



Bài 7

HỌC BÁN GIÁM SÁT

Giảng viên: ThS. Lê Thị Thùy Trang

Email: trangltt@dainam.edu.vn

Điện thoại: 0966730396



CASE STUDY



Hành trình giải cứu Nam Chill

- ❖ Nam Chill là học bá trong ngành fintech.
- ❖ Những ngày đầu đi làm ở ngân hàng Super, Nam đúng kiểu “tân binh tóc thẳng”: đầu tóc bóng bẩy, áo sơ mi kẻ sọc phẳng phiu, nhìn hồ sơ nào cũng gật gù “deal nhẹ”.
- ❖ Cả phòng tín dụng đồn nhau: “Hồ sơ nào vào tay Nam cũng như nước chảy mây trôi, anh em chỉ việc pha trà ngồi xem.”

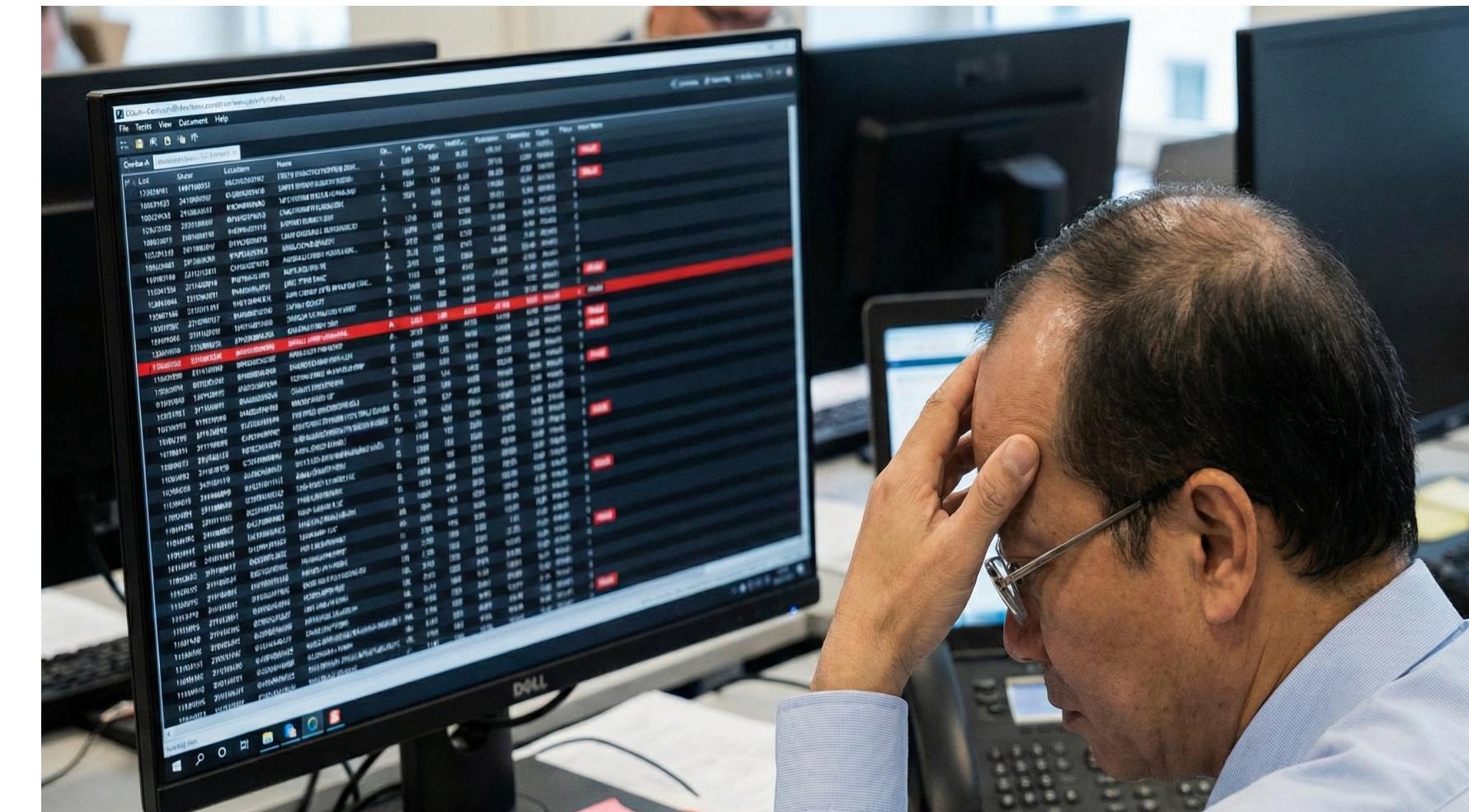


Hành trình giải cứu Nam Chill

❖ Cho đến một ngày đẹp trời, sếp gọi Nam lên phòng họp, chiếu lên màn hình cả núi giao dịch thẻ tín dụng: hàng triệu dòng mà số giao dịch gian lận được gắn nhãn thì ít như... tóc trên đầu sếp tổng.

❖ Sếp ban *thánh chỉ*:

 “Nam à, cậu là siêu sao nên từ hôm nay kiêm luôn thánh bắt gian lận, nhưng chỉ được dùng chỗ nhãn ít ỏi này thôi, còn lại tự tìm cách tận dụng đống dữ liệu chưa nhãn kia nhé.”



Hành trình giải cứu Nam Chill

- ❖ Nam nhìn vào tỷ lệ gian lận bé tí, mô hình nào cũng chỉ muốn đoán: Bình thường hết cho lành, **accuracy đẹp nhưng kẻ gian vẫn nhởn nhơ**.
- ❖ Ngồi ôm đống data mấy đêm liền, tóc Nam từ thẳng băng bắt đầu... xoăn nhẹ, rồi xù dần thành “Nam xù” lúc nào không hay.

Before



After

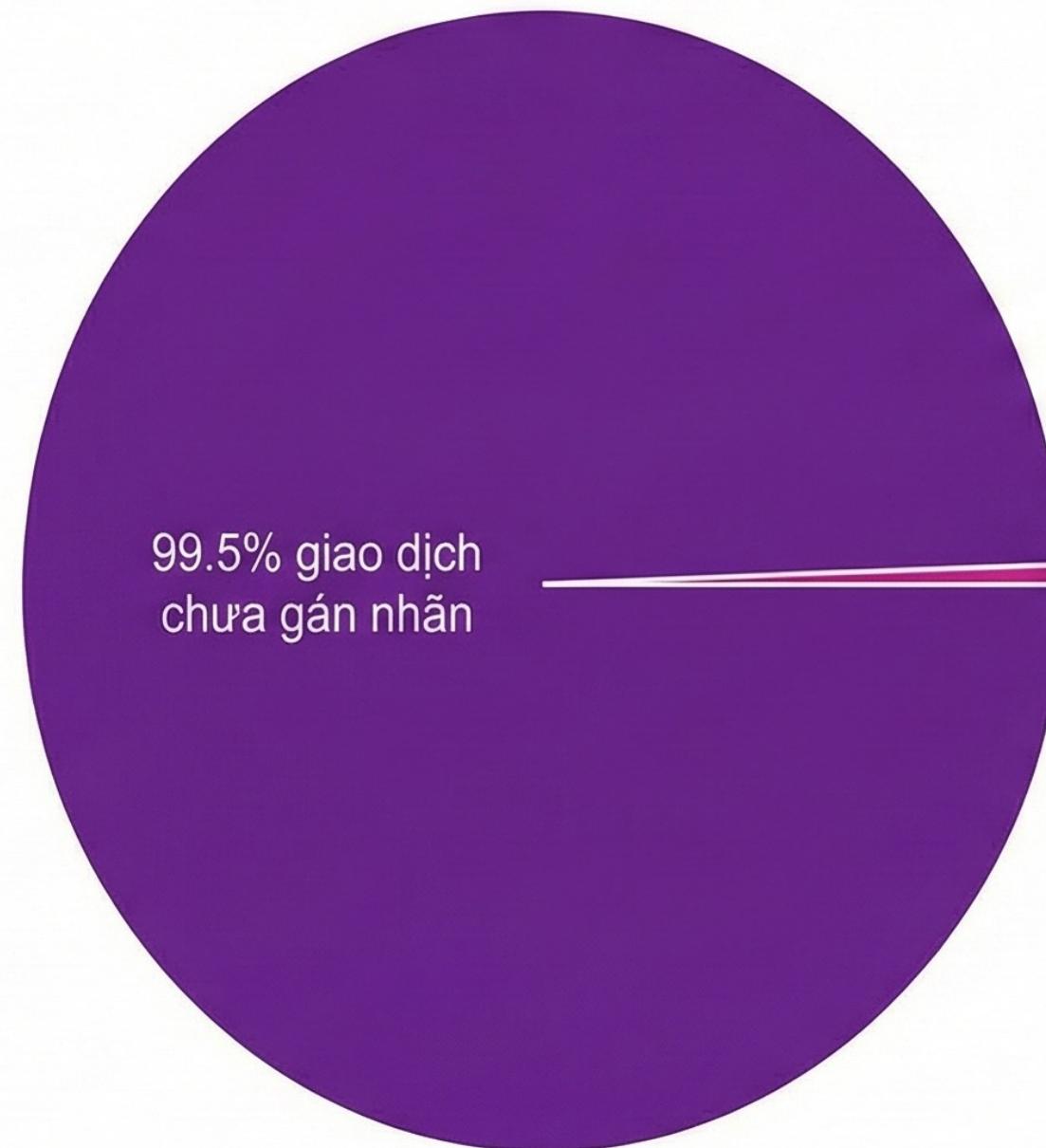


Hành trình giải cứu Nam Chill

- ❖ Sau nhiều đêm try hard với núi dữ liệu giao dịch, Nam Chill nhận ra vấn đề

Ngân hàng Super - Toàn Cảnh Giao Dịch

1,000,000



Chi Tiết 5,000 Giao Dịch Được Gán Nhãn

100

0.5% giao dịch
được gán
nhãn

Trong 5,000
giao dịch này

98% giao dịch
hợp pháp

2% giao dịch
gian lận

Hành trình giải cứu Nam Chill

- ❖ Được sự chỉ giáo của sếp, Nam khám phá chân trời kiến thức mới mang tên:

HỌC BÁN GIÁM SÁT



OUTLINE

1. Học máy bán giám sát
2. Thuật toán tự huấn luyện
3. Thuật toán huấn luyện đồng thời
4. Luyện tập





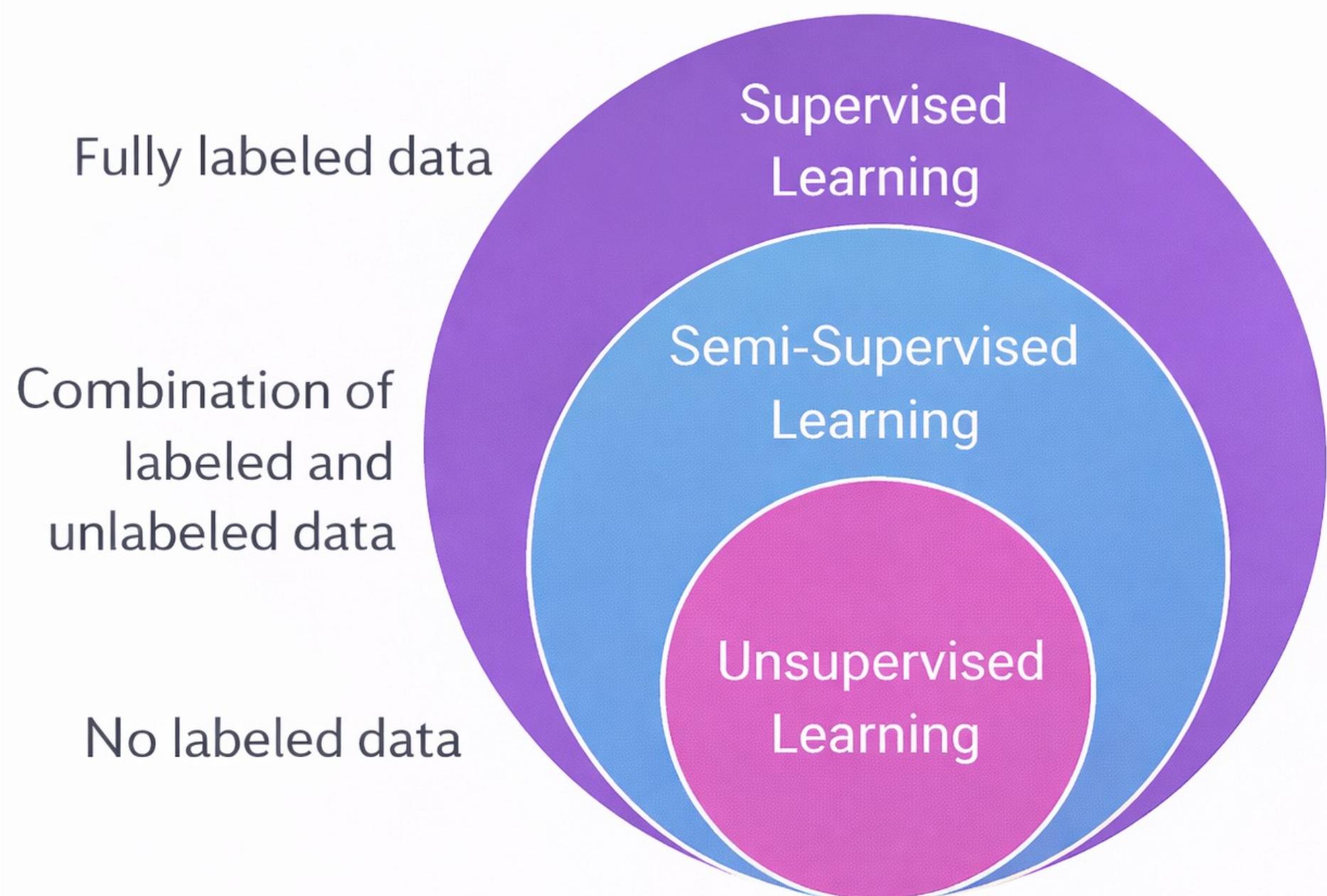
HỌC MÁY BÁN GIÁM SÁT

SEMI-SUPERVISED LEARNING

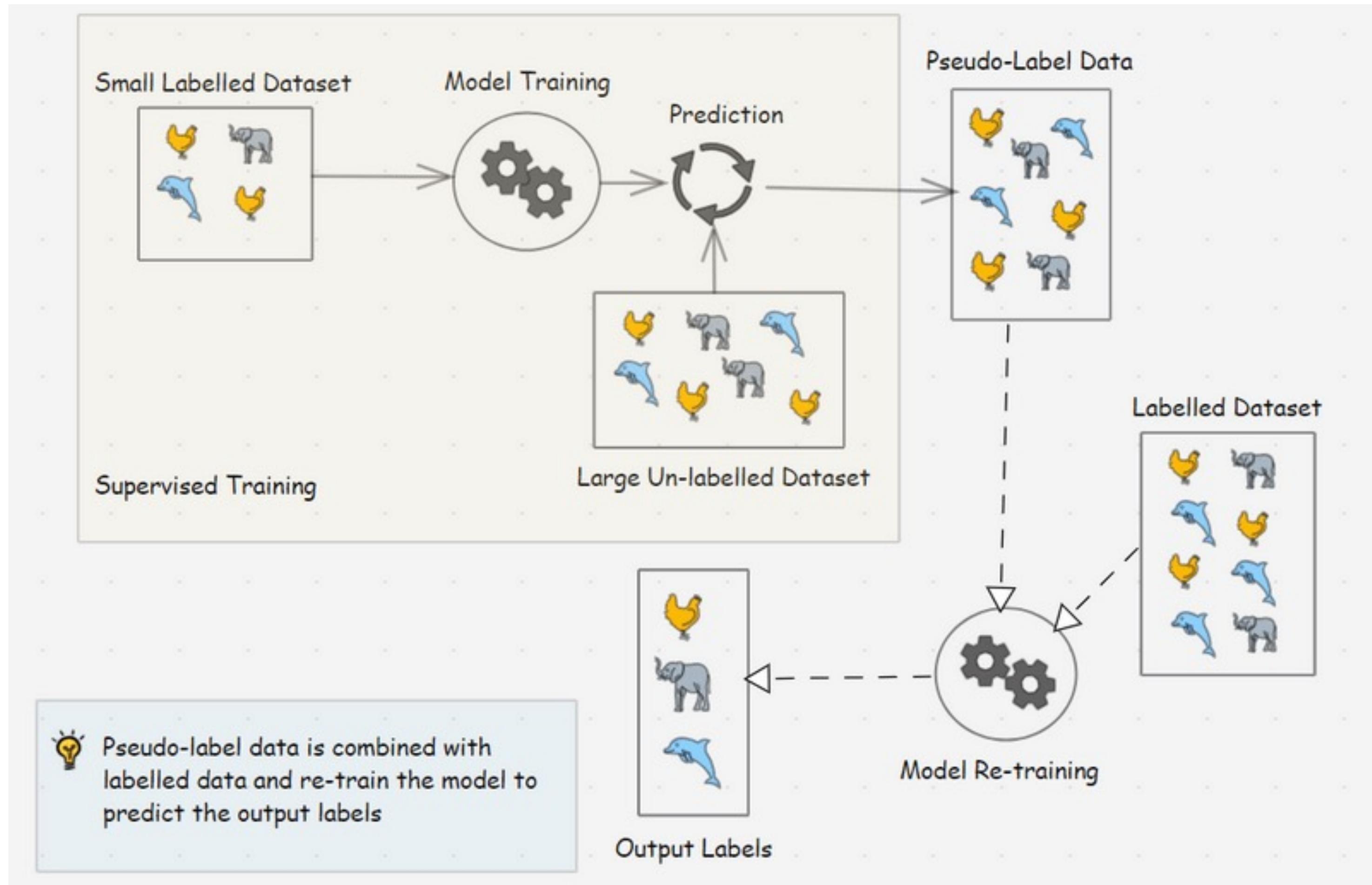
❖ Học bán giám sát (Semi Supervised Learning – SSL): là một kỹ thuật học máy sử dụng **một phần nhỏ dữ liệu được gán nhãn và rất nhiều dữ liệu chưa được gán nhãn** để huấn luyện một mô hình dự đoán.

❖ Một số hướng tiếp cận trong SSL:

- Tự huấn luyện (Self-Training)
- Huấn luyện đồng thời (co-training, tri-training)
- Graph-Based Learning



