

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA KHOA HỌC MÁY TÍNH



BÀI TẬP THỰC HÀNH 1

Giảng Viên: Nguyễn Thị Anh Thư

Mã Lớp: CS313.O21

Nhóm 04

<i>Trương Văn Khải</i>	<i>21520274</i>
<i>Đoàn Nhật Sang</i>	<i>21522542</i>
<i>Lê Ngô Minh Đức</i>	<i>21520195</i>
<i>Lê Minh Quang</i>	<i>21522510</i>
<i>Lê Yến Nhi</i>	<i>21522427</i>
<i>Hoàng Thị Mỹ Hạnh</i>	<i>21522044</i>
<i>Hoàng Tiến Đạt</i>	<i>21520696</i>

TP. Hồ Chí Minh, 2/2024

MỤC LỤC

1. Bộ dữ liệu sử dụng.....	1
2. Giới thiệu bộ dữ liệu sử dụng.....	1
3. Mô tả sơ bộ về tập dữ liệu:.....	1
3.1. Course resources.....	2
3.1.1. Course Info (entity).....	2
3.1.2. Video (entity).....	4
3.1.3. Problem (entity).....	5
3.1.4. School (entity).....	6
3.1.5. Teacher (entity).....	7
3.1.6. Course - Field (relation).....	8
3.1.7. Course - School (relation).....	8
3.1.8. Course - Teacher (relation).....	8
3.1.9. Exercise - Problem (relation).....	9
3.1.10. Video ID - CCID (relation).....	9
3.2. Student behaviors.....	9
3.2.1. Student profile (entity).....	9
3.2.2. Comment (entity).....	10
3.2.3. Reply (entity).....	11
3.2.4. User-video (relation).....	12
3.2.5. User-problem.....	13
3.2.6. User-xiaomu.....	14
3.2.7. Course-comment.....	15
3.2.8. User-comment.....	15
3.2.9. User-reply.....	15
3.2.10. Comment-reply.....	15
3.3. Concepts.....	16
3.3.1 Concept.....	16
3.3.2 Other.....	16
3.3.3 Paper.....	17
3.3.4. Concept-Other.....	19
3.3.5. Concept-Paper.....	19
3.3.6. Concept-Problem.....	19
3.3.7. Concept-Video.....	20
3.3.8. Concept-Comment.....	20
3.4. Prerequisites.....	20

3.4.1. CS.json.....	20
3.4.2. Math.json.....	21
3.4.3. Psy.json.....	21
4. Nhận xét và dự đoán mục tiêu sử dụng bộ dữ liệu.....	22
4.1. Nhận xét.....	22
4.2. Dự đoán mục tiêu sử dụng bộ dữ liệu.....	22

NỘI DUNG

1. Bộ dữ liệu sử dụng

Sau quá trình khảo sát và tìm hiểu, chúng tôi trước quyết định sử dụng bộ dữ liệu: **MOOCCubeX**

2. Giới thiệu bộ dữ liệu sử dụng

Bộ dữ liệu được thu thập từ nền tảng XuetangX - Đây là một trong những đối tác của edX. Tuy hệ thống ra mắt vào tháng 10 năm 2013 nhưng đến ngày 31 tháng 5 năm 2021 đã cung cấp hơn 6.000 khóa học, bao gồm các khóa từ Đại học Thanh Hoa, Đại học Bắc Kinh và các khóa học của edX từ MIT, Stanford, UC Berkeley, ... thu hút 4.500.000 người dùng đăng ký. XuetangX cung cấp đa dạng tài nguyên học tập, cho phép người dùng tự do ghi danh vào các khóa học và tham gia vào quá trình học đầy đủ bao gồm học qua video, làm bài tập và tham gia thảo luận. Các dữ liệu này có mối liên hệ chặt chẽ và được quản lý tốt, nên thường được sử dụng làm cơ sở lý tưởng cho MOOCCubeX.

