

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



HUỲNH THỊ NGỌC CẨM

**TIẾP CẬN HỌC MÁY ĐỂ TUYỂN CHỌN
VÀ TRỰC QUAN HÓA ĐỂ HUẤN LUYỆN
HỌC SINH ĐỘI TUYỂN**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Hà Nội - Năm 2026

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

HUỲNH THỊ NGỌC CẨM

**TIẾP CẬN HỌC MÁY ĐỂ TUYỂN CHỌN
VÀ TRỰC QUAN HÓA ĐỂ HUẤN LUYỆN
HỌC SINH ĐỘI TUYỂN**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Xác nhận của Học viện

Khoa học và Công nghệ



Người hướng dẫn

(Ký, ghi rõ họ tên)

A handwritten signature in blue ink, likely belonging to PGS.TS. Trần Vĩnh Phước.

PGS.TS. Trần Vĩnh Phước

Hà Nội - Năm 2026

LỜI CAM ĐOAN

Tôi cam đoan luận án: "Tiếp cận học máy để tuyển chọn và trực quan hóa để huấn luyện học sinh đội tuyển" là công trình nghiên cứu của chính mình dưới sự hướng dẫn khoa học của thầy PGS.TS. Trần Vĩnh Phước. Luận án có sử dụng tài liệu tham khảo từ nhiều nguồn khác nhau với các dữ liệu trích dẫn được ghi rõ nguồn gốc. Các kết quả nghiên cứu của tôi được công bố chung với các tác giả khác đã được sự nhất trí của đồng tác giả khi đưa vào luận án. Các số liệu, kết quả được trình bày trong luận án là hoàn toàn trung thực và chưa từng được công bố trong bất kỳ một công trình nào khác ngoài các công trình công bố của tác giả. Luận án được hoàn thành trong thời gian tôi làm nghiên cứu sinh tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Hà Nội, ngày 06 tháng 5 năm 2026

Tác giả luận án



Huỳnh Thị Ngọc Cẩm

LỜI CẢM ƠN

Luận án tiến sĩ được thực hiện tại Học viện Khoa học và Công nghệ thuộc Viện Hàn lâm Khoa học và Công nghệ Việt Nam, dưới sự hướng dẫn khoa học tận tình của Thầy PGS.TS Trần Vĩnh Phước. Tác giả kính trọng và chân thành biết ơn sâu sắc quý thầy cô giảng viên và chuyên viên của Học viện Khoa học và Công nghệ, Viện Cơ học và Tin học Ứng dụng, Viện Công nghệ Thông tin đã truyền đạt những kiến thức quý báu cũng như hỗ trợ trong quá trình học tập và nghiên cứu. Đặc biệt, tác giả kính gửi lời cảm ơn đến Thầy TS. Trương Nguyên Vũ, Viện trưởng Viện Cơ học và Tin học ứng dụng - Viện Hàn lâm Khoa học và Công nghệ Việt Nam và Thầy PGS.TS Trần Vĩnh Phước, đã tận tình hướng dẫn và hỗ trợ tác giả trong suốt quá trình nghiên cứu và thực hiện luận án này.

Tác giả trân trọng cảm ơn Ban Lãnh đạo Viện Công nghệ Thông tin, Viện Cơ học và Tin học Ứng dụng, Ban Giám đốc Học viện Khoa học và Công nghệ, Phòng Đào tạo, các Phòng Ban chức năng đặc biệt các nhà giáo, nhà khoa học tại Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã quan tâm giúp đỡ tạo điều kiện thuận lợi về cơ sở vật chất, nguồn học liệu và các thủ tục hành chính cho tác giả trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án này.

Tác giả chân thành cảm ơn Ban giám hiệu Trường THPT chuyên Huỳnh Mẫn Đạt đã luôn động viên, giúp đỡ tác giả trong công tác để tác giả chuyên tâm vào việc nghiên cứu và hoàn thành luận án, trân trọng cảm ơn thầy cô giáo là huấn luyện viên đội tuyển học sinh giỏi cấp quốc gia môn Tin học và các em học sinh của một số trường trung học phổ thông chuyên đã hỗ trợ làm khảo sát để tác giả thực nghiệm kết quả nghiên cứu của luận án. Tác giả gửi lời cảm ơn sâu sắc đến gia đình, chồng và con, đã luôn bên cạnh hỗ trợ, động viên và tạo điều kiện thuận lợi cho tác giả hoàn thành luận án. Cuối cùng, tác giả cảm ơn các bạn cùng NCS đã luôn chia sẻ, trao đổi, thảo luận trong quá trình học tập và nghiên cứu.

Hà Nội, ngày 06 tháng 5 năm 2026

Tác giả luận án



Huỳnh Thị Ngọc Cẩm

MỤC LỤC

MỞ ĐẦU	1
1. Lý do chọn đề tài.....	1
2. Mục tiêu của luận án	2
3. Đối tượng nghiên cứu của luận án	3
4. Phạm vi nghiên cứu của luận án	3
5. Phương pháp nghiên cứu	4
6. Các đóng góp của luận án.....	5
6.1. Kết quả nghiên cứu nội dung 1	5
6.2. Kết quả nghiên cứu nội dung 2	6
6.3. Kết quả nghiên cứu nội dung 3	6
7. Cấu trúc luận án	6
CHƯƠNG 1. TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU GIÁO DỤC VÀ	
HỌC KHÁM PHÁ ĐƯỢC LIÊN QUAN.....	9
1.1. Giới thiệu	9
1.2. Khai phá dữ liệu giáo dục.....	10
1.3. Dữ liệu giáo dục	11
1.4. Mục tiêu của những bài toán khai phá dữ liệu giáo dục	12
1.5. Những tiếp cận của khoa học máy tính được liên quan.....	15
1.5.1. Tiếp cận Support Vector Machine (SVM)	16
1.5.2. Tiếp cận cây quyết định.....	17
1.5.3. Tiếp cận clustering (gom nhóm)	18
1.5.4. Tiếp cận phân cụm K-means	18
1.5.5. Xử lý dữ liệu	19
1.6. Phương pháp giáo dục	20
1.7. Nguyên lý học tập	20
1.8. Học khám phá	21
1.8.1. Khái niệm.....	21
1.8.2. Mô hình học khám phá.....	21
1.8.3. Phương pháp học khám phá.....	22
1.8.4. Tính chất của học khám phá	24

1.9. Học khám phá có hướng dẫn (Guided Discovery Learning)	26
1.10. Trực quan hóa.....	27
CHƯƠNG 2. DATASET_NĂNG LỰC THÍ SINH DỰ THI HỌC SINH GIỎI	29
2.1. Giới thiệu	29
2.2. Các yếu tố năng lực thí sinh	31
2.2.1. Năng lực chuyên môn.....	32
2.2.2. Kỹ năng.....	32
2.2.3. Cách sống.....	34
2.3. Câu hỏi khảo sát (Questionnaire).....	35
2.4. Thu thập dữ liệu	38
2.5. Tiền xử lý dữ liệu	38
2.5.1. Chuẩn bị dữ liệu	38
2.5.2. Xử lý dữ liệu	39
2.6. Định lượng tập giá trị các biến năng lực của không gian năng lực.....	40
2.6.1. Định lượng các biến năng lực học tập.....	41
2.6.2. Định lượng các biến năng lực phi học tập.....	42
2.7. Định lượng giá trị các thành phần phi học tập của vector năng lực của người được khảo sát	46
2.8. Dataset	47
2.9. Kết luận chương 2	48
CHƯƠNG 3. TIẾP CẬN HỌC MÁY ĐỂ CHỌN HỌC SINH GIỎI VÀO ĐỘI TUYỂN DỰ THI HỌC SINH GIỎI	49
3.1. Giới thiệu	49
3.2. Những công trình nghiên cứu được liên quan	50
3.3. Phát biểu bài toán	51
3.3.1. Câu hỏi nghiên cứu.....	51
3.3.2. Mô tả bài toán.....	51
3.3.3. Tiếp cận Winner-domain	51
3.4. Thực nghiệm và đánh giá kết quả thực nghiệm.....	60
3.5. Kết luận chương 3	78

CHƯƠNG 4. TIẾP CẬN HỌC KHÁM PHÁ CÓ HƯỚNG DẪN TRỰC	
QUAN ĐỂ HUẤN LUYỆN ĐỘI TUYỂN	83
4.1. Học khám phá có hướng dẫn	83
4.1.1. Giới thiệu	83
4.1.2. Lợi ích của học khám phá có hướng dẫn.....	84
4.1.3. Mô hình học khám phá có hướng dẫn.....	85
4.2. Học khám phá có hướng dẫn trực quan	89
4.2.1. Người học.....	89
4.2.2. Bài giảng	89
4.2.3. Tính tương thích.....	90
4.2.4. Tính trực quan	90
4.2.5. Tính phức tạp	90
4.2.6. Người dạy.....	90
4.2.7. Bài giảng trực quan	91
4.3. Đánh giá hệ thống học khám phá có hướng dẫn trực quan.....	95
4.4. Áp dụng phương pháp học khám phá có hướng dẫn trực quan để huấn luyện đội tuyển môn Tin học.....	97
4.5. Kết luận chương 4	99
KẾT LUẬN.....	101
DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ	104
TÀI LIỆU THAM KHẢO	105
PHỤ LỤC	114

DANH MỤC CHỮ VIẾT TẮT

Viết tắt	Tiếng anh	Diễn giải
EDM	Educational Data Mining	Khai phá dữ liệu giáo dục
DL	Discovery Learning	Học khám phá
SVM	Support Vector Machine	Máy vector hỗ trợ, một tiếp cận học máy.
ZPD	Zone of Proximal Development	Vùng phát triển gần trong miền phát triển kiến thức.
AAL	Actual Ability Level	Mức khả năng hiện tại
PAL	Potential Ability Level	Mức khả năng tiềm năng được phát triển khi có người hướng dẫn.
FN	False Negative	Âm tính sai: hệ thống kết luận “không thắng”, thực tế là “thắng”. Bỏ sót một trường hợp thực sự thắng.
FNR	Tỉ lệ False Negative	Tỉ lệ âm tính sai
FP	False Positives	Dương tính sai: hệ thống kết luận “thắng”, nhưng thực tế “không thắng”
FPR	Tỉ lệ False Positive	Tỉ lệ dương tính sai
TN	True Negative	Âm tính đúng: hệ thống kết luận “không thắng”, thực tế là “không thắng”
TNR	Tỉ lệ True Negative	Tỉ lệ âm tính đúng
TP	True Positives	Dương tính đúng: hệ thống kết luận “thắng”, thực tế đúng là “thắng”
TPR	Tỉ lệ True Positive	Tỉ lệ dương tính đúng

DANH MỤC BẢNG

Bảng 2.1. Xử lý dữ liệu bị sai Trường hợp 1 (Họ và tên trong cột 3 của bảng đã được thay đổi, không phải tên thật)	39
Bảng 2.2. Xử lý dữ liệu bị thiếu Trường hợp 2	40
Bảng 2.3. Ánh xạ biến đổi đáp án các biến năng lực học tập $F_s s=1,..,18$ trong kết quả hồi đáp Questionnaire của người được khảo sát.	41
Bảng 2.4. Kết quả số hóa các biến năng lực học tập	42
Bảng 2.5. Ma trận hồi đáp về biến năng lực phi học tập $s=19,20,..,40$ của người được khảo sát $m m=1,2,..,M$	43
Bảng 2.6. Kết quả thực nghiệm lượng hóa các biến năng lực phi học tập của những người được khảo sát	47
Bảng 2.7. Mô hình dataset.....	47
Bảng 3.1. Unnormalized confusion matrix	54
Bảng 3.2. Kết quả thực nghiệm mô hình Winner-domain kết hợp Winner-cosin do luận án đề xuất áp dụng cho các kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và năm học 2024-2025.....	60
Bảng 3.3. Kết quả phát sinh Winner-domains cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024	62
Bảng 3.4. Kết quả chọn Winner-domains tốt cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024	63
Bảng 3.5. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_2^2	63
Bảng 3.6. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_2^2 theo thuật toán Winner-cosin.....	64
Bảng 3.7. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_2^1	64
Bảng 3.8. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_2^1 theo thuật toán Winner-cosin.....	65
Bảng 3.9. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_7^3	65

Bảng 3.10. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_7^3 theo thuật toán Winner-cosin.....	66
Bảng 3.11. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_5^4	66
Bảng 3.12. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_5^4 theo thuật toán Winner-cosin.....	67
Bảng 3.13. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023- 2024 của những đội viên được chọn bằng mô hình Winner-domain kết hợp Winner-cosin	67
Bảng 3.14. Kết quả dự đoán bằng mô hình Random Forest cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024	67
Bảng 3.15. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của những đội viên được dự đoán bằng mô hình Random Forest.....	68
Bảng 3.16. So sánh kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 đội viên đã chọn theo phương pháp truyền thống với quyết định chọn của mô hình Winner-domain kết hợp Winner-cosin và với kết quả dự đoán của mô hình Random Forest	69
Bảng 3.17. Kết quả phát sinh Winner-domains cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025	70
Bảng 3.18. Kết quả chọn Winner-domains tốt cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025	71
Bảng 3.19. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_4^2	72
Bảng 3.20. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_3^1 theo thuật toán Winner-cosin.....	73
Bảng 3.21. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_4^2	73
Bảng 3.22. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_4^2 theo thuật toán Winner-cosin.....	74
Bảng 3.23. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_6^3	74

Bảng 3.24. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_6^3 theo thuật toán Winner-cosin.....	75
Bảng 3.25. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của những đội viên được chọn bằng mô hình Winner-domain kết hợp Winner-cosin	75
Bảng 3.26. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_4^4 theo thuật toán Winner-cosin.....	76
Bảng 3.27. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của những đội viên được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin.	76
Bảng 3.28. Kết quả dự đoán bằng mô hình Random Forest cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025	77
Bảng 3.29. Thực nghiệm mô hình Random Forest dự đoán kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 đội viên đã chọn bằng phương pháp truyền thống	77
Bảng 3.30. So sánh kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 đội viên đã chọn theo phương pháp truyền thống với quyết định chọn của mô hình Winner-domain kết hợp Winner-cosin và với kết quả dự đoán của mô hình Random Forest	78
Bảng 3.31. Tóm tắt so sánh kết quả thực nghiệm của luận án với kết quả của những phương pháp truyền thống	81
Bảng 3.32. So sánh kết quả các mô hình tuyển chọn và dự đoán kết quả của 2 kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và năm học 2024-2025	82

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Sơ đồ thực hiện khai phá dữ liệu giáo dục (EDM) [6].....	10
Hình 1.2. Quy trình gom nhóm dữ liệu giáo dục [5].	18
Hình 1.3. Hệ thống thị giác thu thập, biến dữ liệu thành thông tin/tri thức [1]	28
Hình 2.1 Quy trình thiết lập dataset của luận án	30
Hình 2.2. Khối 3D hồi đáp - định lượng các giá trị của biến năng lực phi học tập $F_{19}, F_{20}, \dots, F_{40}$	44
Hình 3.1. Quy trình thiết lập Winner-domain [CT.2].....	55
Hình 3.2. Quy trình chọn đội viên từ tập ứng viên.....	80
Hình 4.1. Cấu trúc của hệ thống học khám phá có hướng dẫn trực quan.....	91
Hình 4.2. Dòng thông tin trong quy trình học của người học trong hệ thống học khám phá có hướng dẫn trực quan.....	92
Hình 4.3. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài học và hướng dẫn theo các bước.....	93
Hình 4.4. Thiết kế bài giảng trực quan theo tiếp cận bài toán và hướng dẫn theo các bước.....	95
Hình 4.5. Mô hình đánh giá hiệu năng của hệ thống học khám phá có hướng dẫn trực quan áp dụng cho 1 bài giảng.	96

MỞ ĐẦU

1. Lý do chọn đề tài

Mỗi năm, Bộ Giáo dục và Đào tạo Việt Nam tổ chức kỳ thi học sinh giỏi cấp quốc gia dành cho học sinh các trường trung học phổ thông, tham gia theo từng môn học như Tin học, Toán học, Vật lý, Hóa học, Sinh học, v.v. Để chuẩn bị cho kỳ thi này, các trường trung học phổ thông thành lập đội tuyển học sinh giỏi cho từng môn thi và giao cho một giáo viên phụ trách với chức danh huấn luyện viên. Mỗi huấn luyện viên có trách nhiệm tuyển chọn các học sinh đủ năng lực vào đội tuyển và tổ chức bồi dưỡng, huấn luyện chuyên sâu theo yêu cầu của kỳ thi trước khi các đội tuyển tham gia cuộc thi.

Nhiều năm qua, huấn luyện viên tuyển chọn đội viên bằng các cuộc thi tuyển theo chuyên môn và huấn luyện đội viên theo phương pháp truyền thống với người dạy là trung tâm. Kết quả là nhiều trường không nhận được hoặc nhận được rất ít giải học sinh giỏi cấp quốc gia. Phân tích kết quả còn nhận thấy có một vài học sinh giỏi có kết quả tốt khi tuyển vào đội cũng như khi được huấn luyện nhưng không thành công tại kỳ thi học sinh giỏi cấp quốc gia. Do đó, hầu hết các huấn luyện viên của các trường đều mong muốn có phương pháp và quy trình tuyển chọn đội viên một cách khoa học và phương pháp huấn luyện phù hợp với những yêu cầu của kỳ thi.

Công việc thành lập và huấn luyện đội tuyển đang áp dụng tại nhiều trường gồm:

- Tuyển học sinh giỏi vào các đội tuyển học sinh giỏi dự thi học sinh giỏi cấp quốc gia theo từng môn thi.
- Ngay sau khi thành lập được đội tuyển, huấn luyện viên tổ chức huấn luyện đội viên để dự thi.

Với những yêu cầu về thành tích của trường và những thách thức trong công tác tuyển chọn và huấn luyện đội viên đội tuyển, luận án đã đặt ra các nhiệm vụ nghiên cứu:

1. Làm thế nào hiểu được các yếu tố năng lực cần có của một thí sinh tham dự kỳ thi học sinh giỏi cấp quốc gia của môn thi mà đội mình sẽ dự thi?
2. Làm thế nào để tuyển chọn đúng học sinh giỏi có năng lực đáp ứng đúng yêu cầu của kỳ thi học sinh giỏi cấp quốc gia?

3. Áp dụng phương pháp nào để huấn luyện nâng cao năng lực của đội viên đáp ứng đúng yêu cầu của đề thi học sinh giỏi cấp quốc gia?

Những năm gần đây, khoa học máy tính đã xuyên ngành vào giáo dục nhằm nâng cao chất lượng giáo dục. Tiếp cận học máy để khai phá dữ liệu giáo dục đang được phát triển mạnh với các bài toán chính: (1) Phân tầng học sinh theo năng lực tạo ra những nhóm có năng lực tương đương để áp dụng các phương pháp giảng dạy mới; (2) Dự báo kết quả giáo dục của cơ sở giáo dục để điều chỉnh quản lý nhằm cải thiện kết quả dạy và học; (3) Dự báo kết quả học tập của từng học sinh để can thiệp và ngăn chặn sớm việc bỏ học. Tuy nhiên, chưa có công trình nghiên cứu tuyển chọn học sinh giỏi vào đội tuyển.

Về phương pháp huấn luyện, Bộ Giáo dục và Đào tạo đang đẩy mạnh dự án STEM trong các chương trình phổ thông nhằm hướng tới giảng dạy kết hợp các môn khoa học (science), công nghệ (technology), kỹ thuật (engineering) và toán (mathematic) với nguyên lý cơ bản là đặt người học vào trung tâm của quá trình dạy và học. Mục tiêu của STEM không phù hợp với việc huấn luyện đội tuyển theo từng môn học. Học khám phá có hướng dẫn (guided-discovery learning) có những tính chất phù hợp việc huấn luyện đội tuyển vì đội viên được phát triển tính độc lập trong tư duy và trong hoạt động học tập để tham dự kỳ thi học sinh giỏi cấp quốc gia.

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Luận án xây dựng quy trình tuyển chọn đội viên đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông và huấn luyện đội tuyển để tham dự kỳ thi học sinh giỏi cấp quốc gia với kỳ vọng thắng được nhiều giải, dựa trên tiếp cận học máy và trực quan hóa dữ liệu của khoa học máy tính.

Nội dung nghiên cứu: Để thực hiện mục tiêu này, luận án xây dựng quy trình thành lập đội tuyển dự thi học sinh giỏi cấp quốc gia môn Tin học theo các bước sau:

- *Bước 1:* Thành lập đội tuyển tiềm năng với số lượng phù hợp bằng cách tuyển chọn từ học sinh giỏi của các lớp học.
- *Bước 2:* Huấn luyện chuyên sâu cho đội tuyển tiềm năng để nâng cao trình độ chuyên môn và kỹ năng dự thi đấu giải.

- *Bước 3:* Thành lập đội tuyển dự thi với số lượng vừa đủ bằng cách tuyển chọn từ đội tuyển tiềm năng.

Triển khai nội dung nghiên cứu cho các bước trên, luận án đã phân tích thành các nhiệm vụ nghiên cứu cụ thể như sau:

- *Nhiệm vụ 1:* Phân tích các yếu tố năng lực của những thí sinh thắng giải (winners) và những thí sinh không thắng giải (non-winners) trong những kỳ thi trước để mô hình hóa toán học mỗi người như một vector toán học trong không gian nhiều chiều, mỗi chiều biểu diễn như một biến năng lực. Vì tính chất và nội dung thi của các năm thường có thay đổi, luận án thu thập dữ liệu của những thí sinh đã dự thi học sinh giỏi cấp quốc gia 3 năm gần nhất để thiết kế dataset cho tiếp cận học máy của luận án.
- *Nhiệm vụ 2:* Xây dựng thuật toán và quy trình tuyển chọn học sinh vào đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học. Luận án áp dụng tiếp cận học máy với dataset được thành lập bởi nhiệm vụ 1 để thiết kế thuật toán tuyển chọn đội viên tiềm năng và thuật toán tuyển chọn đội viên đội tuyển học sinh giỏi môn Tin học.
- *Nhiệm vụ 3:* Xây dựng mô hình học khám phá có hướng dẫn trực quan với tiếp cận đánh giá hiệu năng của từng bài giảng đối với đội tuyển và giáo viên.

3. Đối tượng nghiên cứu của luận án

- Năng lực của các thí sinh thuộc các trường trung học phổ thông đã từng tham gia giải học sinh giỏi cấp quốc gia môn Tin học.
- Dữ liệu giáo dục (Educational Data) về năng lực học sinh.
- Các tiếp cận học máy (machine learning) để khai phá dữ liệu giáo dục (EDM: Educational Data Mining)
- Lý thuyết học tập và mô hình học khám phá (DL: Discovery Learning)
- Nguyên lý tiếp nhận và kiến tạo kiến thức của con người (học sinh).
- Nguyên lý tiếp nhận thông tin và kiến tạo kiến thức của con người (học sinh).

4. Phạm vi nghiên cứu của luận án

- Dữ liệu nghiên cứu liên quan thí sinh thắng giải và thí sinh không thắng giải cấp quốc gia môn Tin học của 3 năm trước gần nhất.

- Đặc trưng năng lực học sinh gồm những đặc trưng học tập (learning features) và những đặc trưng phi học tập (non-learning features) được phân tích thành những biến năng lực.
- Mô hình học khám phá có hướng dẫn để huấn luyện đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông.

5. Phương pháp nghiên cứu

- *Phương pháp phân tích - tổng hợp.* Phương pháp phân tích được áp dụng để phân tích năng lực của mỗi học sinh như là một vector năng lực nhiều biến, trong đó có những biến chịu tác động bởi yếu tố học tập ở trường (learning factor) và những biến chịu tác động bởi yếu tố phi học tập (non-learning factor) không do học ở trường mà liên quan cá nhân, gia đình học sinh và môi trường xã hội. Phương pháp phân tích và tổng hợp còn được áp dụng để cấu trúc bài hướng dẫn cho đội viên tiềm năng giải các bài toán theo mô hình học khám phá có hướng dẫn trực quan trong thời gian huấn luyện chuyên sâu.
- *Phương pháp mô hình hóa toán học.* Phương pháp mô hình hóa toán học được áp dụng để mô hình hóa năng lực của một học sinh như một vector năng lực trong không gian năng lực nhiều chiều, mỗi chiều của không gian năng lực biểu diễn một biến năng lực của học sinh.
- *Phương pháp học máy.* Luận án áp dụng phương pháp học máy có giám sát và đề xuất thuật toán Winner-domain để tìm ra những đội viên tiềm năng.
- *Phương pháp cây quyết định.* Luận án áp dụng phương pháp cây quyết định với thuật toán tương đồng cosin (Cosin Similarity) để quyết định chọn vào đội tuyển chính thức của trường từ những đội viên tiềm năng có khả năng thắng giải cao nhất.
- *Phương pháp toán học.* Phương pháp toán học được áp dụng để xây dựng tập dữ liệu dataset, định lượng những hồi đáp khảo sát định tính và giải các bài toán khi áp dụng các tiếp cận Winner-domain và Winner-cosin, cũng như tính hiệu năng của từng bài giảng huấn luyện đội.
- *Phương pháp lập trình.* Phần mềm Python được sử dụng để thiết lập chương trình phân chia ngẫu nhiên dataset, chương trình xây dựng các

Winner-domains, tìm Winner-domain tốt để tuyển ứng viên vào đội tuyển và tìm tương đồng cosin để chọn ứng viên vào đội tuyển.

- *Phương pháp khảo sát.* Phương pháp khảo sát được áp dụng để xây dựng bộ câu hỏi (Questionnaire) và Google Forms được sử dụng để thu thập dữ liệu năng lực của những thí sinh đã dự thi học sinh giỏi cấp quốc gia môn Tin học từ những năm trước, gồm những thí sinh đã thắng giải và những thí sinh không thắng giải.
- *Phương pháp thực nghiệm.* Phương pháp thực nghiệm được áp dụng cho thuật toán Winner-domain để tuyển chọn đội viên tiềm năng và thuật toán Winner-cosin với cây quyết định để chọn đội viên đội tuyển môn Tin học của một trường trung học phổ thông, dựa trên dataset đặc trưng năng lực của các winners và non-winners đã tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học các năm học 2020-2021, 2021-2022, 2022-2023, 2023-2024.

6. Các đóng góp của luận án

Luận án có mục tiêu là xây dựng quy trình tuyển chọn đội viên và huấn luyện đội tuyển môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia với kỳ vọng thắng được nhiều giải, đã áp dụng các tiếp cận học máy của khoa học máy tính để khai phá dữ liệu giáo dục và trực quan hóa dữ liệu để xây dựng phương pháp giáo dục mới. Kết quả nghiên cứu của luận án có ba đóng góp chính như sau:

6.1. Kết quả nghiên cứu nội dung 1

- *Tiếp cận biên soạn câu hỏi khảo sát định tính, đơn giản.* Luận án đã nghiên cứu và phân tích các yếu tố năng lực học sinh, đặc biệt là những yếu tố phi học tập, để biên soạn câu khảo sát bằng ngôn ngữ đơn giản, dễ hiểu với yêu cầu hỏi đáp định tính đơn giản để người được khảo sát có thể trả lời nhanh chóng bằng một dấu X cho mỗi câu khảo sát. Người được khảo sát không bị phiền phức về nội dung và không mất nhiều thời gian để hỏi đáp.
- *Tiếp cận thành lập các biến năng lực và định lượng các hỏi đáp khảo sát định tính.* Luận án đã đề xuất tiếp cận ma trận hỏi đáp và khối 3-D hỏi đáp để lượng hóa các hỏi đáp hoàn toàn định tính của những người được khảo sát. Tiếp cận ma trận hỏi đáp và khối 3-D hỏi đáp được áp dụng để

xác định các giá trị của những tập biến trong không gian năng lực và các giá trị biến năng lực của người được khảo sát, được biểu diễn như vector trong không gian năng lực nhiều chiều.

- *Dataset*. Luận án đã thành lập 2 datasets từ dữ liệu năng lực của một số thí sinh đã tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học những năm trước, gồm: Dataset 2024 được thành lập từ hồi đáp khảo sát năng lực của một số thí sinh kỳ thi năm học 2020-2021, 2021-2022, 2022-2023, và Dataset 2025 được thành lập từ hồi đáp khảo sát năng lực của một số thí sinh kỳ thi năm học 2021-2022, 2022-2023, 2023-2024. Trong các datasets, mỗi mẫu biểu diễn năng lực một thí sinh, mỗi trường biểu diễn một biến năng lực của các thí sinh, tương ứng với một câu khảo sát.

6.2. Kết quả nghiên cứu nội dung 2

Luận án đã đề xuất tiếp cận Winner-domain xác định vùng năng lực của các thí sinh thắng giải (winners) trong không gian năng lực nhiều chiều để xác định miền năng lực có tiềm năng thắng giải học sinh giỏi cấp quốc gia môn Tin học để chọn đội viên tiềm năng. Mô hình này được công bố tại các công trình [CT1] [CT2]. Luận án cũng đã xây dựng cây quyết định với thuật toán Winner-cosin (Cosin Similarity) để quyết định chọn vào đội tuyển những đội viên tiềm năng có khả năng thắng giải cao nhất.

6.3. Kết quả nghiên cứu nội dung 3

Trên cơ sở những nguyên lý học khám phá có hướng dẫn, luận án đã đề xuất mô hình học khám phá có hướng dẫn trực quan để huấn luyện đội viên tiềm năng của đội tuyển môn Tin học trước khi tham dự kỳ thi học sinh giỏi cấp quốc gia. Mô hình này được công bố tại các công trình [CT3] [CT4]. Mô hình học khám phá có hướng dẫn trực quan không chỉ đề cao vai trò trung tâm của người học trong hoạt động dạy và học tại cơ sở giáo dục mà còn áp dụng tiếp cận kỹ thuật trực quan để thiết kế bài giảng trên cơ sở nghiên cứu cơ chế tiếp nhận thông tin và kiến tạo kiến thức của con người. Luận án đã đề xuất tiếp cận đánh giá hiệu năng của mô hình học khám phá có hướng dẫn trực quan đối với từng bài giảng của thầy hướng dẫn.

7. Cấu trúc luận án

Luận án được thiết kế gồm phần mở đầu, bốn chương nội dung, và phần kết luận. Cuối luận án là danh mục các công trình của tác giả được công bố

quốc tế, danh mục các tài liệu tham khảo và phụ lục thực nghiệm. Cụ thể, nội dung của các phần được tóm tắt như sau:

Phần mở đầu. Luận án giới thiệu bối cảnh phát sinh đề tài với 3 thách thức để nghiên cứu. Sau khi xác định mục tiêu nghiên cứu, nội dung nghiên cứu, luận án định hướng nghiên cứu qua các phương pháp nghiên cứu. Cuối chương, luận án tóm tắt các đóng góp chính của luận án.

Chương 1. Tổng quan về khai phá dữ liệu giáo dục và học khám phá được liên quan

Luận án khảo sát tổng quan những tài liệu và công trình nghiên cứu có liên quan đến các chủ đề khoa học được áp dụng trong nghiên cứu của luận án gồm các tiếp cận học máy có giám sát áp dụng trong các bài toán khai phá dữ liệu giáo dục. Luận án nghiên cứu tổng quan về các phương pháp học và nguyên lý trực quan hóa, đặc biệt là mô hình học khám phá có hướng dẫn để đề xuất mô hình học khám phá có hướng dẫn trực quan áp dụng vào việc huấn luyện đội viên tiềm năng dự thi học sinh giỏi cấp quốc gia môn Tin học.

Chương 2. Dataset năng lực thí sinh dự thi học sinh giỏi

Dataset được thiết kế theo quy trình: phân tích đặc trưng năng lực $\rightarrow$ mô hình hóa toán học các đặc trưng năng lực $\rightarrow$ thiết kế bộ câu hỏi khảo sát $\rightarrow$ thu thập và tiền xử lý hồi đáp định tính $\rightarrow$ lượng hóa các hồi đáp $\rightarrow$ định lượng các tập biến năng lực của không gian năng lực và định lượng các biến năng lực của mỗi người được khảo sát $\rightarrow$ thiết lập dataset. Những câu hỏi khảo sát được thiết kế đơn giản, trả lời nhanh để dễ nhận được sự hợp tác của người được khảo sát. Luận án đã đề xuất tiếp cận ma trận hồi đáp và khối 3D hồi đáp để lượng hóa các biến năng lực từ các hồi đáp định tính của người được khảo sát.

Chương 3. Tiếp cận Winner-domain và Winner-cosin

Luận án định nghĩa miền không gian năng lực của những thí sinh từng thắng giải học sinh giỏi cấp quốc gia môn Tin học (Winner-domain), được xác định bởi trọng tâm, bán kính và góc hướng trong không gian năng lực nhiều chiều. Luận án phân hoạch ngẫu nhiên dataset nhiều lần để thiết lập nhiều Winner-domains, kiểm chứng bằng tiếp cận học máy có giám sát, và dựa vào các confusion matrix của chúng để chọn ra các Winner-domains tốt là miền của những đội viên tiềm năng. Luận án cũng đề xuất tiếp cận Winner-cosin và cây

quyết định để chọn đội viên đội tuyển từ những đội viên tiềm năng là những ứng viên có năng lực trong miền không gian của các Winner-domains tốt.

Chương 4. Tiếp cận học khám phá có hướng dẫn trực quan

Luận án nghiên cứu các phương pháp học, phân tích các ưu điểm của từng phương pháp để đề xuất mô hình học khám phá có hướng dẫn trực quan thích nghi với sự phát triển mạnh mẽ và rộng khắp của internet cùng các nguồn dữ liệu phong phú và công cụ thông minh. Luận án đã xây dựng quy trình thiết kế và đánh giá bài giảng áp dụng cho mô hình học khám phá có hướng dẫn trực quan và áp dụng thực tế vào việc huấn luyện đội viên tiềm năng dự thi học sinh giỏi cấp quốc gia môn Tin học.

Phần kết luận. Luận án tóm tắt các đóng góp chính của tác giả trong luận án và những nội dung sẽ phát triển tiếp theo.

CHƯƠNG 1. TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU GIÁO DỤC VÀ HỌC KHÁM PHÁ ĐƯỢC LIÊN QUAN

1.1. Giới thiệu

Công tác quản lý giáo dục tại một cơ sở giáo dục liên quan đến người học và phương pháp học tập, người dạy và phương pháp giảng dạy, đồng thời liên quan đến việc quản lý cơ sở vật chất phục vụ hoạt động giảng dạy, học tập và sinh hoạt chuyên môn. Những năm gần đây, sự phát triển nhanh chóng của khoa học và công nghệ đã tác động mạnh mẽ đến hoạt động giáo dục tại các cơ sở giáo dục, tác động trực tiếp vào quá trình dạy của người dạy và quá trình học của người học. Trong khi đó, nhiều cơ sở giáo dục tại các địa phương vẫn còn duy trì phương pháp giảng dạy với người dạy là trung tâm, người học là người tiếp thu kiến thức từ người dạy, khả năng và kỹ năng tự khám phá của người học chưa được phát huy cao.

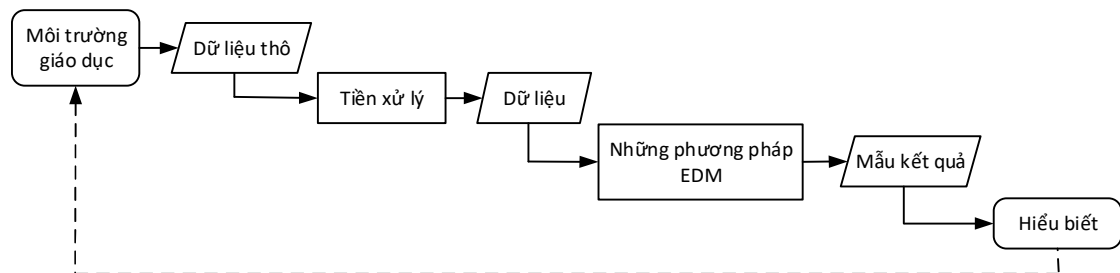
Hầu hết các cơ sở giáo dục đều có mục tiêu tốt đẹp, nhưng sự chuyển biến quá nhanh của công nghệ và của xã hội đang đặt ra cho những người quản lý giáo dục nhiều thách thức. Các cơ sở giáo dục không thể duy trì các phương pháp quản lý giảng dạy truyền thống trước sự phát triển của khoa học và công nghệ. Ngày nay, người quản lý cơ sở giáo dục không thể chờ để đánh giá hoạt động giáo dục thông qua kết quả điểm số của toàn lớp học mà cần phải theo dõi sự phát triển toàn diện của từng người học, đánh giá và kịp thời điều chỉnh quy trình giáo dục, đánh giá tác động của người dạy, của cơ sở vật chất, hoặc phát hiện và hỗ trợ người học trước những tác động của xã hội [1].

Tình trạng hoạt động của các cơ sở giáo dục hiện nay đang đặt ra cho người quản lý những vấn đề rất lớn trong hoạt động dạy và học. Những vấn đề phát sinh trong giáo dục, đặc biệt là vấn đề thành tích giáo dục của các trường không chỉ ảnh hưởng đến người học, đến công tác quản lý giáo dục, mà còn ảnh hưởng đến toàn xã hội. Tại các cơ sở giáo dục, tình trạng sa sút trong học tập và bỏ học của người học đang là vấn đề bức xúc đối với những người quản lý [2, 3]. Nhiều công trình nghiên cứu áp dụng khai phá dữ liệu giáo dục để dự báo những khuynh hướng kết quả học tập của một lớp học hoặc của một học sinh, dự báo những tình huống có thể xảy ra như tình trạng sa sút học tập của người học [1, 4], tình trạng người học bỏ học sau một, hai học kỳ đầu tiên để tìm nguyên nhân và biện pháp khắc phục hoặc cần phải can thiệp kịp thời.

1.2. Khai phá dữ liệu giáo dục

Trong lĩnh vực giáo dục, phân tích và dự đoán năng lực học sinh là chủ đề nghiên cứu đang được phát triển rộng rãi từ những công trình nghiên cứu dành cho các cơ sở giáo dục [3]. Khai phá dữ liệu giáo dục là một nghiên cứu xuyên ngành, tích hợp học máy với thống kê để trích xuất thông tin từ dữ liệu giáo dục, phân tích tâm lý sư phạm, tâm lý nhận biết [5], phân tích những yếu tố nội tại tác động đến kết quả học tập của người học để dự đoán năng lực và phân tích nguyên nhân bỏ học của người học do động lực cá nhân hoặc thói quen, do vấn đề xã hội hoặc tài chính, do thiếu tiến bộ trong học tập hoặc chuyển đổi ngành. Mô hình dự đoán áp dụng confusion matrix để đánh giá mức độ đúng/sai [2].

Khai phá dữ liệu giáo dục (Hình 1.1) hỗ trợ người quản lý cơ sở giáo dục và những người nghiên cứu giáo dục khám phá ra những ý nghĩa ẩn chứa trong dữ liệu giáo dục, dự đoán và nhận dạng người học bỏ trường hoặc ít quan tâm bài học [2]. Khai phá dữ liệu giáo dục có thể dự đoán những khuynh hướng về kết quả giáo dục, những tình huống có thể xảy ra như tình trạng sa sút học tập của người học [1, 4]. Khai phá dữ liệu giáo dục giúp người quản lý phân tích kết quả học tập của người học, hiểu được khuynh hướng và tình huống để kịp thời quyết định chính sách, điều chỉnh phương pháp giảng dạy, ngăn chặn khuynh hướng không tốt, cải thiện kết quả giáo dục.



Hình 1.1. Sơ đồ thực hiện khai phá dữ liệu giáo dục (EDM) [6]

Khoa học dữ liệu với nhiều tiếp cận khai phá dữ liệu được tích hợp vào quy trình dạy và học với mục tiêu chính là cải thiện quy trình [1]. Từ những kết quả khai phá dữ liệu, người quản lý giáo dục hiểu được hoạt động của cơ sở giáo dục để nhận ra khuynh hướng và các quan hệ để ra những quyết định cải thiện kết quả học tập của người học [1]. Những tiếp cận khai phá dữ liệu trích xuất ra thông tin hữu ích giúp người quản lý giáo dục có cơ sở vững chắc cho

những quyết định liên quan công việc xây dựng hoặc cải tiến môi trường giáo dục, điều chỉnh hoặc cải tiến phương pháp giảng dạy [7].

1.3. Dữ liệu giáo dục

Khai phá dữ liệu giáo dục và phân tích việc học của người học được áp dụng trên dữ liệu giáo dục [8]. Bài toán đặt ra là làm thế nào dữ liệu giáo dục có thể được nghiên cứu để làm tăng lợi ích cho giáo dục và cho việc học của người học [8]. Trong khi đó, các cơ sở giáo dục đang lưu trữ một khối lượng lớn dữ liệu liên quan đến hoạt động giáo dục [5], gồm dữ liệu liên quan đến người học như đăng ký học, dữ liệu cá nhân người học, hồ sơ giáo vụ của trường, kết quả kiểm tra bài, kết quả thi cuối kỳ, giữa kỳ, kết quả điểm danh, số lần người học đặt câu hỏi tại lớp học [5, 6] và những thành tích đặc biệt khác,... Dữ liệu liên quan đến cá nhân thầy, đến việc giảng dạy, môi trường học tập cũng là một phần của dữ liệu giáo dục.

Kết quả hoạt động của mỗi cơ sở giáo dục phụ thuộc vào 3 thành phần, người học, người dạy, môi trường giáo dục gồm cơ sở vật chất và công tác quản trị [5]. Đến nay, hầu hết các nghiên cứu khai phá dữ liệu giáo dục ghi nhận năng lực do học tập thể hiện trên kết quả học tập [9] và những yếu tố được sử dụng để dự báo năng lực người học là tuổi, giới tính, thời gian tham gia lớp học, thời gian truy cập internet, sở hữu máy tính, số lớp học đã tham dự [2]. Những yếu tố khác của người học như nền tảng xã hội, nhân chủng học, kiến thức, ứng xử, chất lượng và kết quả học tập trước, kiến thức có sẵn có liên quan với chủ đề bài giảng và những kỹ năng có trước đều tác động đến năng lực người học [3].

Dữ liệu giáo dục được thu thập từ người dạy, người học và người quản trị giáo dục, trong đó nhiều nhất là của người học. Ngoài ra, cơ sở giáo dục có lưu trữ một khối lượng lớn dữ liệu tiềm năng với những định dạng khác nhau, những mức độ tinh/thô khác nhau với nhiều mức độ phân cấp ý nghĩa. Thuộc tính giáo dục gồm: kiểu học, thất bại thi, trang trí phòng học, thời biểu và thời gian thi, e-learning, kết quả học tập, mục tiêu học tập, chỗ ngồi được sắp xếp, tính năng động của người học, web ngữ nghĩa trong giáo dục, học trong lớp, học cộng tác, tiểu sử người học, hệ thống giám hộ thông minh, khả năng tài chính của người học [5].

Tuy nhiên, đến nay, mô hình năng lực người học vẫn còn là thách thức và là một trong những chủ đề nghiên cứu trong khai phá dữ liệu giáo dục [3]. Những yếu tố năng lực như: giới tính, giáo dục và nghề nghiệp của cha mẹ, kỹ năng truyền thông, ứng xử xã hội, tình trạng kinh tế xã hội, thu nhập và chi tiêu của gia đình, thông tin cá nhân, thói quen học tập và thói quen dùng cell phone, tuổi, quan hệ xã hội và học tập, chiến lược học, cách học, thói quen và kỹ năng nhận biết, hoạt động thể dục, kỹ năng cảm xúc, thói quen uống rượu, cách ứng xử khi trả lời câu hỏi đột xuất, tình trạng học bổng, con đầu hoặc con duy nhất trong gia đình, nghề nghiệp của cha mẹ, hoạt động của hệ thống quản lý học [3].

Xử lý dữ liệu là một việc khó, phải chuyển dữ liệu vào dạng thích hợp cho từng bài toán, tập trung vào câu hỏi và chắc chắn [8]. Điểm đặc biệt của các nghiên cứu liên quan dữ liệu giáo dục là không có dataset dùng chung mà mỗi công trình nghiên cứu đều phải tự thu thập dữ liệu, tự thiết kế dataset theo yêu cầu của chủ đề, theo phạm vi của môi trường nghiên cứu với số lượng mẫu tùy theo môi trường giáo dục được khảo sát. Ví dụ từ 87 mẫu đến 162.030 mẫu [2, 10], hoặc trung bình khoảng 300 mẫu cho một môn học [11], hoặc từ 93 đến 404 mẫu [5], v.v...

1.4. Mục tiêu của những bài toán khai phá dữ liệu giáo dục

Những năm gần đây, chủ đề khai phá dữ liệu giáo dục thu hút nhiều tác giả phát triển các thuật toán học máy theo nhiều hướng mục tiêu khác nhau như phân tích và dự đoán năng lực học sinh, nâng cao chất lượng giáo dục của cơ sở giáo dục [12]. Khoa học máy tính với các tiếp cận khai phá dữ liệu đã xuyên ngành vào giáo dục, áp dụng khai phá dữ liệu trên tập dữ liệu giáo dục [12] nhằm tìm ra những ý nghĩa và những quan hệ tiềm ẩn trong dữ liệu giáo dục [2]. Tất cả các cơ sở giáo dục đều mong muốn thành công trong việc giáo dục người học [5]. Mục tiêu cuối cùng của những bài toán khai phá dữ liệu giáo dục là cải thiện kết quả học tập của từng người học, đánh giá kết quả dạy và học của cơ sở giáo dục để làm giảm chi phí giáo dục [13].

Đối với hoạt động giáo dục, những bài toán khai phá dữ liệu giáo dục giúp người quản lý hiểu được môi trường học tập, phát hiện năng lực tốt nhất của người dạy/người học, phát hiện những tiềm năng thất bại của người học để cải thiện môi trường giáo dục, phát triển phương pháp sư phạm [2]. Đối với khoa học máy tính, khai phá dữ liệu giáo dục áp dụng các mô hình học máy để

phân tích và nghiên cứu dữ liệu giáo dục [14, 15]. Các tiếp cận khai phá dữ liệu giáo dục thực hiện các bài toán phân tích dữ liệu giáo dục để dự báo năng lực học tập của người học, tình trạng bỏ học hoặc tình trạng ít quan tâm đến việc học của người học [2].

Để thực hiện mục tiêu cải thiện kết quả học tập của người học, nâng cao hiệu năng quản lý cơ sở giáo dục, các nghiên cứu dựa vào dữ liệu giáo dục thực hiện 2 nhóm bài toán là khai phá dữ liệu giáo dục và phân tích tình hình học tập của người học. Khai phá dữ liệu giáo dục liên quan đến việc phát triển những phương pháp nghiên cứu trên tập dữ liệu mới lạ của môi trường giáo dục, áp dụng các kỹ thuật khai phá dữ liệu trên tập dataset đặc biệt của giáo dục nhằm xử lý những câu hỏi quan trọng trong giáo dục. Do tính chất đặc biệt này, khai phá dữ liệu giáo dục vượt qua thách thức công nghệ bằng cách tìm những mẫu mới trong dữ liệu và phát triển thuật toán mới [8].

Phân tích học tập là đo lường, thu thập, phân tích và báo cáo về người học, ngữ cảnh của việc học tập để hiểu và tối ưu hóa việc học tập, môi trường học tập của người học. Một số nghiên cứu phân tích kết quả học tập của người học để nhận dạng những vấn đề trọng tâm trong dạy và học [3]. Có 3 thành phần chủ yếu trong phân tích học tập là dữ liệu, câu hỏi phân tích và những tác động. Phân tích học tập áp dụng các mô hình dự báo, tập trung nhận biết và hiểu được các thách thức trong giáo dục để làm quyết định nhằm cải thiện kết quả học tập của người học và kết quả hoạt động của cơ sở giáo dục bằng cách tích hợp kỹ thuật với sự phạm [8].

Phân tích và dự báo kết quả học tập của người học theo tiếp cận khai phá dữ liệu giáo dục là hai chủ đề chính được nhiều tác giả quan tâm với nhiều bài toán đa dạng [3, 6, 15, 16]. Đối với cơ sở giáo dục, những kết quả phân tích dữ liệu giáo dục và dự báo kết quả học tập của người học giúp người dạy và người quản lý tự đánh giá để quyết định điều chỉnh công việc của mình [8, 12, 13]. Những bài toán dự đoán được áp dụng với từng lớp học trước khi bắt đầu học kỳ và kiểm tra lại sau khi kết thúc học kỳ [3], tập trung vào những trường hợp sa sút của người học như bỏ học, sao nhãng việc học để có biện pháp hỗ trợ kịp thời [4, 17, 18].

Một yếu tố quan trọng để đánh giá chất lượng cơ sở giáo dục là năng lực do học tập của người học. Hiệu quả của cơ sở giáo dục được phản ánh bởi năng

lực do học tập của người học [9]. Sự thất bại hay thành công của người học là chỉ tiêu lớn cho cơ sở giáo dục. Sự thất bại của người học tác động xấu đến cá nhân vì có thể người học không có việc làm, đồng thời tác động xấu đến cơ sở giáo dục. Dự báo năng lực do học tập giúp cả người dạy và người học tìm học bổng hoặc workshop, đồng thời nhận dạng những người học cần sự quan tâm đặc biệt. Do đó, phân tích và dự báo năng lực do học tập của người học phải luôn được cập nhật, từ cập nhật dataset đến mô hình khai phá dữ liệu [3].

- *Dự báo sớm năng lực do học tập của người học.* Kết quả thi cuối học kỳ được dự báo để cảnh báo kịp thời những người học có thể có kết quả không tốt và khích lệ những người học có năng lực cao [2, 18, 19]. Một số tác giả khai phá dữ liệu giáo dục theo chủ đề dự báo năng lực do học tập của người học để phát triển các tiếp cận như ANN, DT, ET, RF, SVM, kNN, LR. ... trên cơ sở so sánh kết quả của các thuật toán áp dụng trên cùng một tập dữ liệu tự thu thập [2, 9, 13, 20].
- *Phân tích và dự báo năng lực do học tập của người học để can thiệp sớm* [2-4, 9, 10, 18, 21, 22]. Nhiều nghiên cứu áp dụng kỹ thuật khai phá dữ liệu hỗ trợ người dạy và người học. Kết quả nghiên cứu hỗ trợ người dạy thay đổi phương pháp và nội dung giảng dạy để nâng cao hiệu quả. Người học được phát hiện sớm kết quả học tập tốt có thể được tiếp cận những khoản tặng thưởng như học bổng, cơ hội thực tập hoặc hội thảo khoa học. Người học được phát hiện sớm nhu cầu cần được giúp đỡ có thể tránh được nguy cơ bỏ học. Kết quả phân tích năng lực hỗ trợ sinh viên đại học chọn môn học và làm kế hoạch học tập cho từng học kỳ.
- *Cải thiện môi trường và phương pháp giáo dục* [1, 3, 4, 6, 11, 13, 23]. Những kết quả phân tích dữ liệu giáo dục giúp cơ sở đào tạo làm chính sách đánh giá chương trình, tài nguyên học tập và phương pháp giảng dạy để cải thiện hiệu năng giáo dục của cơ sở giáo dục. Phân tích dữ liệu giáo dục cũng hỗ trợ phân tầng năng lực kết quả người học thành những lớp gồm những người học có năng lực tương đương, áp dụng các mô hình giảng dạy mới như học khám phá có hướng dẫn bằng phương pháp trực quan nhằm cải thiện hiệu quả giáo dục.

1.5. Những tiếp cận của khoa học máy tính được liên quan

Các tiếp cận chính của khoa học máy tính được sử dụng nhiều để khai phá dữ liệu giáo dục như support vector machines (SVM), decision tree, clustering, random forest, nearest neighbour, logistic regression, v.v... [2, 5]. Phương pháp có giám sát dự báo giá trị biến ra dựa trên biến vào, phương pháp không giám sát khai phá những mẫu dữ liệu không gán nhãn, tìm mẫu trong dataset không có biến ra để dự báo [9]. Những phương pháp khai phá dữ liệu giáo dục khác với những phương pháp chuẩn của khai phá dữ liệu do bản chất phân cấp và không độc lập của dữ liệu giáo dục [5].

Khai phá dữ liệu là một lĩnh vực của khoa học máy tính nhằm trích xuất thông tin hữu ích và tiềm năng từ tập dữ liệu lớn. Ngày nay, khai phá dữ liệu giáo dục là hướng nghiên cứu mới nổi, phát triển những phương pháp nghiên cứu loại dữ liệu mới lạ từ lĩnh vực giáo dục để hiểu tốt hơn về người học và môi trường người học [24]. Những thuật toán thường dùng trong khai phá dữ liệu giáo dục để dự báo là classification, regression, density estimation, clustering; hoặc để khai phá quan hệ là khai phá luật kết hợp, khai phá tương quan, khai phá mẫu nối tiếp theo mô hình dãy, khai phá dữ liệu nguyên nhân. Hầu hết các cơ sở giáo dục đều đang sở hữu nhiều dữ liệu về giáo dục và được cập nhật thường xuyên.

Khai phá dữ liệu giáo dục và phân tích học tập từ dữ liệu có cùng mục tiêu là cải thiện tình trạng giáo dục của cơ sở giáo dục và kết quả học tập của người học [8]. Khai phá dữ liệu giáo dục sử dụng kỹ thuật khai phá dữ liệu của khoa học máy tính trong môi trường giáo dục. Khai phá dữ liệu giáo dục phân tích, phân lớp, biểu diễn và trích xuất thông tin từ dữ liệu giáo dục [25]. Các tiếp cận khai phá dữ liệu được sử dụng để phân tích và dự báo trong lĩnh vực giáo dục [15]. Mục tiêu khai phá dữ liệu giáo dục là cải thiện và nâng cao chất lượng học của người học và cải thiện sự hiểu biết của người quản lý giáo dục về quy trình học [12].

Khai phá dữ liệu giáo dục có nhiều phương pháp tiếp cận, tùy theo những áp dụng khác nhau. Người dùng cuối của những kết quả khai phá dữ liệu là người học, người dạy, người quản trị và người nghiên cứu. Mô hình người học cung cấp sự hiểu biết tốt hơn về tình trạng học của người học, hệ hỗ trợ quyết định có thể giúp trực tiếp người dạy làm quyết định tốt hơn trong việc cải thiện

quy trình dạy và học [12]. Những tiếp cận thường được sử dụng để dự báo năng lực và tính chất người học là hồi quy và phân lớp, clustering phân nhóm người học theo động lực thấp, hành vi sai, bỏ học, thất bại học tập, v.v..., SVM và cây quyết định dự báo người học bỏ học [12].

Các phương pháp phân lớp (classification), hồi quy (regression), khai phá luật kết hợp, v.v... cũng được áp dụng cho các bài toán phân tích và dự đoán [3, 15]. Thời gian gần đây, các tác giả áp dụng thêm nhiều tiếp cận học máy trong các bài toán khai phá dữ liệu giáo dục để dự đoán năng lực do học tập của người học [2, 17]. Những tiếp cận học máy như học có giám sát, học không giám sát, gán nhãn, v.v.. được áp dụng trong các bài toán khai phá dữ liệu giáo dục gồm:

- Thuật toán phân lớp như: Support vector machine [26], Decision tree [27], K-nearest neighbour [6, 28], Naïve Bayes [17, 19];
- Thuật toán hồi quy [22] như hồi quy tuyến tính (Linear Regression), hồi quy logistics (Logistic Regression);
- Thuật toán phân cụm [29, 30] như K-means clustering, Hierarchical Clustering, DBSCAN [31]; v.v...

1.5.1. Tiếp cận Support Vector Machine (SVM)

SVM là một thuật toán học máy có giám sát áp dụng cho tập dữ liệu với các mẫu đã được gán nhãn, được sử dụng chủ yếu trong các bài toán phân lớp dữ liệu (data classification) và bài toán hồi quy (regression). SVM được thông dụng trong lĩnh vực khai phá dữ liệu vì nó có thể tạo ra các mô hình phân loại mạnh mẽ, từ những tập dữ liệu nhỏ đến những bộ dữ liệu phức tạp và không tuyến tính. Thuật toán SVM phân chia tập dữ liệu thành 2 lớp được ngăn cách bởi siêu phẳng (hyperplane) là đường thẳng trong không gian 2D, là mặt phẳng trong không gian 3D, hoặc là mặt nhiều chiều trong không gian nhiều chiều. [6, 11, 15-17, 22, 26].

Nội dung cơ bản cần phải thực hiện khi áp dụng SVM là tìm một siêu phẳng (hyperplane) tối ưu sao cho khoảng cách từ siêu phẳng này đến các điểm gần nhất của mỗi lớp (gọi là support vectors) được lớn nhất. Biên (margin) được định nghĩa là khoảng cách từ siêu phẳng đến các support vectors. Có hai loại SVM chính:

- *SVM tuyến tính (Linear SVM)*: SVM tuyến tính được sử dụng khi dữ liệu có thể được phân tách bằng một đường thẳng hoặc một siêu phẳng tuyến tính. Thuật toán SVM phải tìm kiếm một siêu phẳng tuyến tính để phân chia dữ liệu.
- *SVM phi tuyến (Non-linear SVM)*: SVM phi tuyến được sử dụng khi dữ liệu không thể phân tách tuyến tính. Trong trường hợp này, SVM sử dụng một hàm kernel để biến đổi không gian đặc trưng của dữ liệu, giúp biến các dữ liệu không tuyến tính thành tuyến tính.

Điểm mạnh chính của thuật toán SVM là có thể phân lớp dữ liệu có số chiều lớn và có khả năng xử lý dữ liệu phi tuyến. Điểm yếu chính của SVM là thời gian xử lý lâu đối với dữ liệu lớn, sai số phân lớp có thể tăng cao nếu phân chia tập dữ liệu nhiều lần thành nhiều lớp. Trong phân tích dữ liệu giáo dục, SVM là một công cụ mạnh để phân lớp dữ liệu và dự đoán hiệu quả [2, 13, 15, 26]. Bài toán dự đoán năng lực học sinh để chọn vào đội tiềm năng, về cơ bản có thể được xem là bài toán phân chia tập dữ liệu của những ứng viên thành 2 lớp, lớp những ứng viên có nhiều khả năng thắng giải và lớp những ứng viên ít có khả năng thắng giải. Do đó, tiếp cận SVM đã được luận án phát triển theo yêu cầu thực tế của bài toán là tuyển chọn một số ít ứng viên vào đội tiềm năng trong số rất nhiều ứng viên.

1.5.2. Tiếp cận cây quyết định

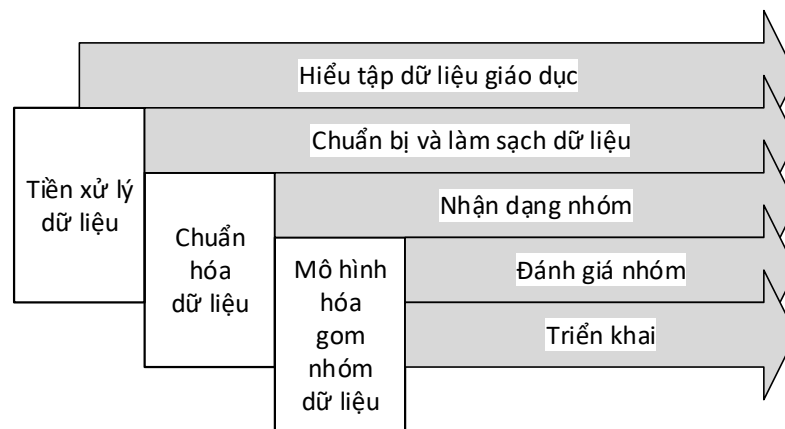
Một trong những kỹ thuật được dùng rộng rãi trong khai phá dữ liệu là những hệ thống phân lớp [32]. Cây quyết định là phương pháp học giám sát phi tham số được dùng để phân lớp dữ liệu [13, 27, 33]. Thuật toán cây quyết định là mô hình phân cấp dự báo biến ra bằng những quyết định nhị phân. Bắt đầu bằng cách thiết lập cây gồm một nút gốc và các nút lá để áp dụng thuật toán gán nhãn theo luật “if-then”. Tùy theo mức độ phức tạp về thuộc tính của tập mẫu cần phân hoạch, cây quyết định có thể có nhiều tầng. Cây quyết định dựa trên mô hình cây phân cấp, mô tả một chuỗi dữ liệu nhị phân bắt đầu từ gốc đến lá cuối cùng.

Bắt đầu phân hoạch tập mẫu với cây có 1 nút gốc và 2 nút lá, trong đó có 1 nút lá được gán thuộc tính 1. Đặt tập mẫu cần phân loại tại nút gốc và phân hoạch theo luật “if-then” để gán nhãn thuộc tính 1 cho những mẫu có cùng thuộc tính với nút lá 1, những mẫu còn lại đưa về nút lá cấp 1 còn lại. Tiếp tục

tương tự với nút lá cấp 1 còn lại này được gán là nút gốc mới để phân hoạch tập mẫu còn lại. Quy trình lặp lại tương tự cho đến khi tập đối tượng còn lại có cùng thuộc tính và được gán nhãn của nút lá cuối cùng.

