

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN
KHOA THƯ VIỆN – THÔNG TIN HỌC



ĐỒ ÁN TỐT NGHIỆP
WEBSITE PHÂN TÍCH THỊ TRƯỜNG
BẤT ĐỘNG SẢN VIỆT NAM

Giảng viên hướng dẫn: ThS. Trần Đình Anh Huy

Nhóm sinh viên thực hiện:

1. Hoàng Xuân Quốc
2. Võ Châu Trường Thịnh
3. Hứa Gia Tường

Lớp: Quản lý thông tin B

Khóa: 2021 – 2025

Thành phố Hồ Chí Minh, tháng 1 năm 2025

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC XÃ HỘI VÀ NHÂN VĂN
KHOA THƯ VIỆN – THÔNG TIN HỌC**

**ĐỒ ÁN TỐT NGHIỆP
WEBSITE PHÂN TÍCH THỊ TRƯỜNG
BẤT ĐỘNG SẢN VIỆT NAM**

Xác nhận của đơn vị thực tập
(Ký, ghi rõ họ tên)

Xác nhận của GVHD đồ án
(Ký, ghi rõ họ tên)

MỤC LỤC

LỜI CAM ĐOAN	5
LỜI CẢM ƠN	6
DANH MỤC TỪ VIẾT TẮT	7
DANH MỤC BẢNG, BIỂU ĐỒ, HÌNH ẢNH	8
TÓM TẮT ĐỒ ÁN THỰC TẬP TỐT NGHIỆP	9
GIỚI THIỆU	10
1. Lý do chọn đề tài	10
2. Mục đích nghiên cứu	10
3. Tình hình nghiên cứu	11
4. Nhiệm vụ nghiên cứu	12
5. Phương pháp nghiên cứu	12
5.1. Phương pháp thu thập dữ liệu.....	12
5.2. Phương pháp phân tích và tổng hợp	13
5.3. Phương pháp thử nghiệm và đánh giá	13
6. Các kết quả đạt được của đề tài	13
7. Kết cấu của đồ án	14
CHƯƠNG I. THỊ TRƯỜNG BẤT ĐỘNG SẢN VÀ LÝ THUYẾT HỌC MÁY	15
1.1. Khái quát về thị trường bất động sản	15
1.1.1. Các khái niệm liên quan.....	15
1.1.2. Các loại hình bất động sản	16
1.1.3. Vai trò của thị trường bất động sản trong nền kinh tế	17
1.2. Đặc điểm của thị trường bất động sản Việt Nam	19
1.2.1. Thực trạng thị trường bất động sản.....	19
1.2.2. Các yếu tố đặc thù của thị trường bất động sản Việt Nam	20
1.2.3. Xu hướng phát triển hiện nay.....	20
1.3. Các yếu tố ảnh hưởng đến thị trường bất động sản	21
1.3.1. Yếu tố kinh tế.....	21
1.3.2. Yếu tố pháp lý	22
1.3.3. Yếu tố xã hội	22
1.3.4. Yếu tố công nghệ	22
1.4. Ứng dụng công nghệ trong lĩnh vực bất động sản	23
1.4.1. Tầm quan trọng của công nghệ	23
1.4.2. Các công nghệ đang được áp dụng.....	23

1.4.3. Thực trạng ứng dụng công nghệ tại Việt Nam	24
1.5. Nhận định và vấn đề cần nghiên cứu	26
1.5.1. Nhận định về tiềm năng của thị trường bất động sản Việt Nam	26
1.5.2 Hướng giải quyết và mục tiêu nghiên cứu của đề tài	27
1.6 Dữ liệu và lý thuyết về các mô hình học máy	27
1.6.1 Dữ liệu sử dụng	27
1.6.2 Hồi quy và ứng dụng trong bất động sản	28
1.6.3 Lý thuyết các mô hình học máy	29
CHƯƠNG II. CHUẨN BỊ VÀ XÂY DỰNG MÔ HÌNH DỰ BÁO	32
2.1 Chuẩn bị dữ liệu	32
2.1.1 Nguồn dữ liệu.....	32
2.1.2 Thu thập dữ liệu	32
2.2 Lưu trữ dữ liệu	39
2.3 Xử lý dữ liệu	40
2.3.1 Xử lý dữ liệu trong cơ sở dữ liệu để lưu trữ	40
2.3.2 Xử lý dữ liệu bên ngoài cơ sở dữ liệu bằng Python để lưu trữ.....	45
2.4 Chuẩn bị mô hình.....	48
2.4.1 Xử lý dữ liệu đầu vào của mô hình.....	48
2.4.2 Thực hiện mô hình học máy	50
2.4.3 Lựa chọn mô hình.....	50
CHƯƠNG III. TRIỂN KHAI MÔ HÌNH TỐI ƯU VÀ XÂY DỰNG WEBSITE.....	53
3.1 Tối ưu mô hình RandomRegressor	53
3.2 Xây dựng website phân tích bất động sản.....	54
3.2.1 Mục tiêu thiết kế.....	54
3.2.2 Yêu cầu thiết kế Website.....	55
3.2.3 Công nghệ sử dụng.....	55
3.2.4. Phương pháp triển khai và quy trình thực hiện	57
3.2.5 Chức năng website.....	60
KẾT LUẬN	64
DANH MỤC TÀI LIỆU THAM KHẢO	66
PHỤ LỤC	67

