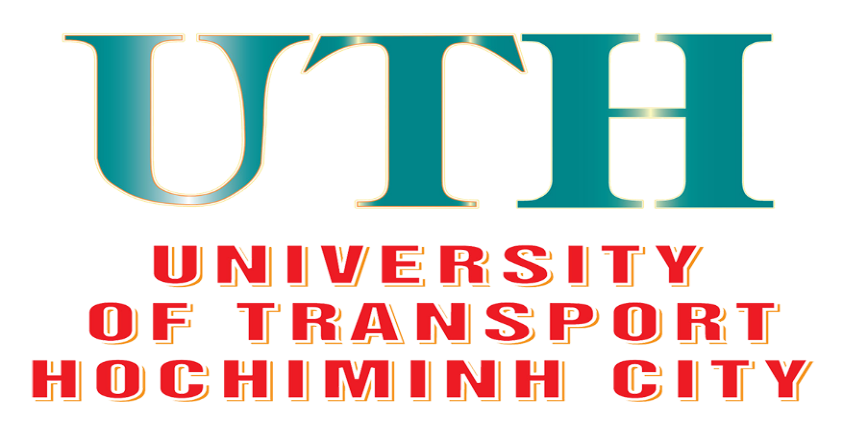
**BỘ GIAO THÔNG VẬN TẢI**



**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI TP. HỒ CHÍ MINH**

**Viện công nghệ thông tin và Điện, điện tử**

------o0o------



***Lập trình ngôn ngữ Python***

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Xử lý dữ liệu với ngôn ngữ Python**

**Đề tài: Xử lý dữ liệu điểm chuẩn đại học tại TP. HCM**

**Giảng viên : ThS. Huỳnh Thanh Việt**

**Mã học phần : 010112410001**

**Nhóm : 07**

**SL thành viên : 04**

**Github:** [**https://github.com/Le-AnV/group\_7\_final\_python**](https://github.com/Le-AnV/group_7_final_python%20)

**TP. HỒ CHÍ MINH, 2025**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tổng hợp đóng góp của các thành viên** | | | |
| STT | Tên thành viên | MSSV | Nội dung công việc |
| 1 | Lê Văn An | 056205001827 | * Crawl data từ các website * Clean data * Trực quan hoá dữ liệu * Machine learning * Đưa dữ liệu vào database * Báo cáo Machine learing |
| 2 | Phạm Gia Bảo | 093205005065 | * Clean data * Trực quan hoá dữ liệu * Báo cáo Clean data |
| 3 | Trần Văn Hưng | 042205003605 | * Clean data * Trực quan hoá dữ liệu * Báo cáo Crawl data |
| 4 | Trần Tô Khắc Huy | 082205009536 | * Trực quan hoá dữ liệu * Báo cáo trực quan hoá dữ liệu |

**Mục lục**

[**Lời mở đầu** 1](#_Toc197624824)

[**Lời cảm ơn** 2](#_Toc197624825)

[**Chương 1: Giới thiệu đề tài** 3](#_Toc197624826)

[1.1 Giới thiệu đề tài 3](#_Toc197624827)

[1.2. Mục tiêu của dự án 3](#_Toc197624828)

[**Chương 2: Crawl Data** 5](#_Toc197624829)

[2.1. Mục tiêu thu thập dữ liệu 5](#_Toc197624830)

[2.2. Phương pháp Crawl dữ liệu 5](#_Toc197624831)

[2.3. Cấu trúc dữ liệu sau khi crawl 5](#_Toc197624832)

[2.4. Nhận xét 6](#_Toc197624833)

[**Chương 3: Clean data** 7](#_Toc197624834)

[3.1. Dữ liệu thô: 7](#_Toc197624835)

[3.2. Dữ liệu đã được làm sạch: 7](#_Toc197624836)

[3.3 Các bước làm sạch dữ liệu: 7](#_Toc197624837)

[*3.3.1. Xử lý dữ liệu thiếu* 7](#_Toc197624838)

[*3.3.2. Chuẩn hóa và bổ sung dữ liệu* 7](#_Toc197624839)

[*3.3.3. Loại bỏ trùng lặp* 8](#_Toc197624840)

[*3.3.4. Kiểm tra định dạng dữ liệu* 8](#_Toc197624841)