1.5.3. Tiếp cận clustering (gom nhóm)

Thuật toán được sử dụng lâu bền nhất trong khai phá dữ liệu giáo dục (từ 1983) là clustering [3, 5]. Tiếp cận clustering được áp dụng để phân nhóm học sinh theo kết quả học tập giúp người dạy hiểu và xác định phương pháp giảng dạy thích hợp. Kỹ thuật clustering được phân lớp mở rộng vào 2 loại, phân cấp và riêng phần. Mục tiêu của phân lớp có giám sát là xác định nhãn cho data chưa được gán nhãn. Phương pháp clustering gom nhóm dữ liệu thành những lớp theo trọng tâm, graph, grid, mật độ, v.v... Quy trình khai phá dữ liệu giáo dục biến đổi dữ liệu thô từ các hệ thống giáo dục bằng cách gom nhóm thành thông tin hữu ích có tiềm năng tác động lớn đến nghiên cứu và thực hành giáo dục (Hình 1.2).



Hình 1.2. Quy trình gom nhóm dữ liệu giáo dục [5].

1.5.4. Tiếp cận phân cụm K-means

Thuật toán phân cụm K-means (K-means clustering) là một thuật toán phân cụm (clustering) được sử dụng rộng rãi trong các bài toán học máy không giám sát. Thuật toán K-means clustering được áp dụng phổ biến trong phân tích dữ liệu giáo dục để phân nhóm học sinh [5, 29, 30, 34-38]. Những kết quả phân tích giúp cơ sở giáo dục phân lớp học sinh theo năng lực, hoặc nhận dạng nhu cầu hỗ trợ của từng nhóm học sinh; giúp người quản lý giáo dục phát hiện sớm các vấn đề liên quan hiệu quả học tập, các yếu tố tiềm năng ảnh hưởng đến thành tích học tập của học sinh như mức độ tham gia lớp học, thời gian học tập,

hoặc kỹ năng làm bài thi; giúp người dạy tự đánh giá hiệu quả của phương pháp giảng dạy.

Về phương diện kỹ thuật, thuật toán K-means clustering chia dữ liệu thành K cụm sao cho các mẫu trong cùng một cụm có tính tương đồng cao về đặc trưng năng lực và ít tương đồng với các cụm khác. Mỗi cụm có một điểm trọng tâm gọi là centroid đại diện cho cụm. Thuật toán bắt đầu bằng cách chọn ngẫu nhiên K điểm làm trọng tâm K cụm ban đầu. Sau đó, mỗi điểm dữ liệu được gán vào cụm có trọng tâm gần nhất. Trọng tâm của mỗi cụm được tính lại bằng cách lấy trung bình các điểm trong cụm đó và tiếp tục lặp lại quy trình này cho đến khi các trọng tâm cụm không thay đổi nhiều hoặc đạt đến số lần lặp tối đa.

1.5.5. Xử lý dữ liệu

Dữ liệu bị thiếu là một vấn đề phổ biến trong quá trình thu thập dữ liệu. Xử lý giá trị thiếu là điền vào hoặc loại bỏ các giá trị bị khuyết nhằm đảm bảo dữ liệu đầy đủ và phù hợp cho việc phân tích hoặc huấn luyện mô hình [39]. Một số phương pháp phổ biến để xử lý dữ liệu thiếu:

- Loại bỏ các quan sát có giá trị thiếu: Đây là phương pháp đơn giản nhất, trong đó các dòng dữ liệu chứa giá trị thiếu được loại bỏ khỏi tập dữ liệu.

- Loại bỏ các biến có tỷ lệ giá trị thiếu cao: Trong trường hợp một biến có tỷ lệ thiếu lớn hoặc không mang lại thông tin quan trọng cho phân tích, việc loại bỏ biến này có thể được xem xét nhằm giảm nhiễu và cải thiện hiệu suất mô hình.

- Thay thế giá trị thiếu bằng các thống kê đặc trưng: Điền giá trị thiếu bằng các thống kê như trung bình (mean), trung vị (median), hoặc giá trị xuất hiện nhiều nhất (mode), được tính trên toàn bộ tập dữ liệu hoặc trong từng nhóm phân loại cụ thể.

- Sử dụng kỹ thuật hồi quy hoặc mô hình dự đoán: Các kỹ thuật hồi quy hoặc mô hình dự đoán có thể được sử dụng để ước lượng giá trị thiếu dựa trên các biến khác có sẵn.

- Sử dụng phương pháp đặc biệt cho dữ liệu bị thiếu: Một số trường hợp đặc thù, có thể sử dụng các phương pháp chuyên sâu hơn như Hot-deck imputation, Multiple imputation hoặc Interpolation, tùy thuộc vào loại dữ liệu và yêu cầu của bài toán.

1.6. Phương pháp giáo dục

Ngày nay, lý thuyết học tập được áp dụng trong môi trường giáo dục để cải thiện khả năng người học và kỹ năng người dạy [5], phương pháp giáo dục đang chuyển sang các mô hình người học là trung tâm, người dạy không “đọc sách giùm bạn” mà hướng dẫn người học tự khám phá tri thức mới. Mô hình học khám phá có hướng dẫn triển khai áp dụng cho cơ sở giáo dục gặp các thách thức: (1) kiến thức sẵn có của người học trong lớp học không tương đồng; (2) người thầy hướng dẫn được đòi hỏi có tiếp cận phù hợp để người học nhanh chóng tiếp thu những phương pháp được hướng dẫn. Luận án này tiếp cận các kỹ thuật trực quan hóa dữ liệu từ khoa học máy tính để xây dựng mô hình học khám phá có hướng dẫn trực quan [40-43].

1.7. Nguyên lý học tập

Học tập xảy ra khi người học làm việc hay học để xử lý những vấn đề phức tạp hoặc những bài toán với nội dung hoặc những vấn đề trong vùng phát triển gần trên mức nhận biết (ZPD: Zone of Proximal Development), gọi là vùng phát triển tiềm năng, của người học. Vùng phát triển tiềm năng phụ thuộc vào khả năng hiện có của người học và vấn đề đặt ra hoặc bài toán mà người học phải đối mặt. Vùng phát triển tiềm năng của người học được xác định thông qua cách tự giải bài toán một cách độc lập bằng khả năng thực đang có tới mức phát triển được xác định thông qua cách giải bài toán với sự hướng dẫn của người có khả năng cao hơn [44].

Nếu người học có thể giải bài toán một cách độc lập thì người học sẵn sàng trong mức khả năng thực (AAL: Actual Ability Level). Nếu người học có thể giải bài toán trong sự có mặt của những người hiểu bài toán tốt hơn thì người học đang trong mức khả năng tiềm năng (PAL: Potential Ability Level). Người dạy nên đặt bài toán trong vùng giữa AAL và PAL hoặc trong vùng tiệm cận nhận biết của người học [44]. Bài toán có thể được đặt ra như những câu hỏi xuất phát từ đời sống thực mà mức độ khó của bài toán vẫn trong vùng tiệm cận tư duy của người học nhưng không có thuật toán hay thủ tục cho người học áp dụng được ngay [44].

Để nâng cao khả năng từ mức khả năng thực AAL lên mức khả năng tiềm năng PAL, người học phải tiếp cận thông tin. Người học là người dùng thông tin, tiếp xúc với thông tin trong quy trình học tập, thông tin từ người dạy,

thông tin từ tài liệu học tập được in ấn, thông tin từ tài liệu điện tử, v.v.. [45]. Do đó, người học phải có kỹ năng nhận dạng nhu cầu thông tin, kỹ năng phát triển chiến lược để tìm kiếm thông tin, tìm và truy cập thông tin, tổ chức đánh giá và dùng thông tin chính xác, hiệu quả, hiệu năng và truyền thông thông tin tốt như là một vấn đề có thể được giải tốt [45].

1.8. Học khám phá

1.8.1. Khái niệm

Học khám phá là một mô hình giáo dục mới, trong đó khả năng và sự tự tin của người học được thúc đẩy để tự tìm ra, tự cấu trúc kiến thức mới để lưu trữ trong bộ nhớ [46]. Nhiều tác giả thảo luận rằng phương pháp học của người học tại các cơ sở giáo dục bây giờ đang phát triển từ học có tính cách mô tả sang học khám phá [42, 47], trong đó vị trí trung tâm của những hoạt động giáo dục được chuyển từ người dạy sang người học [48] để nâng cao vai trò của người học trong hoạt động học tập [49]. Nhiều tác giả cho rằng những tính chất cơ bản của mô hình học khám phá là học tích cực, học có ý nghĩa, tự lực để có kết quả [47, 50-52].

Mô hình học khám phá phát triển tính tích cực của người học trong học tập bằng cách tạo môi trường cho người học tự khám phá, tự tìm hiểu, tự nghiên cứu để kết quả học tập được lưu trữ lâu dài trong ký ức, không dễ quên [53, 54]. Trong quy trình học khám phá, người học giữ vai trò tích cực trong thực hiện, tích hợp và tổng quát hóa kiến thức sau khi tự tìm hiểu, tự nghiên cứu kiến thức [55], như vậy cần thiết kế quy trình và môi trường học sao cho người học tham gia tích cực vào quy trình học [56]. Trong mô hình học khám phá, người học có nhiệm vụ khám phá thông tin cuối cùng [56], được khuyến khích quan sát các hoạt động của chính bản thân mình để kết luận [57].

1.8.2. Mô hình học khám phá

Trong mô hình học khám phá, người học tự xây dựng kiến thức bằng cách thực nghiệm để tìm nguyên lý từ thực nghiệm [46]. Nhiệm vụ chính của người học là tìm các tính chất của môn học hoặc của bài toán, những tính chất này người học không được nghe hoặc được người dạy trình bày trực tiếp mà phải tự khám phá thông qua thực nghiệm và suy luận [58]. Người học tự khám phá những gì cần học, không nhất thiết những nội dung khám phá được trong

hoạt động học tập này là mới, mà có thể khám phá ra lời giải đã từng gặp trước đó trong thời gian học tập [59].

Học khám phá đòi hỏi những quy trình quy nạp trong đó thông tin và kiến thức được phát sinh từ những thực nghiệm được thực hiện [60]. Người học xử lý một môn học hoặc bài toán bằng cách dẫn các thực nghiệm, nghĩa là bằng cách xử lý các biến và các thông số trong môn học đó hoặc bài toán đó và quan sát kết quả của những cách xử lý này [60]. Nói cách khác, người học tự rút ra kết luận quy nạp rồi chứng thực kết luận [61]. Những người học biết tự thiết kế và phân tích trường hợp sẽ hiểu khái niệm tốt hơn người học chỉ tập trung vào tài liệu học tập [56].

1.8.3. Phương pháp học khám phá

Phương pháp học khám phá là phương pháp học để kiến thức người học được phát triển từ không biết đến biết bởi chính người học [57]. Phương pháp học khám phá được dùng như một giải pháp khuyến khích người học phải trực tiếp trong quy trình học, bắt đầu từ hoạt động định hướng, hình thành bài toán, đề xuất giả thuyết, thu thập dữ liệu, kiểm tra giả thuyết và hình thành kết luận [62]. Trong quy trình học khám phá, người học phải tham gia tích cực [57], người học tự giải bài toán, tự hình thành câu hỏi, tự trả lời câu hỏi, chủ động thảo luận, giải thích, tranh luận và tích cực động não khi ở lớp [63]. Phương pháp học khám phá giúp người học tự kiến tạo kiến thức bằng những khả năng và kỹ năng theo một quy trình tích cực [64].

Phương pháp học khám phá phải khuyến khích và đáp ứng yêu cầu của những người ham hiểu biết, muốn tìm hiểu để tham gia tích cực vào quy trình học và tham gia trực tiếp vào việc khám phá kiến thức [57]. Quy trình học khám phá cung cấp hoạt động và trải nghiệm để thu hút sự chú ý của người học, cho phép hình thành các khái niệm trừu tượng có ý nghĩa với những hoạt động hiện thực [62]. Phương pháp học khám phá đáp ứng yêu cầu nhân lực của các tổ chức tuyển dụng. Các tổ chức nghề nghiệp đánh giá năng lực người học dựa vào hiệu năng làm việc chuyên môn, tự lực giải quyết công việc, biết gắn kết với các đồng nghiệp và trực tiếp tìm hiểu công việc [50].

- *Phương pháp học khám phá nhấn mạnh kỹ năng lập luận.* Phương pháp học khám phá nhấn mạnh kỹ năng lập luận, đề cao phát triển kỹ năng lập luận của người học khi thực hiện quy trình học khám phá [58], thúc đẩy

người học nâng cao kỹ năng lập luận, phân tích các trường hợp [56] để dễ thành công hơn trong quy trình giải các bài toán hoặc dự án. Người học có kỹ năng lập luận, phân tích các trường hợp tốt [56] có nhiều triển vọng đạt kết quả học tập tốt [61]. Đặc biệt, trong toán học, người học có kỹ năng lập luận thấp sẽ có kết quả học tập không tốt [61].

- *Phương pháp học khám phá nhấn mạnh kỹ năng tư duy phản biện.* Phương pháp học khám phá nhấn mạnh tư duy phản biện, đề cao phát triển kỹ năng của người học về tư duy phản biện và phân tích các trường hợp [56, 58] để nhận dạng và phát biểu bài toán, xác thực và chứng minh kết quả, rút ra kết luận và tổng quát hóa bài toán [57].
- *Phương pháp học khám phá theo nhóm cộng tác.* Thực hiện phương pháp học khám phá theo nhóm cộng tác, người học làm việc theo nhóm về những bài toán và dự án [63]. Phương pháp học khám phá theo nhóm cộng tác khuyến khích người học thảo luận nhóm, chia sẻ thông tin và kiến thức, khuyến khích người học dũng cảm phát biểu ý kiến hoặc truyền thông về một vấn đề [62]. Phương pháp học khám phá theo nhóm cộng tác áp dụng phù hợp cho bài toán đòi hỏi mức khả năng cần thiết PAL cao hơn mức khả năng thực AAL của người học. Phương pháp học khám phá theo nhóm cộng tác tập trung phát triển kỹ năng truyền thông [58].
- *Phương pháp học khám phá bằng thực nghiệm hoặc bằng mô hình.* Phương pháp học khám phá bằng thực nghiệm hoặc bằng mô hình tập trung phát triển kiến thức [58], kiến tạo kiến thức mới trên nền tảng kiến thức cũ đã có. Khi thực hiện quy trình học khám phá, người học tự thiết kế [56] thực nghiệm những hiểu biết đã có trước để củng cố kiến thức và có thể kiểm chứng những khái niệm phát sinh nhằm gia tăng kết quả học tập [64], hoặc mô phỏng các vấn đề của khoa học, kỹ thuật, hoặc các tình huống xã hội để phân tích trường hợp [56] tìm ra những khái niệm chưa biết thông qua khảo sát các mô hình [54]. Những người học thực nghiệm phát sinh giả thiết, thiết kế thực nghiệm, dự đoán kết quả, suy ra dữ liệu và khảo sát lại giả thiết để kiến tạo kiến thức về lĩnh vực đang thực nghiệm.
- *Phương pháp học khám phá dựa trên bài toán.* Phương pháp học khám phá dựa trên bài toán cổ vũ tính tự tin vào năng lực bản thân của người học [53]. Thực hiện phương pháp học khám phá dựa trên bài toán, người

học được phát triển khả năng tự nghiên cứu, tự lập luận hệ thống để mô phỏng, nhận dạng và phát biểu bài toán; người học được khơi gợi tính tò mò và khả năng lập luận để thiết kế thu thập dữ liệu; người học được kích thích tư duy logic để phân tích trường hợp [56] và xử lý dữ liệu.

- *Phương pháp học khám phá dựa trên dự án.* Phương pháp học khám phá dựa trên dự án là một phương pháp học quy nạp, người học tự trình bày những thách thức và tự tra cứu tài liệu trong ngữ cảnh nhân mạnh thách thức [63]. Tương tự học khám phá dựa trên bài toán, khi áp dụng phương pháp học khám phá dựa trên dự án, người học tự nghiên cứu để thiết kế mô hình [56] mô phỏng, nhận dạng và phát biểu mục tiêu dự án; người học vận dụng tư duy logic để phân tích thông tin và tính tỉ mỉ để thu thập dữ liệu; người học áp dụng tư duy logic để tự phân tích mục tiêu dự án và xử lý dữ liệu để đề xuất các giải pháp và quy trình thực hiện dự án; người học áp dụng tư duy phản biện, tư duy logic và tư duy sáng tạo để chứng minh kết quả, phân tích trường hợp [56] để rút ra kết luận, đề xuất quy trình thực hiện và giải pháp khắc phục những khiếm khuyết của dự án [63].

1.8.4. Tính chất của học khám phá

1.8.4.1. Những thuộc tính của học khám phá

- Người học giữ vai trò tích cực trong việc học, chủ động theo nhịp độ riêng để tự khám phá qua thực nghiệm, tư duy, nghiên cứu và đối sánh với kiến thức có sẵn [50].
- Người học tự cấu trúc kiến thức theo thông tin mới và dữ liệu thu thập được trong môi trường học để tích hợp và xây dựng kiến thức mới trên nền kiến thức đang có [50, 60].

1.8.4.2. Học khám phá khác với học truyền thống

- So với mô hình học truyền thống, người học theo mô hình học khám phá tích cực hơn vì người học được khuyến khích học một cách độc lập và là người trực tiếp khám phá những khái niệm và những nguyên lý khác nhau thông qua việc giải bài toán hoặc kết quả từ những đối tượng văn hóa khác nhau [44, 50].
- So với mô hình học truyền thống, mô hình học khám phá giúp người học hiểu sâu hơn các khái niệm và các kiến thức được khám phá từ quy trình học bởi vì người học được trực tiếp trong quy trình khám phá ra khái

niệm chưa biết trước và được yêu cầu thực hiện ý tưởng, tìm các quan hệ để hình thành khái niệm [50].

- So với mô hình học truyền thống, học khám phá làm tăng sự tiếp thu của người học vì người học được rèn luyện kỹ năng nhiều hơn học sự kiện. Người học theo mô hình học khám phá có kỹ năng tư duy sáng tạo cao hơn người học theo mô hình học truyền thống và kết quả học tập của người học trong môi trường học khám phá cao hơn kết quả học tập bằng phương pháp truyền thống [50, 54, 62].
- So với mô hình học truyền thống, học khám phá dựa vào quy trình hơn là dựa vào sự kiện. Trong học khám phá, khái niệm không được trình bày, mà người học tự đặt ra câu hỏi để tìm ra thông tin, khái niệm, nguyên lý chưa biết trước, tự giải quyết chủ đề theo cách của những nhà khoa học [44, 48-50].

1.8.4.3. Lợi ích của học khám phá

- *Học khám phá thúc đẩy việc học của người học.* Học khám phá cho phép người học tìm thông tin để thỏa mãn sự ham muốn tự nhiên của mình [50]. Trong quy trình học khám phá, người học kích hoạt tư duy kiến thức trước rồi ánh xạ vào bài toán được chỉ định để biểu diễn bài toán và phát sinh giả thiết thích hợp để hiểu [65].
- *Học khám phá phát triển kỹ năng tư duy.* Mô hình học khám phá cung cấp cơ hội cho người học phát triển sự sáng tạo, người học tự nghiên cứu, tự tìm hiểu thông tin một cách độc lập thông qua phương pháp, mô hình và môi trường để cải thiện kỹ năng tư duy sáng tạo [62]. Trong khi tự mình áp dụng quy trình học khám phá để tìm ra những khái niệm chưa được biết trước, người học cũng được phát triển tư duy phản biện [66].
- *Học khám phá phát triển tư duy sáng tạo.* Học khám phá khuyến khích tư duy sáng tạo, cải thiện và phát triển kỹ năng tư duy sáng tạo của người học [46, 62, 67, 68]. Khả năng tư duy sáng tạo của người học theo mô hình học khám phá gia tăng có ý nghĩa, đặc biệt là tư duy sáng tạo trong các môn học vật lý và toán.
- *Học khám phá cải thiện kỹ năng nhận dạng và hiểu nhu cầu thông tin.* Người học khám phá được rèn luyện tốt kỹ năng nhận dạng và hiểu nhu cầu thông tin, phát triển chiến lược tìm kiếm thông tin, tìm và truy cập thông tin, tổ chức đánh

giá và dùng thông tin một cách chính xác, hiệu quả và hiệu năng. Truyền thông thông tin tốt để một vấn đề có thể được giải quyết tốt [45]. Kỹ năng nhận biết, kỹ năng tư duy, kỹ năng phản biện, kỹ năng sáng tạo [69], kỹ năng đọc viết [45] là những kỹ năng cần có của người đề xuất và viết dự án nghiên cứu.

- *Học khám phá nâng cao tính hệ thống và tính logic.* Học khám phá nâng cao tính hệ thống và tính logic của người học vì quy trình học khám phá bao gồm lập luận khoa học đúng, xử lý các biến có tính hệ thống, thiết kế định tính và cài đặt thực nghiệm [65].
- *Học khám phá nâng cao kết quả học tập.* Trong hoạt động học khám phá, người học nhớ nhiều nội dung được học hơn là trong hoạt động học truyền thống [48, 50, 63, 64, 68, 70]. Học khám phá cải thiện hoạt động học tập và kết quả học khoa học, học toán học, học ngôn ngữ bao gồm cả văn phạm và viết, và học kỹ thuật.

1.8.4.4. Tư duy sáng tạo của người học theo mô hình học khám phá

Học khám phá phát triển tư duy sáng tạo của người học [46, 62]. Tư duy sáng tạo được hiểu là tư duy theo một hướng khác chỉ được trả lời hoặc đánh giá bằng kết quả thực. Kỹ năng tư duy sáng tạo là kỹ năng nhận biết để đề xuất và phát triển ý tưởng mới từ những ý tưởng đã có trước và kỹ năng giải bài toán theo hướng khác. Khả năng tư duy sáng tạo là kỹ năng lập luận để phát sinh ý tưởng, thực hiện cái gì mới, gốc, khác thường, có giá trị, trừu tượng biểu lộ trong dạng một hay nhiều ý tưởng để tìm ý nghĩa và lời giải cho bài toán một cách sáng tạo mới.

1.9. Học khám phá có hướng dẫn (Guided Discovery Learning)

Trong mô hình học khám phá, người hướng dẫn cần phải lường trước những trạng thái tiêu cực như: quá tải nhận biết, hiểu lầm tiềm năng, phải nhận dạng để tương thích bài toán với năng lực người học [44]. Theo lý thuyết học, mô hình học khám phá là mô hình học nhận biết, người dạy cần tập trung vào việc xây dựng kiến thức kế tiếp trong ký ức người học với thông tin mới được tiếp thu [47]. Do đó, khi thực hiện quy trình học khám phá người dạy cần hướng dẫn người học phương pháp để tìm ra khái niệm hoặc lý thuyết phù hợp tài liệu của chủ đề học [46].

Học khám phá là quy trình nhận biết qua 2 không gian, không gian giả thiết và không gian thực nghiệm. Không gian giả thiết là không gian tìm kiếm, chứa tất cả những luật mô tả hiện tượng có thể quan sát trong lĩnh vực chuyên môn. Không gian thực nghiệm gồm tất cả những thực nghiệm có thể thực hiện trong lĩnh vực chuyên môn. Trong các không gian đó, người học cần được hỗ trợ từ môi trường học để có thể tư duy phản biện và sáng tạo khi giải quyết vấn đề đang gặp, người dạy phải xây dựng các bước hướng dẫn học thuật cụ thể rõ ràng để người học dễ thực hiện [53, 71, 72].

Trong học khám phá, nội dung học không được phân phát trực tiếp bởi người dạy mà được tìm thấy bởi người học trong quy trình học thông qua tình huống vấn đề được người dạy sắp xếp, và người học vẫn cần sự trợ giúp và đây là cái gốc của kết quả học tập [48, 73]. Mức trợ giúp tốt nhất là giữa mức hướng dẫn trực tiếp như phương pháp truyền thống và mức học khám phá không hướng dẫn, nghĩa là mức hướng dẫn cần thiết trong học khám phá có hướng dẫn là từ mức tác động để thay đổi kiến thức người học dựa trên mức kiến thức có sẵn của người học [73]. Học khám phá có hướng dẫn tập trung sự truyền kiến thức giữa người dạy, người học và tài liệu học tập [74] trong một môi trường giáo dục phù hợp để hỗ trợ người học phát huy tính tự lực và các kỹ năng để kiến trúc kiến thức từ mức đang có.

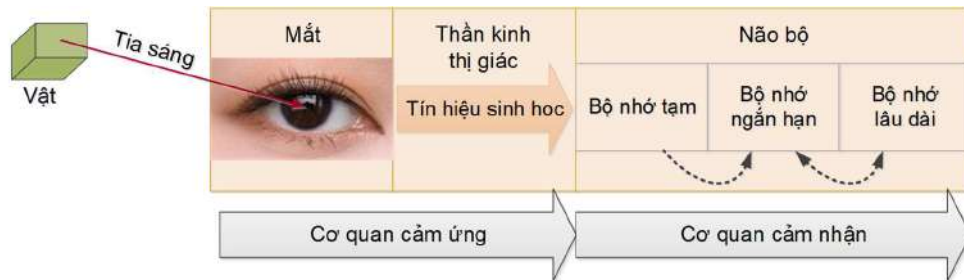
Tóm lại, học khám phá có hướng dẫn hiệu quả hơn học khám phá không hướng dẫn [46]. Người học khám phá có hướng dẫn đạt kết quả tốt hơn và cải thiện được khả năng giải bài toán cao hơn so với người học khám phá không hướng dẫn [44]. Áp dụng học khám phá không hướng dẫn để học thuật toán thì không hiệu quả vì phương pháp học khám phá dựa vào bài toán cần sự thay đổi trong cách tiếp cận để giải bài toán, người học cần được hướng dẫn để tương thích kiến thức có sẵn của mình với bài toán theo quy trình khám phá [47, 71]. Do đó, các chuyên gia giáo dục học không khuyến nghị áp dụng học khám phá không hướng dẫn [63].

1.10. Trực quan hóa

Con người có được thông tin/tri thức thông qua cơ chế tiếp nhận dữ liệu và giải mã dữ liệu, biến dữ liệu thành thông tin/tri thức. Con người tiếp nhận dữ liệu qua 5 kênh giác quan: kênh thị giác với các cơ quan và thần kinh của mắt, kênh thính giác với các cơ quan và thần kinh của tai, kênh xúc giác với

các cơ quan và thần kinh của da, kênh vị giác với các cơ quan và thần kinh của lưỡi, kênh khứu giác với các cơ quan và thần kinh của mũi [1]. Trong đó, thị giác là kênh rộng nhất có khả năng cao nhất để thu thập và giải mã được nhiều dữ liệu thành thông tin/tri thức.

Dữ liệu thu thập được từ các cơ quan giác quan được chuyển thành tín hiệu sinh học để các hệ thần kinh của kênh dẫn đến bộ nhớ tạm của não bộ (Hình 1.3) [1]. Tại bộ nhớ tạm, dữ liệu được phân loại theo ý nghĩa cần thiết hoặc không cần thiết, dữ liệu không cần thiết được loại bỏ, dữ liệu cần thiết được chuyển sang bộ nhớ ngắn hạn để chuyển thành thông tin/tri thức. Tại bộ nhớ ngắn hạn, dữ liệu từ các dạng khác nhau được chuyển về dạng hình ảnh 3D để đối sánh với dữ liệu dạng hình ảnh 3D đã được lưu trữ từ trước trong bộ nhớ lâu dài [1]. Những dữ liệu mới thu thập được chuyển sang hình ảnh 3D, nếu giống với dữ liệu dạng 3D có trước trong bộ nhớ lâu dài, thì dữ liệu mới này được hiểu.



Hình 1.3. Hệ thống thị giác thu thập, biến dữ liệu thành thông tin/tri thức [1]

Tri thức của con người được lưu trữ trong bộ nhớ lâu dài dạng hình ảnh 3D. Con người hiểu được dữ liệu mới nhận được bằng cách đối sánh dữ liệu mới này với tri thức dạng 3D đã lưu trữ trước trong bộ nhớ lâu dài. Khi tiếp nhận dữ liệu mới, bộ nhớ ngắn hạn của con người phải chuyển dữ liệu từ các dạng khác nhau thành dạng hình ảnh 3D. Nếu bộ nhớ ngắn hạn nhận được dữ liệu dạng hình ảnh 3D thì không cần phải tải năng lực bộ nhớ để chuyển đổi từ dạng bất kỳ thành dạng 3D. Như vậy, sự tiếp nhận kiến thức mới được dễ dàng và cần ít thời gian xử lý hơn.

CHƯƠNG 2. DATASET

NĂNG LỰC THÍ SINH DỰ THI HỌC SINH GIỎI

2.1. Giới thiệu

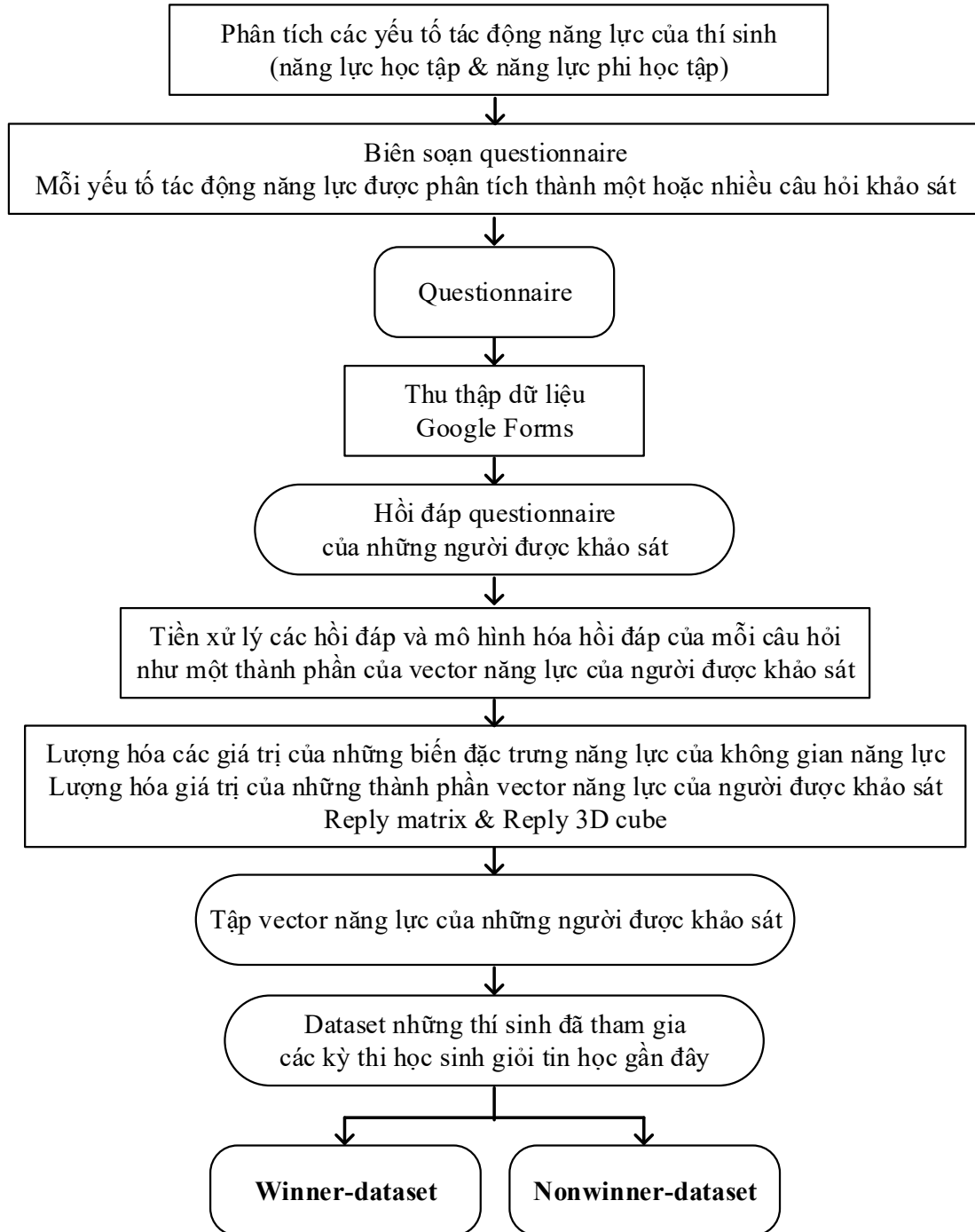
Dataset sử dụng trong một nghiên cứu về dữ liệu giáo dục có những tính chất liên quan đến mục tiêu nghiên cứu, môi trường giáo dục, đối tượng nghiên cứu và người dùng kết quả nghiên cứu.

- *Mục tiêu nghiên cứu:* Những bài toán nghiên cứu về dữ liệu giáo dục hiện nay tập trung vào các chủ đề phân tích việc học và dự báo tình hình học tập. Phân tích việc học để nhận biết tình hình dạy và học, phương pháp giảng dạy, hiệu quả của công tác quản trị, khả năng đáp ứng của cơ sở vật chất giáo dục như tài liệu học tập, trường lớp và trang thiết bị tiện ích cho nhu cầu người học. Dự báo tình hình học tập gồm dự báo kết quả hoạt động dạy và học của cơ sở giáo dục, hoặc dự báo tình hình học tập của từng lớp học, từng người học.
- *Môi trường giáo dục:* Môi trường giáo dục là nơi phát sinh bài toán nghiên cứu như dữ liệu về một trường hoặc lớp cụ thể.
- *Đối tượng nghiên cứu:* Đối tượng nghiên cứu trong môi trường giáo dục gồm người học, người dạy, chương trình học, tài liệu học tập, và cơ sở vật chất với những tiện ích cần thiết.
- *Người dùng kết quả nghiên cứu:* Người dùng kết quả nghiên cứu là người đặt ra yêu cầu nghiên cứu hoặc sử dụng kết quả nghiên cứu cho mục tiêu cụ thể, có thể là người quản trị cơ sở giáo dục, người dạy, hoặc người học.

Các chủ đề nghiên cứu sử dụng dữ liệu giáo dục nhằm mục tiêu giải quyết một bài toán cụ thể tại một cơ sở giáo dục cụ thể. Do đó, những đặc trưng và tính chất của dữ liệu phải phù hợp với mục tiêu của bài toán, môi trường giáo dục, đối tượng nghiên cứu và yêu cầu của người dùng kết quả. Do đó, các nghiên cứu liên quan dữ liệu giáo dục không thể dùng chung dataset. Nói cách khác, mỗi chủ đề nghiên cứu tại một cơ sở giáo dục phải thiết kế xây dựng dataset riêng với số mẫu và biến dữ liệu phụ thuộc vào bài toán nghiên cứu và môi trường nghiên cứu.

Một số công trình cho thấy có những nghiên cứu về tình trạng bỏ học của sinh viên tại một trường sử dụng dataset từ 87 mẫu đến 162.030 mẫu [2, 10]; bài toán dự báo năng lực người học để cải thiện phương pháp giảng dạy được

thực hiện bằng cách phân tầng năng lực theo từng môn học dựa trên dataset khoảng 300 mẫu cho một môn học [11]; bài toán khảo sát động lực học tập và kết quả học tập sử dụng nhiều datasets với số mẫu khác nhau theo ngành học của sinh viên, 93 mẫu, 137 mẫu, 174 mẫu, 404 mẫu [5]; v.v...



Hình 2.1 Quy trình thiết lập dataset của luận án

Luận án nghiên cứu các yếu tố năng lực của thí sinh ảnh hưởng đến kết quả thi học sinh giỏi cấp quốc gia môn Tin học. Dataset của luận án được thiết kế bằng cách phân tích các yếu tố năng lực tác động đến kết quả thi gồm những yếu tố năng lực do học tập (learning factors) và những yếu tố năng lực không do học tập (nonlearning factors). Các yếu tố năng lực này được mô hình hóa thành các biến năng lực, gồm biến năng lực học tập và biến năng lực phi học tập tương ứng với yếu tố năng lực do học tập và yếu tố năng lực không do học tập.

Những yếu tố năng lực học sinh được sử dụng theo từng mục tiêu và yêu cầu của nghiên cứu [10]. Trong đó, những biến năng lực học tập được đánh giá bằng kết quả học tập, những biến năng lực phi học tập được khảo sát thực tế theo từng đối tượng nghiên cứu. Đối tượng nghiên cứu của luận án này là những thí sinh đã tham gia thi học sinh giỏi cấp quốc gia môn Tin học. Những yếu tố năng lực tác động lên mỗi thí sinh với những mức độ khác nhau làm nên kết quả thi. Dataset của luận án là tập mẫu năng lực của nhiều thí sinh được thiết lập theo quy trình Hình 2.1.

2.2. Các yếu tố năng lực thí sinh

Đặc trưng năng lực thí sinh tham gia các kỳ thi học sinh giỏi được các tác giả khảo sát như những yếu tố ảnh hưởng năng lực, là số đo sự thành thạo và kết quả học tập của học sinh trong những lớp học [22]. Năng lực chịu ảnh hưởng bởi nhiều yếu tố khác nhau, không chỉ chịu ảnh hưởng từ việc học tập ở trường thể hiện qua kiến thức chuyên môn có được mà còn có những yếu tố không do học tập thể hiện qua cách ứng xử của học sinh đối với việc học [3, 22, 75-77]. Nói cách khác, năng lực có thể được khảo sát như sự phối hợp của nhiều yếu tố [3, 78]. Số lượng các biến năng lực được xác định theo chủ đề, môi trường giáo dục và mục tiêu nghiên cứu, ví dụ từ 11 đến 32 biến [10].

Tập biến năng lực của các tác giả không giống nhau về nội dung và số lượng, mỗi nghiên cứu sử dụng tập biến năng lực riêng phù hợp với mục tiêu nghiên cứu [10]. Đối tượng nghiên cứu của luận án này là thí sinh dự thi học sinh giỏi cấp quốc gia môn Tin học, được khảo sát như là sự phối hợp của kiến thức, kỹ năng, thói quen, cách ứng xử, nền tảng và sự quan tâm của gia đình, tác động của môi trường xã hội. Dữ liệu các biến năng lực chuyên môn của học sinh có thể được đánh giá bằng cách truy cập vào kết quả học tập. Dữ liệu các

biến năng lực phi học tập của học sinh gồm kỹ năng và cách sống được luận án đánh giá bằng các câu hỏi khảo sát (Questionnaire).

2.2.1. Năng lực chuyên môn

Từ những quan điểm khác nhau, có nhiều định nghĩa về năng lực chuyên môn của học sinh [77, 79]. Luận án này áp dụng định nghĩa năng lực chuyên môn là số đo khả năng những gì học sinh đã học được thông qua quy trình học tập được nhiều người sử dụng [76]. Năng lực chuyên môn của winner có được là do học tập tại các trường lớp, tự học qua sách và các nguồn thông tin. Kết quả này được đánh giá bằng điểm kiểm tra tại lớp học sau mỗi học kỳ. Điểm trung bình cuối học kỳ đánh giá năng lực chuyên môn toàn diện của học sinh. Trong đó, đối với học sinh chuyên Tin học, những môn học như Toán, Tin, Lý, Hóa, Sinh là những môn trọng tâm, hỗ trợ kiến thức dự thi Tin học [3].

2.2.2. Kỹ năng

Trong thời gian thi, những thí sinh của kỳ thi học sinh giỏi phải giải đáp những bài toán đặc biệt hoặc phải xử lý những tình huống phức tạp. Do đó, mỗi thí sinh không chỉ cần có đầy đủ kiến thức chuyên môn của ngành mình mà còn phải có những kỹ năng cần thiết để làm bài thi sao cho bài thi của mình có khả năng cạnh tranh với những bài thi khác về kết quả cũng như về sự thông minh của lời giải. Để chọn một đội viên cho đội tuyển, người thầy huấn luyện viên phải đánh giá được kiến thức cùng với các kỹ năng của ứng viên.

Kỹ năng đam mê. Trước hết đội viên đội tuyển phải là một học sinh ham muốn tham gia kỳ thi học sinh giỏi [3]. Những đội viên đội tuyển môn Tin học phải là những học sinh yêu thích Tin học, có mục tiêu theo đuổi tham gia kỳ thi học sinh giỏi môn Tin học, có thể bỏ qua những công việc khác để tập trung giải các bài toán Tin học khó, chuẩn bị cho kỳ thi.

Kỹ năng tự học. Tất cả những thí sinh thắng giải các kỳ thi học sinh giỏi đều là những học sinh có kỹ năng tự học tốt. Kỹ năng tự học là kỹ năng mà một học sinh tự chủ động tìm kiếm tài liệu để học tập, trau dồi, phát triển kiến thức mới mà không cần sự hướng dẫn trực tiếp của thầy tại lớp học [3]. Trong thời gian học tập ở lớp hoặc huấn luyện tại đội, những đội viên đội tuyển đều phải biết tự tìm thêm những tài liệu chuyên ngành đặc biệt để tự học cùng với những tài liệu học tập tại lớp, tự khám phá những vấn đề mới, phương pháp mới để giải những bài toán mới theo cách sáng tạo riêng.

Kỹ năng tự tin. Một người có kỹ năng tự tin là người tin tưởng vào những việc mình đã làm và bình tĩnh khi đối mặt với thách thức hoặc khó khăn trở ngại bất thường. Kỹ năng tự tin của một thí sinh giúp cho thí sinh xử lý những câu hỏi của đề thi, bình tĩnh đối phó với những vấn đề khó của đề thi [3, 80]. Một người có kỹ năng tự tin quá cao có thể trở thành người tự cao. Một thí sinh tự cao rất dễ bị thất bại trong kỳ thi. Do đó, một đội viên đội tuyển phải là một học sinh tự tin và khiêm tốn.

Kỹ năng nhận biết. Kỹ năng nhận biết của một người liên quan đến những hoạt động và quy trình tiếp nhận, lưu trữ, truy cập và xử lý thông tin [81]. Khả năng nhận biết ảnh hưởng mạnh mẽ đến thành tựu chuyên môn, những học sinh có kỹ năng nhận biết cao có nhiều khả năng đạt kết quả học tập cao [82]. Do đó, nhận biết là kỹ năng cần thiết cho những thí sinh tham dự kỳ thi học sinh giỏi.

Kỹ năng tư duy phản biện. Tư duy phản biện của một người cùng nghĩa với tư duy nhận xét liên quan đến khả năng lý luận, phân tích, đánh giá và bình luận về một vấn đề dựa trên quan điểm hoặc lý thuyết nào đó được áp dụng cho những câu hỏi, những tình huống hoặc những vấn đề thông thường hoặc bất thường [43, 83-85]. Một người có tư duy phản biện luôn muốn tìm ra chân lý hơn là cố tìm lấy chiến thắng khi tranh luận một vấn đề và không dễ bị cuốn theo một khuynh hướng nào [86]. Một người có tư duy phản biện tốt có kỹ năng lắng nghe tất cả những quan điểm khác nhau về một vấn đề để giải quyết vấn đề hoặc ra quyết định [43, 84, 85]. Tư duy phản biện đóng vai trò rất quan trọng đối với học sinh trong bối cảnh thế giới hiện nay - một thế giới đang thay đổi nhanh chóng và không ngừng vận động. Mọi học sinh cần linh hoạt và có khả năng thích ứng. Đặc biệt, các thí sinh dự thi học sinh giỏi muốn thắng giải cần có tư duy phản biện tốt để hiểu và giải quyết những vấn đề một cách độc lập bằng kỹ năng tư duy của mình [84]. Kỹ năng tư duy phản biện của mỗi người được đánh giá dựa trên 6 cấp độ:

- Cấp độ 1: Người không có kỹ năng tư duy phản biện.
- Cấp độ 2: Người có nhận thức về tầm quan trọng của tư duy phản biện.
- Cấp độ 3: Người mới bắt đầu tư duy phản biện.
- Cấp độ 4: Người biết tư duy phản biện trong một số vấn đề thực tế.
- Cấp độ 5: Người có thói quen tư duy phản biện.

- Cấp độ 6: Người biết áp dụng tư duy phản biện để ra quyết định và xử lý thông tin.

Kỹ năng tư duy sáng tạo. Kỹ năng tư duy sáng tạo của một người liên quan đến những kỹ năng nhận dạng, phân tích, suy diễn, đánh giá, tổng hợp [87] và lý luận để phát sinh những ý tưởng mới đối với một việc nào đó [62] hoặc để giải một bài toán có sẵn theo một cách khác, phi truyền thống nhưng tốt hơn [88]. Tính chất của các bài thi trong kỳ thi học sinh giỏi thường đòi hỏi các thí sinh phải giải bài toán một cách sáng tạo bằng kỹ năng tư duy của mình. Kỹ năng tư duy sáng tạo của mỗi người được đánh giá dựa trên 5 cấp độ:

- Cấp độ 1: Người có kỹ năng tư duy sáng tạo là người có kỹ năng thực hiện lại những bài toán đã được giải với sự hiểu biết và phân tích của riêng mình.
- Cấp độ 2: Người có kỹ năng tư duy sáng tạo là người có kỹ năng thực hiện lại những bài toán cũ với lời giải mới dựa trên nền của bài giải trước.
- Cấp độ 3: Người có kỹ năng tư duy sáng tạo là người có kỹ năng thực hiện lại bài toán cũ theo cách tiếp cận hoàn toàn mới.
- Cấp độ 4: Người có kỹ năng tư duy sáng tạo là người có kỹ năng tổng quát hóa thành một loại bài toán và định hướng các tiếp cận để giải loại bài toán.
- Cấp độ 5: Người có kỹ năng tư duy sáng tạo là người có kỹ năng tư duy tổng quát hóa để nâng cấp bài toán, mở ra lĩnh vực mới, hướng khoa học hoặc ngành nghề mới.

Kỹ năng quản lý thời gian. Kỹ năng quản lý thời gian khác với thói quen sử dụng thời gian. Trong phòng thi, thí sinh cần phải có kỹ năng quản lý thời gian để xây dựng kế hoạch tốt nhất cho quy trình chặt chẽ đáp ứng những câu hỏi của đề thi trong một khoảng thời gian có hạn [3]. Kỹ năng quản lý thời gian giúp thí sinh phân chia mức độ quan trọng hoặc mức độ khó của các nội dung trong đề thi, sắp xếp thứ tự ưu tiên giải quyết để tập trung dồn toàn bộ sức lực và trí tuệ vào việc giải bài thi trong một khung thời gian nhất định.

2.2.3. Cách sống

Những thói quen cá nhân và ảnh hưởng của gia đình có những tác động quan trọng đến thời gian học chuyên sâu luyện thi cũng như những kết quả của

đội viên khi tham gia kỳ thi học sinh giỏi cấp quốc gia [3]. Do đó, những yếu tố này cần được khảo sát cho từng ứng viên để chọn vào đội tuyển.

Thói quen sử dụng thời gian. Mỗi đội viên cần phải có chiến lược học tập, sử dụng thời gian cân bằng giữa học và thư giãn [3]. Mỗi đội viên cần có thói quen lập thời gian biểu hợp lý cho các hoạt động trong ngày để nâng cao kết quả học tập trong thời gian luyện thi.

Tiềm năng của gia đình. Yếu tố tiềm năng của gia đình liên quan đến khả năng hỗ trợ học sinh trong quá trình ôn luyện trước khi tham dự kỳ thi học sinh giỏi [3]. Truyền thống học tập của gia đình ảnh hưởng rất lớn đến kết quả học tập và thi của đội viên vì khả năng nhận biết và khả năng tiếp nhận kiến thức của một người có tính di truyền [82]. Trong thực tế, khả năng kinh tế (thu nhập) của gia đình, sự quan tâm và giúp đỡ của gia đình ảnh hưởng tích cực đến kết quả của đội viên trong khi ôn luyện và khi tham dự kỳ thi học sinh giỏi.

Vấn đề tình cảm. Những vấn đề về tình cảm trong quan hệ nam nữ của đội viên ảnh hưởng rất lớn đến kết quả học ôn luyện và kết quả thi học sinh giỏi. Trong quá khứ, Trường đã có trường hợp một đội viên rất giỏi nhưng không thắng giải. Khi tìm hiểu, các huấn luyện viên mới biết là học sinh ấy đang có vấn đề không tốt trong quan hệ tình cảm nam nữ.

Thói quen uống rượu bia. Nhiều tác giả nghiên cứu thấy rằng thói quen uống rượu bia của học sinh có tác động không tốt đến kết quả học tập [3].

Thói quen chơi game. Nhiều tác giả nghiên cứu thấy rằng những học sinh bị mê hoặc ghiền chơi game dễ dẫn đến kết quả học tập không tốt [3].

2.3. Câu hỏi khảo sát (Questionnaire)

Luận án thực nghiệm để tuyển chọn đội viên đội tuyển dự thi học sinh giỏi cấp quốc gia môn Tin học. Luận án áp dụng phương pháp khảo sát bằng Questionnaire để thu thập dữ liệu năng lực của từng thí sinh đã từng tham gia các kỳ thi học sinh giỏi cấp quốc gia môn Tin học. Tuy nhiên, quá trình thu thập dữ liệu gặp phải một số thách thức:

- *Quyền riêng tư của người được khảo sát.* Nội dung thu thập có thể xâm phạm quyền riêng tư của người được khảo sát (winners và non-winners). Sự hợp tác của người được khảo sát quyết định sự thành công của việc thu thập dữ liệu. Giải pháp của luận án là biên soạn câu hỏi khảo sát đáp ứng yêu cầu về các yếu tố tác động năng lực. Ngoài ra, câu hỏi phải dễ

được tiếp nhận và trả lời bởi người được khảo sát, tránh những câu hỏi về đời tư và gia đình.

- *Thời gian trả lời và yêu cầu đầu tư của người được khảo sát.* Người được khảo sát không có nhiều thời gian để trả lời các câu hỏi. Các câu hỏi khảo sát cần phải được biên soạn sao cho người được khảo sát dễ tiếp nhận và không mất nhiều thời gian để trả lời. Giải pháp của luận án là biên soạn câu hỏi bằng những từ ngữ đơn giản dễ hiểu. Các câu hỏi khảo sát không mất thời gian suy nghĩ và cách trả lời đơn giản, chỉ cần một dấu TIC (X) cho mỗi câu hỏi khảo sát. Phương pháp biên soạn câu hỏi này đã dẫn đến thách thức là các hồi đáp hoàn toàn định tính. Do đó, luận án đã đề xuất các tiếp cận:
 - *Lượng hóa tập giá trị của các biến năng lực.* Mỗi biến năng lực được xác định bởi một tập giá trị cần phải được định lượng trong không gian năng lực nhiều chiều. Luận án đã đề xuất các tiếp cận lượng hóa giá trị của các tập biến năng lực.
 - *Lượng hóa các giá trị của vector năng lực của người được khảo sát.* Luận án đã đề xuất các tiếp cận định lượng giá trị các thành phần năng lực của vector biểu diễn năng lực của người được khảo sát.
- *Vị trí cư trú rời rạc của những người được khảo sát.* Những thí sinh đã từng tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học đến từ nhiều tỉnh, địa phương khác nhau trên cả nước, không tập trung tại một khu vực địa lý nhất định. Để khắc phục khó khăn này, luận án tận dụng sự hỗ trợ của các đồng nghiệp là huấn luyện viên đội tuyển Tin học tại các trường trung học phổ thông có học sinh tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học. Việc thu thập dữ liệu chỉ thực hiện đối với các thí sinh đã tham dự kỳ thi trong vòng ba năm gần đây. Các huấn luyện viên đóng vai trò kết nối, hỗ trợ liên hệ và chuyển bảng Questionnaire đến các thí sinh đã tham dự kỳ thi quốc gia môn Tin học. Về kỹ thuật, luận án sử dụng Google Forms để gửi Questionnaire đến những người cần khảo sát là thí sinh đã dự thi trong 3 năm gần nhất.

Dữ liệu năng lực được thu thập bằng bảng Questionnaire gửi đến các winners (thí sinh thắng giải) và non-winners (thí sinh không thắng giải) của

những kỳ thi học sinh giỏi cấp quốc gia môn Tin học 3 năm gần nhất. Những câu khảo sát về năng lực học tập được sử dụng ngôn ngữ hỏi trực tiếp kết quả học tập của các môn học có liên quan đến năng lực chuyên môn cần thiết của môn thi học sinh giỏi cấp quốc gia. Tùy theo mục tiêu của đội tuyển tham dự giải học sinh giỏi cấp quốc gia bộ môn gì thì khảo sát kết quả học tập những môn học thích hợp. Một thách thức lớn khác, những người được khảo sát là những thí sinh đã từng tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học trong 3 năm gần nhất đang sinh sống hoặc học tập rải rác toàn quốc.

Một số tác giả đã chỉ ra những yếu tố năng lực học sinh phù hợp với văn hóa, chủ đề của mình như tuổi tác, nam/nữ, thời gian tham dự lớp học, có truy cập internet không, có sở hữu máy tính không, số lớp học đã tham dự, v.v.. [2]. Nhiều ý kiến cũng cho rằng những tính cách của con người đều thể hiện qua thói quen, qua hành động, v.v... Vì lý do tế nhị, luận án không đặt câu khảo sát hỏi trực tiếp về kỹ năng, về tâm lý giáo dục mà phân tích thành những câu khảo sát hỏi gián tiếp. Những câu khảo sát này hỏi về thói quen, hành vi ứng xử, hoặc sinh hoạt đơn giản hàng ngày. Do đó, người được khảo sát cảm thấy đơn giản dễ hợp tác vì không xâm phạm sâu vào quyền riêng tư cá nhân.

Thực nghiệm thu thập dữ liệu bằng những câu khảo sát trong bảng Questionnaire có 18 câu liên quan năng lực học tập, 22 câu liên quan năng lực phi học tập. Mỗi câu khảo sát được biên soạn ngắn gọn, có một số đáp án với cách hỏi đơn giản, ít mất thời gian đọc và trả lời để người được khảo sát dễ hợp tác. Bảng Questionnaire được trình bày chi tiết trong Phụ lục 1 đã đáp ứng được các thách thức trên.

- *Về nội dung câu khảo sát.* Luận án phân tích mỗi yếu tố năng lực thành một hoặc nhiều câu hỏi rất đơn giản, không đề cập sâu đến những nội dung riêng tư có thể xâm phạm quyền riêng tư của người được khảo sát.
- *Về hình thức câu khảo sát.* Luận án tiết kiệm thời gian của người được khảo sát bằng cách biên soạn nội dung khảo sát thành những câu hỏi ngắn gọn, đơn giản, dễ hiểu, đặc biệt là chỉ trả lời bằng một dấu TIC (X) cho mỗi câu hỏi khảo sát. Questionnaire không có những câu hỏi mà người được khảo sát phải suy nghĩ để viết thành câu trả lời.

- *Về phương pháp tiến hành.* Luận án nhờ sự hỗ trợ của các huấn luyện viên đội tuyển học sinh giỏi môn Tin học của một số trường trung học phổ thông tại các tỉnh để gửi Questionnaire bằng Google Forms đến từng thí sinh được khảo sát.

2.4. Thu thập dữ liệu

- *Đối tượng thu thập:* Dữ liệu được thu thập cho nghiên cứu là đặc trưng năng lực của các winners và non-winners đã từng tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học. Dữ liệu này được thu thập để cập nhật dataset hàng năm sau mỗi kỳ thi học sinh giỏi cấp quốc gia. Dataset có tính động theo thời gian, dataset gồm dữ liệu của những thí sinh đã dự thi học sinh giỏi cấp quốc gia trong 03 năm gần nhất. Các cuộc thi học sinh giỏi cấp quốc gia hàng năm luôn có tính mới, đề thi luôn được cập nhật theo xu hướng thời đại nên dataset phải luôn được cập nhật để đáp ứng được tính thời đại.

- *Phương pháp thu thập:* Luận án áp dụng kỹ thuật của Google Forms và được sự giúp đỡ của các thầy cô huấn luyện viên của các đội tuyển học sinh giỏi môn Tin học. Trong thực nghiệm, luận án được sự ủng hộ của các thí sinh dự thi học sinh giỏi cấp quốc gia môn Tin học các năm học 2020-2021, 2021-2022, 2022-2023, 2023-2024 trả lời Questionnaire. Kết quả, luận án đã thành lập được 2 tập dataset, Dataset 2024 và Dataset 2025.

2.5. Tiền xử lý dữ liệu

2.5.1. Chuẩn bị dữ liệu

Dữ liệu sau khi thu thập được từ các thí sinh tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học các năm học 2020-2021, 2021-2022, 2022-2023 được xử lý để thành lập Dataset 2024 gồm 120 winners và 101 non-winners. Dữ liệu sau khi thu thập được từ các thí sinh tham gia kỳ thi học sinh giỏi cấp quốc gia môn Tin học các năm học 2021-2022, 2022-2023, 2023-2024 được xử lý để thành lập Dataset 2025 gồm 360 winners và 137 non-winners.

Mỗi câu hỏi của bảng Questionnaire có nhiều đáp án, những người được khảo sát được mời đánh dấu TIC (X) vào 1 đáp án phù hợp nhất với bản thân mình. Các hồi đáp của Questionnaire khi nhận lại đều được tiền xử lý thủ công để chọn những hồi đáp hợp lệ [89]. Các hồi đáp Questionnaire được phân thành 2 nhóm: (1) hồi đáp của những winners, (2) hồi đáp của những non-winners,

được xử lý thủ công từng bảng hỏi đáp để chọn những hỏi đáp đúng yêu cầu về quy cách:

- Những hỏi đáp có câu hỏi nhận được nhiều hơn một dấu TIC trả lời hoặc có câu hỏi không có dấu TIC trả lời thì hỏi đáp bị loại ra vì không hợp lệ.
- Những hỏi đáp có ghi chú thêm thì bị loại vì được xem là không hợp lệ.
- Những hỏi đáp có yếu tố năng lực học tập thiếu hỏi đáp phần lớp 11, 12 được xem là hợp lệ.
- Những hỏi đáp có yếu tố năng lực học tập thiếu hoặc thừa các câu hỏi của lớp 10 được xem là không hợp lệ.
- Những hỏi đáp không trả lời đủ các câu hỏi phi học tập lớp 10, lớp 11 thì bị loại.

2.5.2. Xử lý dữ liệu

Khi thu thập dữ liệu từ người được khảo sát có một số trường hợp dữ liệu bị thiếu hoặc dữ liệu bị sai được xử lý như sau:

Trường hợp 1: Hỏi đáp trả lời sai do chủ quan của người được khảo sát.