LỜI CAM ĐOAN

Chúng em, nhóm sinh viên thực hiện đồ án “Website phân tích thị trường bất động sản Việt Nam”, gồm các thành viên:

- | | |
|-------------------------|------------------|
| 1. Võ Châu Trường Thịnh | MSSV: 2156210140 |
| 2. Hoàng Xuân Quốc | MSSV: 2156210125 |
| 3. Hứa Gia Tường | MSSV: 2156210133 |

Xin cam đoan rằng đồ án này là kết quả của sự nỗ lực, nghiên cứu và làm việc nghiêm túc của toàn bộ nhóm dưới sự hướng dẫn của giảng viên ThS. Trần Đình Anh Huy.

Tất cả các thông tin, số liệu và tài liệu được trình bày trong đồ án đều được thu thập từ các nguồn đáng tin cậy, và các trích dẫn đều được thực hiện theo đúng quy định. Nhóm chúng tôi không sao chép, sử dụng trái phép nội dung từ bất kỳ công trình nghiên cứu nào khác.

Chúng em cam kết đồ án này chưa từng được nộp hoặc sử dụng để đánh giá tại bất kỳ cơ sở giáo dục hay tổ chức nào khác.

Nhóm chúng em xin chịu hoàn toàn trách nhiệm về tính trung thực và chính xác của nội dung trong đồ án này trước nhà trường và pháp luật.

Thành phố Hồ Chí Minh, ngày 17 tháng 01 năm 2025

Sinh viên thực hiện

LỜI CẢM ƠN

Trước tiên, nhóm chúng em xin gửi lời cảm ơn chân thành và sâu sắc đến giảng viên hướng dẫn ThS. Trần Đình Anh Huy, người đã tận tình chỉ bảo, hỗ trợ và hướng dẫn chúng em trong suốt quá trình thực hiện đề án “Website phân tích thị trường bất động sản Việt Nam”. Sự chỉ dẫn của thầy/cô không chỉ giúp chúng em hoàn thành tốt đề án mà còn mang lại cho chúng em những kiến thức, kinh nghiệm quý báu phục vụ cho quá trình học tập và phát triển bản thân sau này.

Chúng em cũng xin trân trọng cảm ơn các thầy cô trong Khoa Thư viện - Thông tin học, Trường Đại học Khoa học Xã hội và Nhân văn đã trang bị cho chúng em những kiến thức nền tảng quan trọng trong suốt quá trình học tập tại trường, tạo điều kiện để nhóm chúng em có thể thực hiện đề án này một cách thuận lợi nhất.

Bên cạnh đó, nhóm chúng em cũng muốn gửi lời cảm ơn đến các cá nhân, tổ chức và đơn vị đã cung cấp thông tin, số liệu, tài liệu liên quan, giúp chúng em có nguồn dữ liệu đáng tin cậy để hoàn thiện nội dung nghiên cứu.

Dù đã cố gắng hết sức, nhưng do hạn chế về thời gian và kinh nghiệm, đề án này có thể vẫn còn tồn tại những thiếu sót. Nhóm chúng em rất mong nhận được ý kiến đóng góp từ các thầy cô để có thể hoàn thiện hơn trong tương lai.

Nhóm chúng em xin chân thành cảm ơn!