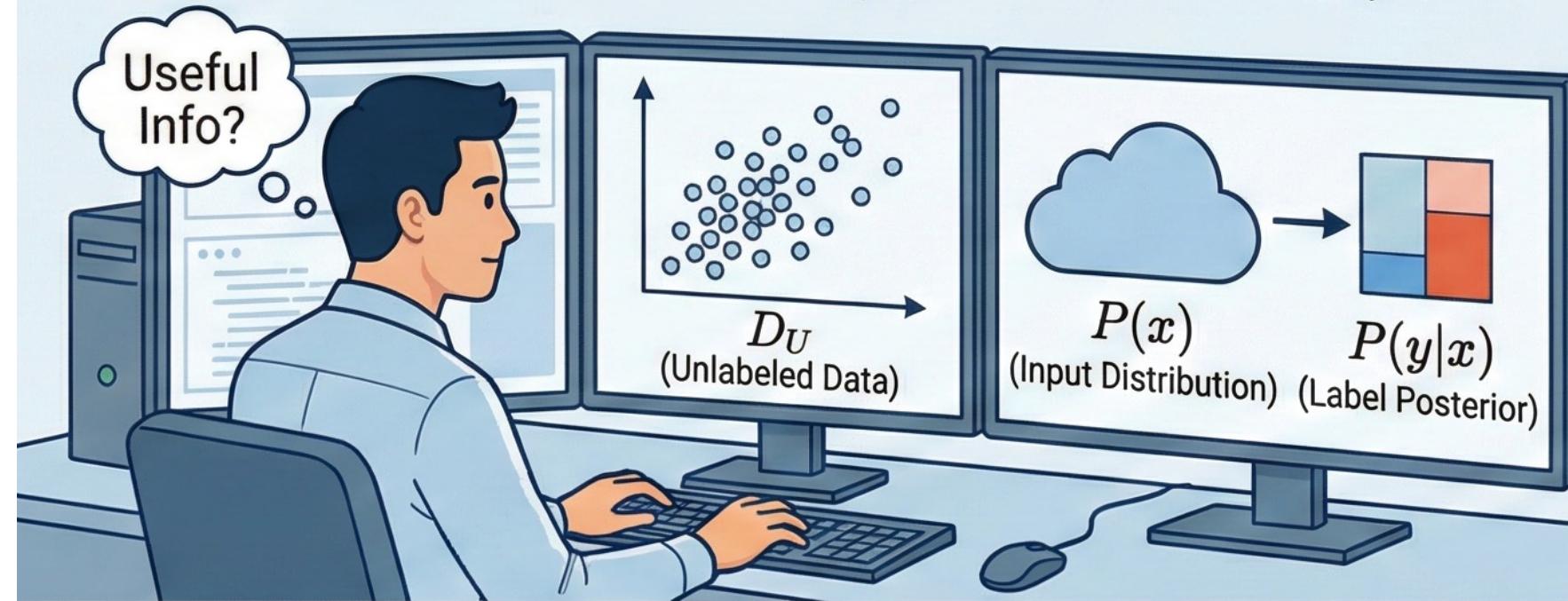
SEMI-SUPERVISED LEARNING



- ❖ **Semi-supervised learning (SSL)** is a machine learning approach where a model is trained on a combination of labeled and unlabeled data. The primary goal is to use the unlabeled data to compensate for the small labeled dataset, enhancing the model's generalization capabilities without incurring high labelling costs.
- ❖ In practice SSL uses a small, labeled dataset to kick-start the learning process. The model then uses its predictions on the unlabeled data to refine its learning, allowing SSL to achieve near-supervised performance levels with a fraction of the labeled data required by traditional supervised learning model.
- ❖ Semi-supervised learning works best when labeled and unlabeled data are similar, sharing common structures or distributions.

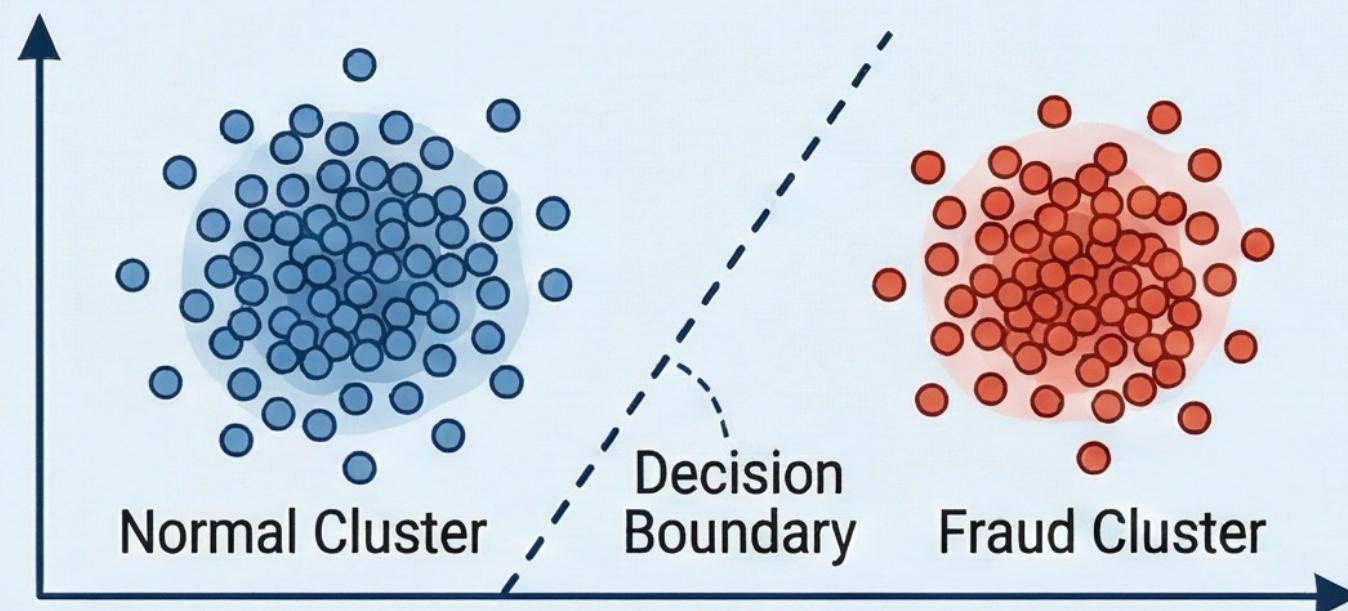
- ❖ Các phương pháp học bán giám sát dựa trên niềm tin: Cách dữ liệu phân bố trong không gian đặc trưng ($P(x)$) giúp ta đoán được nhãn của nó ($P(y/x)$).
- ❖ Ba giả định cốt lõi:
 1. **Giả định độ trơn (Smoothness Assumption):** Nếu hai giao dịch có đặc trưng giống nhau, thì khả năng cao chúng cùng nhãn.
 2. **Giả định phân cụm (Cluster Assumption):** Dữ liệu thường tụ lại thành từng cụm, trong cùng một cụm thì nhãn thường giống nhau.
 3. **Giả định đa tạp (Manifold Assumption):** Dữ liệu nhìn thì nhiều chiều, nhưng thật ra nằm trên một mặt có số chiều thấp hơn

SSL FOUNDATION: $P(x)$ INFORMS $P(y|x)$



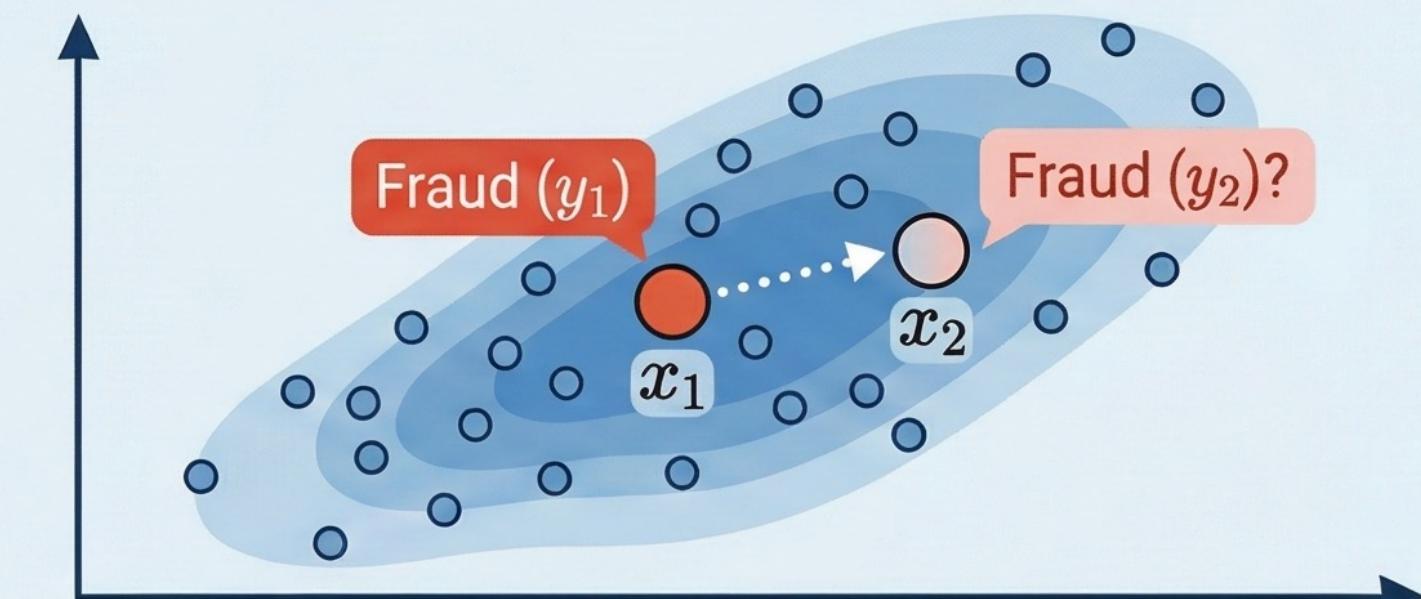
Nam uses input distribution $P(x)$ to infer label distribution $P(y|x)$.

2. CLUSTER ASSUMPTION



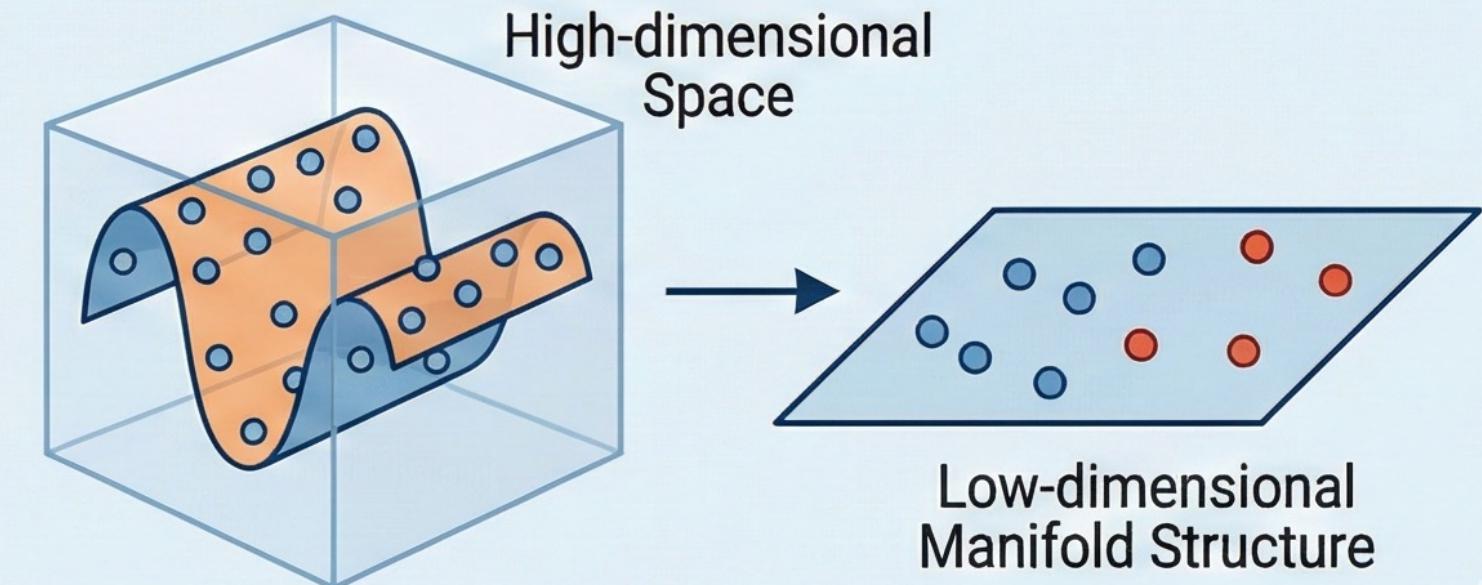
Points in the same cluster likely have the same label.
Boundary avoids dense regions.