Bảng 2.1. Xử lý dữ liệu bị sai Trường hợp 1 (Họ và tên trong cột 3 của bảng đã được thay đổi, không phải tên thật)

Timestamp	Họ và tên	Giới tính	Ngày sinh	Đang là	Trường THPT đã học hoặc đang học	Lớp đang học	Email	Kết quả học	Kết quả học	Kết quả học	Kết quả học
4/9/2024 20:23:59	Nguyễn Sĩ Thái	Nam	18/12/2007	Đúng	THPT chuyên Nguyễn Du - Đắk Lắk	11TH	nguyensithais10@gmail.com	9,6 - 10,0	7,6 - 8,0	7,6 - 8,0	7,6 - 8,0
4/10/2024 0:52:47	wda	Nam	19/4/2024	Đúng	asd	asd	asd	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0
4/10/2024 12:43:16	TPD	Nam	22/2/2004	Đúng	GD	111	Hh@gmail	6,4 - 7,0	6,4 - 7,0	6,4 - 7,0	6,4 - 7,0
4/10/2024 14:14:29	Ngô Xuân Toàn	Nam	4/5/2007	Đúng	Gia đình tphcm	11ctin	Chlengaming@gmail.com	9,6 - 10,0	8,6 - 9,0	8,1 - 8,5	8,1 - 8,5
4/10/2024 16:25:44	Đỗ Thành Trọng	Nam	18/8/2007	Đúng	THPT Chuyên Biên Hòa	11	thanhtrongo45@gmail.com	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0
4/10/2024 16:37:57	Trần Gia Huy	Nam	21/12/2007	Đúng	THPT Chuyên Lê Hồng Phong	11CTINA	trangiahuy211207@gmail.com	9,6 - 10,0	9,1 - 9,5	9,1 - 9,5	9,1 - 9,5
4/10/2024 16:43:02	Mai Nhật Quang	Nam	22/2/2007	Đúng	THPT chuyên Biên Hòa	11	cpptowin@gmail.com	9,6 - 10,0	9,1 - 9,5	9,1 - 9,5	8,6 - 9,0
4/10/2024 17:29:18	Lê Minh Tuấn	Nam	27/8/2007	Đúng	THPT Chuyên Lam Sơn - Thanh Hóa	11 Tin	ugactye12@gmail.com	9,6 - 10,0	9,6 - 10,0	8,6 - 9,0	9,1 - 9,5
4/10/2024 17:33:22	Đoàn Tuấn Anh	Nam	6/4/2006	Đúng	THPT chuyên Lê Hồng Phong	12CTIN	doantuananh06042006@gmail.com	9,6 - 10,0	8,1 - 8,5	8,1 - 8,5	8,1 - 8,5
4/10/2024 17:49:22	Ngô Minh Đức	Nam	13/12/2006	Đúng	Trường THPT Chuyên Hoàng Văn Thụ - Hòa Bình	12 chuyên Tin	ngominhduc006@gmail.com	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0	9,6 - 10,0
4/10/2024 18:06:14	Trịnh Thế Tùng	Nam	3/6/2007	Đúng	THPT Chuyên Lam Sơn - Thanh Hóa	111	trinhthung362007@gmail.com	9,6 - 10,0	9,1 - 9,5	8,6 - 9,0	8,6 - 9,0
4/10/2024 18:42:55	Nguyễn Tiến Thành	Nam	22/10/2007	Đúng	THPT Chuyên Lam Sơn - Thanh Hóa	11	hihihahomot@gmail.com	9,6 - 10,0	9,1 - 9,5	9,1 - 9,5	9,6 - 10,0
4/10/2024 18:58:54	Hoàng Văn Thái	Nam	13/8/2006	Đúng	Trường THPT Chuyên Hoàng Văn Thụ - Hòa Bình	12 Tin	hoangvanthaihb@gmail.com	8,1 - 8,5	8,1 - 8,5	8,1 - 8,5	8,1 - 8,5
4/10/2024 20:08:39	Lê Anh	Nam	5/1/2007	Đúng	Trường THPT Chuyên Lê Khiết - Quảng Ngãi	11	kakakav05@gmail.com	8,6 - 9,0	7,6 - 8,0	7,6 - 8,0	7,6 - 8,0
4/10/2024 20:16:50	Trịnh Quốc Bình	Nam	22/8/2007	Đúng	THPT chuyên Biên Hòa	11 chuyên Tin	trinhquocbinh2007hn@gmail.com	9,6 - 10,0	9,1 - 9,5	8,6 - 9,0	9,1 - 9,5
4/10/2024 21:20:49	Lê Hải Cường	Nam	27/7/2006	Đúng	Trường THPT Chuyên Hoàng Văn Thụ - Hòa Bình	12 chuyên Tin	Lehaicuongho2707@gmail.com	9,1 - 9,5	9,1 - 9,5	8,1 - 8,5	8,1 - 8,5
4/11/2024 12:46:33	Nguyễn Quang Long	Nam	8/5/2008	Đúng	Trường THPT Nguyễn Bình Khiêm	10a3	longnguyenquang234@gmail.com	9,6 - 10,0	7,1 - 7,5	9,1 - 9,5	9,1 - 9,5
4/11/2024 13:09:41	Đặng Quang Minh	Nam	2/4/2006	Đúng	Trường THPT Chuyên Nguyễn Tất Thành - Yên Bái	12	minhdangquang242006@gmail.com	9,6 - 10,0	8,1 - 8,5	8,1 - 8,5	8,1 - 8,5
4/11/2024 13:15:17	Phạm Nhật Quang	Nam	3/8/2005	Không	Trường THPT Chuyên Nguyễn Tất Thành - Yên Bái		pnquang05@gmail.com	9,6 - 10,0	9,1 - 9,5	7,1 - 7,5	8,1 - 8,5
4/11/2024 15:52:31	Nguyễn Thị Hòa	Nữ	24/7/2006	Đúng	Trường THPT Chuyên Bắc Ninh - Bắc Ninh	12Tin	nhhoa77477@gmail.com	9,6 - 10,0	8,6 - 9,0	9,1 - 9,5	9,1 - 9,5
4/11/2024 15:53:54	Nguyễn Minh Đức	Nam		Đúng	Trường THPT Chuyên Bắc Ninh - Bắc Ninh	12 Tin	nguyenninhduc08cbn@gmail.com	9,6 - 10,0	9,6 - 10,0	9,1 - 9,5	9,1 - 9,5
11/4/2024 16:47	Sùng A Sia	Nam			Trường THPT Chuyên Hoàng Văn Thụ - Hòa Bình			9,6 - 10,0	8,6 - 9,0	8,1 - 8,5	8,1 - 8,5
4/11/2024 18:50:38	Ngô Công Đức	Nam	25/8/2007	Đúng	Chuyên Bắc Ninh	11 tin	ngocongduc678@gmail.com	9,6 - 10,0	9,1 - 9,5	9,1 - 9,5	9,6 - 10,0

Dữ liệu hỏi đáp của người được khảo sát được trả lời sai do chủ quan. Trường hợp này xử lý bằng cách loại bỏ dữ liệu của người được khảo sát ra khỏi Dataset (Bảng 2.1).

Trường hợp 2: Hỏi đáp trả lời thiếu do khách quan.

Dữ liệu bị thiếu do khách quan (Bảng 2.2). Ví dụ như người được khảo sát đang học lớp 11 thì không có điểm các môn học ở lớp 12.

Mỗi câu khảo sát của Questionnaire được mô hình hóa toán học như một biến năng lực $F_s \mid s=1,..,40$. Mỗi kết quả hồi đáp Questionnaire từ người được khảo sát (winner hoặc nonwinner) $m \mid m=1,..,M$, với M là số hồi đáp sau khi

được xử lý, được mô hình hóa như một vector $\vec{F}^m | m=1,2,...,M$ năng lực của người được khảo sát m trong không gian 40 chiều.

$$\vec{F}^m = (f_1^m, ..., f_{18}^m, f_{19}^m, ..., f_{40}^m) \equiv (f_s^m | s=1,...,18,19,...,40)$$

Bài toán định lượng biến năng lực của người được khảo sát $m | m=1,2,...,M$ là một ánh xạ các đáp án định tính trong các câu khảo sát của Questionnaire mà người được khảo sát m đã chọn (X) thành vector $\vec{F}^m | m=1,2,...,M$ trong không gian 40 chiều $F_s | s=1,...,40$. Mỗi chiều là một biến năng lực của người được khảo sát m có giá trị được định lượng $f_s^m \in [0,10] \subset \mathbb{R}, \forall s \in \{1,...,40\} | m=1,...,M$. Nói cách khác,

- Input của bài toán định lượng biến năng lực là những hồi đáp Questionnaire định tính của người được khảo sát.
- Output của bài toán định lượng biến năng lực là các tập giá trị định lượng của các biến năng lực $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=1,...,40$ và vector năng lực của người được khảo sát $\vec{F}^m = (f_s^m | s=1,...,18,19,...,40)$ với $m=1,2,...,M$.

Bài toán được phân tích thành 2 outputs:

- Output 1: Tập giá trị của biến năng lực $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=1,...,40$.
- Output 2: Vector đặc trưng năng lực $\vec{F}^m = (f_s^m | s=1,...,18,19,...,40)$ của người được khảo sát $m | m=1,2,...,M$.

Output 1 được thực hiện bằng cách phân chia $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=1,...,40$ thành 2 nhóm với 2 tiếp cận định lượng tương ứng, nhóm biến năng lực học tập $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=1,...,18$ và nhóm biến năng lực phi học tập $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=19,...,40$.

2.6.1. Định lượng các biến năng lực học tập

Bảng 2.3. Ánh xạ biến đổi đáp án các biến năng lực học tập $F_s | s=1,...,18$ trong kết quả hồi đáp Questionnaire của người được khảo sát.

Đáp án các biến năng lực học tập của người được khảo sát trong Questionnaire	Giá trị $f_{s,\beta} \beta=1,2,..$ của biến năng lực học tập $F_s s=1,...,18$ được ánh xạ từ đáp án Questionnaire của người được khảo sát
[0.0 – 5.0]	1

(5.0 – 7.5]	5
(7.5 – 8.0]	6
(8.0 – 8.5]	7
(8.5 – 9.0]	8
(9.0 – 9.5]	9
(9.5 – 10.0]	10

Bảng 2.4. Kết quả số hóa các biến năng lực học tập

id	win	female	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
1	1	0	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
2	1	0	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
3	1	0	10	10	10	10	10	10	10	10	9	10	10	9	10	10	9	10	10	9
4	1	0	10	10	10	9	10	9	10	9	10	10	10	9	10	10	10	10	10	10
5	1	0	10	10	10	10	10	10												
6	1	0	10	10	10	10	10	10	10	10	10	9	10	10	10	10	9	10	10	9
7	1	0	10	10	10	10	10	10	10	10	10	9	9	10						
8	1	0	10	9	10	10	10	10	10	10	10	10	10	10						
9	1	0	10	10	10	10	10	9	10	10	10	10	10	10						
10	1	0	10	10	9	10	10	9	10	10	10	10	10	10	10	10	10	10	10	10

Mỗi đặc trưng năng lực học tập được mô hình hóa như một biến năng lực và được biểu diễn như một chiều năng lực $F_s | s=1,..,18$ có nhiều giá trị $f_{s,\beta} | \beta=1,2,..$ trong không gian năng lực $F_s = \{f_{s,\beta}\} | \beta=1,2,.. | s=1,..,40$. Trong đó, mỗi chiều $F_s | s=1,..,18$ biểu diễn năng lực học tập được định lượng bởi tập giá trị $\{f_{s,\beta}\} = \{1,5,6,7,8,9,10\} | \beta=1,2,..$ được ánh xạ từ kết quả hồi đáp trong những Questionnaire của những người được khảo sát theo Bảng 2.3 và kết quả là Bảng 2.4.

2.6.2. Định lượng các biến năng lực phi học tập

Tất cả các hồi đáp Questionnaire thu thập từ các winners và non-winners tham gia khảo sát được sử dụng để xác định những giá trị các biến năng lực phi học tập. Đáp án của các Questionnaires từ câu 19 đến câu 40 là những đáp án định tính được chuyển thành biến năng lực định lượng $F_s | s=19,..,40$. Mỗi biến gồm một tập giá trị cần được xác định.

- *Tập giá trị của những biến năng lực phi học tập.* Trong không gian năng lực, mỗi biến năng lực phi học tập $F_s | s=19,..,40$ tương ứng với 1 câu hỏi định tính $s=19,..,40$ trong Questionnaire. Tập giá trị định lượng của một biến năng lực phi học tập được chuyển đổi từ những hồi đáp Questionnaire của những người được khảo sát (winner hoặc nonwinner).

Kết quả, mỗi biến năng lực phi học tập $F_s | s=19,...,40$ có một tập giá trị định lượng là $F_s = \{f_{s,\beta} \in [0,10] \subset \mathbb{R} | \beta=1,2,.. \}$.

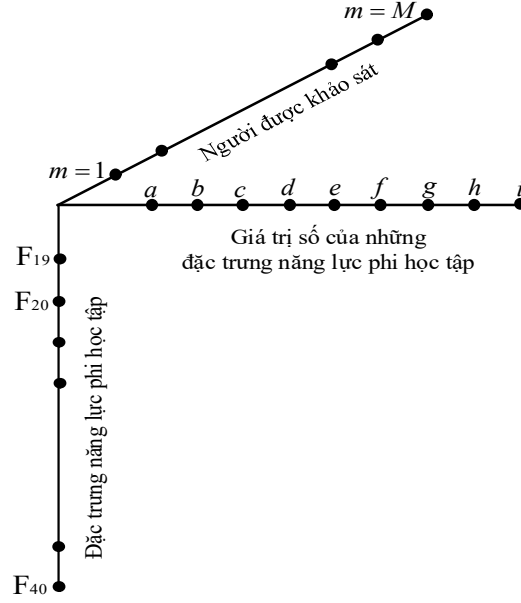
Bảng 2.5. Ma trận hỏi đáp về biến năng lực phi học tập $s=19,20,...,40$ của người được khảo sát $m | m=1,2,...,M$.

Câu 19	a_{19}^m	b_{19}^m	c_{19}^m	d_{19}^m	e_{19}^m	f_{19}^m			
Câu 20	a_{20}^m	b_{20}^m	c_{20}^m	d_{20}^m					
Câu 21	a_{21}^m	b_{21}^m	c_{21}^m	d_{21}^m	e_{21}^m	f_{21}^m	g_{21}^m	h_{21}^m	
Câu 22	a_{22}^m	b_{22}^m	c_{22}^m	d_{22}^m	e_{22}^m				
Câu 23	a_{23}^m	b_{23}^m	c_{23}^m	d_{23}^m					
Câu 24	a_{24}^m	b_{24}^m	c_{24}^m						
Câu 25	a_{25}^m	b_{25}^m							
Câu 26	a_{26}^m	b_{26}^m	c_{26}^m						
Câu 27	a_{27}^m	b_{27}^m	c_{27}^m	d_{27}^m	e_{27}^m	f_{27}^m	g_{27}^m	h_{27}^m	
Câu 28	a_{28}^m	b_{28}^m	c_{28}^m	d_{28}^m					
Câu 29	a_{29}^m	b_{29}^m							
Câu 30	a_{30}^m	b_{30}^m							
Câu 31	a_{31}^m	b_{31}^m							
Câu 32	a_{32}^m	b_{32}^m	c_{32}^m	d_{32}^m	e_{32}^m	f_{32}^m	g_{32}^m	h_{32}^m	i_{32}^m
Câu 33	a_{33}^m	b_{33}^m	c_{33}^m	d_{33}^m	e_{33}^m	f_{33}^m	g_{33}^m	h_{33}^m	
Câu 34	a_{34}^m	b_{34}^m							
Câu 35	a_{35}^m	b_{35}^m	c_{35}^m	d_{35}^m					
Câu 36	a_{36}^m	b_{36}^m	c_{36}^m						
Câu 37	a_{37}^m	b_{37}^m	c_{37}^m	d_{37}^m					
Câu 38	a_{38}^m	b_{38}^m	c_{38}^m	d_{38}^m					
Câu 39	a_{39}^m	b_{39}^m	c_{39}^m	d_{39}^m	e_{39}^m				
Câu 40	a_{40}^m	b_{40}^m	c_{40}^m	d_{40}^m	e_{40}^m				

- Các thành phần phi học tập của vector năng lực. Các thành phần phi học tập $f_s^m | s=19,...,40$ của vector năng lực người được khảo sát m được chuyển đổi từ các đáp án định tính của người này tại những câu khảo sát 19, 20,..., 40. Kết quả, ta có vector năng lực của người được khảo sát $m | m=1,2,...,M$ là $\vec{F}^m = (f_s^m \in F_s | s=1,...,18,19,...,40 | m=1,...,M)$.
- Ma trận hỏi đáp (Bảng 2.5): Hỏi đáp Questionnaire của một người được khảo sát được thiết lập thành ma trận với giá trị **1** được gán vào đáp án có đánh dấu (X) và giá trị **0** được gán vào đáp án không có đánh dấu ().

Những câu khảo sát trong Questionnaire không được biên soạn đủ số đáp án so với ma trận thì vị trí tương ứng trong ma trận hồi đáp được gán bằng 0 (ô mờ trong ma trận của Bảng 2.5).

Giá trị định lượng của biến năng lực phi học tập



Hình 2.2. Khối 3D hồi đáp - định lượng các giá trị của biến năng lực phi học tập $F_{19}, F_{20}, \dots, F_{40}$

Áp dụng khối 3D hồi đáp định lượng biến năng lực phi học tập (Hình 2.2), các giá trị của biến năng lực phi học tập $F_s | s=19, \dots, 40$ tương ứng câu khảo sát $s=19, \dots, 40$ là:

$$a_s = \frac{\delta}{M} \sum_1^M a_s^m | a_s^m \in \{0,1\} | m=1,2,\dots,M; s=19,\dots,40$$

$$b_s = \frac{\delta}{M} \sum_1^M b_s^m | b_s^m \in \{0,1\} | m=1,2,\dots,M; s=19,\dots,40$$

$$c_s = \frac{\delta}{M} \sum_1^M c_s^m | c_s^m \in \{0,1\} | m=1,2,\dots,M; s=19,\dots,24,26,\dots,28,32,33,35,\dots,40$$

$$d_s = \frac{\delta}{M} \sum_1^M d_s^m | d_s^m \in \{0,1\} | m=1,2,\dots,M; s=19,\dots,23,27,28,32,33,35,37,\dots,40$$

$$e_s = \frac{\delta}{M} \sum_1^M e_s^m | e_s^m \in \{0,1\} | m=1,2,\dots,M; s=19,21,22,27,32,33,39,40$$

$$f_s = \frac{\delta}{M} \sum_1^M f_s^m | f_s^m \in \{0,1\} | m=1,2,\dots,M, s=19,21,27,32,33$$

$$g_s = \frac{\delta}{M} \sum_1^M g_s^m \mid g_s^m \in \{0,1\} \mid m=1,2,...,M; s=21,27,32,33$$

$$h_s = \frac{\delta}{M} \sum_1^M h_s^m \mid h_s^m \in \{0,1\} \mid m=1,2,...,M, s=21,27,32,33$$

$$i_s = \frac{\delta}{M} \sum_1^M i_s^m \mid i_s^m \in \{0,1\} \mid m=1,2,...,M, s=32$$

Giá trị của biến năng lực phi học tập được chuẩn hóa sao cho $a_s, b_s, c_s, d_s, e_s, f_s, g_s, h_s, i_s \in [0,10] \subset \mathbb{R} \mid \forall s \in \{19,20,...,39,40\}$ bằng cách gán δ bằng một số thích hợp tùy theo số người tham gia khảo sát. Thực nghiệm của luận án chọn $\delta=10$.

Theo Questionnaire của luận án, tập giá trị của các biến năng lực phi học tập

$$F_s = \{a_s, b_s, c_s, d_s, e_s, f_s, g_s, h_s, i_s \in [0,10] \subset \mathbb{R}\} \equiv \{f_{s,\beta} \in [0,10] \subset \mathbb{R} \mid \beta=1,2,...\} \mid s=19,...,40$$

trong không gian năng lực là:

$$F_{19} = \{a_{19}, b_{19}, c_{19}, d_{19}, e_{19}, f_{19}\} \equiv \{f_{19.1}, f_{19.2}, f_{19.3}, f_{19.4}, f_{19.5}, f_{19.6}\}$$

$$F_{20} = \{a_{20}, b_{20}, c_{20}, d_{20}\} \equiv \{f_{20.1}, f_{20.2}, f_{20.3}, f_{20.4}\}$$

$$\begin{aligned} F_{21} &= \{a_{21}, b_{21}, c_{21}, d_{21}, e_{21}, f_{21}, g_{21}, h_{21}\} \\ &\equiv \{f_{21.1}, f_{21.2}, f_{21.3}, f_{21.4}, f_{21.5}, f_{21.6}, f_{21.7}, f_{21.8}\} \end{aligned}$$

$$F_{22} = \{a_{22}, b_{22}, c_{22}, d_{22}, e_{22}\} \equiv \{f_{22.1}, f_{22.2}, f_{22.3}, f_{22.4}, f_{22.5}\}$$

$$F_{23} = \{a_{23}, b_{23}, c_{23}, d_{23}\} \equiv \{f_{23.1}, f_{23.2}, f_{23.3}, f_{23.4}\}$$

$$F_{24} = \{a_{24}, b_{24}, c_{24}\} \equiv \{f_{24.1}, f_{24.2}, f_{24.3}\}$$

$$F_{25} = \{a_{25}, b_{25}\} \equiv \{f_{25.1}, f_{25.2}\}$$

$$F_{26} = \{a_{26}, b_{26}, c_{26}\} \equiv \{f_{26.1}, f_{26.2}, f_{26.3}\}$$

$$\begin{aligned} F_{27} &= \{a_{27}, b_{27}, c_{27}, d_{27}, e_{27}, f_{27}, g_{27}, h_{27}\} \\ &\equiv \{f_{27.1}, f_{27.2}, f_{27.3}, f_{27.4}, f_{27.5}, f_{27.6}, f_{27.7}, f_{27.8}\} \end{aligned}$$

$$F_{28} = \{a_{28}, b_{28}, c_{28}, d_{28}\} \equiv \{f_{28.1}, f_{28.2}, f_{28.3}, f_{28.4}\}$$

$$F_{29} = \{a_{29}, b_{29}\} \equiv \{f_{29.1}, f_{29.2}\}$$

$$F_{30} = \{a_{30}, b_{30}\} \equiv \{f_{30.1}, f_{30.2}\}$$

$$F_{31} = \{a_{31}, b_{31}\} \equiv \{f_{31.1}, f_{31.2}\}$$

$$\begin{aligned}
F_{32} &= \{a_{32}, b_{32}, c_{32}, d_{32}, e_{32}, f_{32}, g_{32}, h_{32}, i_{32}\} \\
&\equiv \{f_{32.1}, f_{32.2}, f_{32.3}, f_{32.4}, f_{32.5}, f_{32.6}, f_{32.7}, f_{32.8}, f_{32.9}\} \\
F_{33} &= \{a_{33}, b_{33}, c_{33}, d_{33}, e_{33}, f_{33}, g_{33}, h_{33}\} \\
&\equiv \{f_{33.1}, f_{33.2}, f_{33.3}, f_{33.4}, f_{33.5}, f_{33.6}, f_{33.7}, f_{33.8}\} \\
F_{34} &= \{a_{34}, b_{34}\} \equiv \{f_{34.1}, f_{34.2}\} \\
F_{35} &= \{a_{35}, b_{35}, c_{35}, d_{35}\} \equiv \{f_{35.1}, f_{35.2}, f_{35.3}, f_{35.4}\} \\
F_{36} &= \{a_{36}, b_{36}, c_{36}\} \equiv \{f_{36.1}, f_{36.2}, f_{36.3}\} \\
F_{37} &= \{a_{37}, b_{37}, c_{37}, d_{37}\} \equiv \{f_{37.1}, f_{37.2}, f_{37.3}, f_{37.4}\} \\
F_{38} &= \{a_{38}, b_{38}, c_{38}, d_{38}\} \equiv \{f_{38.1}, f_{38.2}, f_{38.3}, f_{38.4}\} \\
F_{39} &= \{a_{39}, b_{39}, c_{39}, d_{39}, e_{39}\} \equiv \{f_{39.1}, f_{39.2}, f_{39.3}, f_{39.4}, f_{39.5}\} \\
F_{40} &= \{a_{40}, b_{40}, c_{40}, d_{40}, e_{40}\} \equiv \{f_{40.1}, f_{40.2}, f_{40.3}, f_{40.4}, f_{40.5}\}
\end{aligned}$$

2.7. Định lượng giá trị các thành phần phi học tập của vector năng lực của người được khảo sát

Giá trị các thành phần năng lực phi học tập ($f_{19}^m, \dots, f_{40}^m$) của người được khảo sát $m | m = 1, 2, \dots, M$ (Bảng 2.6) trong vector năng lực $\vec{F}^m = (f_s^m | s = 1, \dots, 18, 19, \dots, 40)$ được xác định:

$$\begin{aligned}
f_{19}^m &= a_{19}^m \cdot a_{19} + b_{19}^m \cdot b_{19} + c_{19}^m \cdot c_{19} + d_{19}^m \cdot d_{19} + e_{19}^m \cdot e_{19} + f_{19}^m \cdot f_{19} \\
f_{20}^m &= a_{20}^m \cdot a_{20} + b_{20}^m \cdot b_{20} + c_{20}^m \cdot c_{20} + d_{20}^m \cdot d_{20} \\
f_{21}^m &= a_{21}^m \cdot a_{21} + b_{21}^m \cdot b_{21} + c_{21}^m \cdot c_{21} + d_{21}^m \cdot d_{21} + e_{21}^m \cdot e_{21} + f_{21}^m \cdot f_{21} + g_{21}^m \cdot g_{21} + h_{21}^m \cdot h_{21} \\
f_{22}^m &= a_{22}^m \cdot a_{22} + b_{22}^m \cdot b_{22} + c_{22}^m \cdot c_{22} + d_{22}^m \cdot d_{22} + e_{22}^m \cdot e_{22} \\
f_{23}^m &= a_{23}^m \cdot a_{23} + b_{23}^m \cdot b_{23} + c_{23}^m \cdot c_{23} + d_{23}^m \cdot d_{23} \\
f_{24}^m &= a_{24}^m \cdot a_{24} + b_{24}^m \cdot b_{24} + c_{24}^m \cdot c_{24} \\
f_{25}^m &= a_{25}^m \cdot a_{25} + b_{25}^m \cdot b_{25} \\
f_{26}^m &= a_{26}^m \cdot a_{26} + b_{26}^m \cdot b_{26} + c_{26}^m \cdot c_{26} \\
f_{27}^m &= a_{27}^m \cdot a_{27} + b_{27}^m \cdot b_{27} + c_{27}^m \cdot c_{27} + d_{27}^m \cdot d_{27} + e_{27}^m \cdot e_{27} + f_{27}^m \cdot f_{27} + g_{27}^m \cdot g_{27} + h_{27}^m \cdot h_{27} \\
f_{28}^m &= a_{28}^m \cdot a_{28} + b_{28}^m \cdot b_{28} + c_{28}^m \cdot c_{28} + d_{28}^m \cdot d_{28} \\
f_{29}^m &= a_{29}^m \cdot a_{29} + b_{29}^m \cdot b_{29} \\
f_{30}^m &= a_{30}^m \cdot a_{30} + b_{30}^m \cdot b_{30} \\
f_{31}^m &= a_{31}^m \cdot a_{31} + b_{31}^m \cdot b_{31}
\end{aligned}$$

$$\begin{aligned}
f_{32}^m &= a_{32}^m \cdot a_{32} + b_{32}^m \cdot b_{32} + c_{32}^m \cdot c_{32} + d_{32}^m \cdot d_{32} + e_{32}^m \cdot e_{32} + f_{32}^m \cdot f_{32} + g_{32}^m \cdot g_{32} + h_{32}^m \cdot h_{32} + i_{32}^m \cdot i_{32} \\
f_{33}^m &= a_{33}^m \cdot a_{33} + b_{33}^m \cdot b_{33} + c_{33}^m \cdot c_{33} + d_{33}^m \cdot d_{33} + e_{33}^m \cdot e_{33} + f_{33}^m \cdot f_{33} + g_{33}^m \cdot g_{33} + h_{33}^m \cdot h_{33} \\
f_{34}^m &= a_{34}^m \cdot a_{34} + b_{34}^m \cdot b_{34} \\
f_{35}^m &= a_{35}^m \cdot a_{35} + b_{35}^m \cdot b_{35} + c_{35}^m \cdot c_{35} + d_{35}^m \cdot d_{35} \\
f_{36}^m &= a_{36}^m \cdot a_{36} + b_{36}^m \cdot b_{36} + c_{36}^m \cdot c_{36} \\
f_{37}^m &= a_{37}^m \cdot a_{37} + b_{37}^m \cdot b_{37} + c_{37}^m \cdot c_{37} + d_{37}^m \cdot d_{37} \\
f_{38}^m &= a_{38}^m \cdot a_{38} + b_{38}^m \cdot b_{38} + c_{38}^m \cdot c_{38} + d_{38}^m \cdot d_{38} \\
f_{39}^m &= a_{39}^m \cdot a_{39} + b_{39}^m \cdot b_{39} + c_{39}^m \cdot c_{39} + d_{39}^m \cdot d_{39} + e_{39}^m \cdot e_{39} \\
f_{40}^m &= a_{40}^m \cdot a_{40} + b_{40}^m \cdot b_{40} + c_{40}^m \cdot c_{40} + d_{40}^m \cdot d_{40} + e_{40}^m \cdot e_{40}
\end{aligned}$$

Bảng 2.6. Kết quả thực nghiệm lượng hóa các biến năng lực phi học tập của những người được khảo sát

F19	F20	F21	F22	F23	F24	F25	F26	F27	F28	F29	F30	F31	F32	F33	F34	F35	F36	F37	F38	F39	F40
1.9	4	3	3	3	6	7	4	5	5	7	6	6	6	7	5.5	5	6.8	7	4	6.6	6
6	5	3	5	3	6	7	4	5	5	7	4	4	6	7	5.5	5	6.8	7	4	6.6	6
6	5	0.33	5	3	6	7	4	5	5	7	6	6	6	1	5.5	5	3	7	4	6.6	6
6	4	3	5	3	6	7	5	5	3.3	7	6	4	6	7	4.5	3.3	6.8	1	5.8	3	6
1.9	5	3	3	2.5	6	7	4	5	3.3	7	4	6	6	7	4.5	3.3	6.8	7	4	6.6	6
6	5	3	5	3	6	7	4	5	5	3	6	6	6	7	5.5	5	6.8	7	4	6.6	6
6	4	3	5	3	3	7	5	5	5	7	6	6	1.5	2	5.5	5	6.8	7	5.8	6.6	6
6	5	3	0.2	3	6	7	4	5	5	7	6	4	6	7	4.5	3.3	6.8	7	4	6.6	6
6	5	3	5	3	6	3	5	1.5	5	7	6	6	6	7	4.5	3.3	3	7	5.8	6.6	6
1.9	5	3	3	3	3	7	4	5	5	7	6	6	6	7	5.5	5	6.8	2	4	3	3

2.8. Dataset

Dataset là tập những vector biểu diễn năng lực (học tập và phi học tập) của M người được khảo sát, winners và non-winners. Sau khi xử lý các hồi đáp Questionnaire, dataset được tổ chức như một bảng dữ liệu gồm 41 cột, trong đó cột 1 chỉ thị số thứ tự của người được khảo sát $m | m = 1, 2, \dots, M$ và 40 cột chỉ thị các biến năng lực của M người được khảo sát với M dòng, mỗi dòng biểu diễn vector năng lực của một người được khảo sát (Bảng 2.7).

$$\vec{F}^m = (f_1^m, \dots, f_{40}^m) | m = 1, 2, \dots, M$$

Bảng 2.7. Mô hình dataset

Người được khảo sát m	f_1	..	f_{18}	f_{19}	..	..	f_{40}
$m = 1$	f_1^1						f_{40}^1
$m = 2$							
..							
..							
$m = M$	f_1^M						f_{40}^M

2.9. Kết luận chương 2

Chương 2 của luận án thành lập dataset năng lực thí sinh đã thi quốc gia môn Tin học bằng cách phân tích các đặc trưng năng lực học tập thành 18 câu hỏi năng lực học tập, và phân tích tâm lý giáo dục, sư phạm thành 22 câu hỏi năng lực phi học tập để khảo sát nhiều thí sinh đã dự thi học sinh giỏi cấp quốc gia môn Tin học. Sau khi tiền xử lý, luận án sử dụng hồi đáp định tính của các winners và non-winners được khảo sát để thành lập dataset.

Mỗi câu hỏi được mô hình hóa như một thành phần của vector năng lực của người được khảo sát. Biến năng lực học tập được định lượng bằng kết quả học tập, biến năng lực phi học tập được định lượng bằng các tiếp cận ma trận hồi đáp và khối 3D hồi đáp do luận án đề xuất. Bộ câu hỏi của Questionnaire do luận án biên soạn có những đặc điểm:

1. *Nội dung đơn giản.* Questionnaire do luận án biên soạn có nội dung đơn giản, đặc biệt là những câu hỏi phi học tập được phân tích đầy đủ nhưng không xâm phạm đến quyền riêng tư của người được khảo sát.
2. *Cấu trúc đơn giản, trả lời đơn giản.* Questionnaire do luận án biên soạn có cấu trúc đơn giản, người được khảo sát không cần nhiều thời gian để đọc hiểu và trả lời. Người được khảo sát chỉ trả lời bằng một dấu TIC (X) cho mỗi câu hỏi.
3. *Lượng hóa đặc trưng năng lực.* Luận án đã đề xuất tiếp cận ma trận hồi đáp và khối 3D hồi đáp, để lượng hóa các đặc trưng năng lực từ những dấu (X) của hồi đáp khảo sát. Tiếp cận lượng hóa đặc trưng năng lực do luận án đề xuất và thực hiện có tính linh hoạt, có thể áp dụng cho các Questionnaires của các kỳ thi, cấp thi, hoặc môn thi khác nhau.
4. *Linh hoạt.* Tiếp cận ma trận hồi đáp và khối 3D hồi đáp hỗ trợ sự linh hoạt trong soạn thảo Questionnaire, có thể điều chỉnh nội dung cũng như số câu khảo sát theo tính chất của từng môn thi, cấp thi theo từng năm thi một cách dễ dàng.

Tóm lại, luận án đã đề xuất tiếp cận xây dựng dataset linh hoạt (smart dataset) phù hợp với yêu cầu thành lập đội tuyển dự thi các giải. Tiếp cận dataset linh hoạt cung cấp quy trình thành lập tập dữ liệu năng lực của các thí sinh dự các kỳ thi học sinh giỏi. Quy trình có thể được áp dụng để thành lập các đội tuyển theo từng môn học, từng kỳ thi, và cấp thi khác nhau.

CHƯƠNG 3. TIẾP CẬN HỌC MÁY ĐỂ CHỌN HỌC SINH GIỎI VÀO ĐỘI TUYỂN DỰ THI HỌC SINH GIỎI

3.1. Giới thiệu

“Tuyển chọn học sinh vào đội tuyển dự thi học sinh giỏi” là một bài toán ra quyết định, trong đó cần lựa chọn một số lượng hạn chế học sinh có năng lực nổi trội nhất từ một tập lớn các ứng viên để đại diện cho trường tham dự kỳ thi học sinh giỏi. Những thí sinh muốn thắng giải học sinh giỏi phải có năng lực đáp ứng yêu cầu của kỳ thi. Áp dụng triết lý “gần mực thì đen, gần đèn thì sáng”, luận án đề xuất “Tuyển chọn học sinh vào đội tuyển dự thi học sinh giỏi” dựa trên sự tương đồng về năng lực, năng lực học tập và năng lực phi học tập, giữa ứng viên với những thí sinh đã từng thắng giải (winners).

Những năm gần đây, khoa học máy tính với các kỹ thuật khai phá dữ liệu (data mining) và học máy (machine learning) đã xuyên ngành vào giáo dục để giải những bài toán cải thiện và nâng cao chất lượng học tập của học sinh [90]. Theo xu thế đó, luận án áp dụng các kỹ thuật khai phá dữ liệu và học máy để giải bài toán “Tuyển chọn học sinh vào đội tuyển dự thi học sinh giỏi cấp quốc gia môn Tin học”. Luận án đánh giá mức độ tương đồng về suất năng lực của ứng viên với năng lực của những winners và áp dụng cây quyết định chọn những ứng viên có năng lực tương đồng cao nhất với winners vào đội tuyển để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học.

Mỗi mẫu của dataset đều đã được gán nhãn winner hoặc non-winner, nên luận án áp dụng học máy có giám sát. Tuy nhiên, do khách quan như đã trình bày trong chương trước, số lượng mẫu của dataset bị hạn chế. Luận án khắc phục điểm yếu này bằng cách phân hoạch ngẫu nhiên dataset thành training set và testing set nhiều lần $p | p = 1, 2, \dots, P$ và đánh giá kết quả Winner-domain bằng confusion matrix để chọn trường hợp tốt nhất (Winner-domain tốt). Mặc khác, Winner-domain cũng được chứng minh tập lỗi bằng thực nghiệm với tập non-winners có số mẫu nhiều hơn số mẫu của testing set.

Luận án đã mô hình hóa năng lực mỗi học sinh như một vector nhiều chiều, mỗi chiều biểu diễn một biến năng lực. Các biến năng lực của học sinh gồm những biến liên quan kiến thức chuyên môn, gọi là biến năng lực học tập (learning variables) và những biến liên quan đến gia đình, liên quan đến kỹ năng, cách sống cá nhân, thói quen và ứng xử trong môi trường xã hội, gọi

chung là biến năng lực phi học tập (non-learning variables) [2]. Luận án áp dụng tiếp cận phân lớp của kỹ thuật học máy (machine learning) [91] và tiếp cận phân tích thống kê để đánh giá mức độ tương tự của ứng viên với winners để tìm ra ứng viên có độ tương đồng cao nhất với winners.

3.2. Những công trình nghiên cứu được liên quan

Những năm gần đây, khoa học máy tính đã xuyên ngành vào khoa học giáo dục với những áp dụng tiếp cận học máy vào việc xử lý dữ liệu giáo dục (educational data) nhằm nâng cao chất lượng giáo dục. Bài toán thu hút nhiều tác giả nghiên cứu là “Khai phá dữ liệu giáo dục để phân tích, dự báo tình trạng học tập của học sinh”.

Một cách tổng quát, bài toán tuyển chọn học sinh giỏi vào đội tuyển có thể xem như bài toán đánh giá năng lực học sinh để dự đoán khả năng thắng giải của mỗi học sinh ứng viên. Trong những năm gần đây, như là một trào lưu khoa học máy tính xuyên ngành vào giáo dục, nhiều tác giả đã áp dụng các tiếp cận học máy để giải bài toán đánh giá năng lực học sinh cho nhiều mục tiêu khác nhau [9, 22, 33, 75, 92]. Trong trào lưu đó, nhiều tác giả đã áp dụng tiếp cận học máy để dự báo năng lực học sinh [2, 75, 78, 91-93].

Các bài toán thu hút nhiều tác giả thuộc chủ đề Phân tích và dự báo năng lực học sinh, hoặc dự báo năng lực học sinh để can thiệp sớm. Trong đó, các tiếp cận data mining dùng để phân tích tình trạng học tập của học sinh nhằm cải thiện tình trạng dạy và học. Bài toán cần thiết của người quản lý giáo dục là dự báo kết quả dạy và học để phân tầng trình độ học sinh dựa trên học máy có giám sát và các tiếp cận data mining để áp dụng mô hình học khám phá [11].

Luận án đề xuất tiếp cận Winner-domain và tiếp cận Winner-cosin để đánh giá độ tương đồng năng lực về suất và về khuynh hướng của ứng viên với winners, cây quyết định dựa trên mức độ tương đồng về khuynh hướng năng lực của ứng viên với tất cả winners trong winner-dataset để tuyển chọn. Ứng viên được quyết định chọn vào đội khi được cả hai tiếp cận Winner-domain và Winner-cosin đánh giá mức độ tương đồng năng lực với các winners của winner-dataset cao nhất.

- *Tiếp cận Winner-domain.* Tiếp cận Winner-domain được phát triển từ tiếp cận khai phá dữ liệu SVM [26, 94] để thiết lập domain trong không gian năng lực và đánh giá khả năng thắng giải của ứng viên dựa trên mức

độ tương đồng năng lực của ứng viên với các winners. Từ nhiều Winner-domains do phân hoạch dataset nhiều lần (P lần), tiếp cận chọn ra một số rất ít Winner-domains tốt để làm miền chọn ứng viên.

- *Tiếp cận Winner-cosin.* Tiếp cận Winner-cosin áp dụng độ tương đồng cosin [95, 96] để đánh giá mức độ tương đồng về hướng năng lực của ứng viên với các winners. Mỗi ứng viên trong Winner-domain tốt được đánh giá mức độ tương đồng hướng năng lực với các winners bằng cosin trung bình của ứng viên với tất cả các winners của winner-dataset.

3.3. Phát biểu bài toán

3.3.1. Câu hỏi nghiên cứu

Bài toán được đặt ra nhằm trả lời câu hỏi “Ứng viên nào có nhiều khả năng thắng giải?”. Trong thực tế, có rất nhiều học sinh giỏi có thể tham gia đội tuyển. Tuy nhiên, luận án thiết lập các thuật toán quyết định tuyển chọn những ứng viên có năng lực cao hơn vừa đủ số lượng đội viên cần thiết và chấp nhận để lại những ứng viên khác.

3.3.2. Mô tả bài toán

Input:

- Tập những thí sinh đã thắng giải, Winner-dataset $W = \{w_1, \dots, w_N\}$, trong đó mỗi winner w_n được biểu diễn như một vector toán học trong không gian năng lực S chiều.
- Tập những thí sinh đã dự thi nhưng không thắng giải, Nonwinner-dataset $U = \{u_1, \dots, u_M\}$, trong đó mỗi nonwinner u_m được biểu diễn như một vector toán học trong không gian năng lực S chiều.
- Tập những ứng viên $X = \{x_1, \dots, x_L\}$ muốn được chọn vào đội tuyển, trong đó mỗi ứng viên x_l được biểu diễn như một vector toán học trong không gian năng lực S chiều.

Output:

Tập những ứng viên được quyết định chọn vào đội tuyển $Y_{team} = \{y_1, \dots, y_{K_{team}}\}$.

3.3.3. Tiếp cận Winner-domain

Thuật toán Winner-domain được đề xuất để xác định miền không gian năng lực tập trung nhiều winners của 3 năm trước. Những ứng viên có năng lực

trong miền không gian tập trung nhiều winners được xem là có độ tương đồng cao với các winners. Tiếp cận Winner-domain [CT.1, CT.2] áp dụng thuật toán Winner-domain xác định nhiều Winner-domains đối với một tập winner-dataset bằng cách phân hoạch nhiều lần và tìm ra những Winner-domain tốt xác định miền không gian tập trung winners cao nhất.

3.3.3.1. Winner-domain

Luận án phân hoạch dataset thành tập mẫu thắng giải (winner-dataset) $W = \{w_1, \dots, w_N\}$ gồm những mẫu biểu diễn năng lực của các winners và tập mẫu không thắng giải (nonwinner-dataset) $U = \{u_1, \dots, u_M\}$ gồm những mẫu biểu diễn năng lực của các non-winners. Luận án đề xuất thuật toán tạo ra nhiều Winner-domains (miền năng lực của những winners) để chọn đội viên tiềm năng. Những ứng viên có vector năng lực trong Winner-domains thì được dự đoán có nhiều khả năng thắng giải. Áp dụng tiếp cận học máy, luận án sử dụng những training sets được thành lập từ winner-dataset $W = \{w_1, \dots, w_N\}$ để thành lập Winner-domains chọn đội viên tiềm năng.

Chọn ngẫu nhiên khoảng 70% số mẫu của winner-dataset $W = \{w_1, \dots, w_N\}$ nhiều lần $p | p = 1, 2, \dots, P$ để thành lập các tập huấn luyện (training sets) $V^p = \{w_1, \dots, w_I\} | p = 1, 2, \dots$; khoảng 30% mẫu còn lại được dùng làm winner testing set $V_{test}^p | p = 1, 2, \dots$. Chọn ngẫu nhiên J mẫu của nonwinner-dataset $U = \{u_1, \dots, u_M\}$ nhiều lần $p | p = 1, 2, \dots, P$ để được những nonwinner testing sets $U_{test}^p = \{u_1, \dots, u_J\} | p = 1, 2, \dots$ tương ứng. Mỗi lần $p | p = 1, 2, \dots, P$ chọn được một winner testing set $V_{test}^p | p = 1, 2, \dots$ và một nonwinner testing set $U_{test}^p | p = 1, 2, \dots$ để thành lập $Testing\ set \equiv V_{test}^p \cup U_{test}^p | p = 1, 2, \dots$. Mỗi training set $V^p = \{w_1, \dots, w_I\} | p = 1, 2, \dots$ được đặc trưng bởi trọng tâm w_c^p , bán kính R^p và góc hướng θ^p .

- Trọng tâm w_c^p của training set: $w_c^p = [f_{1,c}^p; \dots; f_{s,c}^p; \dots; f_{S,c}^p]$

trong đó $f_{s,c}^p = \frac{1}{I} \sum_{i=1}^I f_{s,i}^p | s = 1, 2, \dots, S$, với I là số mẫu của training set, S là số biến năng lực.

- Bán kính R^p của training set: $R^p = d_{\max}^p(w_c^p, w_i^p), \forall i$
 với $d_{c,i}^p \equiv d(w_c^p, w_i^p) = \sqrt{(f_{1,c} - f_{1,i})^2 + \dots + (f_{S,c} - f_{S,i})^2}$
- Góc hướng θ^p của training set V^p . Góc hướng của training set là góc hướng tối đa $\theta^p = (\theta_{c,i}^p)_{\max}, \forall i$ của các mẫu trong tập training set so với trọng tâm.

$$\cos \theta_{c,i}^p = \frac{w_c^p \cdot w_i^p}{|w_c^p| |w_i^p|} = \frac{f_{1,c}^p \cdot f_{1,i}^p + \dots + f_{S,c}^p \cdot f_{S,i}^p}{\sqrt{(f_{1,c}^p)^2 + \dots + (f_{S,c}^p)^2} \sqrt{(f_{1,i}^p)^2 + \dots + (f_{S,i}^p)^2}}$$

3.3.3.2. Winner-domain tốt

Mỗi training set V^p có thể phát sinh nhiều Winner-domains $D_k^p | k=1,2,\dots$ bằng cách thay đổi các tính chất của training set. Mỗi Winner-domain có:

- Trọng tâm w_c^p cùng với trọng tâm của Training set
 $V^p = \{w_1, \dots, w_I\} | p=1,2,\dots$
- Bán kính $R_k^p = \alpha_k \cdot R^p$ với $\alpha_k < 1 | k=1,2,\dots$
- Góc hướng $\theta_k^p \leq \theta^p = (\theta_{c,i}^p)_{\max} | \forall i, i=1,\dots, I_k^p$ với I_k^p là số mẫu của Winner-domain $D_k^p | k=1,2,\dots$

Mỗi Winner-domain $D_k^p | k=1,2,\dots$ chứa các winners w_i có
 $(d_{c,i}^p < R_k^p) \wedge (\cos \theta_{c,i}^p < \cos \theta_k^p)$

Vì $\alpha_k < 1$, nên có $\bar{I}_k^p$ mẫu của training set ở ngoài Winner-domain, số mẫu $\bar{I}_k^p$ của training set ở ngoài Winner-domain tăng lên khi $\alpha_k \rightarrow 0$. Nếu chọn α_k lớn thì có nhiều khả năng xác định nhầm mẫu nonwinner thành winner. Nếu chọn α_k nhỏ thì bỏ sót nhiều mẫu winner.

Các Winner-domains được đánh giá hiệu năng bằng các hệ số của unnormalized confusion matrix 2 x 2 (Bảng 3.1) [2]:

Bảng 3.1. Unnormalized confusion matrix

Tập dữ liệu kiểm tra	Kết luận: Thắng giải (cột 1 của matrix)	Kết luận: Không thắng giải (cột 2 của matrix)
Lớp thắng giải (hàng 1 của matrix) $V = \{v_i, i=1, \dots, I\}$	TP : Số mẫu $v_i, v_i \in V$ được đoán ĐÚNG là thắng giải	FN : Số mẫu $v_i, v_i \in V$ được đoán SAI là không thắng giải (bị bỏ sót)
Lớp không thắng giải (hàng 2 của matrix) $U = \{u_j, j=1, \dots, J\}$	FP : Số mẫu $u_j, u_j \in U$ được đoán NHẦM là thắng giải	TN : Số mẫu $u_j, u_j \in U$ được đoán ĐÚNG là không thắng giải

- Bắt đầu với: $TP=0; FN=0; FP=0; TN=0$
- For $i = 1, 2, \dots, I$: If $(d_{i,c}^p < d^p) \wedge (\cos \theta_{i,c}^p > \cos \theta^p)$ then $TP := TP + 1$
- For $i = 1, 2, \dots, I$: If $(d_{i,c}^p > d^p) \vee (\cos \theta_{i,c}^p < \cos \theta^p)$ then $FN := FN + 1$
- For $j = 1, 2, \dots, J$: If $(d_{j,c}^p < d^p) \wedge (\cos \theta_{j,c}^p \geq \cos \theta^p)$ then $FP := FP + 1$
- For $j = 1, 2, \dots, J$: If $(d_{j,c}^p > d^p) \vee (\cos \theta_{j,c}^p < \cos \theta^p)$ then $TN := TN + 1$

Độ tin cậy của các Winner-domains được đánh giá bởi [2, 75, 78]:

- Tỷ lệ winners của training set thuộc Winner-domain: $WTR = \frac{I - \bar{I}}{I}$
- Tỷ lệ chọn đúng: Precision = $TPR = \frac{TP}{TP + FP}$
- Tỷ lệ chọn nhầm: $FPR = \frac{FP}{TP + FP}$
- Tỷ lệ bỏ sót: $FNR = \frac{FN}{TP + FN} = \frac{FN}{I}$

Các giá trị TPR và FNR phụ thuộc vào giá trị $\alpha_k | 0 < \alpha_k < 1$ được chọn, với α_k nhỏ thì tỷ lệ bỏ sót FNR cao, tỷ lệ chọn nhầm FPR thấp. Trong thực tế, số học sinh cần chọn vào đội tuyển không nhiều, luận án chọn α_k theo chiến lược “chọn đúng vừa đủ, không chọn hơn là chọn nhầm”.

Các Winner-domains $D_k^p(w_c^p, R_k^p, \theta_k^p)$ được đánh giá theo các hệ số TP , FP , FN của ma trận unnormalized confusion matrix [2]. Những Winner-domains có TP cao, FP thấp, FN thấp, được kiểm tra tính lỗi. Những Winner-

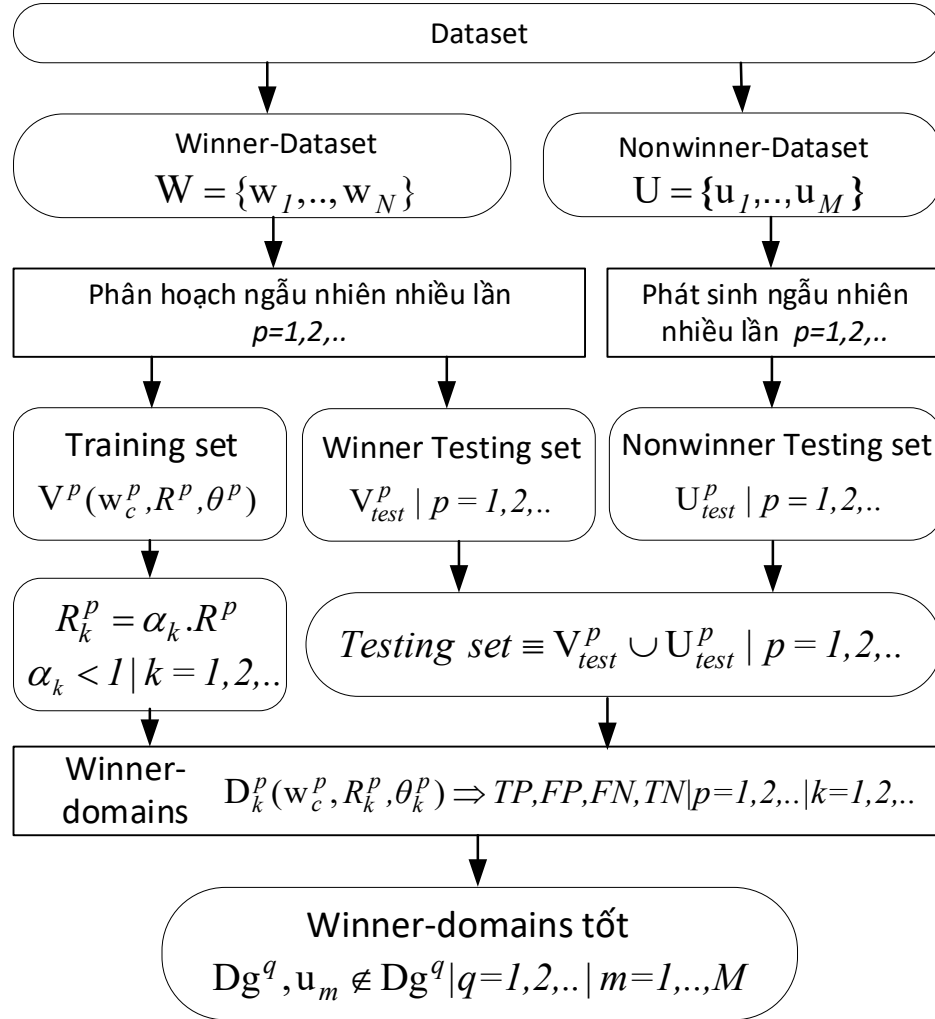
domains có TP cao, FP thấp, FN thấp không chứa nonwinner $u_m, \forall u_m \in U | m=1, \dots, M$ thì được chọn làm Winner-domain tốt $Dg^q, u_m \notin Dg^q | m=1, \dots, M | q=1, 2, \dots$

3.3.3.3. Đội viên tiềm năng

Mỗi ứng viên $x_l \in X$ được đánh giá khả năng thắng giải bằng cách đối sánh năng lực với các Winner-domains tốt $Dg^q | q=1, 2, \dots$ trong không gian năng lực, nếu $x_l \in \exists Dg^q | q=1, 2, \dots$ thì x_l được chọn làm đội viên tiềm năng.

3.3.3.4. Quy trình thiết lập Winner-domain tốt

Quy trình thiết lập các Winner-domains và Winner-domains tốt được tóm tắt trong Hình 3.1.



Hình 3.1. Quy trình thiết lập Winner-domain [CT.2]

3.3.3.5. Thuật toán Winner-domain

Input:

Winner-Dataset: $W = \{w_1, \dots, w_N\}$

Nonwinner-Dataset: $U = \{u_1, \dots, u_M\}$

Ứng viên: $X = \{x_1, \dots, x_L\}$

Mỗi mẫu của các tập dữ liệu có S biến $F_0, F_1, \dots, F_S$

Output:

Đội tiềm năng: $Y = \{y_1, \dots, y_K\}$

Giải thuật:

//Phân hoạch dữ liệu ngẫu nhiên Winner-Dataset $W = \{w_1, \dots, w_N\}$:

For $p = 1 \rightarrow P$ do

- Training set $V^p = \{w_1, \dots, w_L\}$, có trọng tâm w_c^p , bán kính R^p , và góc hướng θ^p .

- Winner Testing set V_{test}^p .

- Nonwinner testing set U_{test}^p .

// Sinh và đánh giá Winner-domain

For $k = 1 \rightarrow K$ do

- $R_k^p = \alpha_k \cdot R^p \mid \alpha_k < 1$

- Winner-domain: $D_k^p(w_c^p, R_k^p, \theta_k^p)$

- Kiểm thử với $Testing\ set^p \equiv V_{test}^p \cup U_{test}^p$

- Kiểm tra tính lỗi:

If $\exists u_m \in D_k^p(w_c^p, R_k^p, \theta_k^p) \mid u_m \in U \rightarrow$ Loại bỏ D_k^p

// không có tính lỗi

Else $u_m \notin D_k^p(w_c^p, R_k^p, \theta_k^p) \mid \forall u_m \in U \rightarrow$

$D_k^p(w_c^p, R_k^p, \theta_k^p)$ // có tính lỗi.

- Đánh giá hiệu năng $TP, FP, FN, TN \mid p \mid k$ của

$D_k^p(w_c^p, R_k^p, \theta_k^p)$ lỗi:

If $FP \neq 0$:

Loại bỏ $D_k^p(w_c^p, R_k^p, \theta_k^p)$ $k \leftarrow K+1$ // Continue
 Else $FP=0$:
 Chọn $D_k^p(w_c^p, R_k^p, \theta_k^p)$ là Winner-domain tốt
 $Dg^q(w_c^q, R_k^q, \theta_k^q)$ và Break //Thoát khỏi vòng lặp
 End For
 // Điều kiện dừng cho vòng lặp p
 If tìm được Winner-domain tối ưu then
 Break
 End If
 End For

Chọn ứng viên vào đội tiềm năng:

Ứng viên: $X = \{x_1, \dots, x_L\}$

For $l=1$ đến L :

- Nếu $x_l \in \exists Dg^q(w_c^q, R_k^q, \theta_k^q)$: x_l được chuyển vào đội tiềm năng $Y = \{y_1, \dots, y_K\}$.
- Nếu $x_l \notin Dg^q(w_c^q, R_k^q, \theta_k^q) | \forall q$: x_l được loại bỏ.

Kết quả: Đội tiềm năng $Y = \{y_1, \dots, y_K\}$

*** Đánh giá độ phức tạp thuật toán Winner-domain:**

Thuật toán có 2 vòng lặp lồng nhau: Vòng lặp ngoài lặp theo p với p đi từ $1 \rightarrow P$, vòng lặp bên trong lặp theo K với k đi từ $1 \rightarrow K$.

Phân hoạch ngẫu nhiên winner-dataset nhiều lần $p=1,2,\dots,P$: Phân hoạch ngẫu nhiên tập winner-dataset thành những training sets $V^p | p=1,2,\dots,P$ và những winner testing sets $V_{test}^p | p=1,2,\dots,P$.

Độ phức tạp cho P lần phân hoạch: $O(P.(N+M).S)$ (3.1)

Mỗi lần phân hoạch là một vòng lặp để tạo ra K Winner-domains.

Phát sinh nhiều miền năng lực Winner-domains: Phát sinh nhiều miền năng lực Winner-domains cho từng training set $V^p | p=1,2,\dots,P$. Mỗi training set $V^p | p=1,2,\dots,P$ được thay đổi bán kính R^p thành nhiều bán kính $R_k^p = \alpha_k R^p$. Với mỗi vòng lặp tạo training set $V^p | p=1,2,\dots,P$, thực hiện vòng lặp tính $R_k^p = \alpha_k R^p$

theo $k=1,2,...,K$ để phát sinh ra K Winner-domains khác nhau từ cùng một training set. Vòng lặp theo $k=1,2,...,K$ được dừng lại khi tỉ lệ chọn nhầm $FPR=0$.

- Tính toán các tham số đặc tính $(w_c^p, R_k^p, \theta_k^p)$ của từng Winner-domain tương ứng với training set có bán kính R^p :
 - Độ phức tạp tính trọng tâm w_c^p phụ thuộc vào tổng và trung bình của các đặc trưng S trên tập V^p . Chi phí là $N.S$
 - Tính toán bán kính R^p và góc hướng θ^p đòi hỏi các phép tính trên S của tập V^p .
 - Độ phức tạp của việc tính tham số ban đầu: $O(N.S)$
- Lặp K lần để tạo nhiều miền $D_k^p \mid k=1,2,...,K$: Mỗi lần lặp chỉ thay đổi bán kính $R_k^p = \alpha_k R^p$

$$\text{Độ phức tạp là: } O(K.N.S) \quad (3.2)$$

Xây dựng tập kiểm thử testing set, đánh giá hiệu năng của từng Winner-domain và chọn Winner-domain tốt: Tập kiểm thử tổng hợp là $Testing\ set \equiv V_{test}^p \cup U_{test}^p \mid p=1,2,...$ Mỗi lần cần đánh giá toàn bộ mẫu kiểm thử $N+M$

Đánh giá hiệu năng: Mỗi miền năng lực D_k^p được đánh giá bằng cách duyệt toàn bộ dữ liệu kiểm thử $N+M$ mẫu. Xác định vị trí của một mẫu đối với Winner-domain $D_k^p(w_c^p, R_k^p, \theta_k^p)$ bằng cách tính (tính khoảng cách, so sánh bán kính, so sánh góc hướng) trên S đặc trưng. Chi phí cho một mẫu là $O(S)$.

$$\text{Chi phí cho một miền } D_k^p: O((N+M).S)$$

$$\text{Chi phí cho tất cả } K \text{ miền } D_k^p: O(K(N+M).S)$$

$$\text{Độ phức tạp của } P \text{ vòng lặp: } O(P.K.(N+M))$$

Bước chọn ứng viên: Kiểm tra tập ứng viên kích thước L , mỗi ứng viên cần $O(S)$ cho mỗi Winner-domain và L nhỏ không đáng kể

$$\text{Độ phức tạp thời gian của thuật toán Winner-domain:}$$

$$O(P.K.(N+M).S)$$

3.3.3.6. Thuật toán Winner-cosin

Tiếp cận Winner-cosin đánh giá độ tương đồng về hướng năng lực của đội viên tiềm năng $Y = \{y_1, \dots, y_K\}$ đối với winner-dataset $W = \{w_1, \dots, w_N\}$ theo thuật toán tương đồng cosin (CosinSimilarity):

$$\text{CosinSimilarity}(y_k, w_n) \equiv \cos \theta_{k,n} = \frac{y_k \cdot w_n}{|y_k| |w_n|} = \frac{f_{1,k} \cdot f_{1,n} + \dots + f_{S,k} \cdot f_{S,n} + \dots + f_{S,k} \cdot f_{S,n}}{\sqrt{(f_{1,k})^2 + \dots + f_{S,k})^2} \sqrt{(f_{1,n})^2 + \dots + (f_{S,n})^2}}, \forall n$$

Thuật toán tương đồng cosin xác định độ tương đồng về hướng năng lực của mỗi đội viên tiềm năng $y_k | k=1..K$ đối với winner-dataset $W = \{w_1, \dots, w_N\}$

$$\text{CosinSimilarity}(y_k, W) \equiv \frac{1}{N} \sum_{n=1}^N \cos(y_k, w_n)$$

Input: Winner-Dataset: $W = \{w_1, \dots, w_N\}$

Đội tiềm năng: $Y = \{y_1, \dots, y_K\}$

Output: Đội tuyển: $Z_{team} = \{z_1, \dots, z_{K_{team}}\}$

Giải thuật:

For $k=1$ đến $k=K$:

$$\text{Cos}k \equiv \text{CosinSimilarity}(y_k, W) \equiv \frac{1}{N} \sum_{n=1}^N \cos(y_k, w_n) | k=1, \dots, K$$

Áp dụng thuật toán Bubble sort sắp xếp tập $Y = \{y_1, \dots, y_K\}$ thành tập

$\rightarrow \downarrow Z = \{z_1, \dots, z_K\}$ theo giá trị $\text{Cos}k$ giảm dần.

Kết quả:

Chọn $K_{team} | K_{team} \leq K$ đội viên tiềm năng là những đội viên có giá trị

$\text{Cos}k$ cao vào đội tuyển dự thi: $Z = (z_1, \dots, z_{K_{team}}) | K_{team} \leq K$

Độ phức tạp của thuật toán Winner-cosin

Ký hiệu và mô tả:

L : số lượng ứng viên.

