chương trình đào tạo và có cố vấn riêng biệt, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của lớp. Chọn vào ô vuông để đánh dấu một hoặc nhiều lớp muốn xóa và nhấn nút xóa.

Danh sách lớp

Thêm lớp mới

<input type="checkbox"/>	Mã lớp	Tên lớp	Mã cố vấn	Mã chương trình đào tạo	Tên chương trình đào tạo
<input type="checkbox"/>	DA18TTA	Công nghệ thông tin lớp A khóa 2018	Nguyễn Hoàng Duy Thiện	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa
<input type="checkbox"/>	DA18TTB	Công nghệ thông tin lớp B khóa 2018	Nguyễn Ngọc Đan Thanh	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa

Hình 4.5 Chức năng quản lý lớp học

Chức năng quản lý khoa: Nhấn vào nút thêm lớp và tiến hành thêm khoa mới, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của khoa. Chọn vào ô vuông để đánh dấu một hoặc nhiều khoa muốn xóa và nhấn nút xóa.

Danh sách khoa

Thêm Khoa

<input type="checkbox"/>	Mã khoa	Tên khoa	Hành động
<input type="checkbox"/>	KTCN	Khoa kỹ thuật và công nghệ	Sửa

Xóa

Hình 4.6 Chức năng quản lý khoa

Chức năng quản lý ngành: Nhấn vào nút thêm lớp và tiến hành thêm ngành mới, mỗi ngành sẽ phải thuộc một bộ môn , nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của ngành. Chọn vào ô vuông để đánh dấu một hoặc nhiều ngành muốn xóa và nhấn nút xóa.

Danh sách ngành

Thêm ngành mới

<input type="checkbox"/>	Mã ngành	Tên ngành	Mã bộ môn	Hành động
<input type="checkbox"/>	CNTT	Ngành công nghệ thông tin	Bộ môn công nghệ thông tin	Sửa

Xóa

Hình 4.7 Giao diện chức năng quản lý ngành

Giao diện chức năng quản lý chương trình đào tạo: Nhấn vào nút thêm chương trình đào tạo và tiến hành thêm chương trình đào tạo mới, mỗi chương trình đào tạo sẽ phải thuộc một ngành, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của chương trình đào tạo. Chọn vào ô vuông để đánh dấu một hoặc nhiều chương trình đào tạo muốn xóa và nhấn nút xóa.

Danh sách chương trình đào tạo

Thêm chương trình đào tạo

<input type="checkbox"/>	Mã chương trình đào tạo	Tên chương trình đào tạo	Ngày quyết định	Số quyết định	Loại hình đào tạo	Mã ngành	Tổng số tín chỉ	Hành động
<input type="checkbox"/>	CNTT-2019	Chương trình đào tạo ngành công nghệ thông tin năm 2019	June 14, 2019	373/QĐ-ĐHTV	Đào tạo hệ chính quy	Ngành công nghệ thông tin	138	Sửa
<input type="checkbox"/>	CNTT-2020	Chương trình đào tạo ngành công nghệ thông tin khóa 2020	June 1, 2024	Số 3455/ QĐ - ĐHTV	Đào tạo hệ chính quy	Ngành công nghệ thông tin	150	Sửa

Xóa

Hình 4.8 Giao diện chức năng chương trình đào tạo

Giao diện chức năng quản lý học kỳ chương trình đào tạo: Nhấn vào nút thêm học kỳ chương trình đào tạo và tiến hành thêm học kỳ chương trình đào tạo mới, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của học kỳ chương trình đào tạo. Chọn vào ô vuông để đánh dấu một hoặc nhiều học kỳ chương trình đào tạo muốn xóa và nhấn nút xóa.