DANH MỤC TỪ VIẾT TẮT

STT	Từ viết tắt	Từ viết đầy đủ
1	MAPE	Mean Absolute Percentage Error
2	ERD	Entity-Relationship Diagram(Biểu đồ thực thể mối quan hệ)
3	AI	Artificial Intelligence (Trí tuệ nhân tạo)
4	ML	Machine Learning (Học máy)
5	DF	DataFrame (Cấu trúc dữ liệu)

DANH MỤC BẢNG, BIỂU ĐỒ, HÌNH ẢNH

Bảng 1. 1 Các giai đoạn phát triển của thị trường bất động sản Việt Nam.....	19
Bảng 3. 1 Các ngôn ngữ sử dụng trong website.....	55
Bảng 3. 2 Các thư viện sử dụng trong website.....	56
Hình 2. 1 Cài đặt thư viện thu thập dữ liệu	33
Hình 2. 2 Thiết lập việc chèn dữ liệu thu thập được vào cơ sở dữ liệu.....	35
Hình 2. 3 Khởi tạo hàm thu thập thông tin chi tiết.....	36
Hình 2. 4 Duyệt các trang con trong danh sách.....	37
Hình 2. 5 Tiến hành thu thập dữ liệu	38
Hình 2. 6 ERD cơ sở dữ liệu	40
Hình 2. 7 Dữ liệu từ trang web	40
Hình 2. 8 Kết quả.....	43
Hình 2. 9 Cài đặt thư viện xử lý dữ liệu	45
Hình 2. 10 Kết nối cơ sở dữ liệu.....	46
Hình 2. 11 Thêm cột tọa độ vào csdl.....	46
Hình 2. 12 Trực quan hóa bất động sản.....	47
Hình 2. 13 Kết quả thu được khi đánh tọa độ.....	47
Hình 2. 14 Cài đặt các thư viện cần thiết	48
Hình 2. 15 Dữ liệu Địa phương	48
Hình 2. 16 File dữ liệu gốc	49
Hình 2. 17 Tối giản dữ liệu.....	49
Hình 2. 18 Kết hợp dữ liệu	49
Hình 2. 19 Tiến hành xử lý.....	50
Hình 2. 20 MAPE RandomForestRegressor	50
Hình 2. 21 MAPE LinearRegression.....	50
Hình 3. 1 Chương 3 – import thư viện	53
Hình 3. 2 Chương 3 – chuyển đổi logarit dữ liệu.....	53
Hình 3. 3 Chương 3 – Chuyển đổi dữ liệu	53
Hình 3. 4 Chương 3 – tách dữ liệu	53
Hình 3. 5 Chương 3 – Xử lý dữ liệu.....	53
Hình 3. 6 Chương 3 – Định nghĩa tham số và mô hình.....	53
Hình 3. 7 Chương 3 – tìm kiếm tham số tối ưu.....	54
Hình 3. 8 Chương 3 – huấn luyện.....	54
Hình 3. 9 Chương 3 – đánh giá mô hình	54
Hình 3. 10 Cấu trúc thư mục dự án	58
Hình 3. 11 Các hàm trong module view.....	59
Hình 3. 12 Các hàm trong module model.....	60

TÓM TẮT ĐỒ ÁN THỰC TẬP TỐT NGHIỆP

Đồ án “Website phân tích thị trường bất động sản Việt Nam” được thực hiện nhằm xây dựng một nền tảng trực tuyến cung cấp thông tin và công cụ hỗ trợ phân tích thị trường bất động sản tại Việt Nam. Đây là một lĩnh vực có vai trò quan trọng trong nền kinh tế nhưng hiện nay vẫn tồn tại nhiều hạn chế như thiếu minh bạch thông tin, khó khăn trong việc tiếp cận dữ liệu đáng tin cậy, và sự khan hiếm các công cụ phân tích xu hướng giá.

Mục tiêu của đồ án là thiết kế và triển khai một website với các tính năng chính bao gồm: hiển thị giá bất động sản theo khu vực, dự đoán được giá nhà và so sánh bất động sản theo nhu cầu người dùng. Bên cạnh đó, website còn tích hợp các công cụ phân tích chuyên sâu để hỗ trợ người dùng ra quyết định trong các giao dịch bất động sản.

Quá trình thực hiện đồ án bao gồm:

- Thu thập và phân tích dữ liệu.
- Xây dựng mô hình học máy dự đoán giá bất động sản
- Thiết kế giao diện và hệ thống chức năng của website.
- Xây dựng và thử nghiệm website.