N : số lượng winner của winner-dataset.

S : số chiều không gian năng lực.

K : số lượng đội viên tiềm năng.

K_{team} : số lượng đội viên đội tuyển cần chọn.

Bước 1: Tính độ tương đồng cosin theo thuật toán Winner-cosin

Mỗi đội viên tiềm năng được tính tương đồng cosin với N winners của winner-dataset. Độ phức tạp là: $O(K.N.S)$

Bước 2: Danh sách đội viên tiềm năng

Độ phức tạp chọn đội viên tiềm năng là: $O(K.\log K)$

Bước 3: Quyết định chọn đội viên đội tuyển

Độ phức tạp sau khi chọn K_{team} đội viên tiềm năng đầu tiên của chuỗi thứ tự $\text{Cos } k$ là $O(K_{team})$ không đáng kể so với độ phức tạp sắp xếp vì $K_{team} \ll K$

Độ phức tạp tổng thể của thuật toán Winner-cosin: $O(L.(S+P)+K.N.S+K.\log K)$.

3.4. Thực nghiệm và đánh giá kết quả thực nghiệm

Được sự hỗ trợ của những huấn luyện viên các đội tin học và những thí sinh đã dự thi tin học cấp quốc gia các kỳ 2020-2021, 2021-2022, 2022-2023, 2023-2024 của nhiều trường tại nhiều tỉnh, luận án đã xây dựng được dataset để áp dụng cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm 2024 và dataset áp dụng cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm 2025. Trên cơ sở dataset thành lập được bằng Questionnaire cùng với các tiếp cận ma trận hỏi đáp và khối 3D hỏi đáp, luận án đã sử dụng tiếp cận Winner-domain, thuật toán Winner-domain kết hợp thuật toán Winner-cosin để tuyển chọn đội viên đội tuyển Tin học.

Kết quả tuyển chọn này được phản ánh bằng kết quả các kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm 2024 và năm 2025 áp dụng thực nghiệm tại một trường trung học phổ thông chuyên ở miền tây. Lịch sử của trường cho thấy rằng các đội tin học của trường từ trước đến nay không thắng giải hoặc chỉ thắng vài giải học sinh giỏi cấp quốc gia môn Tin học. Để chuẩn bị kỳ thi 2024 và 2025, Trường sau khi tuyển đội viên bằng phương pháp truyền thống, áp dụng mô hình Winner-domain kết hợp Winner-cosin để đánh giá lại năng lực đội viên của đội tuyển môn Tin học và huấn luyện theo mô hình học khám phá có hướng dẫn trực quan được kết quả như sau:

Bảng 3.2. Kết quả thực nghiệm mô hình Winner-domain kết hợp Winner-cosin do luận án đề xuất áp dụng cho các kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và năm học 2024-2025

Kết quả kỳ thi năm học 2023-2024			Kết quả kỳ thi năm học 2024 – 2025		
Nhì	Ba	Khuyến khích	Nhì	Ba	Khuyến khích
1	1	3	1	2	3

THỰC NGHIỆM KỲ THI NĂM HỌC 2023-2024

Thực nghiệm biến đổi 10 đội viên được lãnh đạo tỉnh chọn bằng phương pháp truyền thống để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 thành những ứng viên của mô hình Winner-domain kết hợp Winner-cosin. Thực nghiệm sử dụng Dataset 2024 là dữ liệu năng lực một số thí sinh đã dự thi học sinh giỏi cấp quốc gia môn Tin học các năm học 2020-2021, 2021-2022, 2022-2023, gồm 120 winners của Winner Dataset 2024 và 101 non-winners của Nonwinner Dataset 2024. Mô hình Winner-domain kết hợp Winner-cosin do luận án đề xuất thực nghiệm trên 10 đội viên này như là những ứng viên của mô hình. Sau khi có kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024, luận án thực hiện thêm mô hình học máy Random Forest dự đoán kết quả của những đội viên do tỉnh chọn để so sánh với mô hình Winner-domain kết hợp Winner-cosin của luận án đã thực hiện trước với cùng một Dataset và cùng tập năng lực của các ứng viên (đội viên do tỉnh chọn).

1. Thực nghiệm với mô hình Winner-domain kết hợp Winner-cosin

Bước 1. Phát sinh Training set và Testing set

Winner Dataset 2024 được phân hoạch ngẫu nhiên 4 lần để được các Training set $W^p(w_c^p, \theta^p, R^p) \mid p=1,2,3,4$; mỗi Training set có 84 mẫu; và các Winner Testing set tương ứng $V^p \mid p=1,2,3,4$, mỗi Winner Testing set có 36 mẫu. Mỗi Testing set tương ứng với Training set gồm 36 mẫu của Winner Testing set tương ứng $V^p \mid p=1,2,3,4$ và 36 mẫu $U^p \mid p=1,2,3,4$ được phát sinh ngẫu nhiên tương ứng từ Nonwinner Dataset, Testing set = $V^p \cup U^p \mid p=1,2,3,4$ tương ứng với Training set $W^p \mid p=1,2,3,4$.

Bước 2. Phát sinh Winner-domain:

Những Winner-domains được phát sinh từ những training sets. Trong thực nghiệm này, mỗi Training set $W^p(w_c^p, \theta^p, R^p) \mid p=1,2,3,4$ được phát sinh 08 Winner-domains $D_k^p(w_c^p, \theta_k^p, R_k^p \mid p=1,2,3,4 \mid k=1,2,...,8$. Trong đó, $R_k^p = \alpha_k \cdot R^p \mid k=1,2,...,8$. Mỗi Winner-domain được đánh giá WTR , TPR , FPR , FNR và $accuracy$.

Bảng 3.3. Kết quả phát sinh Winner-domains cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024

p	$R^p / \cos \theta^p$	k	α_k	R_k^p	$\cos \theta_k^p$	Tỉ lệ $winner \in D_k^p$ WTR	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Độ chính xác mẫu thắng được chọn: $accuracy$
1	9.3064 / 0.9817	1	0.9	8.37576	0.9851	0.952381	1.0	0.1389	0.0	0.878049
		2	0.87	8.096568	0.9875	0.928571	0.9722	0.0	0.0278	1.0
		3	0.85	7.9104399	0.9875	0.928571	0.9722	0.0	0.0278	1.0
		4	0.83	7.7243119	0.9875	0.928571	0.9722	0.0	0.0278	1.0
		5	0.80	7.44512	0.9878	0.904762	0.9444	0.0	0.0556	1.0
		6	0.78	7.258992	0.9885	0.869048	0.9444	0.0	0.0556	1.0
		7	0.75	6.9798	0.9898	0.821429	0.8889	0.0	0.1111	1.0
		8	0.73	6.793672	0.9898	0.785714	0.8889	0.0	0.1111	1.0
2	9.3872 / 0.9814	1	0.9	8.44848	0.9849	0.952381	1.0	0.1389	0.8611	0.878049
		2	0.87	8.166864	0.9873	0.928571	1.0	0.0	0.0	1.0
		3	0.85	7.97912	0.9873	0.928571	1.0	0.0	0.0	1.0
		4	0.83	7.791376	0.9873	0.928571	1.0	0.0	0.0	1.0
		5	0.80	7.50976	0.9874	0.892857	0.9722	0.0	0.0278	1.0
		6	0.78	7.322016	0.9883	0.869048	0.9722	0.0	0.0278	1.0
		7	0.75	7.0404	0.9894	0.821429	0.9444	0.0	0.0556	1.0
		8	0.73	6.852656	0.9898	0.77381	0.9444	0.0	0.0556	1.0
3	9.39 / 0.9813	1	0.9	8.451	0.9847	0.964286	0.9722	0.1389	0.0278	0.875
		2	0.87	8.1693	0.9856	0.952381	0.9722	0.1389	0.0278	0.875
		3	0.85	7.98150	0.9875	0.940476	0.9722	0.0556	0.0278	0.945946
		4	0.83	7.7937	0.9875	0.940476	0.9722	0.0556	0.0278	0.940476
		5	0.80	7.5120	0.9875	0.904762	0.9444	0.0556	0.0556	0.944444
		6	0.78	7.3242	0.9882	0.869048	0.9444	0.0278	0.0556	0.971429
		7	0.75	7.0425	0.9894	0.833333	0.9444	0.0	0.0556	1.0
		8	0.73	6.8547	0.9903	0.797619	0.8889	0.0	0.1111	1.0
4	9.3955 / 0.9815	1	0.9	8.45595	0.9847	0.97619	0.9444	0.0556	0.0556	0.944444
		2	0.87	8.174085	0.9872	0.952381	0.9444	0.0278	0.0556	0.971429
		3	0.85	7.986175	0.9872	0.952381	0.9444	0.0278	0.0556	0.971429
		4	0.83	7.798265	0.9872	0.952381	0.9444	0.0278	0.0556	0.971429
		5	0.80	7.5164	0.9883	0.916667	0.8611	0.0	0.1389	1.0

	6	0.78	7.32849	0.9884	0.904762	0.8611	0.0	0.1389	1.0
	7	0.75	7.0466250	0.9894	0.857143	0.8333	0.0	0.1667	1.0
	8	0.73	6.858715	0.9898	0.821429	0.8333	0.0	0.1667	1.0

Bước 3. Chọn Winner-domain tốt

Khảo sát tính chất của 32 Winner-domains được phát sinh, luận án chọn 04 domains được cho là tốt hơn những domains khác.

Bảng 3.4. Kết quả chọn Winner-domains tốt cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024

D_k^p	$R^p / \cos\theta^p$	α_k	R_k^p	$\cos\theta_k^p$	Tỉ lệ $winner \in D_k^p$ WTR	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Đánh giá mức độ tốt
D_2^2	9.3872 / 0.9814	0.87	8.166864	0.9873	0.928571	1.0	0.0	0.0	1
D_2^1	9.3064 / 0.9817	0.87	8.096568	0.9875	0.928571	0.9722	0.0	0.0278	2
D_7^3	9.39 / 0.9813	0.75	7.0425	0.9894	0.833333	0.9444	0.0	0.0556	3
D_5^4	9.3955 / 0.9815	0.80	7.5164	0.9883	0.916667	0.8611	0.0	0.1389	4

Bước 4. Chọn đội viên theo mô hình Winner-domain kết hợp Winner-cosin với 10 ứng viên

Với 10 ứng viên của mô hình Winner-domain kết hợp Winner-cosin $x_l \in X \mid l=1,2,...,10$ (đội viên được chọn theo phương pháp truyền thống) tham gia kỳ thi 2024, luận án đã khảo sát năng lực bằng Questionnaire trước kỳ thi. Từ những hồi đáp khảo sát, mỗi ứng viên được biểu diễn như một vector trong không gian năng lực. Vector năng lực của mỗi ứng viên trước kỳ thi được đối sánh với 04 Winner-domain tốt D_2^2 , D_2^1 , D_7^3 , D_5^4 để tuyển chọn.

Bảng 3.5. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_2^2

Ứng viên $x_l \in X$	$d_{c,r}^2$	$\cos\theta_{c,r}^2$	R_2^2	$\cos\theta_2^2$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^2 \leq R_2^2) \wedge (\cos\theta_{c,r}^2 \geq \cos\theta_2^2)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_l	5.2401	0.9943	8.166864	0.9873	Thắng giải	Thắng giải

x_2	3.9725	0.9968	8.166864	0.9873	Thắng giải	Thắng giải
x_3	6.3011	0.9923	8.166864	0.9873	Thắng giải	Thắng giải
x_4	5.9336	0.9925	8.166864	0.9873	Thắng giải	Thắng giải
x_5	4.9533	0.9952	8.166864	0.9873	Thắng giải	Thắng giải
x_6	8.9745	0.9834	8.166864	0.9873	Không thắng giải	Không thắng giải
x_7	9.5336	0.982	8.166864	0.9873	Không thắng giải	Không thắng giải
x_8	10.3148	0.979	8.166864	0.9873	Không thắng giải	Không thắng giải
x_9	9.6898	0.9812	8.166864	0.9873	Không thắng giải	Không thắng giải
x_{10}	10.1392	0.9789	8.166864	0.9873	Không thắng giải	Không thắng giải

Bảng 3.6. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_2^2 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^2$	$\cos \theta_{c,r}^2$	R_2^2	$\cos \theta_2^2$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_2	3.9725	0.9968	8.166864	0.9873	0.9982	Thắng giải
x_1	5.2401	0.9943	8.166864	0.9873	0.9970	Thắng giải
x_5	4.9533	0.9952	8.166864	0.9873	0.9964	Thắng giải
x_3	6.3011	0.9923	8.166864	0.9873	0.9949	Thắng giải
x_4	5.9336	0.9925	8.166864	0.9873	0.9933	Thắng giải

Bảng 3.7. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_2^1

Ứng viên $x_l \in X$	$d_{c,r}^1$	$\cos \theta_{c,r}^1$	R_2^1	$\cos \theta_2^1$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^1 \leq R_2^1) \wedge (\cos \theta_{c,r}^1 \geq \cos \theta_2^1)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_1	5.2153	0.9944	8.096568	0.9875	Thắng giải	Thắng giải
x_2	4.0795	0.9966	8.096568	0.9875	Thắng giải	Thắng giải
x_3	6.166	0.9926	8.096568	0.9875	Thắng giải	Thắng giải
x_4	5.9835	0.9924	8.096568	0.9875	Thắng giải	Thắng giải
x_5	4.9837	0.9951	8.096568	0.9875	Thắng giải	Thắng giải
x_6	8.9535	0.9834	8.096568	0.9875	Không thắng giải	Không thắng giải
x_7	9.4386	0.9823	8.096568	0.9875	Không thắng giải	Không thắng giải

x_8	10.2329	0.9793	8.096568	0.9875	Không thắng giải	Không thắng giải
x_9	9.5519	0.9817	8.096568	0.9875	Không thắng giải	Không thắng giải
x_{10}	10.1469	0.9787	8.096568	0.9875	Không thắng giải	Không thắng giải

Bảng 3.8. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_2^1 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^1$	$\cos \theta_{c,r}^1$	R_2^1	$\cos \theta_2^1$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_2	4.0795	0.9966	8.096568	0.9875	0.9982	Thắng giải
x_1	5.2153	0.9944	8.096568	0.9875	0.9970	Thắng giải
x_5	4.9837	0.9951	8.096568	0.9875	0.9964	Thắng giải
x_3	6.166	0.9926	8.096568	0.9875	0.9949	Thắng giải
x_4	5.9835	0.9924	8.096568	0.9875	0.9933	Thắng giải

Bảng 3.9. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_7^3

Ứng viên $x_l \in X$	$d_{c,r}^3$	$\cos \theta_{c,r}^3$	R_7^3	$\cos \theta_7^3$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^3 \leq R_7^3) \wedge (\cos \theta_{c,r}^3 \geq \cos \theta_7^3)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_1	5.1145	0.9946	7.0425	0.9894	Thắng giải	Thắng giải
x_2	3.9746	0.9968	7.0425	0.9894	Thắng giải	Thắng giải
x_3	6.1655	0.9927	7.0425	0.9894	Thắng giải	Thắng giải
x_4	6.1293	0.992	7.0425	0.9894	Thắng giải	Thắng giải
x_5	4.9654	0.9952	7.0425	0.9894	Thắng giải	Thắng giải
x_6	9.0775	0.9831	7.0425	0.9894	Không thắng giải	Không thắng giải
x_7	9.5807	0.9818	7.0425	0.9894	Không thắng giải	Không thắng giải
x_8	10.6108	0.9776	7.0425	0.9894	Không thắng giải	Không thắng giải
x_9	9.6011	0.9816	7.0425	0.9894	Không thắng giải	Không thắng giải
x_{10}	10.3167	0.9781	7.0425	0.9894	Không thắng giải	Không thắng giải

Bảng 3.10. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_7^3 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^3$	$\cos \theta_{c,r}^3$	R_7^3	$\cos \theta_7^3$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_2	3.9746	0.9968	7.0425	0.9894	0.9982	Thắng giải
x_1	5.1145	0.9946	7.0425	0.9894	0.9970	Thắng giải
x_5	4.9654	0.9952	7.0425	0.9894	0.9964	Thắng giải
x_3	6.1655	0.9927	7.0425	0.9894	0.9949	Thắng giải
x_4	6.1293	0.992	7.0425	0.9894	0.9933	Thắng giải

Bảng 3.11. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 ứng viên bằng Winner-domain tốt D_5^4

Ứng viên $x_l \in X$	$d_{c,r}^4$	$\cos \theta_{c,r}^4$	R_5^4	$\cos \theta_5^4$	Tiếp cận Winner-domain dự đoán khả năng thắng giải ($d_{c,r}^4 \leq R_5^4$) $\wedge$ ($\cos \theta_{c,r}^4 \geq \cos \theta_5^4$)	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_1	5.252	0.9943	7.5164	0.9883	Thắng giải	Thắng giải
x_2	3.9945	0.9968	7.5164	0.9883	Thắng giải	Thắng giải
x_3	6.2854	0.9924	7.5164	0.9883	Thắng giải	Thắng giải
x_4	6.0282	0.9923	7.5164	0.9883	Thắng giải	Thắng giải
x_5	5.1153	0.9949	7.5164	0.9883	Thắng giải	Thắng giải
x_6	9.0467	0.9832	7.5164	0.9883	Không thắng giải	Không thắng giải
x_7	9.4341	0.9825	7.5164	0.9883	Không thắng giải	Không thắng giải
x_8	10.4278	0.9786	7.5164	0.9883	Không thắng giải	Không thắng giải
x_9	9.625	0.9816	7.5164	0.9883	Không thắng giải	Không thắng giải
x_{10}	10.2653	0.9784	7.5164	0.9883	Không thắng giải	Không thắng giải

Bảng 3.12. Đánh giá mức độ tương đồng và hướng năng lực của 05 ứng viên được chọn từ Winner-domain tốt D_5^4 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^4$	$\cos \theta_{c,r}^4$	R_5^4	$\cos \theta_5^4$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2023-2024
x_2	3.9945	0.9968	7.5164	0.9883	0.9982	Thắng giải
x_1	5.252	0.9943	7.5164	0.9883	0.9970	Thắng giải
x_5	5.1153	0.9949	7.5164	0.9883	0.9964	Thắng giải
x_3	6.2854	0.9924	7.5164	0.9883	0.9949	Thắng giải
x_4	6.0282	0.9923	7.5164	0.9883	0.9933	Thắng giải

Thực nghiệm với mô hình Winner-domain kết hợp Winner-cosin chọn được 05/10 ứng viên dự thi học sinh giỏi cấp quốc gia môn Tin học năm 2024 (năm học 2023-2024):

Bảng 3.13. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023- 2024 của những đội viên được chọn bằng mô hình Winner-domain kết hợp Winner-cosin

Số lượng đội viên dự thi	Số lượng đội viên được Winner-domain kết hợp Winner-cosin dự đoán thắng giải	Số lượng đội viên thực tế thắng giải	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Tỉ lệ đoán đúng không thắng giải: TNR
10	05	05	100%	0%	0%	100%

2. Thực nghiệm với mô hình Random Forest

Mô hình Random Forest được triển khai trên cùng Dataset (Winner Dataset và Nonwinner Dataset) và cùng tập năng lực của 10 ứng viên của kỳ thi 2024 (năm học 2023-2024).

Bảng 3.14. Kết quả dự đoán bằng mô hình Random Forest cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024

Đội viên $x_l \in X$	Dự đoán của mô hình Random Forest về khả năng thắng giải	Kết quả thi học sinh giỏi cấp quốc gia môn Tin học kỳ thi năm 2024 (năm học 2023-2024)
x_1	Thắng giải	Thắng giải
x_2	Thắng giải	Thắng giải

x_3	Thắng giải	Thắng giải
x_4	Thắng giải	Thắng giải
x_5	Thắng giải	Thắng giải
x_6	Thắng giải	Không thắng giải
x_7	Thắng giải	Không thắng giải
x_8	Thắng giải	Không thắng giải
x_9	Không thắng giải	Không thắng giải
x_{10}	Không thắng giải	Không thắng giải

Bước 1. Tạo tập con dữ liệu

Lấy mẫu ngẫu nhiên từ tập dữ liệu huấn luyện gốc (Dataset gồm 120 winners và 101 non-winners).

Bước 2. Huấn luyện

Với mỗi tập con dữ liệu và tại mỗi nút của cây, chọn ngẫu nhiên một tập con đặc trưng và tìm điểm phân tách tốt nhất trên tập con đặc trưng đó.

Bước 3. Dự đoán

- Thu thập các dự đoán từ tất cả các cây và chọn lớp được bỏ phiếu nhiều nhất.
- Sử dụng mô hình Random Forest đã được huấn luyện lại để dự đoán kết quả thắng giải cho 10 ứng viên dự thi học sinh giỏi cấp quốc gia môn Tin học (Bảng 3.14).

Bảng 3.15. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của những đội viên được dự đoán bằng mô hình Random Forest

Số lượng đội viên dự thi	Số lượng đội viên được Random Forest dự đoán thắng giải	Số lượng đội viên thực tế thắng giải	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Tỉ lệ đoán đúng không thắng giải: TNR
10	08	05	62.55%	0.6%	0%	0.4%

3. So sánh thực nghiệm kết quả kỳ thi học sinh giỏi cấp quốc gia năm học 2023-2024 của 2 mô hình

Kết quả thực nghiệm của mô hình Winner-domain kết hợp Winner-cosin được so sánh với kết quả thực nghiệm của mô hình Random Forest áp dụng cho cùng 1 tập ứng viên dự thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và cùng Dataset 2024.

Bảng 3.16. So sánh kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 của 10 đội viên đã chọn theo phương pháp truyền thống với quyết định chọn của mô hình Winner-domain kết hợp Winner-cosin và với kết quả dự đoán của mô hình Random Forest

Tiêu chí so sánh	Mô hình Winner-domain kết hợp Winner-cosin chọn đội viên	Mô hình Random Forest dự đoán ứng viên thắng giải
Đội viên đã được chọn theo mô hình truyền thống = Ứng viên input của các mô hình máy tính	10	
Số ứng viên được chọn làm đội viên hoặc dự đoán thắng giải	05	08
Số đội viên thắng giải thực tế	05	05
Tỉ lệ đánh giá đúng thắng giải: <i>TPR</i>	100%	62.55%
Tỉ lệ đánh giá nhầm thắng giải: <i>FPR</i>	0%	0.6%
Tỉ lệ bỏ sót: <i>FNR</i>	0%	0%
Tỉ lệ đánh giá hoặc đoán đúng không thắng giải: <i>TNR</i>	1%	0.4%

THỰC NGHIỆM KỲ THI NĂM HỌC 2024-2025

Thực nghiệm biến đổi 10 đội viên được lãnh đạo tỉnh và trường đã chọn bằng phương pháp truyền thống để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 thành những ứng viên của mô hình Winner-domain kết hợp Winner-cosin. Thực nghiệm sử dụng Dataset 2025 là dữ liệu năng lực một số thí sinh đã dự thi học sinh giỏi cấp quốc gia môn Tin học các năm 2021-2022, 2022-2023, 2023-2024 gồm 360 winners của Winner Dataset 2025 và 137 non-winners của Nonwinner Dataset 2025. Mô hình Winner-

domain kết hợp Winner-cosin do luận án đề xuất thực nghiệm trên 10 đội viên này như là những ứng viên của mô hình. Sau khi có kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025, luận án thực hiện thêm mô hình học máy Random Forest dự đoán kết quả của những đội viên do tỉnh chọn để so sánh với mô hình Winner-domain kết hợp Winner-cosin của luận án đã thực hiện trước với cùng một Dataset và cùng tập năng lực của các ứng viên (đội viên do tỉnh chọn).

1. Thực nghiệm với mô hình Winner-domain kết hợp Winner-cosin

Bước 1. Phát sinh Training set và Testing set

Winner Dataset 2025 được phân hoạch ngẫu nhiên 4 lần để được các Training set $W^p(w_c^p, \theta^p, R^p) \mid p=1,2,3,4$, mỗi Training set có 252 mẫu; và các Winner Testing set tương ứng $V^p \mid p=1,2,3,4$, mỗi Winner Testing set có 108 mẫu. Mỗi Testing set tương ứng với Training set gồm 108 mẫu của Winner Testing set tương ứng $V^p \mid p=1,2,3,4$ và 108 mẫu $U^p \mid p=1,2,3,4$ được phát sinh ngẫu nhiên tương ứng từ Nonwinner Dataset, Testing set = $V^p \cup U^p \mid p=1,2,3,4$ tương ứng với Training set $W^p \mid p=1,2,3,4$.

Bước 2. Phát sinh Winner-domain

Winner-domains được phát sinh từ những training sets. Trong thực nghiệm này, mỗi Training set $W^p(w_c^p, \theta^p, R^p) \mid p=1,2,3,4$ được phát sinh 08 Winner-domains $D_k^p(w_c^p, \theta_k^p, R_k^p \mid p=1,2,3,4 \mid k=1,2,...,8$. Trong đó, $R_k^p = \alpha_k \cdot R^p \mid k=1,2,...,8$. Mỗi Winner-domain được đánh giá WTR , TPR , FPR , FNR và $accuracy$.

Bảng 3.17. Kết quả phát sinh Winner-domains cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025

p	$\frac{R^p}{\cos \theta^p}$	k	α_k	R_k^p	$\cos \theta_k^p$	Tỉ lệ $winner \in D_k^p$ WTR	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Độ chính xác mẫu thắng được chọn: $accuracy$
1	$\frac{9.3563}{0.9816}$	1	0.77	7.204351	0.9884	0.876984	0.8704	0.0185	0.1296	0.979167
		2	0.75	7.017225	0.9892	0.849206	0.8611	0.0093	0.1389	0.989362

		3	0.73	6.830099	0.9898	0.642857	0.8611	0.0	0.1389	1.0
		4	0.71	6.642973	0.9903	0.611111	0.8426	0.0	0.1574	1.0
		5	0.7	6.54941	0.9905	0.587302	0.8426	0.0	0.1574	1.0
		6	0.69	6.455847	0.9911	0.571429	0.8426	0.0	0.1574	1.0
		7	0.68	6.362284	0.9911	0.56746	0.8426	0.0	0.1574	1.0
		8	0.67	6.268721	0.9915	0.551587	0.8426	0.0	0.1574	1.0
2	9.5701 / 0.9802	1	0.77	7.368977	0.988	0.694444	0.9444	0.037	0.0556	0.962264
		2	0.75	7.177575	0.9888	0.662698	0.9167	0.0278	0.0833	0.970588
		3	0.73	6.986173	0.9893	0.642857	0.8704	0.0093	0.1296	0.989474
		4	0.71	6.794771	0.9903	0.603175	0.8704	0.0	0.1296	1.0
		5	0.70	6.69907	0.9903	0.595238	0.8611	0.0	0.1389	1.0
		6	0.69	6.603369	0.9905	0.579365	0.8519	0.0	0.1481	1.0
		7	0.68	6.507668	0.991	0.559524	0.8426	0.0	0.1574	1.0
		8	0.67	6.411967	0.991	0.559524	0.8241	0.0	0.1759	1.0
3	9.5734 / 0.9802	1	0.77	7.371518	0.9879	0.694444	0.9259	0.1111	0.0741	0.892857
		2	0.75	7.18005	0.9884	0.654762	0.9074	0.0556	0.0926	0.942308
		3	0.73	6.988582	0.9892	0.638889	0.8889	0.0278	0.1111	0.969697
		4	0.71	6.797114	0.9897	0.607143	0.8611	0.0093	0.1389	0.989362
		5	0.70	6.70138	0.9902	0.599206	0.8333	0.0093	0.1667	0.989011
		6	0.69	6.605646	0.9907	0.587302	0.8333	0.0	0.1667	1.0
		7	0.68	6.509912	0.9907	0.583333	0.8241	0.0	0.1759	1.0
		8	0.67	6.414178	0.9913	0.56746	0.8148	0.0	0.1852	1.0
4	9.6581 / 0.9799	1	0.77	7.436737	0.9881	0.710317	0.8889	0.0648	0.1111	0.932039
		2	0.75	7.243575	0.9886	0.662698	0.8796	0.037	0.1204	0.959596
		3	0.73	7.050413	0.9891	0.654762	0.8519	0.0185	0.1481	0.978723
		4	0.71	6.857251	0.9901	0.638889	0.8148	0.0	0.1852	1.0
		5	0.70	6.76067	0.9903	0.630952	0.8148	0.0	0.1852	1.0
		6	0.69	6.664089	0.9905	0.623016	0.7963	0.0	0.2037	1.0
		7	0.68	6.567508	0.991	0.603175	0.787	0.0	0.213	1.0
		8	0.67	6.470927	0.9911	0.595238	0.7778	0.0	0.2222	1.0

Bước 3. Chọn Winner-domain tốt

Khảo sát tính chất của 32 Winner-domains được phát sinh, luận án chọn được 4 domains được cho là tốt hơn những domains khác.

Bảng 3.18. Kết quả chọn Winner-domains tốt cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025

D_k^p	d^p / $\cos \theta^p$	α_k	R_k^p	$\cos \theta_k^p$	Tỉ lệ $winner \in D_k^p$ WTR	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sọt: FNR	Đánh giá mức độ tốt
D_4^2	9.5701 / 0.9802	0.71	6.794771	0.9903	0.603175	0.8704	0.0	0.1296	1
D_3^1	9.3563 / 0.9816	0.73	6.830099	0.9898	0.642857	0.8611	0.0	0.1389	2
D_6^3	9.5734 / 0.9802	0.69	6.605646	0.9907	0.587302	0.8333	0.0	0.1667	3
D_4^4	9.6581 / 0.9799	0.71	6.857251	0.9901	0.638889	0.8148	0.0	0.1852	4

Bước 4. Dự đoán kết quả thi của 10 ứng viên

Với 10 ứng viên $c_r \in E \mid r=1,2,\dots,10$ được chọn theo phương pháp truyền thống để tham gia kỳ thi năm 2025 (năm học 2024-2025), luận án đã khảo sát năng lực bằng Questionnaire trước kỳ thi. Từ những hồi đáp khảo sát, mỗi ứng viên được biểu diễn như một vector trong không gian năng lực. Vector năng lực của mỗi ứng viên trước kỳ thi được đối sánh với 04 Winner-domain tốt $D_4^2, D_3^1, D_6^3, D_4^4$ để dự đoán kết quả.

Bảng 3.19. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_4^2

Ứng viên $x_l \in X$	$d_{c,r}^2$	$\cos \theta_{c,r}^2$	R_4^2	$\cos \theta_4^2$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^2 \leq R_4^2) \wedge (\cos \theta_{c,r}^2 \geq \cos \theta_4^2)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_1	6.6789	0.9911	6.794771	0.9903	Thắng giải	Thắng giải
x_2	6.1896	0.9921	6.794771	0.9903	Thắng giải	Thắng giải
x_3	5.7225	0.9938	6.794771	0.9903	Thắng giải	Thắng giải
x_4	5.0968	0.9946	6.794771	0.9903	Thắng giải	Thắng giải
x_5	6.6512	0.991	6.794771	0.9903	Thắng giải	Thắng giải
x_6	4.1603	0.9966	6.794771	0.9903	Thắng giải	Thắng giải
x_7	8.9596	0.9837	6.794771	0.9903	Không thắng giải	Không thắng giải
x_8	9.8723	0.9802	6.794771	0.9903	Không thắng giải	Không thắng giải
x_9	9.8563	0.9813	6.794771	0.9903	Không thắng giải	Không thắng giải
x_{10}	9.9016	0.9808	6.794771	0.9903	Không thắng giải	Không thắng giải

Bảng 3.20. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_3^1 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^1$	$\cos \theta_{c,r}^1$	R_3^1	$\cos \theta_3^1$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_6	4.2134	0.9965	6.830099	0.9898	0.9966	Thắng giải
x_4	5.2136	0.9943	6.830099	0.9898	0.9946	Thắng giải
x_3	5.7487	0.9937	6.830099	0.9898	0.9938	Thắng giải
x_2	6.2077	0.992	6.830099	0.9898	0.9921	Thắng giải
x_1	6.5222	0.9915	6.830099	0.9898	0.9911	Thắng giải
x_5	6.5609	0.9913	6.830099	0.9898	0.9910	Thắng giải

Bảng 3.21. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_4^2

Ứng viên $x_l \in X$	$d_{c,r}^2$	$\cos \theta_{c,r}^2$	R_4^2	$\cos \theta_4^2$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^2 \leq R_4^2) \wedge (\cos \theta_{c,r}^2 \geq \cos \theta_4^2)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_1	6.6789	0.9911	6.794771	0.9903	Thắng giải	Thắng giải
x_2	6.1896	0.9921	6.794771	0.9903	Thắng giải	Thắng giải
x_3	5.7225	0.9938	6.794771	0.9903	Thắng giải	Thắng giải
x_4	5.0968	0.9946	6.794771	0.9903	Thắng giải	Thắng giải
x_5	6.6512	0.991	6.794771	0.9903	Thắng giải	Thắng giải
x_6	4.1603	0.9966	6.794771	0.9903	Thắng giải	Thắng giải
x_7	8.9596	0.9837	6.794771	0.9903	Không thắng giải	Không thắng giải
x_8	9.8723	0.9802	6.794771	0.9903	Không thắng giải	Không thắng giải
x_9	9.8563	0.9813	6.794771	0.9903	Không thắng giải	Không thắng giải
x_{10}	9.9016	0.9808	6.794771	0.9903	Không thắng giải	Không thắng giải

Bảng 3.22. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_4^2 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^2$	$\cos \theta_{c,r}^2$	R_4^2	$\cos \theta_4^2$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_6	4.1603	0.9966	6.794771	0.9903	0.9965	Thắng giải
x_4	5.0968	0.9946	6.794771	0.9903	0.9943	Thắng giải
x_3	5.7225	0.9938	6.794771	0.9903	0.9937	Thắng giải
x_2	6.1896	0.9921	6.794771	0.9903	0.992	Thắng giải
x_1	6.6789	0.991	6.794771	0.9903	0.9915	Thắng giải
x_5	6.6512	0.991	6.794771	0.9903	0.9913	Thắng giải

Bảng 3.23. Đánh giá kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 ứng viên bằng Winner-domain tốt D_6^3

Ứng viên $x_l \in X$	$d_{c,r}^3$	$\cos \theta_{c,r}^3$	R_6^3	$\cos \theta_6^3$	Tiếp cận Winner-domain dự đoán khả năng thắng giải $(d_{c,r}^3 \leq R_6^3) \wedge (\cos \theta_{c,r}^3 \geq \cos \theta_6^3)$	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_1	6.5495	0.9916	6.605646	0.9907	Thắng giải	Thắng giải
x_2	6.1337	0.9923	6.605646	0.9907	Thắng giải	Thắng giải
x_3	5.7245	0.994	6.605646	0.9907	Thắng giải	Thắng giải
x_4	5.1956	0.9944	6.605646	0.9907	Thắng giải	Thắng giải
x_5	6.5808	0.9913	6.605646	0.9907	Thắng giải	Thắng giải
x_6	4.161	0.9966	6.605646	0.9907	Thắng giải	Thắng giải
x_7	8.9834	0.9838	6.605646	0.9907	Không thắng giải	Không thắng giải
x_8	9.912	0.9802	6.605646	0.9907	Không thắng giải	Không thắng giải
x_9	9.9868	0.981	6.605646	0.9907	Không thắng giải	Không thắng giải
x_{10}	9.9515	0.9808	6.605646	0.9907	Không thắng giải	Không thắng giải

Bảng 3.24. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_6^3 theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^3$	$\cos \theta_{c,r}^3$	R_6^3	$\cos \theta_6^3$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_6	4.161	0.9966	6.605646	0.9907	0.9966	Thắng giải
x_4	5.1956	0.9944	6.605646	0.9907	0.9944	Thắng giải
x_3	5.7245	0.994	6.605646	0.9907	0.994	Thắng giải
x_2	6.1337	0.9923	6.605646	0.9907	0.9923	Thắng giải
x_l	6.5495	0.9916	6.605646	0.9907	0.9916	Thắng giải
x_5	6.5808	0.9913	6.605646	0.9907	0.9913	Thắng giải

Bảng 3.25. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của những đội viên được chọn bằng mô hình Winner-domain kết hợp Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^4$	$\cos \theta_{c,r}^4$	R_4^4	$\cos \theta_4^4$	Tiếp cận Winner-domain đánh giá khả năng thắng giải $(d_{c,r}^4 \leq R_4^4) \wedge (\cos \theta_{c,r}^4 \geq \cos \theta_4^4)$	Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025
x_l	6.5575	0.9914	6.857251	0.9901	Thắng giải	Thắng giải
x_2	6.0974	0.9923	6.857251	0.9901	Thắng giải	Thắng giải
x_3	5.7256	0.9937	6.857251	0.9901	Thắng giải	Thắng giải
x_4	5.0647	0.9946	6.857251	0.9901	Thắng giải	Thắng giải
x_5	6.5242	0.9913	6.857251	0.9901	Thắng giải	Thắng giải
x_6	4.2287	0.9964	6.857251	0.9901	Thắng giải	Thắng giải
x_7	8.8427	0.9841	6.857251	0.9901	Không thắng giải	Không thắng giải
x_8	9.7693	0.9805	6.857251	0.9901	Không thắng giải	Không thắng giải
x_9	9.875	0.9811	6.857251	0.9901	Không thắng giải	Không thắng giải
x_{l0}	9.7845	0.9812	6.857251	0.9901	Không thắng giải	Không thắng giải

Bảng 3.26. Đánh giá mức độ tương đồng về hướng năng lực của 06 ứng viên được chọn từ Winner-domain tốt D_4^A theo thuật toán Winner-cosin

Ứng viên $x_l \in X$	$d_{c,r}^4$	$\cos \theta_{c,r}^4$	R_4^4	$\cos \theta_4^4$	Đánh giá mức độ tương đồng Winner-cosin	Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025
x_6	4.2287	0.9964	6.857251	0.9901	0.9964	Thắng giải
x_4	5.0647	0.9946	6.857251	0.9901	0.9946	Thắng giải
x_3	5.7256	0.9937	6.857251	0.9901	0.9937	Thắng giải
x_2	6.0974	0.9923	6.857251	0.9901	0.9923	Thắng giải
x_1	6.5575	0.9914	6.857251	0.9901	0.9914	Thắng giải
x_5	6.5242	0.9913	6.857251	0.9901	0.9913	Thắng giải

Bảng 3.27. Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của những đội viên được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin

Số lượng đội viên dự thi	Số lượng đội viên được Winner-domain kết hợp Winner-cosin dự đoán thắng giải	Số lượng đội viên thực tế thắng giải	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Tỉ lệ đoán đúng không thắng giải: TNR
10	06	06	100%	0%	0%	1%

2. Thực nghiệm với mô hình Random Forest

Mô hình Random Forest được triển khai trên cùng Dataset (Winner Dataset và Nonwinner Dataset) và cùng tập năng lực của 10 ứng viên của kỳ thi năm học 2024-2025.

Bước 1. Tạo tập con dữ liệu

Lấy mẫu ngẫu nhiên từ tập dữ liệu huấn luyện gốc (Dataset gồm 360 winners và 137 non-winners).

Bước 2. Huấn luyện

Với mỗi tập con dữ liệu và tại mỗi nút của cây, chọn ngẫu nhiên một tập con đặc trưng và tìm điểm phân tách tốt nhất trên tập con đặc trưng đó.

Bước 3. Dự đoán

- Thu thập các dự đoán từ tất cả các cây và chọn lớp được bỏ phiếu nhiều nhất.
- Sử dụng mô hình Random Forest đã được huấn luyện lại để dự đoán kết quả 'thắng giải' cho 10 ứng viên dự thi học sinh giỏi cấp quốc gia môn Tin học (Bảng 4.28).

Bảng 3.28. Kết quả dự đoán bằng mô hình Random Forest cho kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025

Ứng viên $x_i \in X$	Dự đoán của mô hình Random Forest về khả năng thắng giải	Kết quả thi học sinh giỏi cấp quốc gia năm học 2024-2025
x_1	Thắng giải	Thắng giải
x_2	Thắng giải	Thắng giải
x_3	Thắng giải	Thắng giải
x_4	Thắng giải	Thắng giải
x_5	Thắng giải	Thắng giải
x_6	Thắng giải	Thắng giải
x_7	Thắng giải	Không thắng giải
x_8	Không thắng giải	Không thắng giải
x_9	Không thắng giải	Không thắng giải
x_{10}	Thắng giải	Không thắng giải

Bảng 3.29. Thực nghiệm mô hình Random Forest dự đoán kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 đội viên đã chọn bằng phương pháp truyền thống

Số lượng đội viên dự thi	Số lượng đội viên được Random Forest dự đoán thắng giải	Số lượng đội viên thực tế thắng giải	Tỉ lệ đoán đúng thắng giải: TPR	Tỉ lệ đoán nhầm thắng giải: FPR	Tỉ lệ bỏ sót: FNR	Tỉ lệ đoán đúng không thắng giải: TNR
10	08	06	75.0%	0.5%	0%	0.5%

3. So sánh thực nghiệm kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 2 mô hình

Đối sánh với kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 với các kết quả của mô hình Winner-domain kết hợp Winner-cosin và mô hình Random Forest được thực nghiệm trên cùng Dataset 2025 và cùng tập ứng viên là những đội viên do tỉnh chọn.

Bảng 3.30. So sánh kết quả kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2024-2025 của 10 đội viên đã chọn theo phương pháp truyền thống với quyết định chọn của mô hình Winner-domain kết hợp Winner-cosin và với kết quả dự đoán của mô hình Random Forest

Tiêu chí so sánh	Mô hình Winner-domain kết hợp Winner-cosin	Mô hình Random Forest
	10	
Số ứng viên được chọn làm đội viên hoặc dự đoán thắng giải	06	08
Số đội viên thực tế thắng giải	06	06
Tỉ lệ chọn đúng hoặc đoán đúng thắng giải: <i>TPR</i>	100%	75.0%
Tỉ lệ chọn nhầm hoặc đoán nhầm thắng giải: <i>FPR</i>	0%	0.5%
Tỉ lệ bỏ sót: <i>FNR</i>	0%	0%
Tỉ lệ đánh giá hoặc đoán đúng không thắng giải: <i>TNR</i>	1%	0.5%

3.5. Kết luận chương 3

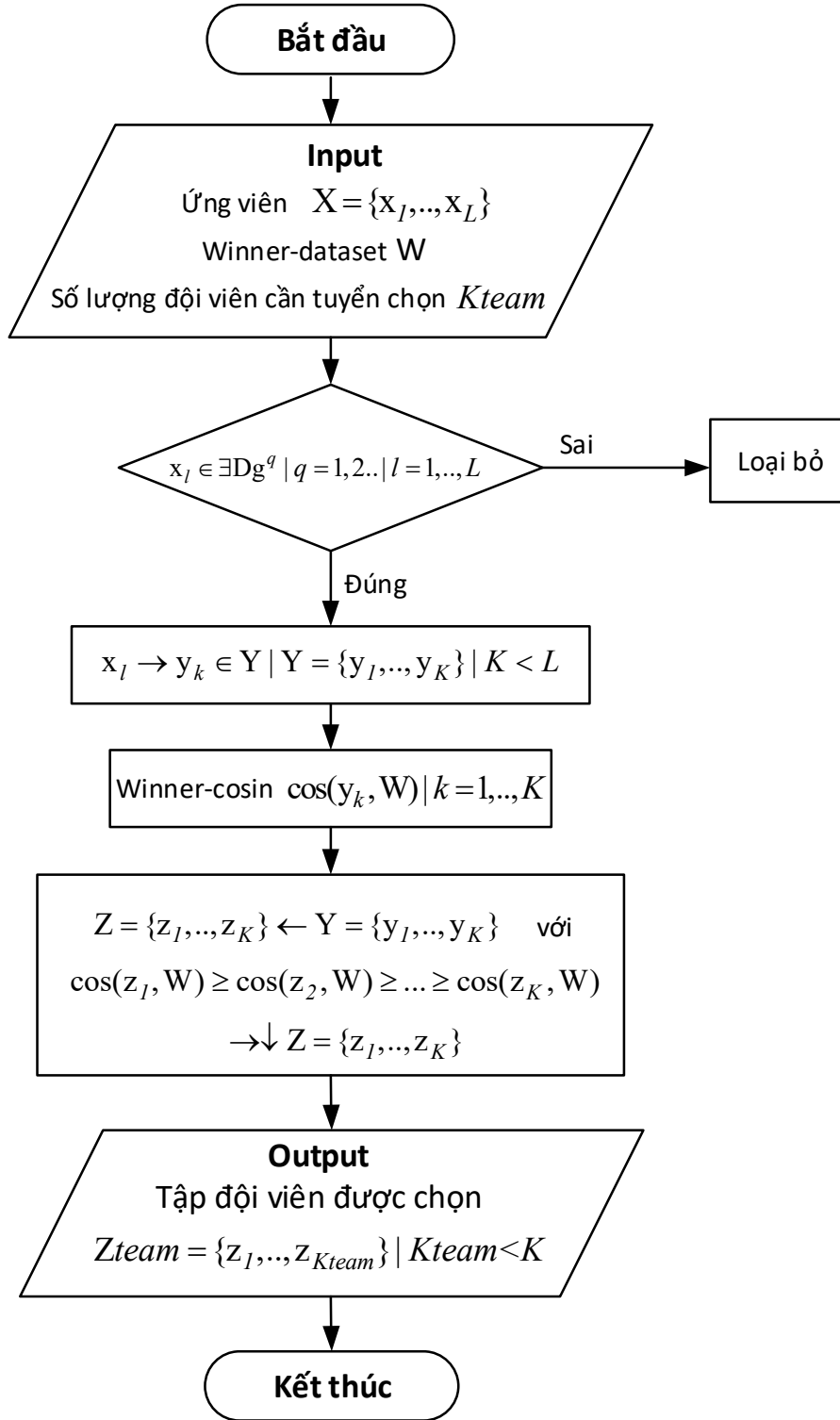
Áp dụng tiếp cận học máy, chương 3 đã xây dựng sơ đồ quyết định chọn những học sinh ứng viên vào đội tuyển trên cơ sở mô hình hóa năng lực mỗi học sinh là một vector nhiều chiều, mỗi chiều biểu diễn một biến năng lực trong không gian năng lực nhiều chiều. Luận án đề xuất các tiếp cận đánh giá độ tương đồng của ứng viên với các winners của vài năm trước gần nhất:

- Tiếp cận Winner-domain đánh giá độ tương đồng về suất năng lực.
- Tiếp cận tương đồng cosin đánh giá độ tương đồng hướng năng lực.

Với số lượng ứng viên rất nhiều so với số lượng đội viên cần tuyển, luận án không gán nhãn từng ứng viên mà áp dụng chiến lược “thà loại bỏ hơn chọn nhầm” để quyết định chọn ứng viên có khả năng thắng giải cao vào đội tuyển. Nghiên cứu đã được áp dụng thực nghiệm tại một trường trung học phổ thông có kết quả tốt so với những năm trước, khi chưa áp dụng kết quả nghiên cứu của luận án. Kết quả thực nghiệm được thống kê trong Bảng 3.31.

Những kết quả nghiên cứu của luận án đã được áp dụng, có đối sánh với phương pháp tuyển chọn và huấn luyện truyền thống cho đội tin học của một trường trung học phổ thông với kết quả kỳ thi học sinh giỏi cấp quốc gia các năm học 2023-2024, 2024-2025. Theo kết quả trình bày trong phụ lục luận án, những đội viên môn Tin học được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin được thắng giải 100% (5/5, 6/6) trong 2 kỳ thi học sinh giỏi cấp quốc gia năm học 2023-2024, 2024-2025, trong khi những đội viên được tuyển chọn bằng phương pháp truyền thống cùng được huấn luyện và dự thi đồng thời thắng giải 50% (5/10) năm học 2023-2024, 60% (6/10) năm học 2024-2025. Trước đó, khi chưa thực nghiệm mô hình Winner-domain kết hợp Winner-cosin, các đội viên được tuyển chọn bằng phương pháp truyền thống cùng được huấn luyện theo phương pháp truyền thống và dự thi thì năm học 2019-2020 thắng giải 0% (0/6), năm học 2020-2021 thắng giải 16,67% (1/6), năm học 2021-2022 thắng giải 50% (3/6), năm học 2022-2023 thắng giải 16,67% (1/6).

Kết quả thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và 2024-2025 của những đội viên được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin cũng được so sánh với mô hình Random Forest trên cùng tập ứng viên và cùng Dataset 2024 và Dataset 2025 cho thấy Tỷ lệ chọn đúng thắng giải của mô hình Winner-domain kết hợp Winner-cosin cao hơn tỷ lệ đoán đúng của mô hình Random Forest (Bảng 3.32)



Hình 3.2. Quy trình chọn đội viên từ tập ứng viên

Bảng 3.31. Tóm tắt so sánh kết quả thực nghiệm của luận án với kết quả của những phương pháp truyền thống

Năm học	Số lượng đội viên dự thi học sinh giỏi cấp quốc gia môn tin học	Đội viên được tuyển chọn bằng phương pháp truyền thống để dự thi học sinh giỏi cấp quốc gia môn tin học				Đội viên được tuyển chọn bằng mô hình Winner-domain kết hợp Winner-cosin để dự thi học sinh giỏi cấp quốc gia môn tin học				Số lượng đội viên đạt giải học sinh giỏi cấp quốc gia môn tin học	Ghi chú
		Số lượng	Đạt giải học sinh giỏi cấp quốc gia		Không đạt giải học sinh giỏi cấp quốc gia	Số lượng	Đạt giải học sinh giỏi cấp quốc gia		Không đạt giải học sinh giỏi cấp quốc gia		
			Số lượng	Tỉ lệ %			Số lượng	Tỉ lệ %			
2019-2020	6	6	0	0%	6	100%				0	
2020-2021	6	6	1	16,67%	5	83,33%				1	
2021-2022	6	6	3	50%	3	50%				3	
2022-2023	6	6	1	16,67%	5	83,33%				1	
2023-2024	10	10	5	50%	5	50%	5	100%	0	5	Mô hình Winner-domain kết hợp Winner-cosin
2024-2025	10	10	6	60%	4	40%	6	100%	0	6	

Thực nghiệm áp dụng cho kỳ thi năm học 2023-2024 với dataset 2024 và tập ứng viên 2024 là những đội viên 2024 đã chọn theo phương pháp truyền thống, kỳ thi năm học 2024-2025 với dataset 2025 và tập ứng viên 2025 là những đội viên 2025 đã chọn theo phương pháp truyền thống. Kết quả được so sánh với 3 mô hình: - mô hình truyền thống, mô hình Winner-domain kết hợp Winner-cosin, mô hình Random Forest theo Bảng 3.32 như sau:

Bảng 3.32. So sánh kết quả các mô hình tuyển chọn và dự đoán kết quả của 2 kỳ thi học sinh giỏi cấp quốc gia môn Tin học năm học 2023-2024 và năm học 2024-2025

Tiêu chí so sánh	2023-2024		2024-2025	
Kỳ thi năm học	Mô hình Winner-domain kết hợp Winner-cosin	Mô hình Random Forest	Mô hình Winner-domain kết hợp Winner-cosin	Mô hình Random Forest
	10		10	
Số ứng viên được chọn làm đội viên hoặc dự đoán thắng giải	05	08	06	08
Số đội viên thực tế thắng giải	05	05	06	06
Tỉ lệ chọn đúng hoặc đoán đúng thắng giải: <i>TPR</i>	100%	62.55%	100%	75.0%
Tỉ lệ chọn nhầm hoặc đoán nhầm thắng giải: <i>FPR</i>	0%	0.6%	0%	0.5%
Tỉ lệ bỏ sót: <i>FNR</i>	0%	0%	0%	0%
Tỉ lệ đánh giá hoặc đoán đúng không thắng giải: <i>TNR</i>	1%	0.4%	1%	0.5%

CHƯƠNG 4. TIẾP CẬN HỌC KHÁM PHÁ CÓ HƯỚNG DẪN TRỰC QUAN ĐỂ HUẤN LUYỆN ĐỘI TUYỂN

4.1. Học khám phá có hướng dẫn

4.1.1. Giới thiệu

Ngày nay, hầu hết mô hình học khám phá có hướng dẫn đều tích hợp học khám phá có hướng dẫn với học cộng tác (collaborative learning) để kết hợp các đặc trưng chính của học khám phá có hướng dẫn và học cộng tác. Học khám phá kết hợp với học cộng tác có 2 hoặc nhiều người cùng học sẽ chia sẻ và xây dựng kiến thức khi làm việc cùng nhau hướng tới lời giải cho bài tập hoặc bài toán [60]. Học khám phá cộng tác có hướng dẫn sử dụng truyền thông để tranh luận, để chia sẻ nên đòi hỏi những người học phải xây dựng nền tảng chung trước khi làm việc cùng nhau [60]. Vì những lợi ích khi kết hợp học khám phá có hướng dẫn với học cộng tác nên khi phát triển học khám phá có hướng dẫn cũng bao gồm luôn thuộc tính của học cộng tác.

Học khám phá có hướng dẫn cải thiện chất lượng quy trình học khám phá [60]. Trong học khám phá có hướng dẫn, quy trình truyền thông góp phần kiến tạo kiến thức mới và xây dựng nền tảng chung là tranh luận và kiểm tra thông tin, hỏi đáp để xác thực [60]. Trong khi tài liệu học tập không trình bày dạng cuối cùng, người học được đòi hỏi phải thực hiện những hoạt động khác nhau để thu thập thông tin, so sánh, phân loại, phân tích, tích hợp, tổ chức tài liệu và làm kết luận [97]. Học khám phá có hướng dẫn và kiến thức có sẵn của người học có quan hệ nhau, kiến thức có sẵn tác động đến sự phát triển kiến thức khi học khám phá có hướng dẫn [58].

Lý thuyết Piaget về phát triển nhận biết cho rằng người học không thể tự xử lý, tự hiểu những thông tin nhận được [98]. Do đó, các cơ sở giáo dục hiện đang phát triển tiếp cận học khám phá với sự hướng dẫn của thầy [44, 74, 87]. Khi thúc đẩy triển khai mô hình học khám phá trong các cơ sở giáo dục, những vấn đề cần được giải quyết ưu tiên là xác định các hợp phần của hệ thống, định nghĩa các chức năng và quan hệ của chúng và quan trọng là phải tổ chức lớp học, thiết kế bài giảng phù hợp với khả năng những người cùng học tiếp thu thông tin và kiến thức mới, đồng thời phải đánh giá hiệu năng của hệ thống.

4.1.2. Lợi ích của học khám phá có hướng dẫn

Mô hình học khám phá có hướng dẫn được đánh giá là tốt hơn mô hình học truyền thống vì nó có nhiều ý nghĩa đối với người học hơn là nhận thông tin một cách đơn giản một chiều từ người dạy [47]. Trong khi đó, mô hình học khám phá không hướng dẫn có thể thất bại vì không hỗ trợ người học tích cực xây dựng kiến thức để hiểu thông tin mới và tích hợp thông tin mới nhận được cho đến khi phân tích ra kiến thức đúng [46]. Mô hình học khám phá có hướng dẫn có thể hỗ trợ người học làm đầy 2 đòi hỏi quan trọng này [46].

Mô hình học khám phá có hướng dẫn có những lợi ích tích cực đối với kết quả học tập của người học [47]:

- Học khám phá giúp người học đương đầu với những ý tưởng hiện hành về những chủ đề mà nhiều phần trong đó có thể là quan niệm sai. Từ đó, người học hóa giải được những quan niệm sai nhờ vào những gì mình quan sát được;
- Đối với người học, khi học khám phá về một lĩnh vực thì dễ hiểu vì những nhiệm vụ khám phá dựa trên những bài toán thực hoặc tình huống thực. Do đó, nội dung học khám phá dễ biểu diễn trực quan và người học có thể nhận ra những nguyên lý khi học và có cách tốt hơn để hình dung cái gì đang xảy ra với một ý tưởng;
- Học khám phá áp dụng học tình huống, học theo ngữ cảnh, với nội dung được học không chỉ là thông tin mà còn là kiến thức được phát sinh từ các tình huống, từ các ngữ cảnh;
- Trong học khám phá, người học tự khám phá ra khái niệm và thông tin mình chưa biết để có kiến thức mới rồi tự xác định làm thế nào để hiểu và sử dụng chúng;
- Học khám phá khuyến khích người học tự đặt câu hỏi, tự giải bài toán, và tự phát triển để làm tăng khả năng của mình trong việc xử lý bài toán. Kết quả là học khám phá khuyến khích người học đi xa hơn trong việc hình thành tri thức của riêng mình [47];
- Học khám phá có hướng dẫn giúp người học tự tin hơn trong cách tiếp cận để giải bài toán [47];

- Học khám phá giúp người học tự khám phá, tìm hiểu những gì cần biết, cần học, không phụ thuộc vào người khác hoặc phụ thuộc duy nhất vào một tài liệu giáo khoa;
- Học khám phá hỗ trợ quan điểm cho rằng khoa học là một quy trình, không phải là những sự kiện vì học khám phá được thiết kế để cung cấp cơ hội và dạy cho người học những bước trong quy trình;
- Học khám phá đặt nhiều trách nhiệm học cho người học. Khi người học được đặt trong vị trí phải chỉ ra bài toán thì người học có nhiều trách nhiệm cho việc học của chính mình không phải phụ thuộc vào người khác.

4.1.3. Mô hình học khám phá có hướng dẫn

Mô hình học khám phá có hướng dẫn kết hợp các mô hình học tích cực là mô hình kết nối 4 hợp phần: người học, tài liệu học tập, người dạy, môi trường học. Phương pháp học khám phá được chấp nhận dựa trên trình bày năng lực và hoạch định hoạt động. Người học theo phương pháp học khám phá nắm vững nội dung tài liệu học tập, áp dụng chiến lược học, áp dụng phương pháp khoa học, dùng tài nguyên học tập trong quy trình học, dùng ngôn ngữ chính xác và thích hợp để kết thúc hoạt động học tập [59].

4.1.3.1. Người học

Học khám phá có hướng dẫn kết hợp những đặc trưng của các mô hình học tích cực, trong đó người học là một thành viên tham gia tích cực vào quy trình học, người học không đóng vai một thùng rỗng được làm đầy bởi người dạy [47]. Mô hình học khám phá cho phép người học khám phá ra quy luật mới và ý tưởng mới, không phải chỉ nhớ cái gì người dạy đã giảng [46]. Người học có tư duy sáng tạo tốt sẽ có nhiều khả năng tốt để giải bài toán [46]. Khi thực hiện tường minh quy trình học khám phá, một câu hỏi của một người học có thể tự nó đóng góp vào việc khám phá [99].

Nhiều người học cùng chủ đề có thể được tổ chức thành những nhóm cộng tác để hỗ trợ quy trình học khám phá có hướng dẫn [99]. Những người học cùng nhóm cải thiện hiệu năng quy trình học khám phá bằng cách cùng nhau phát sinh giả thiết và thiết kế thí nghiệm [99]. Mô hình học khám phá có hướng dẫn nhấn mạnh những chức năng của mô hình học cộng tác để kết quả

học tập của người học được nâng lên do tác động của cộng tác và những lợi ích của thông tin rút ra từ quy trình khám phá [99].

Người học trong mô hình học khám phá cần được hỗ trợ giải thích, hỗ trợ thực nghiệm, hỗ trợ phản hồi. Người học cần được hỗ trợ giải thích để truy cập và kích hoạt kiến thức, phát sinh giả thiết thích hợp và xây dựng sự hiểu biết mạch lạc. Người học cần được hỗ trợ thực nghiệm để tạo khung cho thiết kế thực nghiệm khoa học có tính hệ thống và logic, dự báo và quan sát kết quả và rút ra kết luận hợp lý. Người học cần được hỗ trợ phản hồi để tăng sự tự nhận biết của mình trong quy trình học và tích hợp các tóm tắt phản hồi vào những khám phá của mình [100].

4.1.3.2. Tài liệu học tập

Sự kết nối kiến thức mới với kiến thức có sẵn được chỉ ra bởi người hướng dẫn hoặc được khám phá bởi người học [47]. Tài liệu học tập cần thiết và có ý nghĩa trong hoạt động dạy và học ở trường để cải thiện hiệu năng người dạy và kết quả học của người học [44]. Theo lý thuyết sơ đồ [101], người học thu thập dữ liệu vào bộ nhớ ngắn hạn (short-term memory) đều được chuyển tất cả sang dạng hình ảnh để được hiểu và lưu trữ vào bộ nhớ dài hạn (long-term memory) [102]. Người dạy cần gợi ý để người học tư duy và giải thích những khái niệm khác nhau bằng hình ảnh trực quan và thay đổi phương pháp trực quan tùy theo đối tượng người học tại những giai đoạn khác nhau [71].

Bài giảng theo mô hình học truyền thống không thể áp dụng cho mô hình học khám phá có hướng dẫn. Tài liệu học tập dùng trong mô hình học khám phá có hướng dẫn nhằm nâng cao năng lực với các kỹ năng tư duy sáng tạo, tư duy phản biện và tính tự tin, tự lực để khám phá ra thông tin mới hoặc những khái niệm mà người học chưa được biết trước đó. Lý thuyết dual coding chia quy trình nhận biết làm 2 hệ thống xử lý, hình ảnh và chữ viết [102]. Do đó, tài liệu học tập dùng trong mô hình học khám phá được biên soạn kết hợp ngôn ngữ hình ảnh và chữ viết. Mô phỏng trên máy tính là một công cụ mạnh để hướng dẫn học khám phá trong các lĩnh vực khoa học [100].

