ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

PHẠM THỊ TỐ NGA

MỘT SỐ PHƯƠNG PHÁP ĐẢM BẢO TÍNH CÔNG BẰNG CHO CÁC HỆ THỐNG HỌC MÁY TRONG LĨNH VỰC GIÁO DỤC

TÓM TẮT LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Hà Nội - 2026

Công trình được hoàn thành tại:
Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội.

Người hướng dẫn khoa học: 1. GS. TS. Nguyễn Đức Anh
2. PGS. TS. Phạm Ngọc Hùng

Phản biện 1: PGS. TS. Trần Đăng Hưng

Nơi công tác: Đại học Công nghiệp Hà Nội

Phản biện 2: PGS. TS. Phạm Văn Cường

Nơi công tác: Học viện Công nghệ Bưu chính Viễn thông

Phản biện 3: PGS. TS. Lê Hồng Phương

Nơi công tác: Trường Đại học Khoa học Tự nhiên, ĐHQGHN

Luận án được bảo vệ tại Hội đồng chấm luận án cấp Đại học Quốc Gia vào hồi 14h ngày 25 tháng 6 năm 2026 tại Phòng 212-E3, Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

NGHIÊN CỨU SINH **TẬP THỂ CÁN BỘ HƯỚNG DẪN**

PHẠM THỊ TỐ NGA **GS.TS. NGUYỄN ĐỨC ANH**

PGS.TS. PHẠM NGỌC HÙNG

XÁC NHẬN CỦA ĐƠN VỊ ĐÀO TẠO

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam,
- Trung tâm thông tin - Thư viện, Đại học Quốc gia Hà Nội.

Hà Nội - 2026

Tóm tắt

Trong bối cảnh trí tuệ nhân tạo (AI) và học máy (ML) ngày càng được ứng dụng trong ra quyết định, việc đảm bảo tính công bằng cho các mô hình ML trở nên cấp thiết, đặc biệt trong giáo dục – nơi kết quả dự đoán ảnh hưởng trực tiếp đến cơ hội học tập của người học. Tuy nhiên, dữ liệu mất cân bằng và chứa nhiều thuộc tính nhạy cảm như *giới tính*, *chủng tộc*, *sức khỏe*, *nơi sinh*, v.v. thường chứa những thiên lệch tiềm ẩn. Luận án tập trung đề xuất ba phương pháp nâng cao công bằng cho mô hình học máy trong giáo dục, đặc biệt khi tồn tại đồng thời nhiều thuộc tính nhạy cảm và mối quan hệ giao thoa giữa chúng, bao gồm: (1) *Phương pháp FairEdu* – một kỹ thuật tiền xử lý dựa trên hồi quy đa biến, có khả năng loại bỏ sự phụ thuộc giữa các biến đầu vào và nhiều thuộc tính nhạy cảm đồng thời, (2) *Phương pháp DPF* – một kỹ thuật cân bằng phân bố các tổ hợp thuộc tính nhạy cảm bằng cách sử dụng các mô hình sinh dữ liệu tổng hợp như CTGAN và LLM, và (3) *Phương pháp FairEduPlus* – một khung tiếp cận tích hợp FairEdu và DPF, xử lý công bằng hai chiều: "chiều ngang" (cân bằng dữ liệu) và "chiều dọc" (loại bỏ phụ thuộc). Ngoài ra, luận án còn đề xuất chỉ số đánh đổi (Δ_{trade_off}) nhằm định lượng mức độ mối quan hệ giữa công bằng và hiệu suất giúp lựa chọn mô hình một cách hệ thống.

Các thí nghiệm được triển khai trên bảy bộ dữ liệu giáo dục (gồm sáu bộ công khai và một bộ dữ liệu thực tế từ Trường Đại học Đại Nam), với năm mô hình học máy phổ biến. Kết quả thực nghiệm cho thấy các phương pháp đề xuất đạt hiệu quả vượt trội, cải thiện các chỉ số công bằng đồng thời không làm suy giảm hiệu suất dự đoán của mô hình.

Về mặt lý thuyết, luận án đóng góp một khung phương pháp toàn diện cho bài toán công bằng với đồng thời nhiều thuộc tính nhạy cảm cho các mô hình học máy trong lĩnh vực giáo dục, đồng thời giới thiệu chỉ số đánh đổi công bằng – hiệu suất mới phản ánh rõ mối quan hệ giữa hai mục tiêu này. Về thực tiễn, các phương pháp được đề xuất có thể áp dụng hiệu quả trong các hệ thống giáo dục thông minh, hỗ trợ đánh giá và ra quyết định công bằng, góp phần xây dựng các hệ thống trí tuệ nhân tạo minh bạch, đáng tin cậy và có trách nhiệm xã hội.

Từ khóa: Tính công bằng, Thiên vị, Hệ thống học máy, Trí tuệ nhân tạo, Giáo dục,

GIỚI THIỆU

0.1 Đặt vấn đề

0.1.1 Bối cảnh nghiên cứu về công bằng cho hệ thống học máy trong lĩnh vực giáo dục

Các hệ thống học máy trong giáo dục tiềm ẩn thiên lệch với nhóm nhạy cảm, khiến công bằng trở thành yêu cầu cốt lõi và thúc đẩy các tiếp cận loại bỏ phụ thuộc và sinh dữ liệu tổng hợp cho công bằng giao thoa.

0.1.2 Các thách thức trong nghiên cứu về đảm bảo tính công bằng cho các hệ thống học máy trong lĩnh vực giáo dục

Các thách thức chính trong đảm bảo tính công bằng cho hệ thống học máy trong giáo dục bao gồm: (1) chưa có sự thống nhất về định nghĩa và thước đo công bằng, dẫn đến khó khăn trong đánh giá; (2) hạn chế trong việc xử lý đồng thời nhiều thuộc tính nhạy cảm và công bằng giao thoa; (3) tồn tại sự đánh đổi giữa công bằng và hiệu suất mô hình; (4) dữ liệu giáo dục còn hạn chế về quy mô, mất cân bằng và bị ràng buộc bởi vấn đề riêng tư; và (5) ảnh hưởng của yếu tố con người trong thu thập dữ liệu và thiết kế hệ thống, gây thiếu nhất quán trong đánh giá công bằng.

0.2 Mục tiêu, đối tượng, phương pháp, và phạm vi nghiên cứu

0.2.1 Mục tiêu nghiên cứu

Đề xuất các phương pháp đảm bảo tính công bằng cho hệ thống học máy trong giáo dục, tập trung vào dữ liệu dạng bảng có nhiều thuộc tính nhạy cảm và mất cân bằng; kết hợp kỹ thuật tiền xử lý và dữ liệu tổng hợp

nhằm cải thiện công bằng đa nhóm, đồng thời duy trì hiệu suất và phân tích đánh đổi giữa công bằng–hiệu suất.

0.2.2 Đối tượng và phương pháp nghiên cứu

Đối tượng là các mô hình học máy ứng dụng trong giáo dục (LR, DT, RF, GB, NN). Phương pháp nghiên cứu kết hợp tổng quan lý thuyết và thực nghiệm trên nhiều bộ dữ liệu giáo dục, sử dụng các thước đo công bằng (DI, SPD, AOD, EOD) và hiệu suất (ACC, Recall, Precision, F1).

0.2.3 Phạm vi nghiên cứu

Tập trung vào các bài toán học máy có giám sát trên dữ liệu dạng bảng trong lĩnh vực giáo dục, sử dụng các bộ dữ liệu phổ biến và dữ liệu thực tế, với các thuộc tính nhạy cảm như giới tính, độ tuổi, sức khỏe, khu vực và điều kiện kinh tế. Hình 1 trình bày cây nghiên cứu tổng quát về đảm bảo công bằng cho các hệ thống AI/ML trong lĩnh vực giáo dục.

Hình 1 – Cây nghiên cứu về tính công bằng cho các hệ thống học máy trong giáo dục liên quan đến luận án.

0.3 Các đóng góp chính của luận án

Luận án có bốn đóng góp chính, cụ thể như sau:

(1) Đề xuất *Fairedu* nhằm loại bỏ phụ thuộc giữa đặc trưng và thuộc tính nhạy cảm. (2) Xây dựng *DPF* sử dụng dữ liệu tổng hợp để cân bằng các nhóm nhạy cảm. (3) Phát triển *FaireduPlus* kết hợp hai cơ chế, đảm bảo công bằng hai chiều. (4) Đề xuất chỉ số đánh đổi để phân tích quan hệ công bằng–hiệu suất.

Các đóng góp này tạo thành một khung tiếp cận thống nhất, hỗ trợ đảm bảo công bằng đa thuộc tính cho hệ thống học máy trong giáo dục.