[Kết luận: 8](#_Toc197624842)

[**Chương 4: Data visualization** 9](#_Toc197624843)

[4.1. Phân phối và xu hướng điểm chuẩn theo phương thức xét tuyển 9](#_Toc197624844)

[4.2. Sự thay đổi số lượng ngành và hệ đào tạo 9](#_Toc197624845)

[4.3. Phân tích tổ hợp môn và điểm chuẩn tại các trường ĐHQG-HCM 9](#_Toc197624846)

[**Chương 5: Machine Learning** 11](#_Toc197624847)

[5.1 Tiền xử lý và chuẩn hóa dữ liệu 11](#_Toc197624848)

[5.2 Mã hóa dữ liệu và chia tập huấn luyện 11](#_Toc197624849)

[5.3 Xây dựng và đánh giá mô hình 11](#_Toc197624850)

[**Nguồn dữ liệu** 13](#_Toc197624851)

## 

# Lời mở đầu

Trong thời đại dữ liệu số bùng nổ, việc khai thác và phân tích dữ liệu một cách hiệu quả đã trở thành kỹ năng thiết yếu trong nhiều lĩnh vực – từ kinh doanh, y tế, cho đến giáo dục. Với ưu điểm dễ học, cú pháp rõ ràng và hệ sinh thái thư viện mạnh mẽ, Python hiện là một trong những ngôn ngữ lập trình phổ biến nhất, đặc biệt trong lĩnh vực xử lý và phân tích dữ liệu.

Trong khuôn khổ môn học Lập trình với ngôn ngữ Python, nhóm chúng mình đã thực hiện đề tài: “Phân tích dữ liệu điểm chuẩn của các trường đại học tại TP.HCM”. Đây là một chủ đề thiết thực, gần gũi với học sinh – sinh viên, đặc biệt là những bạn đang chuẩn bị cho kỳ thi tuyển sinh đại học. Mục tiêu của dự án là thu thập dữ liệu điểm chuẩn các trường qua các năm, xử lý và phân tích để tìm ra xu hướng, mối liên hệ giữa điểm chuẩn và các yếu tố như ngành học, khối xét tuyển, trường đại học,...

Thông qua dự án này, nhóm không chỉ vận dụng các kiến thức lập trình căn bản như vòng lặp, xử lý chuỗi, và làm việc với hàm, mà còn tiếp cận và sử dụng các thư viện phân tích dữ liệu quan trọng như pandas, matplotlib, seaborn và scikit-learn. Ngoài ra, việc xử lý dữ liệu thực tế giúp nhóm rèn luyện kỹ năng làm sạch, trực quan hóa dữ liệu và bước đầu áp dụng một số kỹ thuật học máy đơn giản để đưa ra nhận định hoặc dự đoán xu hướng điểm chuẩn trong tương lai.

Đây là cơ hội quý báu để nhóm nâng cao tư duy phân tích, khả năng làm việc nhóm cũng như rèn luyện kỹ năng giải quyết các bài toán thực tế có ý nghĩa xã hội.

# Lời cảm ơn

Nhóm xin chân thành cảm ơn thầy ThS. Huỳnh Thanh Việt, giảng viên phụ trách môn *Lập trình với ngôn ngữ Python*, người đã tận tình giảng dạy và định hướng xuyên suốt quá trình học tập và thực hiện đề tài.

Những kiến thức và kinh nghiệm thực tiễn mà Thầy đã truyền đạt không chỉ giúp nhóm nắm vững nền tảng lập trình Python mà còn tạo động lực để nhóm mạnh dạn áp dụng vào việc xây dựng sản phẩm trò chơi hoàn chỉnh.

Mặc dù nhóm đã cố gắng hoàn thiện báo cáo và sản phẩm một cách tốt nhất, nhưng chắc chắn vẫn còn một số hạn chế. Nhóm rất mong nhận được những góp ý quý báu từ Thầy để có thể tiếp tục cải thiện kỹ năng, nâng cao chất lượng học tập và phát triển hơn trong các dự án tương lai.

Trân trọng cảm ơn!