4.1.3.3. Người dạy

Trong phương pháp học khám phá có hướng dẫn, người dạy cần chuẩn bị tốt cho hoạt động học tập của người học để đạt đến mục tiêu học đã định trước [59]. Trong môi trường học khám phá có hướng dẫn, những hoạt động

của người dạy phải phù hợp với bài giảng (tài liệu học tập) đã soạn trước [59]. Để giúp cho người học học một cách sáng tạo, người dạy khuyến khích, hướng dẫn người học tìm hiểu, đặt câu hỏi, xử lý, nghe và kiểm tra bài toán [46]. Người dạy đòi hỏi người học giải thích ý tưởng riêng của họ, rồi mình xác nhận đúng /sai của ý tưởng đó ngay lập tức [44].

Người dạy hỗ trợ và hướng dẫn người học phát huy tư duy sáng tạo vì tư duy sáng tạo giúp cho người học giải bài toán dễ hơn [46]. Người dạy có kỹ năng sáng tạo tốt mới có thể phát triển tư duy sáng tạo của người học [46]. Người dạy có kỹ năng sáng tạo tốt mới có thể tìm ra những chiến lược mới để phát triển tiềm năng của người học [46]. Người dạy phải cố gắng thực hiện môi trường học dễ chịu, vui vẻ giúp cho người học khai thác tất cả khả năng đang có [46].

Người dạy trợ giúp người học trong môi trường học khám phá có hướng dẫn để thực hiện dạy theo thời gian, nội dung và loại phản hồi [103]. Về thời gian, người dạy hỗ trợ người học bằng cách phản hồi yes/no nhanh ngay khi có câu hỏi từ người học. Về nội dung, người dạy phản hồi cho người học nội dung nào nên tiến hành hoặc không nên tiến hành. Về loại phản hồi người dạy gợi ý cho người học những gì nên hay không nên, khi nào và bao nhiêu [103].

Người dạy trợ giúp người học thực nghiệm trong môi trường học khám phá có hướng dẫn theo 3 mức, tối thiểu, trung bình, cao. Trợ giúp mức tối thiểu, người dạy chỉ hỏi đáp người học những nội dung thực nghiệm rất cơ bản. Trợ giúp mức trung bình, người dạy trợ giúp người học những nội dung thực nghiệm theo yêu cầu người học hoặc khi nhận thấy người học thực nghiệm sai. Trợ giúp mức cao, người dạy thuyết phục người học thực nghiệm nội dung học được để nhận kết quả [103].

Người dạy hỗ trợ người học thực hiện các phương pháp khoa học tương ứng với các giai đoạn của mô hình học khám phá [59]:

- *Quan sát*: Môi trường học hỗ trợ người học quan sát các mô hình mô phỏng chủ đề học tập được thiết kế đẹp mắt, thu hút, dễ khảo sát, gợi sự tò mò để khám phá.
- *Đặt câu hỏi*: Người dạy đặt ra câu hỏi để tạo ra cảm hứng và dẫn tư duy của người học đi dần đến nhận dạng và phát biểu bài toán.

- *Lập luận*: Người dạy hỗ trợ người học bằng cách gợi ý một logic lập luận để tìm dữ liệu, khám phá dần thông tin và các khái niệm ẩn chứa trong mô hình.
- *Cố gắng*: Người dạy khuyến khích và vận động sự cố gắng của từng người học trong suốt quy trình học, nhất là trong giai đoạn xử lý dữ liệu.
- *Phối hợp*: Người dạy hướng dẫn người học phối hợp các vấn đề đã xử lý để xác thực và chứng minh kết quả.
- *Trình bày*: Người dạy tạo ra sự hứng thú trong không gian an toàn và hướng dẫn người học trình bày những nội dung đã thảo luận và kết quả có được.

4.1.3.4. Môi trường học

Học khám phá có hướng dẫn là học theo mẫu phương pháp khoa học để những người học trong những nhóm tìm cách giải bài toán từ mô phỏng, nhận dạng bài toán, thu thập dữ liệu, xử lý dữ liệu, xác thực, rồi rút ra kết luận [44]. Môi trường học cần thực sự an toàn cho người học sáng tạo để người học cảm thấy dễ chịu và yên tâm khi trình bày ý kiến của mình [46].

Các thuộc tính của mô hình học khám phá có hướng dẫn chỉ ra kiến trúc của môi trường học tập [99]:

- *Khung tham chiếu*: Trong môi trường học khám phá có hướng dẫn, 2 hoặc nhiều người học cùng nhau thực hiện nhiệm vụ khám phá bằng cách làm thực nghiệm, suy ra kết quả và chia sẻ suy nghĩ thông qua truyền thông, thảo luận.
- *Công cụ học tập*: Công cụ cộng tác, diễn đàn thảo luận, bảng trắng, kênh “chat” là những dụng cụ hỗ trợ cộng tác, chúng tham gia như một hợp phần quản lý sự tương tác giữa những người học, hỗ trợ người học truyền thông và truyền thông giữa những hợp phần của kiến trúc.
- *Không gian thực nghiệm*: Không gian thực nghiệm biểu diễn miền nhiệm vụ của học khám phá có hướng dẫn, không gian thực nghiệm cũng là nguồn chính cung cấp thông tin cấu trúc khung tham chiếu với những định nghĩa các biến trong miền và cũng là mô hình miền.

4.2. Học khám phá có hướng dẫn trực quan

Lý thuyết sơ đồ chỉ ra rằng con người lưu trữ kiến thức như một mạng sơ đồ 3D [104] trong bộ nhớ dài hạn, con người nhận biết ý nghĩa của dữ liệu mới bằng cách đối sánh với kiến thức dạng sơ đồ có sẵn trong bộ nhớ dài hạn [101, 104-106]. Dữ liệu mới nhận được từ các giác quan truyền đến bộ nhớ ngắn hạn (short-term memory) được biến đổi thành thông tin dạng sơ đồ và đối sánh với kiến thức dạng sơ đồ được lưu trữ trước trong bộ nhớ dài hạn để hiểu; nếu sơ đồ của thông tin mới tương thích với một sơ đồ nào đó trong bộ nhớ dài hạn thì thông tin này được hiểu dễ dàng; trái lại thì thông tin mới không được hiểu hoặc hiểu không đúng [101].

Trong học tập, người học sau khi tiếp nhận bài giảng phải biến đổi thông tin mới thành sơ đồ kiến thức mới để hiểu và lưu trữ trong bộ nhớ dài hạn [107] hoặc sửa chữa những sơ đồ kiến thức có trước [101, 104-106, 108]. Do đó, nếu người học được tiếp nhận dữ liệu dạng sơ đồ để chuyển đổi thành thông tin thì việc học tốt hơn. Luận án tiếp cận khoa học và kỹ thuật trực quan hóa để đề xuất hệ thống học khám phá có hướng dẫn bằng phương pháp trực quan, sử dụng bài giảng trực quan phù hợp với kiến thức dạng sơ đồ có sẵn trong bộ nhớ dài hạn của người học. Hệ thống học khám phá có hướng dẫn trực quan gồm 3 hợp phần chính: người học, người dạy, và bài giảng (Hình 5.1).

4.2.1. Người học

Trong hệ thống học khám phá có hướng dẫn trực quan, người học được phân lớp (nhóm) theo năng lực. Người học trong cùng một lớp có cùng kiến thức có sẵn, kỹ năng và thái độ tương tự nhau. Các đặc trưng như tự vận động (self-motivation) [50, 56], tự tin [80], tự lực [44, 80], tự tìm học [46, 62, 109], tự đầu tư [62], tự phân tích [110], tư duy phản biện [87], tư duy sáng tạo [44, 46, 62], tương tác [44, 60, 99, 111] của những người học trong cùng một lớp thì tương tự với nhau.

4.2.2. Bài giảng

Lý thuyết sơ đồ [101, 104-106, 108] chỉ ra rằng người học tiếp thu bài giảng bằng hình ảnh trực quan dễ hơn những kiểu bài giảng khác. Do đó, bài giảng nên được thiết kế như là những sơ đồ trực quan phù hợp đồng thời nội dung bài học và sơ đồ kiến thức đang lưu trữ trong bộ nhớ dài hạn của não bộ người học. Về phương diện kỹ thuật, biểu diễn trực quan nội dung bài giảng để

tạo ra bài giảng trực quan cũng phát sinh những yếu tố có lợi và bất lợi đối với sự tiếp thu của người học.

4.2.3. Tính tương thích

Tính tương thích liên quan đến mức độ tương thích của đồ thị trực quan trong bài giảng với sơ đồ kiến thức đang có trong não bộ người học [106]. Người học tiếp thu bài học tốt hơn nếu bài giảng trực quan tương thích tốt với sơ đồ kiến thức có sẵn trong não bộ.

4.2.4. Tính trực quan

Tính trực quan của bài giảng tác động đến tải bộ nhớ của người học [107]. Bài giảng có tính trực quan tốt hỗ trợ người học giảm tải nhận biết của não bộ khi xử lý thông tin cần thiết. Nếu bài giảng không có tính trực quan tốt, tải nhận biết của não bộ bị huy động vào những thông tin không cần thiết. Những dấu hiệu trực quan của bài giảng gồm tính mỹ thuật để thu hút người học vào việc học [112, 113], tính định hướng liên quan đến tính logic hướng đến mục tiêu và kết quả của bài học, tính kích thích khơi gợi khả năng người học trích xuất thông tin từ bài giảng.

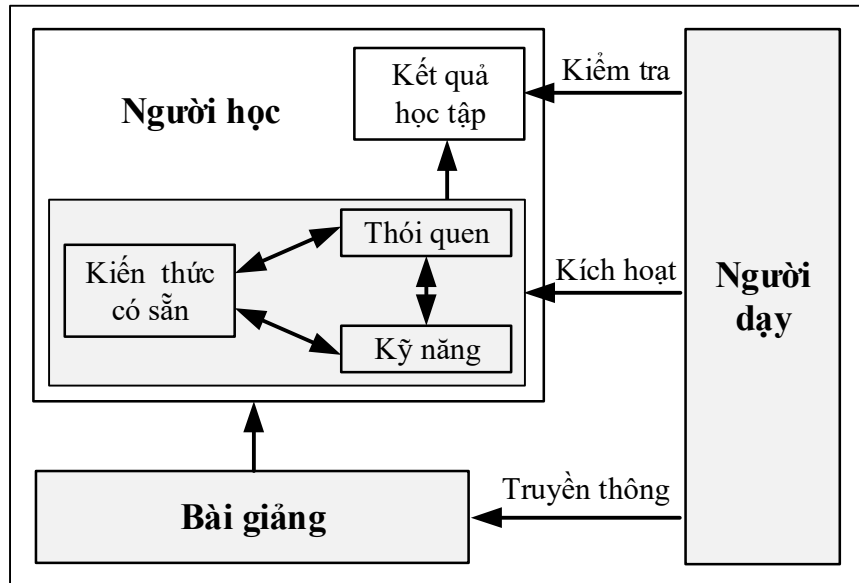
4.2.5. Tính phức tạp

Tính phức tạp của bài giảng tác động đến tải nhận biết của bộ nhớ ngắn hạn của người học. Độ phức tạp của bài giảng càng ít thì bài giảng càng dễ được tiếp thu. Độ phức tạp bài giảng gồm độ phức tạp do tính đa dạng của lĩnh vực học thuật [52, 114, 115], độ phức tạp của vấn đề được trình bày bởi các sơ đồ trực quan biểu diễn những vấn đề khác nhau, độ phức tạp dữ liệu do sự đa dạng của các biến dữ liệu và các quan hệ của chúng, độ phức tạp trực quan do sự khác biệt của các kiểu trực quan, độ phức tạp thời gian tạo ra bởi sự khác biệt về thời gian tham chiếu [107, 114]. Tính phức tạp có thể được giảm thiểu bằng cách chọn các kỹ thuật trực quan phù hợp [52, 113, 115].

4.2.6. Người dạy

Trong hệ thống học khám phá có hướng dẫn trực quan, người dạy không chỉ thiết kế bài giảng trực quan để truyền đạt nội dung bài học đến người học, mà còn kích hoạt quy trình học tập của người học cũng như khuyến khích những thói quen chủ động và kích hoạt những kỹ năng tích cực trong việc học của người học. Người dạy hỗ trợ người học sử dụng sử dụng kiến thức có sẵn để tiếp thu bài giảng và tích hợp vào kiến thức có sẵn trong bộ nhớ dài hạn để hình

thành kiến thức mới. Người dạy cũng hướng dẫn người học kết nối các khoảng cách giữa lý thuyết và thực hành của chủ đề bài học [98] đồng thời phát triển khả năng tư duy phản biện và tư duy sáng tạo [46, 62] trong việc hình thành kiến thức mới [116]. Đối với một hệ thống học khám phá có hướng dẫn trực quan, tính cách cá nhân trong học tập là một đặc trưng quan trọng, người dạy cần phải phát huy những kỹ năng và thói quen của từng người học.

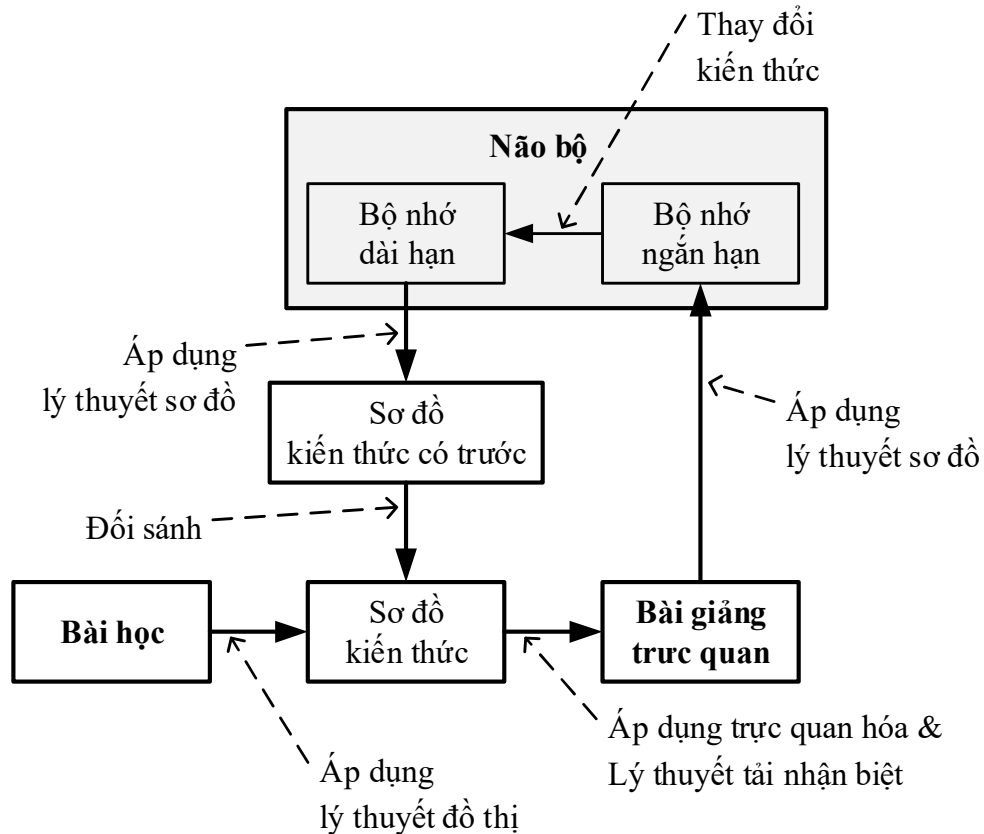


Hình 4.1. Cấu trúc của hệ thống học khám phá có hướng dẫn trực quan

4.2.7. Bài giảng trực quan

4.2.7.1. Dòng thông tin trong quy trình học của người học

Để hiểu được nội dung bài học, người học phải xử lý dòng thông tin đồng thời theo 2 hướng, từ bài học đến bộ nhớ ngắn hạn và từ bộ nhớ ngắn hạn đến bài học [104]. Thiết kế bài giảng trực quan cho hệ thống học khám phá có hướng dẫn trực quan là biến đổi nội dung bài học thành một tập những sơ đồ tri thức được cấu trúc logic như bài giảng, sao cho người học dễ hiểu (Hình 4.2). Sự biến đổi này được thực hiện theo lý thuyết đồ thị (graph theory) [117-119], lý thuyết sơ đồ (schema theory) [101, 104-106, 120], kỹ thuật trực quan hóa [110, 113, 115] và lý thuyết tải nhận biết (cognitive load theory) [107]. Nội dung bài học được phân tích thành những sơ đồ tri thức [117, 121] dựa trên lý thuyết sơ đồ [118, 119], trong đó những sơ đồ kiến thức được thiết kế tương thích với những sơ đồ kiến thức có sẵn của người học đồng thời phù hợp với nội dung bài học.

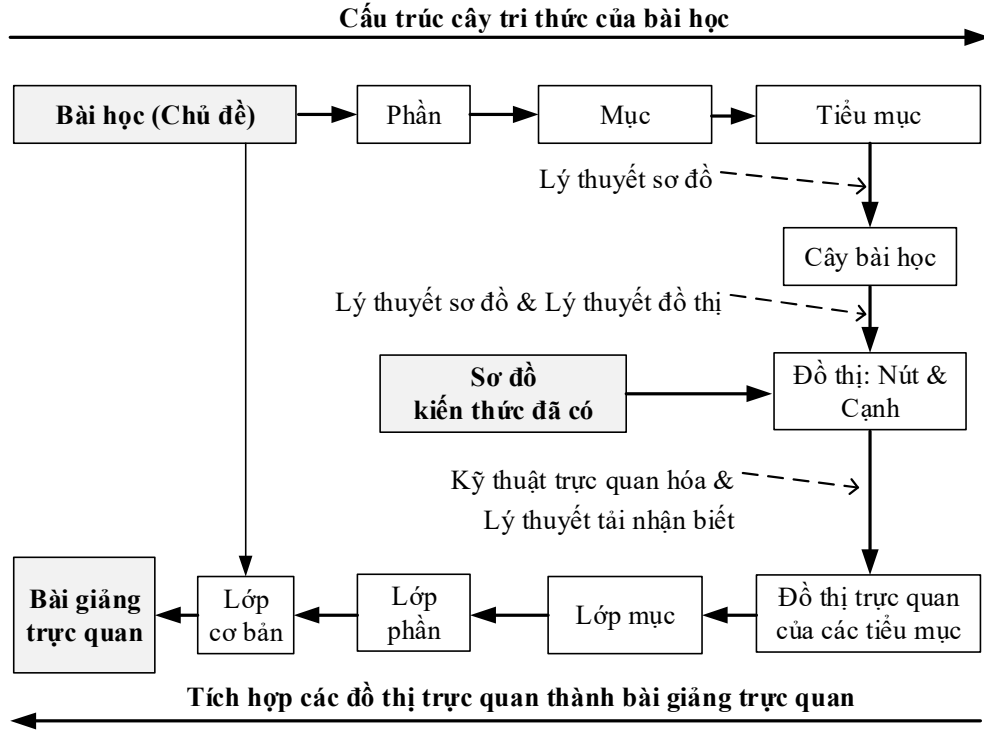


Hình 4.2. Dòng thông tin trong quy trình học của người học trong hệ thống học khám phá có hướng dẫn trực quan

Kết quả học tập của người học được đánh giá bởi sự thay đổi kiến thức trong bộ nhớ dài hạn. Theo lý thuyết tải nhận biết và lý thuyết sơ đồ, con người nhận biết ý nghĩa của dữ liệu bằng cách đối sánh sơ đồ dữ liệu mới nhận với sơ đồ kiến thức có sẵn trong bộ nhớ dài hạn [106, 107]. Kiến thức mới được xử lý tại bộ nhớ ngắn hạn được chuyển đến bộ nhớ dài hạn để lưu trữ. Những bài giảng trực quan được cấu trúc như những sơ đồ kiến thức tương thích với sơ đồ kiến thức có trước giúp người học hiểu được nội dung bài học. Việc học của người học không thành công nếu bộ nhớ ngắn hạn của người học đối sánh sơ đồ kiến thức của bài giảng trực quan không tương thích với sơ đồ kiến thức có sẵn trong bộ nhớ dài hạn. Sự thất bại này có thể do sự nhận dạng không đầy đủ của người giảng về sơ đồ kiến thức có sẵn của người học, hoặc là sự biến đổi nội dung bài học thành sơ đồ kiến thức không hợp lý, hoặc là sự tích hợp không logic những sơ đồ kiến thức vào bài giảng trực quan.

4.2.7.2. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài học

Có 2 tiếp cận thiết kế bài giảng trực quan, tiếp cận bài học và tiếp cận bài toán.



Hình 4.3. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài học và hướng dẫn theo các bước.

Bước 1: Nhận dạng năng lực có sẵn của người học. Năng lực có sẵn của người học gồm kiến thức đã được học, kỹ năng và thói quen mà người dạy cần phải nhận dạng ngay từ lúc mới bắt đầu học. Người dạy tổ chức người học thành những nhóm có năng lực tương tự, nhận dạng sơ đồ kiến thức có sẵn của từng nhóm để soạn bài giảng nhằm tăng hiệu quả khi hướng dẫn. Sơ đồ này được cập nhật sau mỗi bài giảng.

Bước 2: Cấu trúc cây bài học. Người dạy áp dụng lý thuyết đồ thị để phân tích cấu trúc nội dung bài học theo nhiều mức chi tiết kiểu cây phân cấp, gọi là cây bài học. Mức gốc của cây biểu diễn chủ đề của bài học. Mức đề mục biểu diễn lần lượt từng cấp đề mục, tiểu mục, v.v.. như những nút. Mỗi mức có thể có nhiều nút (Hình 4.3).

Bước 3: Thiết kế lớp nền. Lớp nền của bài giảng được thiết kế để làm khung bài giảng gồm tên tựa bài, mục tiêu, thời gian, yêu cầu kết quả học tập, yêu cầu về kỹ năng và kiến thức có sẵn của người học. Lớp nền được thiết kế

không chỉ để tích hợp sơ đồ trực quan của bài giảng mà còn giúp người học lướt tổng quát bài giảng trước khi đi sâu vào chi tiết.

Bước 4: Biểu diễn các mục nhỏ như những sơ đồ trực quan. Những tiêu mục của bài học được biến đổi thành những sơ đồ trực quan theo sơ đồ kiến thức có trước để giúp người học hiểu bài dễ hơn.

Bước 5: Tích hợp sơ đồ trực quan để hình thành bài giảng trực quan. Những sơ đồ trực quan được tích hợp theo các tiêu mục để thiết lập từng lớp tiêu mục. Các lớp tiêu mục được tích hợp một cách logic theo sơ đồ cấu trúc cây để hình thành bài giảng (Hình 4.3).

4.2.7.3. Quy trình thiết kế bài giảng trực quan theo tiếp cận bài toán

Bước 1: Nhận dạng năng lực có sẵn của người học. Năng lực có sẵn của người học gồm kiến thức đã được học, kỹ năng và thói quen mà người dạy cần phải nhận dạng ngay từ lúc mới bắt đầu học. Người dạy tổ chức người học thành những nhóm có năng lực tương tự, nhận dạng sơ đồ kiến thức có sẵn của từng nhóm để soạn bài giảng nhằm tăng hiệu quả khi hướng dẫn. Sơ đồ này được cập nhật sau mỗi bài giảng.

Bước 2: Người dạy phát biểu bài toán và hướng dẫn người học tự khám phá tri thức theo chủ đề nêu ra từ nhiều đề mục của một bài học và của nhiều bài học (Hình 4.4).

Bước 3: Người dạy thiết lập sơ đồ trực quan như một quy trình khám phá tri thức và hướng dẫn người học tự tìm kiếm kiến thức tương ứng từ nhiều nguồn, bao gồm kiến thức có sẵn đến kiến thức tự khám phá mới (Hình 4.4).

Bước 4: Kiểm tra kết quả tự khám phá của người học. Người dạy hướng dẫn người học tự kiểm tra và hệ thống hóa một cách logic để tạo ra đáp án của bài toán.

Bước 5: Tổng quát hóa. Người dạy hướng dẫn người học tổng quát hóa kiểu bài toán thành quy luật để có thể áp dụng cho những trường hợp tương tự [122].

	Bài toán 1	Bài toán 2		Bài toán I-1	Bài toán I
Chuyên đề N					
Chuyên đề N-1					
.....					
Chuyên đề 3					
Chuyên đề 2					
Chuyên đề 1					

Hình 4.4. Thiết kế bài giảng trực quan theo tiếp cận bài toán và hướng dẫn theo các bước

4.3. Đánh giá hệ thống học khám phá có hướng dẫn trực quan

Kết quả của một hệ thống học khám phá có hướng dẫn trực quan gồm tri thức được hướng dẫn và tri thức tự khám phá hình thành trong não bộ của người học do hoạt động dạy và học kết hợp. Kết quả hướng dẫn là phần kiến thức trong bộ nhớ dài hạn của người học thay đổi do sự hướng dẫn của người dạy. Kết quả khám phá liên quan đến thông tin hoặc kiến thức mới được khám phá bởi người học [122]. Hiệu năng (efficiency) của hệ thống được đánh giá dựa trên những kiến thức mà người học đã thêm vào bộ nhớ dài hạn sau mỗi bài học, ký hiệu O .

Đầu vào của hệ thống học khám phá có hướng dẫn trực quan gồm năng lực của não bộ mà người học phải tải và thời gian mà người học cần để phát sinh ra tri thức mới (Hình 4.5). Nói cách khác, kết quả học tập O phụ thuộc vào tải bộ não B và thời gian T (biểu thức 4.1). Những yếu tố đầu vào tác động đến tải não bộ để phát sinh kết quả gồm những tính chất của bài giảng như mức độ tương thích của bài giảng với kiến thức có sẵn của người học M , tính chất trực quan V và độ phức tạp C ; đặc trưng của người học như kiến thức có sẵn K , kỹ năng S và thói quen A ; và năng lực sư phạm của người dạy P (Hình 4.5). Sự

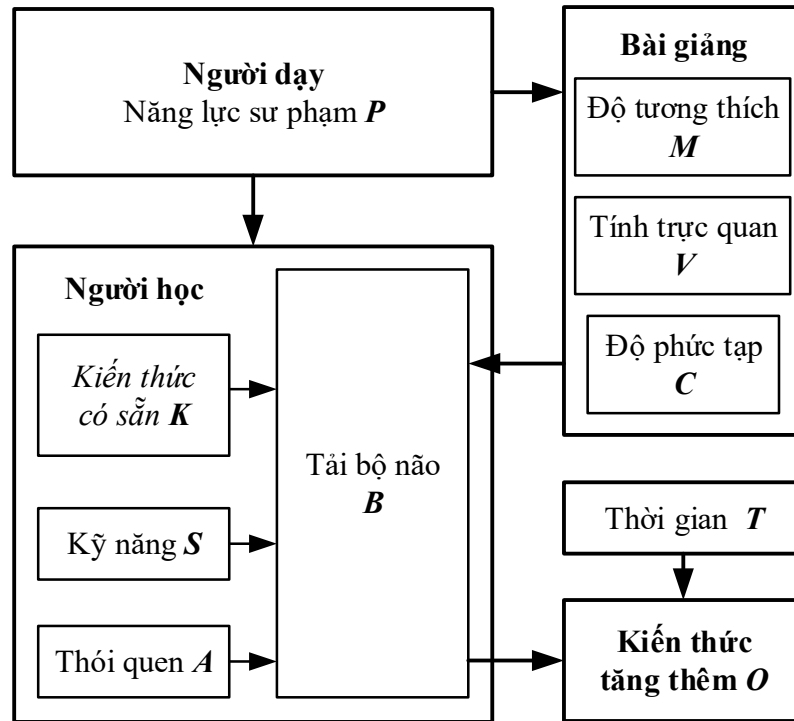
phụ thuộc của tải não bộ vào các tính chất của bài giảng, người học và người dạy được biểu diễn theo biểu thức 4.2 và 4.3.

Trong quy trình học tập, người học huy động tải não bộ trong một khoảng thời gian để tiếp nhận tri thức ẩn chứa trong bài giảng. Hiệu năng của hệ thống học khám phá có hướng dẫn trực quan được định nghĩa là tương quan giữa kết quả học tập và tải não bộ của người học và thời gian mà người học cần để thực hiện việc học (Hình 4.5). Hiệu năng của hệ thống được đánh giá cho mỗi bài giảng vì tính chất khác nhau của các bài giảng trực quan. Biểu thức 4.4 và 4.5 đánh giá hiệu năng của 1 bài giảng trong hệ thống học khám có hướng dẫn trực quan.

$$B \times T \rightarrow O \quad (4.1)$$

$$K \times S \times A \times M \times V \times C \times P \rightarrow B \quad (4.2)$$

$$B \equiv C \times K^{-1} \times S^{-1} \times A^{-1} \times M^{-1} \times V^{-1} \times P^{-1} = C \times (K \times S \times A \times M \times V \times P)^{-1} \quad (4.3)$$



Hình 4.5. Mô hình đánh giá hiệu năng của hệ thống học khám phá có hướng dẫn trực quan áp dụng cho 1 bài giảng.

Áp dụng hệ thống học khám phá có hướng dẫn trực quan cho một nhóm N người học, hiệu năng của hệ thống được tính như sau, với ký hiệu “ $.. \equiv ..$ ” được định nghĩa là tỉ lệ thuận:

$$E \equiv \frac{O}{B \times T} \quad (4.4)$$

$$E \equiv \frac{O \times K \times S \times A \times M \times V \times P}{C \times T} \quad (4.5)$$

Trong đó,

$$O = \text{median}\{o_n \mid n = 1, 2, \dots, N\} \text{ hoặc } O = \left(\frac{1}{N}\right) \sum_{n=1}^N o_n$$

$$K = \text{median}\{k_n \mid n = 1, 2, \dots, N\} \text{ hoặc } K = \left(\frac{1}{N}\right) \sum_{n=1}^N k_n$$

$$S = \text{median}\{s_n \mid n = 1, 2, \dots, N\} \text{ hoặc } S = \left(\frac{1}{N}\right) \sum_{n=1}^N s_n$$

$$A = \text{median}\{a_n \mid n = 1, 2, \dots, N\} \text{ hoặc } A = \left(\frac{1}{N}\right) \sum_{n=1}^N a_n$$

$$M = \text{median}\{h_n \mid n = 1, 2, \dots, N\} \text{ hoặc } M = \left(\frac{1}{N}\right) \sum_{n=1}^N h_n$$

$$V = \text{median}\{v_n \mid n = 1, 2, \dots, N\} \text{ hoặc } V = \left(\frac{1}{N}\right) \sum_{n=1}^N v_n$$

$$C = \text{median}\{c_n \mid n = 1, 2, \dots, N\} \text{ hoặc } C = \left(\frac{1}{N}\right) \sum_{n=1}^N c_n$$

$$P = \text{median}\{p_n \mid n = 1, 2, \dots, N\} \text{ hoặc } P = \left(\frac{1}{N}\right) \sum_{n=1}^N p_n$$

Trong thực tế, kết quả O được đánh giá bằng kiểm tra người học sau khi bài được giảng xong, các biến K, S, A được đánh giá bằng kiểm tra đầu vào, các biến M, V, C, P được đánh giá bằng khảo sát người học, thời gian T được ghi nhận cho mỗi bài giảng. Biểu thức (4.5) có thể được biến đổi toán học để có hiệu năng của hệ thống học khám phá có hướng dẫn trực quan đối với một bài giảng như sau:

$$E = \frac{O + K + S + A + M + V + P}{C + T} \quad (4.6)$$

Giá trị của (5.6) thay đổi từ $\frac{7}{20}$ đến $\frac{70}{2}$ có thể được chuẩn hóa để được:

$$E = \frac{O + K + S + A + M + V + P}{C + T} \times \frac{20}{7} \% \quad (4.7)$$

4.4. Áp dụng phương pháp học khám phá có hướng dẫn trực quan để huấn luyện đội tuyển môn Tin học

Đội tuyển tin học của một trường phổ thông được tuyển chọn từ những học sinh xuất sắc về tin học từ các lớp 10, 11 và 12 để tham gia thi đấu tại các

giải học sinh giỏi. Đội viên đội tuyển môn Tin học phải thi đấu với những học sinh giỏi từ các trường khác nên mỗi đội viên cần có kiến thức tốt và phải có kỹ năng tư duy sáng tạo, tư duy logic, tư duy phản biện, kỹ năng sử dụng thời gian, đặc biệt là sự tự tin, tự lực trong khi thi đấu. Mô hình học khám phá có hướng dẫn là mô hình phù hợp để huấn luyện đội viên đội tuyển vì người học khám phá không chỉ được học kiến thức mà còn được rèn luyện các kỹ năng thông qua việc tìm ra khái niệm hoặc trả lời những vấn đề đã định hướng [62]. Được huấn luyện trong mô hình học khám phá có hướng dẫn, các đội viên đội tuyển được xây dựng nền tảng chung trước khi làm việc cùng nhau và được huấn luyện truyền thông để chia sẻ những khám phá mới [60].

Nhiều nghiên cứu cho thấy rằng những công cụ học toán dùng trong các trường không còn phù hợp với chương trình, do đó phải phát triển công cụ học toán để giúp thực hiện quy trình học [49]. Khoa học về học tập dùng mô hình học khám phá có hướng dẫn phát huy tối đa khả năng của người học để định vị chủ đề và nghiên cứu có hệ thống, phản biện, sáng tạo và logic sao cho người học được nâng cao kiến thức với thái độ và các kỹ năng được thay đổi trong ứng xử với đề thi [62]. Trong đó, tư duy phản biện là nền tảng thiết lập kết quả học tập mà người học nên đạt tới thông qua kinh nghiệm ở trường [123], tư duy sáng tạo cần có để tìm ra ý tưởng mới có lợi hơn cho việc giải bài toán [62].

Bài thi tin học tương tự như những bài toán, những vấn đề của bài toán được trình bày như những câu hỏi xuất phát từ trong chính các bài toán hoặc từ đời sống thực tế bao gồm những sự kiện và môi trường văn hóa có thể được mô hình hóa toán học. Nếu người học có chiến lược tốt để giải bài toán thì bài toán không còn là bài toán mà là bài tập [44]. Khả năng giải bài toán có thể được định nghĩa như là khả năng của người học hiểu bài toán, hoạch định chiến lược giải bài toán, thực hiện những chiến lược hoàn thành được chọn lọc và tái kiểm tra cách giải bài toán để làm những lời giải kế tiếp theo cách khác hoặc phát triển lời giải bài toán [44].

Các bước giải bài toán gồm [44]:

- *Bước 1: Nhận dạng và hiểu bài toán.* Khi tiếp nhận bài toán, người học, với kiến thức và tư duy của mình, nhận dạng bài toán, biết và hiểu rõ mục tiêu, yêu cầu của bài toán.

- *Bước 2: Hoạch định cách giải.* Với tư duy logic và tư duy sáng tạo, người học nhận biết làm thế nào để sử dụng nguồn dữ liệu, xử lý dữ liệu, kết nối dữ liệu với yêu cầu của bài toán, phát sinh ý tưởng để giải và hoạch định cách giải bài toán.
- *Bước 3: Thực hiện kế hoạch.* Sau khi hoạch định cách giải, người học áp dụng tư duy phản biện để kiểm tra lại kế hoạch và bằng sự tự tin của mình, người học triển khai quy trình giải bài toán như đã hoạch định tại bước 2.
- *Bước 4: Kiểm tra kết quả.* Sau khi giải xong bài toán, người học kiểm tra lại tính logic của lời giải và kết nối kết quả với mục tiêu bài toán, kể cả việc kết luận và tổng quát hóa kết quả bài toán. Trong huấn luyện, người học có thể thảo luận với người hướng dẫn hoặc những người bạn cùng học cộng tác.

4.5. Kết luận chương 4

Học tập theo mô hình học khám phá có hướng dẫn, người học không chỉ được người dạy hướng dẫn cấu tạo kiến thức mới, nâng cao kỹ năng tư duy sáng tạo, kỹ năng tư duy phản biện, kỹ năng logic trong lập luận, khả năng tự tin và tự lực để có thể tự giải các bài toán khi không có người hướng dẫn. Thế kỷ 21 với sự đa dạng về phương tiện truyền thông cung cấp nhiều dữ liệu và thông tin, đặc biệt là các phương tiện chia sẻ kiến thức như, Google, ChatGPT, DeepSeek, v.v.. mô hình học khám phá dần thay thế mô hình học khái niệm (exploratory) truyền thống.

Luận án đã đề xuất mô hình học khám phá có hướng dẫn trực quan, trong đó kết hợp với các tính chất của phương pháp học cộng tác và học theo tiếp cận bài toán. Mô hình học khám phá có hướng dẫn trực quan này không chỉ phù hợp sự phát triển của xã hội, sự phát triển của các nguồn dữ liệu, nguồn thông tin, các phương tiện kỹ thuật trên internet và đặc biệt là phù hợp để huấn luyện đội viên đội tuyển môn Tin học của các trường phổ thông tham dự thi đấu tại các kỳ thi học sinh giỏi cấp quốc gia.

Tuy nhiên, trong thực tế, nhiều người dạy không muốn áp dụng mô hình học khám phá vì một số lý do cá nhân [50]. Những rào cản này có thể khắc phục bằng cách kích thích động lực học tập của người học, giúp người học tiếp

thu tốt nội dung và ghi nhớ lâu dài những khái niệm tự khám phá. Những rào cản này là [50]:

- Nhiều người dạy cho rằng phương pháp học khám phá không bao quát nội dung bài giảng và lo ngại nội dung môn học không được học tập đầy đủ vì người dạy không thể cho người học sử dụng tất cả thời gian được quy định để khám phá nội dung mà người dạy chịu trách nhiệm. Rào cản này có thể được khắc phục bởi người dạy bằng cách phối hợp năng lực người học với tài liệu học tập được soạn bài giảng và thời gian tự học có thể truy cập vào các nguồn trên internet.
- Học khám phá đòi hỏi quá nhiều sự chuẩn bị của người dạy và thời gian học của người học. Rào cản này có thể khắc phục vì người dạy có thể hướng dẫn người học kỹ năng nghiên cứu và khuyến khích, hướng dẫn người học nghiên cứu chủ đề đồng thời áp dụng kỹ năng này cho bài mới mà không cần thêm nhiều thời gian để xử lý.
- Quy mô lớp học không thích hợp để áp dụng học khám phá vì quy mô lớp quá lớn hoặc quá nhỏ, lớp quá lớn không thể áp dụng tương tác 1-1 giữa những người học trong khi học khám phá, lớp quá nhỏ thì kinh nghiệm thu được từ sự tương tác giữa những người học bị giới hạn. Rào cản này có thể khắc phục bằng cách chia lớp lớn thành những nhóm có quy mô thích hợp tùy theo nội dung của chủ đề.

KẾT LUẬN

Luận án đã thực hiện được những kết quả sau:

1. Dataset. *Luận án đã thiết lập tập dữ liệu (dataset) năng lực thí sinh đã tham gia các kỳ thi học sinh giỏi cấp quốc gia môn Tin học, gồm những thí sinh thắng giải (winners) và những thí sinh không thắng giải (non-winners).*

Luận án đã xây dựng mô hình vector đặc trưng năng lực của những thí sinh gồm những biến năng lực học tập (learning features) và những biến năng lực phi học tập (non-learning features). Dữ liệu được thu thập bằng Questionnaire với mỗi câu hỏi được khảo sát tương ứng một biến năng lực. Sau khi phân tích những đặc trưng năng lực của con người, luận án đã biên soạn Questionnaire gồm 18 câu hỏi liên quan biến năng lực học tập và 22 câu hỏi liên quan biến năng lực phi học tập. Các câu hỏi đã vượt qua các thách thức về tính riêng tư của nội dung được hỏi và thời gian trả lời. Những câu hỏi được soạn thảo rất đơn giản, chỉ yêu cầu hỏi đáp định tính dễ dàng không mất thời gian.

Luận án đã đề xuất tiếp cận ma trận hỏi đáp và khối 3D hỏi đáp để định lượng các biến năng lực của không gian năng lực và định lượng các biến năng lực của vector năng lực của người được khảo sát. Kết quả là luận án đã xây dựng được 2 dataset để làm thực nghiệm: Dataset 2024 gồm 221 mẫu (120 winners và 101 non-winners) của những thí sinh các kỳ thi năm học 2020-2021, 2021-2022, 2022-2023; Dataset 2025 gồm 497 mẫu (360 winners và 137 non-winners) của những thí sinh các kỳ thi năm học 2021-2022, 2022-2023, 2023-2024.

Phương pháp thiết lập dataset của luận án có tính linh hoạt phù hợp tiếp cận Winner-domain, tiếp cận Winner-cosin và phù hợp với đối tượng tạo mẫu là thí sinh tham gia các kỳ thi học sinh giỏi cấp quốc gia. Dataset linh hoạt có thể thay đổi dễ dàng về số mẫu, có thể được cập nhật mỗi năm về số lượng và nội dung của biến dữ liệu, có thể cập nhật theo yêu cầu thành lập đội tuyển của môn thi, cấp thi, kỳ thi.

2. Thuật toán Winner-domain và thuật toán Winner-cosin. *Luận án đã xây dựng thuật toán và quy trình tuyển chọn học sinh vào đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp quốc gia môn Tin học.*

Luận án đã áp dụng nguyên lý “gần mực thì đen gần đèn thì sáng” và chiến lược “thà bỏ lại hơn chọn nhầm” để xây dựng tiếp cận chọn đội viên từ

nhiều ứng viên cùng là học sinh giỏi của trường để dự thi học sinh giỏi cấp quốc gia môn Tin học. Kết quả nghiên cứu gồm:

- *Winner training set*. Luận án áp dụng nguyên lý của tiếp cận SVM để xác định winner training set là miền không gian năng lực tập trung các học sinh giỏi đã từng thắng giải học sinh giỏi cấp quốc gia môn Tin học, được xác định bởi trọng tâm, khoảng cách tối đa và góc hướng (khuyh hướng) tối đa.
- *Winner-domain tốt*. Luận án phân hoạch ngẫu nhiên winner dataset nhiều lần để có nhiều winner training sets. Với mỗi winner training set, luận án thay đổi bán kính để tìm những Winner-domains có những tỉ lệ chọn nhầm khác nhau và những tỉ lệ chọn đúng khác nhau. Những Winner-domains có tỉ lệ chọn nhầm bằng 0, tỉ lệ chọn đúng cao và không chứa bất kỳ non-winner nào được chọn làm Winner-domain tốt để tuyển chọn đội viên tiềm năng. Những ứng viên có vector năng lực trong miền không gian của Winner-domain tốt được chọn vào đội tiềm năng.
- *Winner-cosin*. Luận án áp dụng độ tương đồng cosin (Cosin Similarity) giữa vector năng lực của ứng viên với tất cả vector năng lực của winners.
- Tiếp cận cây quyết định giúp luận án chọn đội viên đội tuyển từ đội tiềm năng.

3. Mô hình học khám phá có hướng dẫn trực quan

- *Mô hình học khám phá có hướng dẫn trực quan*. Luận án tiếp cận phương pháp giáo dục với người học là trung tâm và mô hình học khám phá để xây dựng mô hình học khám phá có hướng dẫn trực quan áp dụng cho các cơ sở giáo dục. Mô hình được thực nghiệm trong việc huấn luyện đội viên đội tuyển môn Tin học.
- *Bài giảng trực quan trong mô hình học khám phá có hướng dẫn trực quan*. Luận án tiếp cận lý thuyết sơ đồ, lý thuyết tải nhận biết của con người trong quy trình tiếp nhận và lưu trữ kiến thức và kỹ thuật trực quan để xây dựng quy trình và phương pháp biên soạn bài giảng trực quan áp dụng cho mô hình học khám phá có hướng dẫn trực quan.
- *Hiệu năng (efficiency) của một bài giảng trực quan*. Luận án đã xây dựng mô hình toán học để đánh giá hiệu năng của từng bài giảng trực quan

trong mô hình học khám phá có hướng dẫn trực quan. Mô hình đánh giá liên quan nội dung bài giảng và người dạy.

Trong thời gian tới, luận án sẽ được tiếp tục nghiên cứu để:

- Cập nhật giá trị biến năng lực thí sinh của những kỳ thi mới nhất;
- Nghiên cứu các biến năng lực với bộ Questionnaire phù hợp với những môn thi khác, kỳ thi khác;
- Phát triển và so sánh thêm với các mô hình học máy khác.

**DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ
TẠP CHÍ KHOA HỌC QUỐC TẾ**

[CT.1]	Huynh, C.N.T., Nguyen, H.V., Tran, P.V., Ngo, D.N.T., Tran, T.V., Nguyen, H.T. (2023). An Approach to Selecting Students Taking Provincial and National Excellent Student Exams. In: Phan, C.V., Nguyen, T.D. (eds) Context-Aware Systems and Applications. ICCASA 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 475. Springer, Cham, ISBN 978-3-031-28815-9. https://doi.org/10.1007/978-3-031-28816-6_12 . (Scopus) .
[CT.2]	Huynh, C.N.T., Tran, P. & Tran, T. The Approach of Winner-domain to Selecting Members for High School Team Participating in National Excellent Student Competition. <i>Mobile Netw Appl</i> 29 , 306–313 (2024). https://doi.org/10.1007/s11036-024-02342-8 . Volume 29, Issue 1, pp. 306-313. (SCIE, Q2 / 2024), (Scopus) .
[CT.3]	Huynh, C.N.T., Van Thi Tran, A., Bui, T.T., Nguyen, H.T., Tran, P.V. (2024). Applying Guided Discovery Learning to Enhance the Achievement of Information Technology Team. In: Cong Vinh, P., Thanh Tung, N. (eds) Context-Aware Systems and Applications. ICCASA 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 579. Springer, Cham. https://doi.org/10.1007/978-3-031-58878-5_14 , Springer, Cham. ISBN: 978-3-031-58877-8 (Scopus) .
[CT.4]	Cam Ngoc Thi Huynh, Hong Thi Nguyen, Phuoc Vinh Tran, “Efficiency of Visually-Guided Discovery-Learning System for High School”, International Conference on Energy, Infrastructure and Environmental Research (EIER 2025), E3S Web of Conferences 626, 04002 (2025), eISSN: 2267-1242. (Scopus) .

TÀI LIỆU THAM KHẢO

- [1] L. Barbeiro, A. Gomes, F. B. Correia, and J. Bernardino, "A Review of Educational Data Mining Trends," *rocedia Computer Science* vol. 237, pp. 88–95, 2024.
- [2] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 11, pp. 1-19, 2022.
- [3] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Educational and Information Technologies*, vol. 26, no. 1, pp. 205-240, 2021.
- [4] M. P. Colpo, T. T. Primo, M. S. d. Aguiar, and C. Cechinel, "Educational Data Mining for Dropout Prediction: Trends, Opportunities, and Challenges," *Revista Brasileira de Informática na Educação (RBIE)- Brazilian Journal of Computers in Education*, vol. 32, pp. 220-256, 2024.
- [5] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *Ieee Access*, vol. 5, pp. 15991-16005, 2017.
- [6] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in *IOP conference series: materials science and engineering*, Hashim, A. S., Awadh, W. A., & Hamoud, A. K., 2020, vol. 928, no. 3, p. 032019: IOP Publishing.
- [7] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications* vol. 33, pp. 135–146, 2007.
- [8] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowledge Discovery*, pp. 1-21, 2020.
- [9] S. Alturki and N. Alturki, "Using Educational Data Mining to Predict Students' Academic Performance for Applying Early Interventions," *Journal of Information Technology Education: Innovations in Practice*, vol. 20, no. 2021, pp. 121-137, 2021.
- [10] Á. Kocsis and G. Molnár, "Factors influencing academic performance and dropout rates in higher education," *Oxford Review of Education*, vol. 51, no. 3, pp. 414-432, 2025.
- [11] Y. Chen, J. Sun, J. Wang, L. Zhao, X. Song, and L. Zhai, "Machine Learning-Driven Student Performance Prediction for Enhancing Tiered Instruction," *arXiv preprint arXiv:2502.03143*, 2025.
- [12] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Education and Information Technologies*, vol. 23, no. 1, pp. 537-553, 2018.
- [13] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis," *Frontiers in psychology*, vol. 12, p. 698490, 2021.
- [14] C. o. Romero and S. a. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics — Part C: Applications and Reviews*, vol. 40, no. 6, pp. 601-618, 2010.

- [15] G. Mahajan and B. Saini, "Educational Data Mining: A state-of-the-art survey on tools and techniques used in EDM," *International Journal of Computer Applications & Information Technology*, vol. 12, no. 1, pp. 310-316, 2020.
- [16] K. A. Nisha, U. Kulsum, S. Rahman, M. F. Hossain, P. Chakraborty, and T. Choudhury, "A Comparative Analysis of Machine Learning Approaches in Personality Prediction Using MBTI," *Computational Intelligence in Pattern Recognition*, pp. 13-23, 2022.
- [17] F. Ofori, D. E. Maina, and D. R. Gitonga, "Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review," *Journal of Information and Technology*, vol. 4, no. 1, pp. 33-55, 2020.
- [18] Valentim Realinho 1, *, Jorge Machado 2, Luís Baptista 2 and Mónica V. Martins "Predicting Student Dropout and Academic Success," *Data*, vol. 7, no. 146, pp. 1-17, 2022.
- [19] F. Ornelas and C. Ordonez, "Predicting student success: A naïve bayesian application to community college data," *Technology, Knowledge and Learning*, vol. 22, pp. 299-315, 2017.
- [20] Z. Ibrahim and D. Rusli, "Predicting students academic performance: comparing artificial neural network, decision tree and linear regression," in *21st Annual SAS Malaysia Forum*, Shangri-La Hotel, Kuala Lumpur, 2007, pp. 1-6.
- [21] L. Bäumke, C. Grunschel, and M. Dresel, "Student dropout at university: a phase-orientated view on quitting studies and changing majors," *European Journal of Psychology of Education*, vol. 37, pp. 853–876, 2022.
- [22] A. Alshanqiti and A. Namoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification," *IEEE Access*, vol. 8, pp. 203827-203844, 2020.
- [23] W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad, "Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions," in "Scientific Reports,," 2025, vol. 15(1).
- [24] Baker, R. SJD, and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *Journal of educational data mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [25] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, pp. 1432–1462, 2013.
- [26] S. Amarappa and S. V. Sathyanarayana, "Data classification using Support vector Machine (SVM), a simplified approach," *International Journal of Electronics and Computer Science Engineering*, vol. 3, no. 4, pp. 435-445, 2014.
- [27] S. z. Križanic', "Educational data mining using cluster analysis and decision tree technique: A case study," *International Journal of Engineering Business Management*, vol. 12, pp. 1-9, 2020.
- [28] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest

- and Neural Network: A Review," *Journal of Data Analysis and Information Processing*, vol. 8, pp. 341-357, 2020.
- [29] A. Talasbek, A. Serek, M. Zhaparov, S. Moo-Yoo, Y. K. Kim, and G.-H. Jeong, "Personality Classification Experiment by Applying k-Means Clustering," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 16, pp. 162-177, 2020.
 - [30] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE access*, vol. 8, pp. 80716 - 80727, 2020.
 - [31] A. S. Rao, B. S. Kamath, R. R, S. Chowdhury, and S. A, "Use of Artificial Neural Network in Developing a Personality Prediction Model for Career Guidance: A Boon for Career Counselors," *International Journal of Control and Automation*, vol. 13, no. 4, pp. 391 - 400, 2020.
 - [32] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends (JASTT)*, vol. 02, no. 01, pp. 20-28, 2021.
 - [33] E. F. Agyemang *et al.*, "Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application," *Computer Engineering and Intelligent Systems* vol. 15, no. 1, pp. 86-102, 2024.
 - [34] Z. Mushtaq, S. Ashraf, and N. Sabahat, "Predicting MBTI Personality type with K-means Clustering and Gradient Boosting," in *IEEE 23rd international multitopic conference (INMIC) 2020*, pp. 1-5: IEEE.
 - [35] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," in *International Conference on Solid State Devices and Materials Science*, 2012, vol. Physics Procedia, pp. 1104 – 1109: Elsevier.
 - [36] K. A. A. Nazeer and M. P. Sebastian, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm," in *World Congress on Engineering WCE 2009*, London, U.K, 2009, vol. 1, pp. 1-3.
 - [37] Q. Zhou and B. Sun, "Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem," *Data and Information Management*, vol. 8, no. 3, p. 100064, 2024.
 - [38] M.-S. Yang and I. Hussain, "Unsupervised Multi-View K-Means Clustering Algorithm," *IEEE Access*, vol. 11, pp. 13574-13593, 2023.
 - [39] H. Tsegaye, A. Demelash, D. Aklilu, and B. Girma, "Determinants of pediatrics emergency mortality at comprehensive specialized hospitals of South nation nationalities and people region, Ethiopia, 2022: unmatched case-control study," *BMC Pediatrics*, vol. 23, no. 192, pp. 1-9, 2023.
 - [40] Z.-J. Liu, V. Levina, and Y. Frolova, "Information Visualization in the Educational Process: Current Trends," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 13, pp. 49-62, 2020.
 - [41] Y. Mosina, "The Role of Visualization in Education Process," *Інноватика у вихованні*, vol. 11, no. 2, pp. 203-208, 2020.
 - [42] A. Kristiyajati and A. Wijaya, "The Effectiveness of Visualization of Proofs in Learning Mathematics by Using Discovery Learning Viewed from Conceptual Understanding," *Southeast Asian Mathematics Education Journal*, vol. 9, no. 1, 2019.

- [43] K. Shatri and K. Buza, "The Use of Visualization in Teaching and Learning Process for Developing Critical Thinking of Students," *European Journal of Social Sciences: Education and Research*, vol. 4, no. 1, pp. 71-74, 2017.
- [44] R. E. Simamora, S. Saragih, and Hasratuddin, "Improving Students' Mathematical Problem Solving Ability and Self-Efficacy through Guided Discovery Learning in Local Culture Context," *International Electronic Journal of Mathematics Education*, vol. 14, no. 1, pp. 61-72, 2019.
- [45] R. A. Aziz, E. Tarmedi, and Y. Kusmarni, "Developing Students' Information Literature Skill Through The Application of Learning Discovery Learning Model in Social Studies Learning," *International Journal Pedagogy of Social Studies*, vol. 3, no. 1, pp. 9-20, 2018.
- [46] M. H. Rahman, "Using Discovery Learning to Encourage Creative Thinking," *International Journal of Social Sciences & Educational Studies*, vol. 4, no. 2, pp. 98-103, 2017.
- [47] M. D. Svinicki, "A Theoretical Foundation for Discovery Learning," *Advances in Physiology Education*, vol. 20, no. 1, pp. S4-S7, 1998.
- [48] T. A. Prasetya and C. T. Harjanto, "Improving Learning Activities and Learning Outcomes Using the Discovery Learning Method " *Journal of Mechanical Engineering Education*, vol. 5, no. 1, pp. 59-66, 2020.
- [49] Y. Kartika, N. M. Hutapea, and K. Kartini, "Mathematical Learning Development using Discovery Learning Model to Improve Mathematical Understanding Skills of Students," *Journal of Educational Sciences*, vol. 4, no. 1, pp. 124–132, 2020.
- [50] J. A. Castronova, "Discovery Learning for the 21st Century: What is it and how does it compare to traditional learning in effectiveness in the 21st Century?," *Action research exchange*, vol. 1, no. 1, pp. 1-12, 2002.
- [51] H. T. Nguyen, A. V. T. Tran, T. A. T. Nguyen, L. T. Vo, and P. V. Tran, "Multivariate Cube for Representing Multivariable Data in Visual Analytics," in *Context-Aware Systems and Applications*, Thu Dau Mot, Vietnam, 2016, vol. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering: Springer.
- [52] H. T. Nguyen, A. V. T. Tran, T. A. T. Nguyen, L. T. Vo, and P. V. Tran, "Multivariate cube integrated retinal variable to visually represent multivariable data," *EAI Endorsed Transactions on Context-aware Systems and Applications*, vol. 4, no. 12, pp. 1-8, 2017.
- [53] M. R. Ramdhani and B. U. a. S. Subanti, "Discovery Learning with Scientific Approach on Geometry," *Journal of Physics: Conference Series*, vol. 895, p. 012033 2017.
- [54] R. D. Anggraini, A. Murni, and Sakur, "Differences in students' learning outcomes between discovery learning and conventional learning models " *Journal of Physics Conference Series*, vol. 1088, no. 1, p. 012070, 2018.
- [55] E. N. Khabibah, M. Masykuri, and Maridi, "The Effectiveness of Module Based on Discovery Learning to Increase Generic Science Skills," *Journal of Education and Learning*, vol. 11, no. 2, pp. 146-153, 2017.
- [56] D. Druckman and N. Ebner, "Discovery Learning in Management Education: Design and Case Analysis," *Journal of Management Education*, pp. 1 –28, 2017.

- [57] A. G. Balim, "The Effects of Discovery Learning on Students' Success and Inquiry Learning Skills " *Eurasian Journal of Educational Research*, no. 35, pp. 1-20, 2009.
- [58] H. Gijlers and T. d. Jong, "The Relation between Prior Knowledge and Students' Collaborative Discovery Learning Processes," *Journal of Research in Science Teaching* vol. 42, no. 3, pp. 264–282, 2005.
- [59] A. In'am and S. Hajar, "Learning Geometry through Discovery Learning Using a Scientific Approach," *International Journal of Instruction*, vol. 10, no. 1, pp. 55-70, 2017.
- [60] N. Saab, W. R. v. Joolingen, and B. H. A. M. v. Hout-Wolters, "Communication in collaborative discovery learning," *British Journal of Educational Psychology*, vol. 2005, no. 75, pp. 603–621, 2005.
- [61] S. Maarif, "Improving Junior High School Students' Mathematical Analogical Ability Using Discovery Learning Method," *International Journal of Research in Education and Science (IJRES)*, vol. 2, no. 1, pp. 114- 124, 2016.
- [62] S. Suwandari, M. Ibrahim, and W. Widodo, "Application of Discovery Learning to Train the Creative Thinking Skills of Elementary School Student," *International Journal of Innovative Science and Research Technology*, vol. 4, no. 12, 2019.
- [63] J. R. Goldberg and M. L. Nagurka, "Enhancing the Engineering Curriculum: Defining Discovery Learning at Marquette University," in *The 2012 Frontiers in Education Conference: Soaring to New Heights in Engineering Education*, Seattle, Washington, 2012, pp. 405-410: IEEE.
- [64] J. Treadwell and K. J. Walters, "The Impact of Discovery Learning in Writing Instruction on Fifth-Grade Student Achievement," *Journal of Elementary and Secondary Education*, vol. 3, no. 4, pp. 2-29, 2012.
- [65] D. J. Reid, J. Zhang, and Q. Chen, "Supporting scientific discovery learning in a simulation environment," *Journal of Computer Assisted Learning*, vol. 19, pp. 9-20, 2003.
- [66] I. N. Suardana, I. W. Redhana, and N. P. M. Yunithasari, "Students' critical thinking skills comparison in discovery learning based on constructing concept mapping and mind mapping," *Journal of Physics: Conference Series (ICMScE 2019)*, vol. 1521, p. 042089, 2020.
- [67] M. M. Chusni, S. Saputro, S. B. Rahardjo, and Suranto, "Student's Critical Thinking Skills Through Discovery Learning Model Using E-Learning on Environmental Change Subject Matter," *European Journal of Educational Research*, vol. 10, no. 3, pp. 1123 - 1135, 2021.
- [68] G. Mustikaningrum, Widiyanto, and N. Mediatati, "Application of The Discovery Learning Model Assisted by Google Meet to Improve Students' Critical Thinking Skills and Science Learning Outcomes," *International Journal of Elementary Education*, vol. 5, no. 1, pp. 30-38, 2021.
- [69] Y. Arifani, "The Implementation of Team-Based Discovery Learning to Improve Students' Ability in Writing Research Proposal," *International Education Studies*, vol. 9, no. 2, pp. 111-119, 2016.
- [70] A. K. A. Mahmoud, "The Effect of Using Discovery Learning Strategy in Teaching Grammatical Rules to first year General Secondary Student on

- Developing Their Achievement and Metacognitive Skills," *International Journal of Innovation and Scientific Research*, vol. 5, no. 2, pp. 146-153, 2014.
- [71] K. Yamashita, R. Fujioka, S. Kogure, Y. Noguchi, T. Konishi, and Y. Itoh, "Practices of algorithm education based on discovery learning using a program visualization system," *Research and Practice in Technology Enhanced Learning*, vol. 11, no. 15, pp. 1-19, 2016.
 - [72] W. R. v. Joolingen, "Designing for Collaborative Discovery Learning," in *International Conference on Intelligent Tutoring Systems*, Berlin, Heidelberg, 2000, pp. 202-211: Springer Berlin Heidelberg.
 - [73] A. Borek, B. M. McLaren, M. Karabinos, and D. Yaron, "How Much Assistance is Helpful to Students in Discovery Learning?," *Proceedings of the Fourth European Conference on Technology Enhanced Learning, Learning in the Synergy of Multiple Disciplines (EC-TEL 2009)*, LNCS 5794, pp. 391-404, 2009.
 - [74] Yuliana, Tasari, and S. Wijayanti, "The Effectiveness of Guided Discovery Learning to Teach Integral Calculus for the Mathematics Students of Mathematics Education Widya Dharma University," *Journal of Mathematics Education*, vol. 6, no. 1, pp. 1-10, 2017.
 - [75] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in *2nd International Scientific Conference of Al-Ayen University*, 2020, vol. 928, p. 032019: IOP Publishing.
 - [76] I. García-Martínez, J. M. Fernández-Batanero, J. Fernández-Cerero, and S. P. León, "Analysing the Impact of Artificial Intelligence and Computational Sciences on Student Performance: Systematic Review and Meta-analysis," *Journal of New Approaches in Educational Research*, vol. 12, no. 1, pp. 171-197, 2023.
 - [77] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis," *Frontiers in Psychology* 12:698490., vol. 12, p. 698490, 2021.
 - [78] H. Hassan, N. B. Ahmad, and S. Anuar, "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," *Journal of Physics: Conference Series* vol. 1529, no. 5, p. 052041, 2020.
 - [79] Á. Kocsis and G. Molnár, "Factors influencing academic performance and dropout rates in higher education," *Oxford Review of Education*, pp. 1-19, 2024.
 - [80] M. R. Ramdhani, "Discovery Learning with Scientific Approach on Geometry," *Journal of Physics: Conference Series*, vol. 895, 2017.
 - [81] T. Bayne *et al.*, "What is cognition?," *Current Biology*, vol. 29, no. 13, pp. R608-R615, 2019.
 - [82] M. Malanchini, K. Rimfeld, A. G. Allegrini, S. J. Ritchie, and R. Plomin, "Cognitive ability and education: how behavioural genetic research has advanced our knowledge and understanding of their association," *Neuroscience and Biobehavioral Reviews*, pp. 1-52, 2020.

