

CHƯƠNG 4

TRÍ TUỆ NHÂN TẠO TRONG SẢN XUẤT

M.Econ. Đặng Thiện Tâm

ĐỊNH HƯỚNG CHƯƠNG 4

Ba chương đầu đã xây dựng nền tảng. Nhà máy thông minh là một trạng thái năng lực đo được bằng hệ thống chỉ số vận hành (Chương 1), được cấu thành từ các trụ cột công nghệ chỉ tạo giá trị khi được tích hợp (Chương 2), và vận hành thông qua dòng thông tin chạy ngược chiều dòng vật chất từ tín hiệu nhu cầu tới giao hàng (Chương 3). Chương 4 bổ sung lớp cuối cùng là **năng lực biến dữ liệu thành quyết định**.

Trọng tâm của chương không phải là dạy người học lập trình thuật toán, vốn là công việc của nhà khoa học dữ liệu, mà là trang bị **bốn năng lực quản trị** mà một nhà quản trị vận hành bắt buộc phải có khi làm việc với trí tuệ nhân tạo. *Năng lực thứ nhất* là đọc hiểu bản đồ khái niệm: phân biệt được trí tuệ nhân tạo, học máy, học sâu, AI tạo sinh và AI tác tử; biết loại bài toán nào cần loại kỹ thuật nào; và nhận ra khi nào một đề xuất đang dùng thuật ngữ thời thượng để che giấu sự thiếu vắng của bài toán thực. *Năng lực thứ hai* là hiểu vòng đời của một dự án dữ liệu, với nhận thức tinh tế rằng phần lớn công sức nằm ở khâu chuẩn bị dữ liệu chứ không ở khâu mô hình hóa. *Năng lực thứ ba*, và là năng lực phân biệt rõ nhất nhà quản trị với người dùng công nghệ, là đánh giá mô hình bằng ngôn ngữ quản trị: không hỏi độ chính xác bao nhiêu phần trăm mà hỏi chi phí của một dương tính giả so với một âm tính giả trong bối cảnh nhà máy này là bao nhiêu. *Năng lực thứ tư* là xây dựng luận chứng kinh tế và quản trị rủi ro, với đầy đủ chi phí duy trì mô hình, hạng mục bị bỏ sót nhiều nhất. So với các tài liệu cùng chủ đề, phiên bản này bổ sung **năm quy trình chuẩn hóa kèm sơ đồ**: cây quyết định lựa chọn nhóm bài toán học máy, quy trình sáu bước gắn nhãn dữ liệu huấn luyện, quy trình sáu bước giám sát và duy trì mô hình sau triển khai, quy trình bảy bước thẩm định một đề xuất dự án AI, và quy trình sáu bước triển khai AI có trách nhiệm. Phiên bản này cũng bổ sung **sáu công cụ định lượng** chuyển các luận điểm từ mô tả sang tính toán: đồ thị lạm phát kỹ thuật, minh họa số về mất cân bằng lớp, đường suy giảm do trôi dạt dữ liệu, đường đánh đổi giữa độ nhạy và độ chuẩn xác, đường tổng chi phí theo ngưỡng quyết định, và biểu đồ so sánh giá trị hiện tại ròng theo ba kịch bản rủi ro.

Chương khép lại bằng phần trình bày thẳng thắn về **giới hạn của AI trong nhà máy**, bởi biết khi nào không nên dùng AI cũng là một năng lực chuyên môn, và bằng lộ trình bốn lớp năng lực tổng hợp toàn chương.

MỤC TIÊU HỌC TẬP CỦA CHƯƠNG 4

Sau khi học xong chương này, người học có thể:

1. **Phân biệt** được trí tuệ nhân tạo, học máy, học sâu, AI tạo sinh và AI tác tử; **giải thích** được khác biệt về mục đích sử dụng của ba dạng AI trong nhà máy.
2. **Nhận diện** được hiện tượng lạm phát kỹ thuật và **lập luận** được cho việc chọn mức phức tạp mô hình tương xứng với bài toán.
3. **Phân loại** được một bài toán vận hành cụ thể vào các nhóm bài toán học máy bằng cây quyết định bốn câu hỏi.
4. **Mô tả** được sáu bước của vòng đời dự án dữ liệu và **giải thích** vì sao khâu chuẩn bị dữ liệu chiếm phần lớn công sức.
5. **Nhận diện** được bốn vấn đề chất lượng dữ liệu sản xuất và **đề xuất** biện pháp xử lý cho từng vấn đề.
6. **Thực hiện** được quy trình sáu bước gán nhãn dữ liệu huấn luyện và **giải thích** vì sao mức đồng thuận giữa những người gán nhãn lại quan trọng.
7. **Thực hiện** được quy trình sáu bước giám sát và duy trì mô hình sau triển khai.
8. **Tính toán và diễn giải** được các chỉ số đánh giá mô hình từ ma trận nhầm lẫn; **giải thích** được vì sao độ chính xác vô dụng khi dữ liệu mất cân bằng.
9. **Lựa chọn** được ngưỡng quyết định dựa trên chi phí tương đối của dương tính giả và âm tính giả.
10. **Phân tích** được sáu nhóm ứng dụng AI trọng yếu theo cơ chế, dữ liệu cần thiết, chỉ số cải thiện và điều kiện áp dụng; **định vị** được chúng theo giá trị và điều kiện dữ liệu.
11. **Vận dụng** được bốn cấp độ phân tích để chẩn đoán một tập dữ liệu ca sản xuất thực và **đề xuất** hành động cải tiến có ước lượng lợi ích.
12. **Xây dựng** được luận chứng kinh tế cho một dự án AI với đầy đủ chi phí duy trì mô hình, kèm phân tích độ nhạy theo kịch bản.
13. **Vận dụng** được quy trình bảy bước thẩm định một đề xuất dự án AI.
14. **Đánh giá** được yêu cầu về AI có trách nhiệm gồm tính giải thích được, thiên lệch dữ liệu, trách nhiệm giải trình và khung pháp lý áp dụng cho hàng xuất khẩu.

15. **Thiết kế** được kiến trúc con người trong vòng lặp phù hợp với hậu quả và khả năng đảo ngược của hành động.

16. **Nhận diện** được các tình huống mà AI không phải là giải pháp phù hợp.

Chương 4 đáp ứng CLO.2 và CLO.3, ở mức Bloom từ Phân tích đến Đánh giá.

Bảng 4.1. Một số thuật ngữ nền tảng của Chương 4

Thuật ngữ tiếng Việt	Thuật ngữ tiếng Anh	Ý nghĩa ngắn gọn
Học máy	Machine learning	Nhóm phương pháp để hệ thống rút ra quy luật từ dữ liệu thay vì được lập trình quy luật sẵn.
Học sâu	Deep learning	Nhánh học máy dùng mạng nơ ron nhiều lớp, mạnh với dữ liệu ảnh, âm thanh và chuỗi phức tạp.
AI tạo sinh	Generative AI	Nhóm mô hình tạo ra nội dung mới gồm văn bản, hình ảnh, mã lệnh và phương án thiết kế.
AI tác tử	Agentic AI	Hệ thống AI tự lập kế hoạch nhiều bước và thực thi tác vụ qua nhiều hệ thống.
Mô hình nền công nghiệp	Industrial foundation model	Mô hình lớn được huấn luyện trên dữ liệu đặc thù sản xuất, làm nền cho nhiều ứng dụng.
Học có giám sát	Supervised learning	Học từ dữ liệu đã gán nhãn để dự đoán nhãn cho dữ liệu mới.
Phát hiện bất thường	Anomaly detection	Học thế nào là bình thường từ dữ liệu vận hành ổn định rồi cảnh báo khi có sai lệch.
Mất cân bằng lớp	Class imbalance	Tình trạng lớp cần phát hiện, thường là lỗi hoặc hỏng hóc, chiếm tỷ lệ rất nhỏ trong dữ liệu.
Trôi dạt dữ liệu	Data drift	Hiện tượng phân phối dữ liệu thay đổi theo thời gian khiến mô hình mất độ chính xác.
Dương tính giả	False positive	Mô hình báo có vấn đề trong khi thực tế không có.

Thuật ngữ tiếng Việt	Thuật ngữ tiếng Anh	Ý nghĩa ngắn gọn
Âm tính giả	False negative	Mô hình báo bình thường trong khi thực tế có vấn đề.
Ngưỡng quyết định	Decision threshold	Mức xác suất tối thiểu để mô hình tuyên bố là có vấn đề.
Con người trong vòng lặp	Human in the loop	Kiến trúc trong đó con người phê duyệt hoặc can thiệp vào quyết định của hệ thống AI.
Tính giải thích được	Explainability	Khả năng trình bày lý do dẫn tới một kết luận của mô hình theo cách con người hiểu được.
Vận hành mô hình học máy	MLOps	Tập thực hành duy trì, giám sát và tái huấn luyện mô hình trong môi trường vận hành thực.
Đường cơ sở đơn giản	Simple baseline	Quy tắc ngưỡng hoặc mô hình thống kê cổ điển dùng làm mốc so sánh bắt buộc.

4.1. Dẫn nhập: bốn năng lực và tinh thần hoài nghi có căn cứ

4.1.1. Khoảng cách giữa quan tâm và giá trị

Trong ba năm gần đây, trí tuệ nhân tạo đã chuyển từ chủ đề nghiên cứu thành hạng mục ngân sách trong hầu hết kế hoạch chuyển đổi số của doanh nghiệp sản xuất. Tuy nhiên, khoảng cách giữa mức độ quan tâm và mức độ tạo ra giá trị vẫn rất lớn. Báo cáo ngành năm 2026 cho thấy một mô thức lặp lại: tỷ lệ doanh nghiệp *dự định* triển khai AI tác tử trong hai năm tới rất cao, nhưng tỷ lệ doanh nghiệp *đã có nền tảng dữ liệu đủ điều kiện* thì thấp hơn nhiều lần. Nút thắt được xác định nhất quán không phải là thuật toán mà là **thiếu hụt nền tảng dữ liệu**.

Nhận định này không mới đối với người học đã đi qua ba chương đầu. Chương 1 đã nêu chuỗi giá trị dữ liệu tuân theo nguyên tắc mắt xích yếu nhất; Chương 2 đã chỉ ra rằng đa số dự án thất bại ở khâu tích hợp; Chương 3 đã cho thấy chất lượng dữ liệu chủ quyết định thành bại của hệ thống điều độ. Trí tuệ nhân tạo không phải là ngoại lệ mà là **trường hợp cực đoan nhất của quy luật đó**: mô hình học máy khuếch đại mọi khiếm khuyết của dữ liệu đầu vào, và làm điều đó một cách khó phát hiện, vì đầu ra của mô hình luôn có vẻ hợp lý ngay cả khi sai.

Chương này vì vậy được viết theo tinh thần **hoài nghi có căn cứ**: không phủ nhận giá trị thực của AI trong sản xuất, vốn đã được chứng minh trong nhiều ứng dụng cụ thể, nhưng liên tục đặt câu

hỏi về điều kiện áp dụng, về cách đo lường lợi ích, và về những tình huống mà giải pháp đơn giản hơn cho kết quả tốt hơn với chi phí thấp hơn.

4.1.2. Bốn năng lực quản trị mà chương này xây dựng



Hình 4.1. Bốn năng lực quản trị mà chương này xây dựng

Nguồn: Tác giả xây dựng.

Đề cương học phần đặt chuẩn đầu ra CLO.3 ở mức *vận dụng* và *đánh giá*: sinh viên phải vận dụng được việc ra quyết định dựa trên dữ liệu trong vận hành nhà máy thông minh nhằm nâng cao hiệu quả. Chương 4 là chương gánh trọng trách chính cho chuẩn đầu ra này, thông qua bài thực hành bắt buộc trên dữ liệu thật ở mục 4.6.2 và bài tập xây dựng luận chứng kinh tế ở mục 4.7. Người học cần chuẩn bị máy tính cá nhân có phần mềm bảng tính và công cụ trực quan hóa dữ liệu cho buổi thực hành.

4.2. Bản đồ khái niệm: từ trí tuệ nhân tạo tới AI tác tử

4.2.1. Quan hệ bao hàm giữa các khái niệm

Trí tuệ nhân tạo là khái niệm rộng nhất, chỉ mọi hệ thống thực hiện các nhiệm vụ thường được coi là đòi hỏi trí tuệ con người. **Học máy** là tập con, gồm các phương pháp mà hệ thống rút ra quy luật từ dữ liệu thay vì được lập trình quy luật sẵn. **Học sâu** là tập con của học máy, sử dụng mạng nơ ron nhiều lớp; ưu thế của nó thể hiện rõ nhất với dữ liệu phi cấu trúc như ảnh, âm thanh và chuỗi phức tạp. **AI tạo sinh** là nhóm ứng dụng của học sâu tạo ra nội dung mới. **AI tác tử** không phải một tầng công nghệ mới mà là một **kiến trúc sử dụng**: mô hình được trao công cụ, quyền truy cập hệ thống và khả năng lập kế hoạch nhiều bước để tự thực thi một quy trình công việc.

QUAN HỆ BAO HÀM GIỮA CÁC CẤP ĐỘ TỰ ĐỘNG HÓA VÀ VỊ TRÍ CỦA SIÊU TỰ ĐỘNG HÓA



Hình 4.2. Quan hệ bao hàm giữa các khái niệm và vị trí của kiến trúc AI tác tử

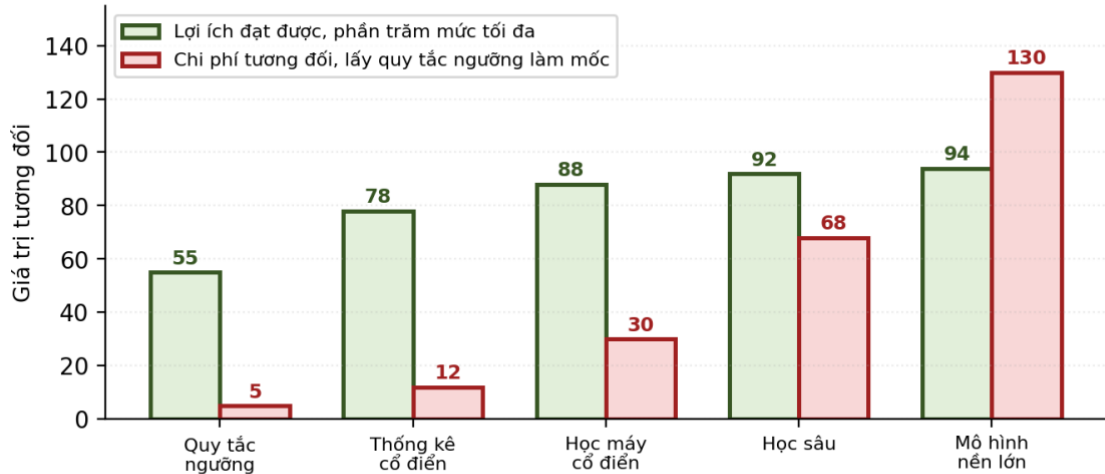
Nguồn: Tác giả xây dựng.

4.2.2. Lạm phát kỹ thuật

Một hệ quả quan trọng của cấu trúc bao hàm ở Hình 4.2 là **không phải mọi bài toán AI đều cần học sâu, và rất nhiều bài toán vận hành được giải tốt hơn bằng thống kê cổ điển**. Dự báo nhu cầu cho một mã hàng có mùa vụ ổn định thường được xử lý tốt bằng mô hình san mũ ba tham số đã học ở Chương 3, với chi phí bằng một phần nhỏ so với mạng nơ ron và với ưu điểm quyết định là *giải thích được*.

Việc chọn công cụ phức tạp hơn mức cần thiết là một dạng lãng phí mà chương này gọi là **lạm phát kỹ thuật**. Hiện tượng này có thể được minh họa bằng số thay vì chỉ mô tả.

LẠM PHÁT KỸ THUẬT: LỢI ÍCH BẢO HÒA TRONG KHI CHI PHÍ TĂNG NHANH



Số liệu minh họa cho một bài toán phân loại điển hình trong nhà máy. Điểm cần rút ra là hình dạng của hai đường: lợi ích bảo hòa nhanh trong khi chi phí tiếp tục tăng, nên mức phức tạp tối ưu hiếm khi nằm ở cột cuối cùng.

Hình 4.3. Lạm phát kỹ thuật: lợi ích bảo hòa trong khi chi phí tăng nhanh

Nguồn: Tác giả xây dựng với số liệu minh họa cho một bài toán phân loại điển hình trong nhà máy.

Hình 4.3 hàm chứa hai thông điệp quản trị. *Thứ nhất*, khoảng cách lợi ích giữa thống kê cổ điển và mô hình nền lớn chỉ khoảng 16 điểm phần trăm, trong khi khoảng cách chi phí là hơn mười lần. *Thứ hai*, và quan trọng hơn, mỗi bước tăng độ phức tạp đều kéo theo một khoản chi phí ẩn không xuất hiện trong hình: mô hình càng phức tạp thì càng khó giải thích, càng khó bảo trì và càng cần nhân lực chuyên môn hiếm.