1. SMOOTHNESS ASSUMPTION



Similar transactions (close in space) share labels.
Nam propagates labels to nearby points.

3. MANIFOLD ASSUMPTION



Data lies on a lower-dimensional manifold.
Learning its structure helps find rules.

Hành trình giải cứu Nam Chill

- ❖ Nam Chill lần lượt học và áp dụng:





ĐẠI NAM
UNIVERSITY

THUẬT TOÁN TỰ HUẤN LUYỆN



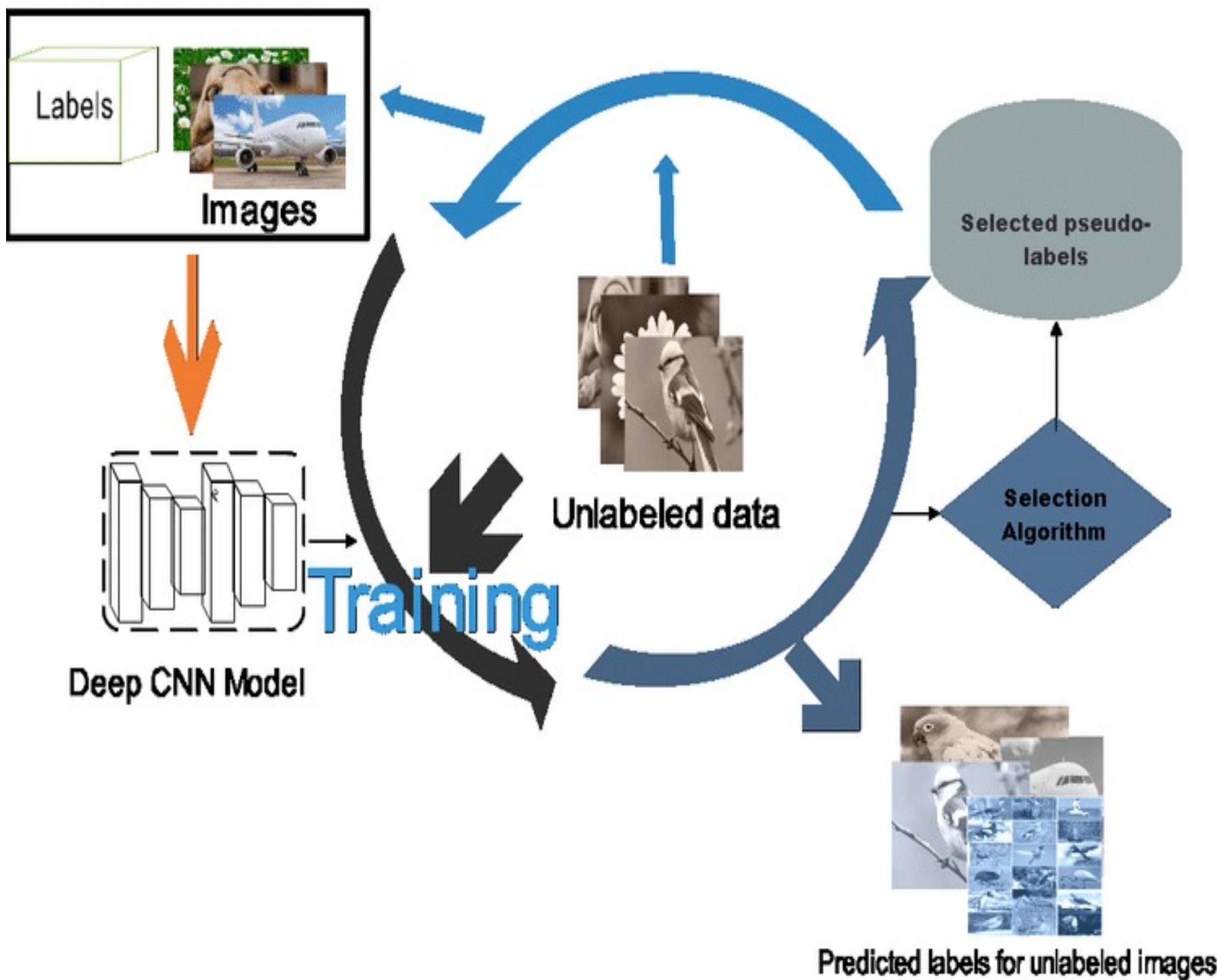
Hành trình giải cứu Nam Chill

- ❖ Trong những đêm dài tại Ngân hàng Super, Nam Chill nảy ra một ý tưởng đơn giản nhưng đầy tiềm năng: "Nếu mình đã dạy mô hình cách nhận diện một số loại gian lận cơ bản, tại sao không để nó tự tìm kiếm các trường hợp tương tự trong đống dữ liệu chưa gán nhãn, rồi dùng chính những phát hiện đó để dạy lại cho nó?"

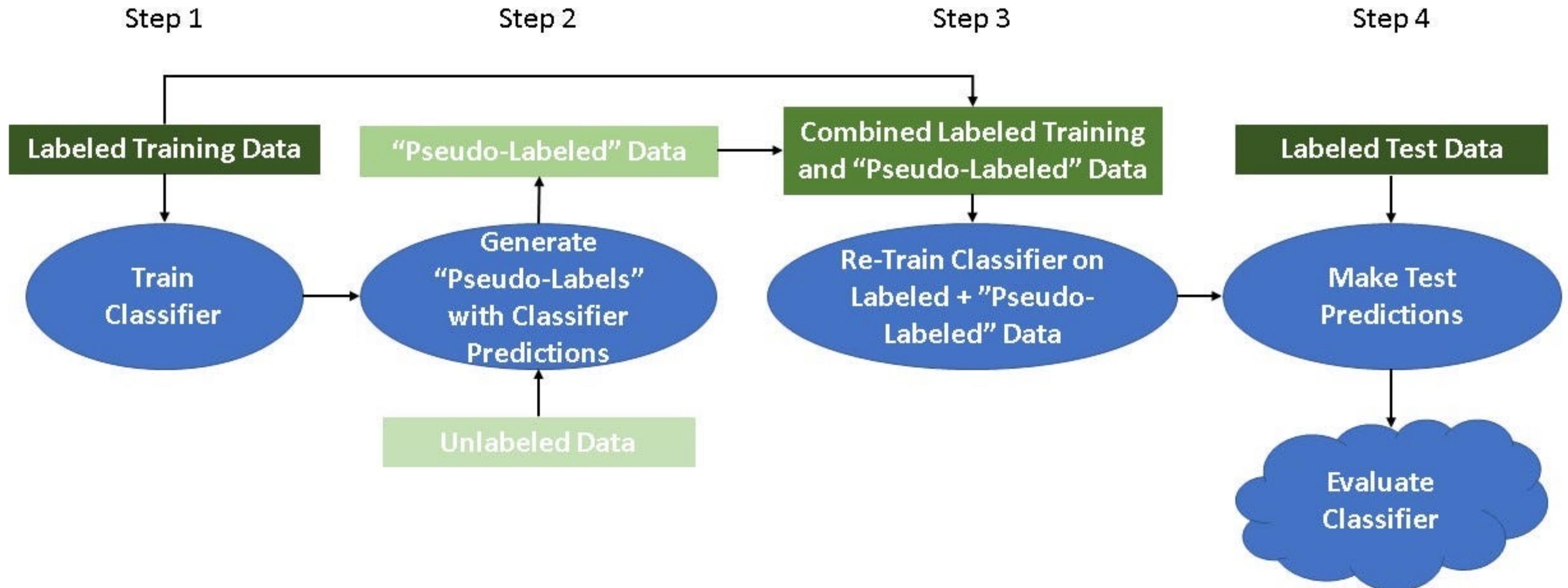


SELF-TRAINING

- ❖ Thuật toán tự huấn luyện (Self-training) hoạt động theo nguyên lý “Tin vào những dự đoán mà mình tự tin nhất.”
- ❖ Ban đầu, thuật toán huấn luyện một mô hình trên tập dữ liệu nhỏ đã được gán nhãn.
- ❖ Sau đó nó sử dụng mô hình này để dự đoán nhãn cho dữ liệu chưa được gán nhãn.
- ❖ Các dự đoán có độ tin cậy cao được thêm vào tập dữ liệu đã được gán nhãn cho các lần huấn luyện tiếp theo.



SELF-TRAINING



- ❖ x : vector đặc trưng của một giao dịch
- ❖ $y \in \{0, 1\}$: là nhãn (0: Bình thường, 1: Gian lận)
- ❖ $D_L = \{(x_i, y_i)\}_{i=1}^{n_L}$: tập dữ liệu có nhãn
- ❖ $D_U = \{(x_j)\}_{j=1}^{n_U}$: tập dữ liệu chưa có nhãn
- ❖ Mô hình phân lớp ở vòng t , f_t có đầu ra là:

$$f_t(x) = P(y = 1 | x) = P(Fraud | x)$$

- ❖ Xác suất dự đoán cho điểm chưa có nhãn: $p_j = f_t(x_j)$
- ❖ Độ tin cậy: $conf(x_j) = \max(p_j, 1 - p_j)$
- ❖ Ngưỡng tin cậy: $\tau \in (0.5, 1)$. Chỉ chọn pseudo—label nếu $conf(x_j) > \tau$

❖ Bước 1: Huấn luyện có giám sát.