- [83] E. R. Lai, "Critical Thinking: A Literature Review," *Pearson's Research Reports*, vol. 6, no. 40, pp. 1-41, 2011.
- [84] U. Toharudin, "Critical Thinking and Problem Solving Skills: How these Skills are needed in Educational Psychology?," *International Journal of Science and Research* vol. 66, no. 3, pp. 2004-2007, 2017.
- [85] G. R. Akramova, "Modern Approaches to Development Critical Thinking of Students," *Eastern European Scientific Journal* vol. 5, 2017.
- [86] J. Maynes, "Critical Thinking and Cognitive Bias," *Informal Logic*, vol. 35, no. 2, pp. 183---203, 2015.
- [87] K. Yuliani and S. Saragih, "The Development of Learning Devices Based Guided Discovery Model to Improve Understanding Concept and Critical Thinking Mathematically Ability of Students at Islamic Junior High School of Medan " *Journal of Education and Practice*, vol. 6, no. 24, pp. 116-128, 2015.
- [88] O. W. A. Gafour and W. A. S. Gafour, "Creative Thinking skills – A Review article," *Journal of Education and E-Learning*, vol. 4, pp. 44-58, 2020.
- [89] J. R. R. a. S. V. Cristóbal Romero, "A Survey on Pre-Processing Educational Data," in *Educational data mining: applications and trends., 2013. 29-64.*, A. Peña-Ayala, Ed.: pringer International Publishing Switzerland 2014, pp. 29-64.
- [90] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: a survey of the last 10 years," *Education and Information Technologies*, vol. 23, pp. 537-553, 2018.
- [91] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review," *Applied sciences*, vol. 10, no. 3, p. 1042, 2020.
- [92] T. S. R. S. Krishna, J. S. H. Padavala, V. R. Poreddy, and K. B. Laxmi, "Student's Performance Prediction," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* vol. 9, no. V, pp. 524-528, 2021.
- [93] X. Zhang, G. Sun, Y. Pan, H. Sun, Y. He, and J. Tan, "Students performance modeling based on behavior pattern," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 5, pp. 1659-1670, 2018.
- [94] D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification: A Review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81-90, 2021.
- [95] J. Xie, M. Wang, S. Xu, and P. W. Grant, "The Unsupervised Feature Selection Algorithms Based on Standard Deviation and Cosine Similarity for Genomic Data Analysis," *Frontiers in Genetics*, vol. 12, no. 684100, pp. 1-17, 2021.
- [96] M. A. Rifai and I. G. Anugrah, "Semantic Search for Scientific Articles by Language Using Cosine Similarity Algorithm and Weighted Tree Similarity," *Journal Of Development Research*, vol. 5, no. 2, pp. 106-114, 2021.
- [97] R. D. Anggraini, A. Murni, and Sakur, "Differences in students' learning outcomes between discovery learning and conventional learning models," *Journal of Physics: Conference Series*, vol. 1088, no. 1, p. 012070, 2018.
- [98] M. A. Weegar and D. Pacis, "A Comparison of Two Theories of Learning -- Behaviorism and Constructivism as applied to Face-to-Face and Online Learning " in *E-leader conference*, Manila, 2012.

- [99] W. R. v. Joolingen, "Designing for Collaborative Discovery Learning," in *Intelligent Tutoring Systems*, Montreal, Canada, 2000, vol. Lecture Notes in Computer Science pp. 202-211: Springer.
- [100] D. J. Reid, J. Zhang, and Q. Chen, "Supporting scientific discovery learning in a simulation environment," *Journal of Computer Assisted Learning*, vol. 2003, no. 19, pp. 9-20, 2003.
- [101] S. A. Widmayer, "Schema Theory: An Introduction," *Retrieved December*, vol. 26, 2004.
- [102] L. P. Rieber, S.-C. Tzeng, and K. Tribble, "Discovery learning, representation, and explanation within a computer-based simulation: finding the right mix," *Learning and Instruction*, vol. 14, pp. 307–323, 2004.
- [103] A. Borek, B. M. McLaren, M. Karabinos, and D. Yaron, "How Much Assistance is Helpful to Students in Discovery Learning?," in *the Fourth European Conference on Technology Enhanced Learning, Learning in the Synergy of Multiple Disciplines (EC-TEL 2009)*, Nice, France, 2009, vol. LNCS 5794, pp. 391-404.
- [104] S. An, "Schema Theory in Reading," *Theory and Practice in Language Studies*, vol. 3, no. 1, pp. 130-134, 2013.
- [105] M. A. Arbib, "Schema Theory," *The encyclopedia of artificial intelligence*, vol. 1992, no. 2, pp. 1427-1443, 1992.
- [106] M. v. Kesteren, "Schemas in the brain: Influences of prior knowledge on learning, memory, and education," Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands, 2013.
- [107] J. Sweller, P. Ayres, and S. Kalyuga, *Cognitive load theory* (Explorations in the Learning Sciences, Instructional Systems and Performance Technologies). Springer, 2011.
- [108] F. Sun, "The Application of Schema Theory in Teaching College English Writing," *Theory and Practice in Language Studies*, vol. 4, no. 7, pp. 1476-1482, 2014.
- [109] A. Kristiyajati and A. Wijaya, "The Effectiveness of Visualization of Proofs in Learning Mathematics by Using Discovery Learning Viewed from Conceptual Understanding," *Southeast Asian Mathematics Education Journal*, vol. 9, no. 1, pp. 37-44, 2019.
- [110] H. T. Nguyen, T. X. Le, P. V. Tran, and D. V. Pham, "An Approach of Taxonomy of Multidimensional Cubes Representing Visually Multivariable Data," in *8th EAI International Conference, ICCASA 2019 and 5th EAI International Conference, ICTCC 2019*, My Tho City, Vietnam, 2019, vol. LNICST 298, pp. 90-104: Springer.
- [111] Y. Arifani, "The Implementation of Team-Based Discovery Learning to Improve Students' Ability in Writing Research Proposal," *International Education Studies*, vol. 9, no. 2, 2016.
- [112] P. E. Parrish, "Aesthetic principles for instructional design," *Educational Technology Research and Development*, vol. 57, no. 4, pp. 511-528, 2009.
- [113] H. T. Nguyen *et al.*, "Integrating Retinal Variables into Graph Visualizing Multivariate Data to Increase Visual Features " in *8th EAI International*

- Conference on Context-Aware Systems and Applications - ICCASA 2019*, My Tho, Vietnam, 2019, vol. LNICST 298, pp. 74-89: Springer.
- [114] W. Huang, P. Eades, and S.-H. Hong, "Measuring effectiveness of graph visualizations: A cognitive load perspective," *Journal of Information Visualization*, vol. 8, no. 3, pp. 139-152, 2009.
 - [115] P. V. Tran and T. X. Le, "Approaching human vision perception to designing visual graph in data visualization," *Concurrency Computation: Practice Experience*, vol. 33, no. 2, pp. 1-17, 2020.
 - [116] A. Guney and S. Al, "Effective learning environments in relation to different learning theories " *Procedia - Social and Behavioral Sciences*, vol. 46, pp. 2334 – 2338, 2012.
 - [117] S. Auer, V. Kovtun, M. Prinz, A. Kasprzik, and M. Stocker, "Towards a Knowledge Graph for Science," in *The 8th International Conference on Web Intelligence, Mining and Semantics*, Novi Sad, Serbia, 2018, pp. 1-6.
 - [118] M. S. Rahman, *Basic Graph Theory* (Undergraduate Topics in Computer Science). Springer 2017, p. 173.
 - [119] R. J. Wilson, *Introduction to Graph Theory*, Fourth edition ed. Longman, 1998.
 - [120] M. L. Wood, D. S. Stoltz, J. V. Ness, and M. A. Taylor, "Schemas and Frames," *Sociological Theory*, vol. 36, no. 3, pp. 244 –261, 2018.
 - [121] M. Rizun, "Knowledge Graph Application in Education: a Literature Review," *Folia Oeconomica - Acta Universitatis Lodziensis*, vol. 3, no. 342, pp. 7-19, 2019.
 - [122] H. T. Nguyen, T. M. T. Pham, T. A. T. Nguyen, A. V. T. Tran, P. V. Tran, and D. V. Pham, "Two-Stage Approach to Classifying Multidimensional Cubes for Visualization of Multivariate Data," in *7th EAI International Conference, ICCASA 2018 and 4th EAI International Conference, ICTCC 2018*, Viet Tri, Vietnam, 2019, vol. LNICST 266, pp. 70-80: Springer.
 - [123] I. N. Suardana, I. W. Redhana, and N. P. M. Yunithasari, "Students' critical thinking skills comparison in discovery learning based on constructing concept mapping and mind mapping," in *International Conference on Mathematics and Science Education 2019 (ICMScE 2019)*, 2019, vol. 1521, no. 2020.

PHỤ LỤC
QUESTIONNAIRE KHẢO SÁT NĂNG LỰC THÍ SINH
ĐÃ THI HỌC SINH GIỎI CẤP QUỐC GIA MÔN TIN HỌC

Câu hỏi đặc trưng năng lực học tập F_1 : Điểm trung bình môn Toán cả năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

*Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.***

Câu khảo sát đặc trưng năng lực học tập F_2 : Điểm trung bình môn Tin cả năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

*Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.***

Câu khảo sát đặc trưng năng lực học tập F_3 : Điểm trung bình môn Lý cả năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

*Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.***

Câu khảo sát đặc trưng năng lực học tập F_4 : Điểm trung bình môn Hóa cả năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

*Người được khảo sát trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.***

Câu khảo sát đặc trưng năng lực học tập F_5 : Điểm trung bình môn Sinh cả năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_6 : Điểm trung bình tất cả các môn học (Kết quả học tập toàn năm) năm lớp 10:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_7 : Điểm trung bình môn Toán cả năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_8 : Điểm trung bình môn Tin cả năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0
0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_9 : Điểm trung bình môn Lý cả năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{10} : Điểm trung bình môn Hóa cả năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu hỏi khảo sát trung năng lực học tập F_{11} : Điểm trung bình môn Sinh cả năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu hỏi khảo sát trung năng lực học tập F_{12} : Điểm trung bình tất cả các môn học (Kết quả học tập toàn năm) năm lớp 11:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{13} : Điểm trung bình môn Toán cả năm (hoặc học kỳ 1) lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{14} : Điểm trung bình môn Tin cả năm (hoặc học kỳ 1) lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{15} : Điểm trung bình môn Lý cả năm (hoặc học kỳ 1) lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{16} : Điểm trung bình môn Hóa cả năm (hoặc học kỳ 1) lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{17} : Điểm trung bình môn Sinh cả năm (hoặc học kỳ 1) lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án.**

Câu khảo sát đặc trưng năng lực học tập F_{18} : Điểm trung bình tất cả các môn học (Kết quả học tập toàn năm hoặc học kỳ 1) năm lớp 12:

0.0 - <7.0	7.0 - <7.5	7.5 - <8.0	8.0 - < 8.5	8.5 - <9.0	9.0 - <9.5	9.5 - 10.0

Người được khảo sát được mời trả lời bằng cách đánh dấu (X) vào ô trống của bảng ở dòng dưới tương ứng số điểm đạt được ở dòng trên, **chọn tối đa 01 đáp án**.

Câu khảo sát đặc trưng năng lực phi học tập F_{19} : Tại sao em muốn vào đội tuyển và dự thi thắng giải học sinh giỏi cấp quốc gia về Tin học?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình

$$F_{19}^m = \{a_{19}^m, b_{19}^m, c_{19}^m, d_{19}^m, e_{19}^m, f_{19}^m\}.$$

- a. Vì em muốn tạo ra niềm vui và sự hạnh diện cho gia đình để không phụ lòng cha mẹ, người thân, thầy cô, bạn bè.

Trả lời: Em chọn câu trả lời này ()

Nếu đáp án này được **chọn** thì giá trị $a_{19}^m = 1$ nếu đáp án này không được chọn thì $a_{19}^m = 0$.

- b. Vì em muốn vào trường đại học danh tiếng mà em và gia đình mong muốn để được học đại học và được học bổng.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{19}^m = 1$, nếu không được chọn thì $b_{19}^m = 0$.

- c. Vì em muốn có được tiền thưởng của nhà nước và của trường.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{19}^m = 1$, nếu không được chọn thì $c_{19}^m = 0$.

- d. Vì em muốn chinh phục các bài toán khó để khẳng định năng lực của bản thân và được tỏa sáng hơn bạn bè.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{19}^m = 1$, nếu không được chọn thì $d_{19}^m = 0$.

- e. Vì em muốn thỏa mãn niềm đam mê tin học.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $e_{19}^m = 1$, nếu không được chọn thì $e_{19}^m = 0$.

- f. Vì em theo xu hướng chung của nhiều bạn.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $f_{19}^m = 1$, nếu không được chọn thì $f_{19}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{20} : Đối với một bài toán, giáo viên hướng dẫn thuật toán để giải với yêu cầu người học hiểu được và vận dụng tốt thuật toán đó khi giải mà không nhất thiết phải đạt điểm tối đa, nhưng thuật toán đó chưa phải tốt nhất, chỉ có thể đạt tối đa 80% test. Trong tình huống đó, em làm gì?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{20}^m = \{a_{20}^m, b_{20}^m, c_{20}^m, d_{20}^m\}$.

- a. Em không cần tìm hiểu thêm, chỉ thực hiện theo yêu cầu giáo viên.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{20}^m = 1$, nếu không được chọn thì $a_{20}^m = 0$.

- b. Em khắc sâu bài toán đó, trong thời gian học luôn suy nghĩ làm cách gì tối ưu hóa bài toán đó để đạt đến 100% test.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{20}^m = 1$, nếu không được chọn thì $b_{20}^m = 0$.

- c. Em tìm cách giải đơn giản hơn chỉ cần đạt 50% test là được, để dành thời gian giải bài khác khó hơn.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{20}^m = 1$, nếu không được chọn thì $c_{20}^m = 0$.

- d. Em nghiên cứu kỹ về bài toán đó để khám phá thêm thuật toán ưu việt hơn hoặc ít nhất phải tìm được lý do, hoặc những khó khăn tồn tại của bài toán.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{20}^m = 1$, nếu không được chọn thì $d_{20}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{21} : Em có lời khuyên nào dành cho các bạn học sinh đang ôn thi học sinh giỏi?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo

nên biến đặc trưng năng lực phi học tập của mình

$$F_{21}^m = \{a_{21}^m, b_{21}^m, c_{21}^m, d_{21}^m, e_{21}^m, f_{21}^m, g_{21}^m, h_{21}^m\}.$$

- a. Bạn hãy học thật thuộc các chương trình hoặc các câu lệnh chính trong chương trình đã được giáo viên dạy.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $a_{21}^m = 1$, nếu không được chọn thì $a_{21}^m = 0$.

- b. Bạn viết lại thật nhiều lần các chương trình đã học để nhớ code.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $b_{21}^m = 1$, nếu không được chọn thì $b_{21}^m = 0$.

- c. Ngoài những bài toán giáo viên đưa ra, nếu có thời gian rảnh thì bạn tự tìm thêm các bài toán khác để thử viết chương trình giải nó để vừa rèn luyện kỹ năng lập trình vừa tăng thêm kiến thức mới.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $c_{21}^m = 1$, nếu không được chọn thì $c_{21}^m = 0$.

- d. Bạn học thuộc các dạng thuật toán đã học và luôn giải bài toán với thuật toán ưu việt nhất.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $d_{21}^m = 1$, nếu không được chọn thì $d_{21}^m = 0$.

- e. Bạn phải hiểu rõ và vận dụng được các thuật toán đã học vào việc giải các bài toán để đạt được kết quả tốt nhất có thể.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $e_{21}^m = 1$, nếu không được chọn thì $e_{21}^m = 0$.

- f. Bạn hãy tự viết chương trình để giải các bài toán chứ không nên sao chép chương trình có sẵn khác mặc dù bạn đã hiểu code.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $f_{21}^m = 1$, nếu không được chọn thì $f_{21}^m = 0$.

- g. Bạn chỉ nên xem tài liệu chính thống được giáo viên cung cấp để đảm bảo kiến thức chuẩn xác nhất.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $g_{21}^m = 1$, nếu không được chọn thì $g_{21}^m = 0$.

- h. Ngoài những bài toán giáo viên đưa ra, bản thân bạn phải tự học, phải tích cực, chủ động học hỏi thêm những kiến thức từ nhiều nguồn như từ bạn bè, thầy cô khác, website, tài liệu không do thầy cô cung cấp, diễn đàn.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $h_{21}^m = 1$, nếu không được chọn thì $h_{21}^m = 0$.*

Câu khảo sát đặc trưng năng lực phi học tập F_{22} : Nếu là giáo viên dạy bồi dưỡng học sinh giỏi em sẽ dạy học sinh của mình như thế nào?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{22}^m = \{a_{22}^m, b_{22}^m, c_{22}^m, d_{22}^m, e_{22}^m\}$*

- a. Em giao bài toán kèm code, sau đó giảng cho người học hiểu về bài toán và thuật toán dựa vào code đã có sẵn để người học dễ hiểu.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $a_{22}^m = 1$, nếu không được chọn thì $a_{22}^m = 0$.*

- b. Em giao bài toán kèm code giải của bài toán đó rồi để người học tự nghiên cứu.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $b_{22}^m = 1$, nếu không được chọn thì $b_{22}^m = 0$.*

- c. Em giao bài toán, hướng dẫn thuật toán để người học tự viết code dựa trên thuật toán đã hướng dẫn.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $c_{22}^m = 1$, nếu không được chọn thì $c_{22}^m = 0$.*

- d. Em giao bài toán, để người học tự suy nghĩ, đưa ra thuật toán và tự viết code để giải bài toán.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $d_{22}^m = 1$, nếu không được chọn thì $d_{22}^m = 0$.*

- e. Em giao bài toán, cho người học tự giải bằng các phương pháp khác nhau, sau đó phân tích và chỉ ra phương pháp tốt nhất.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $e_{22}^m = 1$, nếu không được chọn thì $e_{22}^m = 0$.*

Câu khảo sát đặc trưng năng lực phi học tập F_{23} : Khi em tự tìm tòi, nghiên cứu và phát hiện ra thuật toán mới hay hơn, ưu việt hơn thuật toán mà giáo viên đã dạy, hoặc phát hiện ra một bài toán hay hơn thì em sẽ làm gì?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{23}^m = \{a_{23}^m, b_{23}^m, c_{23}^m, d_{23}^m\}$.

- a. Em không cần phải trao đổi với giáo viên và bạn bè, vì kiến thức đó chưa hẳn giáo viên và bạn bè hiểu được hoặc chấp nhận.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{23}^m = 1$, nếu không được chọn thì $a_{23}^m = 0$.

- b. Em trao đổi với giáo viên và bạn bè ngay, để xem họ có thể hiểu và phân tích thêm bài toán hoặc thuật toán đó hay không.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{23}^m = 1$, nếu không được chọn thì $b_{23}^m = 0$.

- c. Em chỉ trao đổi với giáo viên mà không cần trao đổi với các bạn.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{23}^m = 1$, nếu không được chọn thì $c_{23}^m = 0$.

- d. Em chỉ trao đổi với các bạn mà không cần trao đổi với giáo viên.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{23}^m = 1$, nếu không được chọn thì $d_{23}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{24} : Em nghĩ mình như thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{24}^m = \{a_{24}^m, b_{24}^m, c_{24}^m\}$.

- a. Em là người *trầm tĩnh, bình tĩnh* xử lý công việc trong mọi tình huống.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{24}^m = 1$, nếu không được chọn thì $a_{24}^m = 0$.

- b. Em là người thường *bị run và lo sợ* khi đứng trước một tình huống bất thường.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{24}^m = 1$, nếu không được chọn thì $b_{24}^m = 0$.

- c. Em là người có tính *bốc đồng*, xử lý công việc theo cảm tính.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $f_c^{24} = 1$, nếu không được chọn thì $f_c^{24} = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{25} : Em nghĩ mình như thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{25}^m = \{a_{25}^m, b_{25}^m\}$.

- a. Em là người *năng động* thích tìm phương pháp mới và thích tìm những bài toán mới lạ để giải.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{25}^m = 1$, nếu không được chọn thì $a_{25}^m = 0$.

- b. Em là người *hoạt bát*, thích tranh luận và phô trương những gì mình làm được.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{25}^m = 1$, nếu không được chọn thì $b_{25}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{26} : Em tự cho mình là người như thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{26}^m = \{a_{26}^m, b_{26}^m, c_{26}^m\}$.

- a. Em là người xử lý công việc theo *tình cảm*, chỉ thích học vấn đề em thích và chỉ thích giải những bài toán em thích.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{26}^m = 1$, nếu không được chọn thì $a_{26}^m = 0$.

- b. Em là người xử lý công việc theo *lý trí*, em hăng hái giải bất kỳ bài toán nào mà em nhận được bằng mọi phương pháp có thể, với mục tiêu giải thành công bài toán.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{26}^m = 1$, nếu không được chọn thì $b_{26}^m = 0$.

- c. Em là người cá tính, chỉ muốn giải bài toán theo cách riêng của mình với mục tiêu phải là giải hoàn chỉnh.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{26}^m = 1$, nếu không được chọn thì $c_{26}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{27} : Em đã từng đi thi học sinh giỏi môn Tin học chưa?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{27}^m = \{a_{27}^m, b_{27}^m, c_{27}^m, d_{27}^m, e_{27}^m, f_{27}^m, g_{27}^m, h_{27}^m\}$.

- a. Em đã thi học sinh giỏi môn Tin học cấp tỉnh và thắng giải.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{27}^m = 1$, nếu không được chọn thì $a_{27}^m = 0$.

- b. Em đã thi học sinh giỏi môn Tin học cấp tỉnh nhưng không thắng giải.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{27}^m = 1$, nếu không được chọn thì $b_{27}^m = 0$.

- c. Em đã thi học sinh giỏi môn Tin học Olympic và thắng giải.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{27}^m = 1$, nếu không được chọn thì $c_{27}^m = 0$.

- d. Em đã thi học sinh giỏi môn Tin học Olympic nhưng không thắng giải.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{27}^m = 1$, nếu không được chọn thì $d_{27}^m = 0$.

- e. Em đã từng thi trượt học sinh giỏi môn Tin học cấp quốc gia trước khi thắng giải học sinh giỏi môn Tin học cấp quốc gia.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $e_{27}^m = 1$, nếu không được chọn thì $e_{27}^m = 0$.

- f. Em chưa từng dự thi học sinh giỏi môn Tin học cấp quốc gia trước khi thắng giải học sinh giỏi môn Tin học cấp quốc gia.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $f_{27}^m = 1$, nếu không được chọn thì $f_{27}^m = 0$.

- g. Em chưa từng thắng bất kỳ giải học sinh giỏi môn Tin học nào.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $g_{27}^m = 1$, nếu không được chọn thì $g_{27}^m = 0$.

- h. Em chưa từng thi học sinh giỏi.

Trả lời: Em chọn câu trả lời này ()

Câu khảo sát đặc trưng năng lực phi học tập F_{28} : Em làm gì khi tiếp nhận đề thi?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{28}^m = \{a_{28}^m, b_{28}^m, c_{28}^m, d_{28}^m\}$.

- a. Em triển khai làm ngay theo trình tự từng câu, từng ý trong đề thi.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{28}^m = 1$, nếu không được chọn thì $a_{28}^m = 0$.

- b. Em đọc kỹ đề thi để tìm ra những cụm từ then chốt trong đề thi.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{28}^m = 1$, nếu không được chọn thì $b_{28}^m = 0$.

- c. Em phân tích đề thi để tìm ra input và output của bài toán, input và output của từng phần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{28}^m = 1$, nếu không được chọn thì $c_{28}^m = 0$.

- d. Nếu không tìm ra output của bài toán, em vẫn tiến hành làm bài vì lo sợ thiếu thời gian.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{28}^m = 1$, nếu không được chọn thì $d_{28}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{29} : Em làm gì sau khi tìm ra input và output của bài toán?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{29}^m = \{a_{29}^m, b_{29}^m\}$.

- a. Em phân tích đề thi cùng với input và output mà em mới tìm ra để nhận biết thách thức của bài toán và tìm giải pháp để vượt qua thách thức.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $a_{29}^m = 1$, nếu không được chọn thì $a_{29}^m = 0$.

- b. Em tiến hành làm bài mà không cần nhận biết thách thức của bài toán vì lo sợ thiếu thời gian.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $b_{29}^m = 1$, nếu không được chọn thì $b_{29}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{30} : Em làm gì khi tiến hành giải bài toán?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào đáp án phù hợp nhất và chỉ chọn tối đa 01 đáp án, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{30}^m = \{a_{30}^m, b_{30}^m\}$.

- a. Em tập trung làm bài không quan tâm đến thách thức của bài toán.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $a_{30}^m = 1$, nếu không được chọn thì $a_{30}^m = 0$.

- b. Em tiến hành làm bài đồng thời tìm cách nhận biết thách thức của bài toán để linh hoạt áp dụng phương pháp giải phù hợp.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $b_{30}^m = 1$, nếu không được chọn thì $b_{30}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{31} : Em làm thế nào để giải bài toán?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào đáp án phù hợp nhất và chỉ chọn tối đa 01 đáp án, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{31}^m = \{a_{31}^m, b_{31}^m\}$.

- a. Trước khi tiến hành giải bài toán, em lập quy trình để giải bài toán bằng cách vẽ lên giấy hoặc hình dung trong trí của mình.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $a_{31}^m = 1$, nếu không được chọn thì $a_{31}^m = 0$.

- b. Em tiến hành làm bài mà không cần lập quy trình giải, cứ làm xong nội dung trước thì trí não của em tự động phát sinh cách giải quyết nội dung kế tiếp cho đến khi làm xong hết đề thi.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $b_{31}^m = 1$, nếu không được chọn thì $b_{31}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{32} : Thời gian chơi game của em trong tuần như thế nào?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên **biến đặc trưng năng lực phi học tập của mình**
 $F_{32}^m = \{a_{32}^m, b_{32}^m, c_{32}^m, d_{32}^m, e_{32}^m, f_{32}^m, g_{32}^m, h_{32}^m, i_{32}^m\}$.*

- a. Em chơi game tùy thích mọi lúc mọi nơi, không theo bất kỳ lịch hay kế hoạch nào.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $a_{32}^m = 1$, nếu không được chọn thì $a_{32}^m = 0$.

- b. Em chơi game mỗi ngày tùy theo ý thích của em.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $b_{32}^m = 1$, nếu không được chọn thì $b_{32}^m = 0$.

- c. Em chơi game 2 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $c_{32}^m = 1$, nếu không được chọn thì $c_{32}^m = 0$.

- d. Em chơi game 3 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $d_{32}^m = 1$, nếu không được chọn thì $d_{32}^m = 0$.

- e. Em chơi game 4 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $e_{32}^m = 1$, nếu không được chọn thì $e_{32}^m = 0$.

- f. Em chơi game 5 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được chọn thì $f_{32}^m = 1$, nếu không được chọn thì $f_{32}^m = 0$.

- g. Em chơi game 6 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $g_{32}^m = 1$, nếu không được chọn thì $g_{32}^m = 0$.

- h. Em chơi game 7 ngày/1 lần.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $g_{32}^m = 1$, nếu không được chọn thì $g_{32}^m = 0$.

- i. Em không chơi game.

Trả lời: Em chọn câu trả lời này ()

Câu khảo sát đặc trưng năng lực phi học tập F_{33} : Thời gian mỗi lần chơi game của em như thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên **biến đặc trưng năng lực phi học tập của mình**
 $F_{33}^m = \{a_{33}^m, b_{33}^m, c_{33}^m, d_{33}^m, e_{33}^m, f_{33}^m, g_{33}^m, h_{33}^m\}$.

- a. Em chơi game mỗi lần khoảng 1 giờ và em tự dừng.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{33}^m = 1$, nếu không được chọn thì $a_{33}^m = 0$.

- b. Em chơi game mỗi lần khoảng 2 - 3 giờ và em tự dừng.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{33}^m = 1$, nếu không được chọn thì $b_{33}^m = 0$.

- c. Em chơi game mỗi lần khoảng 3 - 4 giờ và em tự dừng

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{33}^m = 1$, nếu không được chọn thì $c_{33}^m = 0$.

- d. Em chơi game mỗi lần khoảng 4 - 5 giờ và em tự dừng.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{33}^m = 1$, nếu không được chọn thì $d_{33}^m = 0$.

- e. Em chơi game mỗi lần khoảng 5 - 6 giờ và em tự dừng.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $e_{33}^m = 1$, nếu không được chọn thì $e_{33}^m = 0$.

- f. Em chơi game bất kỳ lúc nào em thích và chỉ dừng khi ba mẹ hoặc người thân nhắc nhở.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $f_{33}^m = 1$, nếu không được chọn thì $f_{33}^m = 0$.*

- g. Em chơi game bất kỳ lúc nào em cảm thấy rồi rãnh (free) và dừng khi em cảm thấy chơi đủ rồi.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $g_{33}^m = 1$, nếu không được chọn thì $g_{33}^m = 0$.*

- h. Em không chơi game.

Trả lời: Em chọn câu trả lời này ()

Câu khảo sát đặc trưng năng lực phi học tập F_{34} : Em có lịch làm việc, học tập, giải trí, ngủ nghỉ trong ngày, trong tuần không?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{34}^m = \{a_{34}^m, b_{34}^m\}$.*

- a. Em thường xuyên làm việc, học tập, giải trí, ngủ nghỉ trong ngày, trong tuần theo lịch đã soạn trước hợp lý, cân đối thời gian hợp lý giữa các nội dung và phù hợp với bản thân cũng như gia đình.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $a_{34}^m = 1$, nếu không được chọn thì $a_{34}^m = 0$.*

- b. Em làm việc, học tập, giải trí, ngủ nghỉ bất kỳ lúc nào em muốn, không theo lịch thời gian cố định.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $b_{34}^m = 1$, nếu không được chọn thì $b_{34}^m = 0$.*

Câu khảo sát đặc trưng năng lực phi học tập F_{35} : Em có chơi thể thao không?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{35}^m = \{a_{35}^m, b_{35}^m, c_{35}^m, d_{35}^m\}$.*

- a. Em chơi thể thao hoặc tập thể dục theo lịch ngày và tuần của em.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{35}^m = 1$, nếu không được chọn thì $a_{35}^m = 0$.

- b. Em chơi thể thao hoặc tập thể dục tùy hứng, bất cứ lúc nào em muốn.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{35}^m = 1$, nếu không được chọn thì $b_{35}^m = 0$.

- c. Em không chơi thể thao hoặc tập thể dục vì không có thời gian mặc dù em rất thích.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{35}^m = 1$, nếu không được chọn thì $c_{35}^m = 0$.

- d. Em không chơi thể thao hoặc tập thể dục vì em không thích.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{35}^m = 1$, nếu không được chọn thì $d_{35}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{36} : Mức độ kiên trì, say mê làm việc code trên máy tính của em như thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{36}^m = \{a_{36}^m, b_{36}^m, c_{36}^m\}$.

- a. Em say mê và kiên trì ngồi code hoặc làm việc trên máy tính, không để ý thời gian ăn, nghỉ ngơi. Ngoài bài tập của thầy cô giao, em còn tìm thêm nhiều bài khác để làm.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{36}^m = 1$, nếu không được chọn thì $a_{36}^m = 0$.

- b. Em chỉ ngồi code trên máy tính khi thầy cô giao bài, yêu cầu làm bài vì em còn phải học những môn học khác.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{36}^m = 1$, nếu không được chọn thì $b_{36}^m = 0$.

- c. Em chỉ ngồi vào máy tính để code khi được thầy cô nhắc nhở.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{36}^m = 1$, nếu không được chọn thì $c_{36}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{37} : Khả năng tài chính của gia đình em thế nào?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{37}^m = \{a_{37}^m, b_{37}^m, c_{37}^m, d_{37}^m\}$.

- a. Gia đình em là gia đình giàu có so với dân cư của địa phương, em không mất thời gian vào các công việc của gia đình, em có thời gian thoải mái trong việc tự học, tự nghiên cứu, em không phải lo lắng về chi phí học thêm, mua máy tính hoặc tài liệu học tập.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $a_{37}^m = 1$, nếu không được chọn thì $a_{37}^m = 0$.

- b. Gia đình em là gia đình khá giả so với dân cư của địa phương, em có tham gia một số ít việc nhà nhưng có đủ thời gian để tự học, tự nghiên cứu, em có thể có tiền học thêm nhưng không nhiều, em có thể mua máy tính hoặc tài liệu học tập.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $b_{37}^m = 1$, nếu không được chọn thì $b_{37}^m = 0$.

- c. Gia đình em là gia đình bình thường so với dân cư của địa phương, em phải làm hầu hết việc nhà, có ít khả năng về tiền để học thêm, em có thể mua máy tính hoặc tài liệu học tập.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $c_{37}^m = 1$, nếu không được chọn thì $c_{37}^m = 0$.

- d. Gia đình em là gia đình nghèo so với dân cư của địa phương, em phải làm việc nhà và phụ giúp cha mẹ làm thêm việc kiếm tiền sống, hoàn toàn không có tiền để học thêm và không thể mua máy tính hoặc tài liệu học tập thêm.

Trả lời: Em chọn câu trả lời này ()

Nếu câu này được **chọn** thì $d_{37}^m = 1$, nếu không được chọn thì $d_{37}^m = 0$.

Câu khảo sát đặc trưng năng lực phi học tập F_{38} : Ba mẹ em có muốn em tham gia thi và thắng giải học sinh giỏi môn Tin học không?

Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{38}^m = \{a_{38}^m, b_{38}^m, c_{38}^m, d_{38}^m\}$.

- a. Ba mẹ em rất muốn em được tham gia đội tuyển để thi và muốn em thắng giải học sinh giỏi môn Tin học cấp quốc gia.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $a_{38}^m = 1$, nếu không được chọn thì $a_{38}^m = 0$.*

- b. Ba mẹ em muốn em tham gia đội tuyển để thi học sinh giỏi môn Tin học nhưng không quan trọng việc em có thắng giải hay không.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $b_{38}^m = 1$, nếu không được chọn thì $b_{38}^m = 0$.*

- c. Ba mẹ em không quan tâm đến việc em có được vào đội tuyển để dự thi học sinh giỏi hay không.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $c_{38}^m = 1$, nếu không được chọn thì $c_{38}^m = 0$.*

- d. Ba mẹ em hoàn toàn không để ý đến việc học và kết quả học tập của em.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $d_{38}^m = 1$, nếu không được chọn thì $d_{38}^m = 0$.*

Câu khảo sát đặc trưng năng lực phi học tập F_{39} : Em đã có bạn (hay người yêu) chưa?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{39}^m = \{a_{39}^m, b_{39}^m, c_{39}^m, d_{39}^m, e_{39}^m\}$*

- a. Em đã có người yêu.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $a_{39}^m = 1$, nếu không được chọn thì $a_{39}^m = 0$.*

- b. Em mới có tình ý với một bạn.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $b_{39}^m = 1$, nếu không được chọn thì $b_{39}^m = 0$.*

- c. Em chưa quan tâm đến việc yêu đương nam nữ.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $c_{39}^m = 1$, nếu không được chọn thì $c_{39}^m = 0$.*

- d. Em đã chia tay với người yêu của em lâu rồi.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $d_{39}^m = 1$, nếu không được chọn thì $d_{39}^m = 0$.*

- e. Em vừa mới chia tay với người yêu của em.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $d_{39}^m = 1$, nếu không được chọn thì $d_{39}^m = 0$.*

Câu khảo sát đặc trưng năng lực phi học tập F_{40} : Người yêu (hoặc bạn nam nữ) của em có muốn em tham gia thi và thắng giải học sinh giỏi môn Tin học không?

*Người được khảo sát $m, m=1, \dots, M$ được mời đọc đáp án và trả lời bằng cách đánh dấu (X) vào **đáp án phù hợp nhất** và **chỉ chọn tối đa 01 đáp án**, để tạo nên biến đặc trưng năng lực phi học tập của mình $F_{40}^m = \{a_{40}^m, b_{40}^m, c_{40}^m, d_{40}^m, e_{40}^m\}$*

- a. Người yêu của em rất muốn em được tham gia đội tuyển để thi và muốn em thắng giải học sinh giỏi môn Tin học cấp quốc gia.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $a_{40}^m = 1$, nếu không được chọn thì $a_{40}^m = 0$.*

- b. Người yêu của em muốn em tham gia đội tuyển để thi học sinh giỏi môn Tin học nhưng không quan trọng việc em có thắng giải hay không.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $b_{40}^m = 1$, nếu không được chọn thì $b_{40}^m = 0$.*

- c. Người yêu của em không quan tâm đến việc em có được vào đội tuyển để dự thi học sinh giỏi hay không.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $c_{40}^m = 1$, nếu không được chọn thì $c_{40}^m = 0$.*

- d. Người yêu của em hoàn toàn không muốn em tham gia đội tuyển và không muốn em dự thi học sinh giỏi.

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $d_{40}^m = 1$, nếu không được chọn thì $d_{40}^m = 0$.*

- e. Em không có người yêu (hoặc bạn nam nữ).

Trả lời: Em chọn câu trả lời này ()

*Nếu câu này được **chọn** thì $e_{40}^m = 1$, nếu không được chọn thì $e_{40}^m = 0$.*

Số: 205/QĐ-HVKHCN

Hà Nội, ngày 11 tháng 03 năm 2026

QUYẾT ĐỊNH
Về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện

GIÁM ĐỐC
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Căn cứ Quyết định số 364/QĐ-VHL ngày 01/3/2025 của Chủ tịch Viện Hàn lâm Khoa học và Công nghệ Việt Nam về việc ban hành Quy chế Tổ chức và hoạt động của Học viện Khoa học và Công nghệ;

Căn cứ Thông tư số 08/2017/TT-BGDĐT ngày 04/4/2017 của Bộ Giáo dục và Đào tạo về việc ban hành Quy chế Tuyển sinh và Đào tạo trình độ Tiến sĩ;

Căn cứ Quyết định số 1948/QĐ-HVKHCN ngày 28/12/2018 của Giám đốc Học viện Khoa học và Công nghệ về việc ban hành Quy định đào tạo trình độ tiến sĩ;

Căn cứ Quyết định số 2019/QĐ-HVKHCN ngày 10/12/2020 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận nghiên cứu sinh đợt 2 năm 2020;

Căn cứ Quyết định số 1353/QĐ-HVKHCN ngày 10/12/2025 của Giám đốc Học viện Khoa học và Công nghệ về việc gia hạn lần 2 thời gian học tập cho nghiên cứu sinh Huỳnh Thị Ngọc Cẩm;

Xét đề nghị của Trưởng phòng Đào tạo.

QUYẾT ĐỊNH:

Điều 1. Thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện cho nghiên cứu sinh Huỳnh Thị Ngọc Cẩm với đề tài:

Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Danh sách thành viên Hội đồng đánh giá luận án kèm theo Quyết định này.

Điều 2. Hội đồng có trách nhiệm đánh giá luận án tiến sĩ theo đúng quy chế hiện hành của Bộ Giáo dục và Đào tạo, của Học viện Khoa học và Công nghệ. Quyết định có hiệu lực tối đa 90 ngày kể từ ngày ký. Hội đồng tự giải thể sau khi hoàn thành nhiệm vụ.

Điều 3. Trưởng phòng Tổ chức - Hành chính, Trưởng phòng Đào tạo, Trưởng phòng Kế toán, các thành viên có tên trong danh sách Hội đồng và nghiên cứu sinh có tên tại Điều 1 chịu trách nhiệm thi hành Quyết định này. /.

Nơi nhận:

- Như Điều 3;
- Lưu hồ sơ NCS;
- Lưu: VT, ĐT.PQ.15.



GIÁM ĐỐC

GS. TS. Vũ Đình Lâm



**DANH SÁCH HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ
CẤP HỌC VIỆN**

*(Kèm theo Quyết định số 205/QĐ-HVKHCN ngày 11/03/2026
của Giám đốc Học viện Khoa học và Công nghệ)*

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Đề tài: Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh
đội tuyển

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Người hướng dẫn chính: PGS.TS. Trần Vĩnh Phước,

Trường Đại học Tài Nguyên và Môi trường TP.HCM
Bộ Nông nghiệp và Môi trường

TT	Họ và tên, học hàm, học vị	Ngành	Cơ quan công tác	Trách nhiệm trong Hội đồng
1	PGS.TS. Nguyễn Long Giang	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Chủ tịch
2	TS. Nguyễn Như Sơn	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Phản biện 1
3	PGS. TS. Lê Kim Hùng	Truyền thông thông tin và điện tử	Trường Đại học Công nghệ thông tin, Đại học Quốc gia TP.HCM	Phản biện 2
4	PGS.TS. Trần Trung Duy	Công nghệ thông tin	Học viện Bưu chính viễn thông, Bộ Khoa học và Công nghệ	Phản biện 3
5	TS. Nguyễn Tấn Nô	Kỹ thuật điều khiển và Tự động hoá	Trường Đại học Sư phạm Kỹ thuật Vĩnh Long, Bộ Giáo dục và Đào tạo	Ủy viên - Thư ký
6	PGS.TS. Phạm Thế Bảo	Khoa học máy tính	Trường Đại học Sài Gòn, Ủy ban nhân dân TP.HCM	Ủy viên
7	PGS.TS. Nguyễn Tấn Cẩm	Công nghệ thông tin	Trường Đại học Công nghệ thông tin, Đại học Quốc gia TP.HCM	Ủy viên

Hội đồng gồm 07 thành viên./.

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

ĐẢNG CỘNG SẢN VIỆT NAM

Tp.HCM, ngày 31 tháng 03 năm 2026

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ
(Cấp Học viện)

Tên nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Đề tài: Tiếp cận học máy để tuyển chọn và trực quan hóa đề huấn luyện học sinh đội tuyển

Chuyên ngành: Khoa học máy tính

Mã số: 9 48 01 01

Người nhận xét: PGS.TS. Nguyễn Long Giang, Chủ tịch Hội đồng

Cơ quan công tác của người nhận xét: Viện Công nghệ thông tin, Viện Hàn lâm KH&CN Việt Nam.

NỘI DUNG NHẬN XÉT

1. Ý nghĩa khoa học và thực tiễn của đề tài luận án

Luận án xây dựng quy trình tuyển chọn đội viên đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông và huấn luyện đội tuyển để tham dự kỳ thi học sinh giỏi cấp quốc gia với kỳ vọng thắng được nhiều giải, dựa trên tiếp cận học máy và trực quan hóa dữ liệu trong khoa học máy tính. Các kết quả nghiên cứu của đề tài có ý nghĩa khoa học và thực tiễn trong lĩnh vực nghiên cứu về khai phá dữ liệu giáo dục.

2. Sự hợp lý và độ tin cậy của các phương pháp nghiên cứu

Phương pháp nghiên cứu của luận án tập trung nhiều về lý thuyết kết hợp cài đặt, thử nghiệm. Về lý thuyết, luận án xây dựng, đề xuất mô hình chọn học sinh giỏi dựa vào học máy và học khám phá có hướng dẫn trực quan. Mô hình được thử nghiệm, đánh giá trên các bộ dữ liệu được xây dựng nhằm minh chứng tính hiệu

quả của mô hình đề xuất. Do đó, về cơ bản phương pháp nghiên cứu của luận án là hợp lý và tin cậy.

3. Đánh giá các kết quả đạt được, nêu những đóng góp mới và giá trị của các đóng góp đó

Luận án nghiên cứu, giải quyết bài toán tuyển chọn ứng viên và huấn luyện đội tuyển dự thi học sinh giỏi dựa vào năng lực (học tập, phi học tập) và mô hình học khám phá có hướng dẫn bằng kỹ thuật trực quan, các kết quả chính của luận án bao gồm:

- 1) Xây dựng bộ dữ liệu năng lực của học sinh (gồm năng lực học tập, phi học tập) với 40 thuộc tính, trong đó 18 thuộc tính của năng lực học tập, 22 thuộc tính của năng lực phi học tập dựa trên mở rộng nghiên cứu số [42]
- 2) Sử dụng kỹ thuật winner-domain xác định vùng năng lực của các thí sinh thắng giải (winners) trong không gian năng lực nhiều chiều để xác 6 định miền năng lực có tiềm năng thắng giải học sinh giỏi cấp quốc gia môn Tin học.
- 3) Đề xuất mô hình học khám phá có hướng dẫn bằng kỹ thuật trực quan để huấn luyện đội viên đội tuyển môn Tin học trước khi tham dự kỳ thi học sinh giỏi cấp quốc gia.

Các đóng góp của luận án có ý nghĩa khoa học và thực tiễn trong lĩnh vực nghiên cứu về khai phá dữ liệu giáo dục.

4. Những ưu điểm và thiếu sót, những điểm cần được bổ sung và sửa chữa

Về cơ bản, cấu trúc và bố luận án trình bày logic, hợp lý. Bài toán giải quyết có ý nghĩa khoa học và thực tiễn trong lĩnh vực nghiên cứu về khai phá dữ liệu giáo dục.

Một số ý kiến chỉnh sửa, bổ sung luận án như sau

- 1) Nên xem xét chỉnh sửa lại tên các đề mục của luận án cho hợp lý, logic, không dùng các câu hỏi ở tên chương/mục như Chương 2, không trích dẫn tài liệu tham khảo ở tên mục (Chương 1). Chỉnh sửa lại tên chương cho hợp lý (ví dụ tên Chương 1, Chương 3 nên đặt là xây dựng bộ dữ liệu...)

- 2) Phần mở đầu, cần mô tả rõ hơn bài toán cần giải quyết về khoa học máy tính và bức tranh nghiên cứu liên quan đến bài toán lựa chọn ứng viên cho đội tuyển và bài toán huấn luyện đội tuyển theo mô hình học khám phá. Cần cô đọng lại các đóng góp chính của luận án về mô hình/Phương pháp (hiện tại quá dài và lan man)
- 3) Chương 1: Cần trình bày cô đọng vào bài toán cần giải quyết trong khai phá dữ liệu giáo dục và lược bớt các nội dung tổng quan về khai phá dữ liệu giáo dục và các phương pháp học máy. Cần bổ sung phần các nghiên cứu liên quan, phân tích các hạn chế, tồn tại của các nghiên cứu trước đây từ đó làm rõ các vấn đề nghiên cứu, các thách thức còn tồn tại của luận án. Phần Chương 2 về học khám phá cần trình bày cô đọng và cân nhắc gộp vào chương 1 về cơ sở lý thuyết.
- 4) Về Chương 3 xây dựng bộ dữ liệu, cần trình bày cô đọng về quy trình xây dựng và đặc tả bộ dữ liệu sau khi được xây dựng. Phần trình bày các biến quá dài dòng và nên cô đọng lại. Nên bổ sung các thử nghiệm, đánh giá để bảo đảm tính xác thực của dữ liệu.
- 5) Về kết quả ở Chương 4, cần làm rõ các đóng góp, cải tiến của luận án về mô hình Winner-domains (cải tiến gì so với mô hình trong 106, 107). Chưa rõ mô hình học máy được sử dụng ở chương này như thế nào? Về kết quả thực nghiệm, mô tả cô đọng bộ dữ liệu và kịch bản thử nghiệm, phân tích tính hiệu quả của phương pháp đề xuất so với các phương pháp khác ?
- 6) Về kết quả ở chương 5, nên có mô hình tổng thể đề xuất trước khi trình bày các nội dung chi tiết. Nên xem xét bổ sung thử nghiệm/so sánh/đánh giá với các kỹ thuật khác để làm rõ tính hiệu quả của mô hình học khám phá được trình bày.

5. Đánh giá về sự trùng lặp của luận án so với các đồ án, luận văn hay công trình khoa học đã công bố trong và ngoài nước

Theo hiểu biết của người nhận xét, kết quả của luận án không trùng lặp với các đồ án, luận văn hay công trình khoa học đã công bố trong và ngoài nước.

6. Nhận xét về chất lượng các bài báo khoa học đã được công bố và khẳng định các bài báo đó chứa đựng nội dung chủ yếu của luận án.

Các kết quả của luận án được công bố trong các bài báo đăng trên kỷ yếu, hội thảo chuyên ngành, trong đó có 01 bài báo trên tạp chí SCIE-Q2 và 03 bài báo trên Kỷ yếu hội thảo quốc tế. Luận án đáp ứng yêu cầu về công bố khoa học.

7. Tính trung thực trong việc trích dẫn các công trình đã được NCS công bố trong và ngoài nước, tài liệu tham khảo

Các kết quả nghiên cứu của luận án đã được trích dẫn trung thực trong các công trình công bố.

6. Kết luận

- Luận án đáp ứng được các yêu cầu về nội dung và hình thức đối với một luận án tiến sĩ, chuyên ngành Khoa học máy tính.
- Đồng ý cho NCS được bảo vệ tại Hội đồng chấm luận án cấp Học viện.
- Đề nghị NCS chỉnh sửa, bổ sung luận án theo các đóng góp nêu trên.

Tp.HCM, ngày tháng năm 2026

Người nhận xét



Nguyễn Long Giang

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

ĐẢNG CỘNG SẢN VIỆT NAM

Tp.HCM, ngày 31 tháng 03 năm 2026

*

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài luận án: TIẾP CẬN HỌC MÁY ĐỂ TUYỂN CHỌN VÀ TRỰC QUAN HÓA ĐỀ
HUẤN LUYỆN HỌC SINH ĐỘI TUYỂN

Chuyên ngành: Khoa học Máy tính

Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn: PGS.TS. Trần Vĩnh Phước

Người nhận xét: TS. Nguyễn Như Sơn, Phản biện 1

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KH&CN Việt Nam.

NỘI DUNG NHẬN XÉT

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Từ nhiều năm qua, các cơ sở giáo dục đã và đang áp dụng mô hình học truyền thống - mô hình học mô tả (expository learning model) với người dạy là trung tâm của hoạt động dạy và học, người học thụ động trong việc tiếp nhận thông tin từ người dạy, mô hình này có một số nhược điểm. Từ đó, luận án thực hiện xây dựng mô hình Học khám phá (discovery learning), là một mô hình học tập phát huy sự tích cực tự khám phá của người học để kiến thức được lưu trữ lâu dài trong não bộ. Trên cơ sở bộ dữ liệu của học sinh thực hiện các kỳ thi học sinh giỏi môn Tin học, luận án đã nghiên cứu xây dựng thuật toán và qui trình tuyển chọn học sinh giỏi vào đội tuyển tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi tin học cấp quốc gia.

Xuất phát từ những vấn đề nêu trên cho thấy, việc nghiên cứu của luận án có tính thực tiễn thời sự đối với ngành giáo dục và có tính khoa học.

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Theo hiểu biết của người nhận xét, kết quả của luận án không trùng lặp với các công trình, luận án đã công bố ở trong và ngoài nước.

3. Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với chuyên ngành và mã số chuyên ngành.

Tên đề tài phù hợp với nội dung luận án. Nội dung luận án phù hợp với chuyên ngành và mã số chuyên ngành Khoa học Máy tính.

4. Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

Phương pháp nghiên cứu của luận án tập trung vào nghiên cứu lý thuyết và thực nghiệm. Các đề xuất, cải tiến trong luận án đều được thử nghiệm nhằm minh chứng cho tính hiệu quả của phương pháp. Do đó, phương pháp nghiên cứu sử dụng trong luận án là hợp lý, cơ bản bảo đảm tính khoa học và độ tin cậy.

5. Kết quả nghiên cứu mới của tác giả.

Đóng góp mới của tác giả tập trung vào các nội dung chính sau đây:

1. Thành lập dataset thí sinh dự thi và đề xuất mô hình winner-domain: Dataset được thành lập theo môn học, theo cấp thi và theo thời gian sẽ thi. Trong đó, mỗi thí sinh được biểu diễn như một vector năng lực nhiều biến, dataset của một môn học là tập dữ liệu năng lực thí sinh dự thi của một môn học đó, một cấp thi trong vài năm trước đó gần nhất.

Luận án áp dụng tiếp cận học máy trong khoa học máy tính để thành lập winner-domain của môn thi, cấp thi, năm thi dựa trên dataset tương ứng. Luận án áp dụng độ tương đồng cosin (Cosin Similarity) giữa vector năng lực của ứng viên với tất cả vector năng lực của winners. Tiếp cận cây quyết định giúp luận án chọn đội viên cho đội tuyển từ những ứng viên có đặc trưng năng lực trong miền không gian của Winner-domain tốt.

2. Mô hình học khám phá có hướng dẫn trực quan. Luận án xây dựng mô hình học khám phá có hướng dẫn trực quan với qui trình thiết kế bài giảng trực quan. Mô hình học khám phá có hướng dẫn trực quan được thực nghiệm để huấn luyện đội tuyển tin học.

Các nghiên cứu đã được đăng tải trên các hội nghị và tạp chí chuyên ngành.

6. Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án.

Luận án có bố cục cơ bản hợp lý, cơ bản đảm bảo mạch logic giữa các chương mục. Kết luận của luận án phù hợp với các vấn đề luận án đã đặt ra. Tóm tắt luận án phản ánh nội dung của luận án.

Một số ý kiến nhận xét về chỉnh sửa, bổ sung luận án.

- Hiện còn có 5 chương (khá nhiều), có thể xem xét gộp chương 1 (11 trang) và chương 2 (10 trang) là các cơ sở lý thuyết ban đầu, hiện các chương còn khá ngắn?
- Tên các tiêu đề mục nhỏ không nên đề tham khảo đến các tài liệu []?
- Các chương nên có phần Kết luận chương?
- Các kết quả chính ở các chương cuối cần tham chiếu đến các công trình đã xuất bản có nội dung tương ứng?
- Mô hình thuật toán đề xuất ở chương 3 cần được trình bày theo chuẩn của khoa học máy tính, với Input, Output, có Begin – End, Return Output..., có đánh giá độ phức tạp O của thuật toán?
- Bài toán thực nghiệm ở chương 3 hiện chưa rõ có công trình nào tương tự hay không để thực hiện việc so sánh về các đề xuất, kiểm chứng kết quả thực nghiệm thêm trong luận án?
- Luận án không được sử dụng các Công trình công bố trong Tài liệu tham khảo, ví dụ [106] (hình 4.1 trang 55), cần rà soát toàn bộ và sử dụng tham chiếu đến công trình tương ứng [CT2]?
- Thuật toán Winner-cosin ở 4.3.4 cần trình bày rõ ràng theo mẫu đã nói ở trên?
- Kết luận chương 4 cần súc tích ngắn gọn, các Bảng 4.31, 4.32 chi tiết cần trình bày ở phần trên.
- Mô hình học khám phá có hướng dẫn trực quan đề xuất ở chương 5 hiện chưa có các số liệu thực nghiệm để kiểm chứng mô hình?. Vấn đề so sánh các mô hình khác hiện cũng chưa có đề cập đến rõ ràng, do có liên quan đến các phương pháp đào tạo trong giáo dục.

- Phần Kết luận nên viết ngắn gọn hơn các đóng góp chính của luận án (không nhắc lại các kết quả thực nghiệm), cần có mục nêu hướng phát triển tiếp theo chi tiết hơn?.
- Các tài liệu tham khảo hiện tập trung nhiều cho chuyên ngành Giáo dục, cần rà soát nhấn mạnh chuyên ngành về Khoa học máy tính trong những năm gần đây.

7. Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố.

Các kết quả nghiên cứu này đã được công bố trên các tạp chí và được trao đổi trong các báo cáo tại những hội thảo khoa học chuyên ngành bao gồm 04 công trình khoa học, trong đó có 01 bài báo tạp chí quốc tế thuộc danh mục SCIE, 03 bài báo kỷ yếu hội thảo quốc tế trong danh mục Scopus.

NCS là tác giả chính đầu tiên trong tất cả 04 công trình. Nội dung của các công trình phù hợp với nội dung nghiên cứu của luận án và là các kết quả chính của luận án.

8. Kết luận

Luận án của NCS Huỳnh Thị Ngọc Cẩm cơ bản đáp ứng các yêu cầu đối với một luận án tiến sĩ chuyên ngành Khoa học Máy tính. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án. Luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị Tiến sĩ.

TP.HCM, ngày 31 tháng 03 năm 2026

Người nhận xét



TS. Nguyễn Như Sơn

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: TIẾP CẬN HỌC MÁY ĐỂ TUYỂN CHỌN VÀ TRỰC QUAN
HÓA ĐỂ HUẤN LUYỆN HỌC SINH ĐỘI TUYỂN

Ngành: Khoa học máy tính Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn:

Người phản biện: PGS.TS Lê Kim Hùng

Cơ quan công tác: Trường đại học Công nghệ Thông tin

Nội dung nhận xét:

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

1.1. Tính cần thiết và thời sự

Đề tài luận án tập trung vào bài toán tuyển chọn và huấn luyện đội tuyển học sinh giỏi môn Tin học cấp quốc gia - một bài toán mang tính thường niên tại các trường Trung học phổ thông (THPT) chuyên ở Việt Nam. Công tác bồi dưỡng học sinh giỏi quốc gia là nhiệm vụ trọng tâm được Bộ Giáo dục và Đào tạo đặc biệt quan tâm, đồng thời đang chịu áp lực lớn khi quy trình tuyển chọn truyền thống chủ yếu dựa vào kinh nghiệm cá nhân của giáo viên và điểm thi các vòng loại - dễ dẫn tới sai số chủ quan, bỏ sót ứng viên tiềm năng, và khó mở rộng ở quy mô hệ thống. Vì vậy, hướng tiếp cận dựa trên học máy (Machine Learning) và khai phá dữ liệu giáo dục (Educational Data Mining - EDM) để hỗ trợ ra quyết định là phù hợp với xu thế chuyển đổi số trong giáo dục, đặc biệt trong bối cảnh các công nghệ AI đang được đẩy mạnh ứng dụng vào giáo dục bậc trung học từ 2022 trở lại đây.

Đồng thời, đề tài kết hợp hai trụ cột: (i) thuật toán học máy có giám sát để dự đoán khả năng đạt giải; (ii) mô hình học khám phá có hướng dẫn trực quan (Guided Discovery Learning with Visualization) để hỗ trợ huấn luyện - cho thấy định hướng liên ngành giữa Khoa học máy tính và Khoa học giáo dục. Đây là hướng tiếp cận toàn diện, phù hợp với xu hướng quốc tế về AI-in-Education (AIED) và Learning Analytics (LA).

1.2. Ý nghĩa khoa học

Về mặt khoa học, luận án đề xuất: thuật toán Winner-domain và Winner-cosin kết hợp cây quyết định (Chương 4), mô hình toán học lượng hóa hồi đáp định tính bằng ma trận và khối 3D hồi đáp (Chương 3), cũng như mô hình đánh giá hiệu năng bài giảng $E = O/(B \cdot T)$ và biến thể mở rộng (Chương 5). Các đóng góp này nếu được kiểm chứng chặt chẽ có thể đóng góp vào nhánh one-class/centroid-based classification và phân tích dữ liệu khảo sát trong ngữ cảnh giáo dục.