Kết quả của đồ án là một website hoạt động ổn định, cung cấp thông tin minh bạch, trực quan và dễ sử dụng. Ngoài ra, đồ án cũng đánh giá hiệu quả của hệ thống, đồng thời đề xuất các hướng phát triển trong tương lai để dự báo giá bất động sản hoặc phát triển ứng dụng di động.

Với đồ án này, nhóm thực hiện không chỉ ứng dụng các kiến thức đã học vào thực tế mà còn đóng góp giải pháp công nghệ hữu ích cho một lĩnh vực có tiềm năng phát triển cao tại Việt Nam.

GIỚI THIỆU

1. Lý do chọn đề tài

Thị trường bất động sản là một trong những lĩnh vực quan trọng, đóng vai trò thúc đẩy sự phát triển kinh tế của Việt Nam. Tuy nhiên, thực trạng hiện nay cho thấy thị trường bất động sản Việt Nam vẫn còn tồn tại nhiều hạn chế, như thiếu sự minh bạch về thông tin, khó khăn trong việc tiếp cận dữ liệu chính xác, và sự khan hiếm các công cụ hỗ trợ phân tích chuyên sâu. Điều này không chỉ gây khó khăn cho người mua, người bán mà còn ảnh hưởng đến các nhà đầu tư trong việc đưa ra quyết định chính xác.

Sự phát triển của công nghệ thông tin, đặc biệt là các nền tảng trực tuyến và công cụ phân tích dữ liệu, đã mang lại những giải pháp hiệu quả trong nhiều lĩnh vực, bao gồm bất động sản. Các website hỗ trợ phân tích thị trường bất động sản không chỉ giúp người dùng tiếp cận thông tin dễ dàng hơn mà còn giúp họ hiểu rõ hơn về xu hướng giá cả, tình trạng cung cầu và các yếu tố ảnh hưởng khác. Tuy nhiên, tại Việt Nam, số lượng các nền tảng như vậy vẫn còn hạn chế, hoặc chưa đáp ứng được đầy đủ nhu cầu của người dùng.

Nhận thấy những thách thức và tiềm năng phát triển của lĩnh vực này, nhóm chúng tôi quyết định thực hiện đề tài “Website phân tích thị trường bất động sản Việt Nam”. Đề tài hướng đến việc xây dựng một nền tảng trực tuyến không chỉ cung cấp thông tin chính xác, đáng tin cậy mà còn tích hợp các công cụ phân tích chuyên sâu nhằm hỗ trợ người dùng ra quyết định.

Với đề tài này, nhóm không chỉ mong muốn ứng dụng những kiến thức đã học vào thực tế mà còn đóng góp một giải pháp công nghệ hữu ích, hỗ trợ sự phát triển minh bạch và bền vững của thị trường bất động sản Việt Nam.

2. Mục đích nghiên cứu

Cung cấp nền tảng phân tích dữ liệu đáng tin cậy về bất động sản.

Hỗ trợ người dùng ra quyết định thông qua thông tin thị trường chính xác.

Thứ nhất, đề tài được thực hiện nhằm cung cấp thông tin minh bạch và đáng tin cậy về thị trường bất động sản Việt Nam. Thông qua việc xây dựng một nền tảng trực tuyến tổng hợp dữ liệu từ các nguồn uy tín, website sẽ giúp người dùng dễ dàng tiếp cận thông tin về giá cả, xu hướng và các khu vực tiềm năng, giải quyết được thực trạng thiếu minh bạch trên thị trường hiện nay.

Thứ hai, mục tiêu quan trọng của đề tài là hỗ trợ người dùng đưa ra các quyết định chính xác trong các giao dịch bất động sản. Bằng việc phát triển các công cụ phân tích và trực quan hóa dữ liệu, người dùng có thể so sánh giá, phân tích xu hướng và dự đoán biến

động thị trường, từ đó giảm thiểu rủi ro và tối ưu hóa lợi ích trong quá trình mua, bán hoặc đầu tư.

Thứ ba, đề tài hướng đến việc ứng dụng các công nghệ hiện đại vào lĩnh vực bất động sản. Những công nghệ như trí tuệ nhân tạo (AI) và học máy (Machine Learning) được kỳ vọng sẽ giúp xử lý dữ liệu hiệu quả, cung cấp phân tích chuyên sâu, và nâng cao trải nghiệm người dùng trên website.