0.4 Bố cục của luận án

Luận án gồm sáu chương, được sắp xếp từ cơ sở lý thuyết đến các đóng góp phương pháp và kết quả nghiên cứu. Chương 1 trình bày bối cảnh, mục tiêu và đóng góp. Chương 1 tổng quan về công bằng trong học máy giáo dục, bao gồm các phương pháp, dữ liệu tổng hợp và đánh đổi công bằng–hiệu suất. Ba chương tiếp theo, Chương 2, 3, 4 lần lượt giới thiệu các phương pháp đề xuất: *Fairedu* (loại bỏ phụ thuộc đặc trưng–thuộc tính nhạy cảm), *DPF* (cân bằng dữ liệu bằng sinh dữ liệu tổng hợp), và *FaireduPlus* (kết hợp hai cơ chế và đề xuất chỉ số đánh đổi). Chương 4.6 tổng kết kết quả, nêu hạn chế và định hướng nghiên cứu tiếp theo.

Chương 1

TỔNG QUAN VỀ VIỆC ĐẢM BẢO TÍNH CÔNG BẰNG CHO CÁC HỆ THỐNG TRÍ TUỆ NHÂN TẠO/HỌC MÁY TRONG LĨNH VỰC GIÁO DỤC

1.1 Trí tuệ nhân tạo, học máy và vai trò trong giáo dục

Phần này trình bày vai trò của trí tuệ nhân tạo và học máy trong giáo dục hiện đại, tập trung vào các khái niệm nền tảng, cơ chế hoạt động, ứng dụng tiêu biểu và các tiêu chí đánh giá hiệu suất, làm cơ sở cho việc phân tích hiệu quả và công bằng của các hệ thống giáo dục thông minh.

1.1.1 Các khái niệm

Phần này giới thiệu các khái niệm cơ bản của trí tuệ nhân tạo và học máy, bao gồm vai trò của dữ liệu, các nhóm thuật toán chính và cơ chế đánh giá mô hình, nhằm xây dựng nền tảng lý thuyết cho các nghiên cứu đảm bảo công bằng trong học máy.

1.1.2 Đánh giá hiệu suất của các mô hình AI/ML

Phần này trình bày các chỉ số đánh giá hiệu suất phổ biến như *"độ chuẩn xác"*, *"độ chính xác"*, *"độ hồi tưởng"* và *"điểm số F1"*, đồng thời nhấn mạnh sự cần thiết của việc sử dụng kết hợp nhiều chỉ số, đặc biệt trong bối cảnh dữ liệu giáo dục mất cân bằng.

1.2 Công bằng trong các hệ thống học máy

1.2.1 Khái niệm về công bằng trong các hệ thống học máy

Phần này trình bày các khái niệm nền tảng về công bằng trong học máy, bao gồm công bằng cá nhân, công bằng nhóm và công bằng nhóm con, cùng các độ đo phổ biến như *"tác động khác biệt"*, *"hiệu số chênh lệch thống kê"*, *"chênh lệch trung bình xác suất"* và *"chênh lệch cơ hội công bằng"*, đồng thời nhấn mạnh việc lựa chọn định nghĩa công bằng cần gắn với bối cảnh và mục tiêu ứng dụng.

1.2.2 Khái niệm thiên vị trong AI/ML

Phần này làm rõ thiên vị như nguồn gốc chính gây bất công trong các hệ thống AI/ML, cho thấy vai trò của việc nhận diện và giảm thiểu thiên vị trong quá trình thiết kế các mô hình công bằng và có trách nhiệm.

1.2.3 Thuộc tính nhạy cảm trong nghiên cứu về công bằng

Phần này giới thiệu khái niệm thuộc tính nhạy cảm và vai trò của chúng trong nghiên cứu công bằng, nhấn mạnh sự cần thiết phải xử lý đồng thời nhiều thuộc tính nhạy cảm để nhận diện và giảm thiểu thiên vị giao thoa trong các bài toán giáo dục.

1.2.4 Độ đo công bằng trong các hệ thống học máy

Phần này giới thiệu các độ đo công bằng theo nhóm nhạy cảm như *"tác động khác biệt"*, *"hiệu số chênh lệch thống kê"*, *"chênh lệch trung bình xác suất"*, *"chênh lệch cơ hội công bằng"* và các biến thể dựa trên ROC, đồng thời nhấn mạnh rằng việc lựa chọn độ đo cần phù hợp với ngữ cảnh ứng dụng và mục tiêu công bằng.

1.2.5 Các hướng tiếp cận đảm bảo tính công bằng cho các hệ thống học máy

Phần này trình bày ba hướng tiếp cận chính gồm tiền xử lý dữ liệu, can thiệp trong quá trình học và hậu xử lý đầu ra, tạo thành khung tham chiếu phổ biến cho thiết kế và đánh giá công bằng trong AI/ML.

1.2.6 Những thách thức trong việc đảm bảo tính công bằng trong các hệ thống học máy

Phần này chỉ ra các thách thức cốt lõi gồm khó khăn trong định nghĩa và đo lường công bằng, hạn chế đánh giá theo nhiều thuộc tính nhạy cảm, mối quan hệ đánh đổi giữa công bằng và hiệu suất, và ảnh hưởng của yếu tố con người trong triển khai hệ thống.

1.3 Sinh dữ liệu tổng hợp trong đảm bảo tính công bằng cho các hệ thống học máy

1.3.1 Các khái niệm

Dữ liệu tổng hợp là dữ liệu nhân tạo mô phỏng phân bố thống kê của dữ liệu thực, được sử dụng để mở rộng tập huấn luyện, bảo vệ quyền riêng tư và tăng tính đại diện cho các nhóm nhạy cảm nhằm hỗ trợ đánh giá và cải thiện công bằng.

1.3.2 Kỹ thuật sinh dữ liệu tổng hợp

Các kỹ thuật sinh dữ liệu tổng hợp gồm phương pháp thống kê và phương pháp học sâu như GAN (CTGAN, PATEGAN) và LLM, trong đó GAN và LLM đặc biệt phù hợp để sinh dữ liệu bảng phức tạp và cân bằng nhóm trong bối cảnh giáo dục.

1.3.3 Thách thức và cơ hội của dữ liệu tổng hợp trong việc đảm bảo tính công bằng

Dữ liệu tổng hợp mang lại cơ hội cải thiện công bằng và đa dạng dữ liệu, nhưng cũng đối mặt với nguy cơ khuếch đại thiên lệch nếu thiếu thiết kế có chủ đích và cơ chế kiểm soát công bằng phù hợp.

1.4 Tổng quan nghiên cứu về công bằng cho các hệ thống học máy ứng dụng trong lĩnh vực giáo dục

1.4.1 Những thuật toán học máy phổ biến được sử dụng trong bối cảnh giáo dục

Các nghiên cứu cho thấy các thuật toán truyền thống, dễ diễn giải như *Hồi quy logistic*, *Rừng ngẫu nhiên* và *Cây quyết định* được sử dụng phổ biến nhất trong giáo dục, trong khi học sâu chưa thể hiện ưu thế rõ rệt trên dữ liệu bảng.

1.4.2 Những vấn đề phổ biến được đề cập khi nghiên cứu AI/ML trong giáo dục

Nghiên cứu AI/ML trong giáo dục chủ yếu tập trung vào dự đoán kết quả học tập, đánh giá công bằng, chấm điểm tự động và hỗ trợ ra quyết định, với trọng tâm ngày càng tăng vào đánh đổi giữa công bằng và hiệu suất.

1.4.3 Định nghĩa về công bằng và thiên vị trong các hệ thống học máy trong giáo dục

Công bằng nhóm được sử dụng phổ biến hơn công bằng cá nhân trong giáo dục, phản ánh đặc trưng dữ liệu mất cân bằng và nhu cầu xử lý bất bình đẳng ở cấp độ nhóm.

1.4.4 Đặc điểm chính của các bộ dữ liệu dùng trong nghiên cứu AI/ML trong giáo dục

Phần lớn nghiên cứu sử dụng dữ liệu đóng do yêu cầu bảo mật, với các thuộc tính nhạy cảm phổ biến gồm *giới tính*, *chủng tộc* và *tuổi*, gây hạn chế cho khả năng tái lập.

1.4.5 Các phương pháp đảm bảo tính công bằng cho hệ thống học máy trong giáo dục

Các phương pháp đảm bảo công bằng trải rộng từ can thiệp dữ liệu đến huấn luyện và đánh giá mô hình, trong đó việc lựa chọn phụ thuộc mạnh vào dữ liệu và mục tiêu công bằng.

1.4.6 Những độ đo công bằng được sử dụng phổ biến trong giáo dục

Các độ đo công bằng phổ biến trong giáo dục gồm ABROCA, *"tác động khác biệt"*, *"chênh lệch trung bình xác suất"*, *"chênh lệch cơ hội công bằng"* và *"hiệu số chênh lệch thống kê"*, cho phép cải thiện công bằng mà không nhất thiết làm suy giảm mạnh hiệu suất.

1.4.7 Những phương pháp phổ biến nhằm đánh giá công bằng và hiệu suất của các mô hình AI/ML

Công bằng và hiệu suất thường được đánh giá thông qua so sánh trước–sau can thiệp, kiểm định chéo và phân tích theo nhóm con, với nhiều nghiên cứu ghi nhận cải thiện đồng thời cả hai khía cạnh.

