



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY



BKPrice System

Applying MLOps for Enhanced AI Prediction in Real Estate

Giảng viên hướng dẫn

TS. Trần Văn Đặng

Sinh viên thực hiện

Lê Thành Long

Mã số sinh viên

20194099



- Đồ án tốt nghiệp -

Nội dung

- Giới thiệu đề tài
- Hướng tiếp cận
- Phương pháp đề xuất
- Kết quả thực nghiệm
- Đóng góp nổi bật

1. Giới thiệu đề tài

Mức độ quan tâm đất nền cải thiện



Mức độ quan tâm bất động sản
có dấu hiệu phục hồi từ năm 2024

Bất động sản đang được thổi giá

Cần dịch vụ định giá giảm thiểu rủi ro khi mua bất động sản cho người dân

hanoimoi.vn
<https://hanoimoi.vn> > Góc nhìn

Cần trọng chiêu trò "thổi giá" nhà, đất

26/04/2024 — Trong khi Văn phòng đăng ký đất đai Hà Nội xác nhận, giao dịch nhà, đất không tăng bất thường... Chỉ còn ít ngày nữa, Lễ hội Sen Hà Nội năm 2024...

VTV.vn
<https://vtv.vn> > Kinh tế

Giá nhà đất Hà Nội "leo thang"

Tháng 1/2024, một số căn hộ ở đây rao bán khoảng 80 triệu đồng/m² đã là mức giá cao, giờ giao dịch thực hơn 90 triệu đồng/m² là điều không ngoặt...

Laodong.vn
<https://laodong.vn> > bat-dong-san > han-che-dau-co-th...

Hạn chế đầu cơ thổi giá đất nền tràn lan từ thời điểm 1.8...

3 ngày trước — Câu lạc bộ Thanh Hóa vô địch Cúp Quốc gia Casper 2023-2024 khi thắng 9-8 trước Hà Nội FC ở loạt sút luân lưu, sau khi 2 đội hòa nhau 0-0 trong...

1. Giới thiệu đề tài

 **biggee** bigger than the biggest

ĐỊNH GIÁ NHÀ ĐẤT

« »

Địa chỉ gợi nhớ

Số tờ

Số thửa

Khu vực

Diện tích (m²)

Đặc điểm tốt

- Tiếp giáp trên 2 mặt tiền đường hoặc hẻm
- Gần chợ, siêu thị phạm vi 100m
- Nở hậu
- Hướng Nam hoặc Đông Nam

Đặc điểm bất lợi

- BDS tọa lạc tại hẻm cùt
- BDS có đường hướng vào nhà
- Gần mồ mả, nghĩa trang, nhà tang lễ phạm vi 100m
- BDS có hình dáng tốp hậu

Mặt tiền (m)

Chiều dài (m)

Diện tích đất xây dựng (m²)
Nhập 0 nếu là đất trống.

Số tầng
Nhập 0 nếu là đất trống.

- Dịch vụ định giá bất động sản yêu cầu nhiều thông tin **không thân thiện** với người dùng: số tờ, số thửa, ...
- **Chỉ hỗ trợ** định giá khu vực thành phố Hồ Chí Minh



1. Giới thiệu đề tài

Xác định giá trị bất động sản nhanh và chính xác nhất

1,110+ dự án 1,685,000+ BĐS đã xác thực 1,000,000+ lượt định giá

Địa chỉ: Hiện tại, OneHousing chỉ hỗ trợ định giá Nhà phố/Đất nền tại quận Cầu Giấy, Thanh Xuân, Hoàng Mai, Hà Đông, Hai Bà Trưng, Đống Đa, Tây Hồ của Hà Nội và quận Gò Vấp, Thành phố Thủ Đức của Hồ Chí Minh. Các Tỉnh/Thành và Quận/Huyện khác sẽ được cập nhật trong thời gian tới.

Số 75, Ngõ 75, Giải Phóng, P. Đồng Tâm, Q. Hai Bà Trưng, TP. H... [Định giá ngay](#)

Bạn còn 9 lượt định giá trong ngày hôm nay

OneHousing Mua Bán [Hot](#) Ưu đãi Dự án Phân tích & So sánh Định giá Cửa Số BDS

Cần đang bán [Tim kiếm khu vực](#) Theo dõi

Số 75 - Ngõ 75 - Giải Phóng
P. Đồng Tâm - Q. Hai Bà Trưng - TP. Hà Nội
Chưa xác minh căn nhà

Thông tin độc quyền

Tư vấn giá trị tham khảo 1tỷ 692 triệu - 1tỷ 862 triệu
[Xác thực để xem giá chính xác](#)

Gợi ý cho thuê ~ 3.6 triệu/tháng
[Giá trị được Techcombank chấp nhận xem xét vay thế chấp](#)
[Xác thực để xem giá](#)

Bạn có phải là chủ sở hữu cần hộ này?
Tôi là chủ nhà

Bạn đang có nhu cầu gì về bất động sản này?
Hiểu đúng nhu cầu sẽ giúp OneHousing hiển thị những thông tin phù hợp nhất với bạn

Mua Bán Thuê Cho

- Dịch vụ định giá **chỉ dựa vào địa chỉ** và **không đáp ứng được** nhiều nhu cầu của user
- Kết quả định giá ở **mức chủ quan** và được xác thực **định giá thủ công** bằng tư vấn viên
- Chỉ hỗ trợ chủ yếu là **thành phố Hà Nội**



1. Giới thiệu đề tài

Tư vấn viên định giá bất động sản được
cung cấp bởi bên thứ 3

Dễ bị quá tải
Định giá mang tính chủ quan

Công cụ định giá
của Biggee

Yêu cầu nhiều **thông tin đặc thù**
Khó sử dụng
Chỉ hỗ trợ khu vực TP. Hồ Chí Minh

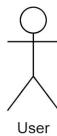
Công cụ định giá
của OneHousing

Không hỗ trợ nhiều thông tin để định giá.
Chỉ dựa vào **địa chỉ**
Chủ yếu hỗ trợ khu vực TP. Hà Nội
Xác thực định giá **thủ công**

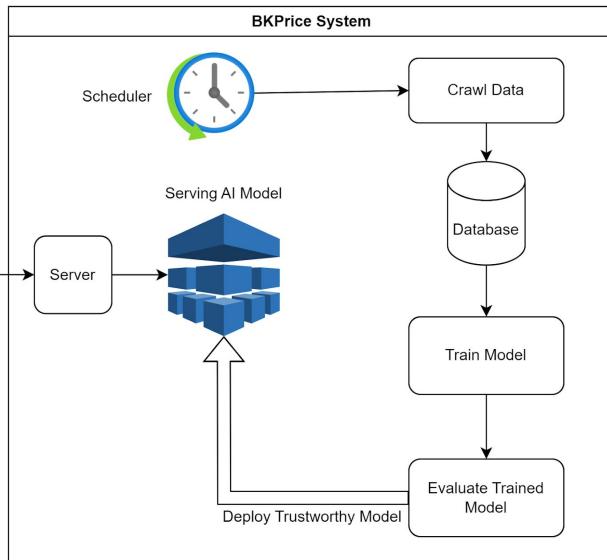
Cần một hệ thống giải quyết nhược điểm của các giải pháp hiện tại



2. Hướng tiếp cận



I want to evaluate the house price on Giai phong street



Hệ thống BKPrice

Chức năng chính:

1. Lập lịch tự động thu thập dữ liệu TP.HN và HCM
2. Giải thuật định giá tin cậy sử dụng đa mức thông tin
3. Giám sát quá trình huấn luyện và đánh giá mô hình tự động

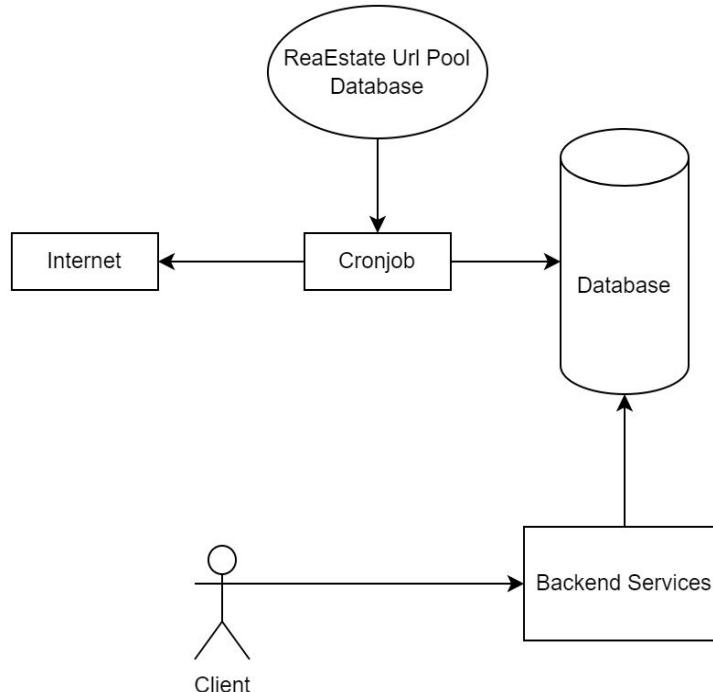
Điểm nổi bật:

1. Định giá cả 2 thành phố HN&HCM
2. Sử dụng thông tin phổ thông làm đầu vào quá trình định giá
3. Quá trình MLOps tự động end-to-end



3. Phương pháp đề xuất

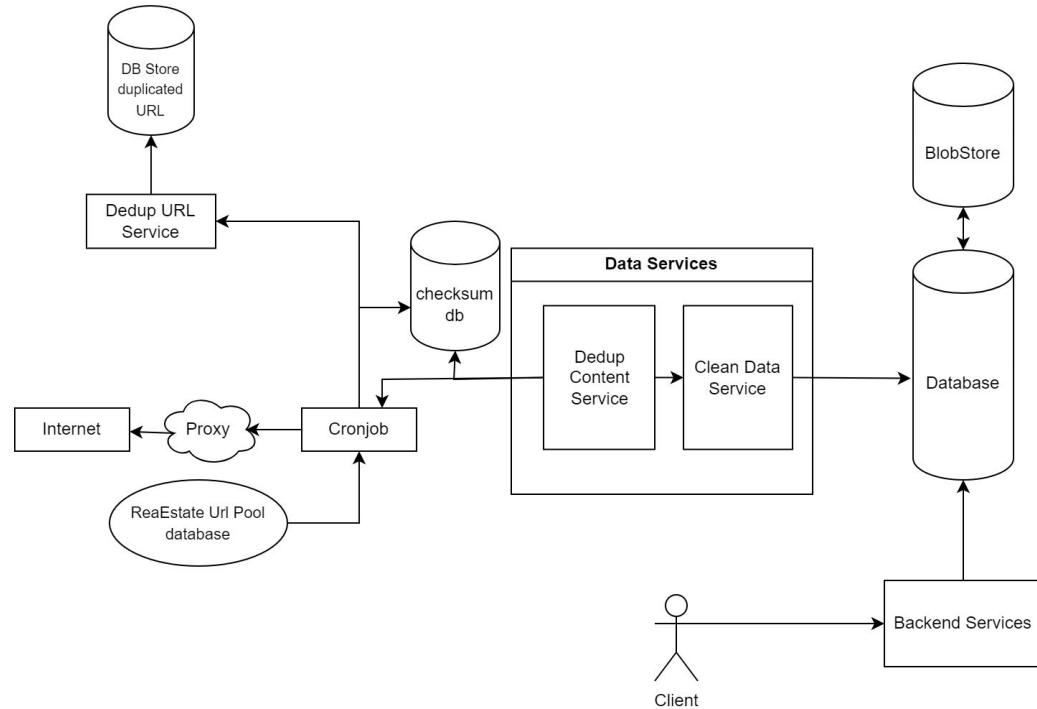
High Level Design



3. Phương pháp đề xuất

> Hệ thống dữ liệu

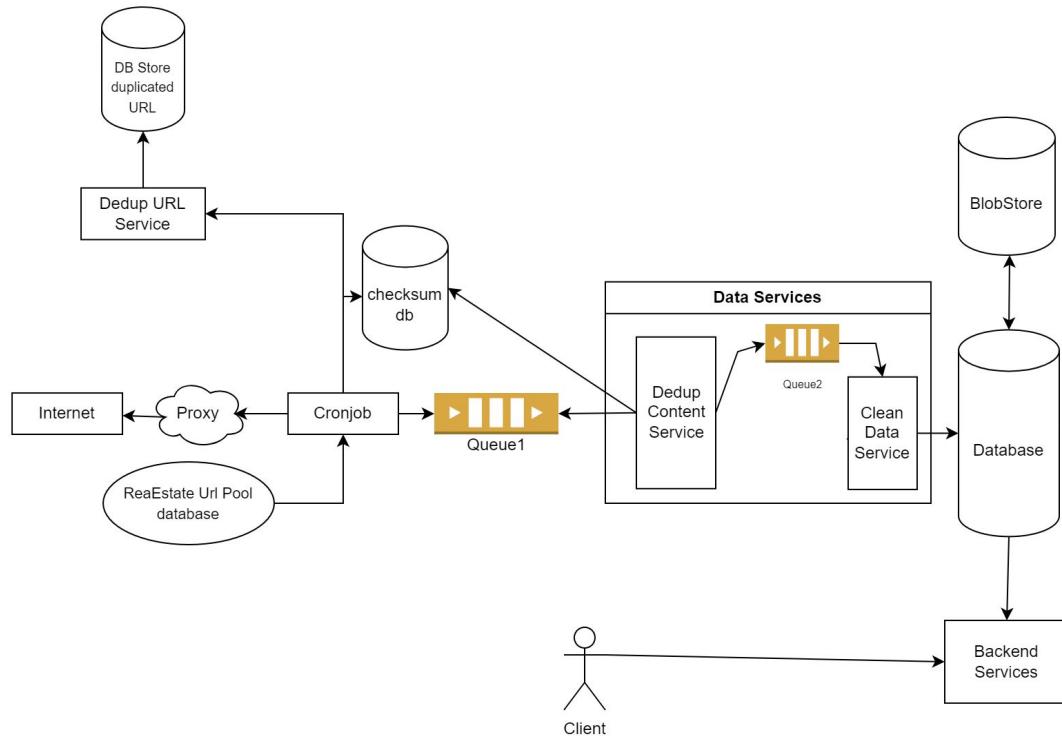
Sơ đồ khối của Hệ thống dữ liệu



3. Phương pháp đề xuất

> Hệ thống dữ liệu

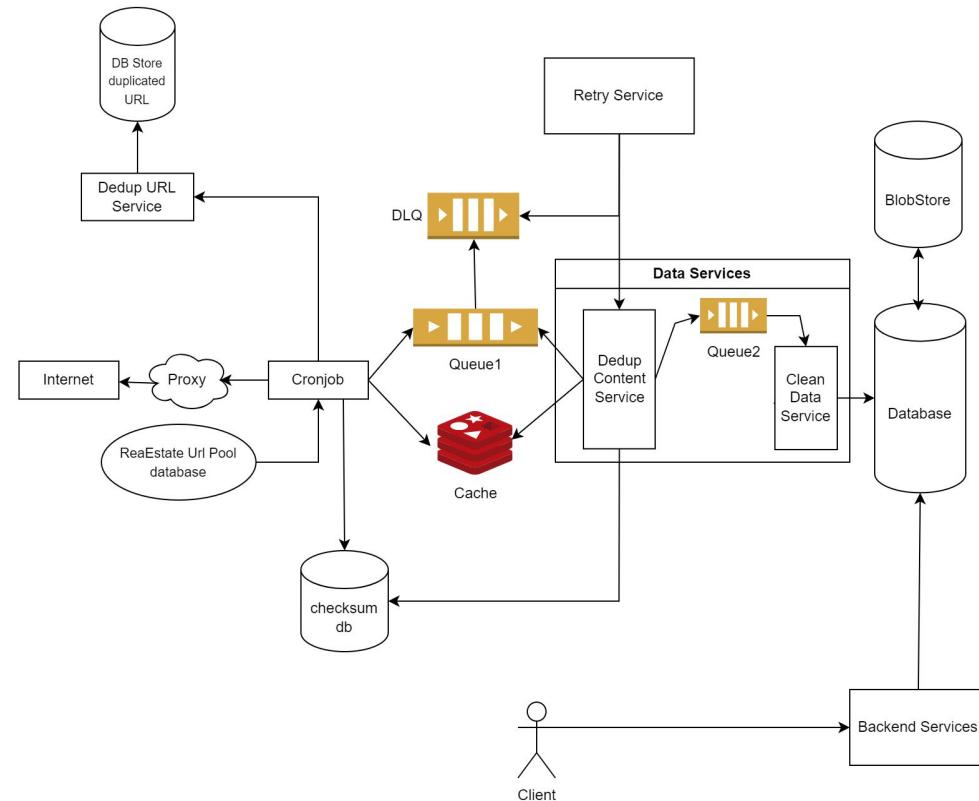
Sơ đồ khái của Hệ
thống dữ liệu



3. Phương pháp đề xuất

> Hệ thống dữ liệu

Sơ đồ khái của Hệ
thống dữ liệu



3. Phương pháp đề xuất

> Hệ thống dữ liệu

 ADD DATA  EXPORT DATA  UPDATE  DELETE

1 - 20 of 40429

```
▶ propertyBasicInfo : Object
  ▶ landType : Object
    ▶ comment : Array (empty)
    status : "UNSELECTED"
    value : "residentialLand"
  ▶ accessibility : Object
  ▶ distanceToNearestRoad : Object
  ▶ frontRoadWidth : Object
  ▶ address : Object
    ▶ comment : Array (empty)
    status : "UNSELECTED"
    ▶ value : Object
      addressDetails : ""
      street : "Lĩnh Nam"
      ward : "Lĩnh Nam"
      district : "Hoàng Mai"
      city : "Hà Nội"
      country : "Việt Nam"
  ▶ description : Object
  ▶ geolocation : Object
  ▶ typeOfRealEstate : Object
  ▶ frontWidth : Object
  ▶ endWidth : Object
  ▶ facade : Object
  ▶ houseDirection : Object
  ▶ landSize : Object
  ▶ contact : Object
  ▶ price : Object
```