# Chương 1: Giới thiệu đề tài

## **1.1 Giới thiệu đề tài**

Trong thời đại mà dữ liệu đóng vai trò then chốt trong mọi lĩnh vực, việc thu thập, xử lý và phân tích thông tin một cách hiệu quả đang ngày càng trở nên quan trọng. Đặc biệt trong lĩnh vực giáo dục, dữ liệu về điểm chuẩn đại học không chỉ phản ánh xu hướng tuyển sinh qua các năm mà còn giúp học sinh, phụ huynh và nhà trường đưa ra những quyết định chính xác hơn trong việc lựa chọn ngành học và định hướng nghề nghiệp.

Xuất phát từ nhu cầu thực tiễn đó, nhóm chúng mình đã thực hiện đề tài: “Phân tích dữ liệu điểm chuẩn các trường đại học tại TP.HCM giai đoạn 2021–2023”. Dự án tập trung vào việc thu thập và phân tích dữ liệu điểm chuẩn từ các trường đại học thuộc khu vực TP.HCM, đồng thời mở rộng đến các yếu tố liên quan như tổ hợp môn thi, mã ngành theo quy định của Bộ Giáo dục và Đào tạo.

Dữ liệu được crawl từ trang thông tin hướng nghiệp học sinh sinh viên. Trong quá trình thực hiện, nhóm đã sử dụng nhiều công cụ và thư viện phổ biến trong Python như:

* BeautifulSoup để thu thập dữ liệu web (Web Scraping)
* Pandas để xử lý và phân tích dữ liệu
* Matplotlib để trực quan hóa thông tin
* Scikit-learn để áp dụng các mô hình Machine Learning nhằm khám phá các yếu tố ảnh hưởng đến điểm chuẩn

Thông qua dự án, chúng mình không chỉ thực hành các kỹ năng thu thập và xử lý dữ liệu thực tế, mà còn nâng cao khả năng tư duy logic, kỹ năng lập trình cũng như hiểu biết sâu sắc hơn về tình hình tuyển sinh của các trường đại học tại TP.HCM trong những năm gần đây.

## **1.2. Mục tiêu của dự án**

Dự án được chia thành 4 giai đoạn chính:

1. Giai đoạn 1: Crawl data
2. Giai đoạn 2: Clean data
3. Giai đoạn 3: Visualization
4. Giai đoạn 4: Machine learning

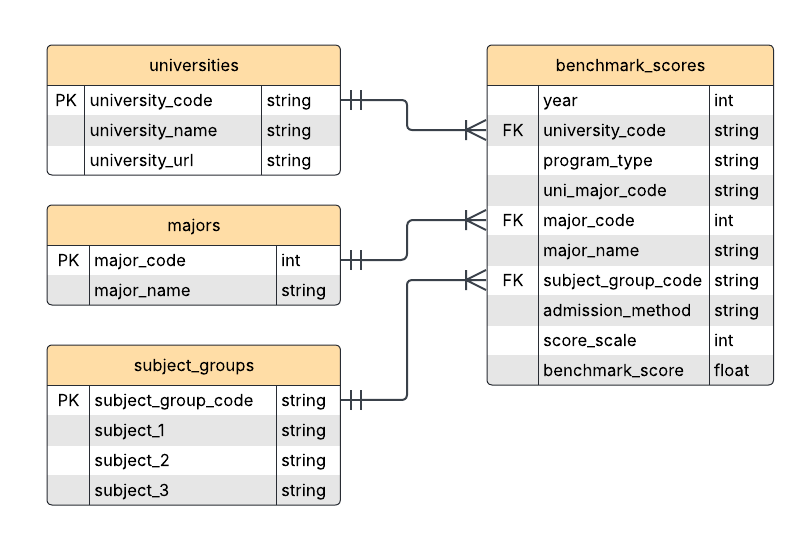
Mỗi giai đoạn sẽ có những mục tiêu khác nhau, nhưng khi kết hợp chúng lại ta sẽ được 1 dự án hoàng chỉnh.

Mục tiêu của giai đoạn 1: Tiếp thu thêm nhiều kiến thức và nghiên cứu để có thể tận dụng được các thư viên như BeautyfulSoup hay Selenium. Từ đó Crawl được dữ liệu 1 cách chính xác nhất. Các dữ liệu đã đào được như: Điểm chuẩn các trường đại học tại thành phố Hồ Chí Minh [1], danh sách tổ hợp môn thi [2], danh sách các nhóm ngành và ngành [3].