Danh sách học kỳ chương trình đào tạo

Thêm học kỳ chương trình đào tạo mới

<input type="checkbox"/>	Mã học kỳ chương trình đào tạo	Tên học kỳ chương trình đào tạo	Hành động
<input type="checkbox"/>	HKI - CNTT	Học kỳ 1 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKII - CNTT	Học kỳ 2 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKIII - CNTT	Học kỳ 3 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKIV - CNTT	Học kỳ 4 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKV - CNTT	Học kỳ 5 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKVI - CNTT	Học kỳ 6 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKVII - CNTT	Học kỳ 7 - Công nghệ thông tin	Sửa
<input type="checkbox"/>	HKVIII - CNTT	Học kỳ 8 - Công nghệ thông tin	Sửa

Xóa

Hình 4.9 Chức năng quản lý chương trình đào tạo

Giao diện chức năng quản lý môn học: Nhấn vào nút thêm môn học và tiến hành thêm môn học mới, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của môn học. Chọn vào ô vuông để đánh dấu một hoặc nhiều môn học muốn xóa và nhấn nút xóa.

Danh sách môn học

[Thêm môn học](#)
[Thêm môn học bằng file](#)
[Tải file mẫu Excel](#)

<input type="checkbox"/>	Mã môn học	Tên môn học	Tín chỉ	Môn tiên quyết	Hành động
<input type="checkbox"/>	110000	Vi tích phân A1	3	Vi tích phân A2	Sửa
<input type="checkbox"/>	110001	Đại số tuyến tính	2	None	Sửa
<input type="checkbox"/>	110002	Vi tích phân A2	2	Toán rời rạc	Sửa
<input type="checkbox"/>	110003	Toán rời rạc	2	Lý thuyết đồ thị	Sửa

Hình 4.10 Chức năng quản lý môn học

Giao diện chức năng quản lý nhóm môn: Nhấn vào nút thêm môn học và tiến hành thêm nhóm môn học mới, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của nhóm môn học. Chọn vào ô vuông để đánh dấu một hoặc nhiều nhóm môn học muốn xóa và nhấn nút xóa.

Danh sách nhóm môn

[Thêm nhóm môn mới](#)

<input type="checkbox"/>	Mã nhóm môn	Tên nhóm môn	Hành động
<input type="checkbox"/>	CN	Kiến thức chuyên ngành	Sửa
<input type="checkbox"/>	CSN	Kiến thức cơ sở ngành	Sửa
<input type="checkbox"/>	GDTC-QPAN	Giáo dục thể chất và giáo dục Quốc phòng - An ninh	Sửa
<input type="checkbox"/>	KH-XH-NV	Kiến thức Khoa học Xã hội và Nhân văn	Sửa
<input type="checkbox"/>	LLTT	Kiến thức lý luận chính trị	Sửa

Hình 4.11 Chức năng quản lý nhóm môn học

Giao diện chức năng quản lý loại học phần: Nhấn vào nút thêm loại học phần và tiến hành thêm loại học phần mới, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của loại học phần. Chọn vào ô vuông để đánh dấu một hoặc nhiều loại học phần muốn xóa và nhấn nút xóa.

Danh sách loại học phần

Thêm loại học phần mới

<input type="checkbox"/>	Mã loại học phần	Tên loại học phần	Hành động
<input type="checkbox"/>	BB	Học phần bắt buộc	Sửa
<input type="checkbox"/>	TC	Học phần tự chọn	Sửa

Xóa

Hình 4.12 Chức năng quản lý loại học phần

Giao diện chức năng quản lý thuộc chương trình đào tạo: Nhấn vào nút thêm và tiến hành thêm thuộc chương trình đào tạo, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của thuộc chương trình đào tạo. Chọn vào ô vuông để đánh dấu một hoặc nhiều thuộc chương trình đào tạo muốn xóa và nhấn nút xóa.