- Ban đầu, Nam huấn luyện f_0 trên D_L . Với bài toán phân loại nhị phân, sử dụng hàm mất mát Binary Cross-Entropy :

$$\mathcal{L}_{sup}(\theta) = -\frac{1}{|D_L|} \sum_{(x,y) \in D_L} [y \log f_\theta(x) + (1 - y) \log (1 - f_\theta(x))]$$

- Ví dụ, Nam huấn luyện mô hình Logistic Regression trên 5000 giao dịch có nhãn.

Với mỗi giao dịch chưa nhãn $x_j \in D_U$ mô hình tính:

$$p_j = f_0(x_j) = P(Fraud | x_j)$$

SELF-TRAINING

❖ Bước 2: Sinh pseudo-label từ dữ liệu chưa nhãn

➤ Nam dùng f_0 để gán nhãn tạm theo quy tắc sau:

Nếu $p_j \geq \tau \Rightarrow$ gán $\hat{y}_j = 1$ (Fraud)

Nếu $p_j \leq 1 - \tau \Rightarrow$ gán $\hat{y}_j = 0$ (Normal)

Nếu $1 - \tau < p_j < \tau \Rightarrow$ bỏ qua (chưa đủ chắc chắn)

➤ Giả sử: Chọn ngưỡng $\tau = 0.95$

Giao dịch A: $P(1 | x_A) = 0.96 \Rightarrow$ gán $\hat{y}_A = 1$ (Fraud)

Giao dịch B: $P(1 | x_B) = 0.52 \Rightarrow$ bỏ qua

Giao dịch C: $P(1 | x_C) = 0.03 \Rightarrow$ gán $\hat{y}_C = 0$ (Normal)

➤ Tập pseudo-labeled tạo được:

$$D_p = \{(x_j, \hat{y}_j) | conf(x_j) \geq \tau, x_j \in D_U\}$$

SELF-TRAINING

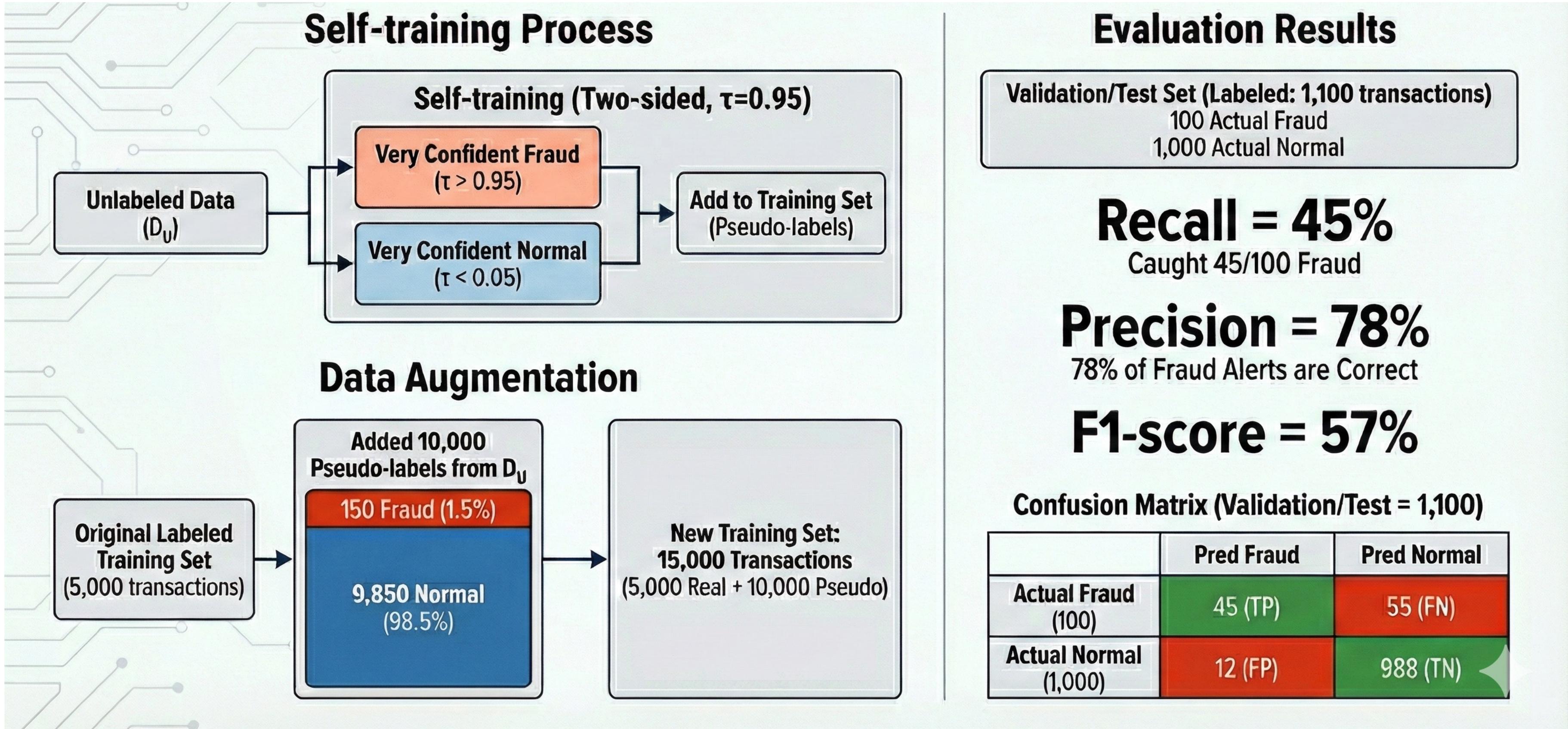
❖ Bước 3: Kết hợp dữ liệu và huấn luyện lại

- Nam gộp tập huấn luyện mới: $D^{(1)} = D_L \cup D_p$
- Huấn luyện lại để thu được mô hình mới f_1 (cùng kiến trúc hoặc có kiến trúc khác)
- Hàm mất mát

$$\mathcal{L}^{(1)}(\theta) = \frac{1}{D^{(1)}} \sum_{(x,y) \in D^{(1)}} [y \log f_\theta(x) + (1 - y) \log (1 - f_\theta(x))]$$

SELF-TRAINING

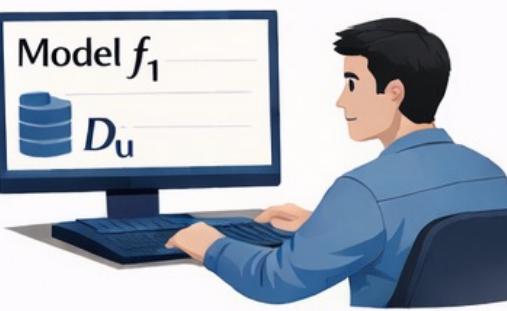
❖ Sau bước 3



❖ Bước 4: Lặp và đánh giá

1. Predict Fraud probability for entire D_u

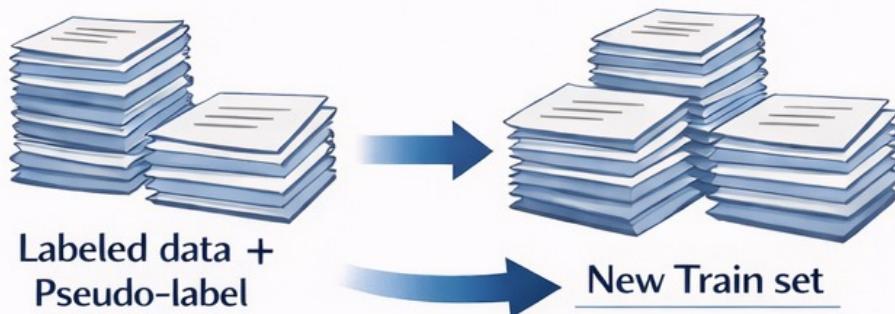
Predict P(Fraud)

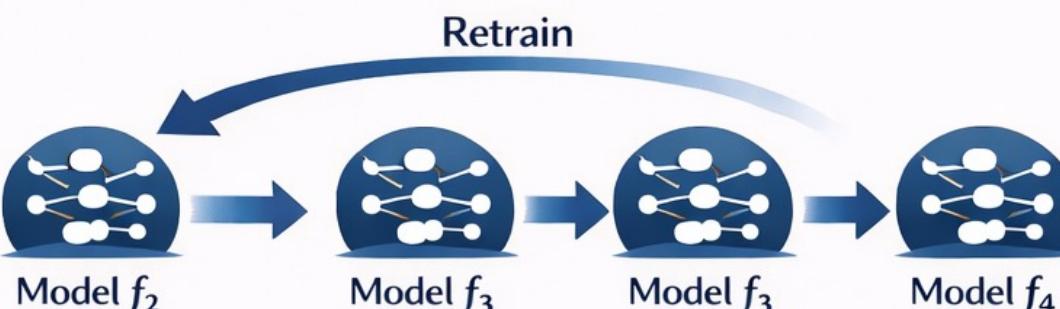
2. Select highly certain transactions by threshold τ (two-sided)