Thứ tư, đề tài nhằm đóng góp vào sự phát triển minh bạch và bền vững của thị trường bất động sản Việt Nam. Bằng cách cung cấp một công cụ hỗ trợ tin cậy cho các bên tham gia thị trường, đề tài kỳ vọng sẽ thúc đẩy sự cạnh tranh công bằng và hiệu quả trong lĩnh vực này.

Cuối cùng, mục tiêu của đề tài là phát triển một giải pháp thực tiễn và khả thi. Sản phẩm cuối cùng sẽ là một website với giao diện thân thiện, dễ sử dụng, đáp ứng nhu cầu thực tế của người dùng và có khả năng mở rộng, nâng cấp để phù hợp với sự phát triển trong tương lai.

3. Tình hình nghiên cứu

Hiện nay, thị trường bất động sản Việt Nam đang trở thành một lĩnh vực được nhiều cá nhân và tổ chức quan tâm, đặc biệt là trong bối cảnh nhu cầu sở hữu, đầu tư và giao dịch bất động sản ngày càng tăng cao. Tuy nhiên, việc nghiên cứu và xây dựng các nền tảng công nghệ hỗ trợ phân tích thị trường bất động sản ở Việt Nam vẫn còn nhiều hạn chế.

Trên thế giới, các website phân tích thị trường bất động sản như Zillow (Hoa Kỳ), Realtor (Hoa Kỳ) hay Rightmove (Anh) đã trở nên phổ biến, cung cấp thông tin minh bạch, công cụ tìm kiếm và phân tích dữ liệu chuyên sâu. Những nền tảng này không chỉ giúp người dùng dễ dàng tiếp cận thông tin mà còn hỗ trợ họ đưa ra các quyết định đầu tư chính xác. Tuy nhiên, tại Việt Nam, những nền tảng tương tự chưa được phát triển đầy đủ hoặc vẫn đang ở giai đoạn sơ khai.

Một số website bất động sản trong nước như Batdongsan.com.vn, Alonhadat.com.vn hay Homedy.com chủ yếu tập trung vào việc đăng tải thông tin rao bán, cho thuê mà chưa chú trọng đến việc phân tích dữ liệu hay hỗ trợ người dùng ra quyết định. Các công cụ như phân tích xu hướng giá, dự báo thị trường, hay so sánh giá theo khu vực vẫn chưa được triển khai một cách toàn diện và hiệu quả.

Ngoài ra, các nghiên cứu học thuật về ứng dụng công nghệ trong lĩnh vực bất động sản tại Việt Nam còn hạn chế. Các nghiên cứu chủ yếu tập trung vào việc phân tích tình hình thị trường mà chưa khai thác sâu vào khía cạnh ứng dụng công nghệ như học máy (Machine Learning) để nâng cao hiệu quả phân tích và dự đoán.

Nhận thấy sự thiếu hụt này, đề tài “Website phân tích thị trường bất động sản Việt Nam” được thực hiện nhằm lấp đầy khoảng trống giữa nhu cầu thực tế và các giải pháp hiện có. Đề tài không chỉ kế thừa các nghiên cứu trước đây mà còn bổ sung các công cụ, chức năng mới nhằm đáp ứng tốt hơn nhu cầu của người dùng.

4. Nhiệm vụ nghiên cứu

Nhóm đã tiến hành phân tích các nguồn thông tin liên quan đến thị trường bất động sản nhằm xác định nhu cầu và hành vi đối với một nền tảng hỗ trợ phân tích trong thị trường bất động sản. Qua quá trình thu thập thông tin dữ liệu từ các nguồn và các bài nghiên cứu cũng như chuyên gia phân tích, nhóm đã phân loại và xác định các yếu tố quan trọng cần tập trung, bao gồm giao diện thân thiện, dữ liệu đáng tin cậy, và các công cụ phân tích mạnh mẽ.

Bên cạnh đó, nhóm đã thiết kế giao diện người dùng với mục tiêu đảm bảo sự trực quan và dễ sử dụng. Giao diện được xây dựng với các tính năng hỗ trợ người dùng như tìm kiếm, lọc, và so sánh bất động sản, cùng với việc tích hợp các biểu đồ, bản đồ trực quan để cung cấp thông tin một cách rõ ràng. Trong giai đoạn tiếp theo, nhóm sẽ tiếp tục tối ưu hóa giao diện dựa trên phản hồi từ người dùng thử nghiệm.