3. Mô tả sơ bộ về tập dữ liệu:

Bộ dữ liệu gồm hai phần chính: Tài nguyên khóa học (Course Resource) và Hành vi học sinh (Student Behavior)

Course Resource:

Phần Tài nguyên khóa học của MOOCCubeX bắt đầu bằng việc thu thập dữ liệu khóa học từ XuetangX. Sau khi loại bỏ các khóa học thử nghiệm và khóa học không còn hoạt động, thông tin chi tiết về 4.216 khóa học đã được thu thập. Ở giai đoạn này, tên và mô tả của mỗi khóa học được lưu trữ dưới dạng văn bản, và mỗi khóa học được gán một mã id. Các khóa học trong MOOCs không độc lập với nhau. Một khóa học bao gồm nhiều chương giảng dạy, và một chương thường bao gồm một loạt video và bài tập. Thông tin có cấu trúc như vậy cũng rất quan trọng, do đó, việc thu thập thông tin liên quan đến khóa học, bao gồm giáo trình của khóa học và danh

sách tài nguyên bao gồm (video, bài tập, và bình luận) được lưu trữ dưới dạng danh sách. Ngoài ra, thông tin về giáo viên và trường đại học của khóa học, cùng với giới thiệu về họ được thu thập từ web. Loại thông tin này có thể xây dựng các mối liên kết cho các khóa học và hỗ trợ các nhiệm vụ liên quan như phát hiện phong cách giảng dạy.

Student behaviours:

Ngoài các nguồn tài nguyên tĩnh, các loại hành vi của sinh viên cũng rất quan trọng cho nghiên cứu học tập thích nghi, giúp mô hình hóa ý định học tập của sinh viên ở các cấp độ nhận thức và các hoạt động xã hội. Do đó, tác giả thu thập các bản ghi chi tiết từ XuetaangX, bao gồm: hồ sơ sinh viên, hành vi xem video, bài tập và thảo luận. Các hành vi này tự nhiên liên kết với các nguồn tài nguyên của khóa học. Mặc dù đã có giấy phép từ nền tảng, tác giả vẫn cần thực hiện các hoạt động giảm nhạy cảm như ẩn danh trong quá trình xử lý dữ liệu.

Mô tả chi tiết của các bảng dữ liệu trong MOOCCubeX như các mục sau:

3.1. Course resources

3.1.1. Course Info (entity)

Mô tả: Thông tin của khóa học và các tài liệu tương ứng

Tên file: **entities/course.json**

Số lượng mẫu: **3781** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
about	Giới thiệu khóa học	string	
id	id của khóa học	string	

field	danh sách các lĩnh vực mà khóa học thuộc về	array của các string	
name	tên của khóa học	string	
prerequisites	mô tả những kiến thức tiên quyết	string	
resource	<p>danh sách các tài nguyên, có thể là một video hoặc một nhóm các bài tập. object gồm các trường:</p> <ul style="list-style-type: none"> - resource_id (string): ID của khóa học. Nếu bắt đầu bằng “V_” thì nó là một Video (resource_id được xem như video_id), nếu bắt đầu bằng “Ex_” thì nó là một nhóm Exercise (resource_id được xem như exercise_id). Xem thêm lưu ý dưới bảng này. - chapter (string): số chapter - titles (list<string>): danh sách các tựa đề, gồm chapter title, video title, v.v. 	list của các object	

Lưu ý:

- Nhiều video_ids tương ứng với 1 ccid (1 ccid xác định 1 video). Những video_id này tương ứng với phần trình chiếu của cùng 1 ccid tại những thời

điểm bắt đầu khác nhau, ví dụ (spring 2018 / fall 2020). Sự tương ứng giữa video_id và ccid có thể tìm thấy ở mục video_id-ccid.

- Mỗi tập hợp exercises (exercise) tương ứng với nhiều câu hỏi (problem). Sự tương ứng giữa chúng có thể tìm thấy ở mục exercise-problem.

3.1.2. Video (entity)

Mô tả: Tên video và chú thích. Nội dung của video, khóa học của nó, chương và thứ tự có thể tìm thấy trong file course.json

Tên file: **entities/video.json**

Số lượng mẫu: **59581** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
ccid	id duy nhất của video	string	
name	tên của khóa học	string	
start	thời gian bắt đầu mỗi câu phụ đề của video	list<float>	[0, +vc)
end	thời điểm mỗi câu của phụ đề video kết thúc	list<float>	[0, +vc)
text	nội dung phụ đề từng câu trong video	list<string>	

3.1.3. Problem (entity)

Mô tả: câu hỏi (problem) thực hành của nhóm bài tập. Lưu ý: Mỗi nhóm bài tập sẽ tương ứng với nhiều câu hỏi (problem).