4.2.3. Ba dạng sử dụng AI trong nhà máy

Bảng 4.2. Phân biệt AI dự báo, AI tạo sinh và AI tác tử theo mục đích sử dụng

Tiêu chí	AI dự báo	AI tạo sinh	AI tác tử
Câu hỏi trả lời	Giá trị hoặc nhãn nào sẽ xảy ra	Nội dung nào phù hợp với yêu cầu này	Làm thế nào để hoàn thành quy trình này
Đầu ra	Con số, xác suất, phân loại	Văn bản, hình ảnh, phương án thiết kế, mã lệnh	Chuỗi hành động được thực thi trên nhiều hệ thống
Ứng dụng điển hình	Bảo trì dự báo, phát hiện lỗi, dự báo nhu cầu	Soạn quy trình chuẩn, tóm tắt tài liệu kỹ thuật, hỗ trợ chẩn đoán sự cố	Tự động xử lý đơn hàng, điều phối lệnh bảo trì, tổng hợp báo cáo đa nguồn

Tiêu chí	AI dự báo	AI tạo sinh	AI tác tử
Tính xác định của kết quả	Cao; cùng đầu vào cho cùng đầu ra	Thấp; kết quả biến thiên giữa các lần chạy	Thấp và tích lũy qua nhiều bước
Rủi ro đặc thù	Sai lệch do trôi dạt dữ liệu	Nội dung sai nhưng trình bày thuyết phục	Sai ở một bước lan truyền qua toàn chuỗi hành động
Yêu cầu giám sát của con người	Theo mẫu, định kỳ	Kiểm chứng nội dung trước khi sử dụng	Phê duyệt tại các điểm có hậu quả không đảo ngược

4.2.4. AI tác tử và mô hình nền công nghiệp: trạng thái công nghệ năm 2026

Đây là lĩnh vực đang thay đổi nhanh nhất và cũng dễ bị thổi phồng nhất, nên cần được trình bày với các mốc tham chiếu cụ thể.

Xu hướng thứ nhất là mô hình nền chuyên biệt cho công nghiệp. Thay vì dùng mô hình ngôn ngữ đa dụng, các nhà cung cấp phần mềm công nghiệp đang huấn luyện những mô hình nền có phạm vi hẹp hơn, dựa trên dữ liệu đặc thù sản xuất như nhật ký máy và tham số quy trình. Ưu thế của hướng này gồm ba mặt: dự đoán chính xác hơn nhờ hiểu ngữ cảnh chuyên ngành; tiêu tốn ít tài nguyên tính toán hơn; và kích thước gọn cho phép triển khai ngay tại biên, nơi độ trễ, băng thông và tính riêng tư của dữ liệu là ràng buộc quyết định.

Xu hướng thứ hai là chuẩn hóa giao tiếp giữa các tác tử và với dữ liệu. Các giao thức mở cho phép tác tử của những nhà phát triển khác nhau phối hợp với nhau, và cho phép tác tử kết nối trực tiếp tới nguồn dữ liệu thời gian thực thay vì chỉ dựa trên tri thức đã đóng băng lúc huấn luyện.

Ba rào cản được ghi nhận nhất quán. Một là thiếu hụt nền tảng dữ liệu, được xác định là nút thắt lớn nhất đối với việc mở rộng quy mô. Hai là tính không xác định của mô hình ngôn ngữ: cùng một câu hỏi có thể cho câu trả lời khác nhau, điều này khó chấp nhận trong ngành có yêu cầu kiểm định nghiêm ngặt. Ba là khoảng cách giữa ý định và năng lực.

Kết luận quản trị cho người học: **con người trong vòng lặp vẫn là yêu cầu lõi của AI công nghiệp**, vì một quyết định sai trong môi trường sản xuất có thể làm dừng dây chuyền, hỏng thiết bị và tạo ra vấn đề kiểm toán. Mức tự chủ được trao cho hệ thống phải tỷ lệ nghịch với hậu quả của sai lầm, nguyên tắc sẽ được cụ thể hóa ở mục 4.8.

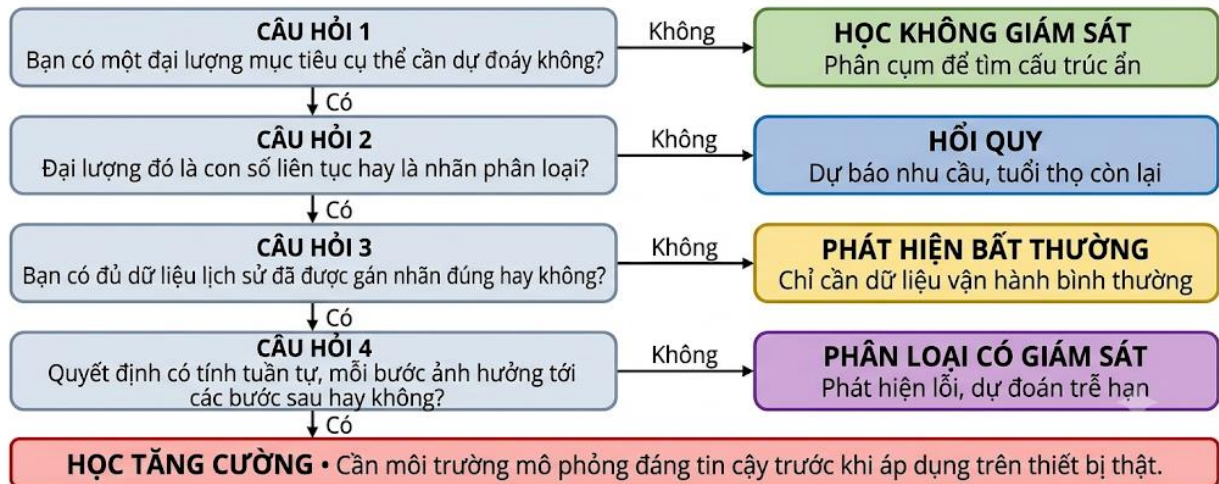
4.2.5. Ba nhóm bài toán học máy và cách chọn nhóm phù hợp

Bảng 4.3. Ba nhóm bài toán học máy và ứng dụng trong nhà máy

Nhóm bài toán	Dạng cụ thể	Ứng dụng trong nhà máy	Yêu cầu dữ liệu
Học có giám sát	Phân loại với nhãn rời rạc	Phân loại sản phẩm đạt hoặc lỗi; phân loại loại lỗi bề mặt; dự đoán máy sẽ hỏng trong bảy ngày tới hay không	Dữ liệu lịch sử kèm nhãn đúng ; đây thường là nút thắt lớn nhất
Học có giám sát	Hồi quy với giá trị liên tục	Dự báo nhu cầu; dự đoán tuổi thọ còn lại của thiết bị; dự đoán độ bền theo tham số quy trình	Dữ liệu lịch sử có giá trị mục tiêu đo được
Học không giám sát	Phân cụm	Nhóm mã hàng có mô thức nhu cầu tương tự; nhóm máy có hành vi vận hành giống nhau	Không cần nhãn; cần dữ liệu đủ đa dạng
Học không giám sát	Phát hiện bất thường	Cảnh báo hành vi vận hành lệch chuẩn khi chưa có lịch sử hỏng hóc đủ nhiều	Dữ liệu giai đoạn vận hành bình thường đủ dài
Học tăng cường	Điều khiển và tối ưu tuần tự	Tối ưu lịch điều độ; điều khiển robot; tối ưu tham số quy trình liên tục	Môi trường mô phỏng đáng tin cậy; khó áp dụng trực tiếp trên thiết bị thật

Bảng trên liệt kê các nhóm, nhưng câu hỏi thực hành mà người học cần trả lời được là: *với một bài toán vận hành cụ thể, nên chọn nhóm nào?* Câu quyết định dưới đây chuẩn hóa lựa chọn đó theo bốn câu hỏi có thứ tự.

CÂY QUYẾT ĐỊNH LỰA CHỌN NHÓM BÀI TOÁN HỌC MÁY



Hình 4.4. Cây quyết định lựa chọn nhóm bài toán học máy

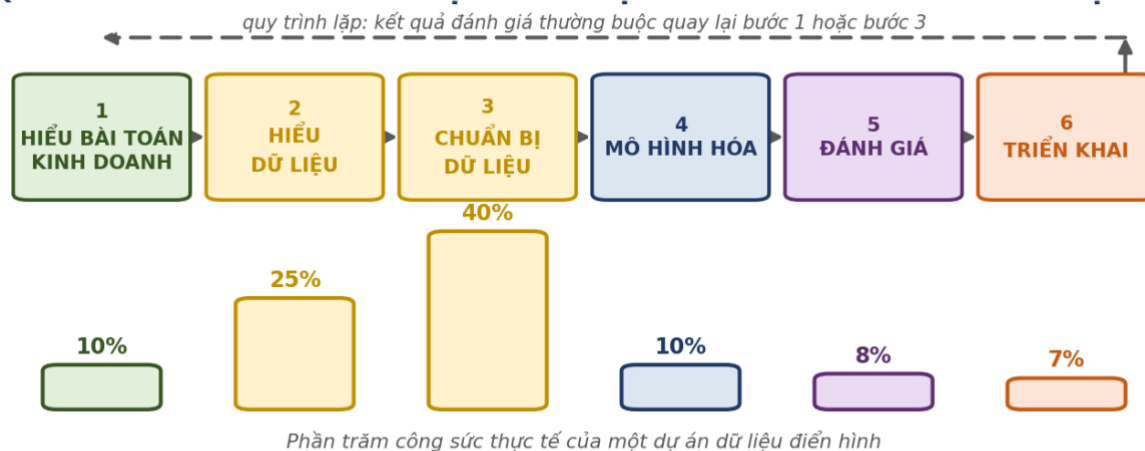
Nguồn: Tác giả xây dựng.

Một chỉ dẫn thực hành quan trọng gắn với câu hỏi thứ ba: **khi doanh nghiệp chưa có lịch sử hỏng hóc được gán nhãn đầy đủ, tình huống rất phổ biến, thì phát hiện bất thường theo hướng học không giám sát là điểm khởi đầu hợp lý hơn nhiều so với bài toán phân loại có giám sát.** Mô hình phát hiện bất thường chỉ cần học thể nào là bình thường từ dữ liệu vận hành ổn định, và cảnh báo khi có sai lệch. Nhược điểm là nó không cho biết *nguyên nhân*, nên vẫn cần chuyên gia diễn giải.

4.3. Vòng đời dự án dữ liệu và chất lượng dữ liệu sản xuất

4.3.1. Sáu bước của vòng đời dự án dữ liệu

QUY TRÌNH SÁU BƯỚC CỦA DỰ ÁN DỮ LIỆU VÀ PHÂN BỐ CÔNG SỨC THỰC TẾ



Khoảng 65 phần trăm công sức nằm ở bước 2 và bước 3 chứ không ở bước mô hình hóa. Đây là lý do nhiều dự án vượt tiến độ nghiêm trọng: kế hoạch được lập theo hình dung chứ không theo thực tế.

Hình 4.5. Quy trình sáu bước của dự án dữ liệu và phân bố công sức thực tế

Nguồn: Tác giả biên vẽ theo Chapman và cộng sự (2000), phần phân bố công sức dựa trên thực tiễn triển khai được ghi nhận trong tài liệu ngành.

Bảng 4.4. Sáu bước của vòng đời dự án dữ liệu trong bối cảnh nhà máy

Bước	Nội dung	Sản phẩm bàn giao	Sai lầm điển hình
1. Hiểu bài toán kinh doanh	Xác định vấn đề vận hành, chỉ số cần cải thiện, mức cải thiện mục tiêu và ràng buộc	Phát biểu bài toán kèm tiêu chí thành công đo được	Bắt đầu từ suy nghĩ chúng ta có dữ liệu, hãy xem tìm được gì, thay vì bắt đầu từ vấn đề
2. Hiểu dữ liệu	Kiểm kê nguồn dữ liệu, đánh giá độ phủ, chất lượng và tần suất	Báo cáo hiện trạng dữ liệu và khoảng trống	Giả định dữ liệu tồn tại và dùng được mà không kiểm tra thực tế
3. Chuẩn bị dữ liệu	Làm sạch, xử lý thiếu, hợp nhất nguồn, tạo biến, gán nhãn	Tập dữ liệu sẵn sàng huấn luyện kèm tài liệu mô tả	Đánh giá thấp khối lượng công việc; không lưu tài liệu về các phép biến đổi đã làm

Bước	Nội dung	Sản phẩm bàn giao	Sai lầm điển hình
4. Mô hình hóa	Chọn thuật toán, huấn luyện, điều chỉnh tham số	Mô hình ứng viên kèm kết quả trên tập kiểm định	Bắt đầu từ mô hình phức tạp thay vì từ đường cơ sở đơn giản
5. Đánh giá	Đối chiếu kết quả với tiêu chí kinh doanh đã nêu ở bước 1	Báo cáo đánh giá theo ngôn ngữ chi phí và lợi ích	Đánh giá bằng chỉ số kỹ thuật mà không quy đổi ra tác động vận hành
6. Triển khai	Đưa vào quy trình làm việc, giám sát, tái huấn luyện	Hệ thống vận hành kèm quy trình xử lý và kế hoạch bảo trì mô hình	Coi triển khai là điểm kết thúc thay vì điểm bắt đầu của vòng đời mô hình

Hai nhận thức cần được nhấn mạnh. *Thứ nhất*, khoảng 65 phần trăm công sức của một dự án nằm ở bước 2 và bước 3, không phải ở bước 4. Điều này trái ngược với hình dung phổ biến và cũng giải thích vì sao nhiều dự án vượt tiến độ nghiêm trọng: kế hoạch được lập theo hình dung chứ không theo thực tế. *Thứ hai*, quy trình này mang tính **lặp** chứ không tuyến tính: kết quả đánh giá ở bước 5 thường buộc quay lại bước 1 để phát biểu lại bài toán, hoặc quay về bước 3 để bổ sung biến.

Một quy tắc thực hành có giá trị cao: **luôn xây dựng một đường cơ sở đơn giản trước**. Đường cơ sở có thể là quy tắc ngưỡng của chuyên gia hoặc mô hình thống kê cổ điển. Nếu mô hình học máy phức tạp không vượt đường cơ sở một cách đáng kể, kết luận đúng là dùng đường cơ sở, vì nó rẻ hơn, nhanh hơn và giải thích được. Đây cũng chính là nội dung định lượng của Hình 4.3.

4.3.2. Bốn vấn đề chất lượng dữ liệu sản xuất

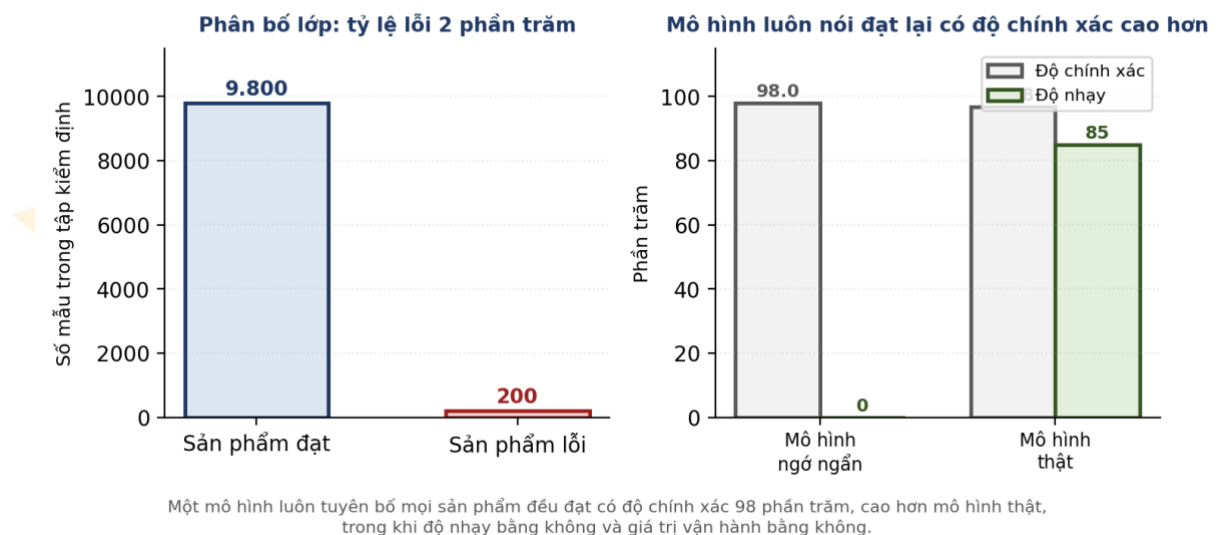
Bảng 4.5. Bốn vấn đề chất lượng dữ liệu đặc thù của môi trường sản xuất

Vấn đề	Biểu hiện	Hậu quả với mô hình	Biện pháp xử lý
Dữ liệu thiếu	Mất bản ghi khi mất mạng; ca đêm không ghi nhận; trường dữ liệu bỏ trống	Mô hình học sai vì mẫu thiếu không ngẫu nhiên, dữ liệu thường thiếu đúng vào lúc có sự cố	Bổ sung cơ chế lưu đệm tại biên; phân tích cơ chế thiếu trước khi điền giá trị thay thế

Vấn đề	Biểu hiện	Hậu quả với mô hình	Biện pháp xử lý
Nhiều cảm biến	Tín hiệu dao động do rung, nhiễu điện, trôi hiệu chuẩn	Mô hình học nhiễu thay vì học tín hiệu; sinh nhiễu cảnh báo giả	Lọc tín hiệu; hiệu chuẩn định kỳ có lưu vết; kiểm tra tính hợp lệ theo miền vật lý
Mất cân bằng lớp	Lỗi và hỏng hóc là sự kiện hiếm, chiếm tỷ lệ rất nhỏ	Mô hình đạt độ chính xác cao bằng cách luôn dự đoán bình thường; vô dụng trong thực tế	Chọn chỉ số đánh giá phù hợp; gán trọng số lớp; lấy mẫu lại; hoặc chuyển sang bài toán phát hiện bất thường
Trôi dữ liệu	Thiết bị lão hóa, đổi nhà cung cấp vật tư, đổi mã hàng, đổi mùa vụ	Mô hình mất dần độ chính xác mà không có cảnh báo rõ ràng	Giám sát phân phối dữ liệu đầu vào và hiệu năng mô hình; lập lịch tái huấn luyện; đặt ngưỡng kích hoạt

Mất cân bằng lớp đáng được minh họa bằng số vì đây là cạm bẫy khiến nhiều hội đồng đầu tư phê duyệt nhầm một mô hình vô dụng.