1.5 Mối quan hệ giữa công bằng và hiệu suất của các hệ thống học máy/AI

Mối quan hệ giữa công bằng và hiệu suất là một chủ đề trung tâm trong nghiên cứu AI/ML, phản ánh nỗ lực dung hòa giữa hai mục tiêu: đảm

bảo công bằng và duy trì hiệu suất của mô hình. Các nghiên cứu cho thấy mối quan hệ này vừa mang tính đánh đổi, vừa có khả năng đồng tối ưu nhờ các kỹ thuật mới.

1.5.1 Sự đánh đổi giữa công bằng và hiệu suất

Việc tăng công bằng đôi khi làm giảm độ chính xác dự đoán, nhưng không phải luôn đối nghịch. Khi dữ liệu được xử lý để loại bỏ thiên vị, mô hình thường tổng quát hóa tốt và ổn định hơn. Nguyên nhân chính của sự đánh đổi nằm ở sai lệch cấu trúc dữ liệu; các kỹ thuật tiền xử lý như cân bằng phân phối hoặc loại bỏ ảnh hưởng nhân quả có thể đồng thời cải thiện cả công bằng và hiệu suất.

1.5.2 Các hướng tiếp cận để xử lý đánh đổi

Ba hướng tiếp cận chính giúp cân bằng giữa công bằng và hiệu suất gồm: (1) *Giảm thiên vị trong huấn luyện* – tối ưu hiệu suất rồi điều chỉnh công bằng; (2) *Tiền xử lý nhân quả* — loại bỏ ảnh hưởng của thuộc tính nhạy cảm; và (3) *Tối ưu đa mục tiêu*. Dù hiệu quả, các kỹ thuật này cần được điều chỉnh linh hoạt theo từng miền dữ liệu để xây dựng mô hình AI/ML công bằng và tin cậy hơn.

1.6 Tổng kết chương

Chương này trình bày nền tảng lý thuyết về công bằng trong hệ thống học máy ứng dụng giáo dục, gồm bốn nội dung: (i) khái niệm và chỉ số hiệu suất (*"độ chuẩn xác"*, *"độ chính xác"*, *"độ hồi tưởng"*, *"điểm số F1"*); (ii) các định nghĩa công bằng tiêu biểu như *công bằng nhóm*, *công bằng cá nhân*, *công bằng nhân khẩu học*, *cơ hội công bằng*, *công bằng theo xác suất*; (iii) ba nhóm phương pháp đảm bảo công bằng – tiền xử lý, trong huấn luyện và hậu xử lý; (iv) mối quan hệ đánh đổi giữa công bằng và hiệu suất, cùng các hướng cân bằng hợp lý như giảm thiên vị, tiền xử lý nhân quả và tối ưu đa mục tiêu. Nội dung này là cơ sở cho các chương sau, được công bố trong hai công trình khoa học của tác giả (Springer LNISO 2023; JSS, Elsevier 2024).

Chương 2

PHƯƠNG PHÁP ĐẢM BẢO TÍNH CÔNG BẰNG NHỜ LOẠI BỎ SỰ PHỤ THUỘC VÀO CÁC THUỘC TÍNH NHẠY CẢM TRONG BỘ DỮ LIỆU HUẤN LUYỆN

2.1 Giới thiệu

Chương này giải quyết câu hỏi nghiên cứu *RQ1* thông qua việc đề xuất phương pháp *FairEdu*, một kỹ thuật tiền xử lý dựa trên hồi quy đa biến nhằm loại bỏ đồng thời sự phụ thuộc giữa đặc trưng đầu vào và nhiều thuộc tính nhạy cảm trong các hệ thống học máy giáo dục.

Các nghiên cứu tiêu biểu gồm: (i) LTDD của Li và cộng sự (2022), sử dụng hồi quy tuyến tính để loại bỏ phụ thuộc giữa đặc trưng và thuộc tính nhạy cảm nhưng chỉ xử lý từng thuộc tính riêng lẻ; và (ii) công bằng giao thoa của Chen và cộng sự (2024), đánh giá trên các nhóm kết hợp của nhiều thuộc tính nhưng gặp hạn chế về bùng nổ tổ hợp và dữ liệu thưa, mất cân bằng.

Để khắc phục các hạn chế này, luận án đề xuất khái niệm *công bằng đồng thời* và phương pháp *Fairedu*, mở rộng LTDD bằng hồi quy đa biến nhằm loại bỏ đồng thời ảnh hưởng của nhiều thuộc tính nhạy cảm, qua đó cải thiện công bằng trong khi vẫn duy trì hiệu suất mô hình.

2.2 Phương pháp FairEdu

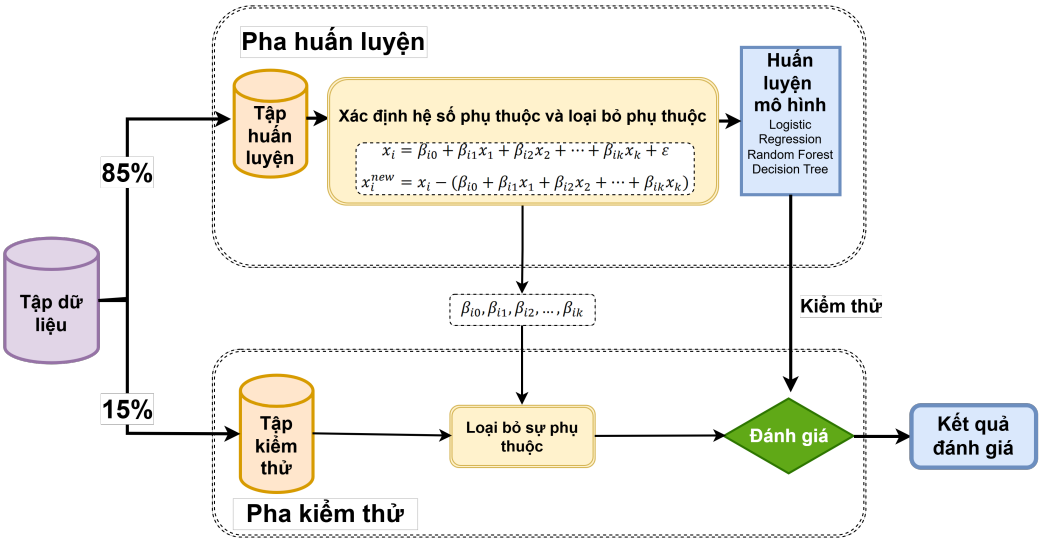
2.2.1 Nguyên lý hoạt động

FairEdu loại bỏ sự phụ thuộc tuyến tính giữa các đặc trưng đầu vào và nhiều thuộc tính nhạy cảm đồng thời thông qua hồi quy đa biến, qua đó

mở rộng LTDD vốn chỉ xử lý từng thuộc tính nhạy cảm riêng lẻ và phù hợp hơn với các bối cảnh công bằng giao thoa trong giáo dục.

2.2.2 Kiến trúc tổng thể

Kiến trúc FairEdu gồm năm bước chính (Hình 2.1): xác định thuộc tính nhạy cảm, xây dựng hồi quy đa biến, loại bỏ thành phần phụ thuộc để tạo tập dữ liệu D', huấn luyện mô hình trên D', và đánh giá công bằng cùng hiệu suất.



Hình 2.1 – Kiến trúc tổng thể của phương pháp FairEdu.

2.2.3 Thuật toán

Thuật toán 2.1 Thuật toán FairEdu (rút gọn)

Require: D_{tr}, D_{test}
Ensure: S_{ML}, $S_{ML}(x^{te})$

for $i = k+1$ to d **do**
  $x_i \leftarrow a_i + b_i^1 x_1 + \cdots + b_i^k x_k + \mu$
  if p-value < 0.05 **then**
    $x_i \leftarrow x_i - (a_i + b_i^1 x_1 + \cdots + b_i^k x_k)$
  end if
end for
Huấn luyện S_{ML} trên D_{tr} đã điều chỉnh; xử lý tương tự D_{test} rồi dự đoán.

2.2.4 Ví dụ minh họa

Phần này minh họa cách áp dụng phương pháp FairEdu cho bài toán dự đoán điểm tốt nghiệp (GPA) của sinh viên Khoa CNTT – Trường Đại học Đại Nam.

2.3 Thực nghiệm

Phần này trình bày thiết lập thực nghiệm nhằm đánh giá hiệu quả của FairEdu trong việc cải thiện tính công bằng cho các hệ thống học máy giáo dục. Thực nghiệm phân tích thiên lệch dữ liệu, khả năng xử lý đa thuộc tính nhạy cảm và mối quan hệ đánh đổi giữa công bằng và hiệu suất.

2.3.1 Dữ liệu

Dữ liệu thực nghiệm gồm ba bộ dữ liệu giáo dục từ Kaggle [1] là *Oulad*, *Student Performance*, *Student Predict Dropout* và một bộ dữ liệu thực tế từ Khoa Công nghệ Thông tin, Trường Đại học Đại Nam (*DNU Data*) [2]. Các thuộc tính số được chuẩn hóa bằng min–max, còn các thuộc tính nhạy cảm được mã hóa nhị phân trước khi huấn luyện.

