

# Unnamed1

April 13, 2025

## **Câu 1 – Lý thuyết và phân tích mô hình (3 điểm)**

Giả sử bạn đang huấn luyện một mô hình phân loại nhị phân để phân biệt giữa email "spam" và "không spam".

a) Trình bày ý nghĩa và cách tính các độ đo sau: Accuracy, Precision, Recall, F1-score. (1.5 điểm)

b) Mô hình của bạn cho kết quả như sau trên tập kiểm tra:

- True Positive (TP): 80
- False Positive (FP): 20
- True Negative (TN): 90
- False Negative (FN): 10

Tính các độ đo trên và nhận xét ngắn gọn về mô hình. (1.5 điểm)

### **0.1 Confusion Matrix**

- True Positive (TP): 80
  - False Positive (FP): 20
  - True Negative (TN): 90
  - False Negative (FN): 10
-

## 0.2 Các công thức đo lường hiệu suất

### 0.2.1 1. Accuracy (Độ chính xác)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- 
- $\text{Accuracy} = (80 + 90) / (80 + 90 + 20 + 10) = 170 / 200 = 0.85$  (85%)
- **Ý Nghĩa:**
  - Accuracy cho biết tỷ lệ dự đoán đúng trên toàn bộ dữ liệu.
  - Đây là một độ đo tổng quát nhưng không phù hợp khi dữ liệu mất cân bằng
  - (ví dụ: 90% là âm tính, 10% là dương tính). ###

$$\text{Precision} = \frac{TP}{TP + FP}$$

- - $\text{Precision} = 80 / (80 + 20) = 80 / 100 = 0.80$  (80%)
  - **Ý nghĩa:**
    - Precision đo lường tỷ lệ các mẫu được dự đoán là dương tính thực sự là dương tính.
    - Thích hợp khi chi phí của sai dương tính (False Positive) là cao.
    - Ví dụ: chẩn đoán ung thư, xét nghiệm sai gây lo lắng và điều trị không cần thiết. ###
- ### 3. Recall (Độ nhạy / True Positive Rate)
- $\text{Recall} = TP / (TP + FN)$
  - $\text{Recall} = 80 / (80 + 10) = 80 / 90 = 0.8889$  (88.89%)
  - **Ý nghĩa:**
    - Recall cho biết trong tất cả các trường hợp dương tính thực sự, mô hình phát hiện được bao nhiêu.

- Quan trọng khi không được bỏ sót các trường hợp dương tính (ví dụ: phát hiện bệnh, gian lận tài chính). ### 4. **F1-score**

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- 
- $F1 = 2 \times (0.80 \times 0.8889) / (0.80 + 0.8889) = 2 \times 0.7111 / 1.6889 = 0.841$
- Ý nghĩa:
  - F1 là một độ đo cân bằng giữa precision và recall.
  - Nó đặc biệt hữu ích khi cần cân bằng giữa sai dương tính và sai âm tính.
  - F1 đạt giá trị cao nhất khi precision và recall đều cao. ### 5. **Specificity (True Negative Rate)**
- $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$
- $\text{Specificity} = 90 / (90 + 20) = 90 / 110 = 0.8182 \text{ (81.82\%)}$
- Ý nghĩa:
  - Specificity cho biết trong số các trường hợp âm tính thực sự, có bao nhiêu được dự đoán đúng.
  - Rất quan trọng trong các ứng dụng cần đảm bảo không gây dương tính giả
  - (ví dụ: xét nghiệm tiền sản, kiểm soát chất lượng).

### 0.3 Tổng kết

Độ đo	Giá trị
Accuracy	85.00%
Precision	80.00%
Recall	88.89%
F1-score	84.10%
Specificity	81.82%

Một mô hình hồi quy được huấn luyện để dự đoán giá nhà dựa trên diện tích và số phòng ngủ. Trên tập kiểm tra, sinh viên thu được các giá trị dự đoán và thực tế như sau:

Giá trị thực tế (triệu VND)	Giá trị dự đoán (triệu VND)
3.0	2.5
4.5	4.0
5.0	5.2
6.0	5.5
3.5	3.0