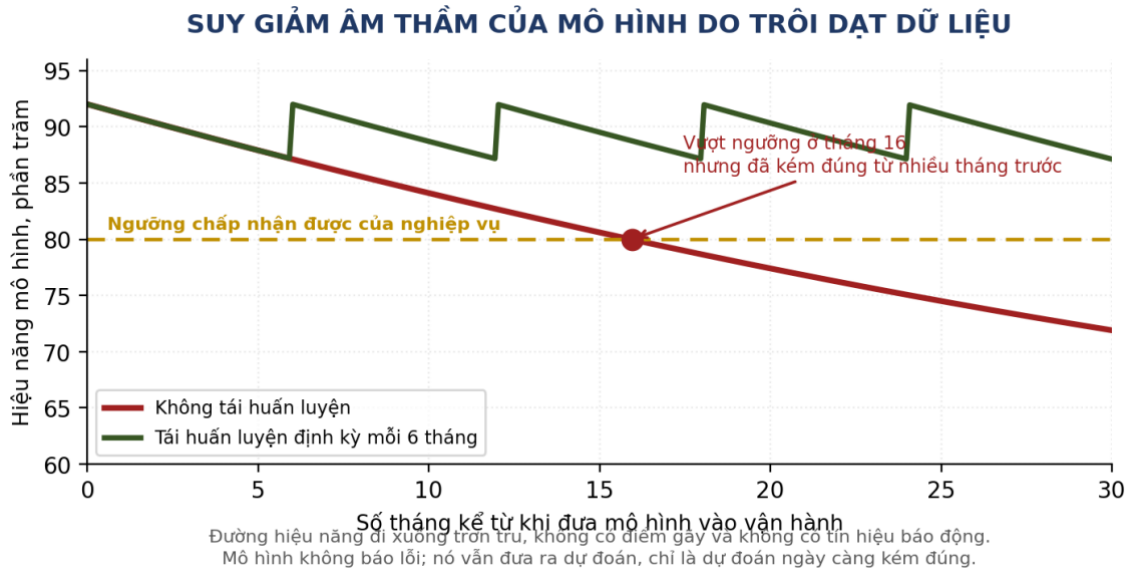
VÌ SAO ĐỘ CHÍNH XÁC VÔ DỤNG KHI DỮ LIỆU MẤT CÂN BẰNG LỚP



Hình 4.6. Vì sao độ chính xác vô dụng khi dữ liệu mất cân bằng lớp

Nguồn: Tác giả tính toán và biên vẽ theo số liệu của ví dụ ở mục 4.4.

Trôi dạt dữ liệu xứng đáng được nhấn mạnh riêng vì nó là nguyên nhân phổ biến nhất khiến một dự án AI thành công ở giai đoạn thí điểm lại suy giảm âm thầm sau khoảng một năm tới một năm rưỡi vận hành. Điểm nguy hiểm nằm ở chỗ mô hình không báo lỗi; nó vẫn đưa ra dự đoán, chỉ là dự đoán ngày càng kém đúng.



Hình 4.7. Suy giảm âm thầm của mô hình do trôi dạt dữ liệu

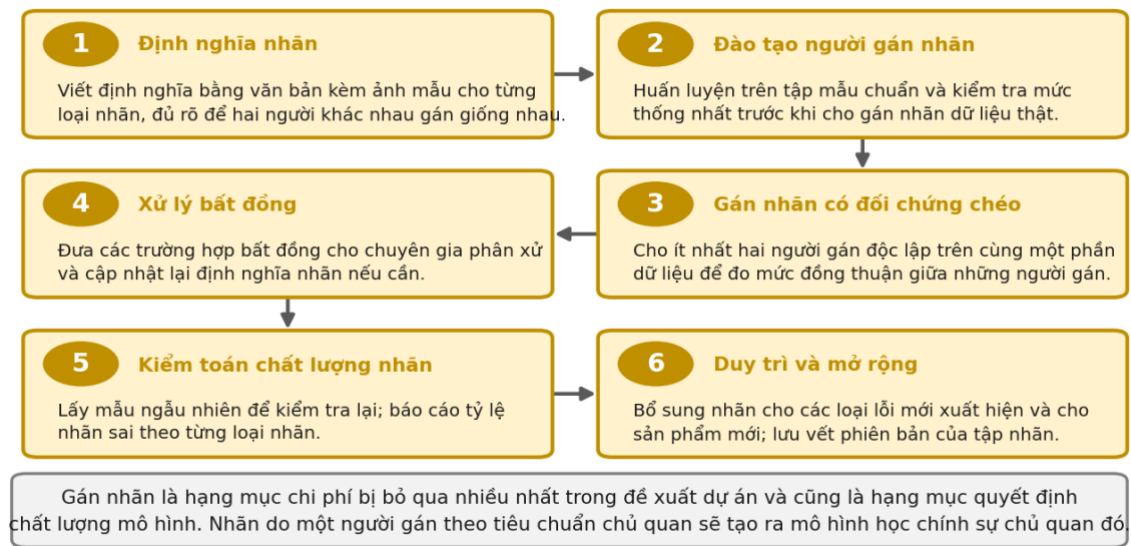
Nguồn: Tác giả mô phỏng minh họa theo cơ chế trôi dạt được mô tả trong Sculley và cộng sự (2015).

Đồ thị cho thấy một đặc điểm mà mô tả bằng lời khó truyền đạt: đường hiệu năng của mô hình không tái huấn luyện đi xuống trơn tru, không có điểm gãy, không có tín hiệu báo động. Đến khi vượt ngưỡng chấp nhận được của nghiệp vụ ở khoảng tháng thứ mười sáu thì mô hình đã kém đúng suốt nhiều tháng trước đó, và những quyết định đã đưa ra trong khoảng thời gian đó không còn cách nào truy hồi. Vì vậy, **kế hoạch giám sát và tái huấn luyện phải là một phần của hợp đồng triển khai, không phải một hạng mục phát sinh sau.**

4.3.3. Quy trình sáu bước gắn nhãn dữ liệu huấn luyện

Trong bốn vấn đề nêu trên, khâu tạo ra nhãn đúng là khâu tốn kém nhất và bị đánh giá thấp nhất. Nhãn không phải là thứ có sẵn trong hệ thống; nó là sản phẩm của một quá trình lao động có tổ chức, và chất lượng của nó đặt trần cho chất lượng của mọi mô hình xây trên đó.

QUY TRÌNH SÁU BƯỚC GÁN NHÂN DỮ LIỆU HUẤN LUYỆN



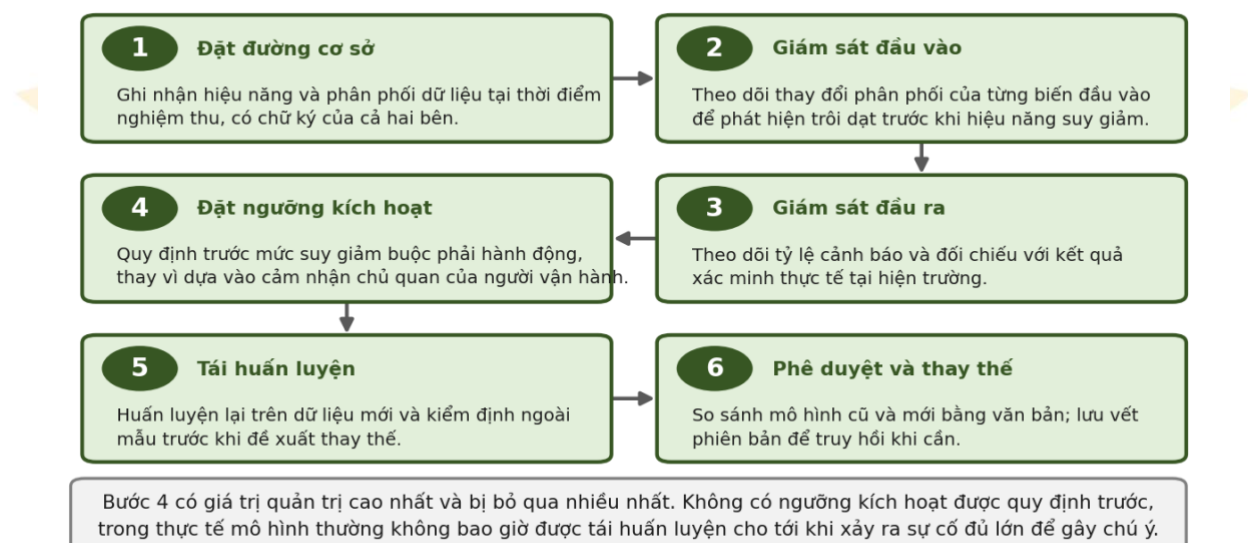
Hình 4.8. Quy trình sáu bước gán nhãn dữ liệu huấn luyện

Nguồn: Tác giả xây dựng.

Bước 3 là bước có giá trị chẩn đoán cao nhất. Khi hai người gán nhãn độc lập trên cùng một tập dữ liệu, mức đồng thuận giữa họ chính là **giới hạn trên của độ chính xác mà bất kỳ mô hình nào cũng có thể đạt được**. Nếu hai kiểm tra viên chỉ thống nhất với nhau ở 75 phần trăm số mẫu, thì việc kỳ vọng một mô hình đạt 95 phần trăm là kỳ vọng không có cơ sở, vì bản thân định nghĩa nhãn còn chưa rõ ràng. Trong trường hợp đó, hành động đúng là quay lại bước 1 để làm rõ định nghĩa, chứ không phải đổi thuật toán.

4.3.4. Quy trình sáu bước giám sát và duy trì mô hình sau triển khai

QUY TRÌNH SÁU BƯỚC GIÁM SÁT VÀ DUY TRÌ MÔ HÌNH SAU TRIỂN KHAI



Hình 4.9. Quy trình sáu bước giám sát và duy trì mô hình sau triển khai

Nguồn: Tác giả xây dựng theo thực hành vận hành mô hình học máy được tổng hợp trong Sculley và cộng sự (2015).

Bảng 4.6. Sản phẩm đầu ra và người chịu trách nhiệm của từng bước

Bước	Sản phẩm đầu ra	Người chịu trách nhiệm điển hình
1. Đặt đường cơ sở	Hồ sơ hiệu năng và phân phối dữ liệu tại thời điểm nghiệm thu	Nhóm triển khai và bên tiếp nhận cùng ký
2. Giám sát đầu vào	Báo cáo định kỳ về thay đổi phân phối biến đầu vào	Người quản lý dữ liệu vận hành
3. Giám sát đầu ra	Báo cáo tỷ lệ cảnh báo và kết quả xác minh thực tế	Bộ phận vận hành sử dụng mô hình
4. Đặt ngưỡng kích hoạt	Quy định mức suy giảm buộc phải hành động	Chủ sở hữu nghiệp vụ của mô hình
5. Tái huấn luyện	Mô hình mới kèm kết quả kiểm định ngoài mẫu	Nhóm phân tích dữ liệu
6. Phê duyệt và thay thế	Biên bản so sánh mô hình cũ và mới; quyết định đưa vào vận hành	Hội đồng quản trị dữ liệu hoặc người phê duyệt được chỉ định

Bước 4 là bước có giá trị quản trị cao nhất và cũng bị bỏ qua nhiều nhất. Không có ngưỡng kích hoạt được quy định trước, việc tái huấn luyện sẽ phụ thuộc vào cảm nhận chủ quan của người vận hành, và trong thực tế điều đó thường có nghĩa là mô hình không bao giờ được tái huấn luyện cho tới khi xảy ra sự cố đủ lớn để gây chú ý.

4.4. Đánh giá mô hình bằng ngôn ngữ quản trị

4.4.1. Ma trận nhầm lẫn và các chỉ số dẫn xuất

Với bài toán phân loại hai lớp, dạng phổ biến nhất trong nhà máy, kết quả của mô hình được tổng hợp trong ma trận nhầm lẫn gồm bốn ô.



Hình 4.10. Ma trận nhầm lẫn và các chỉ số dẫn xuất

Nguồn: Tác giả xây dựng.

Bảng 4.7. Các chỉ số đánh giá mô hình phân loại và ý nghĩa quản trị

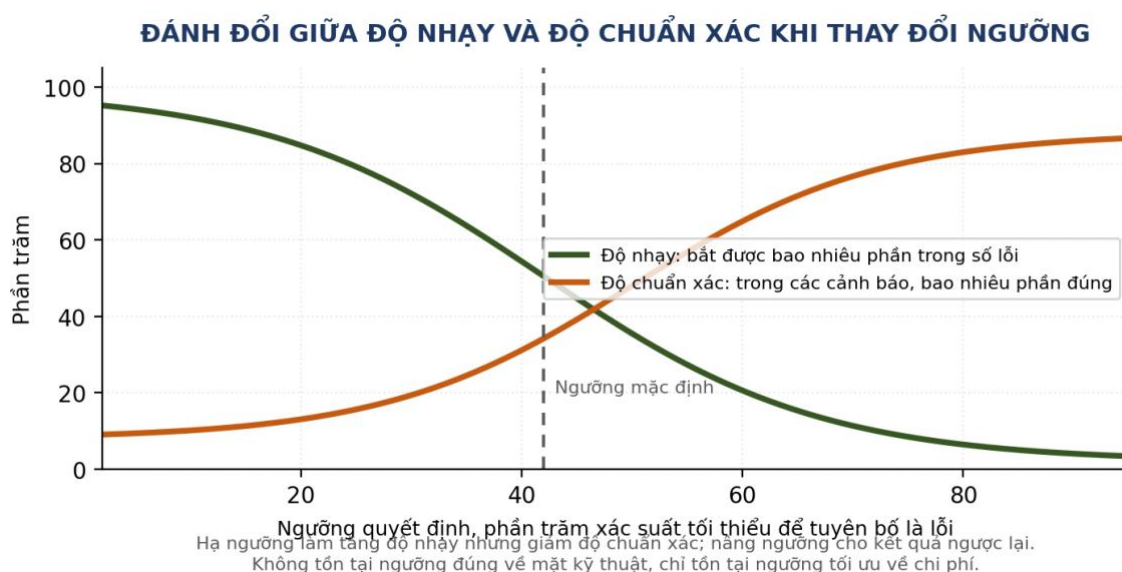
Chỉ số	Công thức	Câu hỏi trả lời	Cạm bẫy
Độ chính xác	Tổng dương tính thật và âm tính thật, chia tổng số mẫu	Tỷ lệ dự đoán đúng nói chung	Vô dụng khi dữ liệu mất cân bằng
Độ nhạy, còn gọi là độ bao phủ	Dương tính thật chia tổng dương tính thật và âm tính giả	Trong tất cả sản phẩm lỗi, mô hình bắt được bao nhiêu phần	Có thể đẩy lên 100 phần trăm bằng cách báo lỗi mọi thứ
Độ chuẩn xác	Dương tính thật chia tổng dương tính thật và dương tính giả	Trong các cảnh báo, bao nhiêu phần là đúng	Có thể đẩy lên cao bằng cách chỉ báo khi rất chắc chắn, đổi lại bỏ sót nhiều
Độ đặc hiệu	Âm tính thật chia tổng âm tính thật và dương tính giả	Trong sản phẩm tốt, mô hình để yên được bao nhiêu phần	Luôn cao khi lớp âm chiếm đa số

Chỉ số	Công thức	Câu hỏi trả lời	Cạm bẫy
Điểm F1	Trung bình điều hòa của độ nhạy và độ chuẩn xác	Cân bằng hai chỉ số trên	Giả định hai loại sai lầm có chi phí ngang nhau, điều thường không đúng

4.4.2. Nguyên tắc trung tâm: chi phí của hai loại sai lầm không bằng nhau

Đây là nội dung quan trọng nhất của mục này và cũng là điểm phân biệt rõ nhất giữa cách tiếp cận của nhà quản trị với cách tiếp cận thuần kỹ thuật.

