**TÓM TẮT BÁO CÁO**

**1. Giới thiệu**

Thị trường bất động sản là một phần quan trọng của nền kinh tế toàn cầu. Quyết định mua hoặc bán nhà là vấn đề tài chính quan trọng, do đó, dự đoán giá nhà chính xác là điều cần thiết. Giá bất động sản liên tục biến động do nhiều yếu tố như vị trí, kích thước, đặc điểm và điều kiện kinh tế vĩ mô. Việc dự đoán chính xác giá nhà là rất quan trọng đối với các bên liên quan như người mua, người bán, đại lý và nhà đầu tư để đảm bảo định giá tài sản phù hợp. Dữ liệu bất động sản có thể nhiễu và có độ biến động cao theo thời gian, bị ảnh hưởng bởi xu hướng thị trường, thay đổi kinh tế và các sự kiện bên ngoài. Độ chính xác cao trong các mô hình dự đoán là rất quan trọng vì sai sót có thể dẫn đến hậu quả tài chính đáng kể. Mặc dù đã có nhiều nghiên cứu sử dụng các kỹ thuật học máy, nhưng vẫn tồn tại những khoảng trống, đặc biệt là nhu cầu phân tích kịch bản thời gian thực và xem xét những thay đổi kinh tế gần đây

**2. Mục tiêu**

* Phát triển mô hình dự báo giá nhà mạnh và chính xác dựa trên dữ liệu thực tế.
* Nâng cao độ chính xác bằng các bước làm sạch dữ liệu, khử ngoại lai, giảm chiều, và kỹ thuật tạo đặc trưng (feature engineering).
* Triển khai mô hình như một ứng dụng web thân thiện, hoạt động thời gian thực trên AWS EC2 để hỗ trợ người dùng định giá tài sản. 8\_Enhanced\_House\_Price\_…

**3. Dữ liệu**  
Dữ liệu được thu thập từ nhiều nguồn trực tuyến, gồm các thuộc tính: *Availability, Area, Room, Society, Bathroom, Balcony, Location* và nhãn *Price*. Sau khi tiền xử lý, tập dữ liệu được chia 80 % huấn luyện, 20 % kiểm thử. 8\_Enhanced\_House\_Price\_…

**4. Phương pháp**

1. **Tiền xử lý & Tạo đặc trưng**
   * Xử lý giá trị thiếu (thay thế bằng median).
   * Sinh thuộc tính mới *Bhk* (số phòng ngủ) từ cột *size*.
   * Giảm chiều: gộp các vị trí có < 10 mẫu thành nhóm “other”.
   * Khử ngoại lai theo các quy tắc nghiệp vụ (diện tích tối thiểu 300 ft² mỗi phòng ngủ, giá/m² ngoài 1 σ, số phòng tắm ≤ phòng ngủ + 1). 8\_Enhanced\_House\_Price\_…
2. **Khảo sát mô hình**
   * Thử năm thuật toán: Linear Regression, Ridge, Lasso, Decision Tree, Random Forest.
   * Đánh giá bằng RMSE và *score* (R²).
3. **Triển khai**
   * Lưu mô hình tốt nhất (*.pickle*), xây REST API Flask, giao diện HTML/CSS/JS.
   * Host trên AWS EC2 để dự báo thời gian thực.

**5. Kết quả**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Thuật toán** | |  | | --- | |  |  |  | | --- | | **R² (*Score*)** | | |  | | --- | |  |  |  | | --- | | **RMSE** | |
| Linear Regression | 0,8581 | 28,56 |
| Ridge Regression | 0,8466 | 29,69 |
| Random Forest | 0,8079 | 33,23 |
| Decision Tree | 0,7817 | 35,43 |
| Lasso Regression | 0,7378 | 38,83 |

Linear Regression vượt trội, cho sai số gốc trung bình thấp nhất và đường *Actual vs. Predicted* thể hiện khớp tốt. 8\_Enhanced\_House\_Price\_…

**6. Kết luận**  
Quy trình toàn diện—từ thu thập, tiền xử lý đến lựa chọn mô hình—giúp đạt độ chính xác ~86 % cho nhiệm vụ định giá nhà. Thêm các tham số chi tiết (Bathroom, Area…) và bước làm sạch nghiêm ngặt đã cải thiện đáng kể so với nghiên cứu trước. 8\_Enhanced\_House\_Price\_…

**7. Ý nghĩa**

* **Thực tiễn**: Cung cấp công cụ định giá tức thời, giảm rủi ro tài chính cho người mua/bán, hỗ trợ môi giới và nhà phát triển ra quyết định.
* **Kỹ thuật**: Chứng minh tầm quan trọng của feature engineering, khử ngoại lai và lựa chọn mô hình phù hợp thay vì chỉ dựa vào thuật toán phức tạp.
* **Nghiên cứu**: Mở hướng tích hợp dữ liệu kinh tế vĩ mô thời gian thực để mô hình thích ứng với biến động thị trường.