| **TRƯỜNG ĐH KHOA HỌC TỰ NHIÊN**  **KHOA CÔNG NGHỆ THÔNG TIN** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do – Hạnh Phúc**  Tp. Hồ Chí Minh, ngày 5 tháng 6 năm 2022 |
| --- | --- |

**TÓM TẮT SEMINAR – KAGGLE COMPETITION**

* **Nội dung chính**: Seminar “Kaggle Competition - House Prices - Advanced regression techniques”.
* **Thời gian**: Chủ Nhật, ngày 5 tháng 6 năm 2022
* **Nội dung seminar**:
* *Chủ đề*: Kaggle Competition - House Prices - Advanced regression techniques
* *Tóm tắt:*
  + Đây là một cuộc thi được tổ chức trên trang web Kaggle, mục tiêu của cuộc thi là tìm ra những mô hình học máy để dự đoán giá nhà dựa trên những thuộc tính được cung cấp. Mô hình sẽ càng được xếp hạng cao nếu sai số giữa dự đoán và thực tế càng nhỏ.
  + Các hạng mục trình bày chính
    - Tổng quan về dữ liệu, phân tích và tiền xử lí dữ liệu.
    - Xây dựng mô hình tiền xử lí dữ liệu.
    - Lựa chọn những mô hình hồi quy để thực hiện dự đoán: Linear Regression, Decision Tree Regression, Ridge Regression, Lasso Lars, SVR, Gradient Boosting, Random Forest Regression, Logistic Regression, Bayesian Ridge.
    - Thực nghiệm: huấn luyện các mô hình trên dữ liệu đã cung cấp, tính toán độ lỗi các mô hình và thực hiện dự đoán trên tập test của Kaggle.
* Tổng kết: sau khi huấn luyện các mô hình trên dữ liệu đã cung cấp, mô hình cho kết quả tốt nhất là Gradient Boosting với số điểm 0.13683 (sử dụng độ đo RMSLE và dữ liệu test của Kaggle).