

BÁO CÁO VỀ MÔ HÌNH

Xây dựng và Triển khai Hệ thống Học máy Ứng dụng

Báo cáo cần tập trung vào quy trình kỹ thuật, giải thích lý do lựa chọn và phân tích kết quả. Một báo cáo cần có các nội dung sau

Cấu trúc nội dung

Giới thiệu bài toán

Trình bày ngắn gọn lại bài toán học máy mà nhóm hướng tới giải quyết

Tổng quan dữ liệu đầu vào

- **Dữ liệu huấn luyện (Train/Validation/Test):** Nêu rõ tỷ lệ chia dữ liệu (ví dụ: 70/15/15 hoặc 80/20). Giải thích lý do chọn tỷ lệ này.
- **Tiền xử lý:** Tóm tắt các bước đã áp dụng lên dữ liệu trước khi đưa vào mô hình (*chuẩn hóa, mã hóa, xử lý dữ liệu thiếu, tăng cường dữ liệu nếu có*).

Lựa chọn Mô hình & Kiến trúc

- **Mô hình sử dụng (*Lưu ý ít nhất 3 mô hình*):** Nêu tên mô hình/thuật toán (ví dụ: Random Forest, SVM, CNN, LSTM, Linear Regression...).
- **Lý do lựa chọn:** Tại sao chọn thuật toán này?
- **Kiến trúc chi tiết (**Đối với Deep Learning**):**
 - Vẽ sơ đồ kiến trúc.
 - Mô tả số lượng tham số.
 - Hàm kích hoạt được sử dụng.

Cấu hình huấn luyện

- **Hàm mất mát:** Sử dụng hàm gì (MSE, Cross-entropy, Hinge loss...) hoặc tự xây dựng? Tại sao?
- **Thuật toán tối ưu:** SGD, Adam, RMSprop...? Tốc độ học (Learning rate) là bao nhiêu? Có sử dụng Learning Rate Scheduler không?
- **Siêu tham số:**
 - Liệt kê các tham số chính (Batch size, Epochs, Number of Trees, Kernel type, Regularization L1/L2...).
 - Phương pháp tinh chỉnh tham số (Grid Search, Random Search hay thử nghiệm thủ công).

Kết quả thực nghiệm

- **Biểu đồ quá trình học (Learning Curves):**
 - Bắt buộc phải có biểu đồ **Loss** và **Accuracy** (hoặc metric khác) trên tập Train và Validation theo từng Epoch.
 - *Yêu cầu:* Nhận xét biểu đồ (Mô hình có hội tụ không? Có bị dao động mạnh không?).
- **Đánh giá trên tập Test:**
 - **Các chỉ số định lượng:** Accuracy, Precision, Recall, F1-Score (đối với bài toán phân loại); RMSE, MAE (đối với bài toán hồi quy).
 - **Confusion Matrix:** Hiển thị ma trận nhầm lẫn để phân tích các lớp bị dự đoán sai nhiều nhất.

Thảo luận & Phân tích lỗi

- **Hiện tượng Overfitting/Underfitting:** Dựa vào kết quả trên tập Train và Test để kết luận mô hình đang ở trạng thái nào. Cách khắc phục đã thực hiện (Dropout, Early Stopping...).
- **Phân tích các trường hợp sai:** Lấy ví dụ cụ thể về dữ liệu mà mô hình dự đoán sai. Đưa ra giả thuyết tại sao sai (do nhiễu, do đặc trưng không rõ ràng...).
- **So sánh:** Lập bảng so sánh hiệu năng giữa các mô hình.

Yêu cầu về trình bày

1. **Văn phong:** Khoa học, khách quan, ngắn gọn, tránh văn nói.
2. **Hình ảnh/Bảng biểu:**
 - Mọi hình ảnh (đồ thị, sơ đồ) phải có chú thích bên dưới.
 - Mọi bảng số liệu phải có tiêu đề bên trên và đơn vị đo lường rõ ràng.
3. **Trích dẫn tài liệu:** Nếu sử dụng kiến trúc mô hình từ bài báo khoa học nào (ví dụ: ResNet, BERT), phải trích dẫn nguồn.