Mục tiêu của giai đoạn 2: Làm sạch dữ liệu là điều dĩ nhiên và ngoài ra ta cần phải xử lý các giá trị không đúng theo định dạng, thêm 1 số cột mới từ dữ liệu sẵn có, xoá các giá trị trùng lập,…

Mục tiêu của giai đoạn 3: Trực quan hoá dữ liệu là cách đơn giản nhất để ta có thể trình bày những sự thú vị nằm trong dữ liệu cho mọi người. Ở phần này, nhóm có đưa ra 1 số biểu đồ liên quan đến sự phân phối và cân bằng của dữ liệu, cũng như là các con số sẽ được trình bày bằng các biểu đồ 1 cách hợp lý, hay ho.

Mục tiêu của giai đoạn 4:Dù là phần nghiên cứu thêm trong dự án, nhưng tiếp xúc sớm với các kiến thức mới không bao giờ là thừa. Trong quá trình tìm tòi và nghiên cứu, nhóm đã tiếp thu thêm nhiều kiến thức cơ bản của máy học. Mục tiêu ở phần này đơn giản chỉ là pre-processing data sau đó lựa chọn mô hình máy học sao cho tối ưu nhất và đánh giá kết quả.



# Chương 2: Crawl Data

## **2.1. Mục tiêu thu thập dữ liệu**

Crawl dữ liệu là bước đầu tiên và quan trọng trong quá trình xây dựng hệ thống phân tích và dự đoán điểm chuẩn của các trường đại học. Mục tiêu của bước này là thu thập đầy đủ dữ liệu điểm chuẩn tuyển sinh đại học của các trường trên địa bàn TP.HCM trong ba năm gần nhất (2021, 2022, 2023). Ngoài ra, nhóm cũng thu thập thông tin về tổ hợp môn xét tuyển để phục vụ cho việc phân tích, trực quan hóa và huấn luyện mô hình học máy. Việc thu thập dữ liệu có hệ thống và chính xác sẽ giúp đảm bảo chất lượng cho toàn bộ quá trình xử lý và phân tích dữ liệu sau này.

## **2.2. Phương pháp Crawl dữ liệu**

Nhóm sử dụng ngôn ngữ lập trình Python để xây dựng các đoạn mã tự động thu thập dữ liệu. Các thư viện chính được sử dụng bao gồm:  
  
- requests: Thư viện gửi các HTTP request để tải nội dung HTML từ trang web.  
- BeautifulSoup: Công cụ phân tích cú pháp HTML và trích xuất dữ liệu cần thiết như tên trường, ngành học, tổ hợp môn, điểm chuẩn.  
- pandas: Dùng để tổ chức và lưu trữ dữ liệu ở định dạng bảng, xử lý các thao tác lọc, nhóm, chuyển đổi dữ liệu.  
  
Ngoài ra, trong một số trường hợp, nhóm sử dụng dữ liệu đã được tổng hợp sẵn ở định dạng bảng (CSV) từ các nguồn trực tuyến. Nhóm viết thêm các đoạn mã để đọc, kiểm tra và lưu trữ các bảng dữ liệu này một cách có hệ thống. Tất cả các dữ liệu crawl được lưu trữ trong thư mục 'raw/', được phân loại theo năm và nội dung để dễ dàng truy cập và xử lý sau này.

## **2.3. Cấu trúc dữ liệu sau khi crawl**

Dữ liệu sau khi được thu thập và lưu trữ được chia thành các tệp riêng biệt theo từng năm và loại thông tin:  
  
- raw\_benchmark\_scores\_hcm\_2021.csv  
- raw\_benchmark\_scores\_hcm\_2022.csv  
- raw\_benchmark\_scores\_hcm\_2023.csv  
- raw\_subject\_groups.csv

Trong đó, ba tệp đầu chứa dữ liệu điểm chuẩn của các trường đại học tại TP.HCM tương ứng với các năm 2021, 2022 và 2023. Tệp 'raw\_subject\_groups.csv' chứa thông tin về các tổ hợp môn xét tuyển như A00 (Toán, Lý, Hóa), D01 (Toán, Văn, Anh), v.v.  
  