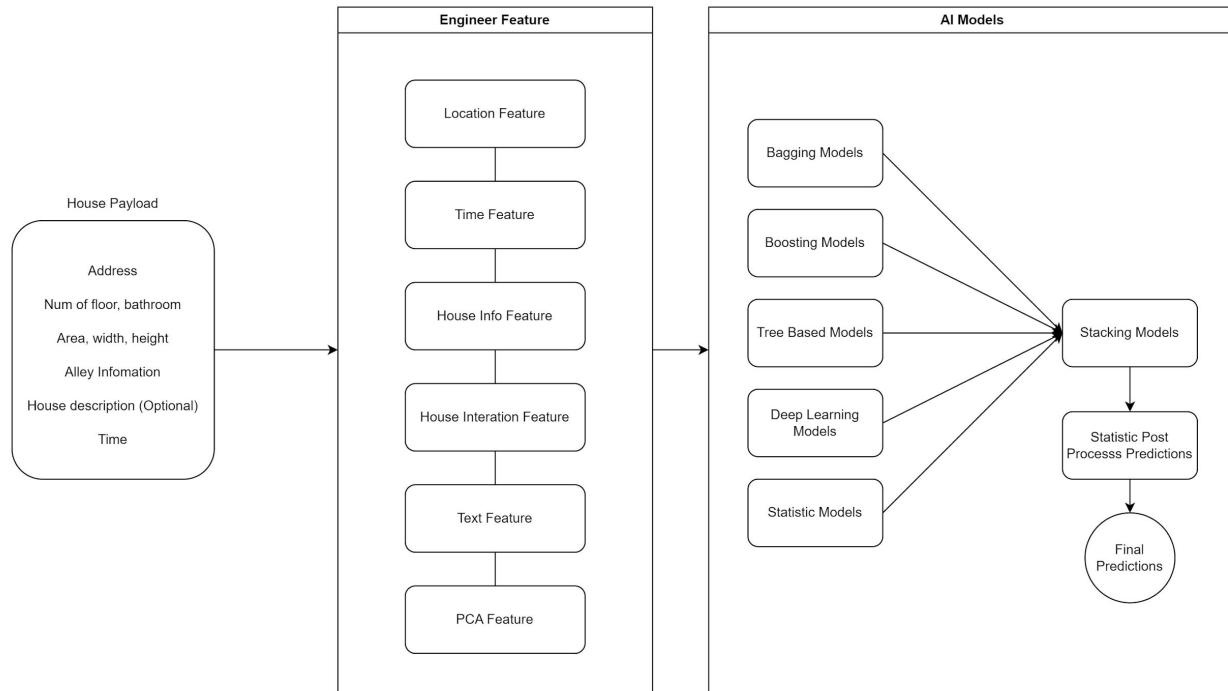
Định dạng
chuẩn của dữ
liệu



3. Phương pháp đề xuất

> Giải thuật định giá

Tổng quan giải
thuật



3. Phương pháp đề xuất

> Giải thuật định giá > Facility Feature

Name	Description
University - School	Trường học
Fuel	Cây xăng
Cafe	Quán cafe
Parking	Bãi đỗ xe
Fast food	Tiệm đồ ăn nhanh
Marketplace	Chợ
Restaurant	Nhà hàng
Hospital	Bệnh viện
Kindergarten	Nhà mẫu giáo
Community centre	Trung tâm công cộng
Police	Khu vực cảnh sát
Place of worship	Nhà thờ
Bank - ATM	Ngân hàng

Bảng 4.1: Danh sách tiện tích công

- **Tạo đặc trưng tiện ích công dựa vào vị trí lat, lon của bất động sản**
- **Đếm số lượng tiện ích công trong bán kính 500m, 1000m và 200m**

3. Phương pháp đề xuất

> Giải thuật định giá > Distance Feature

lat	lon	name	address
10.773358	106.701287	Vincom Đồng Khởi	72 Lê Thánh Tôn, Bến Nghé, Quận 1, Hồ Chí Minh
21.030585	105.802927	Công viên Thủ Lệ	Công viên Thủ Lệ, Đường Bưởi, Ngọc Khánh, Ba Đình, Hà Nội
10.795529	106.716628	Công viên Vinhomes Central Park	208 Nguyễn Hữu Cánh, Phường 22, Quận Bình Thạnh, Hồ Chí Minh
10.809043	106.671617	Công viên Gia Định	Đường Hoàng Minh Giám, Phường 9, Phú Nhuận, Hồ Chí Minh
21.051217	105.835552	Hồ Tây 5	Hồ Tây, Tây Hồ, Hà Nội
21.011068	105.846908	Vincom Bà Triệu	191 Phố Bà Triệu, phường Lê Đại Hành, quận Hai Bà Trưng, Hà Nội
21.010941	105.839811	Công viên Thống Nhất	Công viên Thống Nhất, Đường Lê Duẩn, Lê Đại Hành, Đống Đa, Hà Nội
10.727873	106.717026	SC VivoCity Shopping Center	1058 Nguyễn Văn Linh, KP. 1, P. Tân Phong, Quận 7, TP. HCM
10.772185	106.698227	Công viên trên mây tại Taka Shimaya	92-94 Nam Kỳ Khởi Nghĩa, Phường Bến Nghé, Quận 1, Hồ Chí Minh
21.015025	105.775093	Tops Market The Garden	Tops Market The Garden, Mễ Trì, Từ Liêm, Hà Nội
20.964183	105.852110	Công viên yên sở	Công viên Yên Sở, Quốc lộ 1A, Gamuda Central, Hoàng Mai, Hà Nội
10.858315	106.582624	Công viên cá Koi Rin Rin Park	87/8P Xuân Thới Thượng 6, Ấp Xuân Thới Đông, Huyện Hóc Môn, Hồ Chí Minh
21.003251	105.801975	Trung tâm thương mại Hà Nội Center Point	27 Đường Lê Văn Lương, Nhân Chính, Thanh Xuân, Hà Nội
10.788040	106.700999	Thảo Cầm Viên	2 Nguyễn Bình Khiêm, Phường Bến Nghé, Quận 1, Hồ Chí Minh
21.029358	105.852400	Hồ Gươm	Hồ Gươm, Phố Lê Thái Tổ, Hàng Trống, Hoàn Kiếm, Hà Nội
21.016627	105.782185	Trung tâm thương mại Hà Nội Keangnam	72 Phạm Hùng, Keangnam, Mễ Trì, Từ Liêm, Hà Nội
21.002143	105.812609	Trung tâm thương mại Royal City Hà Nội	72A Đường Nguyễn Trãi, Khu đô thị Royal City, phường Thượng Đình, quận Đống Đa, Hà Nội
21.027274	105.896800	Aeon Mall Long Biên	Aeon Mall Long Biên, Đường Cổ Linh, p. Long Biên, Long Biên, Hà Nội

Tạo đặc trưng khoảng cách dựa vào **khoảng cách bất động sản** tới các địa điểm nổi tiếng.

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

- Tạo đặc trưng tiện ích công dựa vào **vị trí lat, lon** của **tập bất động sản**
- Tạo đặc trưng từ các **bất động sản lân cận** -> Qua đó thể hiện được mối tương quan của các bất động sản với nhau.
- Lấy top K bất động sản gần nhất: **thông tin giá trung bình** của các bất động sản này, **thông tin đường, phường**, và **thông tin khoảng cách** từ bất động sản đến top K bất động sản
- Trong nghiên cứu, sử dụng **K = 10**

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

Bài toán:

- Gọi M là **số lượng bất động sản ($M \geq 300000$)**
- N là số nguyên dương ($N \geq 1000$) thể hiện **số phép tính toán tìm kiếm gần nhất**
- **Thực hiện N lần tìm kiếm bất động sản lân cận tối ưu ?**

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

Ta có thể tìm kiếm trong cơ sở dữ liệu như sau:

- $Q_i : (lat_i, lon_i)$ là câu query thứ i , $1 \leq i \leq N$
- $A_j : (lat_j, lon_j)$ là bất động sản thứ j , $1 \leq j \leq M$
- D_{ij} là khoảng cách giữa 2 điểm không gian (lat_i, lon_i) và (lat_j, lon_j) với $1 \leq i \leq N, 1 \leq j \leq M$
- Với mỗi câu query $1 \leq i \leq N$, gọi A_k là bất động sản có khoảng cách $d = \min_u D_{iu}$, $1 \leq u \leq M$

3. Phương pháp đề xuất

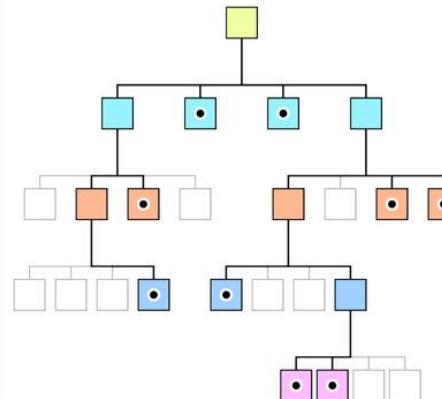
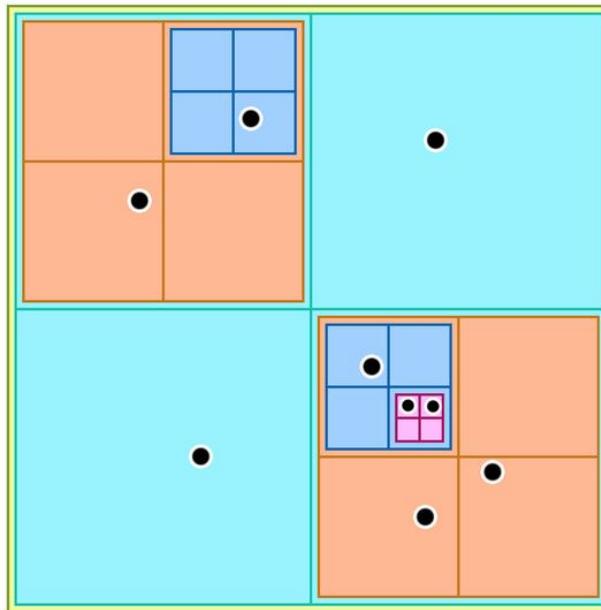
> Giải thuật định giá > Nearest Feature

- + Đối với mỗi query thì hệ thống phải **tìm kiếm toàn bộ M** bất động sản (độ phức tạp là $O(M)$)
 - + Độ phức tạp của giải thuật là **$O(N \times M)$, $N \geq 1000$, $N \geq 100000$**
-  Giải thuật này **không hiệu quả trong thực tế** vì độ phức tạp lớn khi số lượng bất động sản lớn. Cần một giải pháp khác ?

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

- + Đề xuất xử sụng Quadtree để tối ưu quá trình tìm kiếm gần nhất



Cấu trúc cây Quadtree

- phân hoạch vùng không gian hai chiều hiện tại thành **các vùng nhỏ** dựa vào lat, lon
- Mỗi node đại diện cho **1 vùng không gian** chứa một hay nhiều bất động sản

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

- + Mô phỏng M bất động sản thành không gian 2 chiều như trên. Mỗi ô nhỏ sẽ chưa một vài bất động sản.
- + Mỗi câu query (lat_i, lon_i), ta thực hiện quá trình tìm kiếm như sau:
 - Tìm kiếm vị trí của ô hợp lệ cho điểm (lat_i, lon_i) . Do mỗi node trong quadtree có 4 node con tương ứng với 4 vùng không gian, vì vậy cứ mỗi lần lặp tìm kiếm số lượng bất động sản sẽ giảm dần theo cấp số nhân.
 - Tìm kiếm bất động sản trong ô hợp lệ có khoảng cách nhỏ nhất tới (lat_i, lon_i)

3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

Gọi p là số lượng vòng lặp tối đa để tìm kiếm bất động sản. Gọi M_i là số lượng bất động sản trong tập tìm kiếm sau vòng lặp thứ i . Ta có:

$$M_{i+1} = \frac{M_i}{4}, 0 \leq i \leq p - 1, M_0 = M$$

Sau vòng lặp thứ k thì số lượng bất động sản trong mảng tìm kiếm là $M_k = \frac{M}{4^k}$

Điều kiện dừng của thuật toán tìm kiếm khi mỗi node chứa một lượng node $u \geq 1$ vừa đủ được cấu hình từ trước. Do đó ta có điều kiện sau:

$$M_p \geq u \geq 1 \iff \frac{M}{4^p} \geq 1 \iff p \leq \log_4 M$$

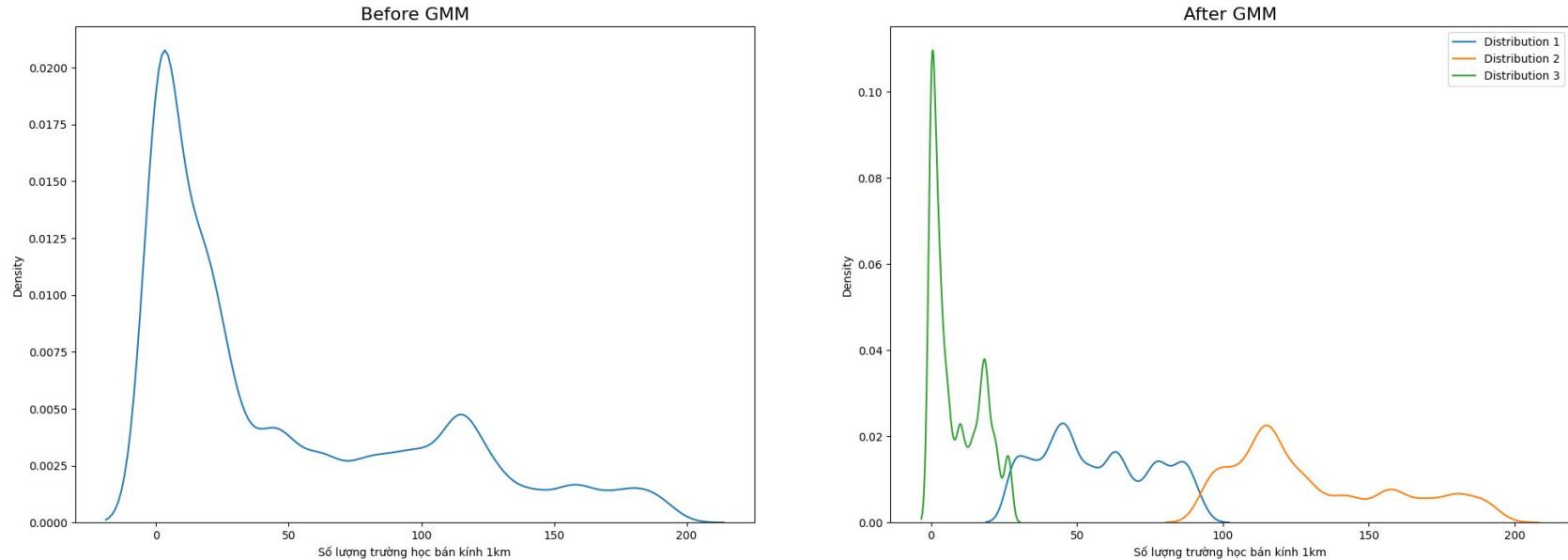
3. Phương pháp đề xuất

> Giải thuật định giá > Nearest Feature

- + Do đó số lượng vòng lặp tối đa là $\text{Log}_4(M)$, hay nói cách khác độ phức tạp của mỗi lần tìm kiếm sẽ là $O(\text{Log}_4 M)$
 - + Độ phức tạp cho N câu lệnh query là $O(N \times \text{Log}_4 M)$, $N \geq 1000$, $N \geq 100000$
-  Giải thuật này **hiểu quả trong thực tế** vì độ phức tạp trong miền tính toán được

3. Phương pháp đề xuất

> Giải thuật định giá > Gaussian Mixture Model Feature



Mô hình hóa mật độ của thuộc tính num_of_marketplace_in_1000m_radius trước và sau khi khớp Gaussian Mixture Model

3. Phương pháp đề xuất

> Giải thuật định giá > Gaussian Mixture Model Feature

- Việc kiểm định thuộc tính của những **mẫu thử mới** thuộc **vùng phân bố nào** của tập training dataset đóng vai trò quan trọng trong việc định giá chính xác hơn.
- **Mô hình hiểu rõ dữ liệu hơn** khi đối với **mỗi thuộc tính**, các mẫu dữ liệu đã được phân loại trên phân đúng phân vùng dữ liệu

3. Phương pháp đề xuất

> Giải thuật định giá > PCA Feature

- Tạo PCA features làm **nổi bật** mối tương quan giữa các thuộc tính và **giảm nhiễu**
- **Bài toán:** Cần bao nhiêu chiều (thành phần) để biểu diễn thông tin của dữ liệu ?
- **Lời giải:**
 - Explained Variance (EV): tỷ lệ phương sai được giải thích, là một thước đo trong PCA thể hiện tỷ lệ **phần trăm biến đổi dữ liệu được giữ lại** bởi một số lượng thành phần chính nhất định.
 - EV cao cho thấy các thành phần đó capture được nhiều thông tin quan trọng trong dữ liệu gốc, EV thấp cho thấy các thành phần đó thể hiện ít thông tin quan trọng
 - **Giữ** những thành phần có **EV cao**
 - **Loại bỏ** những thành phần có **EV thấp**

3. Phương pháp đề xuất

> Giải thuật định giá > PCA Feature

Ta có tập dữ liệu S . Giả sử có thể mô hình hóa thông tin của tập dữ liệu S bởi k chiều. Gọi $\lambda_1, \lambda_2, \dots, \lambda_k$ là k trị riêng của k chiều. Khi đó giá trị *Explained Variance* (EV) của thành phần thứ t có công thức như sau:

$$EV_t = \frac{\lambda_t}{\sum_{i=1}^k \lambda_i}$$

Không mất tính tổng quát giả sử: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{k-1} \geq \lambda_k$.