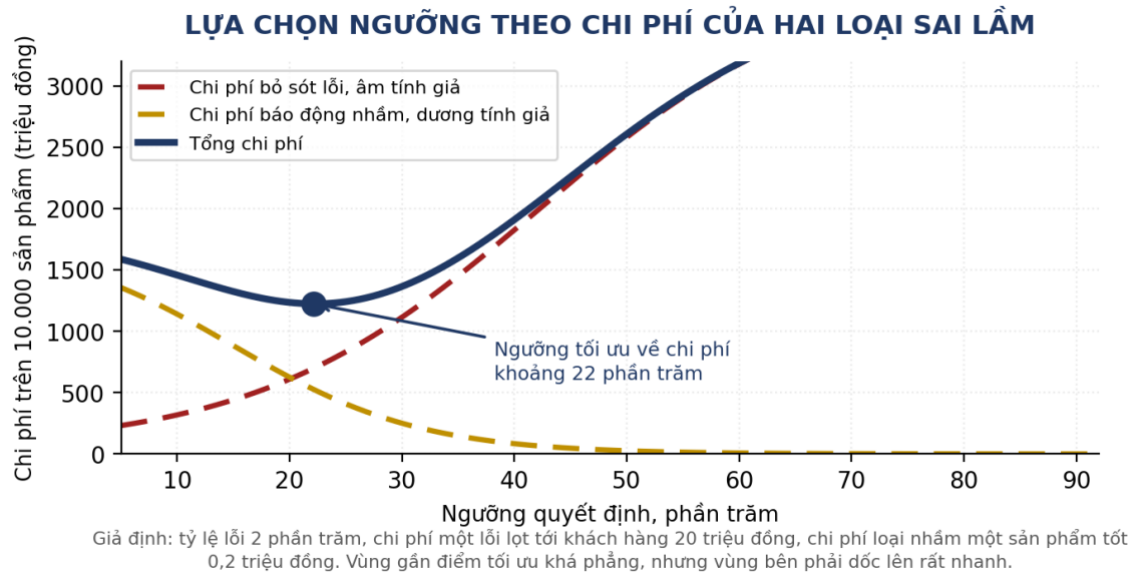
Mọi mô hình phân loại đều có một **ngưỡng quyết định**, tức mức xác suất tối thiểu để tuyên bố là lỗi. Hạ ngưỡng làm tăng độ nhạy nhưng giảm độ chuẩn xác; nâng ngưỡng làm điều ngược lại.



Hình 4.11. Đánh đổi giữa độ nhạy và độ chuẩn xác khi thay đổi ngưỡng

Nguồn: Tác giả xây dựng.

Điểm mấu chốt là **không tồn tại ngưỡng đúng về mặt kỹ thuật, chỉ tồn tại ngưỡng tối ưu về chi phí**, và chi phí đó phụ thuộc hoàn toàn vào bối cảnh vận hành. Trong nhà máy sản xuất linh kiện an toàn cho ô tô hoặc thiết bị y tế, **âm tính giả**, tức để một sản phẩm lỗi lọt tới khách hàng, có thể dẫn tới thu hồi, kiện tụng và mất hợp đồng; chi phí của nó lớn hơn hàng trăm lần so với việc loại nhầm một sản phẩm tốt. Ngược lại, trong dây chuyền sản xuất hàng khối lượng lớn giá trị thấp, **dương tính giả** ở tỷ lệ cao sẽ làm ngừng chuyền liên tục, khiến người vận hành mất tin tưởng và cuối cùng tắt hệ thống cảnh báo, một dạng thất bại thường gặp gọi là *một mối cảnh báo*.



Hình 4.12. Lựa chọn ngưỡng quyết định theo chi phí của hai loại sai lầm

Nguồn: Tác giả tính toán và biên vẽ với giả định tỷ lệ lỗi 2 phần trăm và chi phí âm tính giả gấp 100 lần chi phí dương tính giả.

Hình 4.12 làm rõ ba điều mà bảng số liệu không thể hiện được. *Thứ nhất*, hai đường chi phí thành phần đi ngược chiều nhau, nên tổng chi phí có dạng chữ U và tồn tại một điểm cực tiểu. *Thứ hai*, vùng gần điểm tối ưu khá phẳng, nghĩa là không cần tìm ngưỡng chính xác tuyệt đối; sai lệch nhỏ quanh điểm tối ưu ít gây thiệt hại. *Thứ ba*, vùng bên phải dốc lên rất nhanh: khi chi phí bỏ sót lỗi lớn, **đặt ngưỡng quá cao là sai lầm tốn kém hơn nhiều so với đặt ngưỡng quá thấp.**

VÍ DỤ TÍNH TOÁN VÀ LỰA CHỌN NGƯỠNG THEO CHI PHÍ

Bối cảnh. Hệ thống thị giác máy kiểm tra lỗi bề mặt. Tập kiểm định gồm 10.000 sản phẩm, trong đó 200 sản phẩm thực sự có lỗi, tương ứng tỷ lệ lỗi 2 phần trăm. Kết quả với ngưỡng mặc định: dương tính thật 170; âm tính giả 30; dương tính giả 294; âm tính thật 9.506.

Các chỉ số. Độ chính xác bằng tổng 9.676 chia 10.000, tức **96,76 phần trăm**. Độ nhạy bằng 170 chia 200, tức **85,0 phần trăm**. Độ chuẩn xác bằng 170 chia 464, tức **36,6 phần trăm**. Độ đặc hiệu bằng 9.506 chia 9.800, tức **97,0 phần trăm**. Điểm F1 xấp xỉ **0,51**.

Quan sát thứ nhất là cạm bẫy của độ chính xác. Một mô hình ngớ ngẩn luôn tuyên bố mọi sản phẩm đều tốt sẽ đạt độ chính xác 9.800 chia 10.000, tức **98 phần trăm**, *cao hơn* mô hình thật, trong khi giá trị thực tế bằng không. Đây là minh chứng dứt khoát rằng **độ chính xác không được dùng làm chỉ số đánh giá khi dữ liệu mất cân bằng**, đúng như Hình 4.6 đã thể hiện.

Quan sát thứ hai là quy đổi ra chi phí. Giả sử một sản phẩm lỗi lọt tới khách hàng gây thiệt hại trung bình 20 triệu đồng gồm khiếu nại, thay thế và ảnh hưởng uy tín; một sản phẩm

tốt bị loại nhằm gây thiệt hại 0,2 triệu đồng gồm kiểm tra lại và xử lý. Tổng chi phí sai lầm bằng tích của 30 với 20 cộng tích của 294 với 0,2, tức 600 cộng 58,8 và bằng **658,8 triệu đồng** trên tập 10.000 sản phẩm.

Quan sát thứ ba là hạ ngưỡng. Nếu hạ ngưỡng để độ nhạy tăng lên 95 phần trăm, tức âm tính giả còn 10, cái giá là dương tính giả tăng lên khoảng 900. Tổng chi phí mới bằng 200 cộng 180 và bằng **380 triệu đồng**, giảm **278,8 triệu đồng** so với ngưỡng ban đầu.

Kết luận quản trị. Khi chi phí của âm tính giả lớn gấp 100 lần chi phí của dương tính giả, ngưỡng tối ưu nằm ở vùng độ nhạy cao, tức chấp nhận nhiều cảnh báo sai để không bỏ sót lỗi. **Nhiệm vụ của nhà quản trị không phải là yêu cầu mô hình chính xác hơn mà là cung cấp cho nhóm kỹ thuật hai con số: chi phí của mỗi loại sai lầm.** Không có hai con số đó, việc chọn ngưỡng trở thành quyết định kỹ thuật tùy tiện, và thường là quyết định sai.

Một ràng buộc bổ sung cần kiểm tra. Ở mức 900 dương tính giả trên 10.000 sản phẩm, khối lượng kiểm tra lại phát sinh có thể vượt năng lực của tổ kiểm tra. Nếu vậy, ngưỡng tối ưu về chi phí lý thuyết lại **không khả thi về mặt vận hành**, và lời giải đúng là đầu tư song song vào năng lực xử lý cảnh báo hoặc vào điều kiện quan sát để giảm dương tính giả.

ĐIỂM CẦN NHỚ

Ba câu hỏi bắt buộc phải trả lời trước khi phê duyệt một mô hình phân loại đưa vào vận hành. Thứ nhất, tỷ lệ lớp thiểu số trong dữ liệu là bao nhiêu, và chỉ số đánh giá đã được chọn phù hợp với mức mất cân bằng đó chưa?

Thứ hai, chi phí bằng tiền của một dương tính giả và của một âm tính giả là bao nhiêu, ai đã ước lượng và trên căn cứ nào?

Thứ ba, khi hệ thống phát cảnh báo thì quy trình xử lý là gì, ai chịu trách nhiệm, trong bao lâu, và điều gì xảy ra nếu số cảnh báo vượt quá năng lực xử lý của đội ngũ?

GIÁ TRỊ TẠI TÂM

4.5. Sáu nhóm ứng dụng AI trọng yếu trong nhà máy

4.5.1. Bảng tổng hợp

Bảng 4.8. Sáu nhóm ứng dụng AI trọng yếu: cơ chế, dữ liệu, lợi ích và điều kiện

Ứng dụng	Cơ chế hoạt động	Dữ liệu cần có	Chỉ số cải thiện kỳ vọng	Điều kiện tiên quyết	Rủi ro đặc thù
Bảo trì dự báo	Cảm biến rung, nhiệt độ, dòng điện và phân tích dầu đưa vào mô hình dự đoán thời điểm hỏng	Dữ liệu tình trạng liên tục kèm nhãn sự cố lịch sử ; hồ sơ bảo trì	Giảm dừng máy ngoài kế hoạch; tăng thành phần khả dụng; giảm tồn kho phụ tùng	Thiết bị có chế độ hỏng phát triển dần; đủ số ca hỏng lịch sử; năng lực bảo trì đủ để hành động	Thiết bị ít hỏng thì không đủ nhãn; cảnh báo không ai xử lý kịp
Kiểm tra chất lượng bằng thị giác máy	Camera và học sâu phân loại lỗi bề mặt, kích thước, lắp ráp	Ảnh sản phẩm đạt và lỗi đã gán nhãn; điều kiện chiếu sáng ổn định	Tăng tỷ lệ phát hiện lỗi; giảm chi phí chất lượng kém; giảm nhân lực kiểm tra	Chiếu sáng và vị trí đặt sản phẩm ổn định; đủ mẫu cho từng loại lỗi	Lỗi hiếm không đủ mẫu; đổi mã hàng làm mô hình mất hiệu lực
Tối ưu tham số quy trình	Mô hình hồi quy hoặc học tăng cường tìm điểm vận hành tối ưu theo mục tiêu và ràng buộc	Dữ liệu tham số quy trình gắn với kết quả chất lượng và tiêu hao	Giảm tiêu hao nguyên liệu và năng lượng; tăng tỷ lệ đạt lần đầu	Quá trình đã ổn định; phạm vi tham số an toàn được xác định	Mô hình đề xuất điểm vận hành ngoài vùng dữ liệu đã quan sát
Dự báo nhu cầu bằng học máy	Kết hợp dữ liệu nội bộ với tín hiệu bên ngoài	Lịch sử bán hàng đủ dài; dữ liệu khuyến mãi,	Giảm sai số dự báo và độ chệch; giảm	Dữ liệu sự kiện quá khứ được ghi nhận đầy đủ	Sản phẩm mới không có lịch sử; thay đổi cấu

Ứng dụng	Cơ chế hoạt động	Dữ liệu cần có	Chỉ số cải thiện kỳ vọng	Điều kiện tiên quyết	Rủi ro đặc thù
	đã nêu ở Chương 3	giá, thời tiết, sự kiện	tồn kho an toàn		trúc thị trường
Điều độ và tối ưu logistics nội bộ	Tối ưu tổ hợp hoặc học tăng cường trên môi trường mô phỏng	Dữ liệu chủ chính xác gồm thời gian chu kỳ, ma trận chuyển đổi, năng lực nguồn lực	Rút ngắn thời gian chu kỳ và thời gian chờ; tăng tỷ lệ giao đúng hạn	Dữ liệu chủ đã được kiểm toán như đã nêu ở Chương 3	Lịch tối ưu về lý thuyết nhưng không khả thi ở xưởng
AI tạo sinh và trợ lý số cho vận hành	Mô hình ngôn ngữ truy xuất và tổng hợp tài liệu kỹ thuật, soạn quy trình chuẩn, hỗ trợ chẩn đoán	Kho tài liệu kỹ thuật được số hóa, phân quyền và cập nhật	Rút ngắn thời gian xử lý sự cố; chuẩn hóa tri thức; rút ngắn thời gian đào tạo	Tài liệu đúng phiên bản; cơ chế trích dẫn nguồn	Nội dung sai nhưng trình bày thuyết phục; rò rỉ tri thức nội bộ

ĐỊNH VỊ SÁU NHÓM ỨNG DỤNG THEO GIÁ TRỊ VÀ ĐIỀU KIỆN DỮ LIỆU



Hình 4.13. Định vị sáu nhóm ứng dụng theo giá trị và điều kiện dữ liệu

Nguồn: Tác giả xây dựng.

Cách đọc Hình 4.13 có ý nghĩa thực hành trực tiếp với việc chọn thứ tự triển khai. Các ứng dụng nằm ở **nửa trái** đòi hỏi điều kiện dữ liệu dễ đáp ứng hơn, nên là điểm khởi đầu hợp lý cho phần lớn nhà máy Việt Nam. Các ứng dụng nằm ở **nửa phải**, tiêu biểu là bảo trì dự báo, đòi hỏi dữ liệu có nhãn tích lũy trong thời gian dài, nên nếu triển khai sớm sẽ vướng đúng nghịch lý được phân tích ngay dưới đây.

4.5.2. Ba nhóm cần bình luận sâu

Bảo trì dự báo và nghịch lý dữ liệu. Đây là ứng dụng được nhắc tới nhiều nhất nhưng cũng thất bại nhiều nhất, vì một nghịch lý: mô hình dự báo hồng học cần **nhiều ca hồng lịch sử để học**, nhưng thiết bị càng quan trọng thì doanh nghiệp càng bảo dưỡng kỹ và càng ít hỏng. Kết quả là chính những thiết bị đáng đầu tư nhất lại thiếu dữ liệu nhất. Ba cách xử lý thực tế gồm: bắt đầu bằng **phát hiện bất thường** không cần nhãn; gộp dữ liệu của nhiều thiết bị cùng loại để tăng số ca hồng; hoặc chấp nhận dừng ở mức **bảo trì theo tình trạng** với ngưỡng kỹ thuật, vốn đã tạo phần lớn giá trị với chi phí thấp hơn nhiều.

Thị giác máy và bài toán ổn định điều kiện quan sát. Phần lớn thất bại của dự án thị giác máy không nằm ở mô hình mà nằm ở **vật lý của khâu chụp ảnh**: ánh sáng thay đổi theo giờ trong ngày, sản phẩm đặt lệch, bụi bám ống kính, phản xạ trên bề mặt kim loại. Nguyên tắc thực hành là đầu tư vào hệ thống chiếu sáng và cơ cấu định vị sản phẩm trước khi đầu tư vào mô hình. **Một mô hình trung bình với ảnh tốt luôn thắng một mô hình xuất sắc với ảnh kém.**

AI tạo sinh và ranh giới sử dụng an toàn. Trợ lý số dựa trên mô hình ngôn ngữ tạo giá trị rõ ràng ở ba việc: tra cứu nhanh trong kho tài liệu kỹ thuật lớn, soạn thảo bản nháp quy trình chuẩn để chuyên gia hiệu đính, và gợi ý hướng chẩn đoán sự cố dựa trên mô tả triệu chứng. Nhưng đặc điểm không xác định của mô hình khiến nó **không phù hợp để ra quyết định cuối cùng trong các tình huống có rủi ro an toàn hoặc yêu cầu kiểm định**. Nguyên tắc triển khai bắt buộc là hệ thống phải **trích dẫn được nguồn tài liệu** cho mỗi câu trả lời, và người dùng phải được đào tạo rằng câu trả lời là bản nháp cần kiểm chứng chứ không phải kết luận.