$P < 1 - \tau$ Normal $\rightarrow |\tau| \rightarrow P > \tau$ Fraud

3. Merge pseudo-labels with labeled data to create new Train set

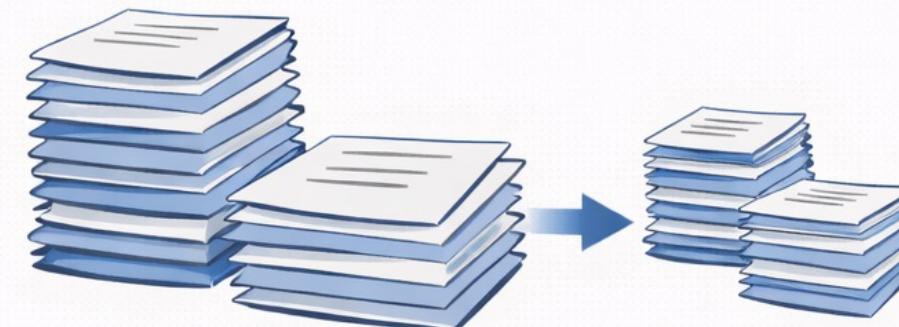


4. Retrain to obtain better model (e.g., f_2, f_3, \dots)



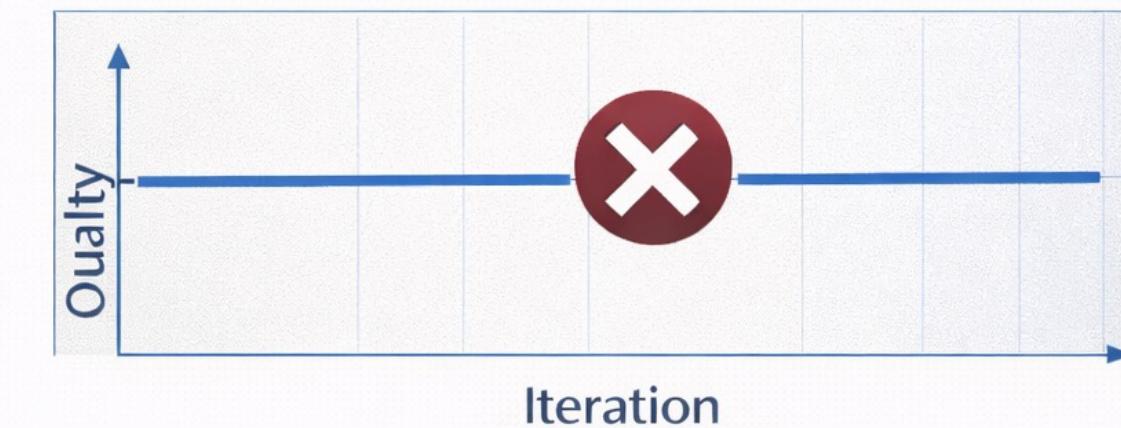
Stop iteration when:

- Number of new pseudo-labels increases very little



Only a few new pseudo-labels

- Quality on validation does not improve



❖ Bước 4: Lắp và đánh giá

- Đánh giá trên tập test có nhãn

Tập test:

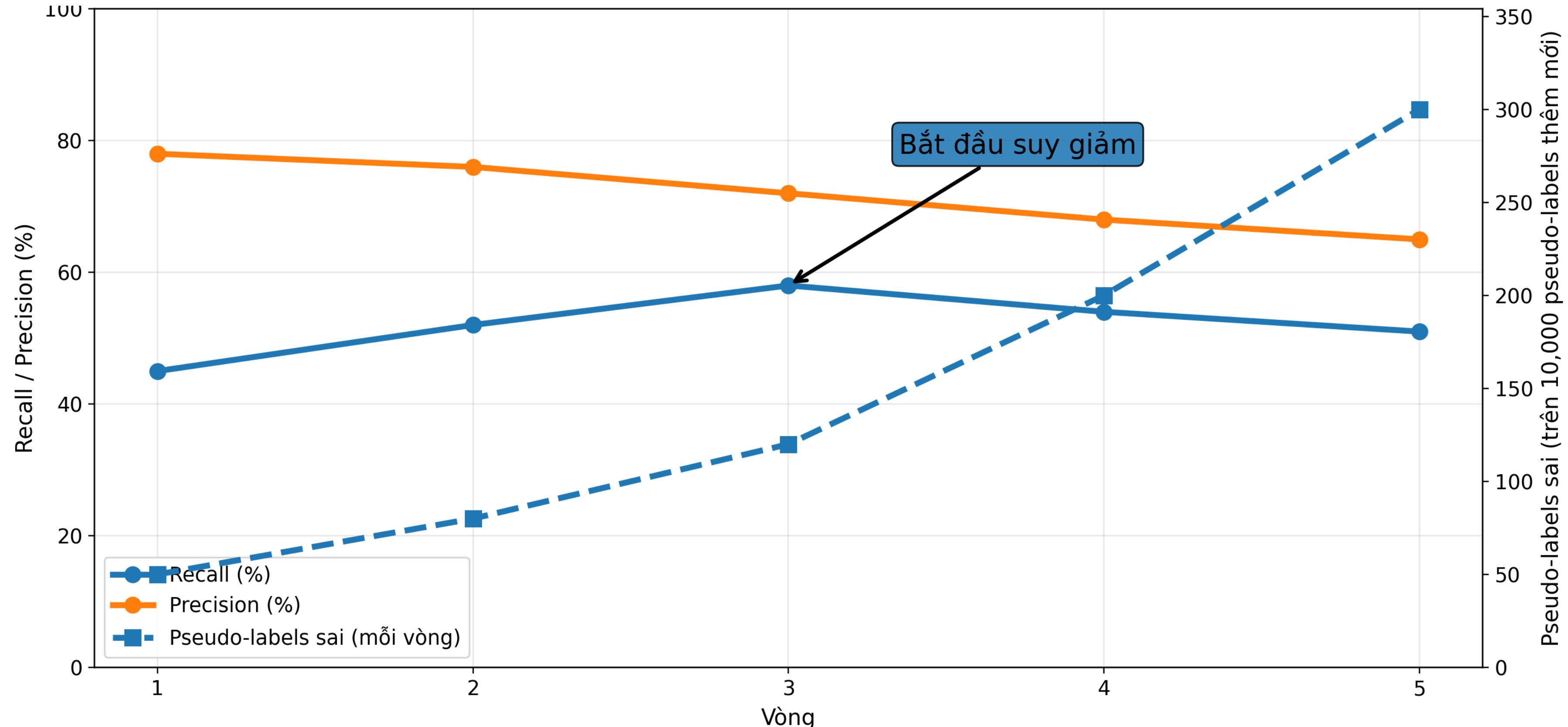
$$D_{\text{test}} = \{(x_k, y_k)\}_{k=1}^{n_{\text{test}}}$$

Dự đoán:

$$\widehat{y}_k = 1 [f_T(x_k) \geq 0.5]$$

Metrics đánh giá: precision, recall

Sau 5 vòng lặp, Nam phát hiện vấn đề lớn

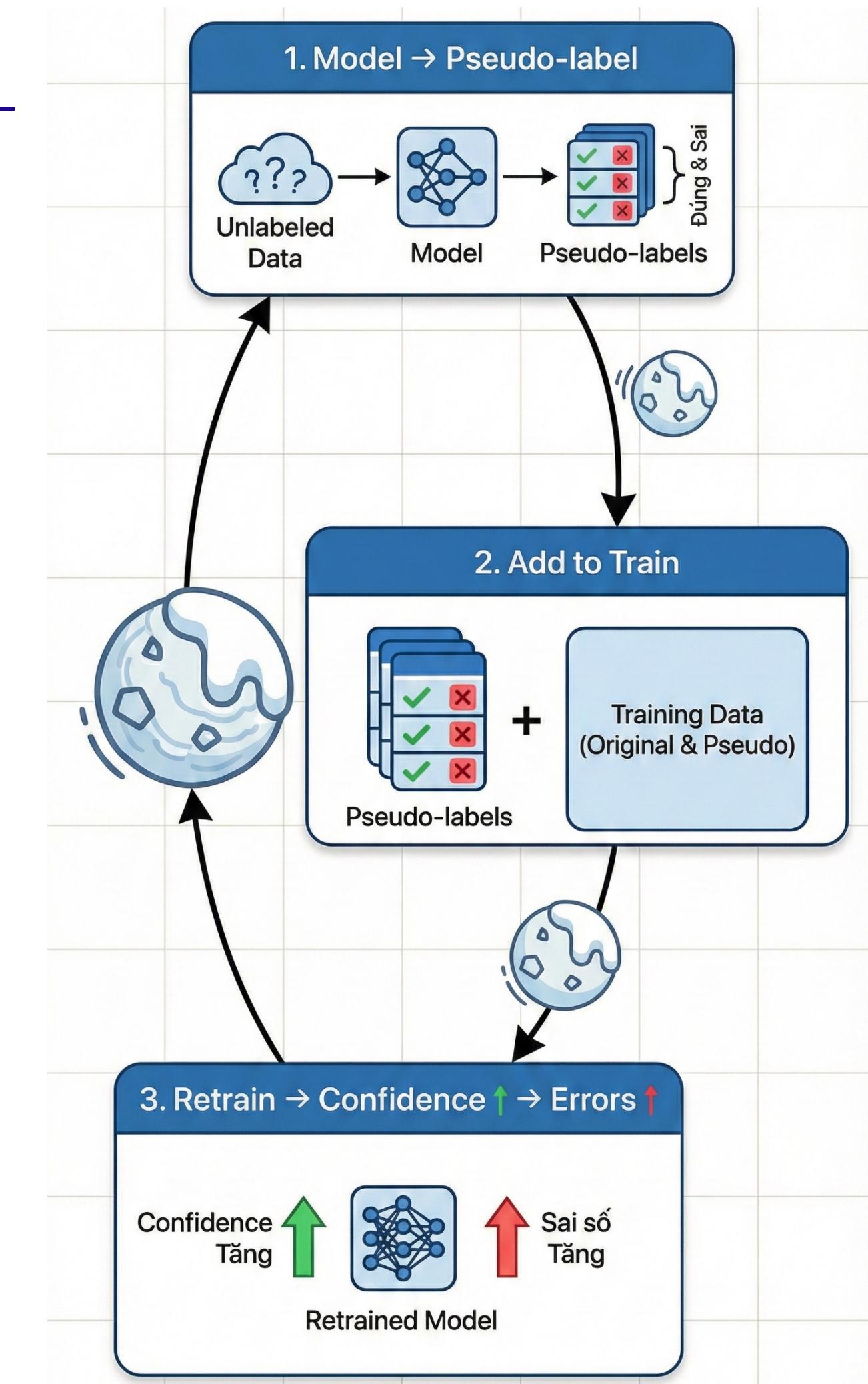


SELF-TRAINING

❖ Vấn đề cốt lõi của self-training:

Confirmation Bias & Error Propagation

- Confirmation bias: mô hình tin vào dự đoán của chính nó → củng cố sai lầm.
- Error propagation: một số pseudo-label sai → vào train → mô hình mới càng tự tin vào sai số.