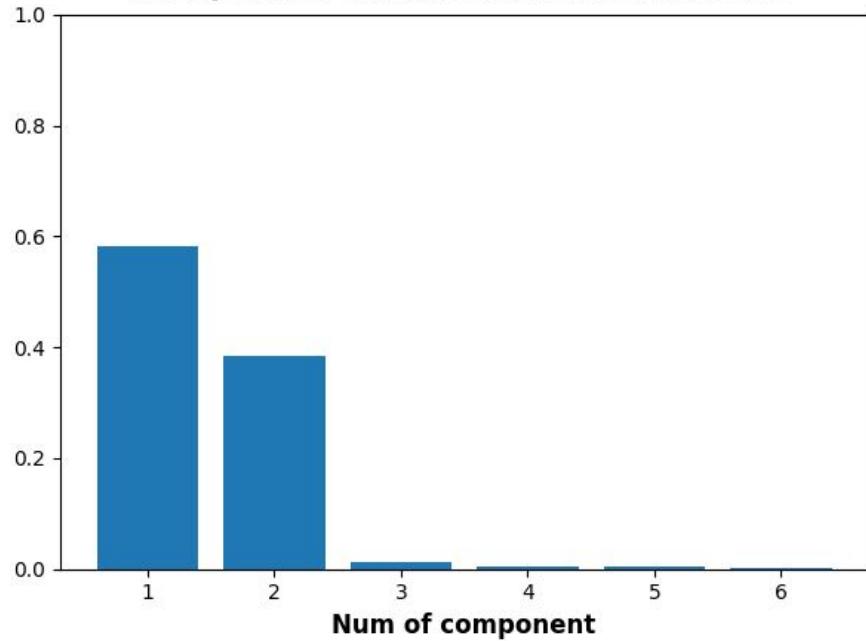
Gọi O phần trăm là ngưỡng thông tin tối thiểu mà phương PCA bảo toàn được. Do đó số lượng thành phần tối ưu d^* của PCA định nghĩa là số nhỏ nhất thỏa mãn:

$$\sum_{i=1}^{d^*} EV_i \geq O$$

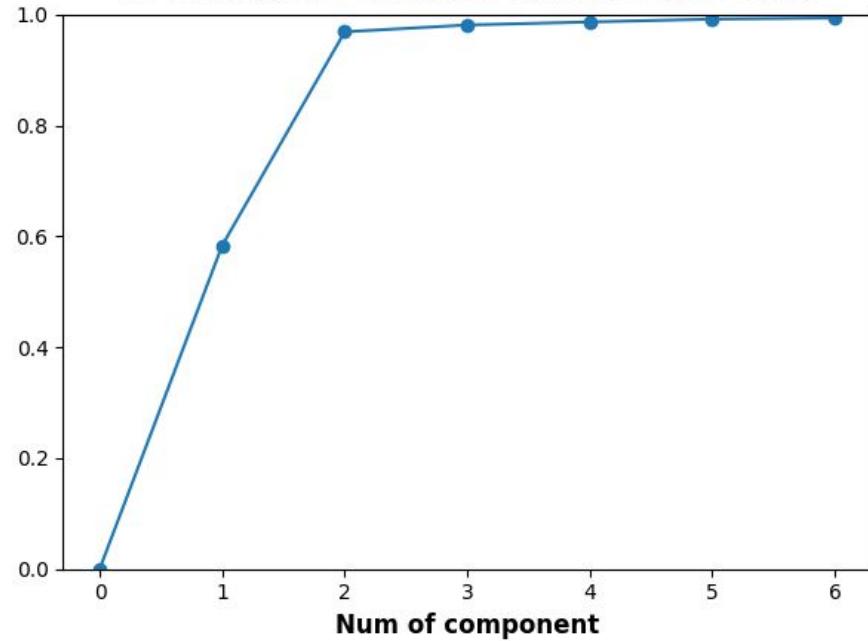
3. Phương pháp đề xuất

> Giải thuật định giá > PCA Feature

% Explained Variance in Realestate Data



% Cumulative Variance in Realestate Data



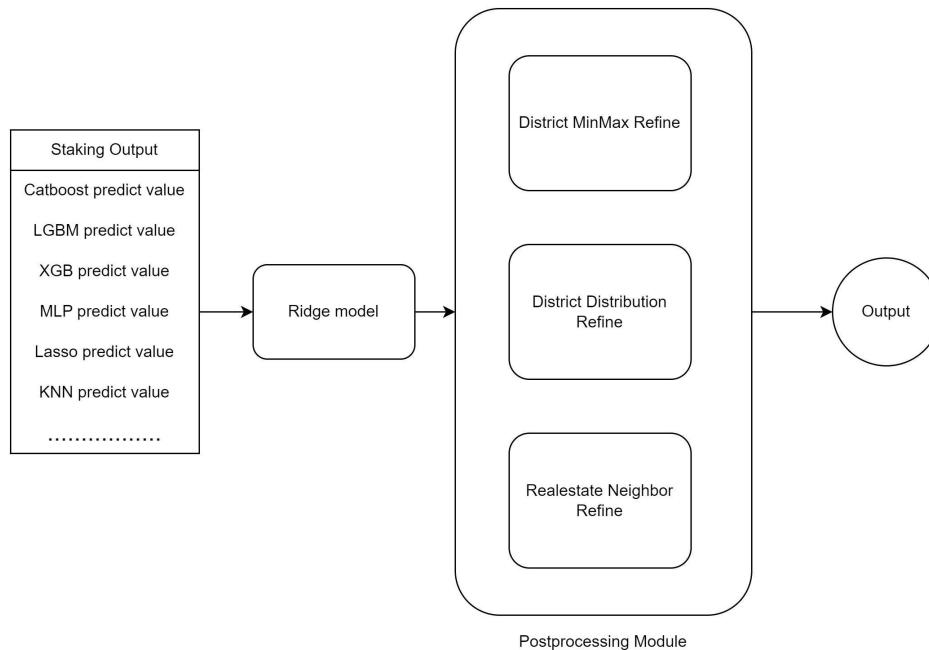
3. Phương pháp đề xuất

> Giải thuật định giá > Other Feature

- Khoảng cách chim bay của bất động sản đến trung tâm của quận
- Tỷ lệ diện tích của bất động sản với tổng diện tích của quận
- Thông tin dân số, mật độ dân số, số lượng phường của quận

3. Phương pháp đề xuất

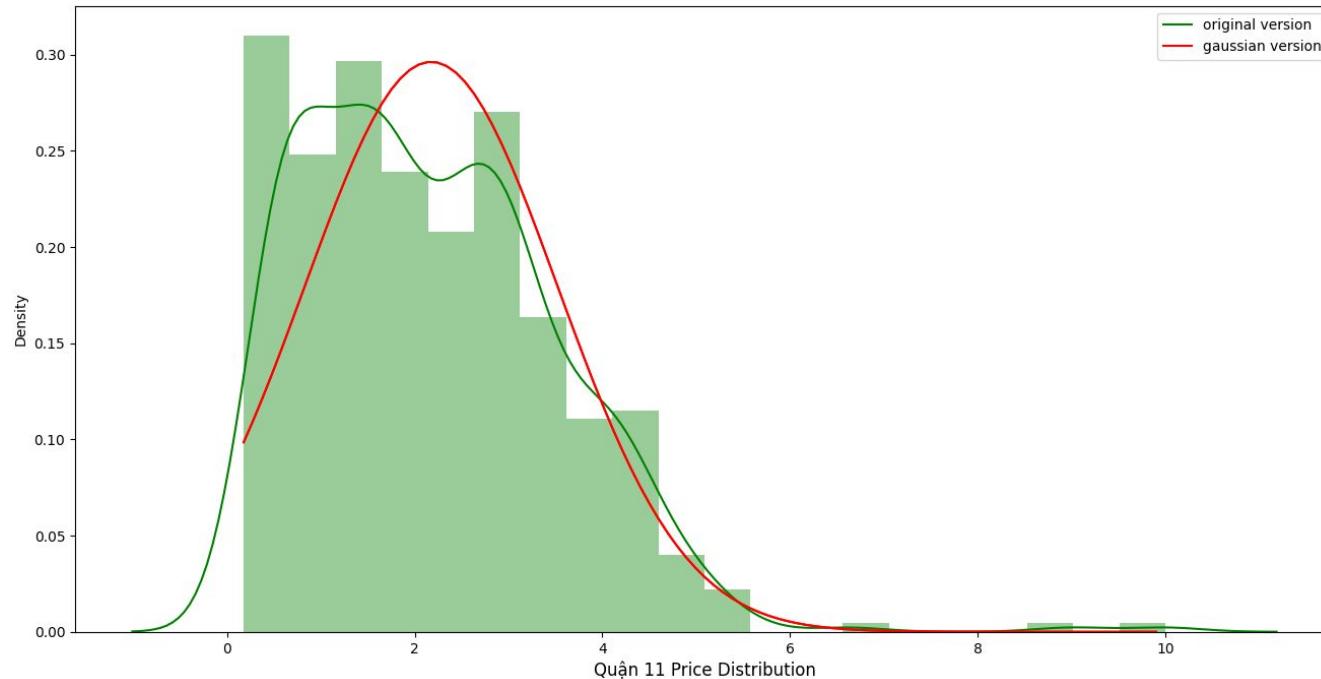
> Giải thuật định giá > Post processing



- Hậu xử lý kết quả dự đoán bằng Min Max District Price
- Hậu xử lý bằng Min Max Top K Nearest Neighbour Realestate
- Hậu xử lý bằng cách điều chỉnh giá dự đoán khớp với phân bố giá của quận trong dữ liệu training

3. Phương pháp đề xuất

> Giải thuật định giá > Post processing

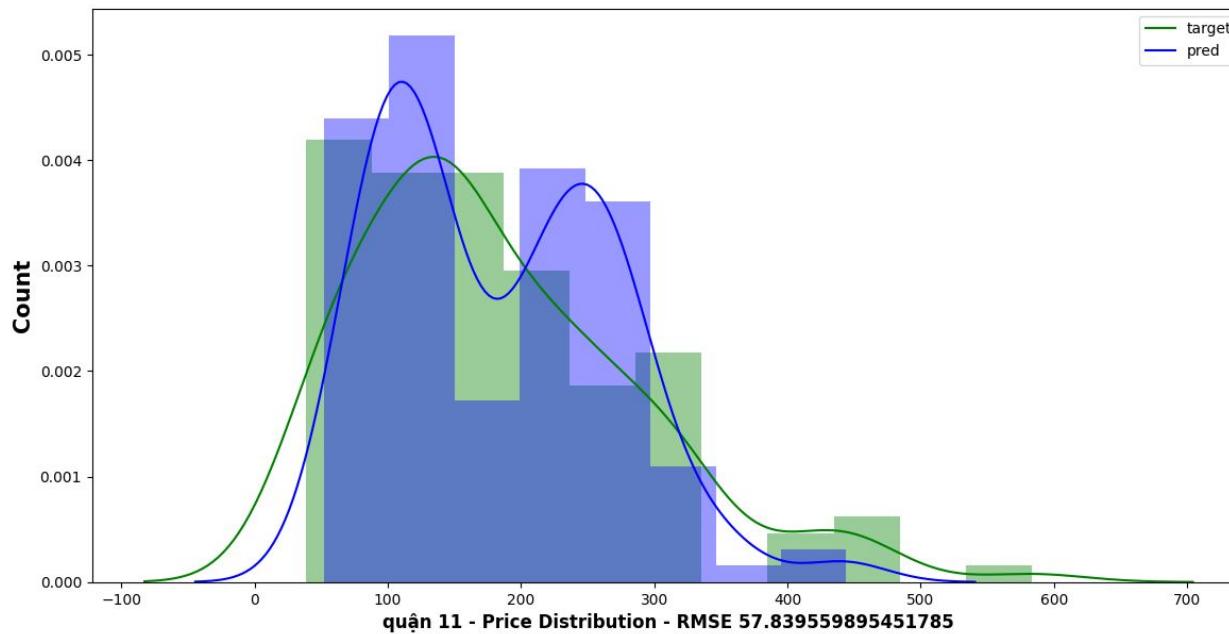


Giá khu vực quận 11 tuân theo phân bố gaussian

3. Phương pháp đề xuất

> Giải thuật định giá > Post processing

Distribution Testing

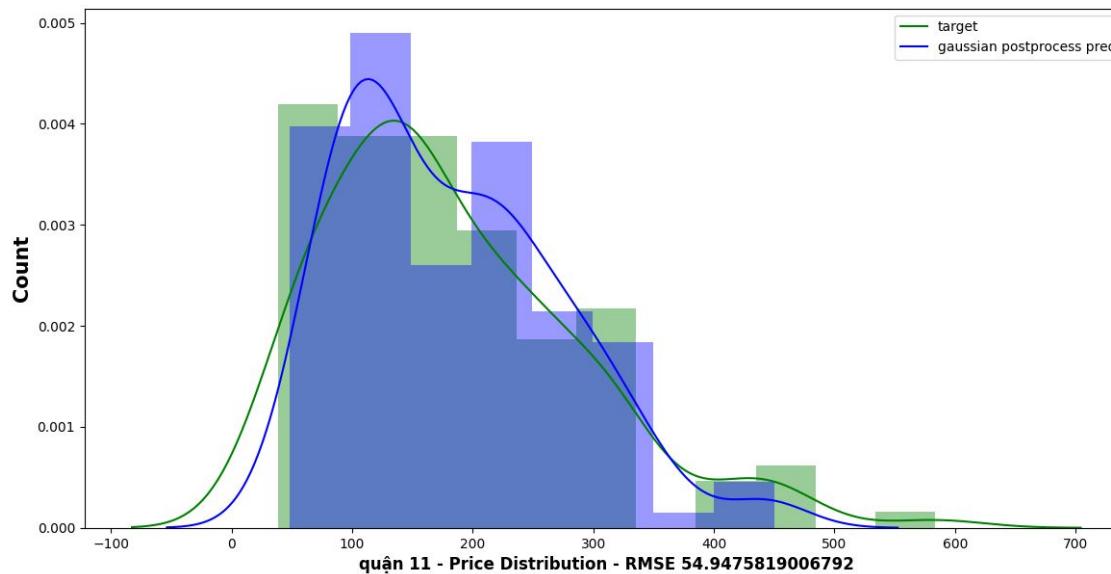


Phân bố giá dự đoán **chưa khớp với phân bố giá** thực tế

3. Phương pháp đề xuất

> Giải thuật định giá > Post processing

Distribution Testing



Phân bố giá dự đoán đã **gần khớp** với phân bố giá thực tế

3. Phương pháp đề xuất

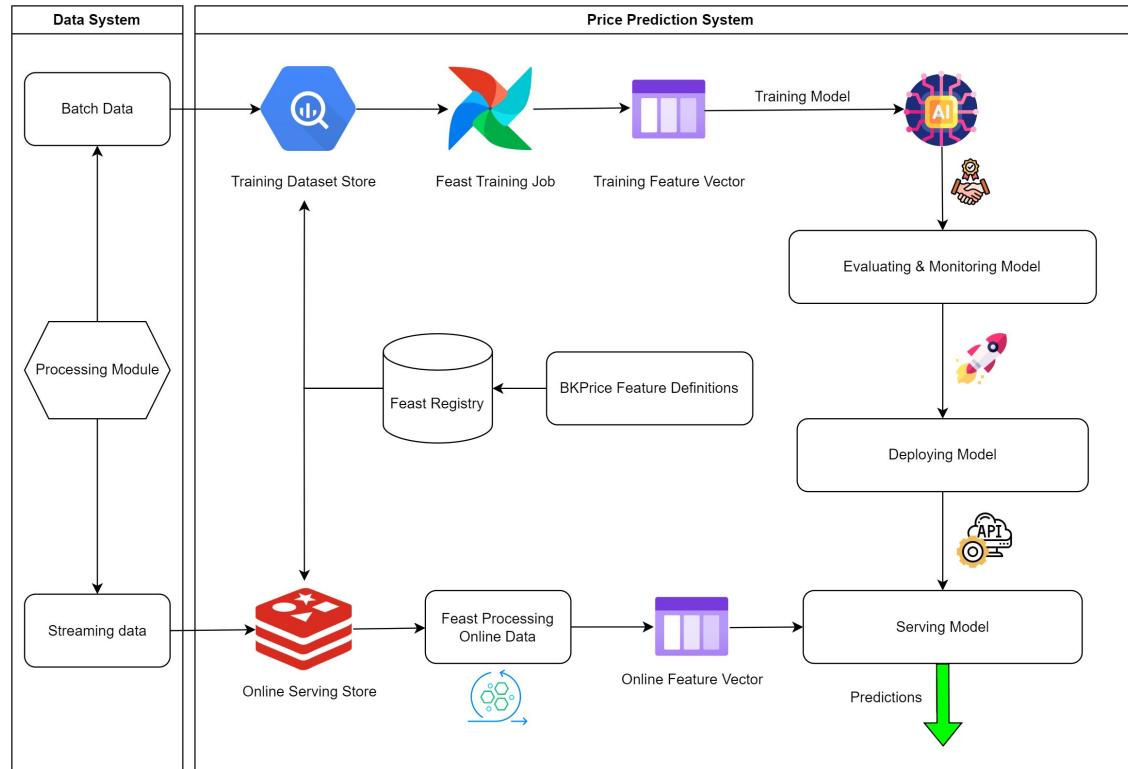
> Giải thuật định giá > Base Models

Model Name	Model Type	Training Device
CatboostRegressor	Boosting Model	GPU
MLPRegressor	Multilayer perceptron Model	GPU
LGBMRegressor	Boosting Model	GPU
XGBRegressor	Boosting Model	GPU
AdaBoostRegressor	Boosting Model	CPU
GradientBoostingRegressor	Boosting Model	CPU
ExtraTreesRegressor	Tree-based Model	CPU
RandomForestRegressor	Tree-based Model	CPU
Lasso	Regression Model	CPU
Ridge	Regression Model	CPU
Linear	Regression Model	CPU
KNeighborsRegressor	Neighbor-based Model	CPU

