

## Báo cáo Tóm tắt Kết quả Dự đoán Giá Nhà ở California

### Đề bài

Dự án này nhằm mục đích xây dựng mô hình Deep Learning để dự đoán giá nhà trung bình ở California dựa trên các đặc điểm của các khối dân cư. Dữ liệu được sử dụng là bộ dữ liệu California Housing từ scikit-learn.

### Mô hình

Mô hình được sử dụng là mạng nơ-ron (Neural Network) với các lớp ẩn (hidden layers) và hàm kích hoạt (activation function) được điều chỉnh thông qua thử nghiệm với nhiều cấu hình siêu tham số khác nhau.

### Siêu tham số

Các cấu hình siêu tham số được thử nghiệm bao gồm:

- **baseline:** Mạng cơ sở với 2 lớp ẩn (64, 32 nơ-ron), hàm kích hoạt ReLU, không có dropout, tốc độ học 0.001.
- **deep\_network:** Mạng sâu hơn với 4 lớp ẩn (128, 64, 32, 16 nơ-ron), hàm kích hoạt ReLU, dropout 0.2, tốc độ học 0.001.
- **regularized:** Mạng với 2 lớp ẩn (64, 32 nơ-ron), hàm kích hoạt ReLU, dropout 0.3, tốc độ học 0.001, L2 regularization 0.001.
- **wider\_network:** Mạng rộng hơn với 2 lớp ẩn (256, 128 nơ-ron), hàm kích hoạt ReLU, dropout 0.2, tốc độ học 0.001, L2 regularization 0.0005.
- **learning\_rate\_tuned:** Mạng với 2 lớp ẩn (64, 32 nơ-ron), hàm kích hoạt ReLU, dropout 0.2, tốc độ học 0.0005, L2 regularization 0.0005.
- **leaky\_relu:** Mạng với 3 lớp ẩn (64, 32, 16 nơ-ron), hàm kích hoạt LeakyReLU, dropout 0.2, tốc độ học 0.001, L2 regularization 0.0001.

### Kết quả

Kết quả được đánh giá dựa trên RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), và R-squared ( $R^2$ ) trên tập kiểm tra.

### Bảng kết quả:

Configuration	Mean RMSE	Std RMSE	Mean MAE	Std MAE	Mean R <sup>2</sup>	Std R <sup>2</sup>
baseline	0.7254	0.0083	0.5273	0.0047	0.6031	0.0080
deep_network	0.7072	0.0091	0.5109	0.0049	0.6242	0.0095
regularized	0.7154	0.0065	0.5199	0.0035	0.6152	0.0061
wider_network	0.7047	0.0087	0.5093	0.0052	0.6272	0.0089
learning_rate_tuned	0.7128	0.0079	0.5178	0.0042	0.6181	0.0074
leaky_relu	0.7091	0.0076	0.5127	0.0039	0.6224	0.0072

**Cấu hình tốt nhất:** wider\_network với RMSE trung bình là 0.7047 và R<sup>2</sup> trung bình là 0.6272.

Nhận xét

- Mô hình Deep Learning đã đạt được hiệu suất tốt trong việc dự đoán giá nhà ở California.
- Cấu hình wider\_network cho thấy kết quả tốt nhất, cho thấy việc tăng số lượng nơ-ron trong các lớp ẩn có thể cải thiện hiệu suất.
- Việc sử dụng dropout và L2 regularization giúp giảm thiểu overfitting.
- Các biểu đồ log từ MLflow cung cấp cái nhìn trực quan về quá trình huấn luyện và đánh giá mô hình.

**Biểu đồ mất mát (loss) và MAE trong quá trình huấn luyện:**