ĐẠI NAM
UNIVERSITY

THUẬT TOÁN HUẤN LUYỆN ĐỒNG THỜI

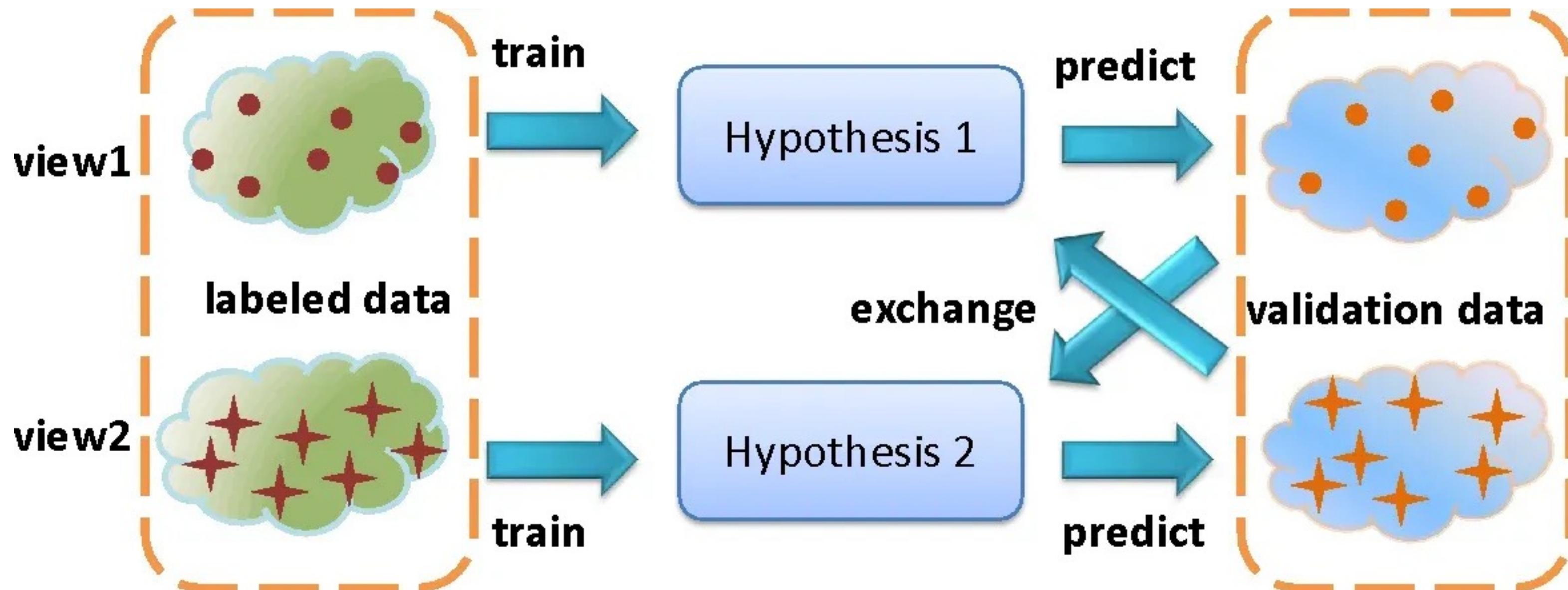
Quay trở lại hành trình giải cứu Nam Chill

- ❖ Sếp gợi ý Nam rằng mỗi giao dịch có thể được mô tả bằng:

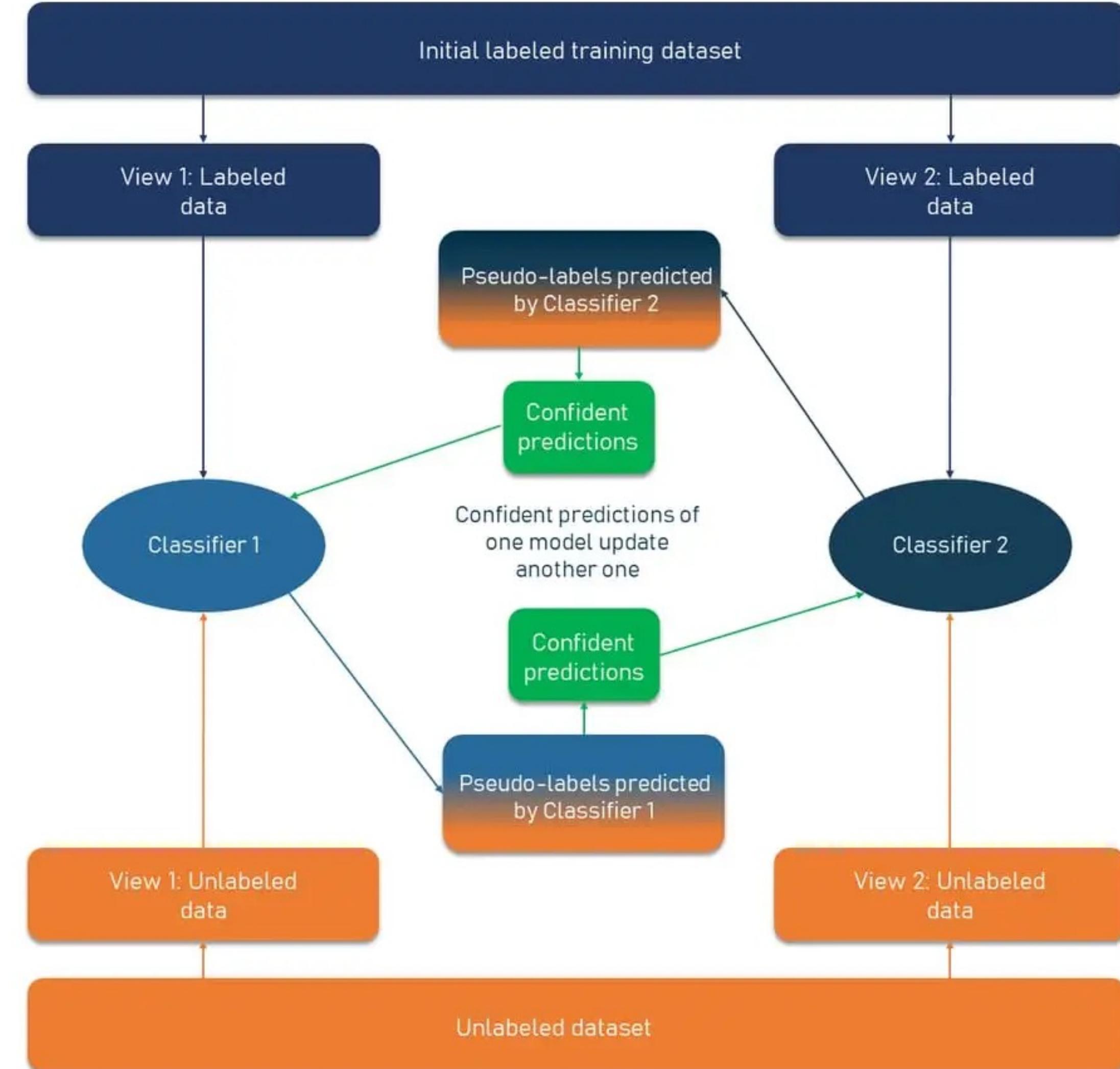
- View 1: Đặc trưng giao dịch
- View 2: Hành vi của khách hàng



- ❖ Thuật toán huấn luyện đồng thời (Co-training) bao gồm việc huấn luyện nhiều mô hình trên tập con hoặc góc nhìn khác nhau của dữ liệu.
- ❖ Mỗi mô hình sẽ gán nhãn cho các trường hợp chưa được gán nhãn dựa trên góc nhìn của nó, và chúng trao đổi dữ liệu đã được gán nhãn để học hỏi lẫn nhau.



CO-TRAINING



- ❖ Mỗi mẫu tách thành hai view:

$$x = (x^{(1)}, x^{(2)})$$

- ❖ Hai mô hình theo vòng lặp t : $h_t^{(1)}$ học trên $x^{(1)}$, $h_t^{(2)}$ học trên $x^{(2)}$.
- ❖ Xác suất gian lận trên mẫu chưa nhãn:

$$p_j^{(v)} = P(y = 1 | x_j^{(v)}, h_t^{(v)}), v \in \{1, 2\}$$

- ❖ Độ tin cậy: $conf^{(v)}(x_j) = \max(p_j^{(v)}, 1 - p_j^{(v)})$
- ❖ Ngưỡng tin cậy: $\tau \in (0.5, 1)$. Chỉ chọn pseudo—label nếu $conf(x_j) > \tau$

Để Co-training thành công, dữ liệu phải thoả mãn hai điều kiện tiên quyết

- ❖ **Tính đầy đủ:** Mỗi view (tập đặc trưng con) phải chứa đủ thông tin để phân loại chính xác nếu có đủ dữ liệu luân luyện.
 - View 1: đặc trưng giao dịch đủ để phát hiện các loại gian lận dựa trên quy tắc như giao dịch vượt hạn mức,...
 - View 2: đủ để phát hiện Account Takeover (ATO) khi thiết bị hoặc IP thay đổi đột ngột
- ❖ **Tính độc lập có điều kiện:** Với mỗi nhãn y cho trước, hai view $x^{(1)}$ và $x^{(2)}$ phải độc lập với nhau:
$$P(x^{(1)} | x^{(2)}, y) \approx P(x^{(1)} | y)$$
 - Nếu biết một giao dịch là gian lận ($y = 1$) việc biết số tiền giao dịch lớn ($x^{(1)}$) không nhất thiết giúp ta suy ra được IP của kẻ gian ($x^{(2)}$).
 - Sự độc lập này đảm bảo rằng sai lầm của mô hình 1 sẽ **không tương quan** chặt chẽ với sai lầm của mô hình 2, cho phép chúng sửa lỗi cho nhau hiệu quả.