Bảng 4.2: Danh sách mô hình cơ sở trong BKPrice Prediction System

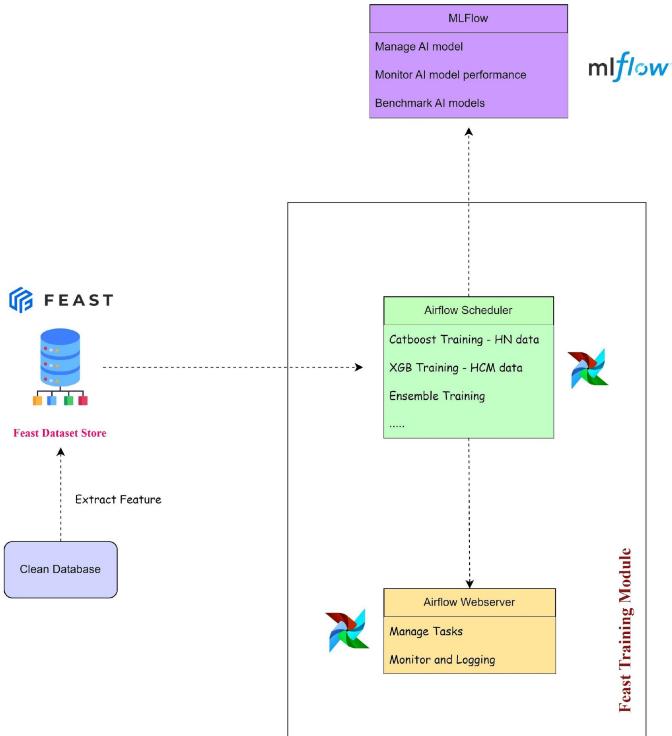
3. Phương pháp đề xuất

› Hệ thống huấn luyện và đánh giá mô hình tự động



3. Phương pháp đề xuất

› Hệ thống huấn luyện và đánh giá mô hình tự động



Tiếp nhận tác vụ và huấn luyện mô hình và
đánh giá mô hình một cách tự động

3. Phương pháp đề xuất

> Hệ thống huấn luyện và đánh giá mô hình tự động

The screenshot shows two main sections of the FEAST interface:

Feature Views (Top):

Name	# of Features
df_hcm_feature_view_v3	188
df_hcm_feature_view_v0	16
df_hn_feature_view_v5	187
target_hn_feature_view_v2	1
df_hn_feature_view_v2	185

Data Sources (Bottom):

Name	Type
/home/long/long/datn-feast/data/target_df_hn_v1.parquet	BATCH_FILE
/mnt/long/long/datn-feast/data/update_data/demo1/target_df_hcm_v2.parquet	BATCH_FILE
/home/long/long/datn-feast/data/data_df_hn_v5.parquet	BATCH_FILE
/mnt/long/long/datn-feast/data/update_data/demo1/data_df_hn_v1.parquet	BATCH_FILE
/home/long/long/datn-feast/data/data_df_hn_v4.parquet	BATCH_FILE

Below the Data Sources table is a table showing feature definitions:

Feature Name	Type	Description	Type
nearest_2_street	INT32	lon	FLOAT
houseDirection	INT32	frontWidth	FLOAT
nearest_1_ward	INT32	lat	FLOAT
nearest_6_district	INT32	w	FLOAT
nearest_3_district	INT32	numberOfBathRooms	FLOAT
nearest_4_ward	INT32	landSize	FLOAT
nearest_8_district	INT32	numberOfBedRooms	FLOAT
gmm_2_component_num_of_cafe_in_1000m_radius	INT32	num_of_marketplace_in_1000m_radius	FLOAT

Quản lý đặc trưng của tập dữ liệu huấn luyện

3. Phương pháp đề xuất

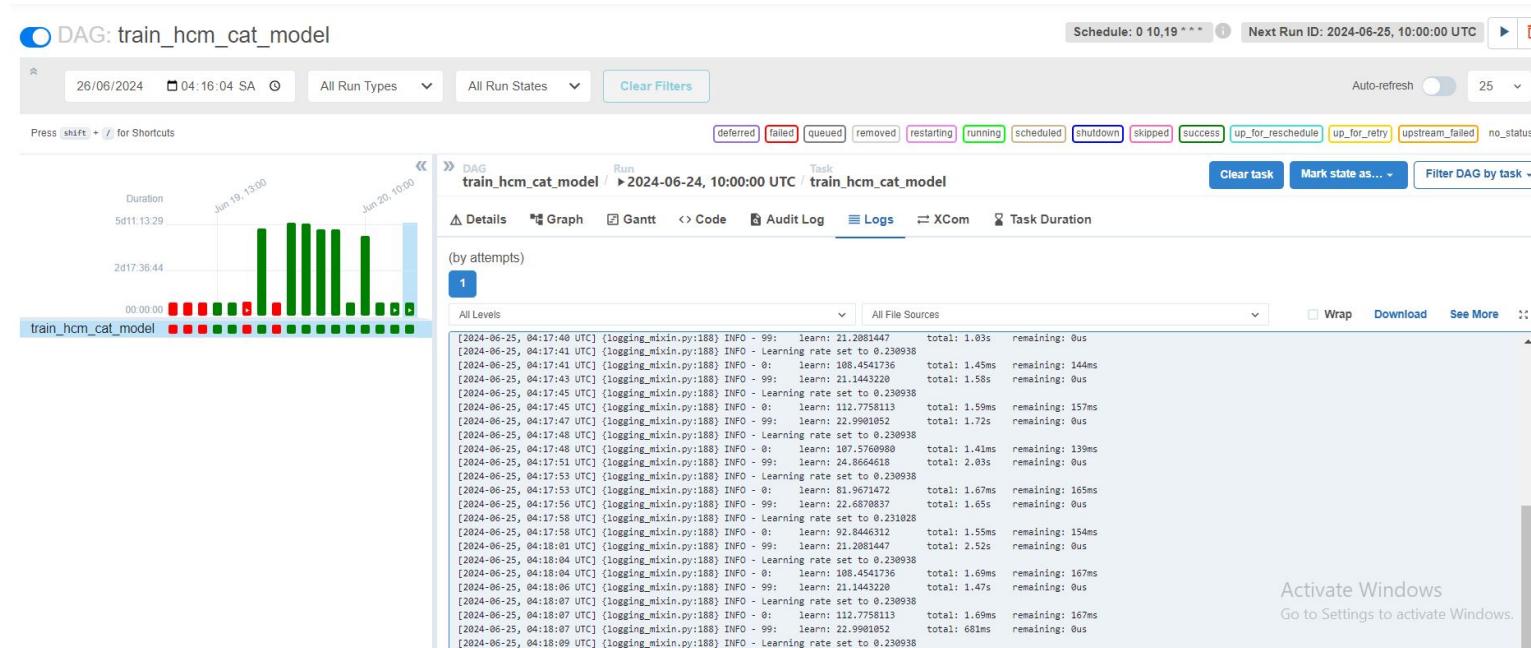
› Hệ thống huấn luyện và đánh giá mô hình tự động

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
ensemble_model	airflow	○○○○○	xjg ****	2024-06-26, 03:02:00			▶ ⚡	...
postgres_operator_dag	airflow	○○○○○	@once	2023-11-24, 00:00:00			▶ ⚡	...
train	airflow	○○○○○	*5 ****	2024-06-26, 03:10:00			▶ ⚡	...
train_hcm_abr_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:10:19	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_cat_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:16:35	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_eir_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:05:08	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_gbr_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:09:49	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_knr_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:09:48	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_la_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:07:36	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_lgbm_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:07:26	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_linear_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:07:25	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_mlp_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:07:24	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_rf_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:07:24	2024-06-25, 10:00:00		▶ ⚡	...
train_hcm_xgb_model	airflow	○○○○○	0 10,19 ***	2024-06-25, 04:10:25	2024-06-25, 10:00:00		▶ ⚡	...

Quản lý tác vụ tự động trong BKPrice System

3. Phương pháp đề xuất

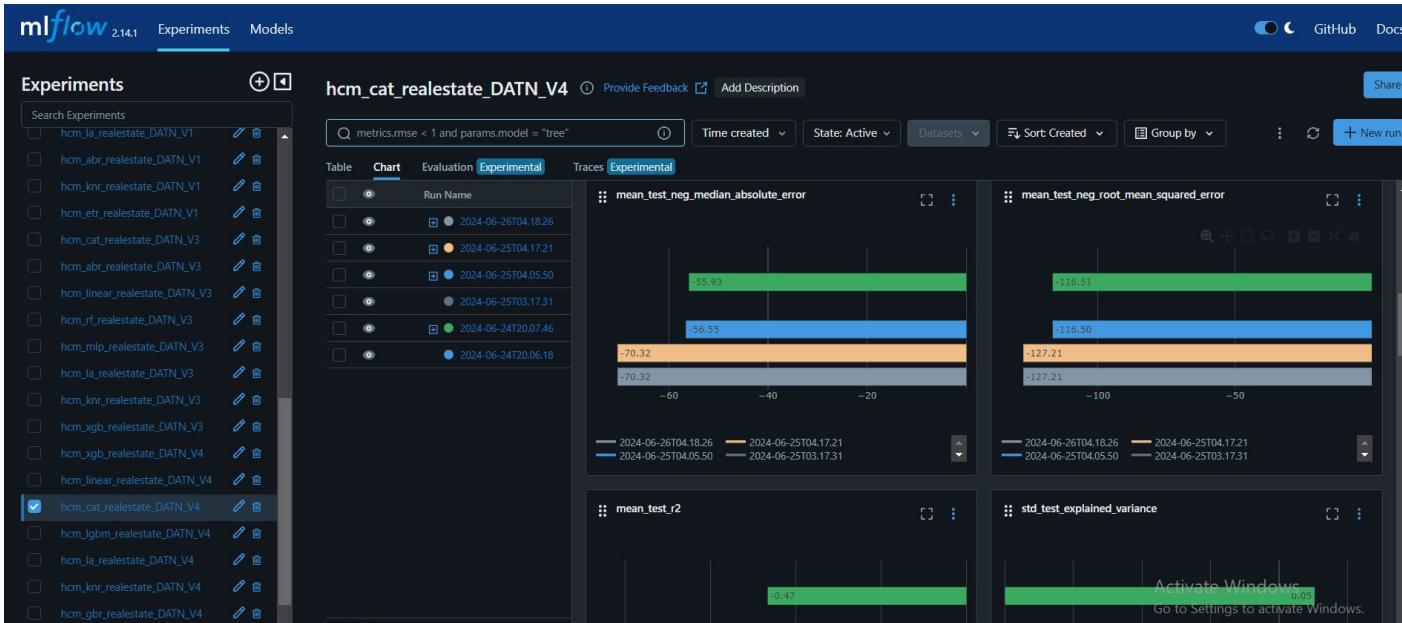
> Hệ thống huấn luyện và đánh giá mô hình tự động



Quản lý từng đơn vị tác vụ và giám sát quá trình chạy của tác vụ

3. Phương pháp đề xuất

➤ Hệ thống huấn luyện và đánh giá mô hình tự động



Màn hình quản lý thời các thời điểm huấn luyện mô hình tự động

3. Phương pháp đề xuất

› Hệ thống huấn luyện và đánh giá mô hình tự động

The screenshot shows the Airflow UI interface for a model named 'hcm_xgb_realestate_DATN_V4' from June 25, 2024, at 04:11:10. The interface is divided into two main sections: 'Overview' and 'Model metrics'.

Overview: This section displays the following parameters and their values:

Parameter	Value
cv	5
error_score	nan
estimator	BKPriceEstimator(pretrained_model_path='/home/long/airflow/dags/models/hcm/xgb/v0/model.joblib', update_model=XGBRegressor(base_score=None, booster=None, callbacks=None, colsample_bytree=None, colsample_bynode=None, colsample_bylevel=None, early_stopping_rounds=None, enable_categorical=False, eval_metric=None, feature_types=None, gamma=None, grow_policy=None, importance_type=None, interaction_constraints=None, learning_rate=0.04, max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None, max_delta_step=None, max_depth=None, max_leaves=None, min_child_weight=None, missing=nan, monotone_constraints=None, multistart_strategy=None, n_estimators=100, n_jobs=None, num_parallel_tree=None, random_state=561, ...))
n_jobs	None

Model metrics: This section displays various performance metrics for the trained model:

Metric	Value
mean_test_max_error	-569.5339454157075
mean_test_neg_median_absolute_error	-57.034395335476255
std_test_neg_root_mean_squared_error	38.09426832422123
std_test_neg_median_absolute_error	8.899537027388563
training_mean_squared_error	12678.05928893058
training_mean_absolute_error	73.07199523768377
training_r2_score	-0.07798196639882549
training_root_mean_squared_error	112.59688845137143
training_score	-112.59688845137143
best_cv_score	-113.76178815921453
std_test_explained_variance	0.09025164849315176
mean_test_neg_root_mean_squared_error	-113.76178815921453
mean_test_r2	-0.40786220650780436
std_test_r2	0.2696689972156514
mean_test_neg_mean_absolute_percentage	-2.327769294459194
std_test_neg_mean_absolute_percentage	1.0784336061750013
mean_test_explained_variance	0.0700136207256711

On the right side of the interface, there are buttons for 'Register model' and 'Activate Windows'. The 'Activate Windows' button includes the text 'Go to Settings to activate Windows.'

Thông tin của một thời điểm huấn luyện mô hình XGB trên bộ dữ liệu thành phố Hồ Chí Minh

3. Phương pháp đề xuất

> Hệ thống tương tác với quy trình MLOps bằng chatbot

- **Mô hình ngôn ngữ lớn hỗ trợ người dùng tương tác tốt hơn với quy trình MLOps của BKPrice System**
- **Đối với những câu hỏi đáp về lĩnh vực MLOps, BKPrice System tận dụng khả năng và kiến thức của mô hình ngôn ngữ lớn để trả lời cho người tương tác**
- **Đối với những truy vấn tương tác với hệ thống MLOps thì sẽ được giải quyết bởi controller prompt. Đây là một prompt làm nổi bật tính tự động hóa gọi chức năng (function calling)**

3. Phương pháp đề xuất

> Hệ thống tương tác với quy trình MLOps bằng chatbot

- **Classifier Prompt:** Bạn là người phân loại, bạn nhận được truy vấn sau từ, người dùng, sử dụng thông tin chi tiết, ...
- **Information Prompt:** Bạn là một chatbot được chuyên hỏi đáp về quy trình MLOPS. Bạn có thể sử dụng những hiểu biết mà tôi cung cấp dưới đây, ...
- **Controller Prompt:** Bạn là người điều khiển, bạn nhận được truy vấn từ người dùng, sử dụng thông tin mà tôi cung cấp và chọn hành động chính tốt nhất từ danh sách những hàm số mà tôi liệt kê dưới đây, ...

3. Phương pháp đề xuất

> Hệ thống tương tác với quy trình MLOps bằng chatbot

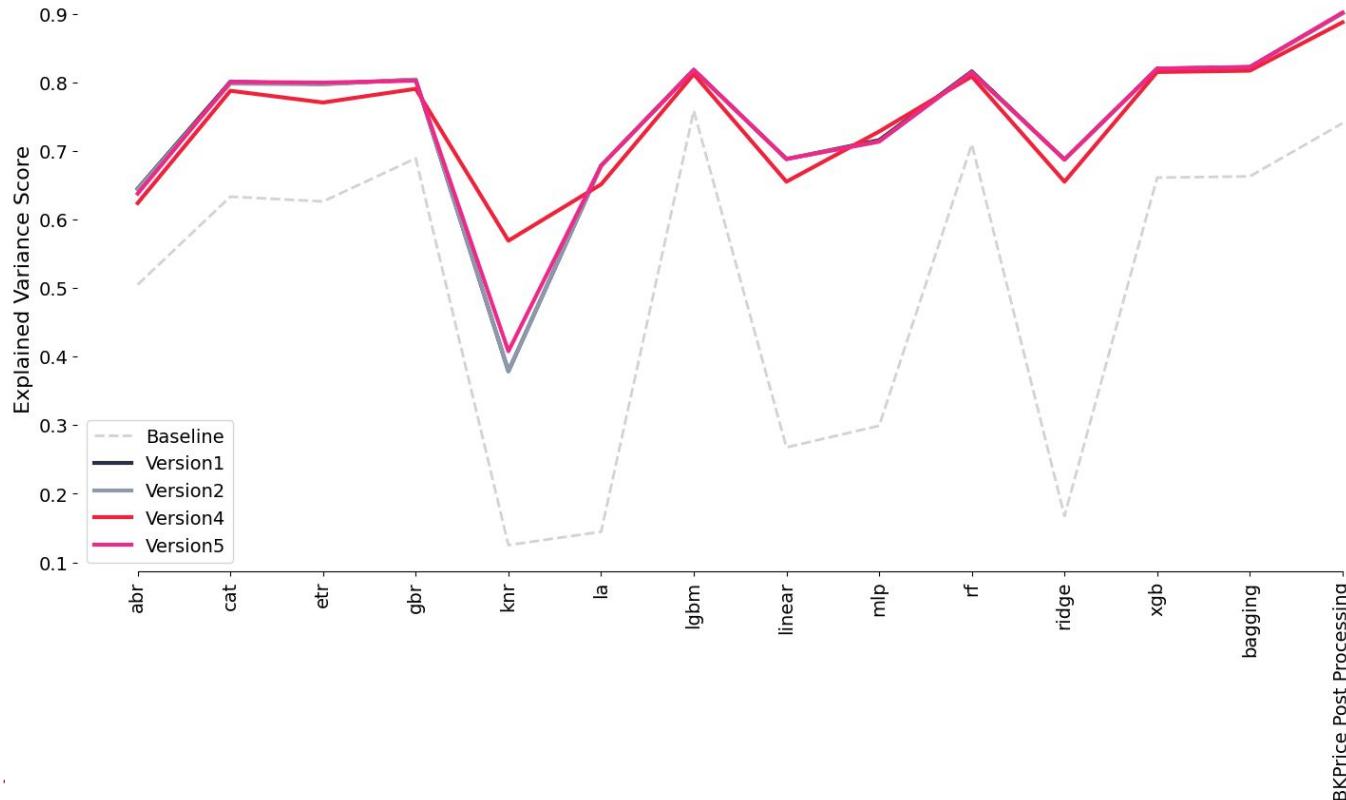
Function Name	Parameter	Description
_crawl_data	+ realestate_source	Tiến hành thu thập dữ liệu
_clean_data		Tiến hành làm sạch dữ liệu từ hàng đợi
_insert_data		Tiến hành đẩy dữ liệu từ hàng đợi vào database
_build_offline_batch_data		Trích xuất đặc trưng, xây dựng tập huấn luyện
_train_price_prediction_model	+ modelname + feature_set_version + city	Huấn luyện mô hình dự đoán giá

