# Ma trận nhầm lẫn

Ma trận nhầm lẫn là một bảng dùng để đánh giá hiệu suất của mô hình phân loại. Nó hiển thị số lượng dự đoán đúng và sai của mô hình so với các phân loại thực tế trong dữ liệu kiểm thử. Ma trận được cấu trúc như sau:

|  |  |  |
| --- | --- | --- |
|  | Dự đoán Dương tính | Dự đoán Âm tính |
| Thực tế Dương tính | True Positive (TP) | False Negative (FN) |
| Thực tế Âm tính | False Positive (FP) | True Negative (TN) |

Các chỉ số chính:  
- Độ chính xác (Accuracy): Tỷ lệ giữa số dự đoán đúng (TP + TN) và tổng số dự đoán.  
 Accuracy = (TP + TN) / (TP + TN + FP + FN)  
- Độ chính xác (Precision): Tỷ lệ giữa số dự đoán dương tính đúng và tổng số dự đoán dương tính.  
 Precision = TP / (TP + FP)  
- Độ nhạy (Recall): Tỷ lệ giữa số dự đoán dương tính đúng và tổng số dương tính thực tế.  
 Recall = TP / (TP + FN)  
- F1 Score: Trung bình điều hòa của độ chính xác và độ nhạy.  
 F1 = 2 \* (Precision \* Recall) / (Precision + Recall)

Ví dụ:  
Với y\_true = [0, 0, 0, 0, 1, 1, 1, 2, 2, 2] và y\_pred = [0, 1, 0, 2, 1, 1, 0, 2, 1, 2], ma trận nhầm lẫn là:  
[[2, 1, 1], [1, 2, 0], [0, 1, 2]]  
Độ chính xác: 0.6 (60%)

Ma trận nhầm lẫn chuẩn hóa:  
Mỗi hàng của ma trận nhầm lẫn được chia cho tổng các phần tử trên hàng đó để có ma trận nhầm lẫn chuẩn hóa, giúp nhìn rõ hơn hiệu suất của mô hình.

Minh họa:  
Ma trận nhầm lẫn thường được minh họa bằng biểu đồ nhiệt để dễ dàng nhận biết hiệu suất của mô hình trên các lớp khác nhau.