**Giới thiệu bài toán (Problem Definition)**

* Mô tả bài toán cụ thể: phân loại/phân cụm/dự đoán/nhận dạng...
* Mục tiêu của bài toán.
* Ứng dụng thực tiễn (nếu có).

**Dữ liệu (Dataset)**

* Nguồn dữ liệu (CSV, Kaggle, thực tế,...)
* Kích thước: số dòng (samples), số cột (features), kiểu dữ liệu.
* Mô tả từng thuộc tính, nhãn.
* Tiền xử lý:  
  + Xử lý missing data.
  + Mã hóa (label encoding, one-hot).
  + Chuẩn hóa/standardization (nếu cần).
  + Phân chia tập train/test (hold-out hoặc cross-validation).

**Lựa chọn mô hình và thuật toán**

* Tại sao chọn mô hình đó? (Perceptron, KNN, SVM, Decision Tree,...)
* Nếu là phân cụm (KMeans, Agglomerative Clustering):  
  + Cách chọn số cụm **k** (dùng **Elbow**, **Silhouette Score**).
  + Trực quan hóa cụm (PCA, TSNE nếu nhiều chiều).
* Nếu là phân loại:  
  + Cách huấn luyện.
  + Tham số: max\_iter, learning\_rate, v.v.

**Huấn luyện và đánh giá**

* Thông số mô hình sau khi huấn luyện (trọng số, bias,...)
* Đánh giá:  
  + Accuracy, Precision, Recall, F1-score (với phân loại).
  + Confusion matrix.
  + Nếu phân cụm: Silhouette Score, nội lực (inertia).
* Vẽ đồ thị: learning curve, decision boundary, confusion matrix.

**Phân tích kết quả**

* Nhận xét mô hình hoạt động tốt ở đâu, còn yếu ở đâu.
* Ảnh hưởng của việc thay đổi tham số (như k, eta, max\_iter).
* Có thể cải thiện gì thêm? (chuyển mô hình, tăng dữ liệu,...)

**Kết luận**

* Tóm tắt kết quả đạt được.
* Hạn chế và hướng phát triển tiếp theo.