Tên file: **entities/problem.json**

Số lượng mẫu: **2454422** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
problem_id	id của vấn đề, bắt đầu bằng Pm_	int	
excercise_id	id của bài tập, bắt đầu bằng Ex_	string	
language	ngôn ngữ mô tả của vấn đề	string	{‘English’, 'Chinese'}
title	tựa của bài tập	string	
content	mô tả vấn đề	string	
option	lựa chọn của vấn đề	object chứa các cặp key, value. Key có kiểu string, là ký hiệu A, B, C, ... của lựa chọn. Value có kiểu string, là mô tả của lựa chọn	Key nhận các giá trị như: “A”, “B”, “C”, “D”, ...
answer	câu trả lời cho câu hỏi	string	

score	điểm cho câu hỏi	float	
type	loại câu hỏi	int	
typetext	loại câu hỏi	string	
location	chương của vấn đề	string	
context_id	các leaf_id liên quan đến vấn đề	list<int>	

3.1.4. School (entity)

Mô tả: thông tin của trường học

Tên file: **entities/school.json**

Số lượng mẫu: **429** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	ID trường, bắt đầu với s_	string	
name	Tên tiếng Trung của trường	string	
name_en	Tên tiếng Anh của trường	string	
sign	Tên viết tắt của tên tiếng Anh của trường	string	
about	Giới thiệu	string	

motto	Châm ngôn của trường	string	
-------	----------------------	--------	--

3.1.5. Teacher (entity)

Mô tả: thông tin của giáo viên

Tên file: **entities/teacher.json**

Số lượng mẫu: **17018** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	ID của giáo viên, bắt đầu với T_	string	
name	Tên tiếng Trung của giáo viên	string	
name_en	Tên tiếng Anh của giáo viên	string	
about	Hồ sơ của giáo viên	string	
job_title	chức danh công việc	string	
org_name	tổ chức liên kết	string	

3.1.6. Course - Field (relation)

Mô tả: Các lĩnh vực mà khóa học thuộc về. Các lĩnh vực của các khóa học được đánh dấu thủ công được lấy từ 88 lĩnh vực trong “Danh mục ngành, chuyên ngành cấp bằng tiến sĩ, thạc sĩ và đào tạo sau đại học” do Bộ Giáo dục ban hành năm 1997. Mỗi khóa học có thể thuộc nhiều lĩnh vực.

Tên file: **relations/course-field.json**

Số lượng mẫu: **632** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
course_id	ID của khóa học	int	
course_name	Tên của khóa học	string	
field	Danh sách lĩnh vực được đánh nhãn thủ công	list<str>, mỗi string đại diện cho một lĩnh vực	

3.1.7. Course - School (relation)

Mô tả: Trường mà khóa học tương ứng được dạy.

Định dạng: {course ID}\t{school ID}

Tên file: **relations/course-school.txt**

Số lượng mẫu: **3983** mẫu

3.1.8. Course - Teacher (relation)

Giảng viên của khóa học.

Định dạng: {course ID}\t{teacher ID}

Tên file: **relations/course-teacher.txt**

Số lượng mẫu: **97192** mẫu

3.1.9. Exercise - Problem (relation)

Mô tả: Tập hợp các vấn đề (câu hỏi) chứa trong bài tập.

Định dạng: {exercise ID}\t{question ID}

Mỗi dòng trong bộ Exercise là một bộ gồm bài tập (exercise) tương ứng với câu hỏi (problem). Ví dụ: Ex_143 Pm_1

Tên file: **relations/exercise-problem.txt**

Số lượng mẫu: **6252830** mẫu

3.1.10. Video ID - CCID (relation)

Mô tả: Video và ccid tương ứng của nó.