NGHIÊN CỨU TÌNH HUỐNG CÓ HƯỚNG DẪN: DỰ ÁN BẢO TRÌ DỰ BÁO

Tình huống. Nhà máy đóng gói có 18 máy chiết rót. Trong 24 tháng qua ghi nhận 31 lần dừng máy ngoài kế hoạch do hỏng bơm, mỗi lần trung bình 6 giờ. Nhà cung cấp đề xuất giải pháp bảo trì dự báo gồm gắn cảm biến rung và nhiệt trên mỗi bơm, phần mềm phân tích trên đám mây, hợp đồng 3 năm.

Nhiệm vụ 1 là xác định dữ liệu. Liệt kê dữ liệu cần có và đánh giá khả năng có sẵn: dữ liệu tình trạng hiện chưa có và phải lắp mới; nhật ký sự cố có ghi ngày giờ, hiện tượng và bộ phận hỏng, cần kiểm tra xem có được ghi chuẩn hóa không; hồ sơ bảo trì và thay thế phụ tùng; dữ liệu tải vận hành và loại sản phẩm đang chạy.

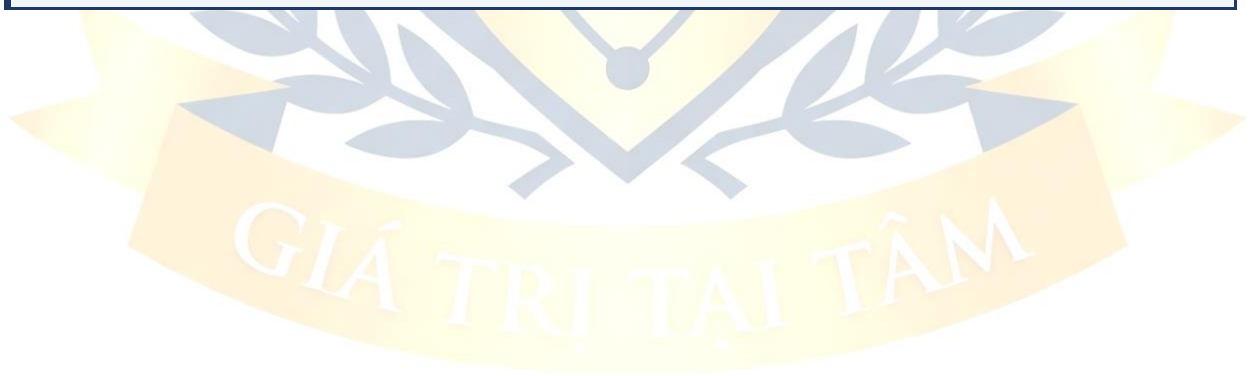
Câu hỏi then chốt: 31 ca hỏng trên 18 máy trong 24 tháng có đủ để huấn luyện mô hình phân loại có giám sát không? Hãy đối chiếu với câu hỏi thứ ba của cây quyết định ở Hình 4.4 và với quy trình gán nhãn ở Hình 4.8.

Nhiệm vụ 2 là nhận diện rào cản. Xem xét ba điểm: nhãn sự cố có đủ chi tiết để phân biệt các chế độ hỏng khác nhau không; tổ bảo trì có năng lực xử lý thêm cảnh báo không; dữ liệu có được phép truyền lên đám mây của nhà cung cấp không và ai sở hữu dữ liệu đó, chiếu theo phần quản trị dữ liệu của Chương 2.

Nhiệm vụ 3 là thiết kế cách đo lợi ích. Xác định chỉ số trước và sau gồm số lần dừng ngoài kế hoạch, tổng giờ dừng, MTTR và khả dụng theo công thức ở Chương 3. Thiết kế nhóm đối chứng bằng cách triển khai trên 9 máy và giữ 9 máy còn lại theo phương thức hiện hành trong 6 tháng để so sánh. **Đây là điểm mà đa số dự án bỏ qua và vì thế không bao giờ chứng minh được lợi ích.**

Nhiệm vụ 4 là đề xuất phương án thay thế. So sánh với phương án chi phí thấp gồm chỉ lắp cảm biến rung với ngưỡng cảnh báo theo tiêu chuẩn rung động và không dùng mô hình học máy; ước tính chi phí và lợi ích của cả hai phương án.

Kết luận cần dẫn dắt: với 31 ca hỏng, bài toán phân loại có giám sát nhiều khả năng thiếu dữ liệu. Lộ trình hợp lý là bắt đầu bằng giám sát tình trạng theo ngưỡng, tích lũy dữ liệu có nhãn trong khoảng một năm tới một năm rưỡi, rồi mới chuyển sang mô hình dự báo.

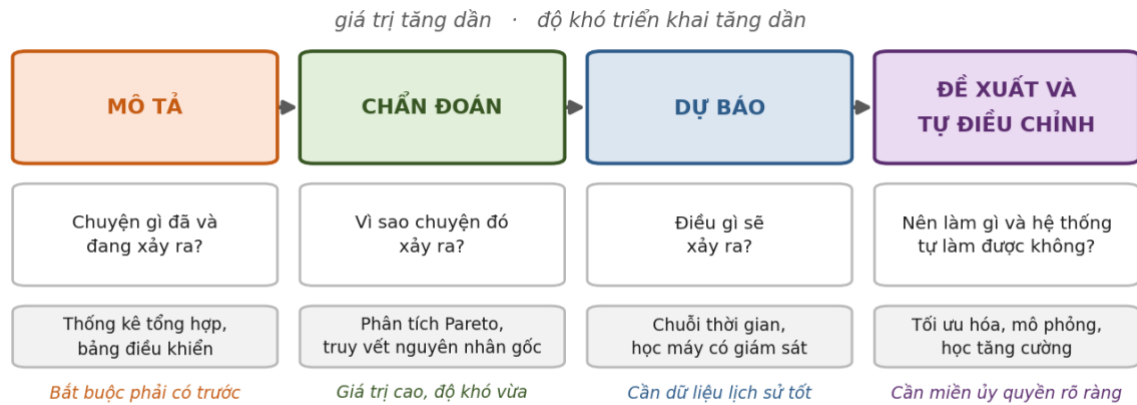


4.6. Bốn cấp độ phân tích và thực hành ra quyết định dựa trên dữ liệu

4.6.1. Bốn cấp độ phân tích

Bốn cấp độ phân tích chính là **trực đọc của khung định vị tình huống ứng dụng** đã học ở Chương 3, và là xương sống của chuẩn đầu ra CLO.3.

BỐN CẤP ĐỘ PHÂN TÍCH TRONG QUAN HỆ GIỮA GIÁ TRỊ VÀ ĐỘ KHÓ



Không thể nhảy cấp. Nhà máy chưa có bảng điều khiển đáng tin cậy ở cấp mô tả thì không thể có mô hình dự báo đáng tin cậy ở cấp ba, vì mô hình cấp ba học từ chính dữ liệu mà cấp một tạo ra.

Hình 4.14. Bốn cấp độ phân tích trong quan hệ giữa giá trị và độ khó triển khai

Nguồn: Tác giả xây dựng, tương thích với trực đọc của khung định vị tình huống ứng dụng trình bày tại Chương 3.

Bảng 4.9. Bốn cấp độ phân tích dữ liệu trong vận hành

Cấp độ	Câu hỏi	Kỹ thuật điển hình	Ví dụ trong nhà máy và mức độ khó
Mô tả	Chuyện gì đã và đang xảy ra	Thống kê tổng hợp, bảng điều khiển, biểu đồ xu hướng	Hiệu suất thiết bị theo máy theo ca; sản lượng theo giờ. Giá trị nền tảng, độ khó thấp, bắt buộc phải có trước
Chẩn đoán	Vì sao chuyện đó xảy ra	Phân tích Pareto, phân tích tương quan, truy vết nguyên nhân gốc	Biểu đồ Pareto nguyên nhân dừng máy; so sánh hiệu năng giữa các ca. Giá trị cao, độ khó trung bình, đòi hỏi dữ liệu có ngữ cảnh

Cấp độ	Câu hỏi	Kỹ thuật điển hình	Ví dụ trong nhà máy và mức độ khó
Dự báo	Điều gì sẽ xảy ra	Mô hình chuỗi thời gian, học máy có giám sát	Dự báo nhu cầu; cảnh báo hỏng hóc; dự đoán trễ đơn hàng. Giá trị cao, độ khó cao, đòi hỏi dữ liệu lịch sử chất lượng
Đề xuất và tự điều chỉnh	Nên làm gì, và hệ thống tự làm được không	Tối ưu hóa, mô phỏng, học tăng cường	Tái lập lịch tự động; tự điều chỉnh tham số trong miền cho phép. Giá trị cao nhất, độ khó cao nhất, đòi hỏi miền ủy quyền rõ ràng

Nguyên tắc trình tự đã nêu ở các chương trước cần được nhắc lại ở đây với một sắc thái mạnh hơn: **không thể nhảy cấp**. Một nhà máy chưa có bảng điều khiển đáng tin cậy ở cấp mô tả thì không thể có mô hình dự báo hỏng hóc đáng tin cậy ở cấp ba, vì mô hình cấp ba học từ chính dữ liệu mà cấp một tạo ra. Đây là lý do phần thực hành của học phần bắt đầu từ cấp mô tả và cấp chẩn đoán, hai cấp tạo ra phần lớn giá trị ban đầu với chi phí thấp nhất.

4.6.2. Bài thực hành bắt buộc trên dữ liệu ca sản xuất

Nội dung dưới đây là hướng dẫn cho buổi thực hành 90 phút. Sinh viên làm việc trên tập dữ liệu ca sản xuất mô phỏng do giảng viên cung cấp, sử dụng phần mềm bảng tính hoặc công cụ trực quan hóa dữ liệu.

Bảng 4.10. Bốn nhiệm vụ của bài thực hành và tiêu chí đánh giá

Nhiệm vụ	Nội dung	Cấp độ phân tích	Tiêu chí đánh giá
a. Tính hiệu suất thiết bị và tách tổn thất	Từ dữ liệu thời gian kế hoạch, thời gian chạy, sản lượng và sản phẩm lỗi, tính khả dụng, hiệu suất, chất lượng; quy đổi ba nhóm tổn thất về cùng một đơn vị	Mô tả	Công thức đúng; quy đổi tồn thất đúng đơn vị; xác định đúng tổn thất lớn nhất

Nhiệm vụ	Nội dung	Cấp độ phân tích	Tiêu chí đánh giá
b. Lập biểu đồ Pareto nguyên nhân dừng máy	Sắp xếp nguyên nhân theo tổng thời gian dừng và theo số lần dừng; vẽ đường tích lũy	Chẩn đoán	Hai cách sắp xếp được trình bày và so sánh; nhận xét về ý nghĩa khác nhau của hai cách
c. Xây dựng bảng điều khiển	Trình bày các chỉ số then chốt cho một đối tượng người dùng cụ thể là quản đốc ca hoặc giám đốc nhà máy	Mô tả và chẩn đoán	Chọn đúng chỉ số cho đúng người dùng; không quá 6 chỉ số; mỗi chỉ số gắn với một hành động khả thi
d. Đề xuất hành động cải tiến	Chọn một tổn thất, đề xuất biện pháp, ước lượng mức cải thiện chỉ số và giá trị bằng tiền	Chẩn đoán hướng tới đề xuất	Lập luận nhân quả rõ ràng; giả định được nêu tường minh; con số ước lượng có căn cứ

Hai lỗi phổ biến mà giảng viên cần lưu ý sinh viên. *Lỗi thứ nhất* là **bảng điều khiển nhồi nhét**, tức đưa mọi chỉ số có thể tính được lên màn hình. Nguyên tắc thiết kế đúng là mỗi chỉ số phải trả lời câu hỏi: nếu chỉ số này xấu đi thì người xem sẽ làm gì? Nếu không có câu trả lời, chỉ số đó không thuộc về bảng điều khiển này. *Lỗi thứ hai* là **ước lượng lợi ích không có giả định tường minh**, chẳng hạn viết cải thiện hiệu suất thiết bị 5 phần trăm mà không nói rõ mức đó đến từ thành phần nào, bằng cơ chế nào và dựa trên căn cứ gì.

4.7. Luận chứng kinh tế cho một dự án AI

4.7.1. Cấu trúc chi phí đầy đủ

Khác biệt lớn nhất giữa luận chứng kinh tế của một dự án AI và của một dự án thiết bị nằm ở chỗ **chi phí của dự án AI không kết thúc khi hệ thống được nghiệm thu**. Mô hình cần được giám sát, tái huấn luyện và hiệu chỉnh suốt vòng đời, và chi phí này thường bị bỏ sót hoàn toàn trong đề xuất ban đầu.



Hình 4.15. Cấu trúc chi phí đầy đủ của một dự án AI trong nhà máy

Nguồn: Tác giả xây dựng, có tham chiếu phân tích nợ kỹ thuật ẩn trong hệ thống học máy của Sculley và cộng sự (2015).

Bảng 4.11. Cấu trúc chi phí đầy đủ và mức độ bị bỏ sót

Nhóm chi phí	Hạng mục	Mức độ bị bỏ sót
Đầu tư ban đầu	Cảm biến, camera, chiếu sáng, thiết bị tính toán biên	Thấp, luôn được tính
Đầu tư ban đầu	Tích hợp với hệ thống hiện có	Cao , thường bị đánh giá thấp
Đầu tư ban đầu	Thu thập và gán nhãn dữ liệu huấn luyện	Rất cao , thường bị bỏ qua hoàn toàn
Đầu tư ban đầu	Thiết kế lại quy trình làm việc và đào tạo người dùng	Cao

Nhóm chi phí	Hạng mục	Mức độ bị bỏ sót
Vận hành thường xuyên	Hạ tầng tính toán, lưu trữ, bản quyền phần mềm	Thấp
Vận hành thường xuyên	Giám sát và tái huấn luyện mô hình	Rất cao , hạng mục bị bỏ sót nhiều nhất
Vận hành thường xuyên	Nhân lực xử lý cảnh báo và ngoại lệ	Cao
Vận hành thường xuyên	Bảo trì cảm biến, hiệu chuẩn, vệ sinh thiết bị quang học	Trung bình
Chi phí ẩn	Thời gian mất năng suất trong giai đoạn chuyển tiếp	Cao

VÍ DỤ XÂY DỰNG LUẬN CHỨNG KINH TẾ CHO DỰ ÁN THỊ GIÁC MÁY

Bối cảnh. Nhà máy sản xuất 120.000 sản phẩm mỗi năm, tỷ lệ lỗi thực tế 2 phần trăm, tức 2.400 sản phẩm lỗi mỗi năm. Kiểm tra thủ công hiện phát hiện được 70 phần trăm số lỗi, tức 720 sản phẩm lỗi lọt tới khách hàng mỗi năm, đồng thời loại nhầm khoảng 1 phần trăm hàng tốt. Hai kiểm tra viên toàn thời gian với chi phí đầy đủ 11 triệu đồng mỗi người mỗi tháng. Thiệt hại trung bình của một sản phẩm lỗi lọt tới khách hàng là 2 triệu đồng; chi phí xử lý một sản phẩm tốt bị loại nhầm là 0,2 triệu đồng.

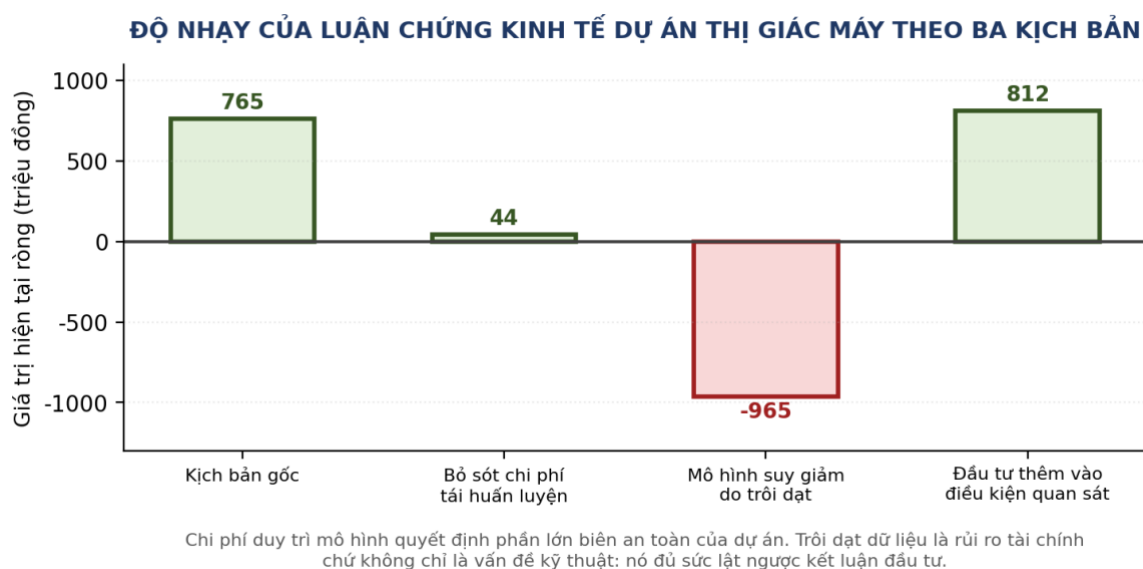
Phương án đầu tư. Hệ thống thị giác máy nâng tỷ lệ phát hiện lỗi lên 95 phần trăm, nhưng tỷ lệ loại nhầm hàng tốt tăng lên 3 phần trăm. Đầu tư ban đầu 1.800 triệu đồng gồm camera, chiếu sáng, máy tính biên, tích hợp, gán nhãn dữ liệu và đào tạo. Chi phí vận hành 150 triệu đồng mỗi năm gồm hạ tầng, hiệu chuẩn, giám sát và tái huấn luyện mô hình. Sau triển khai vẫn giữ 1 kiểm tra viên để xử lý cảnh báo. Vòng đời 5 năm, suất chiết khấu 12 phần trăm mỗi năm, hệ số hiện giá dòng tiền đều 5 năm là 3,6048.