2.3.2 Lựa chọn mô hình học máy

Các mô hình được sử dụng gồm *Hồi quy logistic*, *Cây quyết định*, *Rừng ngẫu nhiên*, *Tăng cường gradient*, và *Mạng nơ ron thần kinh*.

2.3.3 Thiết lập thực nghiệm

Các thí nghiệm được thiết lập với dữ liệu chuẩn hóa (Min–Max Scaling) và mã hóa nhị phân cho các thuộc tính nhạy cảm, chia thành 85% huấn luyện và 15% kiểm tra. Phương pháp FairEdu được áp dụng để loại bỏ phụ thuộc trong dữ liệu huấn luyện, sau đó đánh giá mô hình trên dữ liệu kiểm tra đã điều chỉnh, lặp lại 100 lần để đảm bảo độ tin cậy kết quả.

1. Kaggle.com
2. https://dainam.edu.vn/en

2.3.4 Chỉ số đánh giá

Các thước đo công bằng gồm *"tác động khác biệt"* (DI), *"hiệu số chênh lệch thống kê"* (SPD), *"chênh lệch trung bình xác suất"* (AOD), *"chênh lệch cơ hội công bằng"* (EOD); các chỉ số hiệu suất gồm *"độ chuẩn xác"* (ACC), *"độ hồi tưởng"* (Recall), *"độ chính xác"* (Precision), *"điểm số F1"* (F1-score).

2.4 Kết quả thực nghiệm

2.4.1 Thiên vị hệ thống trong dữ liệu giáo dục

Kết quả phân tích theo các chỉ số công bằng ($| 1 - DI |$, SPD, AOD, EOD) cho thấy các bộ dữ liệu giáo dục đều tồn tại thiên vị với mức độ khác nhau giữa các thuộc tính nhạy cảm. Trong đó, *giới tính* có mức biến thiên lớn, phản ánh ảnh hưởng không ổn định đến kết quả mô hình; *tình trạng nợ* ghi nhận mức thiên vị cao nhất; trong khi *tuổi*, *tình trạng khuyết tật*, *khu vực* và *sức khỏe* có mức thiên vị thấp hơn và ổn định hơn. Nhìn chung, không có thuộc tính nào hoàn toàn không thiên vị, cho thấy công bằng cần được đánh giá đồng thời trên nhiều thuộc tính thay vì chỉ tập trung vào một yếu tố riêng lẻ.

2.4.2 Ảnh hưởng của mô hình học máy đến mức độ công bằng

Kết quả thực nghiệm cho thấy mức độ công bằng phụ thuộc đáng kể vào mô hình học máy, ngay cả khi sử dụng cùng dữ liệu và thuộc tính nhạy cảm. Các mô hình phi tuyến như *Rừng ngẫu nhiên* và *Cây quyết định* thường tạo ra mức độ thiên lệch cao hơn trong một số trường hợp (đặc biệt với *tình trạng nợ*), trong khi *Hồi quy logistic* thể hiện công bằng tốt hơn ở một số thuộc tính khác (như *tình trạng khuyết tật*). Nhìn chung, không có mô hình nào luôn đảm bảo công bằng tối ưu trên mọi thuộc tính, cho thấy sự cần thiết của các phương pháp can thiệp nhằm kiểm soát thiên lệch một cách hệ thống.

2.4.3 Khả năng xử lý đồng thời nhiều thuộc tính nhạy cảm của FairEdu

Kết quả so sánh với các phương pháp như Reweighing, FairSMOTE và LTDD trên nhiều mô hình và bộ dữ liệu cho thấy Fairedu cải thiện vượt trội các chỉ số công bằng (DI, SPD, AOD, EOD) trong phần lớn các trường hợp. So sánh chi tiết với LTDD (trên mô hình *Hồi quy logistic*, chạy lặp 100 lần) cho thấy các kết quả đều có ý nghĩa thống kê ($p < 0.05$). Cụ thể, với hai chỉ số $|\ 1 - DI\ |$ và SPD, Fairedu đạt tỷ lệ Thắng/Hòa/Thua so với mô hình gốc lần lượt là 8/0/1 và 9/0/0; so với LTDD là 5/0/4 và 6/2/1 trên tổng số 9 phép so sánh. Nhìn chung, Fairedu thể hiện ưu thế rõ rệt trong việc cải thiện công bằng một cách ổn định và có khả năng tổng quát cao khi xử lý đồng thời nhiều thuộc tính nhạy cảm.

2.4.4 Mối quan hệ giữa công bằng và hiệu suất dự đoán

Kết quả thực nghiệm cho thấy Fairedu duy trì hiệu suất ổn định trong khi cải thiện đáng kể công bằng. Cụ thể, với ACC, Fairedu tốt hơn trong 11/27 trường hợp và mức suy giảm tối đa chỉ 5.71%; với Recall, vượt Origin 6 lần và LTDD 12 lần, mức giảm lớn nhất khoảng 8%. Đối với F1-score, Fairedu cải thiện trong 6–9 trường hợp, trong khi Precision thể hiện rõ ưu thế với 17 lần vượt Origin và 19 lần vượt LTDD; điển hình tăng từ 0.929 lên 0.962 ($\approx 3.5\%$) trên mô hình RF với dữ liệu DNU-BP. Nhìn chung, Fairedu đạt được sự cân bằng tốt giữa công bằng và hiệu suất, với mức đánh đổi nhỏ và có thể kiểm soát.

2.5 Thảo luận

2.5.1 Thảo luận các phát hiện chính từ kết quả thực nghiệm

Kết quả thực nghiệm cho thấy thiên vị tồn tại trên tất cả các thuộc tính nhạy cảm và không có thuộc tính nào luôn công bằng, do đó cần đánh giá đồng thời nhiều thuộc tính. Mức độ công bằng phụ thuộc mạnh vào mô hình, trong đó *Hồi quy logistic* thường công bằng hơn *Rừng ngẫu nhiên* và *Cây quyết định*. Fairedu chứng minh hiệu quả trong việc xử lý đồng thời

nhiều thuộc tính nhạy cảm thông qua loại bỏ phụ thuộc dữ liệu, đồng thời duy trì hiệu suất với mức suy giảm nhỏ.

2.5.2 Hạn chế của nghiên cứu

Fairedu dựa trên hồi quy tuyến tính nên còn hạn chế trong việc mô hình hóa quan hệ phi tuyến và xử lý dữ liệu lớn hoặc dữ liệu phi số. Kết quả thực nghiệm chủ yếu trên dữ liệu giáo dục và các mô hình truyền thống nên tính khái quát còn hạn chế. Ngoài ra, việc lựa chọn thước đo công bằng và sự đánh đổi công bằng–hiệu suất có thể ảnh hưởng đến kết luận nghiên cứu.

2.5.3 Ý nghĩa và ứng dụng thực tiễn

Fairedu có thể được áp dụng như một bước tiền xử lý để giảm thiên vị trong các hệ thống tuyển sinh, đánh giá học tập và dự đoán nguy cơ bỏ học. Phương pháp góp phần nâng cao tính minh bạch, công bằng và độ tin cậy của hệ thống AI trong giáo dục, đồng thời hỗ trợ các tổ chức giáo dục thực hiện các mục tiêu về bình đẳng và hòa nhập.

2.6 Tổng kết chương

Chương này đề xuất phương pháp *Fairedu* – kỹ thuật tiền xử lý dựa trên hồi quy đa biến nhằm xử lý đồng thời nhiều thuộc tính nhạy cảm, cải thiện công bằng và hạn chế đánh đổi với hiệu suất.

Kết quả thực nghiệm trên nhiều bộ dữ liệu và mô hình cho thấy Fairedu giúp giảm đáng kể các chỉ số thiên lệch (DI, SPD, AOD, EOD) trong khi vẫn duy trì hiệu suất ổn định, thậm chí cải thiện trong một số trường hợp.

Phương pháp cũng thể hiện tiềm năng ứng dụng trong các bài toán giáo dục như tuyển sinh, đánh giá học tập và dự đoán rủi ro, đồng thời có thể mở rộng sang các lĩnh vực khác. Kết quả nghiên cứu đã được công bố trên tạp chí *Expert Systems with Applications* (Q1, 2025).

Chương 3

PHƯƠNG PHÁP ĐẢM BẢO TÍNH CÔNG BẰNG NHỜ CÂN BẰNG DỮ LIỆU DỰA TRÊN KỸ THUẬT SINH DỮ LIỆU TỔNG HỢP

3.1 Giới thiệu

Trong giáo dục, đảm bảo công bằng cho các hệ thống học máy là cần thiết nhưng gặp thách thức do dữ liệu mất cân bằng và tồn tại nhiều thuộc tính nhạy cảm.

Các nghiên cứu liên quan tập trung vào hai hướng chính: (i) phương pháp Reweighing điều chỉnh trọng số để cân bằng phân phối giữa các nhóm, nhưng không giải quyết được vấn đề thiếu dữ liệu; và (ii) các phương pháp sinh dữ liệu tổng hợp như TabFairGAN, giúp tăng đại diện cho nhóm thiểu số nhưng còn hạn chế trong kiểm soát phân phối giữa các nhóm giao thoa khi số thuộc tính tăng.

