

# Bài tập cuối kỳ 2024

Môn học: Thiết kế và phân tích thực nghiệm

## 1 Dữ liệu và bài toán cần thực hiện

**ys1a** là bộ dữ liệu về ứng suất phát sinh (yield strength) của các hợp kim entropy cao (high-entropy), trong đó biến **ys** là ứng suất phát sinh, đo bằng MPa (megapascal), **composition** là công thức hóa học của hợp kim. Các biến **vec**, **deltachi**, **delta**, **deltahmix**, **deltasmix** được tính toán từ công thức hóa học. Chúng ta cần dự đoán **ys** từ **vec**, **deltachi**, **delta**, **deltahmix**, và **deltasmix**. Để xây dựng các mô hình dự đoán, cần chia ngẫu nhiên tập dữ liệu **ys1a** thành hai tập con là tập dữ liệu huấn luyện và tập dữ liệu kiểm thử. Các độ đo của mô hình dự đoán được sử dụng trong bài tập này là RMSE (Root Mean Squared Error) và MAPE (Mean Absolute Percentage Error).

## 2 Thí nghiệm CRD

Hãy thiết kế và phân tích thí nghiệm CRD (Completely Randomized Design) để so sánh hai độ đo RMSE và MAPE của mô hình dự đoán **ys** sử dụng hồi qui tuyến tính khi độ lớn của tập dữ liệu kiểm thử là 10%, 20%, và 30% của **ys1a**.

## 3 Thí nghiệm CRFD

Hãy thiết kế và phân tích thí nghiệm CRFD (Completely Randomized Factorial Design) để so sánh hai độ đo RMSE và MAPE của mô hình dự đoán **ys** sử dụng SVR (Support Vector Regression) với các điều kiện sau:

- Sử dụng RBF kernel (Radial Basis Function)
- Các siêu tham số:  $C$  nhận giá trị 100, 200 và  $\gamma$  nhận giá trị 0.1, 0.5
- Tập dữ liệu kiểm thử chiếm 20% của **ys1a**

## 4 Yêu cầu chung cho thiết kế và phân tích thí nghiệm

- Cần xác định rõ ràng random seed để có thể lặp lại kết quả thí nghiệm
- Mức ý nghĩa  $\alpha = 0.05$ , power = 0.8
- Xác định rõ các biến độc lập, biến phụ thuộc, đơn vị thí nghiệm
- Ước tính số lần lặp (replication) căn cứ vào mức ý nghĩa, power, độ phân giải... bằng cách sử dụng một cách thích hợp các hàm **Fpower1**, **Fpower2** của thư viện **daewr**
- Phân tích kết quả thí nghiệm bằng phương pháp ANOVA (bỏ qua kiểm tra các giả định của mô hình tuyến tính)
- Có thể thực hiện thí nghiệm bằng ngôn ngữ Python hoặc R

## 5 Nộp bài

- Mỗi nhóm gửi một file ZIP chứa báo cáo định dạng PDF, code và dữ liệu kết quả thí nghiệm định dạng CSV vào địa chỉ email `chaunh@vnu.edu.vn`, với subject là [TKPTTN2024] <Tên nhóm>. Ví dụ: [TKPTTN2024] Nhóm 2.
- Hạn nộp bài: **23:59 ngày 25/12/2024**.