Mỗi dòng dữ liệu trong các tệp CSV bao gồm các trường thông tin chính như:  
- Tên trường đại học  
- Mã ngành  
- Tên ngành  
- Tổ hợp môn xét tuyển  
- Điểm chuẩn tuyển sinh  
- Năm tuyển sinh  
  
Việc cấu trúc dữ liệu rõ ràng và đầy đủ ngay từ đầu giúp tiết kiệm đáng kể thời gian xử lý và làm sạch dữ liệu trong các bước sau.

## **2.4. Nhận xét**

Bước crawl dữ liệu đã được thực hiện một cách kỹ lưỡng và có hệ thống. Nhóm không chỉ đảm bảo dữ liệu đầy đủ về mặt số lượng mà còn chú trọng đến chất lượng, độ chính xác và khả năng mở rộng về sau. Việc sử dụng các công cụ lập trình tự động giúp tiết kiệm thời gian, đồng thời giảm thiểu sai sót so với việc nhập tay thủ công.  
  
Dữ liệu sau khi crawl đã sẵn sàng cho bước tiếp theo là tiền xử lý (preprocessing), nơi dữ liệu sẽ được làm sạch, chuẩn hóa và trích xuất các đặc trưng cần thiết để đưa vào mô hình học máy. Có thể nói, việc thu thập dữ liệu thành công đã đặt nền móng vững chắc cho toàn bộ dự án.

# Chương 3: Clean data

## **3.1. Dữ liệu thô:**

Cấu trúc dữ liệu: Gồm 7 cột:  
university\_code, specialized\_major, major\_name, major\_code, subject\_group\_code, benchmark\_score, note

Vấn đề gặp phải trong dữ liệu thô:

* Nhiều giá trị thiếu ở cột specialized\_major.
* Cột note cũng có nhiều giá trị rỗng.
* Không có thông tin về năm tuyển sinh (year), loại hình chương trình (program\_type), phương thức xét tuyển (admission\_method), hoặc thang điểm (score\_scale).
* Không thống nhất định dạng khóa chính ngành giữa các trường.

## **3.2. Dữ liệu đã được làm sạch:**

* Số cột tăng từ 7 lên 10, bao gồm các trường thông tin mới như:
* year: Năm tuyển sinh
* program\_type: Loại hình chương trình đào tạo
* uni\_major\_code: Mã tổ hợp ngành - trường
* admission\_method: Phương thức xét tuyển
* score\_scale: Thang điểm sử dụng

Không còn dòng nào bị thiếu dữ liệu ở tất cả các cột.

## **3.3 Các bước làm sạch dữ liệu:**

### ***3.3.1. Xử lý dữ liệu thiếu***

* Thay thế giá trị trống trong specialized\_major bằng dữ liệu suy luận từ các trường khác hoặc gán nhãn mặc định.
* Loại bỏ hoặc giữ lại giá trị note tùy theo mức độ cần thiết.

### ***3.3.2. Chuẩn hóa và bổ sung dữ liệu***

* Thêm các cột mới như:
* year: Thêm dựa vào nguồn hoặc metadata gốc.
* admission\_method, program\_type: Gán theo cấu trúc ngành hoặc loại hình đào tạo.
* score\_scale: Chuẩn hóa thang điểm (ví dụ: thang 30).

### ***3.3.3. Loại bỏ trùng lặp***

* Kiểm tra và loại bỏ các dòng bị trùng lặp về university\_code, major\_code, subject\_group\_code, và benchmark\_score.

### ***3.3.4. Kiểm tra định dạng dữ liệu***

* Chuyển đổi các cột số như benchmark\_score và score\_scale về kiểu float/int đúng chuẩn.
* Kiểm tra ký tự đặc biệt, lỗi mã hóa

## **Kết luận:**

* Dữ liệu thô có nhiều thiếu sót và không đồng nhất về định dạng.
* Sau khi làm sạch, dữ liệu trở nên đầy đủ, có thể sử dụng cho các bước phân tích, thống kê, hoặc trực quan hóa tiếp theo.
* Việc bổ sung các thông tin như year, program\_type, và chuẩn hóa uni\_major\_code đã tăng cường khả năng phân tích đa chiều và ứng dụng trong thực tế.