Để khắc phục, nghiên cứu đề xuất *DPF*, kết hợp phân chia dữ liệu theo nhóm giao thoa và sinh dữ liệu tổng hợp nhằm tái cân bằng dữ liệu một cách hiệu quả và có kiểm soát.

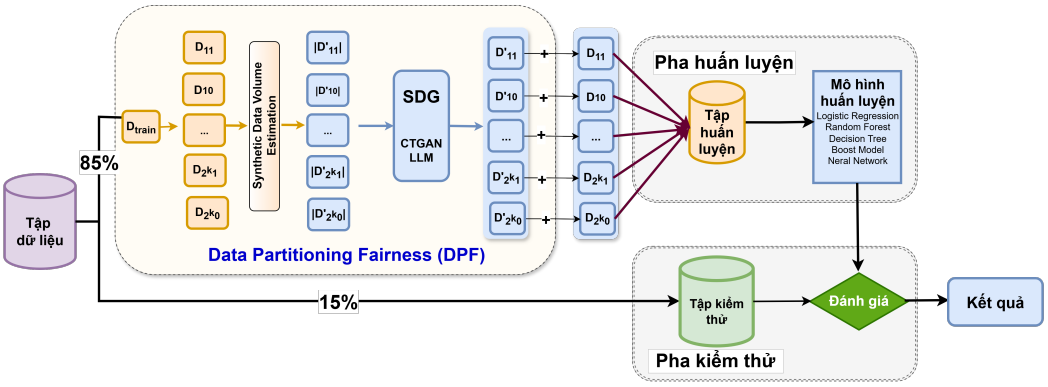
3.2 Phương pháp DPF

3.2.1 Nguyên lý hoạt động

DPF chia dữ liệu thành 2^{k+1} nhóm con theo tổ hợp các thuộc tính nhạy cảm và nhãn, sau đó áp dụng sinh dữ liệu tổng hợp độc lập trên từng nhóm để đạt tỷ lệ phân bố mong muốn. Quá trình sinh dữ liệu chỉ thực hiện với các nhóm có đủ mẫu (≥ 2) nhằm đảm bảo độ tin cậy và tránh suy diễn quá mức. Mục tiêu là cân bằng tỷ lệ giữa các nhóm nhạy cảm và nhãn, từ đó giảm thiên lệch và giúp mô hình học máy học được phân phối dữ liệu công bằng hơn.

3.2.2 Kiến trúc tổng thể

Kiến trúc tổng thể của phương pháp DPF được minh họa trong Hình 3.1. Toàn bộ quy trình thực hiện DPF được tổ chức thành bốn bước chính: chuẩn bị dữ liệu, cân bằng dữ liệu, huấn luyện mô hình và kiểm tra, đánh giá.



Hình 3.1 – Kiến trúc tổng thể của phương pháp DPF

3.2.3 Thuật toán DPF

Thuật toán 3.1 Thuật toán DPF — (rút gọn)

Require: Tập D_{tr} với k thuộc tính nhạy cảm $(x_1, \ldots, x_k)$ và nhãn $y \in \{0, 1\}$
Ensure: Tập cân bằng D'_{tr}, mô hình S_{ML}, dự đoán $S_{ML}(x^{te})$

1: Chia D_{tr} thành 2^{k+1} nhóm D_{ij} theo tổ hợp $(x_1, \ldots, x_k, y)$
2: Loại bỏ các thuộc tính nhạy cảm và sinh thêm dữ liệu bằng SDG (CTGAN hoặc LLM)
3: Hợp nhất các D'_{ij} thành D'_{tr}, huấn luyện S_{ML}, và dự đoán trên D_{test}
return D'_{tr}, S_{ML}, $S_{ML}(x^{te})$

3.2.4 Ví dụ minh họa

Ví dụ trên dữ liệu DNU.

3.3 Thực nghiệm

3.3.1 Dữ liệu

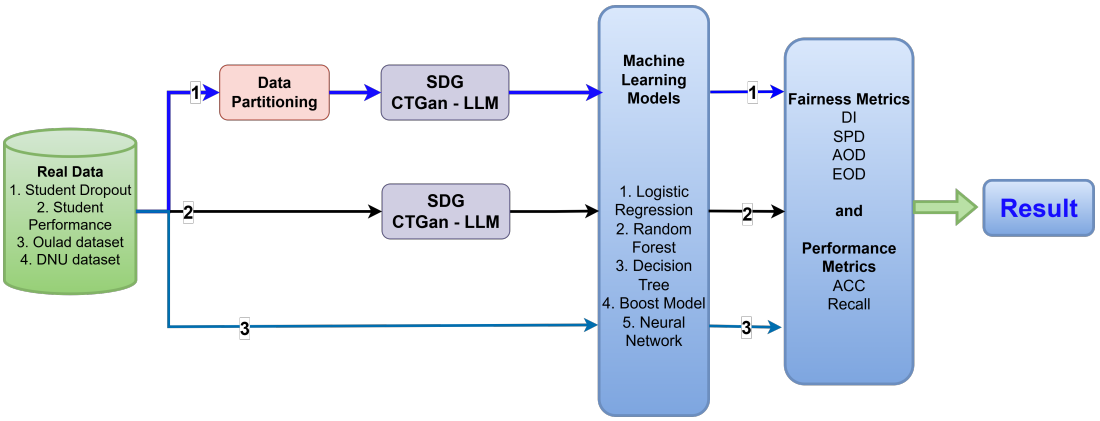
Bốn bộ dữ liệu sử dụng trong nghiên cứu này đã được trình bày chi tiết trong Mục 2.3.1, bao gồm: *Student Performance*, *Student Predict Dropout*, *Oulad*, và *DNU Data*.

3.3.2 Mô hình và kỹ thuật sinh dữ liệu

Năm mô hình học máy *Hồi quy logistic*, *Rừng ngẫu nhiên*, *Cây quyết định*, *Tăng cường gradient*, *Mạng nơ ron thần kinh* và hai kỹ thuật sinh dữ liệu (*CTGAN*, *LLM*) được sử dụng để cân bằng dữ liệu.

3.3.3 Thiết lập thực nghiệm

Nghiên cứu xây dựng ba cấu hình thực nghiệm, từ không can thiệp đến can thiệp bằng phân vùng dữ liệu và sinh dữ liệu tổng hợp theo nhóm con, nhằm đánh giá hiệu quả của phương pháp trong việc cải thiện công bằng cho các mô hình học máy (xem Hình 3.2.



Hình 3.2 – Tổng quan ba cấu hình thực nghiệm đánh giá hiệu quả DPF

3.3.4 Chỉ số đánh giá

Sáu chỉ số được sử dụng: bốn chỉ số công bằng DI, SPD, AOD, EOD và hai chỉ số hiệu năng ACC và *recall* để đánh giá toàn diện hiệu quả và tính công bằng của DPF.

3.4 Kết quả thực nghiệm

3.4.1 Đặc điểm của dữ liệu tổng hợp trong lĩnh vực giáo dục

Kết quả cho thấy dữ liệu tổng hợp sinh bằng CTGAN và LLM giúp cải thiện đáng kể sự cân bằng giữa các nhóm giao thoa, đặc biệt trên các bộ dữ liệu mất cân bằng như *Student Predict Dropout* và *DNU Data*. Phương pháp DPF phân tách dữ liệu theo nhóm nhạy cảm và sinh dữ liệu có kiểm soát, giúp tỷ lệ giữa các nhóm trở nên tương đồng hơn.

3.4.2 Hiệu quả của phương pháp DPF trong cải thiện công bằng giao thoa

3.4.2.1 Đánh giá thông qua các chỉ số công bằng

Kết quả so sánh cho thấy sự khác biệt rõ rệt giữa các cấu hình và phụ thuộc vào kỹ thuật sinh dữ liệu. Với CTGAN, SDG dẫn đầu ở các chỉ số phân phối ($|\ 1 - DI\ |$: 23 lần, SPD: 21 lần), trong khi DPF vượt trội ở các chỉ số sai số (AOD: 22 lần, EOD: 17 lần), cho thấy hai phương pháp tối ưu theo các khía cạnh khác nhau. Với LLM, DPF thể hiện ưu thế toàn diện khi dẫn đầu ở cả bốn chỉ số ($|\ 1 - DI\ |$: 19, SPD: 19, AOD: 14, EOD: 16), cho thấy khả năng cải thiện công bằng ổn định và cân bằng hơn.

3.4.2.2 Ảnh hưởng của mô hình học máy

Kết quả cho thấy hiệu quả công bằng phụ thuộc vào cả mô hình và phương pháp sinh dữ liệu. Với LLM, *Hồi quy logistic* và *Mạng nơ ron thần kinh* đạt kết quả đồng đều và cải thiện rõ rệt trên nhiều chỉ số, trong khi

Rừng ngẫu nhiên và *Cây quyết định* có sự thay đổi theo từng chỉ số cụ thể. CTGAN giúp *Rừng ngẫu nhiên* và *Tăng cường gradient* đạt kết quả ổn định hơn, đặc biệt ở các chỉ số sai số (AOD, EOD), trong khi LLM cải thiện tốt hơn các chỉ số phân phối (DI, SPD) ở một số mô hình. Nhìn chung, không có sự kết hợp nào tối ưu cho mọi trường hợp, cho thấy cần lựa chọn phương pháp sinh dữ liệu phù hợp với từng mô hình và mục tiêu công bằng cụ thể.