Một nhiệm vụ quan trọng khác mà nhóm đã thực hiện là nghiên cứu và phát triển các công cụ phân tích và dự đoán giá trị bất động sản. Nhóm đã bước đầu triển khai các thuật toán để phân tích dữ liệu, cung cấp thông tin về giá cả, xu hướng thị trường, và các yếu tố ảnh hưởng khác. Trong giai đoạn tới, nhóm sẽ hoàn thiện công cụ dự đoán giá trị bất động sản dựa trên dữ liệu lịch sử và các yếu tố thời gian thực như vị trí, diện tích, và tiện ích xung quanh.

Song song với đó, nhóm cũng đã thu thập và xử lý dữ liệu từ các nguồn đáng tin cậy như API của các sàn giao dịch bất động sản và các báo cáo thị trường. Dữ liệu đã được chuẩn hóa để đảm bảo tính nhất quán và loại bỏ các thông tin không chính xác. Trong thời gian tới, nhóm sẽ tiếp tục cập nhật và mở rộng cơ sở dữ liệu để hỗ trợ các chức năng phân tích và dự đoán.

Cuối cùng, nhóm đã thực hiện các đợt kiểm thử ban đầu để đánh giá hiệu suất hoạt động và mức độ thân thiện của website. Các bước tiếp theo sẽ tập trung vào việc hoàn thiện tính năng, mở rộng khả năng phân tích, và lập báo cáo tổng hợp để đánh giá toàn diện hiệu quả của dự án.

5. Phương pháp nghiên cứu

5.1. Phương pháp thu thập dữ liệu

Dữ liệu được thu thập từ trang web batdongsan.com và thông tin công khai từ các sàn giao dịch bất động sản uy tín tại Việt Nam. Đồng thời, nhóm tiến hành lấy dữ liệu từ

các đối tượng liên quan, bao gồm người mua nhà, nhà đầu tư và chuyên gia trong ngành. Những thông tin này giúp nhóm hiểu rõ hơn về nhu cầu của người dùng và xác định các tính năng cần thiết cho website.

5.2. Phương pháp phân tích và tổng hợp

Thông qua việc phân tích các yếu tố ảnh hưởng đến giá cả, các đặc điểm nổi bật của thị trường bất động sản, nhóm đưa ra các giải pháp phù hợp để thiết kế và triển khai website. Kết quả phân tích cũng được tổng hợp lại để làm cơ sở xây dựng các tính năng như bản đồ giá, phân tích xu hướng và dự đoán biến động thị trường.

5.3. Phương pháp thử nghiệm và đánh giá

Nhóm sử dụng Python và thư viện máy học để xây dựng mô hình dự đoán giá bất động sản, cụ thể là áp dụng các thuật toán phân loại, hồi quy. Trước tiên, dữ liệu được chia thành hai tập: tập huấn luyện và tập kiểm tra với tỷ lệ 80:20 để đảm bảo mô hình được đánh giá công bằng. Các bước tiền xử lý dữ liệu bao gồm xử lý giá trị khuyết thiếu, mã hóa các biến phân loại, và chuẩn hóa dữ liệu để tăng hiệu quả của mô hình.

Trong quá trình thử nghiệm, nhóm thực hiện tinh chỉnh các siêu tham số (hyperparameters) của mô hình, bao gồm số lượng cây ($n_estimators$), độ sâu tối đa của cây (max_depth), và các tham số khác, bằng cách sử dụng GridSearchCV để tìm kiếm cấu hình tốt nhất. Sau khi huấn luyện, mô hình được đánh giá trên tập kiểm tra dựa trên các chỉ số như Mean Absolute Error (MAE), Mean Squared Error (MSE), và R-squared (R^2).

Để đảm bảo kết quả đánh giá khách quan, chúng tôi thực hiện quá trình kiểm tra chéo (cross-validation) với $k = 5$. Điều này giúp giảm thiểu ảnh hưởng của dữ liệu ngẫu nhiên và cung cấp một cái nhìn toàn diện về hiệu suất của mô hình trên toàn bộ tập dữ liệu.