Thuộc chương trình đào tạo

Thêm chi tiết

Thêm bảng file

Tải file mẫu Excel

<input type="checkbox"/>	Mã thuộc chương trình đào tạo	Mã học kỳ chương trình đào tạo	Mã môn học	Mã loại học phần	Mã nhóm môn	Mã chương trình đào tạo	
<input type="checkbox"/>	TTCDT01	Học kỳ 1 - Công nghệ thông tin	Vi tích phân A1	Học phần bắt buộc	Kiến thức Toán - Tin học - Khoa học tự nhiên	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa
<input type="checkbox"/>	TTCDT02	Học kỳ 1 - Công nghệ thông tin	Đại số tuyến tính	Học phần bắt buộc	Kiến thức Toán - Tin học - Khoa học tự nhiên	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa
<input type="checkbox"/>	TTCDT03	Học kỳ 1 - Công nghệ thông tin	Triết học Mác - Lênin	Học phần bắt buộc	Kiến thức lý luận chính trị	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa
<input type="checkbox"/>	TTCDT04	Học kỳ 1 - Công nghệ thông tin	HP3 (3TC): QS chung và chiến thuật	Học phần bắt buộc	Giáo dục thể chất và giáo dục Quốc phòng - An ninh	Chương trình đào tạo ngành công nghệ thông tin năm 2019	Sửa

Hình 4.13 Chức năng quản lý thuộc chương trình đào tạo

Giao diện chức năng quản lý sinh viên: Nhấn vào nút thêm sinh viên và tiến hành thêm sinh viên, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của sinh viên. Chọn vào ô vuông để đánh dấu một hoặc nhiều sinh viên muốn xóa và nhấn nút xóa.

Danh sách sinh viên

Thêm sinh viên

Thêm sinh viên bằng file

Tải file mẫu Excel

<input type="checkbox"/>	Mã số sinh viên	Tên sinh viên	Địa chỉ	Giới tính	Ngày sinh	Email	Mã lớp	
<input type="checkbox"/>	110118001	Lê Minh Dũng	Trà Vinh	Nam	June 23, 2024	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018	Sửa
<input type="checkbox"/>	110118002	Lê Đặng Hoàng Quy	Trà Vinh	Nam	June 6, 2024	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018	Sửa
<input type="checkbox"/>	110118006	Nguyễn Kiều Anh	Trà Vinh	Nữ	Jan. 1, 2000	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018	Sửa
<input type="checkbox"/>	110118007	Nguyễn Chí Bảo	Trà Vinh	Nam	Jan. 1, 2000	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018	Sửa

Hình 4.14 Chức năng quản lý sinh viên

Giao diện chức năng quản lý cố vấn học tập: Nhấn vào nút thêm cố vấn học tập và tiến hành thêm cố vấn học tập, có thể lựa chọn thêm bằng excel nếu như cần thêm một lượng lớn dữ liệu mới, khi nhấn vào nút tải mẫu excel sẽ tiến hành tải file excel mẫu và tiến hành điền các thông tin cần thiết vào mẫu sau đó tiến hành thêm cố vấn học tập bằng excel vừa điền. Nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của cố vấn học tập. Chọn vào ô vuông để đánh dấu một hoặc nhiều cố vấn học tập muốn xóa và nhấn nút xóa.

Thêm giảng viên

Thêm giảng viên bằng file

Tải file mẫu Excel

<input type="checkbox"/>	Mã số giảng viên	Tên giảng viên	Địa chỉ	Giới tính	Ngày sinh	Email	
<input type="checkbox"/>	GV01	Nguyễn Hoàng Duy Thiện	Trà Vinh	Nam	June 14, 2024	NHDT@gmail.com	Sửa
<input type="checkbox"/>	GV02	Nguyễn Ngọc Đan Thanh	Trà Vinh	Nữ	June 13, 2024	NNDT@gmail.com	Sửa

Xóa

Hình 4.15 Chức năng quản lý cố vấn học tập

Giao diện chức năng quản lý tài khoản: Nhấn vào nút thêm tài khoản và tiến hành thêm tài khoản, nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của tài khoản. Chọn vào ô vuông để đánh dấu một hoặc nhiều tài khoản muốn xóa và nhấn nút xóa.