Tính lợi ích. *Giảm lỗi lọt:* trước là 720 sản phẩm, sau là 120 sản phẩm; tiết kiệm bằng tích của 600 với 2, tức **1.200 triệu đồng mỗi năm**. *Tăng chi phí loại nhầm:* hàng tốt là 117.600 sản phẩm; trước 1 phần trăm là 1.176 sản phẩm, sau 3 phần trăm là 3.528 sản phẩm; phát sinh thêm bằng tích của 2.352 với 0,2, tức **470,4 triệu đồng mỗi năm**. *Giảm nhân lực:* một kiểm tra viên tương ứng **132 triệu đồng mỗi năm**.

Dòng tiền. Lợi ích ròng năm bằng 1.200 trừ 470,4 cộng 132 trừ 150, tức **711,6 triệu đồng mỗi năm**. Thời gian hoàn vốn giản đơn bằng 1.800 chia 711,6, xấp xỉ **2,53 năm**. Giá trị hiện tại ròng bằng tích của 711,6 với 3,6048 rồi trừ 1.800, xấp xỉ **cộng 765 triệu đồng**. Dự án khả thi.

Phân tích độ nhạy với ba kịch bản. *Kịch bản 1 là bỏ sót chi phí tái huấn luyện.* Nếu chi phí vận hành thực tế là 350 triệu thay vì 150 triệu, lợi ích ròng còn 511,6 và giá trị hiện tại ròng xấp xỉ **cộng 44 triệu đồng**, gần như hòa vốn. *Kịch bản 2 là mô hình suy giảm do trôi dạt.* Nếu sau 18 tháng tỷ lệ phát hiện giảm từ 95 xuống 85 phần trăm vì không tái huấn luyện, lỗi lọt tăng lên 360 sản phẩm mỗi năm, lợi ích giảm 480 triệu và **giá trị hiện tại ròng chuyển sang âm**. *Kịch bản 3 là giảm dương tính giả.* Nếu đầu tư thêm vào chiếu sáng và cơ cấu định vị để hạ tỷ lệ loại nhầm từ 3 xuống 1,5 phần trăm, tiết kiệm thêm khoảng 235 triệu mỗi năm, đủ bù cho chi phí tái huấn luyện đầy đủ ở Kịch bản 1.

Kết luận quản trị. Ba bài học. *Một*, chi phí duy trì mô hình quyết định phần lớn biên an toàn của dự án, nên bỏ sót nó là bỏ sót yếu tố có thể lật ngược kết luận. *Hai*, trôi dạt dữ liệu là rủi ro tài chính chứ không chỉ là vấn đề kỹ thuật, đúng như đường cong ở Hình 4.7 đã minh họa. *Ba*, khoản đầu tư bổ sung vào điều kiện quan sát vật lý có tỷ suất hoàn vốn cao hơn khoản đầu tư vào mô hình tinh vi hơn.



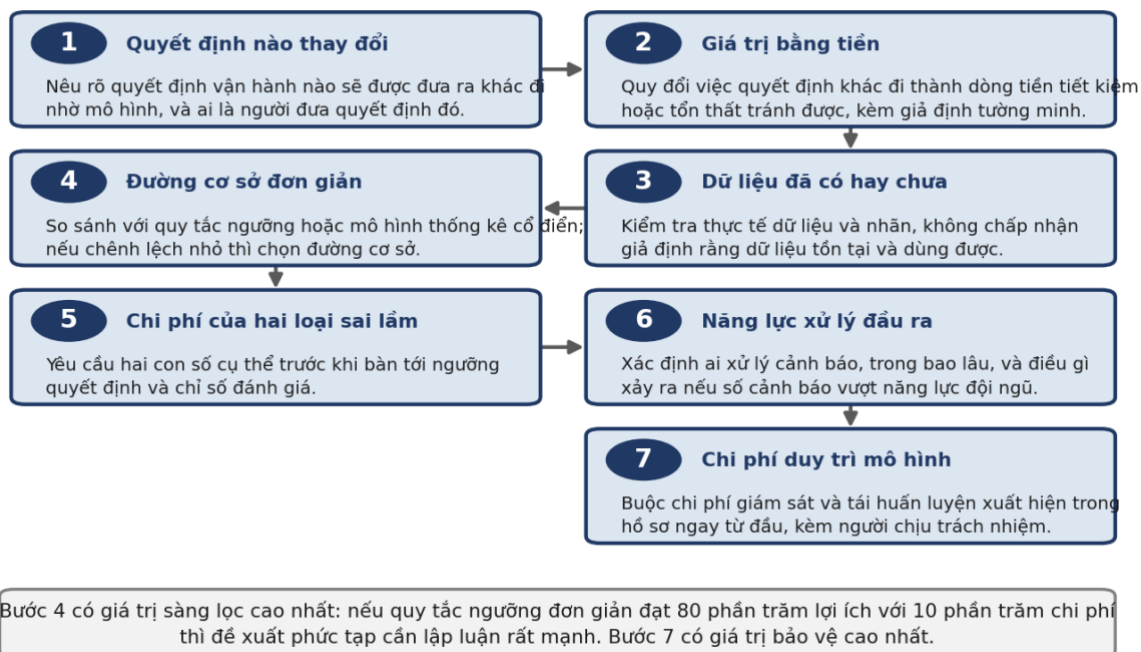
Hình 4.16. Độ nhạy của luận chứng kinh tế theo ba kịch bản

Nguồn: Tác giả tính toán và biên vẽ theo số liệu của ví dụ trên.

4.7.2. Quy trình thẩm định một đề xuất dự án AI

Khi tiếp nhận một đề xuất dự án AI, nhà quản trị nên yêu cầu câu trả lời cho bảy câu hỏi theo đúng thứ tự dưới đây. Thứ tự có ý nghĩa: các câu hỏi đầu sàng lọc những đề xuất không có bài toán thực, còn các câu hỏi cuối sàng lọc những đề xuất có bài toán nhưng thiếu năng lực thực thi.

QUY TRÌNH BẢY BƯỚC THẨM ĐỊNH MỘT ĐỀ XUẤT DỰ ÁN AI



Hình 4.17. Quy trình bảy bước thẩm định một đề xuất dự án AI

Nguồn: Tác giả xây dựng.

Câu hỏi ở bước 4 có **giá trị sàng lọc cao nhất**: nếu một quy tắc ngưỡng đơn giản đạt 80 phần trăm lợi ích với 10 phần trăm chi phí, thì đề xuất mô hình phức tạp cần lập luận rất mạnh mới được chấp nhận. Câu hỏi ở bước 7 có **giá trị bảo vệ cao nhất**, vì nó buộc chi phí duy trì mô hình xuất hiện trong hồ sơ ngay từ đầu thay vì phát sinh sau khi ngân sách đã đóng.

4.8. Quản trị trí tuệ nhân tạo có trách nhiệm

4.8.1. Ba yêu cầu cốt lõi

Tính giải thích được. Trong môi trường sản xuất, một quyết định không giải thích được là một quyết định không thể cải tiến và không thể quy trách nhiệm. Khi mô hình cảnh báo một máy sắp hỏng, kỹ thuật viên cần biết cảnh báo dựa trên dấu hiệu nào để kiểm tra đúng chỗ. Khi hệ thống loại một lô sản phẩm, bộ phận chất lượng cần biết đặc trưng nào dẫn tới kết luận đó để truy nguyên nhân gốc. Đây là lý do trong nhiều tình huống nhà máy, **một mô hình đơn giản giải thích được có giá trị vận hành cao hơn một mô hình phức tạp chính xác hơn vài phần trăm nhưng không giải thích được**, luận điểm đã được Rudin (2019) trình bày một cách hệ thống.

Thiên lệch dữ liệu. Mô hình học từ dữ liệu quá khứ và vì vậy tái tạo cả những khiếm khuyết của quá khứ. Trong nhà máy, thiên lệch xuất hiện dưới các dạng cụ thể: dữ liệu chỉ thu thập ở ca ngày nên mô hình hoạt động kém ở ca đêm; dữ liệu chỉ từ một dòng sản phẩm nên mô hình sai với dòng khác; nhãn lỗi do một kiểm tra viên gán nên mô hình học theo tiêu chuẩn chủ quan của người đó, đúng vấn đề mà bước 3 của quy trình gán nhãn ở Hình 4.8 nhằm phát hiện. Thiên lệch cũng có chiều liên quan tới con người: mô hình đánh giá năng suất cá nhân có thể phản ánh sự khác biệt về máy móc và ca làm việc hơn là về năng lực người lao động, và dùng nó để xét thưởng phạt là một sai lầm quản trị nghiêm trọng.

Trách nhiệm giải trình. Câu hỏi cần trả lời trước khi triển khai chứ không phải sau khi xảy ra sự cố: khi mô hình sai và gây thiệt hại hoặc mất an toàn thì ai chịu trách nhiệm, người phê duyệt triển khai, người vận hành đã tin theo khuyến nghị, hay nhà cung cấp mô hình? Câu trả lời phải nằm trong hợp đồng và trong quy trình nội bộ.

4.8.2. Khung tham chiếu quốc tế và ràng buộc pháp lý

ISO/IEC 42001 là tiêu chuẩn về hệ thống quản lý trí tuệ nhân tạo, cung cấp khung quản lý vòng đời hệ thống AI trong tổ chức gồm chính sách, phân định vai trò và trách nhiệm, đánh giá rủi ro và tác động, quản lý dữ liệu, giám sát hiệu năng và cải tiến liên tục. Đây là công cụ hữu ích cho doanh nghiệp muốn chứng minh năng lực quản trị AI với khách hàng quốc tế.

Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu, tức Quy định (EU) 2024/1689, phân loại hệ thống AI theo mức rủi ro và áp nghĩa vụ tương ứng.

BỐN MỨC RỦI RO THEO ĐẠO LUẬT TRÍ TUỆ NHÂN TẠO CỦA LIÊN MINH CHÂU ÂU

RỦI RO KHÔNG CHẤP NHẬN ĐƯỢC	Bị cấm hoàn toàn	Chấm điểm xã hội; thao túng hành vi; nhận dạng cảm xúc tại nơi làm việc
RỦI RO CAO	Nghĩa vụ nặng: hồ sơ kỹ thuật, quản lý rủi ro, chất lượng dữ liệu, giám sát của con người	AI là thành phần an toàn của máy móc; dùng trong tuyển dụng và đánh giá lao động
RỦI RO GIỚI HẠN	Nghĩa vụ minh bạch	Trợ lý số trả lời người dùng phải cho biết đang tương tác với hệ thống AI
RỦI RO TỐI THIỂU	Không có nghĩa vụ riêng	Bảo trì dự báo; kiểm tra chất lượng bằng thị giác máy; tối ưu điều độ

Phần lớn ứng dụng AI trong nhà máy thuộc nhóm rủi ro tối thiểu. Doanh nghiệp Việt Nam chịu ràng buộc gián tiếp qua hai đường: sản phẩm có tích hợp AI đưa vào thị trường châu Âu, và yêu cầu chứng minh tuân thủ từ khách hàng.

Hình 4.18. Bốn mức rủi ro theo Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu

Nguồn: Tác giả biên vẽ theo Quy định (EU) 2024/1689.

Doanh nghiệp sản xuất Việt Nam chịu ràng buộc gián tiếp qua hai đường: khi sản phẩm có tích hợp AI được đưa vào thị trường châu Âu, và khi khách hàng châu Âu yêu cầu chứng minh tuân thủ trong chuỗi cung ứng. Điểm đáng lưu ý về mặt thực hành là **phần lớn ứng dụng AI trong nhà máy thuộc nhóm rủi ro tối thiểu**, nên gánh nặng tuân thủ thấp hơn nhiều so với hình dung phổ biến. Rủi ro pháp lý thực sự tập trung ở hai nhóm hẹp: AI là thành phần an toàn của máy móc, và AI dùng trong tuyển dụng hoặc đánh giá người lao động.

Về mốc thời gian áp dụng, trạng thái tại thời điểm biên soạn là tháng 8 năm 2026. Các nghĩa vụ về minh bạch, trong đó có yêu cầu cho người dùng biết họ đang tương tác với hệ thống AI, áp dụng từ ngày 2 tháng 8 năm 2026. Đối với hệ thống rủi ro cao, lịch trình đang được điều chỉnh: gói sửa đổi Digital Omnibus đề xuất lùi thời hạn tuân thủ đối với hệ thống rủi ro cao thuộc Phụ lục III sang tháng 12 năm 2027, và đối với AI tích hợp trong sản phẩm thuộc Phụ lục I sang tháng 8 năm 2028. Một thỏa thuận sơ bộ giữa các cơ quan lập pháp châu Âu đạt được vào tháng 5 năm 2026, nhưng cần được thông qua chính thức mới có hiệu lực pháp lý.

Lưu ý bắt buộc: đây là lĩnh vực pháp lý đang thay đổi. Giảng viên phải rà soát trạng thái hiện hành trước mỗi học kỳ; người học không được sử dụng các mốc trên cho mục đích tuân thủ thực tế mà không kiểm chứng lại.

4.8.3. Nguyên tắc phân bổ mức tự chủ

Nguyên tắc thực hành tổng hợp lại toàn bộ mục này: **mức tự chủ trao cho hệ thống AI phải tỷ lệ nghịch với hậu quả của sai lầm và tỷ lệ thuận với khả năng đảo ngược của hành động.**

MA TRẬN PHÂN BỐ MỨC TỰ CHỦ CHO HỆ THỐNG AI TRONG NHÀ MÁY



Hình 4.19. Ma trận phân bố mức tự chủ cho hệ thống AI trong nhà máy

Nguồn: Tác giả xây dựng.

Ma trận cho biết *mức* tự chủ phù hợp, nhưng chưa cho biết *cách* hiện thực hóa nó trong kiến trúc hệ thống. Sơ đồ dưới đây bổ sung phần còn thiếu bằng cách đặt các điểm phê duyệt vào đúng vị trí trên chuỗi từ dữ liệu tới hành động.

KIẾN TRÚC CON NGƯỜI TRONG VÒNG LẶP VÀ CÁC ĐIỂM PHÊ DUYỆT BẮT BUỘC



Hình 4.20. Kiến trúc con người trong vòng lặp và các điểm phê duyệt bắt buộc

Nguồn: Tác giả xây dựng.

Ba điểm thiết kế cần nhấn mạnh. *Thứ nhất*, khuyến nghị của mô hình phải đi kèm giải thích, nếu không thì điểm phê duyệt thứ nhất trở thành một thao tác bấm nút hình thức. *Thứ hai*, điểm phê duyệt thứ hai đặt trước hành động chứ không sau, vì với hành động không đảo ngược được thì việc phát hiện sai sau khi thực hiện không còn giá trị. *Thứ ba*, mỗi lần phê duyệt hoặc bác bỏ đều được M.Econ Dang Thien Tam

ghi nhận, tạo ra đồng thời hai giá trị là dữ liệu huấn luyện cho chu kỳ sau và hồ sơ trách nhiệm giải trình khi cần truy vết quyết định.

4.8.4. Quy trình sáu bước triển khai AI có trách nhiệm

QUY TRÌNH SÁU BƯỚC TRIỂN KHAI AI CÓ TRÁCH NHIỆM TRONG NHÀ MÁY



Hình 4.21. Quy trình sáu bước triển khai AI có trách nhiệm trong nhà máy

Nguồn: Tác giả xây dựng, có tham chiếu ISO/IEC 42001 và Quy định (EU) 2024/1689.

Bước 1 phải đi trước thiết kế chứ không phải sau khi hệ thống đã vận hành, vì nghĩa vụ của nhóm rủi ro cao bao gồm cả những yêu cầu về hồ sơ kỹ thuật và chất lượng dữ liệu **không thể bổ sung hồi tố**. Một hệ thống được xây dựng mà không lưu vết nguồn gốc dữ liệu huấn luyện sẽ không thể chứng minh tuân thủ về sau, dù dữ liệu đó trên thực tế có chất lượng tốt.

4.9. Giới hạn của trí tuệ nhân tạo trong nhà máy

Ba giới hạn cấu trúc cần được người học nắm vững, vì nhận ra chúng là điều kiện để tránh đầu tư sai.