3.4.2.3 Ảnh hưởng của đặc trưng bộ dữ liệu

Hiệu quả của DPF phụ thuộc rõ rệt vào đặc điểm của bộ dữ liệu. Với các bộ dữ liệu tương đối cân bằng như *Oulad* và *Student Performance*, DPF giúp duy trì mức công bằng tốt với các chỉ số thấp trên hầu hết mô hình. Ngược lại, với các bộ dữ liệu mất cân bằng như *Student Predict Dropout* và *DNU Data*, mức độ công bằng giảm đáng kể, đặc biệt ở một số thuộc tính nhạy cảm như *tình trạng nợ*, với giá trị $|\ 1 - DI\ |$ tăng cao. Kết quả cho thấy hiệu quả của DPF không chỉ phụ thuộc vào phương pháp mà còn chịu ảnh hưởng mạnh từ phân bố dữ liệu, mô hình học máy và từng thuộc tính nhạy cảm cụ thể.

3.4.3 Tác động của DPF đến hiệu suất dự đoán của mô hình

Kết quả cho thấy Origin đạt ưu thế về ACC (26 lần thắng), trong khi SDG vượt trội về Recall (22 lần), phản ánh khả năng đánh đổi giữa độ chính xác và khả năng phát hiện. DPF thể hiện vai trò cân bằng khi đạt 13 lần thắng về ACC và 11 lần về Recall, duy trì hiệu suất cạnh tranh mà không suy giảm đáng kể. Nhìn chung, DPF là giải pháp hiệu quả giúp cân bằng giữa công bằng và hiệu suất trong các hệ thống học máy.

3.5 Thảo luận

3.5.1 Phân tích và tổng hợp các phát hiện chính

DPF cải thiện công bằng giao thoa rõ rệt, và đạt hiệu quả ổn định hơn khi kết hợp với LLM với mức đánh đổi hiệu suất nhỏ.

3.5.2 Hạn chế của nghiên cứu

Hiệu quả của DPF phụ thuộc vào chất lượng dữ liệu tổng hợp và độ đo công bằng, chưa được kiểm chứng trên dữ liệu lớn hoặc phi cấu trúc.

3.5.3 Ý nghĩa và ứng dụng thực tiễn

DPF là kỹ thuật tiền xử lý khả thi cho các hệ thống AI/ML giáo dục và có tiềm năng mở rộng sang các lĩnh vực nhạy cảm khác.

3.6 Tóm tắt chương

Chương này trình bày phương pháp DPF – kỹ thuật cân bằng dữ liệu bằng sinh dữ liệu tổng hợp nhằm cải thiện công bằng giao thoa cho mô hình học máy giáo dục mà vẫn duy trì hiệu suất. Kết quả đã được tổng hợp trong bài báo gửi tạp chí *Information and Software Technology (2025)*.

Chương 4

PHƯƠNG PHÁP ĐẢM BẢO TÍNH CÔNG BẰNG HAI CHIỀU CHO DỮ LIỆU DẠNG BẢNG – CHỈ SỐ ĐÁNH ĐỔI GIỮA CÔNG BẰNG VÀ HIỆU SUẤT

4.1 Giới thiệu

Chương này đề xuất *FairEduPlus*, một khuôn khổ tiền xử lý công bằng hai chiều nhằm cải thiện công bằng và kiểm soát đánh đổi với hiệu suất trong dữ liệu giáo dục.

4.2 Phương pháp FairEduPlus

4.2.1 Ý tưởng chính

FairEduPlus kết hợp cân bằng dữ liệu theo nhóm giao thoa (DPF) và loại bỏ phụ thuộc đặc trưng–thuộc tính nhạy cảm (FairEdu) để đảm bảo công bằng hai chiều.

4.2.2 Kiến trúc tổng thể

FairEduPlus gồm ba bước: cân bằng dữ liệu bằng DPF, hiệu chỉnh đặc trưng bằng FairEdu, và huấn luyện–đánh giá mô hình theo các chỉ số công bằng và hiệu suất.

4.2.3 Thuật toán FairEduPlus

FairEduPlus thực hiện cân bằng dữ liệu theo nhóm giao thoa bằng DPF, sau đó loại bỏ phụ thuộc giữa đặc trưng và thuộc tính nhạy cảm bằng FairEdu, như trình bày trong Thuật toán 4.1.

Thuật toán 4.1 FAIREDUPLUS

1: **Input:** Tập huấn luyện D_{tr}, tập kiểm thử D_{test};
2: **Output:** Tập dữ liệu công bằng D', mô hình S_{ML};
3: Phân D_{tr} thành 2^{k+1} nhóm D_{ij} theo tổ hợp các thuộc tính nhạy cảm và nhãn;
4: Áp dụng FairEdu để loại bỏ phụ thuộc giữa thuộc tính nhạy cảm và không nhạy cảm;
5: Huấn luyện mô hình SML trên $D'tr$;
6: Áp dụng điều chỉnh tương tự cho $Dtest$ và đánh giá theo các chỉ số công bằng và hiệu suất; $D'tr$, SML.

4.3 Thực nghiệm

4.3.1 Dữ liệu, mô hình và độ đo

Sử dụng dữ liệu và các kỹ thuật như đã trình bày ở chương 4.

4.3.2 Thiết lập thực nghiệm

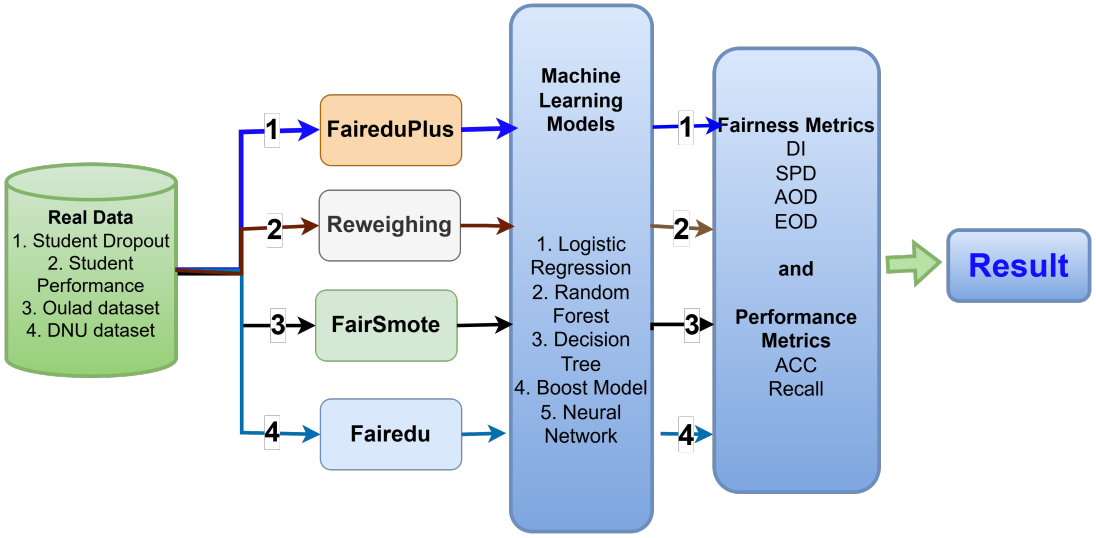
So sánh sáu cấu hình can thiệp công bằng, sử dụng dữ liệu tổng hợp (CTGAN và LLM) để cân bằng các nhóm con và đánh giá công bằng. (Hình 4.1

4.4 Kết quả

4.4.1 Cải thiện công bằng đồng thời của FairEduPlus

FaireduPlus cải thiện công bằng vượt trội, đặc biệt với CTGAN (tỷ lệ thắng $|\,1 - DI\,|$ và SPD lên tới 75.6%–84.4%). Với LLM, phương pháp vẫn duy trì lợi thế nhưng ở mức thấp hơn ($\approx$ 48.9%˘66.7%).

Nhìn chung, FaireduPlus cải thiện công bằng ổn định, hiệu quả nhất khi kết hợp CTGAN.



Hình 4.1 – Tổng quan các cấu hình thực nghiệm đánh giá hiệu quả FaireduPlus

4.4.2 Tác động của FairEduPlus đến hiệu suất dự đoán

FairEduPlus duy trì hiệu suất dự đoán ở mức cạnh tranh và thường cải thiện *"độ hồi tưởng"* dù có thể làm giảm nhẹ *"độ chuẩn xác"*.

4.4.3 Đánh giá mối quan hệ đánh đổi giữa công bằng và hiệu suất

Kết quả cho thấy FairEduPlus đạt sự đánh đổi hợp lý công bằng và hiệu suất.