6. Các kết quả đạt được của đề tài

Đề tài “Website phân tích thị trường bất động sản Việt Nam” đã đạt được những kết quả quan trọng, góp phần giải quyết các vấn đề thực tiễn trong lĩnh vực bất động sản. Trước tiên, nhóm đã xây dựng thành công mô hình dự đoán giá bất động sản dựa trên tập dữ liệu thu thập được. Thứ hai, nhóm xây dựng được một nền tảng trực tuyến với giao diện thân thiện, trực quan, giúp người dùng dễ dàng tra cứu và tìm hiểu thông tin về thị trường bất động sản. Website tích hợp các chức năng thiết yếu như hiển thị giá bất động sản theo khu vực, phân tích và hỗ trợ so sánh bất động sản dựa trên các tiêu chí cụ thể, đáp ứng nhu cầu đa dạng của người dùng.

Ngoài ra, website còn áp dụng công nghệ phân tích dữ liệu hiện đại để xử lý và trực quan hóa thông tin. Nhờ vào các thuật toán và công cụ phân tích, hệ thống có khả năng cung cấp thông tin chi tiết và đáng tin cậy, hỗ trợ người dùng ra quyết định đầu tư hoặc giao dịch chính xác hơn. Đồng thời, nhóm cũng tiến hành thử nghiệm hệ thống trong môi

trường thực tế, thu thập phản hồi từ người dùng để đánh giá mức độ hiệu quả và tính thân thiện của website. Các ý kiến đóng góp đã được áp dụng để cải thiện và tối ưu hóa hệ thống, đảm bảo sản phẩm đạt chất lượng cao nhất khi hoàn thiện.

Kết quả của đề tài không chỉ khẳng định tính khả thi của việc ứng dụng công nghệ trong phân tích thị trường bất động sản mà còn mở ra cơ hội phát triển các tính năng nâng cao như dự đoán giá trị bất động sản bằng trí tuệ nhân tạo (AI) trong tương lai.

7. Kết cấu của đồ án

Bên cạnh phần mở đầu, kết luận và phụ lục, phần nội dung chính của đề tài “Website phân tích thị trường bất động sản Việt Nam” được trình bày trong 3 chương như sau:

Chương 1 tập trung nghiên cứu cơ sở lý luận liên quan đến thị trường bất động sản, bao gồm các khái niệm cơ bản, vai trò và đặc điểm của thị trường bất động sản tại Việt Nam. Đồng thời, chương này phân tích các yếu tố ảnh hưởng đến giá trị bất động sản, như yếu tố pháp lý, kinh tế, xã hội, và môi trường. Ngoài ra, chương 1 cũng đề cập đến thực trạng sử dụng công nghệ trong lĩnh vực bất động sản, làm nền tảng cho việc xây dựng hệ thống phân tích hiện đại. Cuối cùng, nội dung chương này còn trình đến cơ sở lý thuyết về các mô hình học máy sẽ sử dụng trong đề tài.

Chương 2 với nội dung chính là xây dựng mô hình học máy dựa trên dữ liệu thu thập được trước đó. Trong đó, đầu tiên là mô tả cách thu thập dữ liệu từ nguồn batdongsan.com và các nguồn uy tín khác. Tiếp theo, chương này sẽ phân tích các bước để xử lý và làm sạch dữ liệu thu thập, chuẩn bị cho việc thử nghiệm cũng như triển khai mô hình về sau. Cuối cùng, trong chương này sẽ đề cập đến một số thống kê mô tả cơ bản để nắm rõ được các tính chất, đặc điểm của dữ liệu, độ tương thích với các mô hình học máy để chuẩn bị đưa vào thử nghiệm.

Chương 3 tập trung vào việc thử nghiệm mô hình học máy để dự đoán giá bất động sản và tiến hành lựa chọn mô hình phù hợp nhất, sau đó triển khai xây dựng website phân tích dự đoán bất động sản. Đầu tiên là thử nghiệm một số mô hình trong việc dự đoán giá, sau đó dựa vào các chỉ số đánh giá để lựa chọn mô hình tốt nhất đưa vào triển khai thực tế website. Kết quả thử nghiệm mô hình được trình bày rõ ràng trong đề tài. Sau khi có được mô hình tốt nhất, chương này sẽ tiếp tục mô tả về quy trình thiết kế website bao gồm việc phân tích yêu cầu người dùng, các tính năng cốt lõi, lựa chọn công nghệ và công cụ thực hiện phù hợp.

