

Pha huấn luyện

Đầu ra huấn luyện
($\mathbf{y}_{\text{train}}$)

Thông tin biết
trước về dữ liệu

Dữ liệu huấn luyện
thô (đầu vào)

Bộ trích chọn
đặc trưng

Đặc trưng
được trích
chọn ($\mathbf{X}_{\text{train}}$)

Thuật toán
phân loại,
hồi quy,
phân cụm,...

Dữ liệu kiểm tra
thô (đầu vào)

Bộ trích chọn
đặc trưng

Đặc trưng
được trích
chọn (\mathbf{X}_{test})

Thuật toán
phân loại,
hồi quy,
phân cụm,...

Pha kiểm tra

Đầu ra dự đoán
(\mathbf{y}_{test})