Danh sách tài khoản

Thêm tài khoản

<input type="checkbox"/>	Tên tài khoản	Mật khẩu	Phân quyền	Hành động
<input type="checkbox"/>	110118001	123123	student	Sửa

Hình 4.16 Chức năng quản lý tài khoản

Giao diện chức năng quản lý điểm số: Nhấn vào nút thêm cố vấn học tập và tiến hành thêm điểm số, có thể lựa chọn thêm bằng excel nếu như cần thêm một lượng lớn dữ liệu mới, khi nhấn vào nút tải mẫu excel sẽ tiến hành tải file excel mẫu và tiến hành điền các thông tin cần thiết vào mẫu sau đó tiến hành thêm điểm số bằng excel vừa điền. Nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của điểm số. Chọn vào ô vuông để đánh dấu một hoặc nhiều điểm số muốn xóa và nhấn nút xóa.

Bảng điểm

Thêm Bảng điểm

Thêm Bảng điểm bằng file

Tải file mẫu Excel

<input type="checkbox"/>	Điểm hệ 4	Mã số sinh viên	Học kỳ niên khóa	
<input type="checkbox"/>	2.24	110119004	Hockynienkhoa object (HKI 2019-2020)	Sửa
<input type="checkbox"/>	2.24	110119004	Hockynienkhoa object (HKII 2019-2020)	Sửa
<input type="checkbox"/>	2.26	110119004	Hockynienkhoa object (HKI 2020-2021)	Sửa
<input type="checkbox"/>	2.34	110119004	Hockynienkhoa object (HKII 2020-2021)	Sửa

Hình 4.17 Chức năng quản lý điểm số

Giao diện chức năng quản lý chi tiết bảng điểm: Nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của chi tiết bảng điểm . Chọn vào ô vuông để đánh dấu một hoặc nhiều chi tiết bảng điểm muốn xóa và nhấn nút xóa.

Chi tiết bảng điểm					
<input type="checkbox"/>	Điểm hệ 4	Điểm hệ 10	Điểm hệ 10 lần 2	Mã môn học	
<input type="checkbox"/>	2.0	6.1	Null	Vĩ tích phần A1	Sửa
<input type="checkbox"/>	3.5	8.3	Null	Đại số tuyến tính	Sửa
<input type="checkbox"/>	1.5	5.2	Null	Triết học Mác - Lênin	Sửa
<input type="checkbox"/>	2.0	5.8	Null	HP3 (3TC): QS chung và chiến thuật	Sửa
<input type="checkbox"/>	2.5	6.9	Null	Học phần 1 (2TC): Đường lối quốc phòng và an ninh của Đảng Cộng sản Việt Nam	Sửa

Hình 4.18 Chức năng quản lý chi tiết bảng điểm

4.2. Giao diện của cố vấn học tập

Giao diện chức năng quản lý sinh viên: Nhấn vào nút thêm sinh viên và tiến hành thêm điểm số, có thể lựa chọn thêm bằng excel nếu như cần thêm một lượng lớn dữ liệu mới, khi nhấn vào nút tải mẫu excel sẽ tiến hành tải file excel mẫu và tiến hành điền các thông tin cần thiết vào mẫu sau đó tiến hành thêm sinh viên bằng excel vừa điền. Nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của sinh viên. Chọn vào ô vuông để đánh dấu một hoặc nhiều sinh viên muốn xóa và nhấn nút xóa.