Giới hạn thứ nhất là phụ thuộc chất lượng và độ phủ của dữ liệu lịch sử. Mô hình chỉ biết những gì đã xuất hiện trong dữ liệu huấn luyện. Nếu nhà máy chưa từng ghi nhận một chế độ hỏng nào đó, mô hình sẽ không phát hiện được chế độ hỏng ấy. Đây không phải là khiếm khuyết có thể khắc phục bằng thuật toán tốt hơn; đó là **giới hạn nguyên lý**.

Giới hạn thứ hai là khó ngoại suy sang tình huống chưa từng gặp. Mô hình hoạt động đáng tin cậy trong vùng dữ liệu đã quan sát và trở nên bất định khi ra ngoài vùng đó. Điều này đặc biệt nguy hiểm với ứng dụng tối ưu tham số quy trình: mô hình có thể đề xuất một điểm vận hành chưa từng được thử, nơi dự đoán của nó không có căn cứ. Biện pháp bắt buộc là **giới hạn miền đề xuất trong phạm vi dữ liệu đã quan sát** và thử nghiệm từng bước nhỏ.

Giới hạn thứ ba là chi phí duy trì và sự lão hóa của mô hình. Như ví dụ ở mục 4.7 và đường cong ở Hình 4.7 đã cho thấy, một mô hình không được tái huấn luyện sẽ suy giảm âm thầm và có thể lật ngược toàn bộ luận chứng kinh tế của dự án.

Bảng 4.12. Khi nào không nên dùng trí tuệ nhân tạo

Tình huống	Vì sao AI không phù hợp	Nên làm gì thay thế
Quy trình chưa chuẩn hóa và chưa ổn định	Dữ liệu có phương sai lớn và nhiều ngoại lệ; mô hình học nhiều	Áp dụng trình tự tinh gọn 4.0 ở Chương 2: ổn định trước, số hóa sau
Quan hệ nhân quả đã biết rõ và đơn giản	Quy tắc ngưỡng cho kết quả tương đương, rõ hơn và giải thích được	Dùng quy tắc kỹ thuật hoặc kiểm soát quá trình bằng thống kê
Dữ liệu lịch sử quá ít hoặc không có nhãn	Không đủ cơ sở học; mô hình cho kết quả ngẫu nhiên nhưng trông có vẻ hợp lý	Thu thập dữ liệu trước; dùng phát hiện bất thường không cần nhãn
Không ai xử lý được đầu ra của mô hình	Cảnh báo không dẫn tới hành động thì không tạo giá trị	Xây dựng năng lực và quy trình xử lý trước khi đầu tư mô hình
Yêu cầu kiểm định nghiêm ngặt và kết quả phải tái lập được	Mô hình không xác định không đáp ứng yêu cầu kiểm định	Dùng phương pháp xác định; hoặc dùng AI ở vai trò hỗ trợ, con người quyết định

Tình huống	Vì sao AI không phù hợp	Nên làm gì thay thế
Tần suất ra quyết định rất thấp	Chi phí xây dựng và duy trì mô hình không được bù đắp	Phân tích thủ công có cấu trúc, có sự tham gia của chuyên gia

ĐIỂM CẦN NHỚ

Câu hỏi hay nhất mà một nhà quản trị có thể đặt cho đề xuất ứng dụng AI không phải là mô hình này chính xác đến đâu, mà là ba câu hỏi nổi tiếp:

Quyết định nào sẽ được đưa ra khác đi nhờ mô hình này? Giá trị bằng tiền của việc quyết định khác đi là bao nhiêu? Và nếu bỏ mô hình đi, chúng ta mất gì đo được?

Nếu ba câu hỏi này không có câu trả lời cụ thể, đề xuất chưa sẵn sàng để phê duyệt, bất kể công nghệ được đề xuất tiên tiến đến đâu.

4.10. Tổng hợp: lộ trình bốn lớp năng lực

LỘ TRÌNH BỐN LỚP NĂNG LỰC TRIỂN KHAI AI TRONG NHÀ MÁY (GÓC NHÌN QUẢN TRỊ)

mỗi lớp đánh giá mức độ sẵn sàng và giá trị, nên không thể nhảy cóc



Hình 4.22. Lộ trình bốn lớp năng lực triển khai AI trong nhà máy

Nguồn: Tác giả tổng hợp từ nội dung Chương 4.

Lớp thứ nhất là nền dữ liệu. Chuẩn hóa định nghĩa chỉ số, kiểm toán chất lượng dữ liệu và xây dựng quy trình gán nhãn. Lớp này gần như không chứa thành phần AI nào nhưng quyết định toàn bộ giá trị của ba lớp sau.

Lớp thứ hai là phân tích mô tả và chẩn đoán. Bảng điều khiển đáng tin cậy, phân tích Pareto nguyên nhân và đường cơ sở đơn giản. Đây là lớp tạo ra phần lớn giá trị ban đầu với chi phí thấp nhất, và cũng là lớp cung cấp mốc so sánh bắt buộc cho mọi mô hình phức tạp về sau.

Lớp thứ ba là mô hình dự báo. Phát hiện bất thường, bảo trì dự báo, thị giác máy, kèm theo quy trình giám sát và tái huấn luyện. Lớp này chỉ khả thi khi lớp thứ nhất đã tạo ra dữ liệu có nhãn đủ tin cậy.

Lớp thứ tư là đề xuất và tự điều chỉnh. Tối ưu tham số quy trình, điều độ tự động, với kiến trúc con người trong vòng lặp được thiết kế theo hậu quả và khả năng đảo ngược.

Thứ tự này không thể đảo, vì **mỗi lớp học từ chính dữ liệu do lớp trước tạo ra**. Đây là dạng cụ thể của nguyên tắc mắt xích yếu nhất đã nêu từ Chương 1, áp dụng cho lĩnh vực trí tuệ nhân tạo.

4.11. Tóm tắt chương

Chương 4 đã bổ sung lớp năng lực cuối cùng của nhà máy thông minh là năng lực biến dữ liệu thành quyết định, thông qua bốn năng lực quản trị nêu ở Hình 4.1.

Về bản đồ khái niệm, trí tuệ nhân tạo, học máy, học sâu và AI tạo sinh có quan hệ bao hàm, còn AI tác tử là một kiến trúc sử dụng chứ không phải một tầng công nghệ mới. Hệ quả quan trọng là hiện tượng lạm phát kỹ thuật: lợi ích bão hòa nhanh trong khi chi phí tiếp tục tăng, nên mức phức tạp tối ưu hiếm khi nằm ở lựa chọn tiên tiến nhất. Ba dạng sử dụng AI khác nhau về tính xác định của kết quả và về yêu cầu giám sát. Cây quyết định bốn câu hỏi ở Hình 4.4 chuẩn hóa việc chọn nhóm bài toán, trong đó chỉ dẫn quan trọng nhất là khi chưa có nhãn đầy đủ thì phát hiện bất thường là điểm khởi đầu hợp lý hơn phân loại có giám sát.

Về vòng đời dự án dữ liệu, khoảng 65 phần trăm công sức nằm ở hai bước hiểu và chuẩn bị dữ liệu chứ không ở mô hình hóa, và quy trình mang tính lặp chứ không tuyến tính. Bốn vấn đề chất lượng dữ liệu đặc thù là dữ liệu thiếu, nhiễu cảm biến, mất cân bằng lớp và trôi dạt. Mất cân bằng lớp khiến độ chính xác trở thành chỉ số vô dụng, minh chứng là mô hình luôn tuyên bố mọi sản phẩm đều đạt lại có độ chính xác cao hơn mô hình thật. Trôi dạt dữ liệu gây suy giảm âm thầm không có tín hiệu báo động, và trong ví dụ minh họa thì mô hình vượt ngưỡng chấp nhận được ở khoảng tháng thứ mười sáu sau khi đã kém đúng nhiều tháng trước đó. Quy trình sáu bước gán nhãn ở Hình 4.8 xử lý nút thắt tồn kém nhất, với nguyên tắc rằng mức đồng thuận giữa những người gán nhãn chính là giới hạn trên của độ chính xác mà mô hình có thể đạt được. Quy trình sáu bước giám sát ở Hình 4.9 có bước đặt ngưỡng kích hoạt là bước giá trị nhất và bị bỏ qua nhiều nhất.

Về đánh giá mô hình, ma trận nhầm lẫn sinh ra năm chỉ số trả lời những câu hỏi khác nhau, trong đó độ chính xác vô dụng khi dữ liệu mất cân bằng và điểm F1 giả định hai loại sai lầm có chi phí ngang nhau, điều thường không đúng. Không tồn tại ngưỡng đúng về mặt kỹ thuật mà chỉ tồn tại

ngưỡng tối ưu về chi phí; đường tổng chi phí có dạng chữ U với vùng tối ưu khá phẳng nhưng vùng bên phải dốc lên rất nhanh. Nhiệm vụ của nhà quản trị không phải là yêu cầu mô hình chính xác hơn mà là cung cấp hai con số chi phí cho nhóm kỹ thuật, kèm kiểm tra xem ngưỡng tối ưu về chi phí có khả thi về mặt vận hành hay không.

Về ứng dụng và luận chứng kinh tế, sáu nhóm ứng dụng trọng yếu được định vị theo giá trị và mức khắt khe của điều kiện dữ liệu; bảo trì dự báo vướng nghịch lý rằng thiết bị đáng đầu tư nhất lại thiếu dữ liệu nhất; thị giác máy thất bại chủ yếu ở vật lý khâu chụp ảnh chứ không ở mô hình. Bốn cấp độ phân tích không thể nhảy cấp. Cấu trúc chi phí đầy đủ cho thấy gán nhãn dữ liệu và tái huấn luyện mô hình là hai hạng mục bị bỏ sót nhiều nhất; phân tích ba kịch bản chứng minh rằng bỏ sót chi phí tái huấn luyện đưa dự án về gần điểm hòa vốn, còn để mô hình suy giảm do trôi dạt thì lật ngược hẳn kết luận đầu tư. Quy trình bảy bước ở Hình 4.17 có bước so sánh với đường cơ sở đơn giản mang giá trị sàng lọc cao nhất.

Về quản trị có trách nhiệm, ba yêu cầu cốt lõi là tính giải thích được, kiểm soát thiên lệch dữ liệu và phân định trách nhiệm giải trình. Phần lớn ứng dụng AI trong nhà máy thuộc nhóm rủi ro tối thiểu theo phân loại pháp lý hiện hành, nhưng nghĩa vụ của nhóm rủi ro cao không thể bỏ sung hời hợt nên việc phân loại phải đi trước thiết kế. Mức tự chủ trao cho hệ thống phải tỷ lệ nghịch với hậu quả và tỷ lệ thuận với khả năng đảo ngược, và được hiện thực hóa bằng hai điểm phê duyệt đặt đúng vị trí trên chuỗi từ dữ liệu tới hành động.

Ba giới hạn cấu trúc của AI là phụ thuộc dữ liệu lịch sử, khó ngoại suy và chi phí duy trì. Toàn bộ nội dung được tổng hợp thành lộ trình bốn lớp năng lực ở Hình 4.22, với nguyên tắc khép lại chương: **quyết định nào sẽ được đưa ra khác đi, giá trị bằng tiền của việc quyết định khác đi là bao nhiêu, và nếu bỏ mô hình đi thì chúng ta mất gì đo được.**

4.12. Thuật ngữ trọng tâm

Bảng 4.13. Thuật ngữ trọng tâm Chương 4

Thuật ngữ	Giải thích ngắn
Artificial intelligence	Trí tuệ nhân tạo; khái niệm rộng nhất, bao hàm học máy và học sâu.
Machine learning và deep learning	Học máy rút quy luật từ dữ liệu; học sâu là nhánh dùng mạng nơ ron nhiều lớp.
Generative AI	AI tạo sinh; tạo nội dung mới, kết quả không xác định giữa các lần chạy.

Thuật ngữ	Giải thích ngắn
Agentic AI	AI tác tử; một kiến trúc sử dụng chứ không phải một tầng công nghệ mới.
Industrial foundation model	Mô hình nền công nghiệp; phạm vi hẹp, triển khai được tại biên.
Lạm phát kỹ thuật	Chọn công cụ phức tạp hơn mức cần thiết; lợi ích bão hòa trong khi chi phí tăng nhanh.
Supervised và unsupervised learning	Học có giám sát cần nhãn; học không giám sát tìm cấu trúc ẩn.
Anomaly detection	Phát hiện bất thường; điểm khởi đầu hợp lý khi chưa có nhãn sự cố.
Reinforcement learning	Học tăng cường; cần môi trường mô phỏng đáng tin cậy.
CRISP DM	Vòng đời sáu bước của dự án dữ liệu; khoảng 65 phần trăm công sức ở hai bước dữ liệu.
Simple baseline	Đường cơ sở đơn giản; mốc so sánh bắt buộc trước khi chấp nhận mô hình phức tạp.
Class imbalance	Mất cân bằng lớp; khiến độ chính xác trở thành chỉ số vô dụng.
Data drift	Trôi dạt dữ liệu; gây suy giảm âm thầm không có tín hiệu báo động.
Inter annotator agreement	Mức đồng thuận giữa những người gán nhãn; giới hạn trên của độ chính xác mô hình.
MLOps	Vận hành mô hình học máy; giám sát, ngưỡng kích hoạt và tái huấn luyện.
Confusion matrix	Ma trận nhầm lẫn; bốn ô sinh ra toàn bộ các chỉ số đánh giá.
Recall và precision	Độ nhạy và độ chuẩn xác; hai chỉ số đánh đổi nhau theo ngưỡng.
Decision threshold	Ngưỡng quyết định; không có ngưỡng đúng kỹ thuật, chỉ có ngưỡng tối ưu chi phí.

Thuật ngữ	Giải thích ngắn
Alert fatigue	Mệt mỏi cảnh báo; người vận hành mất tin tưởng rồi tắt hệ thống cảnh báo.
Explainability	Tính giải thích được; điều kiện để cải tiến và để quy trách nhiệm.
Data bias	Thiên lệch dữ liệu; mô hình tái tạo khiếm khuyết của quá khứ.
Human in the loop	Con người trong vòng lặp; hai điểm phê duyệt đặt trước hành động không đảo ngược.
ISO/IEC 42001	Tiêu chuẩn hệ thống quản lý trí tuệ nhân tạo trong tổ chức.
EU AI Act	Đạo luật phân loại hệ thống AI theo bốn mức rủi ro với nghĩa vụ tương ứng.
Total cost of ownership của dự án AI	Chi phí không kết thúc khi nghiệm thu; gán nhãn và tái huấn luyện bị bỏ sót nhiều nhất.

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 4

A. Câu hỏi ôn tập

1. Trình bày bốn năng lực quản trị ở Hình 4.1. Vì sao năng lực thứ ba được coi là năng lực phân biệt rõ nhất nhà quản trị với người dùng công nghệ?
2. Vẽ và giải thích quan hệ bao hàm giữa trí tuệ nhân tạo, học máy, học sâu và AI tạo sinh. AI tác tử nằm ở đâu trong sơ đồ này và vì sao?
3. Lạm phát kỹ thuật là gì? Dựa trên Hình 4.3, hãy nêu hai thông điệp quản trị và cho một ví dụ trong bối cảnh dự báo nhu cầu.
4. So sánh AI dự báo, AI tạo sinh và AI tác tử trên sáu tiêu chí. Vì sao rủi ro của AI tác tử mang tính tích lũy?
5. Trình bày hai xu hướng công nghệ và ba rào cản của AI công nghiệp được ghi nhận trong năm 2026.
6. Trình bày bốn câu hỏi của cây quyết định ở Hình 4.4. Vì sao câu hỏi về nhân được đặt sau câu hỏi về đại lượng mục tiêu chứ không phải ngược lại?
7. Vì sao phát hiện bất thường thường là điểm khởi đầu hợp lý hơn phân loại có giám sát trong nhà máy Việt Nam? Nhược điểm của lựa chọn này là gì?
8. Nêu sáu bước của vòng đời dự án dữ liệu và sai lầm điển hình của từng bước. Vì sao quy trình này mang tính lặp?
9. Vì sao phần lớn công sức của dự án nằm ở khâu chuẩn bị dữ liệu? Hệ quả đối với việc lập kế hoạch dự án là gì?
10. Giải thích giá trị sàng lọc của nguyên tắc luôn xây dựng đường cơ sở đơn giản trước.
11. Trình bày bốn vấn đề chất lượng dữ liệu sản xuất và biện pháp xử lý. Vì sao dữ liệu thường thiếu đúng vào lúc có sự cố, và điều đó gây hậu quả gì?
12. Dựa trên Hình 4.6, hãy giải thích bằng số vì sao độ chính xác vô dụng khi dữ liệu mất cân bằng lớp.
13. Dựa trên Hình 4.7, hãy giải thích vì sao trôi dạt dữ liệu nguy hiểm hơn các lỗi kỹ thuật khác. Nêu ba biện pháp phòng ngừa.
14. Trình bày sáu bước của quy trình gán nhãn dữ liệu ở Hình 4.8. Vì sao mức đồng thuận giữa những người gán nhãn lại là giới hạn trên của độ chính xác mô hình?
15. Nếu hai kiểm tra viên chỉ thống nhất với nhau ở 75 phần trăm số mẫu, hành động đúng của nhà quản trị là gì và vì sao?
16. Trình bày sáu bước của quy trình giám sát và duy trì mô hình ở Hình 4.9. Vì sao bước đặt ngưỡng kích hoạt có giá trị quản trị cao nhất?