4.5 Thảo luận

4.5.1 Tổng hợp và diễn giải kết quả theo mục tiêu nghiên cứu

FairEduPlus cải thiện công bằng giao thoa với đánh đổi hiệu suất nhỏ, cụ thể: (1) Cải thiện rõ rệt các chỉ số công bằng, đặc biệt với CTGAN. (2) Duy trì hiệu suất tốt, nổi bật ở Recall. (3) Đạt cân bằng hiệu quả giữa công bằng và hiệu suất.

4.5.2 Hạn chế nghiên cứu

Nghiên cứu có những hạn chế: (1) *Giá trị nội tại:* Kết quả chịu ảnh hưởng bởi tính ngẫu nhiên của sinh dữ liệu và chưa tối ưu toàn diện siêu tham số. (2) *Khả năng khái quát:* Mới đánh giá trên dữ liệu giáo dục, chưa kiểm chứng ở các lĩnh vực khác và các thuộc tính nhạy cảm đa dạng hơn. (3) *Giá trị cấu trúc:* Chỉ sử dụng một số chỉ số công bằng phổ biến, chưa bao quát các khía cạnh công bằng khác. (4) *Độ tin cậy:* Dù có thể tái lập, việc sử dụng LLM có thể gây sai khác nhỏ giữa các lần thực nghiệm.

4.5.3 Hàm ý thực tiễn

(1) FaireduPlus là giải pháp tiền xử lý khả thi, giúp cân bằng giữa công bằng và hiệu suất trong các hệ thống học máy giáo dục. (2) Kết hợp sinh dữ liệu tổng hợp với phân tách nhóm là yếu tố quan trọng để cải thiện công bằng giao thoa. (3) Không có kỹ thuật sinh dữ liệu tối ưu chung; cần lựa chọn phù hợp theo từng bối cảnh (CTGAN ổn định, LLM linh hoạt). (4) Dữ liệu tổng hợp còn hỗ trợ bảo vệ quyền riêng tư và tăng tính minh bạch, phù hợp với yêu cầu đạo đức và pháp lý.

4.6 Tóm tắt chương

Chương này đề xuất *FaireduPlus* – khuôn khổ tiền xử lý kết hợp cân bằng dữ liệu (DPF) và hiệu chỉnh đặc trưng, nhằm xử lý đồng thời thiên lệch về phân bố và cấu trúc dữ liệu.

Kết quả thực nghiệm cho thấy FaireduPlus cải thiện công bằng một cách ổn định, đồng thời duy trì hiệu suất và đạt cân bằng tốt hơn so với các phương pháp đơn lẻ. Hiệu quả phương pháp phụ thuộc vào dữ liệu và kỹ thuật sinh dữ liệu (CTGAN ổn định, LLM linh hoạt).

Nhìn chung, FaireduPlus là phương pháp khả thi và có khả năng mở rộng cho các bài toán học máy trong giáo dục, và đã được báo cáo trong một bài báo đang phản biện tại *Information and Software Technology*.

KẾT LUẬN

Luận án tập trung nghiên cứu và phát triển các phương pháp nhằm đảm bảo tính công bằng trong các mô hình học máy ứng dụng cho lĩnh vực giáo dục. Toàn bộ nội dung nghiên cứu được triển khai xuyên suốt bốn chương chính, tương ứng với bốn hướng đóng góp khoa học có mối liên hệ chặt chẽ và kế thừa lẫn nhau.

Chương **??** trình bày tổng quan toàn diện về vấn đề công bằng trong các hệ thống trí tuệ nhân tạo và học máy, với trọng tâm là bối cảnh giáo dục. Nội dung chương hệ thống hóa khung lý thuyết nền tảng, làm rõ các khái niệm cốt lõi liên quan đến công bằng và thiên lệch, phân loại các hướng tiếp cận đảm bảo công bằng, đồng thời phân tích những thách thức và khoảng trống nghiên cứu còn tồn tại. Các phân tích này đóng vai trò định hướng quan trọng cho việc xác lập các mục tiêu và phương pháp nghiên cứu của luận án.

Trên cơ sở đó, Chương **??** đề xuất phương pháp *Fairedu*, một kỹ thuật tiền xử lý dựa trên hồi quy đa biến nhằm loại bỏ sự phụ thuộc giữa các thuộc tính nhạy cảm và các thuộc tính đầu vào trong dữ liệu huấn luyện. Phương pháp này cho phép xử lý đồng thời nhiều thuộc tính nhạy cảm và đã được kiểm chứng thực nghiệm trên các bộ dữ liệu giáo dục thực tế. Kết quả cho thấy *Fairedu* giúp cải thiện đáng kể các chỉ số công bằng, đồng thời vẫn duy trì hiệu suất dự đoán của mô hình ở mức chấp nhận được.

Chương 3 tập trung khai thác các kỹ thuật sinh dữ liệu tổng hợp để xử lý vấn đề mất cân bằng phân phối trong dữ liệu, từ đó đề xuất phương pháp *DPF*. Phương pháp này hướng tới cân bằng dữ liệu giữa các nhóm giao thoa của nhiều thuộc tính nhạy cảm, qua đó giảm thiểu thiên lệch phát sinh do thiếu tính đại diện. Các kết quả thực nghiệm cho thấy *DPF* đạt được sự cải thiện công bằng rõ rệt so với các cấu hình tham chiếu, đồng thời thể hiện khả năng duy trì hiệu suất và tính mở rộng trong các kịch bản dữ liệu khác nhau.

Trên nền tảng của hai phương pháp trên, Chương **??** đề xuất *FaireduPlus* như một khuôn khổ đảm bảo công bằng hai chiều, kết hợp giữa can thiệp theo chiều dọc (loại bỏ phụ thuộc) và can thiệp theo chiều ngang (cân bằng

phân phối dữ liệu). Phương pháp này cho phép xử lý hiệu quả các tập dữ liệu chứa đồng thời nhiều thuộc tính nhạy cảm và mất cân bằng nghiêm trọng. Thực nghiệm cho thấy *FaireduPlus* không chỉ cải thiện công bằng một cách ổn định mà còn duy trì hiệu suất dự đoán ở mức hợp lý. Bên cạnh đó, chương này còn đề xuất và bước đầu kiểm chứng một *chỉ số đánh đổi tổng hợp* nhằm hỗ trợ định lượng mối quan hệ giữa công bằng và hiệu suất, qua đó hỗ trợ so sánh và lựa chọn mô hình trong các bối cảnh ra quyết định thực tiễn. Mặc dù chỉ số này còn cần tiếp tục được hoàn thiện và đánh giá trên nhiều bối cảnh khác nhau, kết quả ban đầu cho thấy tiềm năng hỗ trợ đánh giá sự đánh đổi giữa hai mục tiêu quan trọng của các hệ thống học máy.

Tổng thể, luận án đã đóng góp một chuỗi giải pháp có tính hệ thống, từ phân tích lý thuyết, đề xuất phương pháp đến đánh giá thực nghiệm, nhằm nâng cao tính công bằng cho các mô hình học máy trong lĩnh vực giáo dục. Các kết quả đạt được không chỉ có ý nghĩa về mặt học thuật mà còn mang lại giá trị ứng dụng thực tiễn trong việc phát triển các hệ thống AI đáng tin cậy, có trách nhiệm và hướng tới người học.

Luận án đã mang đến những đóng góp quan trọng cả về lý thuyết và thực tiễn. Điểm cốt lõi xuyên suốt của nghiên cứu là tập trung giải quyết bài toán đảm bảo tính công bằng cho dữ liệu với đồng thời nhiều thuộc tính nhạy cảm – một thách thức chưa được quan tâm đầy đủ trong các nghiên cứu trước đây. Các phương pháp được đề xuất đều hướng đến xử lý vấn đề này theo những cách tiếp cận khác nhau: (i) Fairedu điều chỉnh dữ liệu huấn luyện nhằm loại bỏ sự phụ thuộc giữa thuộc tính nhạy cảm và biến không nhạy cảm, giúp nâng cao tính công bằng ngay cả khi xét nhiều thuộc tính nhạy cảm đồng thời; (ii) DPF khai thác dữ liệu tổng hợp để cân bằng phân phối giữa các nhóm giao nhau, từ đó khắc phục hiện tượng mất cân bằng dữ liệu khi số lượng thuộc tính nhạy cảm tăng lên; (iii) FaireduPlus tích hợp sức mạnh của hai hướng tiếp cận trên, vừa cân bằng dữ liệu bằng kỹ thuật sinh dữ liệu tổng hợp, vừa điều chỉnh sự phụ thuộc, tạo nên một giải pháp toàn diện, dung hòa công bằng và hiệu suất. Bên cạnh đó, luận án còn đóng góp ở khía cạnh phương pháp luận khi đề xuất và bước đầu kiểm chứng một "chỉ số đánh đổi" nhằm hỗ trợ đánh giá sự cân bằng giữa tính công bằng và hiệu suất của mô hình. Chỉ số này là cơ sở ban đầu cho các nghiên cứu tiếp theo hướng tới xây dựng các thước đo đánh đổi ổn định

và toàn diện hơn. Các kết quả thực nghiệm đã chứng minh rằng ba phương pháp này không chỉ cải thiện rõ rệt mức độ công bằng, mà còn duy trì hiệu suất dự đoán ổn định, đồng thời khẳng định tính khả thi và khả năng mở rộng trong thực tiễn giáo dục – lĩnh vực vốn đặc trưng bởi dữ liệu khan hiếm, mất cân bằng và chứa nhiều thiên lệch.