Danh sách sinh viên							
<div> Thêm sinh viên Thêm sinh viên bằng file Tải file mẫu Excel </div>							
<input type="checkbox"/>	Mã số sinh viên	Tên sinh viên	Địa chỉ	Giới tính	Ngày sinh	Email	Mã lớp
<input type="checkbox"/>	110118001	Lê Minh Dũng	Trà Vinh	Nam	June 23, 2024	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018
<input type="checkbox"/>	110118002	Lê Đặng Hoàng Quy	Trà Vinh	Nam	June 6, 2024	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018
<input type="checkbox"/>	110118006	Nguyễn Kiều Anh	Trà Vinh	Nữ	Jan. 1, 2000	demo@gmail.com	Công nghệ thông tin lớp A khóa 2018
<input type="checkbox"/>	110118007	Nguyễn Chí Bảo	Trà Vinh	Nam	Jan. 1, 2000	demo@gmail.com	Công nghệ thông tin lớp A

Hình 4.19 Chức năng quản lý thông tin sinh viên của cố vấn học tập

Giao diện chức năng quản lý bảng điểm: Nhấn vào nút thêm cố vấn học tập và tiến hành thêm điểm số, có thể lựa chọn thêm bằng excel nếu như cần thêm một lượng lớn dữ liệu mới, khi nhấn vào nút tải mẫu excel sẽ tiến hành tải file excel mẫu và tiến hành điền các thông tin cần thiết vào mẫu sau đó tiến hành thêm điểm số bằng excel vừa điền. Nhấn nút sửa nếu trong quá trình thêm mới có sai sót hoặc có nhu cầu chỉnh sửa các thông tin của điểm số. Chọn vào ô vuông để đánh dấu một hoặc nhiều điểm số muốn xóa và nhấn nút xóa.

Bảng điểm					
<div> Thêm Bảng điểm Thêm Bảng điểm bằng file Tải file mẫu Excel </div>					
<input type="checkbox"/>	Mã bảng điểm	Điểm hệ 4	Mã số sinh viên	Học kỳ niên khóa	
<input type="checkbox"/>	162	2.24	110119004	Hockynienkhoa object (HKI 2019-2020)	Sửa
<input type="checkbox"/>	163	2.24	110119004	Hockynienkhoa object (HKII 2019-2020)	Sửa
<input type="checkbox"/>	164	2.26	110119004	Hockynienkhoa object (HKI 2020-2021)	Sửa

Hình 4.20 Chức năng quản lý điểm số của cố vấn học tập

Giao diện chức năng quản lý chi tiết bảng điểm: Chức năng cho phép sửa, xóa chi tiết bảng điểm của sinh viên thuộc lớp cố vấn học tập quản lý

Chi tiết bảng điểm					
<input type="checkbox"/>	Điểm hệ 4	Điểm hệ 10	Điểm hệ 10 lần 2	Mã môn học	
<input type="checkbox"/>	2.0	6.1	Null	Vi tích phân A1	Sửa
<input type="checkbox"/>	3.5	8.3	Null	Đại số tuyến tính	Sửa
<input type="checkbox"/>	1.5	5.2	Null	Triết học Mác - Lênin	Sửa
<input type="checkbox"/>	2.0	5.8	Null	HP3 (3TC): QS chung và chiến thuật	Sửa

Hình 4.21 Chức năng quản lý chi tiết bảng điểm của cố vấn học tập

4.3. Giao diện của sinh viên

Giao diện chức năng dự đoán điểm số: Hệ thống hiển thị toàn bộ môn học, có thể lọc theo học kỳ chương trình đào tạo để xem môn học đó ở học kỳ nào. Nếu môn học có điểm sẽ hiển thị điểm, nếu môn học chưa có điểm cho phép lựa chọn dự đoán môn học đó. Sau khi nhấn nút dự đoán sinh viên nhận kết quả điểm dự đoán cùng với các thông số đánh giá độ tin cậy

Chọn học kỳ:

Tất cả

Xem

STT	Mã môn học	Tên môn học	Điểm hệ 10
1	110001	Đại số tuyến tính	6.5
2	110002	Ví tích phân A2	<div>Dự đoán</div>
3	110003	Toán rời rạc	<div>Dự đoán</div>
4	110042	Ví tích phân A1	4.9
5	110057	Quy hoạch tuyến tính	<div>Dự đoán</div>
6	110079	Kiến trúc máy tính	<div>Dự đoán</div>
7	120004	Vật lý đại cương	<div>Dự đoán</div>
8	170011	Tiếng Việt thực hành	<div>Dự đoán</div>
9	180001	Tư tưởng Hồ Chí Minh	<div>Dự đoán</div>
10	180050	Triết học Mác - Lênin	6.5