# Chương 4: Data visualization

## **4.1. Phân phối và xu hướng điểm chuẩn theo phương thức xét tuyển**

Dữ liệu điểm chuẩn được phân tích theo ba phương thức phổ biến: Tốt nghiệp THPT và Xét học bạ (thang 30), Đánh giá năng lực (ĐGNL) ĐHQG-HCM (thang 1200). Kết quả cho thấy:

* Phương thức thi THPT có điểm chuẩn tập trung cao, phần lớn từ 20–30 điểm, phản ánh mức độ cạnh tranh lớn.
* Xét học bạ phân tán hơn, nhưng vẫn chủ yếu nằm trong khoảng 18–28 điểm.
* Điểm chuẩn ĐGNL tập trung từ 600–900 điểm, cho thấy sự phân hóa rõ rệt giữa các thí sinh.

So sánh bằng biểu đồ boxplot chỉ ra rằng điểm trung bình phương thức thi THPT cao hơn Xét học bạ, và khoảng IQR của ĐGNL rộng hơn, phản ánh sự đa dạng giữa các ngành/trường.

Bên cạnh đó, xu hướng điểm chuẩn trung bình qua 3 năm cho thấy:

* Thi THPT và Học bạ giảm nhẹ năm 2022, tăng lại năm 2023.
* ĐGNL có xu hướng giảm dần, có thể do số lượng thí sinh tăng hoặc thay đổi đề thi.

## **4.2. Sự thay đổi số lượng ngành và hệ đào tạo**

Số lượng ngành áp dụng mỗi phương thức xét tuyển có sự thay đổi theo thời gian:

* Tốt nghiệp THPT giảm nhẹ, thể hiện xu hướng đa dạng hóa phương thức tuyển sinh.
* Xét học bạ tăng mạnh vào năm 2022, sau đó giảm nhẹ.
* ĐGNL giữ ổn định, được duy trì tại nhiều trường, đặc biệt trong hệ thống ĐHQG-HCM.

Về hệ đào tạo, hệ Đại trà chiếm khoảng 85% tổng số ngành, giữ vai trò chủ đạo. Hệ Chất lượng cao chiếm khoảng 15%, tập trung ở một số trường trong khối ĐHQG-HCM, phục vụ nhóm ngành chuyên sâu, chất lượng cao.

## **4.3. Phân tích tổ hợp môn và điểm chuẩn tại các trường ĐHQG-HCM**

Trong các phương thức có sử dụng tổ hợp môn, 5 tổ hợp phổ biến nhất (A00, A01, B00, C00, D01) chiếm tỷ trọng lớn, phản ánh xu hướng tập trung vào các khối thi truyền thống. Các tổ hợp còn lại được nhóm vào "Khác".

Với các trường thuộc ĐHQG-HCM, điểm chuẩn trung bình có sự khác biệt rõ:

* VNUHCM-UIT và VNUHCM-UT (các ngành kỹ thuật) có điểm chuẩn cao và ổn định.
* VNUHCM-USSH (khối xã hội) có mức điểm chuẩn thấp hơn, phù hợp với đặc thù ngành.
* Điểm trung bình ĐGNL tại các trường này có xu hướng giảm nhẹ theo thời gian

.

# Chương 5: Machine Learning

Trong chương này, chúng ta tiến hành xây dựng mô hình dự đoán điểm chuẩn đại học dựa trên dữ liệu tuyển sinh giai đoạn 2021–2023. Mục tiêu là tạo ra một hệ thống có khả năng ước lượng chính xác mức điểm chuẩn đầu vào dựa trên các yếu tố như mã ngành, tổ hợp xét tuyển, phương thức tuyển sinh, hình thức đào tạo và năm tuyển sinh. Toàn bộ quy trình được thực hiện qua các bước tiền xử lý, mã hóa dữ liệu, xây dựng mô hình và đánh giá hiệu quả dự đoán.