❖ Bước 1: Khởi tạo

- Ban đầu, Nam tách các đặc trưng thành 2 view:
 1. View 1 bao gồm các đặc trưng giao dịch
 2. View 2 bao gồm các hành vi giao dịch
- Huấn luyện 2 mô hình $h_0^{(1)}$ và $h_0^{(2)}$ khác nhau để tăng sự đa dạng
- Với bài toán phân loại nhị phân, sử dụng hàm mất mát Binary Cross-Entropy :

$$\mathcal{L}_{sup}(\theta) = - \frac{1}{|D_L^{(v)}|} \sum_{(x^{(v)}, y) \in D_L^{(v)}} [y \log p^{(v)} + (1-y) \log(1 - p^{(v)})]$$

Với: $p^{(v)} = h_{\theta^{(v)}}(x^{(v)})$

- Chọn ngưỡng $\tau = 0.95$

❖ Bước 2: Mỗi view tự gán pseudo-label trên D_U

➤ Với mỗi $x_j \in D_U$, mô hình view v cho $p_j^{(v)}$

➤ Quy tắc gán nhãn giống self-training:

$$\hat{y}_j^{(v)} = \begin{cases} 1, & p_j^{(v)} \geq \tau \\ 0, & p_j^{(v)} \leq 1 - \tau \\ bỏ qua, & 1 - \tau < p_j^{(v)} < \tau \end{cases}$$

➤ Tập pseudo-labeled do view v chọn:

$$D_{p,t}^{(v)} = \{(x_j, \hat{y}_j^{(v)}) | conf^{(v)}(x_j) \geq \tau\}$$

❖ Bước 3: Trao đổi pseudo-label

➤ View 1 “dạy” view 2 bằng các mẫu rất tự tin:

$$D_{p,t}^{(1 \rightarrow 2)} = \left\{ \left(x_j^{(2)}, \hat{y}_j^{(1)} \right) \mid \left(x_j, \hat{y}_j^{(1)} \right) \in D_{p,t}^{(1)} \right\}$$

➤ Tương tự, view 2 “dạy” view 1:

$$D_{p,t}^{(2 \rightarrow 1)} = \left\{ \left(x_j^{(1)}, \hat{y}_j^{(2)} \right) \mid \left(x_j, \hat{y}_j^{(2)} \right) \in D_{p,t}^{(2)} \right\}$$

➤ Cập nhật tập train:

$$D_{train,t}^{(1)} = D_L^{(1)} \cup D_{p,t}^{(2 \rightarrow 1)},$$

$$D_{train,t}^{(2)} = D_L^{(2)} \cup D_{p,t}^{(1 \rightarrow 2)}$$

❖ Bước 4: Huấn luyện lại 2 mô hình

➤ Huấn luyện lại

$$h_{t+1}^{(1)} = \text{train}(D_{train,t}^{(1)}) , \quad h_{t+1}^{(2)} = \text{train}(D_{train,t}^{(2)})$$

➤ Hàm mất mát:

$$\mathcal{L}_{t+1}^{(v)} = -\frac{1}{|D_{train,t}^{(v)}|} \sum_{(x^{(v)},y) \in D_{train,t}^{(v)}} [y \log p^{(v)} + (1-y) \log(1-p^{(v)})]$$

➤ Tuỳ chọn “an toàn”: giảm trọng số pseudo-label

$$\mathcal{L}^{(v)} = \mathcal{L}_{sup}(D_L^{(v)}) + \lambda \mathcal{L}_{sup}(D_L^{(\bar{v} \rightarrow v)}), \quad 0 < \lambda \leq 1$$

❖ Bước 5: Lắp và tiêu chí dừng

- Lắp lại các bước 2, 3, 4 trong T vòng hoặc đến khi đạt điều kiện dừng
- Mô hình dừng quá trình huấn luyện nếu gặp các trường hợp sau:
 1. $|D_{p,t}^{(1)}| + |D_{p,t}^{(2)}|$ giảm mạnh / ≈ 0 .
 2. Metric validation không cải thiện trong k vòng.
 3. Hai mô hình trở nên quá giống nhau

Quay trở lại hành trình giải cứu Nam Chill

Trong 1 vòng của thuật toán Co-training, Nam thực hiện như sau:

- ❖ Xét giao dịch chưa nhãn $x_j \in D_u$:

- View 1 (giao dịch hiện tại) có:

amount = 50,000,000 VNĐ, time = 3 AM, location = nước ngoài

$$p_j^{(1)} = P(y = 1 | x_j^{(1)}, h_0^{(1)}) = 0.97 \Rightarrow \hat{y}_j^{(1)} = 1.$$

- View 2 (hành vi khách hàng) có:

số dư TB = 5,000,000 VNĐ, lần đầu giao dịch nước ngoài, tần suất tăng đột ngột

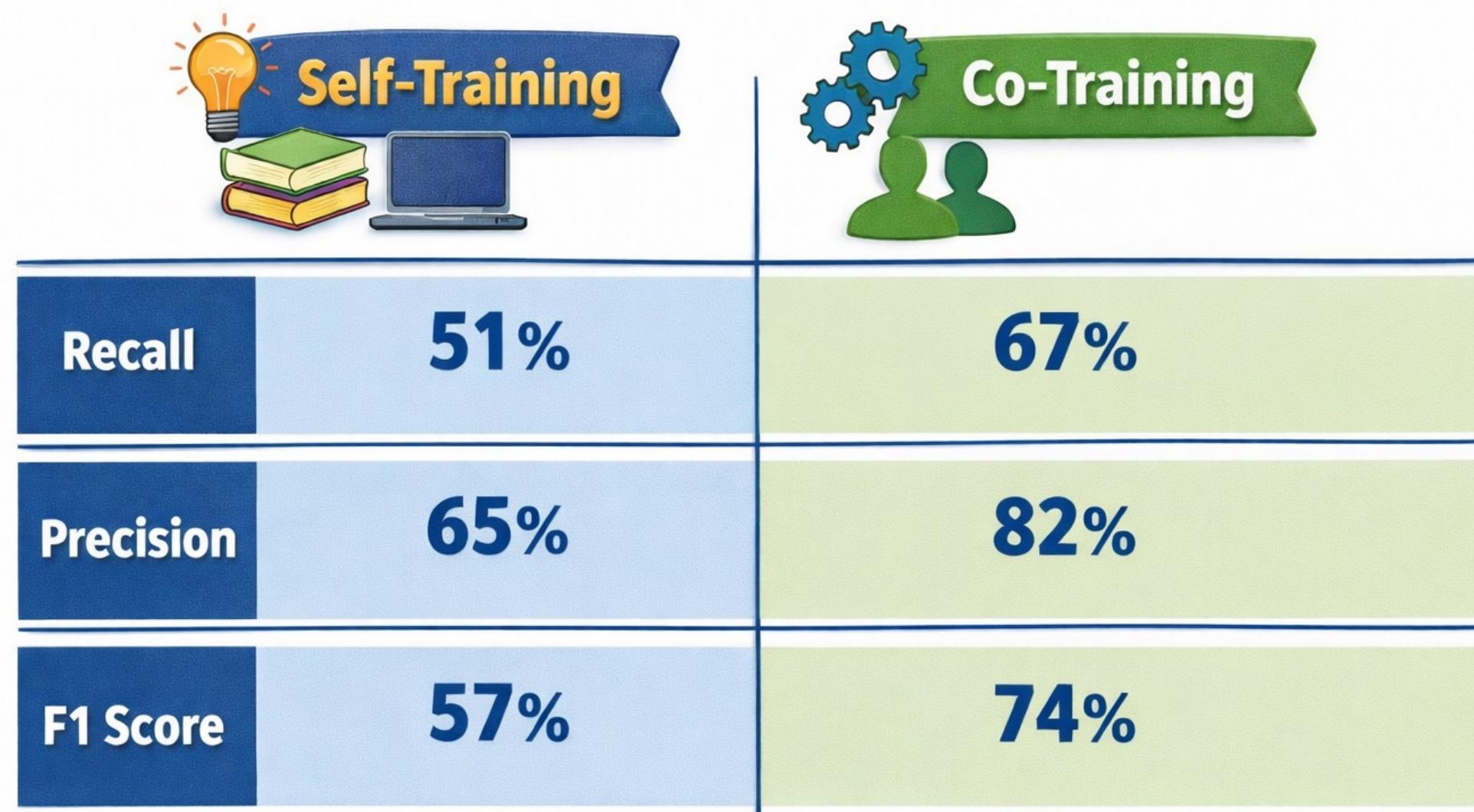
$$p_j^{(2)} = P(y = 1 | x_j^{(2)}, h_0^{(2)}) = 0.95 \Rightarrow \hat{y}_j^{(2)} = 1.$$

- ❖ Cross-teaching: View 1 gửi $(x_j^{(1)}, 1)$ sang train của view 2.

View 2 gửi $(x_j^{(2)}, 1)$ sang train của view 1.

Hành trình giải cứu Nam Chill

- ❖ Sau $T = 10$ vòng, co-training thường tăng recall mà vẫn giữ precision tốt hơn self-training (do giảm bias)



❖ Điều kiện áp dụng Co-training thành công

- Wang & Zhou, 2007) nhấn mạnh trực giác: nếu 2 mô hình ban đầu **không quá tệ** và có **difference/disagreement** **đủ lớn**, bound lỗi có thể giảm.

$$b_1 = \max \left\{ \frac{l b_0 + u a_0 - u d^{(t)}}{l}, 0 \right\}$$

trong đó $d^{(t)}$ là mức bất đồng (disagreement) giữa hai mô hình tại vòng t .

Vậy, $d^{(t)}$ càng lớn \rightarrow Hai mô hình dạy nhau càng nhiều \rightarrow khả năng cải thiện càng cao.

SELF-TRAINING VS CO-TRAINING

Thuật Toán	Ưu Điểm	Nhược Điểm	Ứng Dụng
Self-Training	<p>Đơn giản, không cần chia view</p> <p>Hiệu quả với dữ liệu cân bằng</p>	<p>Dễ lan truyền lỗi nếu threshold không tốt.</p>	Tự cập nhật mô hình từ dữ liệu mới.
Co-Training	<p>Bổ sung góc nhìn, giảm bias.</p> <p>Tốt cho dữ liệu đa chiều.</p>	<p>Cần view độc lập; phức tạp hơn.</p>	Phát hiện pattern phức tạp qua các đặc trưng.







Thank You!

DAI NAM
UNIVERSITY