Hình 4.22 Chức năng dự đoán điểm số sinh viên

Dưới đây là một ví dụ khi nhấn nút dự đoán cho môn Chuyên đề linux và sinh viên sẽ nhận được kết quả hiển thị như sau:

Kết quả dự đoán cho sinh viên: Mạch Khánh Duy

Môn học: Chuyên đề Linux (220034)

Kết quả dự đoán

Phương pháp	Dự đoán	R^2	MSE
Linear Regression	7.3922265773302325	0.6938953672351695	0.06835032631135146
Random Forest	7.734800000000003	0.7192094121605805	0.06269793478991598

Hình 4.23 Kết quả dự đoán của sinh viên

Kết quả dự đoán cho sinh viên được đánh giá dự trên 2 thông số là R^2 và MSE. Hệ số R^2 giao động từ 0 đến 1 (tương đương 0% đến 100%), hệ số càng lớn thì hiệu quả dự đoán càng cao. Hệ số MSE là hệ số cho biết mức độ chênh lệch giữa điểm dự đoán và điểm thực tế, hệ số này chạy càng thấp thì hiệu quả dự đoán càng chính xác.

CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1. Kết luận

Trong quá trình nghiên cứu và xây dựng hệ thống dự đoán điểm tích lũy cho sinh viên Trường Đại học Trà Vinh đã đạt được những mục tiêu và có được những kết quả quan trọng như sau:

- Tích lũy kinh nghiệm: Trong quá trình nghiên cứu tôi đã học được thêm nhiều kiến thức mới, học hỏi được thêm cách sử dụng ngôn ngữ Python cùng với Django, bên cạnh đó còn học hỏi thêm được cách phát triển hệ thống giao diện thân thiện với người dùng.

- Kết quả đạt được: Hoàn thành được các mục tiêu ban đầu đưa ra, xây dựng thành công hệ thống dự đoán điểm cho sinh viên cùng với số phần trăm đáng tin cậy đạt tiêu chuẩn.

Hệ thống dự đoán điểm đối với một số môn chuyên ngành như: Chuyên đề Linux, Lập trình thiết bị di động đã đạt đến độ chính xác trung bình trên 70%, một số môn học khác không nằm trong chuyên ngành như: Quy hoạch tuyến tính, Đại số đại cương cũng đạt đến con số trung bình 60%. Độ chính xác này được xác định dựa vào các chỉ số đánh giá chung cho các mô hình học máy như: R^2 , MSE,...

Đối với khoa học, đóng góp của đề tài này là rất rõ ràng. Bằng cách sử dụng các mô hình học máy và việc dự đoán điểm số không chỉ giúp mở rộng kiến thức mà còn cung cấp một công cụ hữu ích trong việc quản lý và theo dõi điểm số học tập. Đề tài này cũng giúp ích trong việc khẳng định cho sự phát triển của trí tuệ nhân tạo và ứng dụng trí tuệ nhân tạo vào môi trường giáo dục.

Đối với nhà trường, việc cung cấp hệ thống dự đoán điểm góp phần nâng cao chất lượng giáo dục thông qua việc dự đoán chính xác các thông tin về điểm số. Điều này giúp cố vấn học tập sớm nhận biết cần hỗ trợ thêm ở mảng nào và cải thiện chất lượng điểm số cho mỗi sinh viên.

Đối với sinh viên, hệ thống dự đoán điểm này mang lại nhiều lợi ích thiết thực. Sinh viên có thể sử dụng kết quả dự đoán để lập kế hoạch học tập, nhận biết những môn học mà mình cần tập trung hơn và cải thiện điểm số. Hệ thống này cũng giúp sinh viên có

cái nhìn tổng quan về khả năng học tập của mình, từ đó cải thiện phương pháp học tập hiệu quả và đạt được kết quả tốt hơn trong các kỳ thi.