- a) Tính các độ đo sau: MAE, RMSE,  $R^2$  (1.5 điểm)
- b) Nhận xét mô hình dựa vào kết quả trên (0.5 điểm)

```
[ ]: # bảng dữ liệu
# giá trị thực tế
y_true = [3.0, 4.5, 5.0, 6.0, 3.5]
# giá trị dự đoán
y_pred = [2.5, 4.0, 5.2, 5.5, 3.0]

# tính toán sai số
# trung bình tuyệt đối
def mean_absolute_error(y_true, y_pred):
    n = len(y_true)
    total_error = 0
    for i in range(n):
        total_error += abs(y_true[i] - y_pred[i])
    return total_error / n

# căn bậc trung bình bình phương RMSE
def root_mean_squared_error(y_true, y_pred):
    n = len(y_true)
    total_error = 0
    for i in range(n):
        total_error += (y_true[i] - y_pred[i]) ** 2
    return (total_error / n) ** 0.5

# tính R2
```

```

def r_squared(y_true, y_pred):
    n = len(y_true)
    ss_total = 0
    ss_residual = 0
    mean_y_true = sum(y_true) / n
    for i in range(n):
        ss_total += (y_true[i] - mean_y_true) ** 2
        ss_residual += (y_true[i] - y_pred[i]) ** 2
    return 1 - (ss_residual / ss_total)

# tính toán các chỉ số
mae = mean_absolute_error(y_true, y_pred)
rmse = root_mean_squared_error(y_true, y_pred)
r2 = r_squared(y_true, y_pred)
# in kết quả
print("Mean Absolute Error (MAE):", mae)
print("Root Mean Squared Error (RMSE):", rmse)
print("R-squared (R2):", r2)

```

Mean Absolute Error (MAE): 0.44000000000000006  
 Root Mean Squared Error (RMSE): 0.4560701700396552  
 R-squared (R2): 0.8175438596491228

MAE = 0.44 triệu VND, tức sai số trung bình khoảng 440 nghìn đồng, là khá nhỏ.

RMSE = 0.456 triệu VND xác nhận rằng sai số nghiêm trọng không nhiều.

$R^2 = 0.8175$  ( 81.75%) là mức giải thích khá tốt, tức mô hình giải thích được ~82% phương sai trong dữ liệu.

Kết luận: Mô hình dự đoán tương đối chính xác, sai số thấp và độ phù hợp tốt. Có thể cải thiện thêm bằng cách thử mô hình phi tuyến, thêm biến hoặc tính chỉnh tham số.

### Câu 3 – Ứng dụng độ đo để chọn mô hình (2 điểm)

Giả sử bạn đang phát triển một hệ thống phát hiện bệnh từ ảnh X-quang phổi. Bạn đã thử 3 mô hình khác nhau và thu được kết quả như sau trên tập kiểm tra:

Mô hình	Accuracy	Precision	Recall	F1-score
M1	0.95	0.92	0.65	0.76
M2	0.91	0.85	0.80	0.82
M3	0.90	0.75	0.95	0.84

a) Mô hình nào phù hợp hơn nếu bạn ưu tiên phát hiện đúng càng nhiều bệnh nhân thật sự càng tốt? Giải thích lý do. (1 điểm)

b) Mô hình nào nên chọn nếu bạn muốn cân bằng giữa Precision và Recall? (1 điểm)

- a) mô hình phù hợp để phát hiện đúng càng nhiều bệnh nhân thật sự là **M3** bởi vì tỷ lệ độ nhạy của mô hình này là tốt nhất, phù hợp với việc nhận diện sự bất thường trong nhận biết bệnh nhân và bất thường trong gia dịch
- b) nếu muốn chọn mô hình cân bằng giữa **precision** và **recall** thì ta nên chọn mô hình M3 vì đây là mô hình có chỉ số **F1** cao nhất

### Câu 4 – Nhận diện lỗi và cải tiến mô hình (3 điểm)

Bạn nhận thấy mô hình phân loại bạn huấn luyện có Accuracy rất cao (>95%) nhưng F1-score lại thấp (~60%).

a) Giải thích khả năng xảy ra tình trạng này. (1 điểm)

b) Nếu tập dữ liệu bị mất cân bằng, bạn sẽ áp dụng các kỹ thuật nào để cải thiện mô hình? Liệt kê ít nhất 3 kỹ thuật. (1 điểm)

c) Giải thích tại sao Accuracy không phải lúc nào cũng là độ đo tốt nhất. Cho ví dụ minh họa. (1 điểm)

a)Giải thích khả năng xảy ra tình trạng này

**0.3.1** Tình trạng này rất có thể xảy ra do tập dữ liệu bị mất cân bằng (imbalanced dataset).

- Accuracy (Độ chính xác) đo lường tỷ lệ tổng số dự đoán đúng trên toàn bộ tập dữ liệu  $((TP + TN) / \text{Tổng})$ . Khi dữ liệu mất cân bằng, lớp đa số (majority class) chiếm tỷ lệ rất lớn. Nếu mô hình chỉ đơn giản dự đoán hầu hết các điểm dữ liệu thuộc về

lớp đa số, nó vẫn có thể đạt được Accuracy cao, vì số lượng dự đoán đúng cho lớp đa số (TN) là rất lớn.

- F1-score là trung bình điều hòa (harmonic mean) của Precision và Recall. Nó nhạy cảm hơn với hiệu suất của mô hình trên lớp thiểu số (minority class).

+ Precision ( $TP / (TP + FP)$ ): Độ chuẩn xác - Tỷ lệ các điểm được dự đoán là dương tính mà thực sự là dương tính.