Bên cạnh những kết quả đạt được, luận án vẫn còn tồn tại một số hạn chế nhất định. *Thứ nhất*, các phương pháp đề xuất mới chỉ được kiểm chứng trên dữ liệu dạng bảng và các mô hình học máy truyền thống, trong khi chưa được đánh giá trên dữ liệu phi cấu trúc, dữ liệu quy mô lớn cũng như các mô hình học sâu và kiến trúc Transformer, do đó chưa phản ánh đầy đủ tính tổng quát. *Thứ hai*, hiệu quả của các phương pháp DPF và FaireduPlus còn hạn chế khi xử lý các bộ dữ liệu mất cân bằng nghiêm trọng hoặc tồn tại các nhóm giao thoa có rất ít hoặc không có mẫu, do thiếu dữ liệu đại diện cho những trường hợp hiếm. Nghiên cứu chủ yếu đánh giá dữ liệu tổng hợp thông qua hiệu quả của mô hình học máy và các chỉ số công bằng, chưa đánh giá một cách hệ thống các tiêu chí như fidelity giữa dữ liệu sinh và dữ liệu gốc cũng như khả năng tái lập của các mô hình sinh dữ liệu, đặc biệt đối với các mô hình ngôn ngữ lớn. *Thứ ba*, chỉ số đánh đổi giữa công bằng và hiệu suất mới dừng ở mức đề xuất và thử nghiệm ban đầu, đồng thời có thể kém ổn định trong một số trường hợp do sử dụng mức cải thiện tương đối. Bên cạnh đó, việc đánh giá hiệu quả của các phương pháp chủ yếu dựa trên thống kê thắng/thua, chưa kết hợp đầy đủ các kiểm định thống kê và các chỉ số kích thước ảnh hưởng (effect size) để đánh giá mức độ khác biệt giữa các phương pháp. *Thứ tư*, các phương pháp hiện vẫn đòi hỏi nhiều bước can thiệp thủ công và hiệu chỉnh theo từng bộ dữ liệu cụ thể, chưa hình thành một khung giải pháp tổng quát và tự động để triển khai trên nhiều lĩnh vực ứng dụng.

Những hạn chế nêu trên không chỉ cho thấy giới hạn của phạm vi nghiên cứu hiện tại, mà còn mở ra nhiều hướng đi mới để tiếp tục phát triển và hoàn thiện các phương pháp trong tương lai. Trên cơ sở những kết quả đã đạt được trong việc đánh giá và cải thiện tính công bằng của các hệ thống học máy thông qua các phương pháp tiền xử lý như Fairedu, DPF và FaireduPlus, luận án đồng thời đề xuất một số định hướng nghiên cứu tiềm năng nhằm nâng cao hiệu quả và mở rộng khả năng ứng dụng trong thực tiễn. Các định hướng này có thể được xem xét trong những nghiên cứu tiếp theo, bao gồm:

1. *Mở rộng phạm vi dữ liệu và mô hình áp dụng:* Trong phạm vi hiện tại, các phương pháp Fairedu, DPF, và FaireduPlus chủ yếu được thử nghiệm trên các bộ dữ liệu giáo dục với các mô hình truyền thống như *Hồi quy logistic*, *Cây quyết định* và *Rừng ngẫu nhiên*. Trong tương lai, cần tiến hành đánh giá trên nhiều bộ dữ liệu trong các lĩnh vực khác nhau như y tế, tài chính, bao gồm cả dữ liệu lớn và dữ liệu phi cấu trúc. Đồng thời, việc tích hợp với các mô hình hiện đại như *Học sâu* hoặc *Transformer-based* sẽ giúp kiểm chứng tính hiệu quả trong những kiến trúc phức tạp hơn, từ đó đánh giá mức độ tổng quát và khả năng mở rộng của phương pháp. Bên cạnh đó, cần nghiên cứu ảnh hưởng của hiện tượng đa cộng tuyến giữa các thuộc tính nhạy cảm và tích hợp các kỹ thuật xử lý phù hợp nhằm nâng cao tính ổn định của quá trình ước lượng và mở rộng khả năng áp dụng của phương pháp trên các bộ dữ liệu có cấu trúc phức tạp. Bên cạnh đó, cần nghiên cứu tích hợp các kỹ thuật hiệu chỉnh đa kiểm định như Bonferroni hoặc False Discovery Rate (FDR) trong quá trình xác định các thuộc tính cần điều chỉnh nhằm giảm nguy cơ sai số loại I và nâng cao độ tin cậy của phương pháp Fairedu. Ngoài ra, cần nghiên cứu mở rộng các phương pháp đề xuất cho các bài toán phân lớp đa lớp nhằm mở rộng phạm vi ứng dụng của phương pháp.
2. *Phát triển và hoàn thiện các chỉ số công bằng và phương pháp đánh giá:* Các thước đo hiện tại như *"tác động khác biệt"*, *"hiệu số chênh lệch thống kê"*, *"chênh lệch trung bình xác suất"* và *"chênh lệch cơ hội công bằng"* tuy được sử dụng phổ biến nhưng vẫn còn hạn chế trong việc phản ánh đầy đủ các dạng thiên vị phức tạp, đặc biệt khi tồn tại đồng thời nhiều thuộc tính nhạy cảm. Một hướng phát triển quan trọng là đề xuất và kiểm thử các chỉ số công bằng mới, đồng thời hoàn thiện "chỉ số đánh đổi" như một công cụ hỗ trợ lựa chọn mô hình tối ưu giữa công bằng và hiệu suất. Ngoài ra, cần nghiên cứu các phương pháp đánh giá chặt chẽ hơn thông qua việc kết hợp các kiểm định thống kê, khoảng tin cậy và các chỉ số kích thước ảnh hưởng (effect size), nhằm nâng cao độ tin cậy và tính toàn diện trong việc so sánh hiệu quả giữa các phương pháp đảm bảo công bằng. Các nghiên cứu tiếp theo cũng sẽ xem xét áp dụng các kiểm định phi tham số như Wilcoxon signed-rank, các phương pháp bootstrap, đồng thời tích

hợp các kỹ thuật hiệu chỉnh đa kiểm định như Bonferroni hoặc False Discovery Rate (FDR) trong quá trình xác định các thuộc tính cần điều chỉnh của Fairedu nhằm giảm nguy cơ sai lầm và nâng cao độ tin cậy thống kê của phương pháp.

3. *Nghiên cứu giải pháp sinh dữ liệu tổng hợp nhằm cân bằng dữ liệu cho các tình huống đặc biệt:* Một hướng quan trọng là phát triển và tích hợp các kỹ thuật sinh dữ liệu tổng hợp tiên tiến nhằm bổ sung dữ liệu cho các trường hợp thiên lệch trầm trọng hoặc thiếu vắng những tình huống giao thoa quan trọng. Việc này không chỉ giúp cân bằng phân phối dữ liệu mà còn đảm bảo sự hiện diện đầy đủ hơn của các nhóm thiểu số, qua đó nâng cao hiệu quả can thiệp công bằng. Đồng thời, cần xây dựng các tiêu chí đánh giá chất lượng dữ liệu tổng hợp như fidelity, privacy leakage và khả năng tái lập nhằm bảo đảm độ tin cậy khi ứng dụng các mô hình sinh dữ liệu trong các bài toán công bằng.
4. *Xây dựng khung giải pháp công bằng tổng quát và tự động:* Hiện tại, DPF và FaireduPlus vẫn đòi hỏi sự can thiệp thủ công đáng kể trong quá trình xử lý và hiệu chỉnh. Một định hướng dài hạn là phát triển một khung làm việc công bằng tổng quát có khả năng tự động phát hiện, điều chỉnh và đánh giá công bằng trên nhiều loại dữ liệu và mô hình khác nhau. Khung giải pháp này sẽ mở rộng khả năng ứng dụng của DPF và FaireduPlus trong các hệ thống học máy quy mô lớn, đặc biệt tại những lĩnh vực có yêu cầu nghiêm ngặt về đạo đức và công bằng như giáo dục, y tế, tuyển dụng và cấp tín dụng.

Những hướng đi trên không chỉ góp phần khắc phục các hạn chế hiện tại của luận án mà còn tạo tiền đề cho các nghiên cứu chuyên sâu hơn về công bằng trong học máy và trí tuệ nhân tạo. Việc tiếp tục mở rộng, chuẩn hóa và hoàn thiện các giải pháp công bằng sẽ là nền tảng quan trọng để phát triển các hệ thống học máy minh bạch, đáng tin cậy và vì con người, đặc biệt trong những lĩnh vực có tác động xã hội sâu rộng.