17. Vẽ ma trận nhầm lẫn và định nghĩa bốn ô. Viết công thức độ chính xác, độ nhạy, độ chuẩn xác và độ đặc hiệu.
18. Dựa trên Hình 4.11, hãy giải thích cơ chế đánh đổi khi thay đổi ngưỡng quyết định. Vì sao không tồn tại ngưỡng đúng về mặt kỹ thuật?
19. Dựa trên Hình 4.12, hãy nêu ba nhận xét về hình dạng đường tổng chi phí và hàm ý quản trị của từng nhận xét.
20. Vì sao điểm F1 không phải là chỉ số phù hợp trong phần lớn tình huống nhà máy?
21. Một môi cảnh báo là gì? Nó xuất hiện trong bối cảnh nào và gây hậu quả gì cho dự án?
22. Vì sao ngưỡng tối ưu về chi phí lý thuyết có thể không khả thi về mặt vận hành? Nêu hai cách xử lý tình huống đó.
23. Trình bày cơ chế, dữ liệu cần có và điều kiện tiên quyết của ba trong sáu nhóm ứng dụng AI trọng yếu.
24. Dựa trên Hình 4.13, hãy giải thích vì sao các ứng dụng ở nửa trái của ma trận là điểm khởi đầu hợp lý cho phần lớn nhà máy Việt Nam.
25. Giải thích nghịch lý dữ liệu của bảo trì dự báo và ba cách xử lý.
26. Vì sao phần lớn thất bại của dự án thị giác máy không nằm ở mô hình? Hàm ý về thứ tự đầu tư là gì?
27. Nêu ba việc mà AI tạo sinh tạo giá trị rõ ràng trong nhà máy và ranh giới không nên vượt qua.
28. Trình bày bốn cấp độ phân tích và giải thích vì sao không thể nhảy cấp.
29. Nêu hai lỗi phổ biến khi thiết kế bảng điều khiển và nguyên tắc khắc phục.
30. Dựa trên Hình 4.15, hãy nêu bốn hạng mục chi phí bị bỏ sót nhiều nhất trong luận chứng kinh tế của dự án AI và giải thích vì sao chúng bị bỏ sót.
31. Dựa trên Hình 4.16, hãy giải thích vì sao trôi dạt dữ liệu là một rủi ro tài chính chứ không chỉ là vấn đề kỹ thuật.
32. Trình bày bảy bước thẩm định một đề xuất dự án AI. Bước nào có giá trị sàng lọc cao nhất và bước nào có giá trị bảo vệ cao nhất?
33. Vì sao trong nhiều tình huống nhà máy, mô hình đơn giản giải thích được có giá trị cao hơn mô hình phức tạp chính xác hơn?
34. Nêu ba dạng thiên lệch dữ liệu đặc thù của môi trường sản xuất. Vì sao dùng mô hình AI để xét thưởng phạt cá nhân là sai lầm quản trị?
35. Trình bày bốn mức rủi ro theo Đạo luật AI của Liên minh châu Âu kèm ví dụ trong nhà máy. Doanh nghiệp Việt Nam chịu ràng buộc qua những đường nào?

36. Vì sao có thể nói gánh nặng tuân thủ pháp lý đối với phần lớn ứng dụng AI trong nhà máy là thấp hơn hình dung phổ biến? Hai nhóm ứng dụng nào là ngoại lệ?
37. Phát biểu nguyên tắc phân bổ mức tự chủ cho hệ thống AI và giải thích bốn ô của ma trận ở Hình 4.19.
38. Dựa trên Hình 4.20, hãy giải thích vì sao điểm phê duyệt thứ hai phải đặt trước hành động chứ không phải sau, và vì sao khuyến nghị của mô hình bắt buộc phải đi kèm giải thích.
39. Trình bày sáu bước của quy trình triển khai AI có trách nhiệm ở Hình 4.21. Vì sao bước phân loại mức rủi ro phải đi trước thiết kế?
40. Trình bày ba giới hạn cấu trúc của AI trong nhà máy. Vì sao giới hạn thứ nhất không thể khắc phục bằng thuật toán tốt hơn?
41. Nêu ít nhất bốn tình huống mà AI không phải là giải pháp phù hợp, kèm phương án thay thế.
42. Trình bày lộ trình bốn lớp năng lực ở Hình 4.22. Vì sao thứ tự bốn lớp này không thể đảo?
43. Ba câu hỏi nào cần được trả lời trước khi phê duyệt bất kỳ đề xuất ứng dụng AI nào? Vì sao ba câu hỏi đó lại quan trọng hơn câu hỏi về độ chính xác của mô hình?
44. Hãy bình luận nhận định sau: doanh nghiệp nào có nhiều dữ liệu hơn thì đương nhiên có lợi thế hơn trong ứng dụng trí tuệ nhân tạo.

B. Bài tập vận dụng

Bài 1, đánh giá mô hình theo chi phí. Một hệ thống phát hiện lỗi hàn được kiểm định trên 20.000 mối hàn, trong đó 300 mối thực sự có lỗi. Kết quả: dương tính thật 240; âm tính giả 60; dương tính giả 780; âm tính thật 18.920.

- Tính độ chính xác, độ nhạy, độ chuẩn xác, độ đặc hiệu và điểm F1.
- Tính độ chính xác của một mô hình luôn tuyên bố mọi mối hàn đều đạt. Nhận xét và đối chiếu với Hình 4.6.
- Một mối hàn lỗi lọt tới khách hàng gây thiệt hại 35 triệu đồng; một mối hàn tốt bị loại nhầm gây thiệt hại 0,15 triệu đồng. Tính tổng chi phí sai lầm.
- Nếu hạ ngưỡng để độ nhạy tăng lên 96 phần trăm với cái giá là dương tính giả tăng lên 2.400, tổng chi phí thay đổi thế nào? Bạn khuyến nghị ngưỡng nào?
- Ở mức dương tính giả 2.400 trên 20.000 mối hàn, hãy ước lượng khối lượng công việc kiểm tra lại phát sinh mỗi ca và nhận xét về tính khả thi vận hành.
- Vẽ phác đường tổng chi phí theo ngưỡng cho tình huống này và đối chiếu hình dạng với Hình 4.12.

Bài 2, chọn nhóm bài toán học máy. Với mỗi tình huống dưới đây, hãy đi qua cây quyết định ở Hình 4.4, xác định nhóm bài toán phù hợp, nêu rõ dừng ở câu hỏi nào, dữ liệu cần có và một rủi ro chính.

Bảng 4.14. Các tình huống của bài tập 2

Số thứ tự	Tình huống
1	Dự đoán số lượng sản phẩm sẽ bán trong tuần tới cho 200 mã hàng
2	Phân loại ảnh bề mặt sơn thành đạt, lỗi chảy, lỗi bụi và lỗi bọt khí
3	Phát hiện máy nén khí có hành vi khác thường, chưa từng ghi nhận ca hỏng nào
4	Nhóm 400 mã vật tư thành các cụm có mô thức tiêu thụ tương tự để đặt chính sách tồn kho
5	Tìm chuỗi thứ tự chạy đơn hàng tối thiểu hóa tổng thời gian chuyển đổi trên một dây chuyền
6	Dự đoán một đơn hàng có nguy cơ trễ hạn hay không, tại thời điểm 3 ngày trước hạn giao

Bài 3, chẩn đoán chất lượng dữ liệu. Dữ liệu của một nhà máy có các đặc điểm sau: 12 phần trăm bản ghi thiếu trường nguyên nhân dừng máy, tập trung ở ca đêm; cảm biến nhiệt trên máy số 7 cho giá trị lệch trung bình 8 độ C so với các máy cùng loại; trong 18 tháng dữ liệu chỉ có 22 ca hỏng trên tổng 46.000 giờ vận hành; từ tháng thứ 13 nhà máy đổi nhà cung cấp nguyên liệu chính.

- Xác định vấn đề chất lượng dữ liệu tương ứng với từng đặc điểm.
- Với mỗi vấn đề, đề xuất một biện pháp xử lý và nêu rõ hạn chế của biện pháp đó.
- Đánh giá xem tập dữ liệu này có đủ điều kiện để xây dựng mô hình phân loại có giám sát dự đoán hỏng hóc hay không. Nếu không, đề xuất lộ trình.
- Việc đổi nhà cung cấp ở tháng 13 ảnh hưởng thế nào tới cách chia tập huấn luyện và tập kiểm định?

Bài 4, thiết kế quy trình gán nhãn. Một nhà máy điện tử muốn xây dựng hệ thống thị giác máy phát hiện lỗi mối hàn, với 5 loại lỗi cần phân biệt. Hiện có 4 kiểm tra viên với thâm niên khác nhau.

- Áp dụng sáu bước ở Hình 4.8 để thiết kế quy trình gán nhãn cho dự án này.
- Thiết kế cách đo mức đồng thuận giữa những người gán nhãn và nêu ngưỡng bạn cho là chấp nhận được.
- Nếu mức đồng thuận đo được chỉ đạt 68 phần trăm, bạn khuyến nghị làm gì trước khi tiếp tục dự án?

- d) Ước lượng khối lượng công việc gán nhãn cần thiết nếu mục tiêu là 400 mẫu cho mỗi loại lỗi, và đưa khoản này vào cấu trúc chi phí ở Bảng 4.11.

Bài 5, giám sát mô hình sau triển khai. Một mô hình bảo trì dự báo được nghiệm thu với độ nhạy 88 phần trăm và tỷ lệ báo động giả 6 phần trăm.

- Áp dụng sáu bước ở Hình 4.9 để thiết kế kế hoạch giám sát cho mô hình này.
- Đề xuất ngưỡng kích hoạt tái huấn luyện cụ thể và lập luận cho mức ngưỡng đó.
- Nêu ba biến đầu vào cần giám sát phân phối và giải thích vì sao chọn ba biến đó.
- Sau 12 tháng, tỷ lệ báo động giả tăng lên 14 phần trăm trong khi độ nhạy vẫn giữ ở 86 phần trăm. Hãy chẩn đoán nguyên nhân có thể và đề xuất hành động.

Bài 6, bốn cấp độ phân tích trên dữ liệu ca sản xuất. Sử dụng tập dữ liệu ca sản xuất do giảng viên cung cấp.

- Thực hiện bốn nhiệm vụ của Bảng 4.10.
- Xác định nhà máy đang ở cấp độ phân tích nào và nêu bằng chứng.
- Đề xuất một ứng dụng ở cấp độ kế tiếp, kèm điều kiện dữ liệu tiên quyết còn thiếu.
- Ước lượng lợi ích bằng tiền của đề xuất đó với giả định được nêu tường minh.

Bài 7, luận chứng kinh tế cho dự án AI. Một nhà máy dệt đề xuất hệ thống phát hiện lỗi vải bằng thị giác máy. Sản lượng 900.000 mét mỗi năm, tỷ lệ lỗi 1,5 phần trăm. Kiểm tra thủ công phát hiện 65 phần trăm số lỗi và loại nhầm 0,8 phần trăm hàng tốt. Ba kiểm tra viên với chi phí đầy đủ 10 triệu đồng mỗi người mỗi tháng. Một mét vải lỗi lọt tới khách hàng gây thiệt hại 0,9 triệu đồng; một mét vải tốt bị loại nhầm gây thiệt hại 0,08 triệu đồng. Hệ thống đề xuất nâng tỷ lệ phát hiện lên 93 phần trăm nhưng tỷ lệ loại nhầm tăng lên 2,5 phần trăm. Đầu tư ban đầu 2.400 triệu đồng; chi phí vận hành 200 triệu đồng mỗi năm; giữ lại 1 kiểm tra viên. Vòng đời 5 năm, hệ số hiện giá 3,6048.

- Tính lợi ích ròng năm, thời gian hoàn vốn và giá trị hiện tại ròng.
- Xác định mức tỷ lệ phát hiện tối thiểu để dự án hòa vốn, giữ nguyên các giả định khác.
- Áp dụng ba kịch bản của Hình 4.16 cho tình huống này và lập bảng so sánh.
- Đề xuất khoản đầu tư bổ sung có tỷ suất hoàn vốn cao nhất và lập luận.
- Nếu đề xuất bỏ sót hoàn toàn chi phí gán nhãn dữ liệu ước tính 350 triệu đồng, kết luận đầu tư có thay đổi không?

Bài 8, thẩm định một đề xuất dự án AI. Một nhà cung cấp chào bán giải pháp trợ lý số dựa trên mô hình ngôn ngữ để hỗ trợ chẩn đoán sự cố thiết bị, giá trọn gói 1.600 triệu đồng cho ba năm. Nhà máy hiện lưu tài liệu kỹ thuật dưới dạng bản in và tệp rời rạc trên máy tính cá nhân của các kỹ sư, không có hệ thống quản lý phiên bản.

- a) Áp dụng quy trình bảy bước ở Hình 4.17 để thẩm định. Chỉ rõ đề xuất bị chặn ở bước nào và vì sao.
- b) Đề xuất một lộ trình thay thế gồm ba giai đoạn với chi phí thấp hơn.
- c) Xác định các rủi ro đặc thù của AI tạo sinh trong tình huống này và biện pháp kiểm soát tương ứng.
- d) Định vị ứng dụng này trên bốn mức rủi ro ở Hình 4.18 và nêu nghĩa vụ tương ứng.

Bài 9, thiết kế kiến trúc con người trong vòng lặp. Với ba ứng dụng sau, hãy định vị trên ma trận ở Hình 4.19 và thiết kế kiến trúc phê duyệt theo mẫu Hình 4.20: hệ thống tự động dừng chuyển khi phát hiện lỗi chất lượng; hệ thống tự động đặt hàng bổ sung vật tư nhóm C; hệ thống tự động điều chỉnh nhiệt độ lò trong dải cho phép.

- a) Với mỗi ứng dụng, xác định mức tự chủ phù hợp và lập luận theo hai chiều của ma trận.
- b) Xác định vị trí các điểm phê duyệt bắt buộc và nội dung cần phê duyệt tại mỗi điểm.
- c) Nêu thông tin tối thiểu mà giải thích của mô hình phải cung cấp để việc phê duyệt có ý nghĩa.
- d) Thiết kế cơ chế lưu vết phục vụ đồng thời hai mục đích là huấn luyện chu kỳ sau và trách nhiệm giải trình.

Bài 10, bài tập nhóm nối tiếp dự án học phần. Trên cơ sở nhà máy đã khảo sát ở các chương trước, hãy thực hiện các nội dung sau: định vị nhà máy trên lộ trình bốn lớp năng lực ở Hình 4.22 kèm bằng chứng; xác định một ứng dụng AI khả thi nhất và đi hết quy trình bảy bước thẩm định; lập luận chứng kinh tế đầy đủ theo cấu trúc chi phí ở Bảng 4.11 với phân tích ba kịch bản; thiết kế kế hoạch gán nhân và kế hoạch giám sát mô hình; và đánh giá yêu cầu về AI có trách nhiệm áp dụng cho ứng dụng đó. Trình bày trong không quá 8 trang, nêu rõ nguồn dữ liệu và các giả định đã sử dụng.

TÀI LIỆU THAM KHẢO

- Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 41st International Conference on Software Engineering*, tr. 291 đến 300.
- Budde, L., Hänggi, R., Friedli, T., & Rüedy, A. (2023). *Smart Factory Navigator: Identifying and implementing the most beneficial use cases for your company*. Springer Cham.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP DM 1.0: Step by step data mining guide*. SPSS Inc.
- Deshpande, A., Sarkar, B., Dave, D., & Dave, R. (2024). *Advanced manufacturing and supply chain with IoT: Revolutionizing industries through smart technologies and connectivity*. BPB Publications.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
- International Organization for Standardization. (2023). *ISO/IEC 42001: Information technology. Artificial intelligence. Management system*. ISO.
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber physical systems architecture for Industry 4.0 based manufacturing systems. *Manufacturing Letters*, 3, tr. 18 đến 23.
- Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data analytic thinking*. O'Reilly Media.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), tr. 206 đến 215.
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28, tr. 2503 đến 2511.
- Shalev Shwartz, S., & Ben David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data driven smart manufacturing. *Journal of Manufacturing Systems*, 48, tr. 157 đến 169.
- Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: Advantages, challenges and applications. *Production and Manufacturing Research*, 4(1), tr. 23 đến 45.

Liên minh châu Âu. (2024). *Quy định (EU) 2024/1689 về thiết lập các quy tắc hài hòa đối với trí tuệ nhân tạo.*