Tóm lại, hệ thống dự đoán điểm không chỉ góp phần mang lại giá trị cho khoa học trí tuệ nhân tạo mà còn đóng góp tích cực cho nhà trường nâng cao chất lượng giáo dục và mang lại cho sinh viên một công cụ hỗ trợ trong việc phát hiện môn học bị điểm kém từ đó có thể tập trung cải thiện tránh được các trường hợp không mong muốn.

Ưu điểm:

- Hệ thống có hệ thống phản hồi nhanh, tính toán dữ liệu với độ tin cậy cao.
- Giao diện hệ thống hiện tại, dễ dàng thao tác và tiếp cận.

Nhược điểm:

- Chức năng quản lý hệ thống còn nhiều bước và nhiều thao tác riêng lẻ
- Chưa có nhiều biểu đồ thống kê chi tiết cho khâu quản lý

5.2. Hướng phát triển

- Nghiên cứu và tiếp thu thêm các thuật toán dự đoán khác với mô hình phức tạp hơn và đưa ra được kết quả chuẩn xác, đáng tin cậy hơn.

- Áp dụng học sâu vào trong hệ thống dự đoán, sử dụng tăng cường dữ liệu để tạo ra dữ liệu huấn luyện phong phú hơn

- Phát triển hệ thống gợi ý bài học và tài liệu cần học để cải thiện các môn học phát hiện điểm kém

- Tạo ra các báo cáo chi tiết cho giáo viên và phụ huynh về kết quả học tập và dự đoán của học sinh.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] getbootstrap, "what is bootstrap 4," [Online]. Available: <https://getbootstrap.com/docs/4.0/getting-started/introduction/>. [Accessed 29 06 2024].
- [2] Amazon, "Python là gì?," [Online]. Available: <https://aws.amazon.com/vi/what-is/python/>. [Accessed 29 06 2024].
- [3] I. NumFOCUS, "pandas," [Online]. Available: <https://pandas.pydata.org/>. [Accessed 10 7 2024].
- [4] T. M. d. team, "Matplotlib: Visualization with Python," [Online]. Available: <https://matplotlib.org/>. [Accessed 10 7 2024].
- [5] Amazon, "SQL (Ngôn ngữ truy vấn có cấu trúc) là gì?," [Online]. Available: <https://aws.amazon.com/vi/what-is/sql/>. [Accessed 29 6 2024].
- [6] Amazon, "Hồi quy tuyến tính là gì?," [Online]. Available: <https://aws.amazon.com/vi/what-is/linear-regression/>. [Accessed 30 6 2024].
- [7] Viblo, "Sử dụng Amazon SageMaker để triển khai, huấn luyện mô hình máy học Random Forest (Scikit-learn)," [Online]. Available: <https://viblo.asia/p/su-dung-amazon-sagemaker-de-trien-khai-huan-luyen-mo-hinh-may-hoc-random-forest-scikit-learn-3RIL5RKzVbB>. [Accessed 26 06 2024].
- [8] ibm, "What is random forest?," [Online]. Available: <https://www.ibm.com/topics/random-forest>. [Accessed 2024 06 29].
- [9] scikit-learn, "What is scikit-learn ?," [Online]. Available: <https://scikit-learn.org/stable/>. [Accessed 28 06 2024].
- [10] D. W. T. H. R. T. J. T. Gareth James, in *An Introduction to Statistical Learning with Applications in Python*, 2023.
- [11] S. Hartshorn, in *Machine Learning with Random Forests and Decision Trees: A Visual Guide for Beginners*, 2020.
- [12] S. Masís, in *Interpretable Machine Learning with Python: Learn to build interpretable high-performance models with hands-on real-world examples*, 2021.