Định dạng: {Video ID}\t{ccid}

Tên file: **relations/video_id-ccid.txt**

Số lượng mẫu: **2798892** mẫu

3.2. Student behaviors.

3.2.1. Student profile (entity)

Mô tả: Thông tin của học sinh (user)

Tên file: **entities/user.json**

Số lượng mẫu: **3330294** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Id người dùng, bắt đầu bằng "U_"	string	

name	Tên người dùng	string	
gender	Giới tính	int	
school	Tên trường	string	
year_of_birth	Năm sinh	int	
course_order	Các mã khóa học đã chọn	list<int>	
enroll_time	Thời gian đăng kí tương ứng với từng khoá học	list<DateTime>. DateTime có định dạng “YYYY-MM-DD HH:MM:SS”	

3.2.2. Comment (entity)

Mô tả: Thông tin các bình luận của các học sinh (user) đăng tải lên

Tên file: **entities/comment.json**

Dung lượng file: 2.1GB

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Comment ID, bắt đầu bằng “Cm_”	String	

user_id	ID của người dùng đã bình luận, bắt đầu bằng "U_"	Int	
text	Nội dung bình luận	String	
create_time	Thời gian bình luận	DateTime, có định dạng “YYYY-MM-DD HH:MM:SS”	
resource_id	ID của tài nguyên (như video, exercise) mà user bình luận	String	Có thể nhận giá trị null

Lưu ý: Thực tế, trường resource_id không được nhắc đến trong file user-en.md trên github của nhóm tác giả nhưng dữ liệu thực lại có thêm trường này.

```
[{"id": "Cm_187", "user_id": 11731, "text": "资质统建", "resource_id": "V_454874", "create_time": "2019-09-05 17:12:24"}, {"id": "Cm_188", "user_id": 11731, "text": "资质统建", "resource_id": "V_454874", "create_time": "2019-09-05 17:12:24"}, {"id": "Cm_190", "user_id": 11731, "text": "资质统建", "resource_id": "V_454874", "create_time": "2019-09-05 17:12:29"}, {"id": "Cm_192", "user_id": 11731, "text": "资质统建", "resource_id": "V_454874", "create_time": "2019-09-05 17:12:29"}]
```

3.2.3. Reply (entity)

Mô tả: Thông tin của phần trả lời bình luận (reply) của học sinh (user)

Tên file: **entities/reply.json**

Số lượng mẫu: **331011** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Reply ID, bắt đầu bằng	string	

	“Rp_”		
user_id	ID của người dùng đã bình luận, bắt đầu bằng "U_"	string	
text	Nội dung phản hồi	string	
create_time	Thời gian phản hồi	DateTime, có định dạng “YYYY-MM-DD HH:MM:SS”	

3.2.4. User-video (relation)

Mô tả: Tốc độ và các bước nhảy thời gian của người dùng khi xem video.

Tên file: **relations/user-video.json**

Dung lượng: **3 GB**

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
user_id	ID của user, bắt đầu bằng “U_”	string	
seq	Mảng, trình tự người dùng xem video, mỗi đối tượng trong mảng là trình tự thời gian người dùng xem một video nhất định,	list<object>. Mỗi object sẽ gồm 2 trường video_id (string) và segment (list<object>). Mỗi phần tử trong	

	bao gồm thời gian xem video, thời gian bắt đầu và kết thúc của video, và tốc độ xem video, v.v.	segment bao gồm các trường start_point (float), end_point (float), speed (float), local_start_time (int)	
--	---	--	--

Lưu ý: ví dụ về 1 phần tử trong seq.

```
{'video_id': 'V_1395639', 'segment': [{ 'start_point': 100.0, 'end_point': 106.25, 'speed': 1.25, 'local_start_time': 1588438980}, { 'start_point': 180.0, 'end_point': 186.25, 'speed': 1.25, 'local_start_time': 1588439045}]}
```

3.2.5. User-problem

Mô tả: Người dùng làm bài tập nào.

Tên file: **relations/user-problem.json**

Dung lượng: **21 GB**

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
log_id	ID của bản ghi câu hỏi của người dùng, kết hợp với khóa duy nhất của user_id và problem_id	string	
user_id	ID người dùng, bắt đầu bằng U_	string	
problem_id	ID vấn đề, bắt đầu bằng Pm_	string	

is_correct	Câu hỏi có đúng không	bool	0 hoặc 1
attempts	Số lượng câu hỏi đã thử	int	
score	Điểm của người dùng	float	
submit_time	Thời gian làm câu hỏi	DateTime, có định dạng “YYYY-MM-DD HH:MM:SS”	

3.2.6. User-xiaomu

Mô tả: Tương tác của người dùng với Xiaomu (bot QA của XuatangX).