Bảng 4.3: Function Calling List

4. Kết quả thực nghiệm

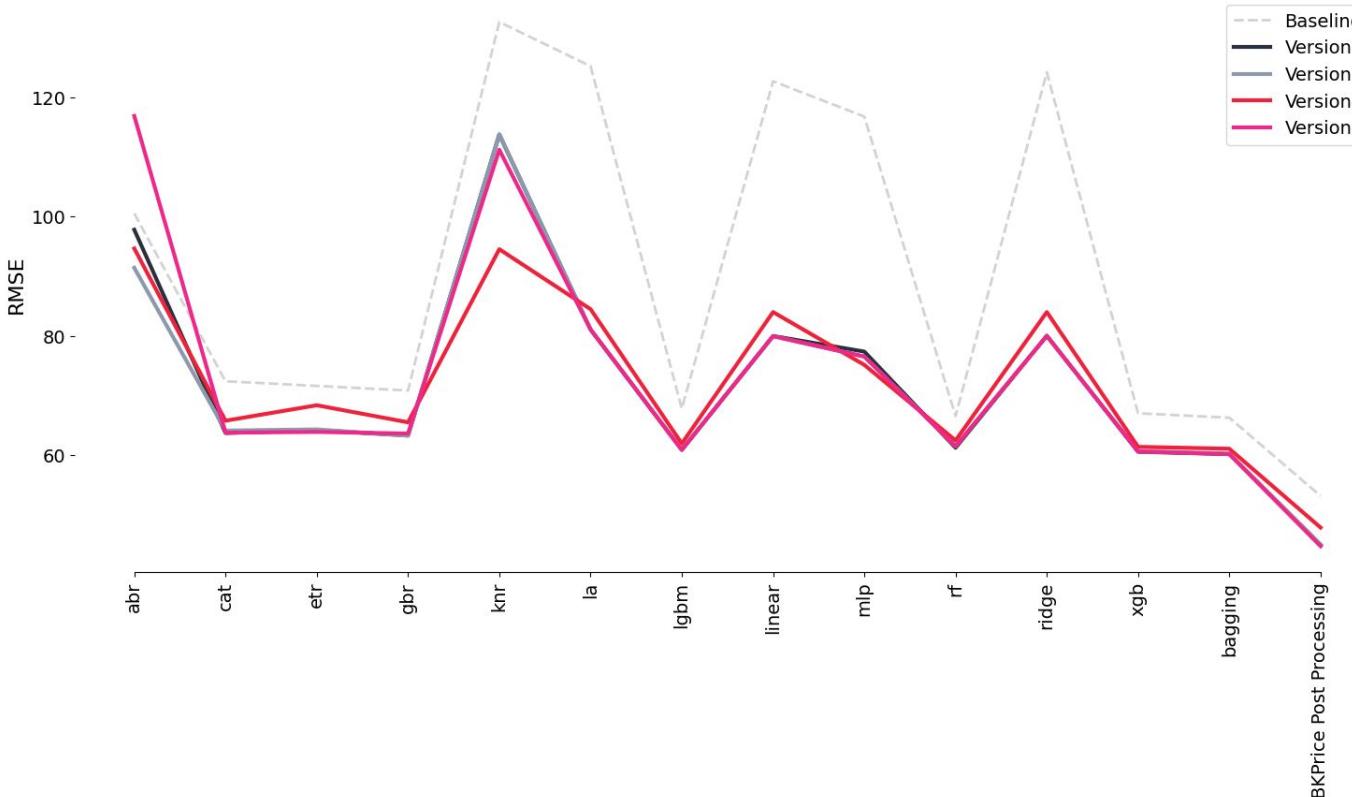
➤ Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó

HCM
EV



4. Kết quả thực nghiệm

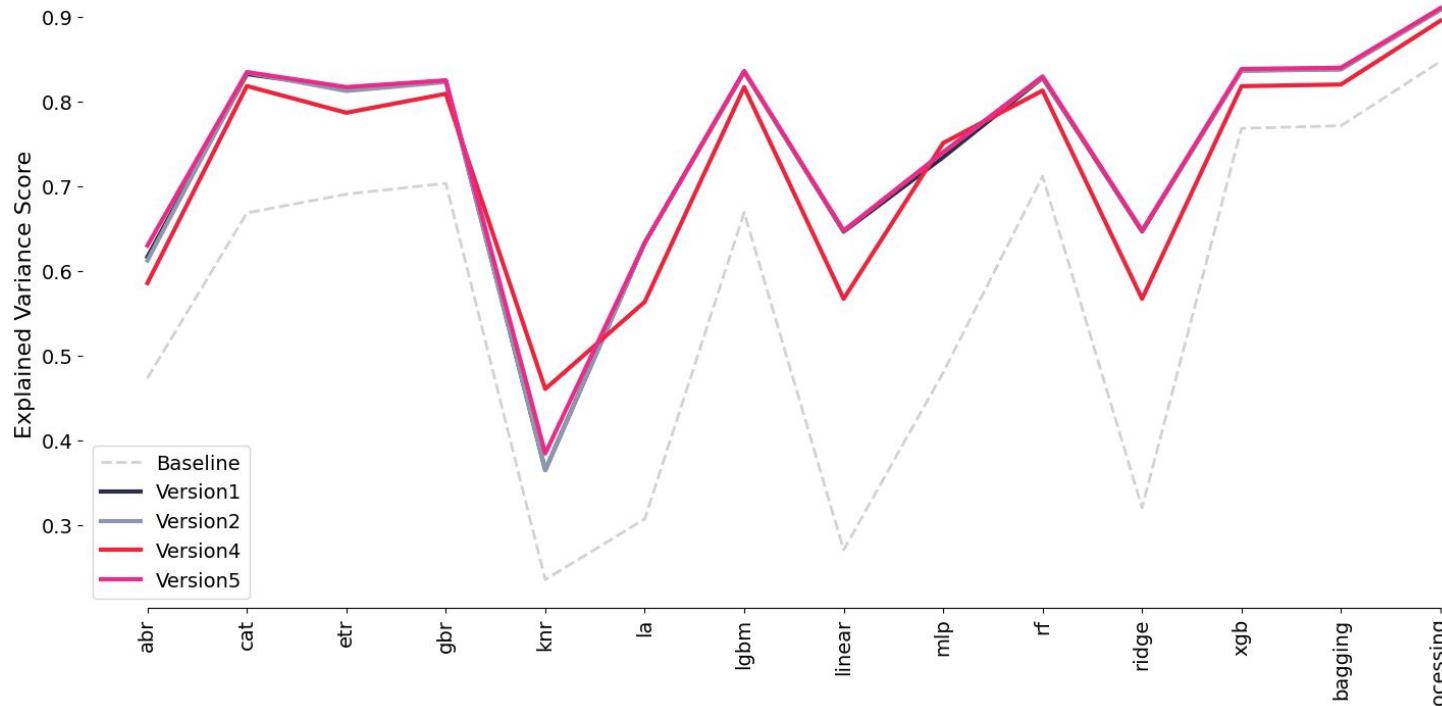
› Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó



4. Kết quả thực nghiệm

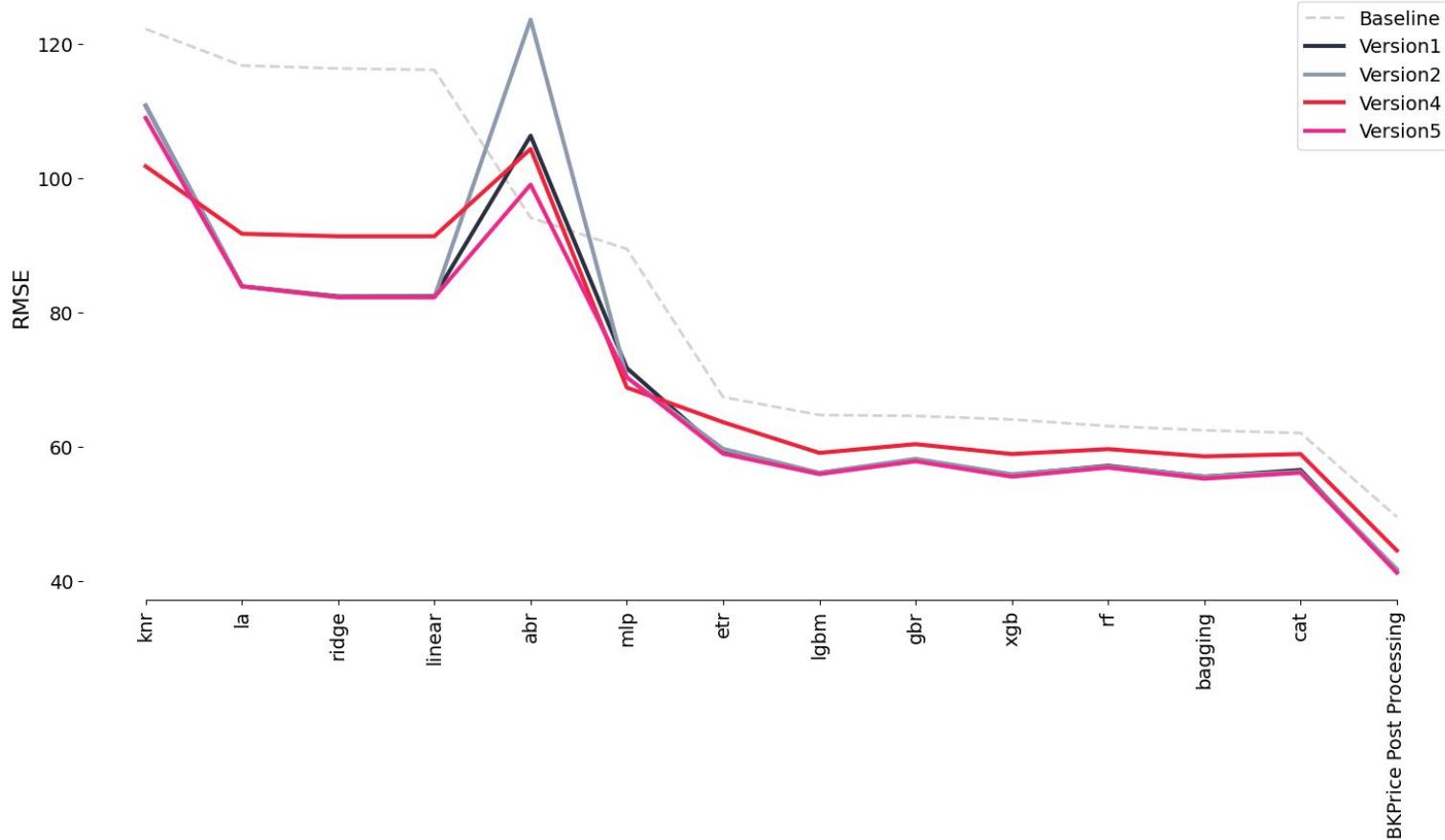
➤ Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó

HN EV



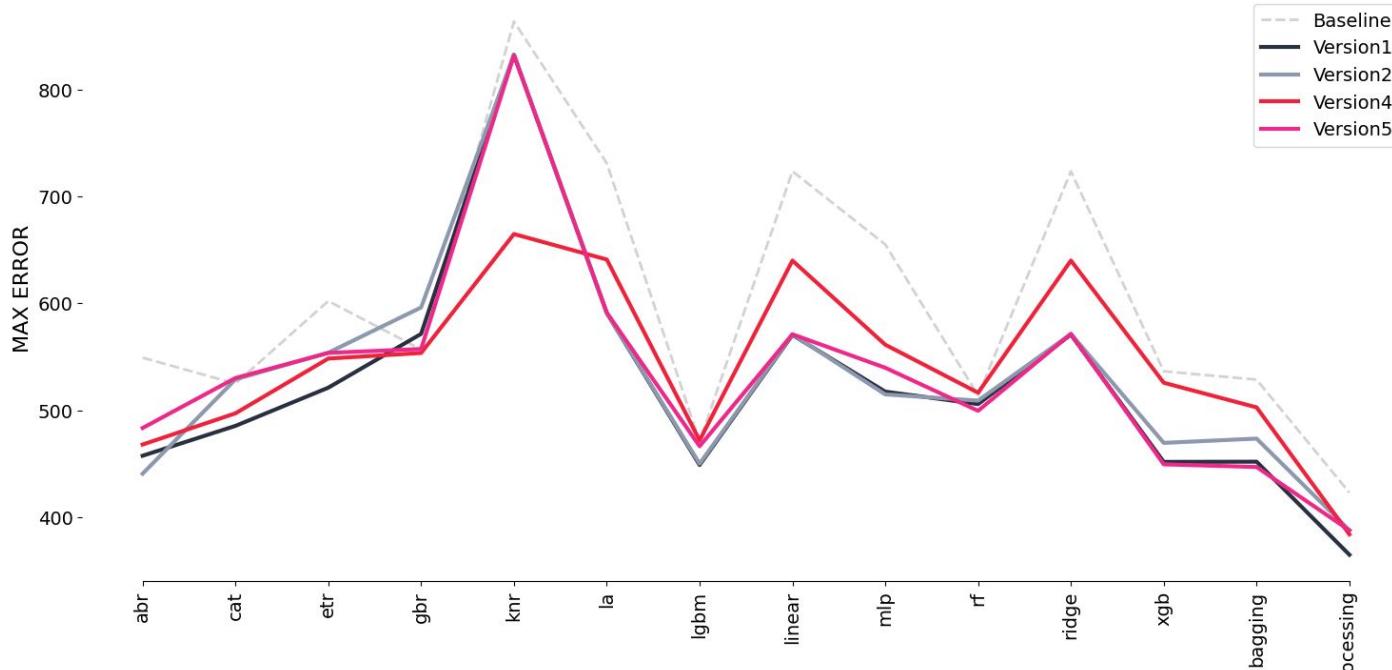
4. Kết quả thực nghiệm

› Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó



4. Kết quả thực nghiệm

› Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó

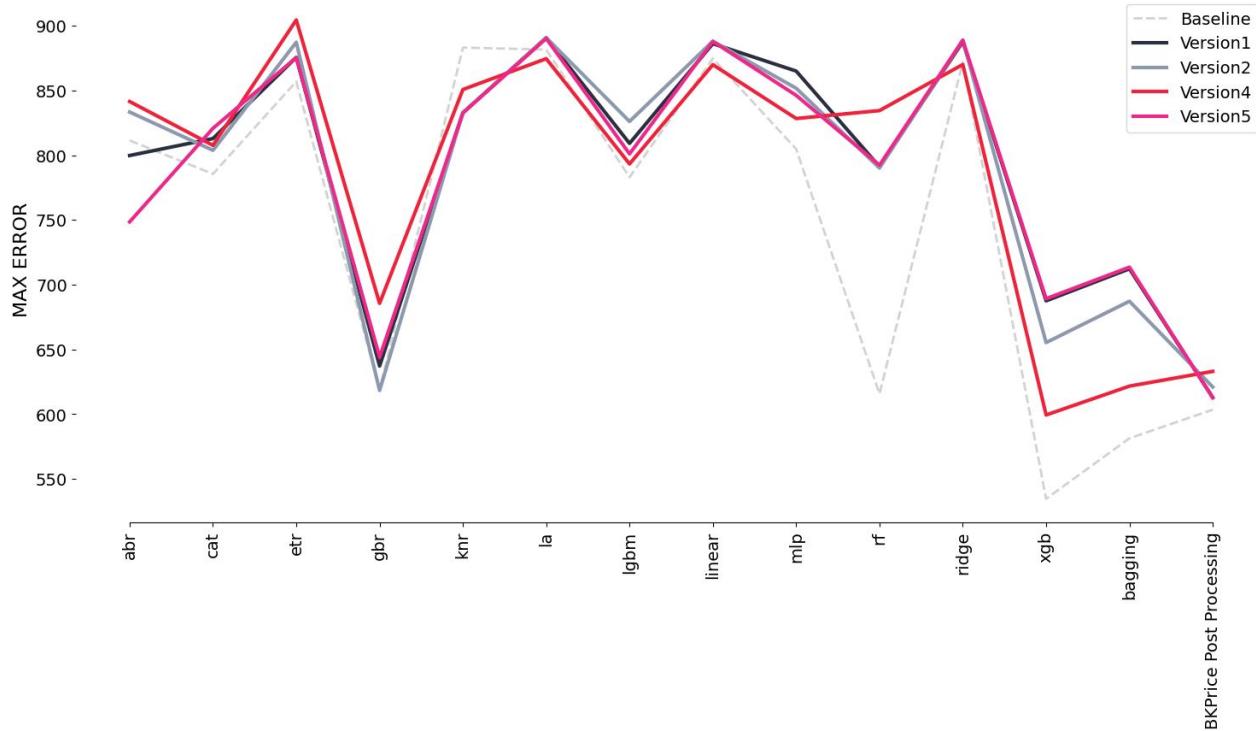


HN
MAX
ERROR

4. Kết quả thực nghiệm

› Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó

HCM
MAX
EROR



4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh kết quả định giá với trang batdongsan.com

Mức giá 5,79 tỷ
Diện tích 58 m²
Phòng ngủ 3 PN
~99.83 triệu/m²
2 tầng

↑ 2,4% Giá tại khu vực này đã tăng trong vòng 1 năm qua. Xem lịch sử giá >

D

Thông tin mô tả

* Bán nhà chính chủ - hẻm xe hơi 5m*.
Tôi cần bán gấp nhà hẻm xe hơi 6m ngay đường Nguyễn Trọng Tuyển. Căn nhà tân huyết, được thiết kế nội thất gỗ lim, nhà thoáng mát, khu vực cực kỳ an ninh, yên tĩnh.