1.3. Ý nghĩa thực tiễn

Luận án có giá trị thực tiễn rõ ràng: dataset thu thập được (Dataset 2024: 221 mẫu; Dataset 2025: 497 mẫu) từ học sinh THPT chuyên nhiều tỉnh thành miền Tây Nam Bộ, và quy trình đã được triển khai tại ít nhất 01 trường THPT chuyên cho 02 kỳ thi HSG quốc gia liên tiếp (2023-2024 và 2024-2025) với kết quả thực tế cải thiện so với phương pháp truyền thống.

1.4. Những điểm cần cải thiện và bổ sung

Đề tăng sức thuyết phục về tính cần thiết, thời sự và ý nghĩa của đề tài, NCS cần bổ sung/cải thiện các điểm sau:

- Phần "Tính cấp thiết" (Mở đầu) chỉ dừng ở mô tả định tính, chưa có phân tích định lượng quy mô vấn đề: số lượng trường THPT chuyên (cả nước có 78 trường), số đội tuyển quốc gia các môn, tỷ lệ đạt giải so với số dự thi, thống kê về sai số tuyển chọn hiện tại. Khi làm rõ các con số này, tầm quan trọng của bài toán mới được định lượng.
- Khoảng trống nghiên cứu (research gap) trong mục 1.4 còn mơ hồ. NCS cần trình bày rõ gap theo bảng so sánh: các công trình quốc tế về "talent identification", "early identification of gifted students using ML", "predicting competition outcome" đang giải quyết gì, bỏ sót gì, và đóng góp cụ thể của luận án lấp khoảng trống nào.
- Thiếu khảo sát các công trình quốc tế chuyên về AI/ML cho tuyển chọn học sinh giỏi, đặc biệt các công trình tại ITS (Intelligent Tutoring Systems), AIED, LAK, EDM conference trong giai đoạn 2020-2025. Chỉ dựa vào EDM tổng quát là chưa đủ để xác lập vị trí khoa học của luận án.
- Bản chất bài toán (phân loại nhị phân với cỡ mẫu dương rất nhỏ, dữ liệu không cân bằng, số chiều $\approx$ số mẫu/5) cần được đặt đúng vào một trong các khung lý thuyết chuẩn: one-class classification, imbalanced learning, positive-unlabeled learning, small-sample learning. NCS chưa thực hiện bước định vị này một cách tường minh, dẫn tới nhiều chọn lựa phương pháp và baseline chưa tối ưu.
- Ý nghĩa thực tiễn cần được trình bày kèm phân tích tác động: ước tính giảm chi phí huấn luyện, giảm số giờ công của giáo viên, tăng tỷ lệ đạt giải, khả năng áp dụng rộng cho các môn học sinh giỏi khác (Toán, Vật lý, Hóa học, Sinh học, Tiếng Anh...) và khả năng chuyên giao quy trình.

2. Sự không trùng lặp của đề tài nghiên cứu và tính trung thực trong trích dẫn

2.1. Đánh giá chung

Qua khảo sát nội dung luận án và đối chiếu với các công trình đã công bố trong và ngoài nước, có thể khẳng định đề tài của NCS không trùng lặp với các luận án tiến sĩ khác đã bảo vệ tại Việt Nam hoặc quốc tế. Thuật toán Winner-domain và Winner-cosin cũng như bộ dataset 2024/2025 do NCS tự xây dựng là tương đối đặc thù. Lời cam đoan được ghi rõ ở đầu luận án; danh mục tài liệu tham khảo có 141 tài liệu, bao gồm cả nguồn tiếng Anh và tiếng Việt.

Tính trung thực trong trích dẫn về cơ bản được bảo đảm.

2.2. Những điểm cần chỉnh sửa, bổ sung về trích dẫn và danh mục TLTK

– Danh mục TLTK còn nhiều lỗi định dạng: tên tạp chí chưa in nghiêng nhất quán; một số tài liệu thiếu DOI; thứ tự trình bày họ tên tác giả không đồng nhất (có chỗ là "Last, F.", có chỗ là "F. Last"); năm xuất bản đôi khi đặt ở vị trí không đồng nhất. NCS cần chuẩn hóa về một kiểu duy nhất (IEEE hoặc APA 7th).

– Có các cặp trích dẫn nghi trùng lặp hoặc lặp nội dung, cần rà soát: [3] và [21] (cùng liên quan đến Zhang et al. về EDM); [5] và [32] (nghi cùng công trình); [8] và [20] (cùng tác giả/năm); cặp [6]/[9] trong phần đề cập đến K-means. Nếu thực sự là hai công trình khác nhau, cần thêm phân biệt (tiêu đề khác, hội nghị/tạp chí khác, số trang khác) để người đọc dễ nhận diện.

– Thiếu các công trình nền tảng của one-class classification và anomaly detection - vốn là lý thuyết gần nhất với Winner-domain, cần đưa vào mục 1.5 và Chương 4.

– Thiếu các công trình quan trọng về XAI (Explainable AI) và giải thích mô hình trong giáo dục - rất liên quan khi ứng dụng ML để tuyển chọn học sinh.

– Thiếu công trình về fairness trong EDM, đặc biệt khi dữ liệu có biến nhạy cảm (giới tính, kinh tế gia đình, dân tộc...).

– Thiếu công trình về small-sample learning, imbalanced learning.

– Các phát biểu mang tính khẳng định nhưng thiếu dẫn nguồn peer-reviewed: (i) mối liên hệ giữa yếu tố di truyền và khả năng học lập trình (mục 3.2.3); (ii) ảnh hưởng của rượu bia, chơi game, mạng xã hội đến kết quả học tập (mục 3.4); (iii) mối quan hệ nhân-quả giữa các yếu tố phi học tập và thành tích tại kỳ thi quốc gia. Cần trích dẫn các nghiên cứu peer-reviewed hoặc có kiểm định thống kê trên cùng đối tượng (học sinh THPT chuyên Việt Nam).

– Một số thuật ngữ trong luận án đang được "đặt mới" bởi NCS (Winner-domain, Winner-cosin, reply cube, response matrix) chưa được khảo sát đầy đủ xem có trùng tên/trùng ý tưởng với thuật ngữ quốc tế nào không. Cần bổ sung một mục "Quy ước thuật ngữ" ở đầu luận án hoặc cuối Chương 1 để định danh rõ các khái niệm mới.

– Tỷ lệ tài liệu tham khảo từ 2023-2025 còn thấp - chưa phản ánh được tính thời sự của lĩnh vực ML/EDM vốn tiến hóa rất nhanh.

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu và giữa nội dung với ngành, mã số chuyên ngành

3.1. Về tên đề tài

Tên đề tài "Tiếp cận học máy để tuyển chọn và trực quan hóa để huấn luyện học sinh đội tuyển" phản ánh hai nội dung chính: (i) học máy cho bài toán tuyển chọn; (ii) trực quan hóa cho bài toán huấn luyện. Về cơ bản, tên đề tài phù hợp với nội dung.

3.2. Về sự phù hợp với ngành và mã số

Ngành Khoa học máy tính (mã số 9 48 01 01): luận án có các nội dung thuật toán (Winner-domain, Winner-cosin, decision tree, KNN-imputer), xử lý dữ liệu, phân tích độ phức tạp, đánh giá hiệu năng bằng metrics chuẩn (TPR, FPR, FNR, Accuracy), sử dụng công cụ Python/scikit-learn... Các nội dung này thuộc phạm vi Khoa học máy tính.

3.3. Những điểm cần làm rõ và cải thiện

– Chương 2 (Lý thuyết học khám phá) và Chương 5 (Mô hình học khám phá có hướng dẫn trực quan) có nội dung thiên về Khoa học giáo dục/Sư phạm học, nặng về lý thuyết Bruner, Dewey, Vygotsky... Cần nêu rõ hơn phần đóng góp "Khoa học máy tính" trong các chương này, ví dụ: mô hình toán học đánh giá hiệu năng bài giảng, thuật toán sinh bài giảng có trực quan hóa, hệ thống hỗ trợ quyết định cho giáo viên, công cụ phần mềm (software artifact) kèm theo. Nếu không, Chương 2 và Chương 5 có nguy cơ bị đánh giá lệch khỏi mã số ngành.

– Hai đóng góp chính (Chương 4 về ML tuyển chọn và Chương 5 về học khám phá trực quan) hiện còn rời rạc về mặt tích hợp hệ thống. Chưa có một chương/mục "end-to-end" trình bày: dữ liệu từ khảo sát → mô hình tuyển chọn → danh sách đội tuyển → thiết kế bài giảng theo mô hình học khám phá → đánh giá hiệu năng. Nếu không có mắt xích này, tên đề tài và mục tiêu tổng quát "tuyển chọn và huấn luyện" chưa được chứng minh về mặt hệ thống.

– Chưa có phần mô tả rõ phần mềm/công cụ mà NCS phát triển (giao diện, kiến trúc, API, repository mở) - điều này đặc biệt quan trọng với luận án ngành KHMT, giúp tăng tính lặp lại (reproducibility) và giá trị chuyên giao.

– Đề nghị NCS bổ sung ở đầu luận án một sơ đồ tổng quan trình bày rõ vai trò Khoa học máy tính ở từng đóng góp, để người đọc thấy rõ tính nhất quán của luận án với ngành và mã số.

4. Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

Luận án đã sử dụng hệ phương pháp đa dạng: phân tích - tổng hợp, mô hình hóa toán học, học máy có giám sát (supervised learning), cây quyết định, khảo sát có cấu trúc, thực nghiệm triển

khai. Công cụ Python, Google Forms, KNNImputer của scikit-learn là phù hợp với thông lệ quốc tế. Tuy nhiên, xét về chuẩn mực của một luận án tiến sĩ ngành Khoa học máy tính ở thời điểm 2026, phương pháp của luận án còn nhiều điểm cần được cải thiện và bổ sung để nâng cao độ tin cậy và tính hiện đại.

4.1. Về thiết kế dataset và lượng hóa hồi đáp định tính (Chương 3)

– Cỡ mẫu nhỏ so với số chiều: Dataset 2024 có $221 \text{ mẫu} \times 40 \text{ đặc trưng}$ (tỷ lệ mẫu/đặc trưng ≈ 5.5), Dataset 2025 có 497×40 (tỷ lệ ≈ 12.4). Theo thông lệ ML, tỷ lệ mẫu/đặc trưng tối thiểu nên là 10-20 và lý tưởng là ≥ 50 để tránh "curse of dimensionality" và overfitting. Cần (i) thực hiện feature selection có hệ thống (filter methods: mutual information, chi-square; wrapper methods: recursive feature elimination; embedded: Lasso); (ii) hoặc feature extraction bằng PCA/Autoencoder; (iii) hoặc dùng regularization mạnh (L1/L2/Elastic Net).

– Chất lượng thang đo khảo sát chưa được kiểm định: với 40 mục câu hỏi đa thành phần (học tập, phi học tập với nhiều tiểu thang đo), NCS chưa báo cáo Cronbach's alpha cho các thang đo con, chưa báo cáo Exploratory Factor Analysis (EFA) hoặc Confirmatory Factor Analysis (CFA) để chứng minh tính hội tụ/phân biệt (convergent/discriminant validity). Với một công cụ đo có ảnh hưởng đến kết quả tuyển chọn, bước kiểm định độ tin cậy thang đo là bắt buộc theo chuẩn nghiên cứu khoa học giáo dục.

– Lượng hóa hồi đáp định tính (mục 3.6 - 3.7): ma trận hồi đáp $F(i,j)$ và khối 3D hồi đáp $B(i,j,k)$ có cách xây dựng ad-hoc. Hằng số chuẩn hóa $\omega = 10$ (trang 67) được NCS "chọn" mà không có cơ sở thống kê/tối ưu. Cần hoặc (i) chứng minh $\omega = 10$ là tối ưu qua grid search trên tập validation, hoặc (ii) thay bằng phương pháp chuẩn hóa có cơ sở như min-max scaling, z-score, rank-based transformation, hoặc target encoding.

– Chưa so sánh với các phương pháp encoding dữ liệu khảo sát chuẩn mực: (i) Likert scale 5-7 điểm + Rasch/IRT model; (ii) one-hot encoding cho biến định danh; (iii) ordinal encoding cho biến thứ bậc; (iv) target/mean encoding có regularization; (v) entity embeddings học được từ neural network. Việc so sánh này sẽ chứng minh hoặc phủ định ưu thế của cách mã hóa mới.

– Xử lý dữ liệu thiếu chỉ trình bày ngắn việc dùng KNNImputer ($K=?$). Cần báo cáo: (i) tỷ lệ thiếu của từng biến; (ii) phân loại kiểu thiếu (MCAR, MAR, MNAR) bằng Little's MCAR test; (iii) so sánh giữa các cách xử lý (xóa hàng thiếu, mean/median imputation, multiple imputation, iterative imputer, KNN với nhiều K); (iv) phân tích độ nhạy (sensitivity analysis) để đánh giá ảnh hưởng của xử lý missing lên kết quả cuối cùng.

– Thiếu kiểm soát thiên lệch khảo sát: (i) self-selection bias (ai tự nguyện trả lời?); (ii) social desirability bias với câu hỏi nhạy cảm (rượu bia, chơi game); (iii) response-style bias (acquiescence, extreme responding). Cần thiết kế câu hỏi đối chứng, câu hỏi ngược, câu "bẫy" để phát hiện các thiên lệch này.

– Thông tin về tỷ lệ đáp viên, trường học tham gia, phân bố địa lý, thời gian thu thập còn mỏng. Cần bảng phân tích mô tả (descriptive statistics) đầy đủ trong Chương 3:

min/max/mean/std/skewness/kurtosis của mỗi biến, phân bố của biến mục tiêu (đạt giải/không đạt giải), kiểm định cân bằng giữa Dataset 2024 và 2025.

4.2. Về thuật toán Winner-domain và Winner-cosin (Chương 4)

– Định vị thuật toán chưa rõ ràng: Winner-domain được mô tả là "vùng chiến thắng" (cầu/nón trong không gian đặc trưng có trọng tâm, bán kính, và "cosin góc hướng"). Về bản chất, cấu trúc này rất gần với (i) one-class SVM với kernel RBF (tạo một "ranh giới" bao quanh lớp dương); (ii) Support Vector Data Description - SVDD (tìm hình cầu nhỏ nhất bao phủ lớp dương); (iii) centroid-based classifier với ngưỡng khoảng cách; (iv) Mahalanobis distance classifier; (v) Gaussian Mixture Model 1-component cho lớp dương. NCS cần thực hiện empirical comparison trực tiếp với các thuật toán này - đây là điều kiện tối thiểu để chứng minh tính mới và ưu thế của đề xuất.

– Thiếu chứng minh tính hội tụ và bounds lý thuyết: công thức độ phức tạp $O(P \cdot K \cdot (N+M) \cdot S)$ (trang 88) chỉ là độ phức tạp tính toán. Cần thêm phân tích về: (i) sample complexity (cần bao nhiêu mẫu để đạt xác suất ϵ -accurate); (ii) khả năng tổng quát hóa (generalization bound); (iii) kỳ vọng PAC-learnability; (iv) sensitivity to hyperparameters.

– Chiến lược đánh giá $P=4$ phân hoạch ngẫu nhiên là quá ít: theo thông lệ ML, cần k-fold cross-validation với $k=5$ hoặc $k=10$, hoặc stratified k-fold (vì dữ liệu mất cân bằng), hoặc repeated k-fold (ví dụ 10×5 -fold CV). Nested cross-validation nên được áp dụng khi có tuning hyperparameters (số Winner-domains Q , bán kính r , ngưỡng cosin θ). Việc chỉ dùng $P=4$ làm cho khoảng tin cậy của các chỉ số rất rộng, khó kết luận ý nghĩa thống kê.

– Tiêu chí chọn Winner-domain "tốt" là $FPR = 0$ trên tập huấn luyện. Điều này dẫn đến: (i) mô hình quá thiên về precision, có nguy cơ bỏ sót nhiều candidate tiềm năng (giảm recall); (ii) dễ overfit tập huấn luyện, đặc biệt khi dữ liệu nhỏ và nhiễu. Cần phân tích trade-off precision-recall, vẽ ROC curve, Precision-Recall curve, và báo cáo thêm các chỉ số phù hợp với dữ liệu mất cân bằng: F1-score, F2-score (ưu tiên recall), Matthews Correlation Coefficient (MCC), Balanced Accuracy, Area Under PR Curve (AUPRC).

– Baseline so sánh còn hạn chế: luận án chỉ so sánh với Random Forest (Bảng 4.32) - không đủ. Cần mở rộng baseline theo 4 nhóm: (a) Các mô hình truyền thống: Logistic Regression (với L1/L2 regularization), k-Nearest Neighbors (các giá trị k khác nhau), Gaussian Naive Bayes, Linear/RBF SVM; (b) Các mô hình cây/ensemble: XGBoost, LightGBM, CatBoost, Gradient Boosting, AdaBoost, Extra Trees; (c) Các mô hình chuyên cho small-sample/imbalanced: SMOTE/ADASYN + classifier, one-class SVM, Isolation Forest, LOF (Local Outlier Factor), SVDD, Elliptic Envelope; (d) Các mô hình neural nhẹ (phù hợp cỡ mẫu nhỏ): TabNet, MLP có regularization mạnh, FT-Transformer, SAINT.

– Thiếu ablation study có hệ thống: (i) chỉ dùng nhóm đặc trưng học tập (academic features) vs chỉ dùng nhóm phi học tập (non-academic) vs kết hợp - để định lượng đóng góp của mỗi nhóm; (ii) với và không có thành phần Winner-cosin - để đo đóng góp của phần hướng cosin; (iii) số Winner-domains "tốt" Q ($Q=1, 2, 4, 8, 16$) - để phân tích độ nhạy; (iv) ảnh hưởng của bán kính r và ngưỡng cosin θ .

– Thiếu phân tích tầm quan trọng đặc trưng (feature importance): với ứng dụng giáo dục, giáo viên cần hiểu vì sao mô hình chọn/không chọn một học sinh. Cần báo cáo: (i) SHAP values (global và local); (ii) permutation importance; (iii) partial dependence plots; (iv) với decision tree - visualize cây và rule extraction.

– Công thức độ phức tạp $O(P \cdot K \cdot (N+M) \cdot S)$ chỉ tính cho giai đoạn huấn luyện. Cần bổ sung (i) độ phức tạp inference (suy luận cho một mẫu mới); (ii) độ phức tạp không gian (memory); (iii) thực đo (wall-clock time) trên phần cứng cụ thể và so sánh với các baseline.

– Tính khả chuyển (transferability): thực nghiệm chỉ triển khai tại 01 trường THPT chuyên thuộc miền Tây Nam Bộ. Đây là điểm yếu lớn về external validity. Cần (i) thực nghiệm tại ít nhất 2-3 trường ở 2-3 vùng miền khác nhau; hoặc (ii) group k-fold cross-validation theo trường; hoặc (iii) leave-one-school-out CV. Đồng thời cần phân tích domain shift: đặc trưng học sinh miền Tây có khác với miền Bắc/Trung không?

– Thiếu phân tích "ca thất bại" (error analysis): những học sinh nào được mô hình dự đoán sai? Các đặc trưng nào làm mô hình dễ nhầm lẫn? Điều này giúp cải thiện thuật toán lập.

4.3. Về mô hình học khám phá có hướng dẫn trực quan (Chương 5)

– Mô hình toán học đánh giá hiệu năng - các công thức (5.1) đến (5.7) có nhiều vấn đề về cơ sở lý thuyết: (i) Công thức $B \cdot T \propto O$ thiếu chứng minh. Vì sao "sự chú ý B nhân với thời gian T tỷ lệ thuận với đầu ra học tập O"? Có nghiên cứu cognitive science nào kiểm chứng không? Đề nghị đối chiếu với các mô hình kinh điển như Cognitive Load Theory (Sweller), Model of School Learning (Carroll), ACT-R (Anderson). (ii) Công thức $E = O/(B \cdot T)$ giả định quan hệ thuần tuyến tính giữa $B \cdot T$ và O - trên thực tế quan hệ này là phi tuyến (sigmoid, có plateau). (iii) Công thức mở rộng dạng "... $\propto K \cdot S \cdot A \cdot M \cdot V \cdot P \cdot O/(C \cdot T)$ " có 7-8 biến nhân vô hướng với đơn vị/thang đo rất khác nhau - không có ý nghĩa vật lý/sự phạm. Đề nghị NCS cân nhắc thay bằng mô hình hồi quy đa biến, mô hình phương trình cấu trúc (SEM), hoặc mô hình hiệu ứng hỗn hợp (mixed effects model).

– Cách chuẩn hóa khoảng giá trị: "khoảng [7/20, 70/2]" (trang 108) và sau đó quy về [0, 100%] thông qua công thức (5.7) - các lựa chọn này ad-hoc, không có cơ sở thống kê. Đề nghị hoặc (i) chuẩn hóa theo min-max trên dữ liệu thực; (ii) hoặc chuẩn hóa theo phần trăm (percentile ranking); (iii) hoặc chuyển sang mô hình xác suất logistic.

– Thiếu thực nghiệm đầy đủ kiểm chứng mô hình hiệu năng: cần thiết kế quasi-experimental design với nhóm thực nghiệm (áp dụng mô hình học khám phá có hướng dẫn trực quan) và nhóm đối chứng (học tập truyền thống), cỡ mẫu đủ ($N \geq 30/\text{nhóm}$), đo đầu ra chuẩn hóa (điểm kiểm tra chuẩn, thời gian hoàn thành, retention sau X tuần), kiểm định t-test/ANOVA với effect size (Cohen's d). Hiện tại mục 5.x chỉ mô tả một vài trường hợp, chưa đủ sức thuyết phục khoa học.

– Thiếu so sánh với các mô hình học tập tích cực khác: Problem-Based Learning (PBL), Inquiry-Based Learning (IBL), Project-Based Learning (PjBL), Flipped Classroom, Gamified Learning, Peer Instruction. Không so sánh, khó khẳng định "học khám phá có hướng dẫn trực quan" là lựa chọn tối ưu.

– Thiếu thảo luận về công nghệ trực quan hóa cụ thể (visualization technology): NCS dùng công cụ gì (Algoviz, Python Turtle, p5.js, Jupyter Notebook, phần mềm tự phát triển)? Các yếu tố UI/UX nào được cân nhắc? Có đo eye-tracking, attention, engagement không? Phần này rất cần thiết để chương 5 có tính khoa học máy tính rõ nét.

– Chưa có đánh giá hiệu năng hệ thống phần mềm: thời gian phản hồi, khả năng mở rộng (scalability) số học sinh/số bài giảng đồng thời, chi phí vận hành. Đây là những chỉ số quan trọng đối với một hệ thống dạy học có trực quan hóa.

4.4. Các vấn đề phương pháp luận xuyên suốt

– Fairness và Bias: luận án dùng nhiều biến phi học tập (giới tính, điều kiện gia đình, thói quen sinh hoạt, ...). Những biến này dễ làm mô hình thiên lệch theo nhóm xã hội. NCS cần thực hiện fairness audit: tính disparate impact, equal opportunity difference, statistical parity khi mô hình được áp dụng cho các nhóm học sinh khác nhau (nam/nữ, thành thị/nông thôn, thu nhập cao/thấp). Nếu phát hiện chênh lệch, cần kỹ thuật giảm thiên lệch (reweighing, fairness-constrained optimization, adversarial debiasing).

– Explainability/Interpretability: với ứng dụng ra quyết định có ảnh hưởng lớn đến cá nhân (tuyển chọn đội tuyển), mô hình "hộp đen" là không chấp nhận được theo xu thế Responsible AI hiện nay. NCS cần bổ sung: (i) giải thích toàn cục (global explanation) - những đặc trưng nào quan trọng nhất; (ii) giải thích cục bộ (local explanation) - vì sao một học sinh cụ thể được chọn/loại; (iii) counterfactual explanations - nếu học sinh A cải thiện đặc trưng X bao nhiêu thì có khả năng được chọn; (iv) rule extraction từ cây quyết định thành quy tắc tường minh cho giáo viên.

5. Kết quả nghiên cứu mới của luận án

Luận án tuyên bố 03 đóng góp mới chính:

– Đóng góp 1: Phương pháp thiết lập dataset năng lực thí sinh (Dataset 2024 với 221 mẫu và Dataset 2025 với 497 mẫu) cùng với cách tiếp cận ma trận hồi đáp (response matrix) và khối 3D hồi đáp (reply cube) để lượng hóa các hồi đáp định tính.

– Đóng góp 2: Thuật toán Winner-domain và Winner-cosin kết hợp cây quyết định (Decision Tree) để tuyển chọn thí sinh đội tuyển.

– Đóng góp 3: Mô hình học khám phá có hướng dẫn trực quan cùng mô hình toán học đánh giá hiệu năng bài giảng.

5.1. Đánh giá về tính mới của từng đóng góp

Đóng góp 1 (Dataset và phương pháp lượng hóa hồi đáp):

– Việc xây dựng dataset dành riêng cho bài toán tuyển chọn HSG môn Tin học là có giá trị ứng dụng. Tuy nhiên, về mặt phương pháp, "ma trận hồi đáp" và "khối 3D hồi đáp" cần được đối

chiều với các cấu trúc dữ liệu đã có trong phân tích khảo sát và ML: tensor analysis / multi-way data analysis (Kolda & Bader 2009), item-response theory (IRT) matrix, multi-label multi-instance learning. Nếu không đặt vào khung đã có, khó khẳng định tính mới về mặt phương pháp.

- Dataset nên được công bố công khai (với anonymization) trên các repository như Harvard Dataverse, Kaggle, Zenodo để cộng đồng có thể kiểm chứng và phát triển - đây là cách tốt nhất để khẳng định đóng góp dataset.

Đóng góp 2 (Thuật toán Winner-domain và Winner-cosin):

- Winner-domain về bản chất là một centroid-based/one-class classifier: vùng không gian đặc trưng được xác định bởi trọng tâm và bán kính. Tính mới cần được phát biểu rõ bằng so sánh định lượng với one-class SVM, SVDD, Isolation Forest. Nếu kết quả tốt hơn có ý nghĩa thống kê thì đây là đóng góp; nếu không thì nên định vị thành "ứng dụng/thích nghi" chứ không phải thuật toán mới.

- Winner-cosin sử dụng cosine similarity - là kỹ thuật quen thuộc trong Information Retrieval và Text Mining từ nhiều thập kỷ. Cần phát biểu rõ: đây là "ứng dụng mới/kết hợp mới" (novel application/combination) của cosine similarity trong bối cảnh tuyển chọn học sinh giỏi, không phải thuật toán mới hoàn toàn.

- Kết hợp Winner-domain + Winner-cosin + Decision Tree tạo thành một pipeline. Đóng góp thật sự có thể nằm ở pipeline này - nếu NCS phát biểu như vậy thì dễ bảo vệ hơn, và cần chứng minh từng thành phần đóng góp giá trị gì (bằng ablation study).

Đóng góp 3 (Mô hình học khám phá có hướng dẫn trực quan và mô hình hiệu năng):

- Mô hình học khám phá có hướng dẫn (Guided Discovery Learning) không mới - đã có từ Bruner (1961) và được nhiều tác giả phát triển. Phần "có hướng dẫn trực quan" cần được định nghĩa chặt chẽ và phân biệt với các mô hình lân cận: Cognitive Tutor, Worked Examples, Scaffolded Learning, Visual Programming Environments.

- Mô hình toán học đánh giá hiệu năng $E = O/(B \cdot T)$ hiện là một khung khái niệm (conceptual framework). Để trở thành kết quả mới có giá trị khoa học, cần (i) cơ sở lý thuyết chặt (đạo hàm từ giả thiết nào?); (ii) kiểm chứng thực nghiệm với cỡ mẫu đủ; (iii) so sánh với các chỉ số hiệu năng học tập hiện có (learning gain, Hake's gain factor, effect size).

5.2. Đánh giá tổng hợp về tính mới

Các đóng góp của luận án có giá trị ứng dụng rõ nét và phù hợp với nhu cầu thực tiễn giáo dục Việt Nam. Tuy nhiên, để đạt ngưỡng "kết quả mới" của luận án tiến sĩ ngành KHMT, cần:

- Phát biểu lại tính mới theo hướng khiêm tốn và chính xác: đóng góp nào là thuật toán mới, đóng góp nào là ứng dụng mới, đóng góp nào là hệ thống mới, đóng góp nào là dataset mới.

- Chứng minh tính mới bằng so sánh trực tiếp với baseline mạnh nhất hiện có (state-of-the-art), không chỉ Random Forest.
- Công khai artifact (dataset, source code, bài giảng, công cụ trực quan hóa) để tính mới được cộng đồng xác nhận.

6. Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

6.1. Ưu điểm

- Đề tài xuất phát từ một nhu cầu thực tiễn cụ thể, có tính ứng dụng cao và có khả năng chuyển giao ra các trường THPT chuyên khác.
- Khối lượng công việc đáng kể: thu thập 2 datasets ở nhiều tỉnh thành, thiết kế bộ 40 câu hỏi khảo sát, xây dựng thuật toán, cài đặt, thực nghiệm tại trường 02 năm liên tiếp.
- Cấu trúc luận án đầy đủ theo chuẩn: Mở đầu, 5 chương, Kết luận, Danh mục công trình, TLTK, Phụ lục.
- Kết quả thực nghiệm tại trường thực nghiệm cho thấy cải thiện so với phương pháp tuyển chọn truyền thống - có bằng chứng thực tiễn.
- Có 04 công trình quốc tế (01 SCIE Q2 + 03 hội thảo Scopus) bao phủ 3 nội dung chính của luận án.
- Kết hợp được hai mảng (ML và Khoa học giáo dục) theo hướng liên ngành - phù hợp với xu thế quốc tế.
- Luận án trình bày chi tiết các bảng biểu thực nghiệm, có số liệu cụ thể cho 2 kỳ thi - thể hiện tính minh bạch.
- Tác giả đã cho thấy sự cân trọng đáng quý trong việc xây dựng khung lý thuyết (Chương 2) trước khi đề xuất thuật toán.

6.2. Nhược điểm về nội dung

- Chương 1 (Tổng quan): độ sâu phân tích tài liệu chưa đạt chuẩn systematic literature review (PRISMA). Thiểu bảng tổng hợp so sánh các công trình tiêu biểu theo các tiêu chí: tác giả/năm, đối tượng, dataset, phương pháp, kết quả, hạn chế. Cần thêm bảng matrix này để người đọc thấy rõ landscape của lĩnh vực.
- Chương 2 và Chương 5: nội dung về Guided Discovery Learning bị lặp ở cả hai chương. Cần tái cấu trúc: Chương 2 chỉ trình bày nền tảng lý thuyết chung (Bruner, Dewey...), Chương 5 dành riêng cho mô hình đề xuất.

- Chương 3: thiết kế dataset trình bày nặng về định nghĩa ký hiệu toán học, làm người đọc khó theo. Có thể đưa các định nghĩa vào phụ lục hoặc một mục "Danh mục ký hiệu".
- Chương 4: phần thực nghiệm chỉ so sánh với Random Forest (Bảng 4.32). Như đã nêu ở Mục 4, đây là hạn chế lớn cần khắc phục bằng cách bổ sung các baseline chuẩn.
- Chương 5: thiếu thực nghiệm định lượng kiểm chứng mô hình đánh giá hiệu năng. Mô hình toán học có cảm giác "tự phát" thay vì được derived từ giả thiết rõ ràng.
- Thiếu chương/mục Thảo luận (Discussion) tổng hợp: so sánh kết quả với kỳ vọng, thảo luận ý nghĩa, hạn chế (limitations), đe dọa đến tính hợp lệ (threats to validity: construct validity, internal validity, external validity, reliability).
- Phần Kết luận chỉ 3 trang - quá mỏng so với nội dung luận án. Cần mở rộng: (i) Tổng kết đóng góp theo 3 đóng góp đã nêu; (ii) Hạn chế của nghiên cứu; (iii) Hướng nghiên cứu tiếp theo cụ thể (ít nhất 5 hướng); (iv) Bài học về phương pháp.
- Hai đóng góp chính (ML tuyển chọn ở Chương 4 và học khám phá trực quan ở Chương 5) thiếu tích hợp end-to-end. Người đọc cảm giác đây là 2 nửa luận án rời rạc được gán lại. Nên bổ sung một sơ đồ tích hợp và một trường hợp đầu-cuối (end-to-end case) để thể hiện tính nhất quán.

6.3. Nhược điểm về kết cấu

- Mục 1.5 "Những tiếp cận của khoa học máy tính được liên quan" giới thiệu SVM, Decision Tree, Clustering, K-means. Tuy nhiên trong luận án không thực sự sử dụng Clustering/K-means ở Chương 4 hay 5. Cần rà soát để mục tổng quan chỉ bao phủ các kỹ thuật thật sự được dùng và thảo luận.
- Một số mục được đánh số chưa nhất quán trong mục lục: ví dụ mục 3.6 và các tiểu mục 3.6.1, 3.6.2; các Bảng trong Chương 4 đôi chỗ có khoảng trắng giữa số chính và số phụ ("Bảng 4. 10" vs "Bảng 4.10").
- Các Hình trong luận án được chèn chưa đồng nhất về vị trí, caption, và chất lượng render.
- Trình tự logic của Chương 3 (dataset) và Chương 4 (thuật toán) có thể cần nhắc đảo: giới thiệu tổng quát bài toán và thuật toán đề xuất trước, sau đó mô tả dataset và thực nghiệm - giúp người đọc dễ nắm được điểm chính trước khi đi vào chi tiết.
- Mục "Danh mục các công trình đã công bố của tác giả" chưa phân loại theo hạng công trình (Q1/Q2/Q3, SCI/SCIE/Scopus, rank hội nghị theo CORE) - điều này làm khó cho hội đồng đánh giá giá trị công bố.

6.4. Nhược điểm về hình thức

- Lỗi chính tả và gõ dấu: "khuyến khích" (thiếu dấu, trang 65), "trực quan" (trang 108), một vài chỗ dùng "nhâm"/"nhâm" không nhất quán, dấu câu có chỗ đặt trước khoảng trắng, có chỗ đặt sau - cần rà soát tổng thể.
- Thuật ngữ Anh-Việt không thống nhất: có chỗ dùng "questionnaire", chỗ "bảng khảo sát", chỗ "bộ câu hỏi"; "học khám phá có hướng dẫn trực quan" đôi khi viết là "học khám phá có hướng dẫn bằng kỹ thuật trực quan"; "Winner-cosin" và "Winner-cosine" đều xuất hiện. NCS cần chọn một thuật ngữ Việt và một thuật ngữ Anh tương ứng, dùng xuyên suốt, và có bảng tra thuật ngữ đầu luận án.
- Nhiều công thức toán học trình bày chưa chuẩn: ký hiệu chồng lên nhau, subscript/superscript khó đọc, thiếu khoảng trắng giữa các ký tự (đặc biệt ở các công thức mục 3.6, 4.3, 5.x). Nên gõ lại bằng LaTeX-style với MathType hoặc Word Equation Editor chuẩn.
- Nhiều bảng có font size quá nhỏ, khó đọc khi in (Bảng 4.3, 4.4, 4.32 và các bảng phụ lục). Cần tăng font lên ít nhất 10pt, nếu bảng quá dài thì xoay landscape hoặc tách bảng.
- Một số hình vẽ (Hình 3.1, 4.1, 5.3...) có mật độ chữ dày, chất lượng render thấp (bị mờ). Cần vẽ lại bằng TikZ/PGFPlots/draw.io ở độ phân giải cao.
- Danh mục bảng/hình đánh số không đồng nhất (Bảng 4.10 / Bảng 4. 10).
- Danh mục tài liệu tham khảo có trích dẫn trùng lặp hoặc gần trùng (như đã nêu: [3]/[21], [5]/[32], [8]/[20]).
- Bảng tóm tắt tiếng Anh (summary/abstract) cần rà soát ngữ pháp và kiểm tra bởi native speaker để đảm bảo độ chuyên nghiệp.

6.5. Khuyến nghị bổ sung/chỉnh sửa trọng tâm trước khi bảo vệ cấp Học viện

Nhóm khuyến nghị quan trọng nhất (bắt buộc phải giải quyết trước/trong buổi bảo vệ):

- (R1) Bổ sung so sánh thực nghiệm với tối thiểu 5-7 baseline (XGBoost, LightGBM, Logistic Regression, One-class SVM, Isolation Forest, MLP, SVDD) trên cùng dataset với cùng metrics.
- (R2) Thay P=4 phân hoạch bằng k-fold CV (k=5 hoặc 10), báo cáo mean $\pm$ std và 95% CI cho tất cả metrics.
- (R3) Thêm kiểm định ý nghĩa thống kê (McNemar hoặc Wilcoxon) khi so sánh các mô hình.
- (R4) Bổ sung ablation study: (i) academic-only vs non-academic-only vs combined features; (ii) có/không có Winner-cosin.
- (R5) Bổ sung phân tích feature importance (SHAP hoặc permutation importance) và giải thích cho top-10 đặc trưng quan trọng nhất.

- (R6) Bổ sung kiểm định độ tin cậy thang đo (Cronbach's alpha, factor analysis) cho 40 mục câu hỏi.
- (R7) Bổ sung thảo luận về fairness, bias, đạo đức AI và bảo vệ dữ liệu cá nhân học sinh.
- (R8) Bổ sung thực nghiệm kiểm chứng mô hình hiệu năng bài giảng ở Chương 5 bằng quasi-experimental design.
- (R9) Bổ sung mục "Hạn chế" (Limitations) và "Hướng nghiên cứu tiếp theo" chi tiết trong phần Kết luận.
- (R10) Rà soát toàn bộ hình thức: sửa lỗi chính tả, chuẩn hóa công thức, thuật ngữ, danh mục TLTK, đánh số bảng/hình.

Nhóm khuyến nghị nâng cao (sau bảo vệ, để hoàn thiện):

- (A1) Công khai dataset (với anonymization) và source code trên GitHub/Zenodo.
- (A2) Phát triển thêm bài báo Q1 (Computers & Education, IEEE TLT, ReCALL, British Journal of Educational Technology).
- (A3) Mở rộng thực nghiệm ra 3-5 trường ở các vùng miền khác nhau.
- (A4) Tích hợp các kỹ thuật hiện đại: Large Language Models cho tạo bài giảng trực quan tự động, AutoML cho tuning mô hình, federated learning để bảo vệ dữ liệu học sinh.

7. Nội dung luận án đã được công bố và giá trị khoa học của các công trình

7.1. Danh mục công trình đã công bố

Luận án đã công bố 04 công trình khoa học quốc tế:

- [CT1] ICCASA 2022, Springer LNICST vol. 475, 2023 (Scopus). Nội dung liên quan đến tuyển chọn học sinh tham gia kỳ thi cấp tỉnh/quốc gia - phục vụ Chương 3, 4 của luận án.
- [CT2] Mobile Networks and Applications, Springer, vol. 29(1), 306-313, 2024 (SCIE, Q2 theo JCR). Nội dung về thuật toán Winner-domain - phục vụ Chương 4.
- [CT3] ICCASA 2023, Springer LNICST vol. 579, 2024 (Scopus). Nội dung về áp dụng Guided Discovery Learning - phục vụ Chương 2 và 5.
- [CT4] EIER 2025, E3S Web of Conferences vol. 626, 2025 (Scopus). Nội dung về hiệu năng hệ thống học khám phá có hướng dẫn trực quan - phục vụ Chương 5.

7.2. Đánh giá giá trị khoa học

Đánh giá theo các tiêu chí: độ phủ của nội dung luận án, uy tín của nơi công bố, vai trò tác giả, và trích dẫn:

– Độ phủ: 04 công trình bao phủ được 3 đóng góp chính của luận án - đạt yêu cầu theo Quy chế đào tạo tiến sĩ hiện hành (Thông tư 18/2021/TT-BGDĐT và các quy định của cơ sở đào tạo).

– Uy tín nơi công bố:

- [CT2] đăng trên Mobile Networks and Applications (Springer, SCIE, Q2 theo JCR) - là điểm cộng đáng kể, thể hiện NCS có năng lực công bố quốc tế ở mức trung bình-khá.
- [CT1], [CT3], [CT4] là công trình hội nghị quốc tế có trong Scopus. Tuy nhiên ICCASA và EIER hiện không nằm trong danh sách các hội nghị rank A/A* theo CORE (Computing Research and Education), và có chỉ số ảnh hưởng (impact) chưa cao.

– Vai trò tác giả: NCS là tác giả chính ở các công trình (first author) - phù hợp yêu cầu.

– Trích dẫn: do các công trình mới được công bố, số trích dẫn còn thấp - điều này là bình thường nhưng nên theo dõi sau bảo vệ.

– Đáp ứng yêu cầu tối thiểu: về mặt số lượng và chất lượng, 04 công trình đáp ứng yêu cầu đầu ra của một luận án tiến sĩ. Tuy nhiên so với chuẩn chung ngày nay của các luận án KHMT tại Việt Nam (thường 2-3 SCI/SCIE hoặc 1 SCI + 2-3 Scopus ở hội nghị rank A), thì công bố của NCS ở mức chấp nhận được chứ chưa thật sự ấn tượng.

7.3. Khuyến nghị về công bố (để phát triển sau bảo vệ)

– Phát triển thêm kết quả Chương 4 (bổ sung các baseline, ablation, statistical tests) để gửi tạp chí Q1: Expert Systems with Applications, Knowledge-Based Systems, IEEE Transactions on Learning Technologies, Computers & Education.

– Phát triển Chương 5 (với thực nghiệm định lượng) để gửi các hội nghị rank A/A* trong lĩnh vực AI/EDM: AIED (International Conference on AI in Education), EDM (Educational Data Mining), LAK (Learning Analytics and Knowledge), L@S (Learning at Scale).

– Xem xét công bố dataset như một "Data Paper" ở các tạp chí Data in Brief, Scientific Data (Nature group), Journal of Open Educational Resources - cách làm giá trị khoa học của dataset được cộng đồng ghi nhận.

8. Kết luận chung

8.1. Tóm lược đánh giá

Luận án của NCS Huỳnh Thị Ngọc Cẩm đã thực hiện một khối lượng công việc nghiêm túc: thiết kế bộ câu hỏi khảo sát 40 mục, thu thập dữ liệu tại nhiều trường THPT chuyên (Dataset 2024: 221 mẫu; Dataset 2025: 497 mẫu), đề xuất thuật toán Winner-domain và Winner-cosin kết hợp cây quyết định, đề xuất mô hình học khám phá có hướng dẫn trực quan cùng mô hình toán học đánh giá hiệu năng bài giảng, và triển khai thực nghiệm tại 01 trường THPT chuyên trong 02 kỳ

thi HSG quốc gia liên tiếp. Luận án có giá trị thực tiễn rõ nét, có 04 công trình quốc tế (trong đó 01 SCIE Q2), cấu trúc tương đối đầy đủ và trình bày bằng tiếng Việt khoa học.

8.2. Các hạn chế cần được NCS làm rõ/bổ sung

Bên cạnh các giá trị nêu trên, luận án còn nhiều hạn chế đáng kể cần được NCS làm rõ, chỉnh sửa và/hoặc bổ sung, cụ thể như đã trình bày chi tiết ở các Mục 1-6. Các hạn chế chính được tóm lược như sau:

- (1) Về thiết kế dataset và công cụ khảo sát: cỡ mẫu nhỏ so với số chiều đặc trưng (curse of dimensionality); chưa kiểm định độ tin cậy thang đo (Cronbach's alpha, factor analysis); hằng số chuẩn hóa ad-hoc; chưa kiểm soát các thiên lệch khảo sát.
- (2) Về đánh giá thuật toán: đánh giá chỉ với P=4 phân hoạch thay vì k-fold cross-validation; baseline so sánh hạn chế (chỉ Random Forest); thiếu kiểm định ý nghĩa thống kê (McNemar, paired t-test); thiếu khoảng tin cậy; thiếu ROC/PR curves; thiếu ablation study và feature importance.
- (3) Về định vị thuật toán: Winner-domain cần được so sánh trực tiếp với các phương pháp kinh điển gần nhất (one-class SVM, SVDD, Mahalanobis classifier, GMM) để khẳng định tính mới; Winner-cosin nên được định vị là ứng dụng mới chứ không phải thuật toán mới.
- (4) Về Chương 5: mô hình toán học đánh giá hiệu năng $E = O/(B \cdot T)$ mang tính hiện tượng luận, thiếu cơ sở lý thuyết chặt chẽ và thực nghiệm kiểm chứng đầy đủ; chưa so sánh với các mô hình học tập tích cực khác.
- (5) Về tính hiện đại: thiếu thảo luận về fairness, bias, explainability/XAI, đạo đức AI, bảo vệ dữ liệu cá nhân học sinh; thiếu tham chiếu các công trình 2023-2025 về LLMs, responsible AI trong giáo dục.
- (6) Về tích hợp hệ thống: hai đóng góp chính (ML tuyển chọn ở Chương 4 và học khám phá trực quan ở Chương 5) thiếu mắt xích tích hợp end-to-end.
- (7) Về hình thức: lỗi chính tả rải rác, thuật ngữ chưa nhất quán, công thức toán học và hình vẽ chưa đạt chất lượng in ấn, danh mục TLTK có trích dẫn trùng lặp.
- (8) Về tính khả chuyển: thực nghiệm chỉ tại 01 trường THPT chuyên - cần đánh giá khả năng tổng quát hóa sang các trường khác vùng miền.

8.3. Về bản tóm tắt luận án

Bản tóm tắt luận án về cơ bản phản ánh trung thực nội dung của luận án, bám sát cấu trúc 5 chương và các đóng góp chính. Tuy nhiên, tóm tắt cũng kế thừa các hạn chế đã nêu (đặc biệt về cách phát biểu tính mới và các kết quả định lượng thiếu khoảng tin cậy). Đề nghị NCS điều chỉnh lại để phản ánh chính xác hơn các chỉ số và có phần nêu rõ hạn chế của nghiên cứu.

8.4. Kết luận và kiến nghị

Đề nghị NCS nghiêm túc tiếp thu và chỉnh sửa theo các góp ý ở trên để luận án đáp ứng được các yêu cầu cơ bản của một luận án tiến sĩ ngành Khoa học máy tính. Đồng ý đề luận án được đưa ra bảo vệ chính thức cấp Học viện nhằm nhận học vị Tiến sĩ Khoa học máy tính.

TpHCM, ngày....tháng....năm 2026

Người nhận xét



PGS.TS. Lê Kim Hùng

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển.**

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn: PGS. TS. Trần Vĩnh Phước

Người nhận xét/ Người phản biện: PGS. TS. Trần Trung Duy

Cơ quan công tác: Học viện Công nghệ Bưu chính Viễn thông, cơ sở tại TP.
Hồ Chí Minh

Nội dung nhận xét:

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Trong bối cảnh đổi mới giáo dục và yêu cầu nâng cao chất lượng đào tạo học sinh giỏi, việc tuyển chọn và huấn luyện học sinh đội tuyển hiện nay vẫn chủ yếu dựa trên kinh nghiệm chủ quan của người huấn luyện, thiếu các công cụ phân tích dữ liệu một cách hệ thống. Điều này dẫn đến nguy cơ bỏ sót nhân tài hoặc chưa khai thác tối đa tiềm năng của học sinh. Sự phát triển mạnh mẽ của Học máy cho phép khai thác dữ liệu học tập, kết quả kiểm tra, hành vi học tập... nhằm phát hiện các mẫu tiềm ẩn và dự đoán năng lực học sinh một cách khách quan. Do đó, việc nghiên cứu phương pháp tiếp cận dựa trên học máy kết hợp trực quan hóa để hỗ trợ tuyển chọn và huấn luyện học sinh đội tuyển là rất cần thiết, góp phần nâng cao hiệu quả và tính khoa học trong công tác này.

Gần đây, khai phá dữ liệu giáo dục thu hút sự quan tâm của các nhà nghiên cứu phát triển các thuật toán học máy theo nhiều hướng mục tiêu như phân tích và dự đoán năng lực học sinh, nâng cao chất lượng giáo dục, nhằm tìm ra những ý nghĩa

và quan hệ tiềm ẩn trong dữ liệu giáo dục. Mục tiêu của bài toán khai phá dữ liệu giáo dục là cải thiện kết quả học tập của từng người học và đánh giá kết quả giảng dạy để làm giảm chi phí giáo dục.

Luận án này xây dựng quy trình tuyển chọn đội viên đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông và huấn luyện đội tuyển để tham dự kỳ thi học sinh giỏi cấp Quốc gia với kỳ vọng đạt được nhiều giải thưởng. Các kết quả dự đoán được thực hiện dựa trên tiếp cận học máy và trực quan hóa dữ liệu trong khoa học máy tính.

Kết luận: Luận án có tính cần thiết, tính thời sự, ý nghĩa khoa học và thực tiễn.

2, Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Theo sự hiểu biết tốt nhất của người phản biện, các kết quả đạt được của Luận án không trùng lặp với các công bố, đề tài, công trình đã công bố trong và ngoài nước trước đây. Hơn nữa, các kết quả trong Luận án đã được công bố trên các tạp chí, hội nghị quốc tế uy tín, và đã được phản biện bởi các chuyên gia trong ngành.

Về danh mục tài liệu tham khảo: Luận án trích dẫn 127 tài liệu tham khảo, các tài liệu tham khảo được trích dẫn đầy đủ và hợp lý trong quyền Luận án; danh mục tài liệu tham khảo tương đối đầy đủ và cập nhật. Tuy nhiên, Nghiên cứu sinh (NCS) cần cập nhật thêm các tài liệu tham khảo mới (công bố trong 03 năm gần nhất) để thay thế cho các tài liệu tham khảo đã cũ.

NCS cần kiểm tra lại thông tin, định dạng của một số tài liệu tham khảo sau: tài liệu số [29], [66], [96]. NCS cũng cần rà soát để thống nhất cách viết tên tạp chí, tên hội nghị.

Trong Chương 1, mặc dù NCS đã khảo sát các công trình liên quan đến khai phá dữ liệu giáo dục, tuy nhiên NCS chưa nêu bật được những điểm mới, điểm khác biệt của Luận án khi so sánh với các công trình này. Do đó, NCS nên thêm phần kết luận chương hoặc bổ sung phần so sánh nhằm nêu bật những điểm mới, ưu điểm/đóng góp chính của Luận án khi so sánh với các nghiên cứu trước đây.



3, Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số Chủ đề của Luận án liên quan đến Học máy để tuyển chọn và trực quan hóa để huấn luyện học sinh đội tuyển. Chủ đề của Luận án hoàn toàn phù hợp với ngành khoa học máy tính, mã số 9 48 01 01 (Trình độ đào tạo Tiến Sĩ).

Nội dung của Luận án nghiên cứu về lý thuyết học khám phá, thiết kế dữ liệu năng lực thí sinh tham gia các kỳ thi học sinh giỏi, từ đó nghiên cứu phương pháp tuyển chọn đội viên đội tuyển học sinh giỏi, và cuối cùng trình bày tiếp cận học khám phá có hướng dẫn trực quan để huấn luyện đội tuyển. Do đó, nội dung của Luận án thống nhất với tên của Luận án.

4, Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu Các phương pháp nghiên cứu được sử dụng trong Luận án, bao gồm:

Phương pháp phân tích tổng hợp, phương pháp mô hình hóa toán học, phương pháp học máy, phương pháp cây quyết định, phương pháp toán học, phương pháp lập trình, phương pháp khảo sát, và phương pháp thực nghiệm. Các phương pháp sử dụng trong Luận án là phù hợp.

Về dữ liệu thu thập: Luận án thu thập dữ liệu thông qua Google Forms với sự hỗ trợ của các huấn luyện viên đội tuyển học sinh giỏi cấp Quốc gia môn Tin học các khóa học từ 2020 đến 2024. Luận án đã thành lập dataset năng lực thí sinh bằng cách phân tích các đặc trưng năng lực học tập thành 18 câu câu hỏi năng lực học tập, và phân tích tâm lý giáo dục, sự phạm thành 22 câu hỏi năng lực phi học tập. Sau khâu tiền xử lý, Luận án sử dụng hồi đáp định tính của các Winners và Non-Winners để thành lập dataset. Các tiếp cận dataset linh hoạt cung cấp quy trình thành lập tập dữ liệu năng lực của các thí sinh theo từng mục khác nhau.

Kết luận: Phương pháp nghiên cứu trong Luận án phù hợp, có tính khoa học. Dữ liệu có độ tin cậy.

Câu hỏi 1: Số lượng Dataset trong Luận án đã đủ lớn cho bài toán huấn luyện và dự đoán kết quả. Tại sao NCS chỉ thu thập dữ liệu của 01 trường. Liệu dữ liệu của 01 trường có thể đưa ra kết quả dự đoán phù hợp cho các trường khác.

5, Kết quả nghiên cứu mới của tác giả

Luận án đã thiết lập được tập dữ liệu năng lực thí sinh đã tham gia các kỳ thi học sinh giỏi cấp quốc gia môn tin học, gồm những thí sinh thắng giải và không thắng giải.

Luận án đã xây dựng thuật toán và quy trình tuyển chọn học sinh vào đội tuyển học sinh giỏi môn Tin học của một trường trung học phổ thông để tham dự kỳ thi học sinh giỏi cấp Quốc gia môn tin học.

Luận án xây dựng mô hình học khám phá có hướng dẫn trực quan áp dụng cho các cơ sở giáo dục. Mô hình được thực nghiệm trong việc huấn luyện đội tuyển môn Tin học.

Luận án tiếp cận lý thuyết sơ đồ, lý thuyết tải nhận biết của con người để xây dựng quy trình và phương pháp biên soạn bài giảng trực quan áp dụng cho mô hình học khám phá có hướng dẫn trực quan

Luận án đã xây dựng mô hình toán học để đánh giá hiệu năng của từng bài giảng trực quan trong mô hình học khám phá có hướng dẫn trực quan.

Câu hỏi 2: Thuật toán và quy trình tuyển chọn học sinh vào đội tuyển được đề xuất có ưu điểm gì so với phương pháp truyền thống, và mức độ cải thiện độ chính xác trong việc dự đoán học sinh đạt giải cần được nhấn mạnh trong Luận án

Câu hỏi 3: Việc áp dụng Lý thuyết sơ đồ và Lý thuyết tải nhận thức trong thiết kế bài giảng trực quan được cụ thể hóa như thế nào để phù hợp với đặc thù môn Tin học? Các vấn đề này cũng cần được làm rõ trong nội dung hoặc phần kết luận của Luận án.

6, Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

Ưu điểm: Luận án dài 140 trang, gồm 05 Chương chính. Bố cục của Luận án chặt chẽ, dễ theo dõi. Luận án được viết và trình bày tốt, mặc dù vẫn còn lỗi trình bày, lỗi diễn đạt.

Nhược điểm: NCS cần hiệu chỉnh những lỗi sau trong Luận án:

- Danh mục từ viết tắt thông thường được trình bày với 03 cột. NCS cần bổ sung thêm cột giảng nghĩa cho các từ viết tắt. Ngoài ra, nhiều từ viết tắt trong Luận án cần được bổ sung vào danh mục và các từ viết tắt cần được sắp xếp theo thứ tự Alpha-Bet

- Thông thường, quyền Luận án thường có phần tóm tắt bằng Tiếng Việt và Tiếng Anh (không phải các quyền tóm tắt)

- NCS cần rà soát lỗi chính tả và lỗi diễn đạt trong toàn bộ Luận án

- NCS nên bổ sung kết luận cho Chương 1, Chương 2

- NCS nên có phần dẫn nhập ở đầu mỗi Chương. Phần dẫn nhập này ngoài mục đích tóm tắt Chương còn là lời kết nối nội dung giữa các Chương với nhau, nhằm tạo ra sự liên mạch về nội dung. Hơn nữa, NCS cần nêu rõ trong phần mở đầu này, kết quả của Chương đã được công bố trong các công trình nào của NCS.

- Mục kết luận của Luận án hầu như chỉ tóm tắt lại các nội dung đã đề cập trong Luận án. Trong mục này, NCS cần đưa ra những kết quả quan trọng đã đạt được trong Luận văn, đưa ra các kiến nghị, giải pháp, ưu/nhược điểm của phương pháp nghiên cứu. Do đó, NCS cần phải hiệu chỉnh mục kết luận này.

- Phần hướng phát triển đề tài cũng cần được viết cụ thể hơn. NCS cần làm rõ hơn các vấn đề cần được nghiên cứu tiếp theo trong tương lai

- NCS cần rà soát lại trình bày các ký hiệu toán học. Nhiều ký hiệu toán học không nằm cùng hàng với text. Kích thước của nhiều ký hiệu toán học cũng lớn hơn so với text.

- NCS nên bổ sung thảo luận/giải thích cho các biểu thức toán học từ trang 44-46

- NCS rà soát lại phần trình bày trong phụ lục

7, Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố

NCS đã công bố 01 bài báo tạp chí SCIE xếp hạng Q2 (Scimago) và 03 bài báo Hội nghị Quốc tế có chỉ số Scopus. Cụ thể:

- 01 công bố trên tạp chí Mobile Networks and Applications, ISSN 1383469X, tạp chí SCIE, Q2

- 01 công bố trên Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 475. Springer, Cham, ISBN 978-3-031-28815-9

- 01 công bố trên Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 579. Springer, Cham., Springer, Cham. ISBN: 978-3-031-58877-8

- 01 công bố trên International Conference on Energy, Infrastructure and Environmental Research (EIER 2025), E3S Web of Conferences 626, 04002 (2025), eISSN: 2267-1242.

Nhận xét:

- Nội dung của các công bố phù hợp với nội dung của Luận án. Nội dung của các công bố có tính gắn kết, được phát triển từ vấn đề cần giải quyết của Luận án.

- NCS là tác giả đầu của tất cả các công trình công bố

- Công bố khoa học đạt yêu cầu về chất lượng (tạp chí quốc tế uy tín, hội nghị quốc tế uy tín, tạp chí và hội nghị có phản biện, có chỉ số SCIE/Scopus)

- NCS cần trình bày cụ thể kết quả của các Chương trong Luận án được công bố trong công trình nào?

- NCS cần có giấy xin phép sử dụng các kết quả trong các công bố của các tác giả ngoài GVHD.

Câu hỏi 4: NCS có đang gửi đăng công trình mới, theo hướng phát triển tiếp đến của Luận án?

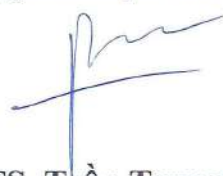
8, Kết luận chung cần khẳng định mức độ đáp ứng các yêu cầu đối với một luận án tiến sĩ. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án hay không; luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ được hay không.

Bản tóm tắt Tiếng Anh và Tiếng Việt đạt yêu cầu về nội dung và hình thức trình bày. Các bản tóm tắt phản ánh trung thành nội dung cơ bản của luận án.

Kết Luận chung: Luận án của NCS đạt yêu cầu Luận Án Tiến Sĩ (NCS cần hiệu chỉnh Luận án theo góp ý của Hội đồng)

TpHCM, ngày.....tháng.....năm 2026

Người nhận xét



PGS.TS. Trần Trung Duy

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển.**

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn:

Người nhận xét: Nguyễn Tấn Nó

Cơ quan công tác: Trường ĐH Sư Phạm Kỹ Thuật Vĩnh Long

Nội dung nhận xét:

1, Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Nội dung nghiên cứu trong luận án có tính cấp thiết cao trong bối cảnh giáo dục hiện nay, đặc biệt là nhu cầu nâng cao chất lượng tuyển chọn và huấn luyện đội tuyển học sinh giỏi quốc gia môn tin học. Việc áp dụng các phương pháp học máy và khai phá dữ liệu giáo dục là xu hướng hiện đại, phù hợp với sự phát triển của khoa học dữ liệu và trí tuệ nhân tạo. Về mặt khoa học: luận án góp phần mở rộng hướng nghiên cứu khai phá dữ liệu giáo dục, đặc biệt trong bài toán tuyển chọn học sinh giỏi môn tin học, vốn chưa được nghiên cứu nhiều. Về ý nghĩa thực tiễn: kết quả nghiên cứu trong luận án có khả năng ứng dụng trực tiếp trong các trường THPT chuyên, giúp nâng cao hiệu quả đào tạo học sinh giỏi, giúp huấn luyện viên tuyển chọn thí sinh một cách khoa học hơn và nâng cao hiệu quả huấn luyện, từ đó tăng số lượng giải thưởng quốc gia.

2, Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Luận án đã lượt khảo các đề tài nghiên cứu liên quan trong và ngoài nước và chỉ ra được khoảng trống nghiên cứu, cụ thể chưa có nghiên cứu nào tập trung vào bài toán tuyển chọn học sinh giỏi bằng học máy. Các nội dung nghiên cứu thể hiện tính mới, không trùng lặp với các công trình đã công bố, việc trích dẫn các tài liệu tương đối đầy đủ, đảm bảo tính khoa học và trung thực.

3, Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số

Nội dung luận án phù hợp với ngành khoa học máy tính (mã số: 9480101), đặc biệt trong lĩnh vực học máy, khai phá dữ liệu và ứng dụng trong giáo dục, phản ánh đúng hai hướng chính: Tuyển chọn học sinh bằng học máy (Winner-domain + Winner-cosin), huấn luyện bằng trực quan hóa và học khám phá có hướng dẫn.

4, Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

Luận án sử dụng các phương pháp hiện đại và phù hợp như: học máy có giám sát, khai phá dữ liệu giáo dục, mô hình hoá toán học, phân tích dữ liệu thực nghiệm trên dataset thực tế nhiều năm. Đặc biệt, tác giả đề xuất các phương pháp mới: Winner-domain và Winner-cosin. Các phương pháp được kiểm chứng bằng thực nghiệm trên dữ liệu thực tế của 2 kỳ thi học sinh giỏi cấp quốc gia (2023-2024 và 2024-2025), đảm bảo độ tin cậy cao.

5, Kết quả nghiên cứu mới của tác giả

Một số đóng góp nổi bật của luận án:

- Xây dựng phương pháp mô hình hóa năng lực học sinh dạng vector nhiều chiều
- Đề xuất thuật toán Winner-domain để xác định vùng năng lực của học sinh có khả năng đạt giải
- Kết hợp Winner-domain + Winner-cosin để tuyển chọn đội tuyển
- Xây dựng mô hình học khám phá có hướng dẫn bằng trực quan hóa cho huấn luyện đội tuyển
- Đề xuất phương pháp lượng hóa dữ liệu khảo sát định tính (ma trận và khối 3D)

6, Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

Ưu điểm : Đối tượng nghiên cứu của luận án là việc lựa chọn học sinh giỏi quốc gia nên có tính mới và ý nghĩa thực tiễn cao, phương pháp nghiên cứu là sự kết hợp tốt giữa lý thuyết và thực nghiệm, với các phương pháp nghiên cứu rõ ràng, logic, nội dung nghiên cứu có đề xuất thuật toán mới, với những dữ liệu được thu thập thực tế và có sần lọc, kiểm chứng cụ thể.

Nhược điểm: Quy mô dataset còn hạn chế (phụ thuộc vào kết quả khảo sát), thiếu sự mở rộng sang các học phần khác ngoài Tin học, thiếu so sánh sâu các mô hình học sâu (deep learning).

7, Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố

Luận án đã có các công trình công bố liên quan (CT1, CT2, CT3, CT4), phản ánh đúng nội dung nghiên cứu chính. Các công trình có giá trị khoa học tốt, đóng góp vào lĩnh vực khai phá dữ liệu giáo dục và ứng dụng học máy.

8, Kết luận chung cần khẳng định mức độ đáp ứng các yêu cầu đối với một luận án tiến sĩ. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án hay không; luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ được hay không.

Luận án đáp ứng đầy đủ yêu cầu của một luận án tiến sĩ. Bảng tóm tắt phản ánh đúng nội dung cơ bản của luận án. Có thể đưa ra bảo vệ cấp Học Viện.

TpHCM, ngày....tháng....năm 2026

Người nhận xét



Nguyễn Tấn Nó

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển.**

Ngành: Khoa học máy tính Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn: PGS.TS. Trần Vĩnh Phước

Người nhận xét/ Người phản biện: Phạm Thế Bảo

Cơ quan công tác: Khoa CNTT, Đại học Sài Gòn

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Bài toán xác định ứng viên chọn vào đội tuyển là một bài toán cần thiết để có thể chọn lựa được các ứng viên phù hợp từ một tập các đối tượng, với điều kiện các ứng viên đội tuyển là những ứng viên tiềm năng đạt giải trong các kỳ thi. Ngoài việc đáp ứng các kỳ thi, bài toán còn có thể mang ý nghĩa lựa chọn nhân tài và định hướng nhân tài phát triển làm hạt giống của khoa học công nghệ, là nhân tố phát triển quốc gia.

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Theo hiểu biết của người nhận xét đến thời điểm nhận xét thì chưa có sự trùng lặp với các công trình, luận án đã công bố. Nội dung rõ ràng có trích dẫn tài liệu tham khảo. Tuy nhiên việc trích dẫn vẫn chưa đầy đủ, còn nhiều chỗ cần xác định của tác giả hay sử dụng thông tin của nhà nghiên cứu khác.

3. Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số

Đề tài phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số.

4. *Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu*

Phương pháp nghiên cứu đáp ứng tính hiện đại và có thể tin cậy. Tuy nhiên, cần xem xét kỹ các nhận xét bên dưới.

5. *Kết quả nghiên cứu mới của tác giả*

NCS đề xuất quy trình và thuật toán để có thể chọn lựa các học sinh cho đội tuyển và ứng viên này có khả năng đạt giải khi đi thi.

6. *Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án*

Luận án có ưu điểm về hướng tiếp cận, tuy nhiên cần xem xét những nhận xét bên dưới.

7. *Nội dung luận án đã được công bố trên tạp chí, kỳ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố*

Các kết quả được công bố trên tạp chí và hội nghị có uy tín, trong đó có 01 bài báo thuộc Scopus Q2 có IF là và 03 bài hội nghị.

8. *Kết luận chung cần khẳng định mức độ đáp ứng các yêu cầu đối với một luận án tiến sĩ. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án hay không; luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ được hay không.*

Luận án đáp ứng một luận án tiến sĩ. Luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ. Luận án cần được phải chỉnh sửa theo góp ý của người nhận xét.

9. *Một số vấn đề tác giả cần xem xét chỉnh sửa.*

a. Hình thức:

- Nên dùng thuật ngữ đã được thống nhất cho phù hợp trong luận án, ví dụ khai phá dữ liệu – khai thác dữ liệu. Trang 1 – “khoa học máy tính đã **xuyên ngành** vào giáo dục”, “áp dụng các **phương pháp giảng dạy mới**” – “phương pháp giảng dạy phù hợp”; trang 2, ...
- Nên dùng từ Việt ngữ cho chính xác, ví dụ trang 20 – “kỹ năng nhân dạng nhu cầu thông tin”, mà là “phát hiện nhu cầu thông tin”; trang 33 – “Cách sống” đây là văn nói, NCS có thể dùng “phong cách sống”...
- Không nên trích dẫn tài liệu tham khảo ở tiêu đề, mục lục.
- Không nên dùng tài liệu tham khảo là các tài liệu công khai ở arXiv, vì đây là nơi không có bất kỳ chuyên gia hay tổ chức nào đánh giá.
- Không nên giải thích Anh ngữ trên tiêu đề, ví dụ mục 2.4, 3.3 – “câu hỏi khảo sát (Questionnaire)”, ...

- Nên chuyển các từ Anh ngữ sang Việt ngữ hoàn toàn khi có từ tương đương, ví dụ “Chương 3. Dataset”, trang 2, ...
- Danh mục viết tắt nên thêm Việt ngữ. Các từ viết tắt phải được viết tắt lần đầu tiên với từ đầy đủ, ví dụ trang 13, chứ không phải dùng từ viết tắt xong mới có từ đầy đủ và viết tắt được trình bày.
- NCS không nên trình bày các công trình của mình trong mục tài liệu tham khảo, mà NCS đã có danh mục các công trình và được chỉ số hóa CT1, CT2, ... nên có thể tham chiếu [CT1] ... trong luận án.
- Công thức, bảng, đồ thị, ... nên đánh số và trình bày vào văn bản.

b. Nội dung:

- Luận án tập trung nghiên cứu phương pháp và thực nghiệm để giải quyết bài toán chọn lựa ứng viên cho đội tuyển học sinh giỏi, chương 2 và 5 không có đóng góp vào luận án, NCS xem xét không nên đưa vào luận án làm giảm giá trị luận án.
- Mục tiêu luận án nên đưa về giải quyết bài toán tổng quát, chọn đội tuyển. Còn thực nghiệm thì chọn đội tuyển tin học. Vì bài toán là xây dựng phương pháp chọn đội tuyển nên các ứng viên không thể là học sinh có thể bỏ học được, do đó mục tiêu “can thiệp và ngăn chặn sớm bỏ học” – trang 2 là không thể.
- Tên luận án “Tiếp cận học máy để tuyển chọn và trực quan hoá đề huấn luyện học sinh đội tuyển” gồm 02 bài toán: tuyển chọn đội tuyển và sau đó trực quan hoá đề huấn luyện, nhưng luận án chưa có phần huấn luyện, và huấn luyện thì như thế nào, có cá nhân hóa đối tượng không? Cũng như sau huấn luyện phải có đánh giá tính hiệu quả. Theo người nhận xét, NCS nên điều chỉnh mục tiêu “Xây dựng mô hình tuyển chọn ứng viên và trực quan hóa năng lực ứng viên để chọn lựa bài học theo hướng cá nhân hóa” thì hợp lý hơn.
- Các câu hỏi để thu thập thông tin của ứng viên mà NCS xây dựng dựa vào cơ sở nào (cần giải thích), vì đây chính là bài toán quan trọng quyết định kết quả phía sau. Đã được những chuyên gia về giáo dục đảm bảo hay không? Có ai kiểm chứng? Đặc biệt là bối cảnh ở Việt Nam, khác với các quốc gia; cho dù cùng Đông Nam Á cũng khác chứ không xem xét đến Châu Á, Châu Âu, Châu Mỹ,
- Trang 10, NCS cho rằng “Kết quả hoạt động của mỗi cơ sở giáo dục phụ thuộc 3 thành phần, người học, người dạy, môi trường giáo dục gồm cơ sở vật chất và

công tác quản lý”, theo hiểu biết của người nhận xét thì ngoài 03 thành phần trên thì thành phần thứ 4 không kém quan trọng là “nội dung – phương pháp”, NCS cần xem xét kỹ.

- Trang 12, NCS cho rằng “một yếu tố quan trọng để đánh giá chất lượng cơ sở giáo dục là năng lực do học tập của người học”, theo người nhận xét thì “một trong những yếu tố quan trọng để đánh giá chất lượng cơ sở giáo dục là phát triển năng lực của người học, năng lực ở đây thể hiện qua kiến thức và kỹ năng”.
- Một số nhận định khá chủ quan của NCS cần xem xét kỹ, ví dụ trang 14 – “ngày nay, khai phá dữ liệu giáo dục là hướng nghiên cứu mới nổi, phát triển những phương pháp nghiên cứu loại dữ liệu mới lạ từ lĩnh vực giáo dục để hiểu tốt hơn về người học và môi trường người học [30]. Những thuật toán thường dùng trong khai phá dữ liệu giáo dục để dự báo là classification, regression, density estimation, clustering; hoặc để khai phá quan hệ là khai phá luật kết hợp, ...”. Các nước phát triển họ đã phát triển đánh giá năng lực của từng học sinh, sinh viên trên các hệ thống của họ để cá thể hóa bài học, bài tập, kiểm tra, đánh giá nhằm phát triển tối ưu nhất cho năng lực riêng biệt của từng cá thể; ví dụ giáo dục phổ thông ở Đức
- Cần có những định nghĩa chính xác, ví dụ trang 19 – NCS dùng tiêu đề là câu hỏi “Học là gì?” nhưng nội dung không trả lời được câu hỏi này; trang 41 và 42 – “đặc trưng năng lực học tập” và “phi học tập”. Trang 55 – “Winner-domain tốt” là gì? Như thế nào là tốt?
- NCS cho rằng “Học khám phá là một mô hình giáo dục mới” – trang 20; thật sự mô hình này xuất hiện 1960, còn ở Việt Nam xuất hiện khoảng 2010 đã được chương trình giáo dục phổ thông 2018 triển khai; đây là phương pháp dạy – học dựa trên thuyết kiến tạo, gồm 05 thành phần: gắn kết, khám phá, giải thích, củng cố, đánh giá. Phương pháp khám phá với học sinh là học giải quyết vấn đề, với giáo viên là dạy học dựa trên truy vấn từ học sinh. NCS cần trình bày cho phù hợp.
- NCS xây dựng bộ dữ liệu cho thực nghiệm, đây là dữ liệu liên quan đến con người nên:
 - Phải được sự đồng ý của cơ quan quản lý các đối tượng được khảo sát.

- Đồng thời phải được sự đồng ý của đối tượng khảo sát (nếu đối tượng trên 18 tuổi), vì đây là đối tượng học sinh phổ thông dưới 18 tuổi nên phải có được sự đồng ý từ cha/mẹ hay người giám hộ của học sinh.
- Trong luận án, NCS dựa trên các tiêu chí tiêu chuẩn từ nhiều nguồn tài liệu tham khảo để xây dựng tiêu chí tiêu chuẩn của mình, nhưng các tiêu chí tiêu chuẩn mà NCS tham khảo đều là tiêu chí tiêu chuẩn cho sinh viên đại học (trên 18 tuổi) và không phải chọn lựa cho đội tuyển học sinh giỏi ở phổ thông. NCS cần tham khảo từ các nghiên cứu của các nhà nghiên cứu giáo dục ở Việt Nam, vì Việt Nam sẽ có những tính chất khác trên thế giới. Và đặc biệt là cho học sinh phổ thông. Điều này dẫn đến các yếu tố ảnh hưởng năng lực học sinh sẽ khác nhau. Với các bạn học sinh mong muốn tham gia đội tuyển thì không thể nào có thói quen “uống rượu bia”. Từ tiêu chí tiêu chuẩn phù hợp mới xây dựng bảng câu hỏi gồm các loại: Câu hỏi đóng: Câu hỏi dạng “Có – Không”, Câu hỏi một lựa chọn, Câu hỏi nhiều lựa chọn, Câu hỏi phân cực, Thang đo xếp hạng, Câu hỏi phức hợp dùng thang đo Likert, Thang đo dự định, ...; Câu hỏi mở: Câu hỏi mở hoàn toàn (tự do trả lời), Câu hỏi mở dạng “hoàn chỉnh câu”, Câu hỏi hồi cố, ... Đồng thời có phương án xử lý câu trả lời qua các phương pháp kiểm tra chéo, Sử dụng câu hỏi gài, ... Cuối cùng là mã hóa dữ liệu. NCS không được nêu thông tin của các bạn học sinh, bảng 3.1 (vi phạm luật định). Những câu hỏi phải được chuyên gia tâm lý học, sự phạm xem xét, ... NCS đưa ra câu hỏi “Em nghĩ mình như thế nào?” chưa phù hợp với học sinh (trang 129), và F_{24} và F_{25} là giống nhau nhưng lại có nhưng đáp án khác nhau hoàn toàn.
- NCS nên tổng quát hóa bài toán và phương pháp giải quyết của mình, ví dụ số lượng câu khảo sát sẽ phụ thuộc chuyên môn vì thế nên là N, không nên là 40. Số 40 là phần thực nghiệm với cụ thể là môn tin học. Các xử lý, tính toán, ... NCS nên chuyển về thuật toán, có đặt tên, để làm nổi bật những đóng góp của luận án.
- NCS nên dựa vào các nghiên cứu khác (về mặt giáo dục học) để có được tỷ lệ của hai loại năng lực học tập và năng lực phi học tập, vì tỷ lệ này cũng sẽ ảnh hưởng kết quả sau này. Từ đây mới có được số lượng câu hỏi phù hợp.
- Các biểu thức, công thức, ... nên trình bày theo đúng quy định của toán; ví dụ trang 44 - “ $a_s = \frac{\delta}{M} \sum_1^M a_s^m \mid a_s^m \in \{0,1\} \mid m = 1,2,\dots,M; s = 19,\dots,40$ ” nên viết

thành $a_s = \frac{\delta}{M} \sum_{s=1}^{40} \sum_{m=1}^M a_s^m$ với $a_s^m \in \mathbb{B}$ – tập *Bool*. Hay trang 63, NCS trình bày “nếu có ít nhất một $Dg^q, \forall q | q = 1, 2, \dots$ chứa $x_l, x_l \in Dg^q, \exists q | q = 1, 2, \dots$ ” vậy q là vô hạn; có thể viết lại $\forall l \in \overline{1..L}$ thì $\exists q \in \overline{1..N}$ sao cho $x_l \in Dg^q \subset X$.

- NCS dựa vào “Áp dụng triết lý “gần mực thì đen, gần đèn thì sáng”, luận án đề xuất ...”; đây là câu tục ngữ của người Việt – đây không phải triết lý. Đây là luận án tiến sỹ khoa học NCS cần dựa vào cơ sở lý luận chặt chẽ. Tại sao NCS lại phân hoạch ngẫu nhiên tập dữ liệu thành tập huấn luyện và tập đánh giá nhiều lần? Dựa trên cơ sở nào? Thật sự đây là phương pháp k-fold giải quyết cho bài toán dữ liệu nhỏ không đủ, hiện nay cũng đã có nhiều phương pháp khác hiệu quả hơn k-fold.
- Mục “4.3 phát biểu bài toán”, theo ý kiến của người nhận xét nên đưa lên đầu luận án. Và các kết quả đưa vào phần thực nghiệm.
- Quy trình thiết lập Winner-domain, trang 55 – hình 4.1, NCS chỉ huấn luyện dữ liệu là Winner-Dataset, đây là mô hình học thiên kiến – chỉ học 01 thông tin duy nhất, nhưng tập test lại có Nonwinner-Dataset, vậy làm sao có thể test. Trong quá trình chọn lựa Winner-domains tốt không có bất kỳ tiêu chí tiêu chuẩn để chọn lựa, đồng thời việc phân hoạch nhiều lần ở bước thứ hai không có ý nghĩa về mặt logic cũng như khoa học, vì sơ đồ xử lý chỉ có một lần làm duy nhất.
- NCS nên trình bày các sơ đồ theo đúng quy định, hình thoi, hình bình hành, hình chữ nhật, hình ellipse, ...
- Tại bước 2 ở thuật toán Winner-domain, với tập Winner-Dataset ban đầu ở bước 1 từ đâu có với đầu vào thuật toán chỉ có tập X , giá trị trọng tâm w_c^p ở đâu có, được tìm như thế nào? Có phải $V^p \in 2^W$. Điều kiện nào đảm bảo tính hội tụ của phương pháp chọn để thuật toán có thể dừng, vì nếu W có 100 phần tử thì (tổng quát là $\sum_{i=0}^N C_N^k = 2^N$) số trường hợp xảy ra là 1,267,650,600,228,229,401,496,703,205,376 phần tử. Đây chính là nút thắt bùng nổ tổ hợp mà NCS cần đánh giá.
- NCS cần xem xét và tính toán độ phức tạp theo đúng phương pháp.
- Thuật toán Winner-cosin là NCS đề xuất thì cần trình bày lý do sử dụng độ đo tương đồng Cosine mà không sử dụng độ đo khác ở đây.

- Với kết quả tại các bảng, ví dụ bảng 4.7, các giá trị R_2^1 và $\cos\theta_2^1$ không có ý nghĩa, NCS cần có phân tích đánh giá
- NCS nên xem xét khi lựa chọn đội tuyển không chỉ có thành viên đội tuyển và không phải thành viên đội tuyển mà còn có ứng viên dự bị ở đội tuyển. Vấn đề có học thêm hay không, tần suất học thêm, ...

TpHCM, ngày 20 tháng 04 năm 2026

Người nhận xét



Phạm Thế Bảo

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: Tiếp cận học máy để tuyển chọn và trực quan hoá đề huấn luyện học sinh đội tuyển.

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Nghiên cứu sinh: Huỳnh Thị Ngọc Cẩm

Người hướng dẫn: PGS. TS. Trần Vĩnh Phước

Người nhận xét/ Người phản biện: PGS. TS. Nguyễn Tấn Cẩm

Cơ quan công tác: Trường Đại học Công nghệ thông tin, ĐHQG TP.HCM.

Nội dung nhận xét:

1, Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án:

Đề tài luận án có tính thời sự và cần thiết, xuất phát từ nhu cầu thực tế của các trường THPT (đặc biệt trường chuyên) trong việc tuyển chọn đội viên và huấn luyện đội tuyển môn Tin học dự thi Học sinh giỏi quốc gia (HSGQG) theo hướng có căn cứ dữ liệu, giảm phụ thuộc cảm tính. Ý nghĩa khoa học thể hiện ở việc kết hợp khai phá dữ liệu giáo dục, học máy có giám sát với mô hình trực quan hóa phục vụ học khám phá có hướng dẫn, tạo ra một quy trình và mô hình có thể kiểm chứng thực nghiệm. Ý nghĩa thực tiễn thể hiện ở các kết quả đối sánh theo kỳ thi thật (các năm gần đây), phục vụ trực tiếp công tác tuyển chọn và huấn luyện đội tuyển.

2, Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Luận án thể hiện hướng tiếp cận tương đối khác biệt so với các công trình phổ biến về dự báo năng lực học sinh, khi tập trung vào độ tương đồng năng lực với nhóm “winners” (đã đạt giải) thông qua hai tiếp cận Winner-domain và Winner-cosin để ra quyết định tuyển chọn. Về trích dẫn, luận án có đối chiếu các hướng nghiên cứu liên quan và nêu rõ các mô hình, tiếp cận do tác giả phát triển, đồng thời chỉ ra các công bố liên quan [CT1]–[CT4]. Tuy nhiên, để tăng tính kiểm chứng, nên chuẩn hóa đồng nhất cách trích dẫn và trình bày rõ hơn ranh giới đóng góp mới so với các hướng học máy đã có.

3, Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số

Tên đề tài phản ánh đúng nội dung triển khai: tiếp cận học máy để tuyển chọn; trực quan hóa, học khám phá có hướng dẫn để huấn luyện đội tuyển. Nội dung luận án thuộc đúng ngành Khoa học máy tính, phù hợp với mã số khi sử dụng các kỹ thuật học máy, xây dựng dataset, đánh giá theo confusion matrix, TPR, FPR... và thiết kế mô hình, qui trình ứng dụng.

4, Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

Phương pháp nghiên cứu có cấu trúc tương đối rõ: thiết kế, tiền xử lý dữ liệu, lượng hóa đặc trưng năng lực, xây dựng dataset, thực nghiệm so sánh với mô hình khác (Random Forest) và đối sánh theo các kỳ thi. Độ tin cậy được củng cố nhờ thực nghiệm nhiều năm (winner, nonwinner qua các năm học) và đánh giá bằng chỉ số định lượng (TPR, FPR, FNR, TNR). Dù vậy, cần mô tả chặt chẽ hơn về quy mô mẫu, cách phân hoạch, đối chứng, và kiểm soát sai lệch (ví dụ: overfitting theo năm, hiệu ứng “đặc thù trường”).

5, Kết quả nghiên cứu mới của tác giả

Luận án nêu ba nhóm đóng góp: (1) tiếp cận thiết kế câu hỏi khảo sát định tính đơn giản và lượng hóa hồi đáp; (2) đề xuất Winner-domain và cây

quyết định Winner-cosin để tuyển chọn; (3) mô hình học khám phá có hướng dẫn bằng kỹ thuật trực quan để huấn luyện đội tuyển. Kết quả thực nghiệm cho thấy mô hình Winner-domain kết hợp Winner-cosin có tỉ lệ chọn đúng thắng giải (TPR) cao hơn Random Forest trong các năm đối sánh, và đạt kết quả thắng giải 100% trên tập đội viên được mô hình chọn trong 2 kỳ thi nêu trong luận án.

6, Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

Luận án có cấu trúc tương đối đầy đủ (mở đầu, các chương nội dung, kết luận), mô tả được quy trình dữ liệu, mô hình và đánh giá; có thực nghiệm gắn với bối cảnh thật của kỳ thi và có đối sánh với mô hình ML phổ biến.

Nên trình bày rõ hơn các giả định, giới hạn áp dụng (khác trường, khác năm, thay đổi đề thi, chuẩn năng lực); mô tả chi tiết hơn về dữ liệu, cách lượng hóa đặc trưng và quy trình tái lập; chuẩn hóa cách trình bày bảng biểu, thuật ngữ và ký hiệu để tăng tính nhất quán.

Các bảng 4.3 đến 4.29 nên được bàn luận chi tiết hơn.

Các thuật toán nên dùng định dạng kiểu thuật toán, mã giả (Ví dụ: Thuật toán trang 39).

Bảng 3.1, và Bảng 3.2 nên thuộc loại Hình.

Phần 3.8 Dataset nên trình bày chi tiết hơn.

7, Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố

Các nội dung chính của như các kết quả về Winner-domain, Winner-cosin được công bố trong [CT1], [CT2] và mô hình học khám phá có hướng dẫn trực quan được công bố trong [CT3], [CT4]. Các công bố này có ý nghĩa hỗ trợ tính mới và độ tin cậy của kết quả nghiên cứu.

8, Kết luận chung cần khẳng định mức độ đáp ứng các yêu cầu đối với một luận án tiến sĩ. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án hay không; luận án có thể đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ được hay không.

Luận án đáp ứng các yêu cầu cơ bản của một luận án tiến sĩ về mục tiêu, nội dung, phương pháp và kết quả; có đóng góp theo hướng kết hợp học máy và trực quan hóa trong bối cảnh khai phá dữ liệu giáo dục, đồng thời có thực nghiệm gắn với kết quả kỳ thi thực tế. Bản tóm tắt, cấu trúc luận án được trình bày theo hướng phản ánh các nội dung chính và đóng góp. Với các yêu cầu chỉnh sửa, bổ sung một số điểm như đã nêu, luận án có thể đưa ra bảo vệ cấp Học viện.

TpHCM, ngày 30 tháng 03 năm 2026

Người nhận xét



PGS. TS. Nguyễn Tấn Cầm

Hà Nội, ngày 23 tháng 4 năm 2026

**QUYẾT NGHỊ CỦA
HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN**

Căn cứ quyết định số 205/QĐ-HVKHCN ngày 11 tháng 03 năm 2026 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện, Hội đồng đã họp vào hồi 08 giờ 30 ngày 23 tháng 04 năm 2026 tại 291 Điện Biên Phủ, Phường Xuân Hoà, Tp.HCM để đánh giá luận án tiến sĩ.

Họ và tên NCS: **Huỳnh Thị Ngọc Cẩm**

Tên đề tài luận án: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển**

Ngành: Khoa học máy tính Mã số: 9 48 01 01

Người hướng dẫn chính: PGS.TS. Trần Vĩnh Phước

**HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN CỦA
NCS HUỖNH THỊ NGỌC CẨM KẾT LUẬN**

1. Tính phù hợp của tên đề tài và sự không trùng lặp về nội dung luận án

- Tên đề tài, nội dung và kết quả nghiên cứu của luận án phù hợp với ngành đào tạo Khoa học máy tính, mã số 9 48 01 01.
- Nội dung của luận án không trùng lặp với các luận án đã bảo vệ và các kết quả nghiên cứu đã công bố trong và ngoài nước.
- Các tài liệu tham khảo của luận án có nội dung phù hợp và đã được trích dẫn trong luận án.

2. Kết quả, ý nghĩa khoa học, thực tiễn của đề tài

Luận án nghiên cứu, đề xuất các mô hình ra quyết định nhằm tuyển chọn và huấn



luyện học sinh đội tuyển. Các đóng góp chính của luận án như sau:

- Xây dựng bộ dữ liệu về năng lực thí sinh với tiếp cận ma trận giải đáp và khối 3D hỏi đáp nhằm tuyển chọn và huấn luyện học sinh đội tuyển.
- Đề xuất thuật toán Winer- domain và Winer -cosin để tuyển chọn học sinh đội tuyển
- Phát triển mô hình học khám phá có hướng dẫn trực quan để huấn luyện học sinh đội tuyển.

Các đóng góp của luận án có ý nghĩa khoa học và thực tiễn trong lĩnh vực khai phá dữ liệu giáo dục

3. Những thiếu sót của luận án, vấn đề cần bổ sung, sửa chữa

NCS cần tiếp thu, rà soát, chỉnh sửa, bổ sung nội dung luận án theo ý kiến đóng góp trong bản nhận xét của các thành viên Hội đồng và Biên bản của Hội đồng đánh giá luận án tiến sĩ cấp Học viện trước khi nộp luận án cho Thư viện Quốc gia Việt Nam, tập trung vào các vấn đề chính như sau:

- Rà soát, chỉnh sửa luận án về trình bày, trích dẫn lại tài liệu tham khảo, đặt tên các đề mục chuẩn hoá các công thức bảng biểu, các sơ đồ hình vẽ về thuật toán.
- Rà soát, chỉnh sửa về xây dựng bộ dữ liệu bao gồm quy trình, phương pháp, tiêu chí đánh giá, giá trị bộ dữ liệu cho cộng đồng
- Rà soát, chỉnh sửa các mô hình đề xuất về Winer- domain, Winner- cosin, mô hình học khám phá có hướng dẫn trực quan để làm nổi bật các đóng góp của luận án.

4. Mức độ đáp ứng yêu cầu của luận án tiến sĩ về cả nội dung và hình thức theo các quy chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo

Luận án của NCS Huỳnh Thị Ngọc Cẩm đáp ứng yêu cầu của một luận án tiến sĩ ngành “Khoa học máy tính”, mã số 9 48 01 01 về nội dung và hình thức theo các quy chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo và của Học viện Khoa học và Công nghệ.

Kết luận:

Kết quả bỏ phiếu đánh giá luận án của Hội đồng: 07/07 thành viên tán thành.

Hội đồng kết luận thông qua luận án, đề nghị Học viện Khoa học và Công nghệ công nhận kết quả bảo vệ và cấp bằng tiến sĩ cho NCS Huỳnh Thị Ngọc Cẩm

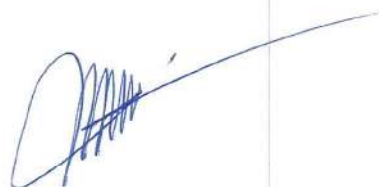
Quyết nghị này được 07/07 thành viên Hội đồng biểu quyết công khai thông qua.

THƯ KÝ



TS. Nguyễn Tấn Nô

CHỦ TỊCH



PGS.TS. Nguyễn Long Giang

XÁC NHẬN CỦA TỔ CHỨC ĐẠO TẠO
PHÓ GIÁM ĐỐC



Nguyễn Thị Trung



Hà Nội, ngày 23 tháng 4 năm 2026

**BIÊN BẢN CỦA
HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN**

Căn cứ quyết định số 205/QĐ-HVKHCN ngày 11 tháng 03 năm 2026 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện, Hội đồng đã họp vào hồi 08 giờ 30 phút ngày 23 tháng 04 năm 2026 tại 291 Điện Biên Phủ, Phường Xuân Hoà, Tp.HCM để đánh giá luận án tiến sĩ.

Nghiên cứu sinh: **Huỳnh Thị Ngọc Cẩm**

Tên đề tài luận án: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển**

Ngành: Khoa học máy tính

Mã số: 9 48 01 01

Người hướng dẫn chính: PGS.TS. Trần Vĩnh Phước

THAM DỰ BUỔI BẢO VỆ GỒM CÓ

- Đại diện cơ sở đào tạo:

1. TS. Trương Nguyên Vũ– Phó Viện trưởng Viện Cơ học

- Thành viên Hội đồng có mặt: 7/7 thành viên

1. PGS.TS. Nguyễn Long Giang, Chủ tịch Hội đồng

2. TS. Nguyễn Như Sơn, Phản biện 1

3. PGS.TS. Lê Kim Hùng, Phản biện 2



- | | |
|----------------------------|----------------|
| 4. PGS.TS. Trần Trung Duy, | Phản biện 3 |
| 5. TS. Nguyễn Tấn Nó, | Ủy viên thư ký |
| 6. PGS.TS. Phạm Thế Bảo, | Ủy viên |
| 7. PGS.TS. Nguyễn Tấn Cầm, | Ủy viên |
- Thành viên Hội đồng vắng mặt: 0
 - Đại diện cán bộ hướng dẫn: PGS.TS. Trần Vĩnh Phước
 - Cùng tham dự buổi bảo vệ còn có nhiều cán bộ nghiên cứu khoa học trong và ngoài Học viện.

TIẾN TRÌNH BUỔI BẢO VỆ

1. Đại diện cơ sở đào tạo, TS. Trương Nguyên Vũ, tuyên bố lý do, giới thiệu đại biểu và đọc quyết định số 205/QĐ-HVKHCN ngày 23 tháng 04 năm 2026 của Giám đốc HVKHCN về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện cho NCS Huỳnh Thị Ngọc Cầm và đề nghị Chủ tịch Hội đồng điều khiển phiên họp.

2. Chủ tịch Hội đồng, PGS.TS. Nguyễn Long Giang, công bố danh sách thành viên có mặt là 07, thông qua chương trình buổi bảo vệ, đề nghị Thư ký thông báo các điều kiện chuẩn bị cho buổi bảo vệ và đọc lý lịch khoa học của NCS.

3. Thư ký Hội đồng, TS. Nguyễn Tấn Nó thông báo các điều kiện cho buổi bảo vệ

- Đọc lý lịch khoa học của NCS Huỳnh Thị Ngọc Cầm.
- Đã nhận đủ 07 nhận xét của các phản biện và các thành viên HĐ.
- Lịch bảo vệ của NCS đã được đăng trên Cổng thông tin điện tử Học viện Khoa học và Công nghệ ngày 01/04/2026.
- Các giấy tờ cần thiết khác.

NCS Huỳnh Thị Ngọc Cầm có đủ các điều kiện về thủ tục để bảo vệ luận án trước Hội đồng đánh giá luận án cấp Học viện.

4. Các thành viên hội đồng và những người tham dự thông qua về lý lịch khoa học và quá trình đào tạo của nghiên cứu sinh.

5. Nghiên cứu sinh Huỳnh Thị Ngọc Cẩm trình bày nội dung luận án trong 30 phút trước Hội đồng. Báo cáo của NCS bao gồm các nội dung chính như sau:

- Mở đầu
- Tổng quan về các công trình khai phá dữ liệu giáo dục được liên quan
- Cơ sở lý thuyết khám phá
- Dataset năng lực thí sinh tham gia các kỳ thi học sinh giỏi.
- Tiếp cận học máy để chọn học sinh giỏi và đội tuyển.
- Tiếp cận học khám phá có hướng dẫn trực quan để hướng dẫn đội tuyển.
- Kết luận và kiến nghị với kết quả nghiên cứu của luận án.
- Trình bày các danh mục các công trình đã công bố liên quan đến nội dung nghiên cứu trong luận án.

6. Các phản biện đọc bản nhận xét và đặt câu hỏi đánh giá luận án của NCS Huỳnh Thị Ngọc Cẩm

1) **Phản biện 1, TS. Nguyễn Như Sơn** đọc nhận xét đánh giá luận án và kết luận (có văn bản kèm theo).

Ưu điểm: Nội dung nghiên cứu của luận án có tính ứng dụng thực tiễn, phù hợp với ngành Khoa học máy tính, không có tính trùng lặp, tên đề tài phù hợp với nội dung được nghiên cứu trong luận án, các công bố phù hợp với hai kết quả chính: xây dựng cơ sở dữ liệu huấn luyện lựa chọn học sinh giỏi và phương pháp giảng dạy, xây thêm mô hình học khám phá hướng dẫn trực quan.

Nhược điểm: Một số phần trình bày còn nhầm chương, nên gộp lại một số chương có mục quá ít, kiến thức nền tảng nên đưa vào phần tổng quan, tổng quan phải nêu ra kết luận, các chương cần nói rõ các công trình trong 4 công trình đã công bố. Chi tiết đóng góp trong chương 03 tập trung tạo bộ dữ liệu, các thuật toán cần

trình bày chuẩn hơn, các công trình công bố không đề vào tài liệu tham khảo, cần trình bày rõ hơn mục 4.31 và chương 4 ngắn gọn, 4.31 nên đưa lên trên và cần có kết quả chính. Chương 5 thực nghiệm đang còn yếu, dữ liệu còn hạn chế, cần làm rõ thêm vấn đề so sánh, kết luận vẫn chưa ổn; nên viết ngắn gọn hai đóng góp chính. Phần mục tiêu, hướng phát triển viết ngắn gọn, chỉ ra một số hướng khả thi.

Kết luận: đủ điều kiện, đạt yêu cầu, các công trình đúng bài toán nghiên cứu của nghiên cứu sinh.

2) **Phản biện 2, PGS.TS. Lê Kim Hùng**, đọc nhận xét đánh giá luận án và kết luận (có văn bản kèm theo).

Ưu điểm: Nghiên cứu có 03 đóng góp chính: Xây dựng bộ dữ liệu dataset; thuật toán winner domain – winner cosin; và mô hình học khám phá; đủ yêu cầu các điều kiện bảo vệ.

Nhược điểm: Cần làm rõ bài toán, thiếu keyword: dataset thực tế, ngữ cảnh sử dụng dataset dựa trên những mẫu nhỏ, cần làm rõ vấn đề xây ra. Làm rõ tính cấp thiết của 3 vấn đề; các trích văn đưa vào từ những bài báo có tính trùng lặp như: 5-31; 8-20... Cấu trúc các chương cần rà soát lại, làm rõ việc thu thập tài liệu dataset, mô hình, bài giảng, số lượng dataset còn hạn chế cần làm gì cho phù hợp. Phần đóng góp 03: chuẩn hóa thêm việc trình bày, mang lại lợi ích gì, nên cấu trúc lại chương 05.

Câu hỏi:

- +) Làm rõ dataset trong lĩnh vực này có công bố hay chưa?
- +) Định nghĩa winner domain – winner cosin phân loại theo loại gì, nằm ở khu vực nào trong machine learning ?
- +) Về mặt đóng góp cộng đồng: có thể công khai nội dung nghiên cứu được hay không? (các thuật toán, source code, dữ liệu...).

Kết luận chung: Đáp ứng yêu cầu cơ bản của luận án ngành khoa học máy tính, cần đọc và tiếp thu ý kiến của phản biện 2.

3) Phản biện 3, PGS.TS. Trần Trung Duy, đọc nhận xét đánh giá luận án và kết luận (có văn bản kèm theo)

Ưu điểm: nội dung luận án phù hợp với chuyên ngành khoa học máy tính.

Khuyết điểm: Cần nêu thêm kết luận chương, còn một số lỗi chính tả, bổ sung danh mục từ viết tắt và sắp xếp theo thứ tự; bổ sung kết luận chương 1, 2 và nên thêm phần dẫn nhập; nên hiệu chỉnh lại các kết luận cho phù hợp, có rất nhiều lỗi trình bày kí hiệu toán học, trang 44, 46 giải thích thêm ý nghĩa, kiểm tra phần phụ lục.

Câu hỏi:

+) Số lượng dataset có đủ chưa? Có phù hợp với ngữ cảnh của các trường khác hay không ?

+) Thuật toán sử dụng trong luận án có ưu điểm gì so với các phương pháp truyền thống ?

+) Lý thuyết sơ đồ có ưu điểm gì ? Dựa vào đâu chọn 18, 22 đặc trưng .

Kết luận: Các công bố phù hợp với nội dung nghiên cứu.

NCS Huỳnh Thị Ngọc Cẩm tiếp thu ý kiến nhận xét của các phản biện và trả lời đầy đủ câu hỏi của các uỷ viên phản biện.

- Về cấu trúc và hình thức trình bày, gộp chương 1 và chương 2: nghiên cứu sinh tiếp thu và hiệu chỉnh.

- Quy mô dataset: thực tế không thu được nhiều mẫu, chỉ thu thập trong thời gian 03 năm gần nhất, trong phạm vi 28 trường chuyên; mặc dù dữ liệu nhỏ nhưng có tính chọn lọc cao; vừa giới hạn thời gian và phạm vi nên dữ liệu không thể lớn hơn được và trên thế giới chưa có dataset công khai.

- Ưu điểm của thuật toán và quy trình tuyển chọn: bài toán chọn những em có tiềm năng dựa trên sử dụng thuật toán có giám sát.

- Áp dụng lý thuyết sơ đồ: nghiên cứu sinh tiếp thu và chỉnh sửa bổ sung.

- Hướng phát triển: nghiên cứu sinh đang viết thêm, bổ sung ít nhất $\frac{1}{2}$ trang.

- Quy trình xây dựng bài giảng: do luận án liên quan ngành Khoa học máy tính không mang tính thiên về giáo dục, nên không trình bày.

- Có thể công khai công trình được không: có thể công bố dataset và các phương pháp nghiên cứu trong luận án.

7. Các thành viên khác trong Hội đồng đưa ra ý kiến nhận xét; Hội đồng và những người tham dự đặt câu hỏi

- **PGS.TS. Phạm Thế Bảo**

Đóng góp của nghiên cứu sinh là phương pháp xây dựng bộ dữ liệu, đặt thành bài toán chung, xây dựng tiêu chí nên tham khảo ý kiến trong lĩnh vực giáo dục học, nên có đánh giá học sinh phổ thông. Cần mã hóa dữ liệu trước khi xử lý; chỉnh sửa một số nội dung: xem lại tập dữ liệu huấn luyện trên tập dữ liệu winner domain – winner cosin.

- **PGS.TS. Nguyễn Tấn Cầm**

Đưa ra một số góp ý: đóng góp chính là xây dựng dataset nên có tiêu chí, đánh giá của chuyên gia, độ phủ, cần làm rõ quy trình. Trình bày bản luận rất ích bằng, đưa ra nhận xét bản luận. Thuật toán sử dụng trình bày theo bước 1, bước 2, bước 3; nên trình bày lại theo kiểu bài báo khoa học. Bảng 3.1, 3.2 là hình chứ không phải bảng.

- **PGS.TS. Nguyễn Long Giang**

Luận án vẫn còn nhiều vấn đề trao đổi về đánh giá giải quyết bài toán lựa chọn. Cần rà soát lại các sơ đồ thể hiện, sơ đồ thuật toán.

Luận án cần trình bày cô đọng về quy trình xây dựng và đặc tả bộ dữ liệu sau khi xây dựng. Phần trình bày các biến quá dài dòng và nên cô đọng lại

Câu hỏi: hai hướng đề xuất hoàn toàn dựa vào ý tưởng nghiên cứu sinh tự đưa ra ? Có áp dụng cho việc lựa chọn học sinh giỏi nói chung hay chỉ đặc thù cho môn tin học, có đặc trưng khác không, cần thu thập dữ liệu khác, câu hỏi khác hay những đặc trưng khác? Machine learning dùng như thế nào? Dùng mô hình học máy là gì?

- **TS. Trương Nguyên Vũ**

Luận án cần phải hiệu chỉnh cách hành văn, trích dẫn, nghiêm túc trong nghiên cứu; nên đưa hướng mở dataset cho mọi người truy cập vào, đóng góp lớn của luận án là sử dụng machine learning cho tập dữ liệu nhỏ.

8. NCS Huỳnh Thị Ngọc Cẩm tiếp thu ý kiến nhận xét của các thành viên của Hội đồng. NCS trả lời các câu hỏi của các thành viên Hội đồng

Nếu sử dụng cho lĩnh vực khác thì cần thay đổi bộ dữ liệu, câu hỏi cũng như đặc trưng cho từng ngành; bộ dữ liệu lúc phân hoạch đã dùng machine learning lấy 70% làm dữ liệu training và 30% dữ liệu để testing, mô hình học dùng key để quyết định.

9. Đại diện tập thể hướng dẫn thầy PGS.TS Trần Vĩnh Phước phát biểu ý kiến bằng văn bản.

10. Hội đồng tiến hành họp riêng để bầu ban kiểm phiếu, bỏ phiếu kín và thảo luận thông qua quyết nghị của Hội đồng.

1) Bầu ban kiểm phiếu gồm:

- Trưởng ban: TS. Nguyễn Tấn Nó
- Ủy viên: PGS.TS. Trần Trung Duy
- Ủy viên: TS. Nguyễn Như Sơn

2) Bỏ phiếu kín và thảo luận thông qua Quyết nghị của Hội đồng.

- Trưởng ban kiểm phiếu, TS. Nguyễn Tấn Nó công bố kết quả kiểm phiếu (có biên bản kiểm phiếu).
- Chủ tịch Hội đồng, PGS.TS. Nguyễn Long Giang, thông qua Quyết nghị (có văn bản kèm theo).

3) Tóm tắt nghị quyết của Hội đồng

3.1. Tính phù hợp của tên đề tài và sự không trùng lặp về nội dung luận án

- Tên đề tài, nội dung và kết quả nghiên cứu của luận án phù hợp với Ngành đào tạo “Khoa học máy tính”, mã số “9 48 01 01”.
- Nội dung của luận án không trùng lặp với các luận án đã bảo vệ và các kết quả nghiên cứu đã công bố trong và ngoài nước.
- Các tài liệu tham khảo của luận án có nội dung phù hợp và đã được trích dẫn trong luận án.

3.2. Các kết quả chính của luận án

Luận án nghiên cứu, đề xuất các mô hình ra quyết định nhằm tuyển chọn và huấn luyện học sinh đội tuyển. Các đóng góp chính của luận án như sau:

- Xây dựng bộ dữ liệu về năng lực thí sinh với tiếp cận ma trận giải đáp và khối 3D hỏi đáp nhằm tuyển chọn và huấn luyện học sinh đội tuyển.
- Đề xuất thuật toán Winer- domain và Winer -cosin để tuyển chọn học sinh đội tuyển
- Phát triển mô hình học khám phá có hướng dẫn trực quan để huấn luyện học sinh đội tuyển.

Các đóng góp của luận án có ý nghĩa khoa học và thực tiễn trong lĩnh vực khai phá dữ liệu giáo dục

3.3. Các điểm cần bổ sung chỉnh sửa trước khi nộp luận án cho Thư viện Quốc gia Việt Nam

NCS cần tiếp thu, rà soát, chỉnh sửa, bổ sung nội dung luận án theo ý kiến đóng

góp trong bản nhận xét của các thành viên Hội đồng và Biên bản của Hội đồng đánh giá luận án tiến sĩ cấp Học viện trước khi nộp luận án cho Thư viện Quốc gia Việt Nam, tập trung vào các vấn đề chính như sau:

- Rà soát, chỉnh sửa luận án về trình bày, trích dẫn lại tài liệu tham khảo, đặt tên các đề mục chuẩn hoá các công thức bảng biểu, các sơ đồ hình vẽ về thuật toán.
- Rà soát, chỉnh sửa về xây dựng bộ dữ liệu bao gồm quy trình, phương pháp, tiêu chí đánh giá, giá trị bộ dữ liệu cho cộng đồng
- Rà soát, chỉnh sửa các mô hình đề xuất về Winer- domain, Winner- cosin, mô hình học khám phá có hướng dẫn trực quan để làm nổi bật các đóng góp của luận án.

3.4. Mức độ đáp ứng yêu cầu của luận án tiến sĩ về cả nội dung và hình thức

- Luận án của NCS Huỳnh Thị Ngọc Cẩm đáp ứng yêu cầu của luận án tiến sĩ ngành “Khoa học máy tính”, mã số “9 48 01 01” về nội dung và hình thức theo các qui chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo, của Học viện Khoa học và Công nghệ.

- Đề nghị Học viện Khoa học và Công nghệ công nhận kết quả bảo vệ và cấp bằng tiến sĩ cho NCS Huỳnh Thị Ngọc Cẩm sau khi chỉnh sửa, bổ sung luận án theo các góp ý của Hội đồng.

11. Tổng kết

- Trưởng ban kiểm phiếu, TS. Nguyễn Tấn Nó công bố kết quả bỏ phiếu đánh giá luận án.

- Chủ tịch Hội đồng, PGS.TS. Nguyễn Long Giang, đọc Quyết nghị của Hội đồng.

- Chủ tịch Hội đồng tuyên bố Hội đồng đã hoàn thành nhiệm vụ và trao lại quyền điều khiển cho Cơ sở đào tạo.

- Các đại biểu và NCS phát biểu ý kiến.

- Đại diện cơ sở đào tạo tuyên bố kết thúc buổi bảo vệ luận án tiến sĩ.

Biên bản họp Hội đồng này được 7/7 ủy viên Hội đồng biểu quyết công khai thông qua.

Buổi họp Hội đồng đánh giá luận án tiến sĩ cấp Học viện kết thúc vào lúc 11 giờ 30 phút, ngày 23 tháng 04 năm 2026.

Thư ký Hội đồng

Chủ tịch Hội đồng



TS. Nguyễn Tấn Nô

PGS.TS. Nguyễn Long Giang

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO

**KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC**



Nguyễn Thị Trung



VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

ĐẢNG CỘNG SẢN VIỆT NAM

Tp.HCM, ngày 6 tháng 5 năm 2026

**BẢN GIẢI TRÌNH CHỈNH SỬA, BỔ SUNG LUẬN ÁN TIẾN SĨ
CẤP HỌC VIỆN**

Ngày 23 tháng 4 năm 2026, Học viện Khoa học và Công nghệ đã tổ chức đánh giá luận án tiến sĩ cấp Học viện cho nghiên cứu sinh Huỳnh Thị Ngọc Cẩm theo Quyết định số 205/QĐ-HVKHCN ngày 11 tháng 3 năm 2026 của Giám đốc Học viện.

Đề tài: **Tiếp cận học máy để tuyển chọn và trực quan hoá để huấn luyện học sinh đội tuyển.**

Ngành: Khoa học máy tính Mã số: 9 48 01 01

Người hướng dẫn khoa học: PGS. TS. Trần Vĩnh Phước

Theo Biên bản của Hội đồng, NCS đã bổ sung và chỉnh sửa luận án các điểm sau đây:

STT	Nội dung đề nghị chỉnh sửa, bổ sung	Nội dung đã được chỉnh sửa, bổ sung (Ghi rõ số trang/chương/mục... đã được chỉnh sửa)
1	Nội dung 1: (Thầy Giang, Thầy Sơn) Đề nghị gộp chương 1 và chương 2.	Nghiên cứu sinh trân trọng tiếp thu và đã gộp nội dung chương 1 và nội dung chương 2 thành: Chương 1. Tổng quan các nghiên cứu và lý thuyết được liên quan trong luận án.
2	Nội dung 2: (Thầy Giang, Thầy Sơn, Thầy Cẩm, Thầy Duy, Thầy Hùng, Thầy Bảo) Cấu trúc, trình bày, và ngôn ngữ trong luận án như tên các chương mục, trùng lặp tài liệu tham khảo.	Nghiên cứu sinh trân trọng cảm ơn Quý Thầy đã chỉ ra cụ thể những thiếu sót về cấu trúc, trình bày, và ngôn ngữ trong luận án, NCS tiếp thu và đã sửa chữa trong tất cả các chương.
3	Nội dung 3: (Thầy Nó, Thầy Duy) Quy mô Dataset và các vấn đề liên quan Dataset.	Theo lý thuyết, các bài toán khai phá dữ liệu sử dụng dataset có nhiều mẫu. Tuy nhiên, theo nhiều công trình được luận án tham khảo, các bài toán toán dự báo dựa vào dữ liệu giáo dục sử dụng dataset không nhiều mẫu, từ vài chục đến vài ngàn mẫu. Đối với tiếp cận của luận án, nếu có dataset lớn cũng tốt. Tuy nhiên, với



		nghiên cứu có tính thực tiễn xã hội này, không thể tạo dataset lớn vì số lượng thí sinh thi mỗi lần không nhiều và ở rải rác cả nước, việc liên lạc thu thập dữ liệu rất khó khăn. Hai dataset sử dụng trong luận án được thu thập từ 28 trường phổ thông trong cả nước. Hơn nữa, dataset của tiếp cận được thành lập mỗi lần tuyển chọn đội viên, và đối tượng khảo sát là những thí sinh vài năm gần nhất. Số lượng mẫu ít, thực ra không trở ngại vì đây là bài toán quyết định chọn ra một số lượng rất ít để làm đội viên so với số lượng ứng viên. Dataset của môn thi nào thì áp dụng cho môn thi đó. Các tiếp cận giống nhau, nhưng không thể áp dụng một dataset dùng chung cho các môn và dùng chung cho các kỳ thi vì dữ liệu thu thập khác nhau, biến năng lực cũng khác nhau. Cho đến nay, luận án chưa tìm thấy công trình liên quan dữ liệu giáo dục có phương pháp định lượng khách quan hay đánh giá độ tin cậy của Dataset. Luận án này đã đề xuất phương pháp định lượng khách quan dataset. Vấn đề đánh giá độ tin cậy của Dataset sẽ được nghiên cứu để phát triển tiếp.
4	Nội dung 4: (Thầy Giang, Thầy Hùng) Cần trình bày cô đọng về quy trình xây dựng bộ dữ liệu, vấn đề định lượng giá trị của các biến.	Quy trình thiết lập dataset được trình bày trong hình 2.1. Các tiếp cận ma trận hồi đáp và khối 3D hồi đáp định lượng khách quan các giá trị của tập biến năng lực và giá trị các biến của vector năng lực trong không gian năng lực nhiều chiều.
5	Nội dung 5: (Thầy Giang) So sánh mô hình Winner-domain trong luận án với các mô hình trong [106] và [107]	Các tài liệu tham khảo [106] và [107] là các công trình công bố của luận án. Luận án đã điều chỉnh cách viết, trong đó các công trình của luận án công bố không xuất hiện trong mục Tài liệu tham khảo.
6	Nội dung 6: (Thầy Duy, Thầy Hùng)	Quy trình tuyển chọn và huấn luyện đội viên đội tuyển, không phải là dự đoán mà là một quy trình tổng hợp từ việc chọn để huấn luyện

	Ưu điểm của thuật toán Winner-domain và quy trình tuyển chọn.	được nghiên cứu theo nhu cầu thực tiễn và đã có kết quả là 2 năm liên Trường HMD từ một trường nhiều năm liên trước đây không có hoặc có rất ít đội viên thắng giải Tin học cấp quốc gia, nay trở thành trường thắng nhiều giải học sinh giỏi Tin học cấp quốc gia, so với số đội viên dự thi là 100% thắng giải. Tiếp cận của luận án là chọn đúng, đủ, không nhầm, sẵn sàng bỏ lại những ứng viên có ít khả năng hơn, chứ không áp dụng dự báo. Quy trình tuyển chọn đội viên được trình bày trong hình 4.2. Hầu hết các phương pháp kinh điển (theo Thầy Hùng) đều áp dụng cho bài toán dự báo và dùng tỉ lệ đúng / sai để so sánh về giá trị khoa học. Luận án phát hiện nhu cầu thực tiễn của Trường, áp dụng khoa học máy tính để giải quyết bài toán thực tiễn của Trường và vì lợi ích của Trường như TBT đã chỉ đạo nghiên cứu khoa học là phải giải bài toán thực tiễn vì nhân dân, đánh giá bằng kết quả thực tiễn (khoa học vị nhân sinh). Do đó, luận án đã đề xuất thuật toán Winner-domain để giải quyết yêu cầu thực tiễn, chưa so sánh với các thuật toán khác về phương diện khoa học. Thuật toán Winner-domain sẽ được nghiên cứu thêm để so sánh với các thuật toán khác về phương diện khoa học (khoa học vị khoa học). NCS tiếp thu các ý kiến của Quý Thầy và sẽ nghiên cứu phát triển về phương diện khoa học.
7	Nội dung 7: (Thầy Duy) Việc áp dụng Lý thuyết sơ đồ và Lý thuyết tải nhận thức trong thiết kế bài giảng trực quan được cụ thể hóa như thế nào để phù hợp với đặc thù môn Tin học? Các vấn đề này cũng cần được làm rõ trong nội dung hoặc phân kết luận của Luận án.	Lý thuyết sơ đồ và lý thuyết tải nhận biết được dùng để xây dựng quy trình bài giảng và các mô hình bài giảng. Theo cơ chế tiếp nhận thông tin và kiến tạo tri thức của con người thì Bài giảng bằng hình ảnh 3D đòi hỏi tải nhận biết của não bộ ít nhất. Các lý thuyết này được cụ thể hóa trong bài giảng thông qua phương pháp mà người hướng dẫn (người thầy) tại các hình 4.2, 4.3, 4.4.

8	Nội dung 8: (Thầy Duy) NCS nên bổ sung thảo luận/giải thích cho các biểu thức toán học từ trang 44-46.	Nghiên cứu sinh tiếp thu ý của Thầy và đã bổ sung giải thích các biểu thức toán học trang 44-46. Vì các hồi đáp khảo sát năng lực phi học tập hoàn toàn định tính. Luận án đã đề xuất các biểu thức này để định lượng một cách khách quan các giá trị của các trục biến năng lực phi học tập và giá trị các biến năng lực phi học tập của từng vector biểu diễn năng lực của thí sinh/học sinh.
9	Nội dung 9: (Thầy Duy) NCS có đang gửi đăng công trình mới, theo hướng phát triển tiếp đến của Luận án?	Nghiên cứu sinh đang viết thêm bài mới liên quan đến nội dung của luận án chưa submit.
10	Nội dung 10: (Thầy Duy) Việc áp dụng Lý thuyết sơ đồ và Lý thuyết tải nhận thức trong thiết kế bài giảng trực quan được cụ thể hóa như thế nào để phù hợp với đặc thù môn Tin học? Các vấn đề này cũng cần được làm rõ trong nội dung hoặc phần kết luận của Luận án.	Lý thuyết sơ đồ và lý thuyết tải nhận biết được áp dụng vào các quy trình xây dựng bài giảng (trang 98) và xây dựng thuật toán đánh giá hiệu năng hướng dẫn của giảng viên (hình 4.5). Câu hỏi của Thầy liên quan nhiều về khoa học giáo dục, luận án của em thuộc ngành khoa học máy tính nên em xin phép không trình bày chi tiết nội dung này.
11	Nội dung 11: (Thầy Giang, Thầy Sơn) Tài liệu tham khảo	Mô hình Winner-domain trong luận án chính là mô hình đã trình bày trong các bài báo. Tuy nhiên, luận án đã đưa nhầm các công trình này vào mục tài liệu tham khảo [106], [107]. Nghiên cứu sinh trân trọng cảm ơn Quý Thầy nhắc nhở và chuyển các công trình này vào mục Công trình công bố. Luận án đã tham khảo 124 tài liệu, trong đó có 68 tài liệu về khoa học máy tính và 56 tài liệu khoa học máy tính xuyên ngành vào giáo dục với những nghiên cứu dựa vào khai phá dữ liệu giáo dục (EDM: Educational Data Mining). Về thời gian xuất bản, luận án đã tham khảo 25 tài liệu xuất bản trong thời gian 2021-2025.

12	Nội dung 12: So sánh với các công trình khoa học máy tính khác. (Thầy Duy, Thầy Sơn) Thuật toán và quy trình tuyển chọn học sinh vào đội tuyển được đề xuất có ưu điểm gì so với phương pháp truyền thống, và mức độ cải thiện độ chính xác trong việc dự đoán học sinh đạt giải cần được nhấn mạnh trong Luận án	Khi thực hiện nghiên cứu để làm luận án, Nghiên cứu sinh đã cố gắng tìm kiếm các công trình tương tự để học tập và so sánh. Tuy nhiên, cho đến nay, Nghiên cứu sinh vẫn chưa tìm thấy công trình nào có cùng mục tiêu. Hầu hết những công trình nghiên cứu ứng dụng khai phá dữ liệu giáo dục EDM (Educational Data Mining), tập dữ liệu được phân hoạch thành 2 miền để thực hiện những bài toán dự báo như dự báo kết quả học tập của học sinh, dự báo những trường hợp học sinh sẽ bỏ học để can thiệp sớm, v.v... Trong khi đó, luận án đề xuất nghiên cứu theo nhu cầu thực tiễn của trường và vì trường (nghiên cứu theo nhu cầu thực tiễn, vì nhân dân) với bài toán chọn thí sinh chứ không phân hoạch tập ứng viên thành 2 tập để gán nhãn. Do đó, Nghiên cứu sinh không có ý định so sánh các thuật toán của những bài toán không cùng mục tiêu không cùng cách tiếp cận. Tuy nhiên, trong thực nghiệm, Nghiên cứu sinh cũng đã áp dụng thử nghiệm thực tế mô hình Random Forest để xem kết quả. So với phương pháp truyền thống, kết quả thực tiễn cho thấy rằng các tiếp cận tuyển chọn học sinh vào đội tuyển kết hợp với huấn luyện theo phương pháp học khám phá có hướng dẫn trực quan do luận án xây dựng đã nâng tầm của Trường từ kết quả không có thí sinh đạt giải hoặc chỉ 1-2 giải của nhiều năm trước thành trường có 100% đội viên đạt giải mỗi năm. Luận án không thực hiện dự đoán học sinh đạt giải để chọn mà tiếp cận của luận án là chọn đúng, chọn đủ, không chọn nhầm người. Do đó, thực tế có rất nhiều học sinh giỏi bị loại ra, không được chọn vào đội. Kết quả thực tiễn đã chứng minh tính đúng của tiếp cận Winner-domain kết hợp Winner-cosin và Học khám phá có hướng dẫn trực quan (Bảng 4.31)
----	--	---

Nghiên cứu sinh trân trọng cảm ơn Quý Thầy trong Hội đồng đánh giá luận án tiến sĩ cấp Học viện đã góp ý và tạo cơ hội cho NCS hoàn thiện luận án của mình.

NGƯỜI HƯỚNG DẪN



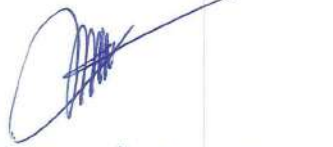
PGS.TS. Trần Vĩnh Phước

NGHIÊN CỨU SINH



Huỳnh Thị Ngọc Cẩm

CHỦ TỊCH HỘI ĐỒNG



PGS.TS. Nguyễn Long Giang

THƯ KÝ HỘI ĐỒNG



TS. Nguyễn Tấn Nô

XÁC NHẬN CỦA HỌC VIỆN
KHOA HỌC VÀ CÔNG NGHỆ

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC



Nguyễn Thị Trung