Tên file: **relations/user-xiaomu.json**

Số lượng mẫu: **108351** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
user_id	ID của user, bắt đầu bằng “U_”	string	
question_type	Loại câu hỏi của user	string	
question	Câu hỏi hỏi bởi user	string	

3.2.7. Course-comment

Mô tả: Bình luận của người dùng liên quan đến khóa học.

Định dạng: {course ID}\t{review ID}.

Tên file: **relations/course-comment.txt**

Số lượng mẫu: **10181950** mẫu

3.2.8. User-comment

Mô tả: Bình luận của Người dùng.

Định dạng: {User ID}\t{Comment ID}.

Tên file: **relations/user-comment.txt**

Số lượng mẫu: **8422134** mẫu

3.2.9. User-reply

Mô tả: Phản hồi Bình luận của Người dùng.

Định dạng: {User ID}\t{Reply ID}.

Tên file: **relations/user-reply.txt**

Số lượng mẫu: **331011** mẫu

3.2.10. Comment-reply

Mô tả: Phản hồi bình luận liên quan đến khái niệm.

Định dạng là {Concept ID}\t{Reply ID}.

Tên file: **relations/comment-reply.txt**

Số lượng mẫu: **370493** mẫu

3.3. Concepts

3.3.1 Concept

Mô tả: thông tin giới thiệu môn học

Tên file: **entities/concept.json**

Số mẫu dữ liệu: **637572** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Mã khái niệm, thể hiện theo format K_{tên khái niệm}_{trường khái niệm}	string	
name	Tên khái niệm (mặc định giống với tên khái niệm trong mã khái niệm)	string	
context	Nội dung của khái niệm được xuất hiện trong “ <u>part of Wiki/Baidu Encyclopedia, Zhihu Q&A</u> ”. Nội dung này bắt buộc có sẵn tên khái niệm	string	

3.3.2 Other

Mô tả: Các tài liệu liên quan được thu thập bên ngoài những khoá học

Tên file: **entities/other.json**

Số mẫu dữ liệu: **210349** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Mã dữ liệu	string	
concept	khái niệm của thông tin của tài liệu thu thập được	string	
type	Nguồn dữ liệu, bao gồm [“zhihu”, ”baike”, “wiki”]	string	[“zhihu”, ”baike”, “wiki”]
content	Nội dung của tài liệu	string	

3.3.3 Paper

Mô tả: Những bài báo khoa học liên quan

Tên file: **entities/paper.json**

Dung lượng: **6.8 GB**

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
id	Mã bài báo	string	
concept	Mã khái niệm	string	
abstract	Phần giới thiệu (abstract) của bài báo	string	[“zhihu”, ”baike”,

			“wiki”]
author	Nội dung của tài liệu	string	
lang	Ngôn ngữ của bài báo	string	“en”: Tiếng Anh “zh”: Tiếng Trung
pages	Số lượng trang	Integer	
num_citation	Số lượng citation (trích dẫn từ bài báo khác) tính trong năm 2020	Integer	
score	Điểm số tương đồng giữa bài báo và khái niệm. Điểm càng cao thì càng liên quan nhiều đến khái niệm	Float	
sourcetype	Nguồn của bài báo (hiện tại tất cả là publication)	String	
title	Tên bài báo	String	
venue	Diễn đàn bài báo được đăng tải	String	

urls	các đường link dẫn đến bài báo	List <String>	
year	năm xuất bản	Year	

3.3.4. Concept-Other

Mô tả: Khái niệm liên quan với tài nguyên ngoài môn học

Định dạng: {concept ID}\t{resource ID}

Tên file: **relations/concept-other.txt**

Số lượng mẫu: **379926** mẫu

3.3.5. Concept-Paper

Khái niệm liên quan với các bài báo ngoài môn học

Định dạng: {concept ID}\t{paper ID}

Tên file: **relations/concept-paper.txt**

Số lượng mẫu: 5410752 mẫu

3.3.6. Concept-Problem

Mô tả: Khái niệm liên quan với các vấn đề

Định dạng: {Concept ID}\t{Question ID}

Tên file: **relations/concept-problem.txt**

Số lượng mẫu: **33180** mẫu

3.3.7. Concept-Video

Mô tả: Khái niệm liên quan đến video

Định dạng: {concept ID}\t{ccid}

Tên file: **relations/concept-video.txt**

Số lượng mẫu: **624683** mẫu

3.3.8. Concept-Comment

Mô tả: Khái niệm liên quan đến bình luận

Định dạng: {concept ID}\t{review ID}

Tên file: **relations/concept-comment.txt**

Số lượng mẫu: **31074** mẫu

3.4. Prerequisites

3.4.1. CS.json

Mô tả: Dự đoán và chú thích của con người về các tiên điều kiện của môn Khoa học Máy tính.