Địa chỉ: 343/ Nguyễn Trọng Tuyển, phường 1, quận Tân Bình.
Diện tích: ngang 4m x 14.5m (58m).
Kết cấu: Nhà 1 trệt, 2 lầu.
Giá còn thương lượng 5 tỷ 790 triệu đồng.
Nhà trong hẻm rộng rãi, hiện nhà tôi đang cho thuê 15 triệu/tháng. Anh chị nào cần ở liền cứ báo với tôi.
Nhà mới sửa lại nên ko cần phải sửa gì, vô là ở ngay.
Khu dân đông, có camera quan sát, an ninh. Tiện kinh doanh mua bán. Gần bách hoá xanh, family mart, gần trường học, chợ.

Góp ý **Chat**

Đặc điểm bất động sản

Diện tích 58 m² Mức giá 5,79 tỷ
Đường vào 5 m Số tầng 2 tầng

Collections Environments History

POST {{domain}}/predict-realestate-batch Send

Params Auth Headers (9) Body Body Pre-req. Tests Settings Cookies Beautify

raw JSON

```
10
11   "body": {
12     "landSize": 58,
13     "city": "hcm",
14     "district": "tân bình",
15     "ward": "Phường 1",
16     "street": "Nguyễn Trọng Tuyển",
17     "prefixDistrict": "quận",
18     "numberOffloors": 1,
19     "numberOfBathRooms": 4,
20     "numberOfLivingRooms": 3,
21     "endWidth": 4,
22     "frontWidth": 4,
23     "frontRoadWidth": 5,
24     "certificateOfLandUseRight": true,
25     "typeOfRealEstate": "privateProperty",
26     "facade": "oneSideOpen",
27     "houseDirection": "east",
28     "facility_check_ok": false,
29     "narrow_alley": 3,
30     "version": "v5"
```

Body 200 OK 27.79 s 145 B Save as example

Pretty Raw Preview Visualize JSON

```
1 [
2   124.59206197863648
3 ]
```

Activate Windows

4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh kết quả định giá với trang batdongsan.com

Chính Chủ cần bán gấp căn nhà tại Giải Phóng Phương Mai. Giá định chuyển nơi ở. Nhà đẹp trung tâm

Đường Giải Phóng, Phường Phương Mai, Đống Đa, Hà Nội

Mức giá Diện tích
17 tỷ 47 m²
~361,7 triệu/m² Mặt tiền 4.07 m



↑ 20,6% Giá tại khu vực này đã tăng trong vòng 1 năm qua.

Xem lịch sử giá >

Thông tin mô tả

Chính Chủ cần bán gấp căn nhà tại Giải Phóng.



Nhà đẹp mới koong - 2 mặt thoáng vĩnh viễn - oto đỗ cửa - ở hoặc cho thuê kinh doanh quá tốt - gần nhiều tiện ích - tt quận
Đống Đa.

Diện tích 47 m mặt tiền 4.07 m giá 17 tỷ.

* Nằm trên trục đường chính nhà dân xây kiên cố.

* Nhà 3 mặt thoáng vĩnh viễn.

* Nội thất full.

* Gần nhiều tiện ích chợ trường học ở rất thuận tiện.

Số đỏ đẹp nở hậu cất két (sẵn giao dịch).

Nhà thiết kế đẹp long lanh thiết kế 5 tầng mỗi tầng 2 phòng sân phơi nhà thờ.

The screenshot shows a Postman API request. The 'Body' tab is selected, displaying a JSON object representing the house details. The response body shows a single value: 388.45039990234375.

```
11   "body": {  
12     "landSize": 47,  
13     "city": "hn",  
14     "district": "đống đa",  
15     "ward": "Phương Mai",  
16     "street": "Giải Phóng",  
17     "prefixDistrict": "quận",  
18     "numberOffloors": 5,  
19     "numberOfBathRooms": 6,  
20     "numberOfLivingRooms": 6,  
21     "endWidth": 4.07,  
22     "frontWidth": 4.07,  
23     "frontRoadWidth": 7,  
24     "certificateOfLandUseRight": true,  
25     "typeOfRealEstate": "townhouse",  
26     "facade": "twoSideOpen",  
27     "houseDirection": "east",  
28     "facility_check_ok": true,  
29     "narrow_alley": -1,  
30     "version": "v5"  
}
```

Body

Pretty Raw Preview Visualize JSON

```
1 [  
2   388.45039990234375  
3 ]
```

Activate Windows

4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh kết quả định giá với trang batdongsan.com

Bán / Hà Nội / Hoàn Kiếm / Nhà riêng tại phố Hàng Bài

Cần bán gấp nhà 4 tầng Dân xây còn mới Phố Hàng Bài Hoàn Kiếm DT 36m², MT hơn 7.1m giá cực tốt.

Phố Hàng Bài, Phường Hàng Bài, Hoàn Kiếm, Hà Nội

Mức giá	Diện tích	Phòng ngủ
12 tỷ	36 m ²	4 PN
~333,33 triệu/m ²	Mặt tiền 7.1 m	

↑ 43,4% Giá tại khu vực này đã tăng trong vòng 1 năm qua.

Xem lịch sử giá >

Thông tin mô tả

Cần bán gấp nhà 4 tầng Dân xây còn mới Phố Hàng Bài Trung Tâm Hoàn Kiếm DT 36m², MT hơn 7.1m giá cực tốt.



- Vị trí quá đinh, trung tâm quận Hoàn Kiếm.

Khu vực Lõi Hoàn Kiếm cực hiếm nhà bán!

- Nhà 3 thoáng Vinh viễn.

Không gian sống tuyệt vời, tràn ngập ánh sáng, giáp mặt phố hàng Bài.

- Nhà xây dựng khung cột bê tông chắc chắn, sạch sẽ.

+ Số đồ chính chủ, sẵn sàng giao dịch ngay.

The screenshot shows a POST request in the Postman application. The 'Body' tab is selected, displaying the following JSON data:

```
11   "id": 36,
12   "landSize": 36,
13   "city": "hn",
14   "district": "hoàn kiếm",
15   "ward": "Hàng Bài",
16   "street": "Hàng Bài",
17   "prefixDistrict": "quận",
18   "numberOfWorkers": 4,
19   "numberOfBathRooms": 4,
20   "numberOfLivingRooms": 4,
21   "endWidth": 7.1,
22   "frontWidth": 7.1,
23   "frontRoadWidth": 20,
24   "certificateOfLandUseRight": true,
25   "typeOfRealEstate": "privateProperty",
26   "facade": "twoSideOpen",
27   "houseDirection": "east",
28   "facility_check_ok": true,
29   "narrow_alley": 0,
30   "version": "v5"
```

The response body is shown below, indicating a successful 200 OK status with a response time of 16.32 seconds and a size of 145 bytes. The response content is a single numerical value: 356.51206771762384.

Body

Pretty Raw Preview Visualize JSON

1 [
2 356.51206771762384
3]

200 OK 16.32 s 145 B Save as example

4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh kết quả định giá với trang batdongsan.com

Nhà trần duy hưng giá hiếm .khu vực ít nhà bán Nhà trần duy hưng giá hiếm .khu vực ít nhà bán

Đường Trần Duy Hưng, Phường Trung Hòa, Cầu Giấy, Hà Nội

Mức giá
30 tỷ
~200 triệu/m²

Diện tích
150 m²
Mặt tiền 8.2 m

↑ 19,9% Giá tại khu vực này đã tăng trong vòng 1 năm qua.

Xem lịch sử giá >

Thông tin mô tả

Bán nhà phố Trần Duy Hưng diện tích 150m² 6 tầng mặt tiền 8,2m.
2 mặt ngõ - ô tô vào nhà - thoáng sáng - khu vip nhất của quận - nhà đồng tiền - dòng tiền lớn.

+ Vị trí:

Trung tâm Cầu Giấy khu vip, xung quanh hàng xóm toàn tòa 9 tầng to vẹt vã, Kinh Doanh sầm uất.

- Hiện trạng nhà 6 tầng, 1 tum thang máy thang bộ

- Pháp lý sạch, sổ đỏ đẹp nở hậu, chính chủ sẵn sàng giao dịch.

Gia chủ đang chào 29 tỷ

Chúc anh chị buôn may bán đắt phát lộc phát tài khi đồng hành cùng em phong bất động sản.

Em Phong chuyên nhà triệu đô chỉ cần anh chị có nhu cầu là em phong tim được.

Mãi yêu.

Đặc điểm bất động sản

The screenshot shows a POST request in Postman to the endpoint `((domain))/predict-realstate-batch`. The Body tab displays the following JSON data:

```
12     ... "landSize": 150,
13     ... "city": "hn",
14     ... "district": "cầu giấy",
15     ... "ward": "Trung Hòa",
16     ... "street": "Trần Duy Hưng",
17     ... "prefixDistrict": "quận",
18     ... "numberOffloors": 6,
19     ... "numberOfBathRooms": 4,
20     ... "numberOfLivingRooms": 4,
21     ... "endWidth": 8.2,
22     ... "frontWidth": 8.2,
23     ... "frontRoadWidth": 4,
24     ... "certificateOfLandUseRight": true,
25     ... "typeOfRealEstate": "privateProperty",
26     ... "facade": "twoSideOpen",
27     ... "houseDirection": "east",
28     ... "facility_check_ok": false,
29     ... "narrow_alley": 3,
30     ... "version": "v5"
```

The response at the bottom shows a status of 200 OK with a response time of 47.01s and a size of 145 B. The response body is a single value: 225.48833803828123.

4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh kết quả định giá với trang batdongsan.com

Đường Nguyễn Thái Học, Phường Cầu Ông Lãnh, Quận 1, Hồ Chí Minh

Mức giá
50 tỷ
~649,35 triệu/m²

Diện tích
77 m²
7 tầng

Phòng ngủ
18 PN

↓ 7,4% Giá tại khu vực này đã giảm trong vòng 1 năm qua.



Xem lịch sử giá >

Thông tin mô tả

- * Chính chủ bán khách sạn đang hoạt động liền kề chợ Bến Thành, trung tâm quận 1, vị trí vàng mặt tiền đường 30m Nguyễn Thái Học, Quận 1.
- * Diện tích sử dụng: 516,60m²; 1 hầm + 6 lầu (có 18 phòng + thang máy xịn); khách sạn mới xây, nội thất mới, xịn... Vào tiếp tục khai thác.
- * Vỉa hè rộng 10m, thuận tiện kinh doanh.
- * Đang có dòng tiền thu hàng tháng: 550tr/tháng.
- * Số hồng hoàn công chuẩn chỉnh, công chứng trong ngày!...



Góp ý



Chat

Đặc điểm bất động sản

	Diện tích	77 m ²
	Đường vào	30 m
	Số phòng ngủ	18 phòng

	Mức giá	50 tỷ
	Số tầng	7 tầng
	Số toilet	19 phòng

The screenshot shows a POST request to a real estate API endpoint. The JSON body of the request contains detailed information about a property located at 30m Nguyen Thai Hoc, District 1, Ho Chi Minh City. The property is a townhouse (townhouse) with one side open (oneSideOpen), 18 bedrooms (numberOfWorkingRooms: 18), 1 bathroom (numberOfBathRooms: 1), 19 living rooms (numberOfLivingRooms: 19), 7 floors (numberOffloors: 7), 5 meters wide (endWidth: 5), 5 meters deep (frontWidth: 5), and a front road width of 30m (frontRoadWidth: 30). It has a certificate of land use right (certificateOfLandUseRight: true) and is categorized as a townhouse (typeOfRealEstate: "townhouse"). The facade is one-side open (facade: "oneSideOpen"), it faces east (houseDirection: "east"), and the facility check is successful (facility_check_ok: true). There are 3 narrow alleys (narrow_alley: 3). The version of the data is v5 (version: "v5").

```
11     "body": {  
12         "landSize": 77,  
13         "city": "hcm",  
14         "district": "quận 1",  
15         "ward": "Cầu Ông Lãnh",  
16         "street": "Nguyễn Thái Học",  
17         "prefixDistrict": "quận",  
18         "numberOffloors": 7,  
19         "numberOfBathRooms": 1,  
20         "numberOfLivingRooms": 18,  
21         "endWidth": 5,  
22         "frontWidth": 5,  
23         "frontRoadWidth": 30,  
24         "certificateOfLandUseRight": true,  
25         "typeOfRealEstate": "townhouse",  
26         "facade": "oneSideOpen",  
27         "houseDirection": "east",  
28         "facility_check_ok": true,  
29         "narrow_alley": 3,  
30         "version": "v5"  
}
```

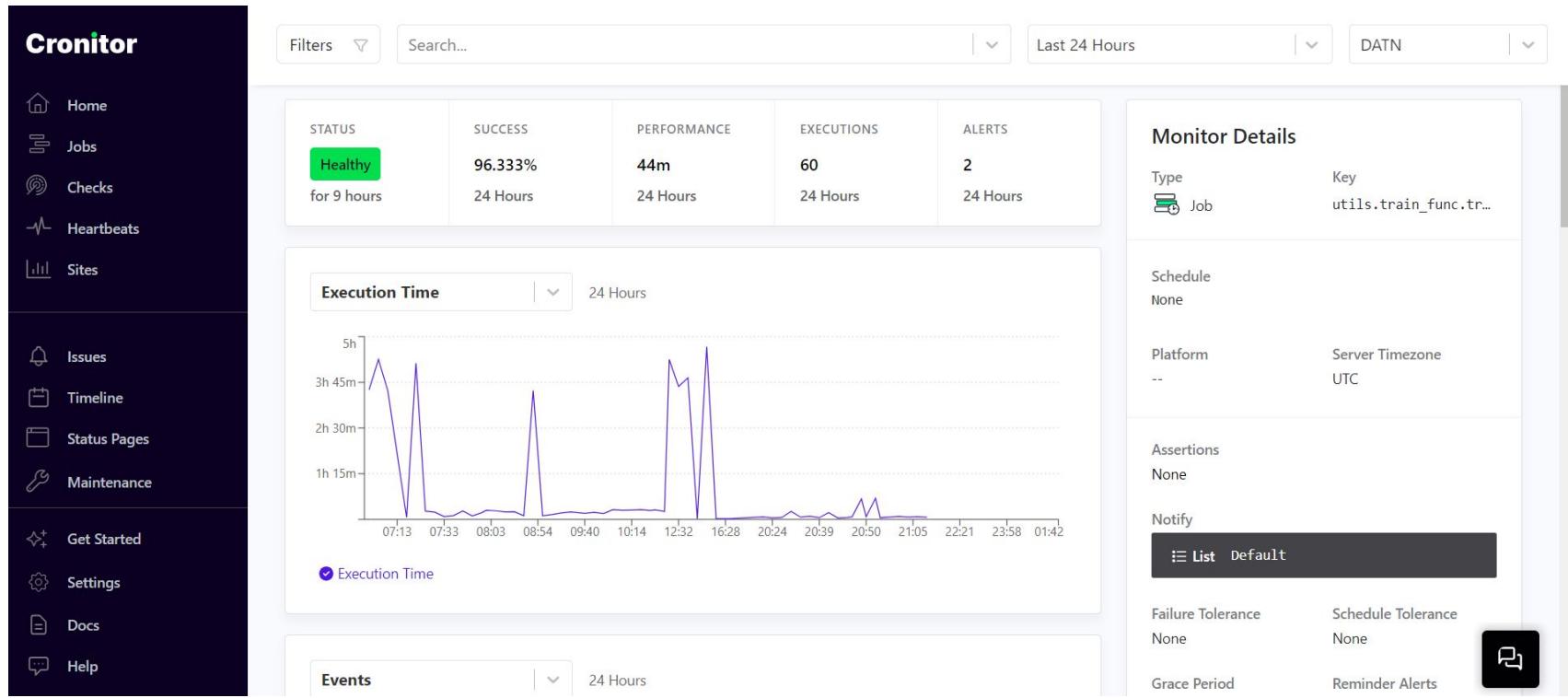
Body

Pretty Raw Preview Visualize JSON

1 [
2 629.7490075192859
3]

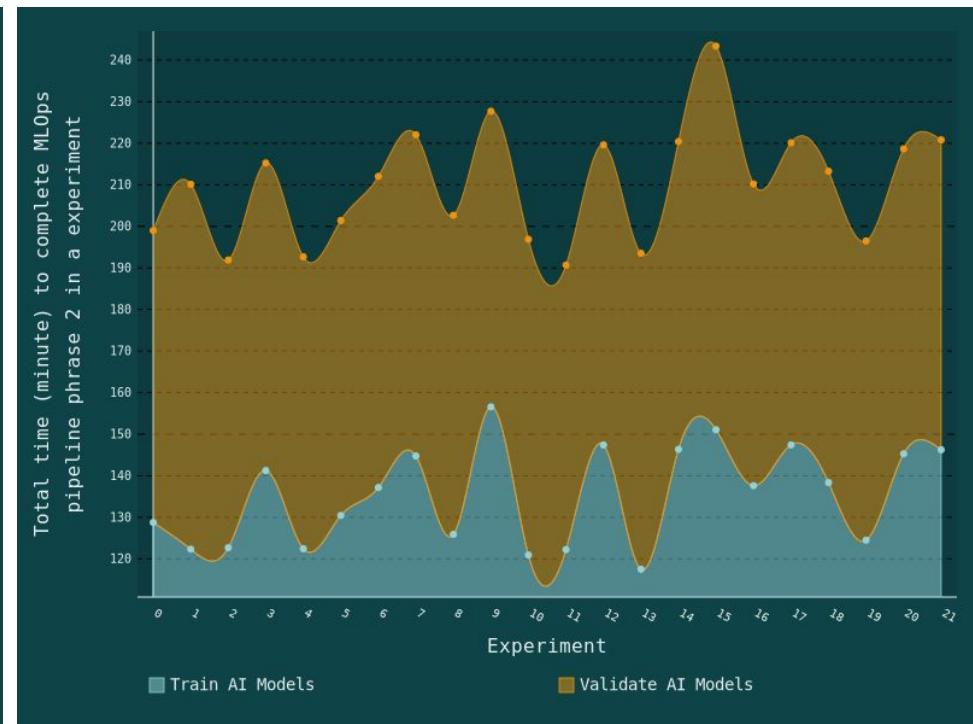
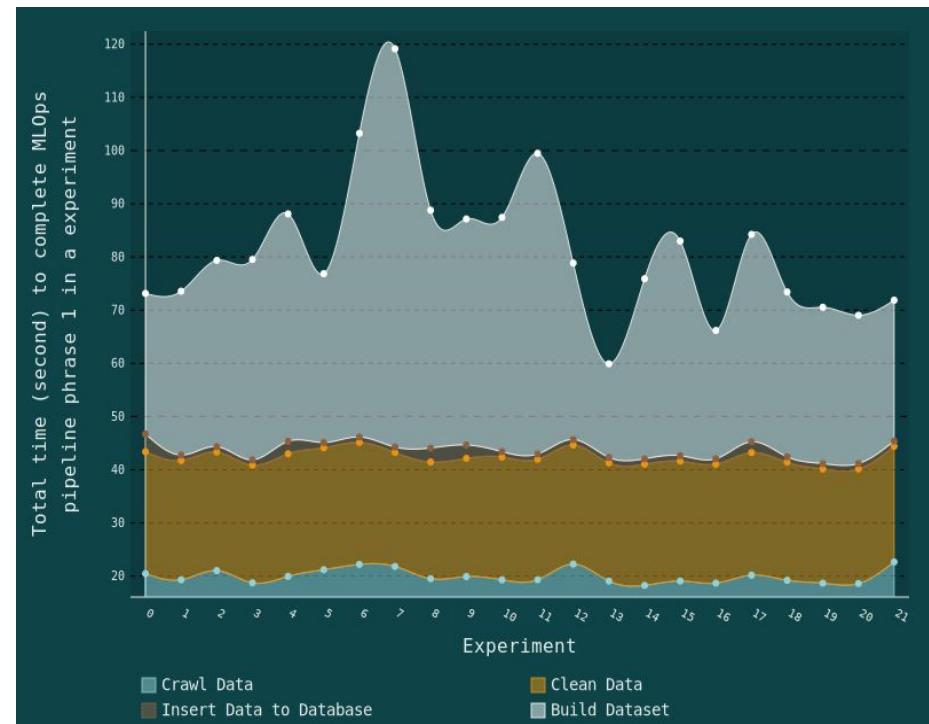
4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm đo tỷ lệ thành công huấn luyện mô hình tự động