## **5.1 Tiền xử lý và chuẩn hóa dữ liệu**

Dữ liệu gốc được nạp từ file CSV đã qua xử lý sơ bộ ở các bước trước. Đầu tiên, các trường có tần suất xuất hiện thấp được gom nhóm vào nhãn “Other” nhằm giảm số lượng biến phân loại cần xử lý. Cụ thể, chỉ giữ lại 20 trường đại học, 40 mã ngành và 20 tổ hợp môn xuất hiện phổ biến nhất. Sau đó, những dòng dữ liệu trùng lặp theo tổ hợp thuộc tính đầu vào được loại bỏ.

Bên cạnh đó, một bước quan trọng khác là chuẩn hóa thang điểm. Một số ngành có điểm chuẩn theo thang 1200, vì vậy cần quy đổi về thang điểm 30 để đảm bảo tính nhất quán. Cuối cùng, các cột không cần thiết như score\_scale, major\_name và uni\_major\_code được loại bỏ khỏi tập dữ liệu.

## **5.2 Mã hóa dữ liệu và chia tập huấn luyện**

Do đầu vào chứa nhiều biến phân loại (categorical), dữ liệu được mã hóa bằng kỹ thuật One-Hot Encoding thông qua ColumnTransformer của scikit-learn. Các đặc trưng sau khi mã hóa được chuyển sang dạng mảng thưa (sparse matrix) để tiết kiệm bộ nhớ.

Dữ liệu sau đó được chia làm ba phần: 70% cho tập huấn luyện, 15% cho tập kiểm tra và 15% còn lại dùng để đánh giá cuối cùng. Việc chia tách này giúp đảm bảo đánh giá mô hình khách quan và tránh hiện tượng overfitting.

## **5.3 Xây dựng và đánh giá mô hình**

Bốn thuật toán hồi quy phổ biến đã được thử nghiệm bao gồm:

* Linear Regression
* Decision Tree Regressor
* Random Forest Regressor
* Gradient Boosting Regressor

Các mô hình được huấn luyện trên tập train và đánh giá bằng hai chỉ số: hệ số xác định R² và lỗi bình phương trung bình (MSE) trên tập validation.

Kết quả cho thấy, Random Forest Regressor có hiệu năng vượt trội hơn các mô hình còn lại với R² cao và MSE thấp nhất. Điều này cho thấy mô hình có khả năng học được các mối quan hệ phi tuyến và phức tạp trong dữ liệu.

Sau khi xác định được mô hình tốt nhất, toàn bộ dữ liệu huấn luyện và validation được gộp lại để huấn luyện một mô hình cuối cùng. Mô hình Random Forest tiếp tục được sử dụng và đánh giá lần nữa trên tập test chưa từng được nhìn thấy.

Trên tập test, mô hình đạt được R² ≈ 0.86 và MSE ≈ 0.32, cho thấy khả năng dự đoán điểm chuẩn khá tốt. Biểu đồ so sánh giữa giá trị thực tế và dự đoán cho thấy đa số điểm dữ liệu nằm gần đường chéo lý tưởng x=y, phản ánh tính chính xác của mô hình.

# Nguồn dữ liệu

**[1]** [**https://huongnghiep.hocmai.vn/diem-chuan/**](https://huongnghiep.hocmai.vn/diem-chuan/)

**[2]** [**https://diemthi.tuyensinh247.com/to-hop-mon.html?gad\_source=1&gbraid=0AAAAA-Z0638cw\_kF0TERH-1l8uApGuPdG&gclid=Cj0KCQjwnui\_BhDlARIsAEo9GuvWhjJaV4i53eIgHVw8rapUIhUU4I\_3dMinLjWrXrzU628Et3HJUdAaApLFEALw\_wcB**](https://diemthi.tuyensinh247.com/to-hop-mon.html?gad_source=1&gbraid=0AAAAA-Z0638cw_kF0TERH-1l8uApGuPdG&gclid=Cj0KCQjwnui_BhDlARIsAEo9GuvWhjJaV4i53eIgHVw8rapUIhUU4I_3dMinLjWrXrzU628Et3HJUdAaApLFEALw_wcB)

**[3]** [**https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-09-2022-TT-BGDDT-Danh-muc-thong-ke-nganh-dao-tao-giao-duc-dai-hoc-516993.aspx**](https://thuvienphapluat.vn/van-ban/Giao-duc/Thong-tu-09-2022-TT-BGDDT-Danh-muc-thong-ke-nganh-dao-tao-giao-duc-dai-hoc-516993.aspx)