Tên file: **prerequisites/cs.json**

Số lượng mẫu: **492102** mẫu

Trường	Nội dung	Kiểu dữ liệu	Miền giá trị
c1	Khái niệm điều kiện tiên quyết	string	
c2	Khái niệm điều kiện sau sửa	string	

	chữ		
ground_truth	Chỉ ra có mối quan hệ sửa chữa tuần tự hay không, 1 có nghĩa là có, 0 có nghĩa là không.	int	0 hoặc 1
text_predict	Cung cấp kết quả dự đoán sử dụng đặc điểm văn bản.	list<float>	
graph_predict	Mức độ tin cậy của dự đoán được đặt được bằng các đặc điểm đồ thị.	list<float>	

3.4.2. Math.json

Mô tả: Chú thích và dự đoán các khái niệm trong lĩnh vực toán học, theo định dạng giống CS.json.

Tên file: **prerequisites/math.json**

Số lượng mẫu: **331202** mẫu

3.4.3. Psy.json

Mô tả: Chú thích và dự đoán các khái niệm trong lĩnh vực tâm lý học, theo định dạng giống CS.json.

Tên file: **prerequisites/psy.json**

Số lượng mẫu: **757771** mẫu

4. Nhận xét và dự đoán mục tiêu sử dụng bộ dữ liệu

4.1. Nhận xét

Sau khi đọc và khảo sát sơ lược bộ dữ liệu, chúng tôi rút ra được một số nhận xét sau:

- Bộ dữ liệu MOOCCubeX là một bộ dữ liệu có mức độ đa dạng cao và bao gồm các thông tin, tài nguyên về giáo dục cũng như một số thông tin có liên quan khác đến học sinh về việc học.
- Đây là một bộ dữ liệu có kích thước lớn, do đó từ bộ dữ liệu này có thể hỗ trợ việc khám phá dữ liệu phục vụ cho các mục đích hỗ trợ học tập với các phương pháp tiếp cận học máy, học sâu, ...
- Nhìn chung, đây là một bộ dữ liệu không đồng nhất nhưng được tổ chức một cách bài bản, linh hoạt với mức độ chi tiết rất cao. Điều này giúp cho việc sử dụng tài nguyên có thể linh hoạt với nhiều mục đích sử dụng khác nhau, đồng thời việc tìm kiếm dữ liệu cũng như thiết lập các mô hình để khai phá dữ liệu cũng dễ dàng hơn.

4.2. Dự đoán mục tiêu sử dụng bộ dữ liệu

Vì đây là một bộ dữ liệu có liên quan đến chủ đề học tập nên chúng tôi dự định sẽ khai thác các bài toán có liên quan đến lĩnh vực “Cố vấn học tập thông minh tại các trường đại học” hay “Cố vấn học tập thông minh cho các nền tảng học tập trực tuyến”. Sau khi tiến hành khảo sát cũng như thảo luận nhóm, chúng tôi đưa ra hai bài toán có thể sẽ giải quyết sau:

1. **Bài toán 1:** Hệ thống khuyến nghị hay cố vấn cho người học đăng ký các môn học, khóa học theo các định hướng chuyên ngành dựa trên hành vi học tập của người học.

2. **Bài toán 2:** Đề xuất video bài học tiếp theo cho người học dựa trên các chuỗi video học từ lịch sử của họ.

Ngoài ra, trong quá trình tìm hiểu của cả nhóm sau này, chúng tôi sẽ cố gắng kết hợp sử dụng các bộ dữ liệu khác cũng như các loại bài toán khác để tiến hành xây dựng hệ thống phù hợp hơn với môi trường học tập ở Việt Nam nói chung, cũng như cho các sinh viên đang học tập tại trường Đại học Công Nghệ Thông Tin nói riêng.