

# Hiểu về RMSE – Root Mean Squared Error trong XGBoost

## 1. RMSE là gì?

**RMSE (Root Mean Squared Error)** – là căn bậc hai của trung bình bình phương sai số, dùng để đo **mức độ sai lệch giữa giá trị thực tế và giá trị dự đoán của mô hình**.

“ □ RMSE càng thấp → mô hình càng chính xác.

## 2. Khi nào dùng RMSE?

RMSE thường được dùng khi bạn giải quyết bài toán **hồi quy** – tức là khi **biến mục tiêu (Y)** là **liên tục**, ví dụ:

- Dự đoán giá nhà
- Dự đoán doanh thu
- Dự đoán nhiệt độ, v.v.

## 3. Công thức RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Trong đó:

- $y_i$ : Giá trị thực tế
- $\hat{y}_i$ : Giá trị dự đoán
- $n$ : Tổng số điểm dữ liệu

## 4. Trực quan hóa RMSE

Hãy tưởng tượng:

- Bạn có trục X và trục Y.

- Đường thẳng màu xanh biểu diễn **mô hình dự đoán**.
- Các **dấu chấm** là **giá trị thực tế**.
- **Mũi tên từ chấm tới đường thẳng** chính là **sai số (error)**.
- RMSE là **trung bình độ dài của các mũi tên đó**, rồi lấy căn bậc hai.

## 5. Ví dụ cụ thể

Giả sử bạn đang dự đoán điểm thi của học sinh, và bạn có:

| Học sinh | Điểm thực tế (Y) | Điểm dự đoán (Ŷ) |
|----------|------------------|------------------|
| A        | 8.0              | 7.5              |
| B        | 6.5              | 7.0              |
| C        | 9.0              | 8.5              |

Tính RMSE:

$$RMSE = \sqrt{\frac{(8.0 - 7.5)^2 + (6.5 - 7.0)^2 + (9.0 - 8.5)^2}{3}} = \sqrt{\frac{0.25 + 0.25 + 0.25}{3}} = \sqrt{0.25} = 0.5$$

👉 **RMSE = 0.5 điểm**, tức là mô hình sai lệch trung bình 0.5 điểm so với thực tế.

## 6. RMSE bao nhiêu là tốt?

Không có “một con số tuyệt đối” nào cho RMSE tốt, vì:

- Phụ thuộc vào **phạm vi dữ liệu** (ví dụ: sai lệch 10 là lớn nếu điểm thi từ 0–10, nhưng nhỏ nếu doanh thu tính bằng triệu đô).
- Phụ thuộc vào **độ khó của bài toán**.

→ 📌 Hãy so sánh RMSE của **mô hình hiện tại với các mô hình khác** để đánh giá.

## 7. RMSE trong XGBoost

XGBoost tự động **tối ưu hóa để giảm RMSE** (hoặc một hàm lỗi tương tự) trong quá trình huấn luyện khi bạn dùng cho bài toán hồi quy.

## Tóm tắt:

| Nội dung | Ý nghĩa                                      |
|----------|--|
| RMSE     | Đo sai số trung bình giữa thực tế và dự đoán |

| Nội dung                       | Ý nghĩa                                |
|--------------------------------|--|
| Thích hợp cho                  | Bài toán hồi quy                       |
| RMSE nhỏ hơn → mô hình tốt hơn | Đúng, nhưng cần phụ thuộc vào ngữ cảnh |
| Dùng để                        | So sánh giữa các mô hình               |

Tác giả: **Đỗ Ngọc Tú**  
Công Ty Phần Mềm **VHTSoft**

Phiên bản #1  
Được tạo 15 tháng 5 2025 10:27:35 bởi Đỗ Ngọc Tú  
Được cập nhật 15 tháng 5 2025 10:31:55 bởi Đỗ Ngọc Tú