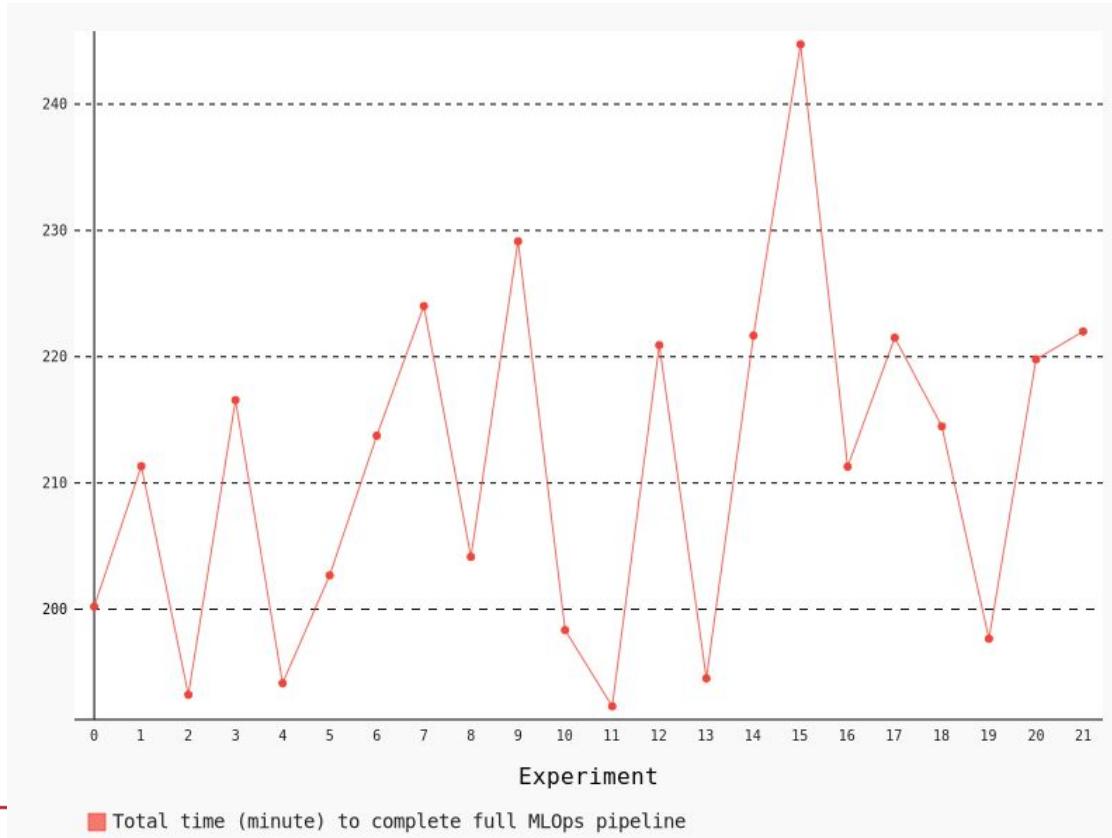
4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm đo thời gian hoàn thành quy trình MLOps hoàn chỉnh



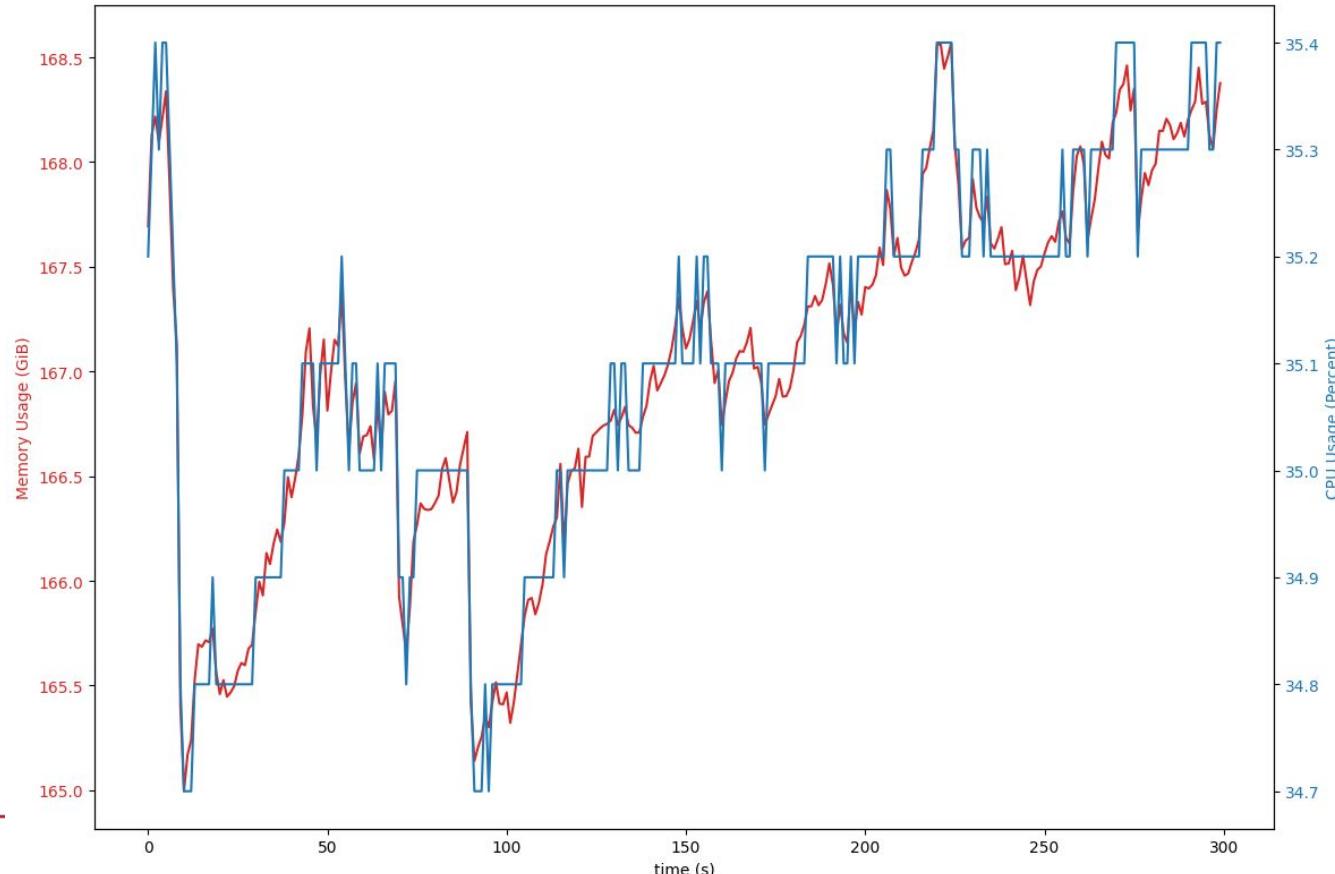
4. Kết quả thực nghiệm

› Hình ảnh thực nghiệm đo thời gian hoàn thành quy trình MLOps hoàn chỉnh



4. Kết quả thực nghiệm

› Hình ảnh thực nghiệm đo thời gian hoàn thành quy trình MLOps hoàn chỉnh



4. Kết quả thực nghiệm

> Hình ảnh thực nghiệm so sánh với các giải thuật định giá trước đó

The screenshot shows a chatbot interface with a white background and a light gray header bar. In the top left corner of the header bar, there is a small icon of a speech bubble with a dot and the text "Chatbot". On the right side of the header bar, there is a rectangular input field containing the text "Hello. What can you do". Below the header, the main area has a light gray background. A message from the bot is displayed in a light gray box: "I can help you build an MLOPs workflow for real estate price prediction. I can help you with data collection, cleaning, transformation, and feature engineering. I can also help you train and evaluate machine learning models, and deploy them to production. Additionally, I can help you monitor the performance of your models and retrain them as needed." Below this message, another message from the user is shown in a light gray box: "Oh good job. What real estate source can you crawl?". At the bottom of the screen, there are three buttons in a horizontal row: "Retry" with a circular arrow icon, "Undo" with a circular arrow icon, and "Clear" with a trash bin icon.

Chatbot

Hello. What can you do

I can help you build an MLOPs workflow for real estate price prediction. I can help you with data collection, cleaning, transformation, and feature engineering. I can also help you train and evaluate machine learning models, and deploy them to production. Additionally, I can help you monitor the performance of your models and retrain them as needed.

Oh good job. What real estate source can you crawl?

In demo version, I only support crawling data from meyland source.

Retry Undo Clear

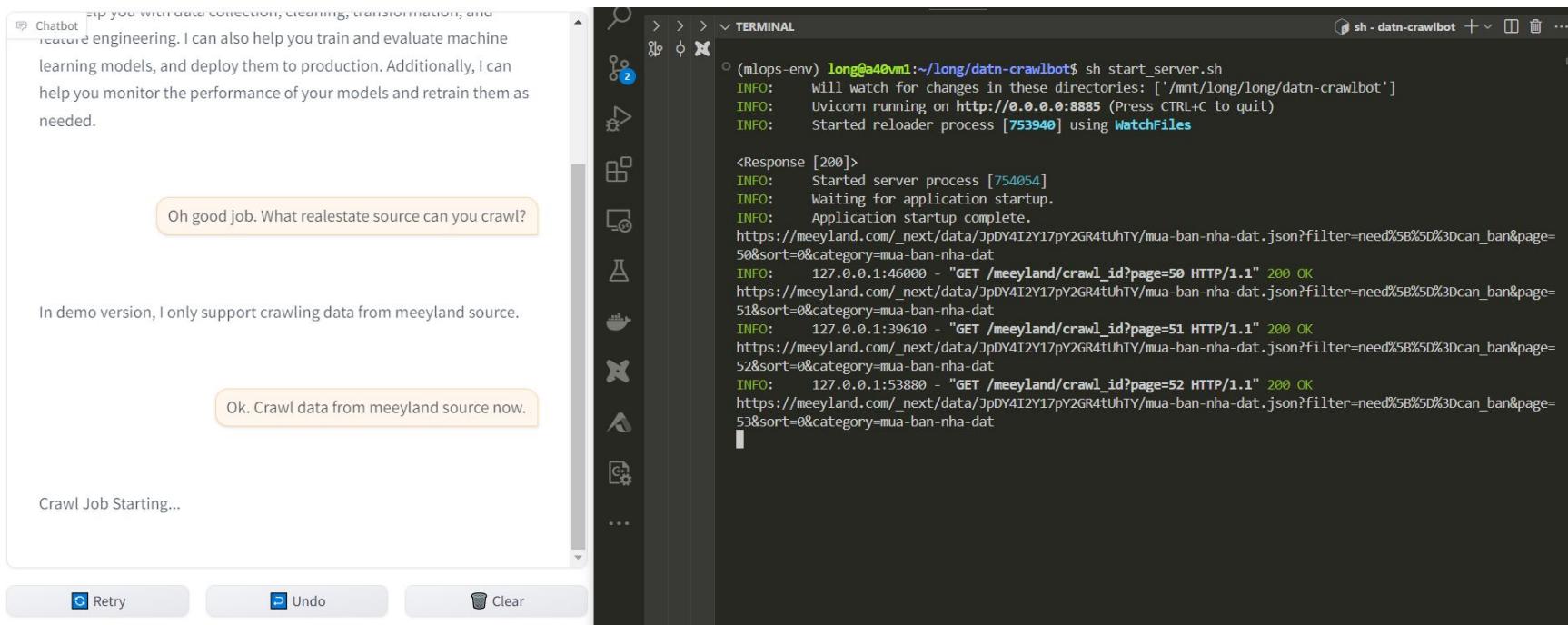
4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot

The screenshot shows a chatbot interface with a red header bar. The header bar contains the text 'Hình ảnh hệ thống chat bot' and a back arrow icon. Below the header is a white main area. In the top left corner of the main area, there is a small button labeled 'Chatbot' with a speech bubble icon. On the right side of the main area, there is a text input field containing the message 'Hello. What can you do'. Below this message, a response from the bot is shown in a yellow box: 'I can help you build an MLOPs workflow for real estate price prediction. I can help you with data collection, cleaning, transformation, and feature engineering. I can also help you train and evaluate machine learning models, and deploy them to production. Additionally, I can help you monitor the performance of your models and retrain them as needed.' Further down, another message from the user is shown in a yellow box: 'Oh good job. What realestate source can you crawl?'. At the bottom of the main area, there is a message from the bot: 'In demo version, I only support crawling data from meyland source.' At the very bottom of the interface, there are three buttons: 'Retry' with a circular arrow icon, 'Undo' with a circular arrow icon, and 'Clear' with a trash bin icon.

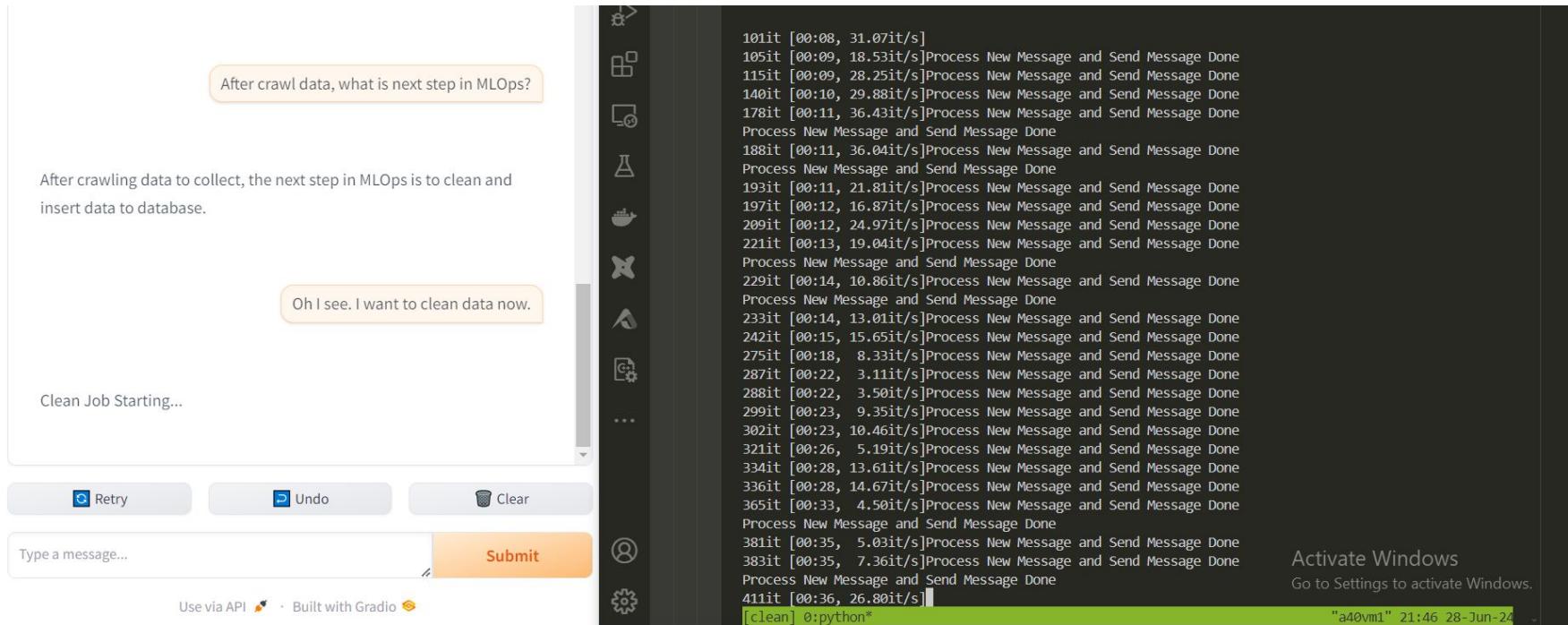
4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