+ Recall ( $TP / (TP + FN)$ ): Độ phủ - Tỷ lệ các điểm thực sự dương tính mà mô hình dự đoán đúng.

- Trong trường hợp dữ liệu mất cân bằng, mô hình có Accuracy cao nhưng F1-score thấp thường là do nó dự đoán kém trên lớp thiểu số. Nó có thể có Recall thấp (bỏ lỡ nhiều điểm thuộc lớp thiểu số - nhiều FN) hoặc Precision thấp (dự đoán sai nhiều điểm thành lớp thiểu số - nhiều FP), hoặc cả hai. Điều này kéo F1-score xuống thấp, mặc dù Accuracy vẫn cao do dự đoán đúng phần lớn các điểm thuộc lớp đa số.

b) Nếu tập dữ liệu bị mất cân bằng bạn sẽ áp dụng các kỹ thuật nào để cải thiện mô hình? Liệt kê.

Có nhiều kỹ thuật để cải thiện mô hình khi gặp dữ liệu mất cân bằng. Dưới đây là ít nhất 3 kỹ thuật phổ biến:

1. Thay đổi tỷ lệ mẫu (Resampling):

- Oversampling (Lấy mẫu quá mức): Tăng số lượng mẫu của lớp thiểu số. Ví dụ: Lặp lại ngẫu nhiên các mẫu thiểu số, hoặc tạo ra các mẫu tổng hợp mới (ví dụ: kỹ thuật SMOTE - Synthetic Minority Over-sampling Technique).
- Undersampling (Lấy mẫu dưới mức): Giảm số lượng mẫu của lớp đa số. Ví dụ: Loại bỏ ngẫu nhiên các mẫu đa số, hoặc sử dụng các kỹ thuật chọn lọc hơn (ví dụ: Tomek links, NearMiss).
- Kết hợp Oversampling và Undersampling.

2. Sử dụng các độ đo đánh giá khác: Thay vì chỉ dựa vào Accuracy, tập trung vào các độ đo phù hợp hơn cho dữ liệu mất cân bằng như Precision, Recall, F1-score, AUC-PR (Area Under the Precision-Recall Curve), G-mean, hoặc Matthews Correlation Coefficient (MCC).

3. Cost-Sensitive Learning (Học nhạy cảm với chi phí): Gán trọng số (chi phí) khác nhau cho việc phân loại sai các lớp. Cụ thể là tăng chi phí/hình phạt cho việc phân loại sai lớp thiểu số. Nhiều thuật toán cho phép thiết lập trọng số lớp (ví dụ: Logistic Regression, SVM, Decision Trees).

4. Sử dụng thuật toán phù hợp: Một số thuật toán hoặc biến thể thuật toán hoạt động tốt hơn với dữ liệu mất cân bằng, ví dụ như các phương pháp Ensemble (Balanced Random Forest, EasyEnsemble, RUSBoost).

c) Giải thích tại sao Accuracy không phải lúc nào cũng là độ đo tốt nhất. Cho ví dụ minh họa.

Accuracy không phải lúc nào cũng là độ đo tốt nhất, đặc biệt là khi tập dữ liệu bị mất cân bằng hoặc khi chi phí của các loại lỗi là khác nhau.

***Giải thích :*** - Accuracy có thể đưa ra một bức tranh sai lệch về hiệu suất của mô hình trên các tập dữ liệu mất cân bằng. Một mô hình có thể đạt Accuracy rất cao chỉ bằng cách dự đoán thiên về lớp đa số, trong khi hoàn toàn thất bại trong việc xác định lớp thiểu số, mà lớp thiểu số này

thường lại là lớp quan trọng hơn (ví dụ: phát hiện bệnh hiếm, giao dịch gian lận). - Ví dụ minh họa: Xét bài toán phát hiện gian lận thẻ tín dụng. Giả sử trong 1000 giao dịch, chỉ có 10 giao dịch là gian lận (lớp thiểu số - 1%) và 990 giao dịch là hợp lệ (lớp đa số - 99%). + Một mô hình đơn giản luôn dự đoán “không gian lận” cho mọi giao dịch sẽ đạt  $\text{Accuracy} = 990/1000 = 99\%$ . + Mặc dù Accuracy rất cao, mô hình này hoàn toàn vô dụng vì nó không phát hiện được bất kỳ giao dịch gian lận nào ( $\text{Recall cho lớp gian lận} = 0/10 = 0\%$ ). + Trong trường hợp này, việc bỏ sót một giao dịch gian lận (False Negative) thường có hậu quả nghiêm trọng hơn nhiều so với việc cảnh báo nhầm một giao dịch hợp lệ (False Positive). Do đó, các độ đo như Recall, Precision, và F1-score sẽ phản ánh đúng hơn hiệu quả thực sự của mô hình trong việc giải quyết vấn đề.