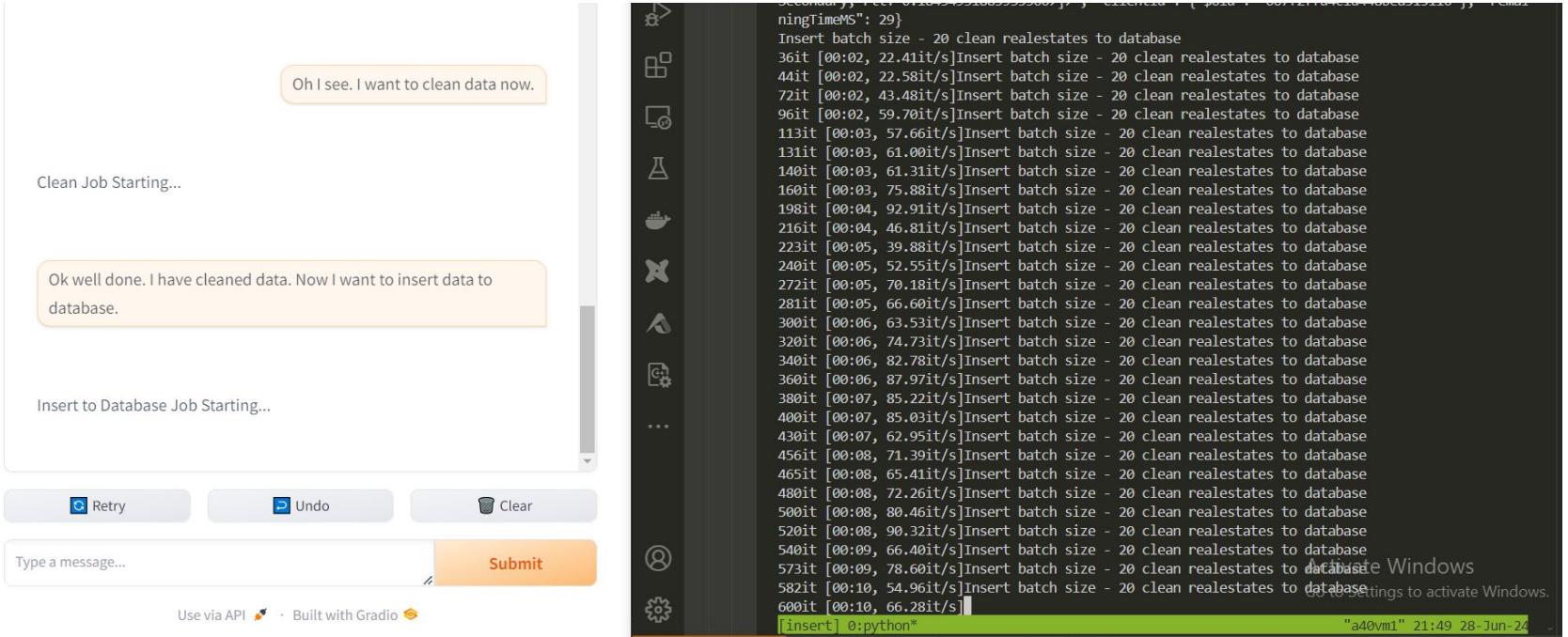
4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



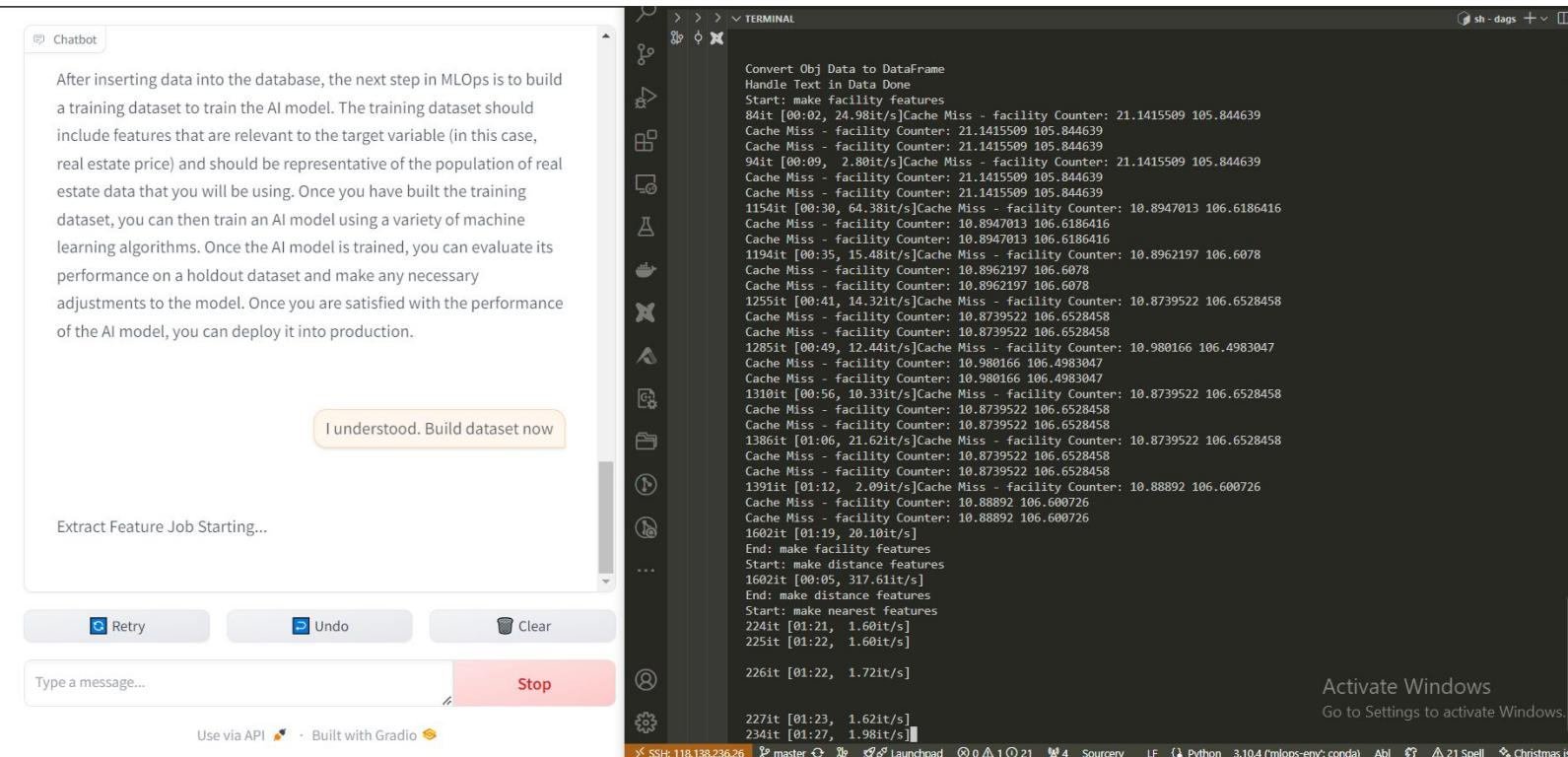
4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



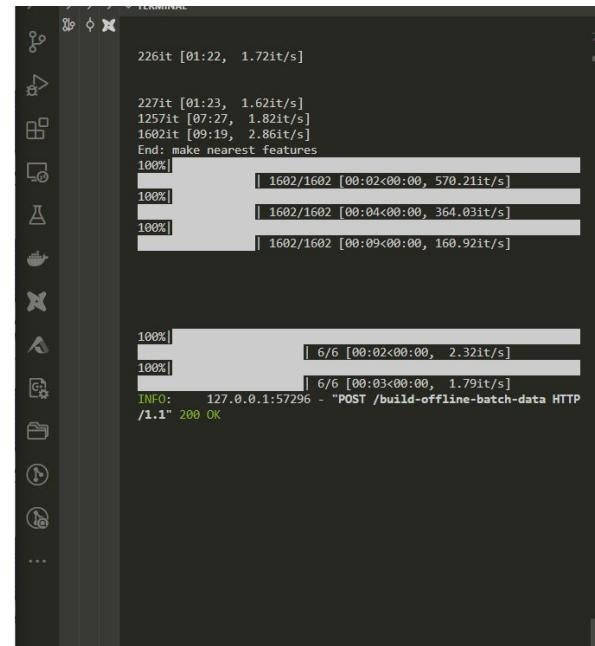
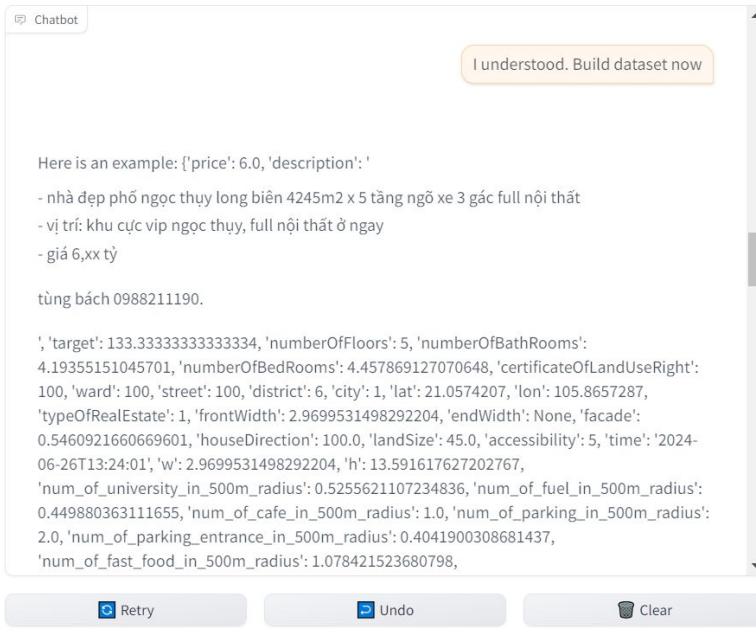
4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot

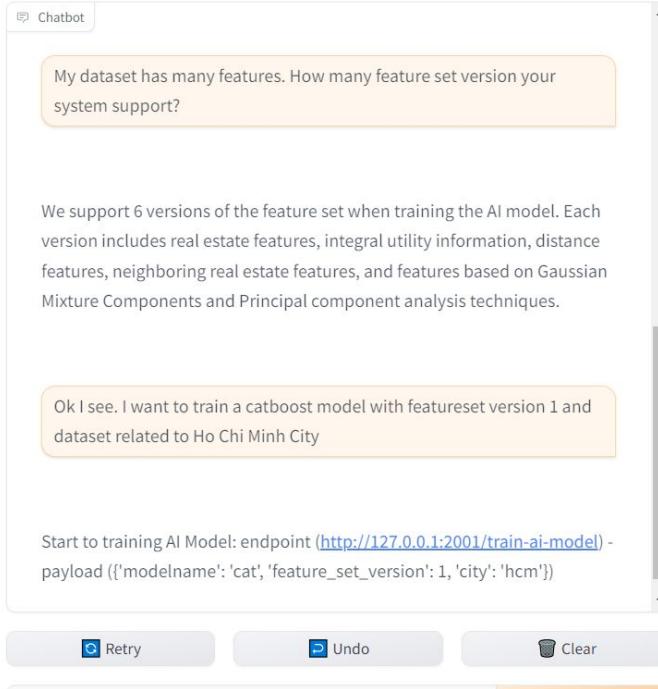
The screenshot shows a chatbot interface with a red header bar containing the title 'Hình ảnh hệ thống chat bot'. The main area is a white window titled 'Chatbot' with a small icon. A message bubble from the user says: 'Ok Done. I want to build a AI model to predict realestate price. What kinds of AI model you support?'. Below this, a response from the bot says: 'Currently, we support these AI model types for real estate price prediction:' followed by a list of supported models.

Currently, we support these AI model types for real estate price prediction:

- Single model:
 - CatBoostRegressor (cat)
 - LGBMRegressor (lgbm)
 - XGBRegressor (xgb)
 - AdaBoostRegressor (abr)
 - GradientBoostingRegressor (gbr)
 - KNeighborsRegressor (knr)
 - Lasso (la)
 - LinearRegression (linear)
 - MLPRegressor (mlp)
 - RandomForestRegressor (rf)
 - Ridge (ridge)
- Ensemble model

4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot



```
[2024-06-28 22:19:34,480: INFO/MainProcess] Task utils.train_func.train_model_by_city_data_and_feature_version[4a035386-6bc0-42cf-a914-c76c7f3278df] received [2024-06-28 22:19:41,891: WARNING/ForkPoolWorker-1] TRAIN_SHAPE: [2024-06-28 22:19:41,891: WARNING/ForkPoolWorker-1] (534, 186) [2024-06-28 22:19:42,067: WARNING/ForkPoolWorker-1] TEST_SHAPE: [2024-06-28 22:19:42,067: WARNING/ForkPoolWorker-1] (9149, 186) [2024-06-28 22:19:57,295: WARNING/ForkPoolWorker-1] Start to calculating correlation map [2024-06-28 22:26:01,911: WARNING/ForkPoolWorker-1] Start to training [2024-06-28 22:26:05,980: WARNING/ForkPoolWorker-1] Start training update model [2024-06-28 22:26:06,266: WARNING/ForkPoolWorker-1] Learning rate set to 0.035792 [2024-06-28 22:26:07,350: WARNING/ForkPoolWorker-1] 0: learn: 83.4027476 total: 680ms remaining: 11m 19s [2024-06-28 22:27:14,274: WARNING/ForkPoolWorker-1] 100: learn: 46.4983215 total: 1m 7s remaining: 9m 58s [2024-06-28 22:28:17,252: WARNING/ForkPoolWorker-1] 200: learn: 29.6708730 total: 2m 9s remaining: 8m 35s [2024-06-28 22:29:23,944: WARNING/ForkPoolWorker-1] 300: learn: 21.0171942 total: 3m 16s remaining: 7m 35s
```

4. Kết quả thực nghiệm

> Hình ảnh hệ thống chat bot

Ok Good. While model is being trained, i want to observe training process. Where can I see process of training?

To observe the training process, you can go to the link <http://localhost:5001/#/experiments/1>. This link will take you to the mlflow interface, where you can see all the experiments that have been run. Each experiment will have a list of runs, and each run will have a list of metrics and parameters. You can click on any run to see more details about it, including the training loss and accuracy graphs.

Retry

Undo

Clear

5. Đóng góp nổi bật

1. Tính mới và tính sáng tạo

Ý tưởng xây dựng một hệ thống có thể **tương tác sâu** với người dùng qua chatbot. Tận dụng đa mức thông tin để huấn luyện mô hình và đưa ra kết quả định giá tin cậy với người dùng. Bên cạnh đó đề xuất 1 hệ thống tự động hóa end-to-end cho bài toán bất động sản

2. Đa dạng công nghệ

Hệ thống BKPrice **kết hợp đa dạng công nghệ** như quản lý feature và quản lý huấn luyện mô hình AI, các công nghệ về tối ưu hóa quá trình xử lý dữ liệu (airflow, kafka), công nghệ function calling trong LLM.

6. Tính ứng dụng

NESTOCK BETA

Tìm kiếm truyền thống

Danh giá bất động sản

Tạo tài khoản Nestock

Nestock Wiki

Vị trí: Hà Nội (Hàng, mảng hoa)

Chi tiết đơn giá BDS

Đơn giá (kết quả) trung bình trong khu vực:
71/m²

Tham khảo bảng giá (tháng 07/2024)

TP.Hà Nội

quận/huyện

Đô thị

Đô thị

Bắc Từ Liêm

Chang Mỹ

Cát Giã

Gia Lâm

Hai Ba Trưng

Hoàn Kiếm

Dịnh giá nhà riêng **Dịnh giá chung cư**

Vị trí
Khu vực ngõ*

Mặt tiền: Ngõ 3 ô tô 7-9m | Ngõ 2 ô tô 5-7m | Ngõ ô tô đồ cửa 4-5m
Ngõ ô tô tránh xe máy 2.5-4m | Ngõ ba gác 2.5-3m | Ngõ **<2.5m**

Vị trí*



Thông tin thửa đất

Diện tích (m²)*

Kích thước mặt tiền (m)

Lộ giới mặt tiền

1 mặt thoáng | 2 mặt thoáng | 3 mặt thoáng | 4 mặt thoáng | Lô góc

Hủy

Dịnh giá

 Tingxuan Sanei Zhang
Active now

A cũng học BK ra

A định gấp xem dự án tới đâu và tđao đổi về cách tiếp cận cho bài toán định giá

Xem có phuocq ánh đầu tư phát triển cùng ko

A thích sản phẩm của bọn e

Nhưng bài toán định giá bds cần tiếp cận kỹ lưỡng hơn

Nếu bọn e quan tâm, mình coa thể gấp trao đổi

A tên Vĩnh nhé

 Trương Trí Vĩnh

em thấy định giá khó ơi là khó

bên CenValua có quảng cáo về mô hình định giá mà bọn em cũng chưa nghĩ ra làm sao để có được 1 mô hình tối ưu

xong bọn em nghĩ chắc do bên đấy có hệ thống chuyên gia

7. Feedback

Nghiên cứu đã được thầy cô và anh chị đánh giá một cách tích cực, bên cạnh đó nghiên cứu còn được góp ý từ hệ thống, đến giải pháp và công nghệ. Dưới đây là danh sách người đã tích cực đánh giá và góp ý cho hệ thống để có thể có được 1 hệ thống BKPrice tương đối hoàn chỉnh như hôm nay:

- + Thầy Trần Văn Đăng tại trường Đại Học Bách Khoa Hà Nội
- + Hệ thống được trình bày cho chú Trương Đức Vĩnh là phó tổng thư ký thường trực của hội môi giới bất động sản Việt Nam và được đánh giá tích cực.
- + TS Tai Le Quy tại trường University of Applied Sciences
- + Giảng viên Nguyễn Đình Toàn tại trường Monash University
- + Anh Phạm Ngọc Lâm hiện tại là Senior Software Engineer tại Grab ở Singapore
- + Anh Lê Quang Hòa hiện tại là Software Engineer tại Airwallex ở Singapore
- + Anh An Vo hiện tại là Software Engineer tại Remitano - Cryptocurrency Exchange ở Thành Phố Hồ Chí Minh và những anh chị trong khóa học System Design của Engineer Pro



Thanks!

Do you have any questions?