Bộ cục báo cáo được thiết kế để đảm bảo sự mạch lạc, đầy đủ, và toàn diện, đáp ứng yêu cầu nghiên cứu khoa học cũng như nhu cầu thực tiễn của đề tài.

CHƯƠNG I. THỊ TRƯỜNG BẤT ĐỘNG SẢN VÀ LÝ THUYẾT HỌC MÁY

1.1. Khái quát về thị trường bất động sản

1.1.1. Các khái niệm liên quan

1.1.1.1. Bất động sản

Điều 105 Bộ luật Dân sự có đề cập đến bất động sản là một trong những tài sản bên cạnh động sản và có thể là tài sản hiện có và tài sản hình thành trong tương lai. Điều 107 Bộ luật Dân sự năm 2015 đã quy định bất động sản gồm: Đất đai; nhà, công trình xây dựng gắn liền với đất đai và tài sản khác gắn liền với đất đai, nhà, công trình xây dựng và theo quy định của pháp luật. Theo Luật Đất đai Việt Nam, bất động sản không chỉ bao gồm quyền sở hữu đất mà còn bao gồm quyền sở hữu đối với các công trình xây dựng, nhà ở hoặc tài sản gắn liền khác thuộc sở hữu hợp pháp của cá nhân, tổ chức.

Nhìn chung, bất động sản (Real Estate) là thuật ngữ dùng để chỉ các tài sản cố định, không thể di dời, bao gồm đất đai và tất cả các tài sản gắn liền vĩnh viễn với mảnh đất đó như nhà cửa, công trình xây dựng, cây cối hoặc các tài nguyên tự nhiên khác. Trong lĩnh vực kinh tế và pháp lý, bất động sản được xem là một loại tài sản quan trọng, có giá trị lớn và được giao dịch phổ biến trên thị trường.

1.1.1.2. Thị trường bất động sản

Cambridge (n.d.) định nghĩa thị trường bất động sản chỉ tất cả những hoạt động có liên quan đến việc mua bán đất đai và các tòa nhà. Ở Việt Nam hiện nay, pháp luật chưa có quy định nào giải thích khái niệm "Thị trường bất động sản" là gì. Tuy nhiên, có thể hiểu thị trường bất động sản là nơi thực hiện giao dịch chuyển nhượng, mua bán, trao đổi và các quyền về bất động sản.

Thị trường bất động sản là một bộ phận quan trọng của nền kinh tế, nơi diễn ra các hoạt động giao dịch liên quan đến bất động sản, bao gồm mua bán, cho thuê, chuyển nhượng, và đầu tư. Đây là nơi gặp gỡ giữa cung và cầu, nơi mà các nhà đầu tư, người bán, người mua, và các tổ chức trung gian thực hiện các hoạt động thương mại nhằm tối ưu hóa lợi ích kinh tế.

Cụ thể, thị trường bất động sản không chỉ bao gồm các giao dịch về đất đai và tài sản gắn liền trên đất mà còn mở rộng đến các hoạt động tài chính liên quan, như vay vốn thế chấp, đầu tư vào quỹ bất động sản (REITs), và các hoạt động quản lý tài sản. Tính chất đa dạng và phức tạp của thị trường này đòi hỏi sự tham gia của nhiều bên liên quan và chịu sự chi phối mạnh mẽ bởi các yếu tố kinh tế, xã hội và pháp luật.

1.1.1.3. Website

Theo Cambridge (n.d.) Website là một tập hợp các trang thông tin trên internet về một chủ đề cụ thể, được xuất bản bởi một cá nhân hoặc tổ chức. Một website thường bao

gồm các nội dung như văn bản, hình ảnh, video, âm thanh, và các tệp tin khác, được tổ chức dưới dạng các trang web liên kết với nhau thông qua các đường dẫn (hyperlinks).

Website là một phương tiện truyền tải thông tin và tương tác phổ biến trong thời đại số, phục vụ nhiều mục đích khác nhau như giải trí, giáo dục, kinh doanh, truyền thông, và dịch vụ công cộng.

1.1.2. Các loại hình bất động sản

Theo Điều 107 Bộ luật Dân sự năm 2015, bất động sản được phân chia thành nhiều loại hình khác nhau, tùy thuộc vào mục đích sử dụng và đặc điểm của từng loại tài sản, cụ thể như sau:

1.1.2.1. Đất đai

Đất đai là loại hình bất động sản cơ bản và quan trọng nhất, được coi là nền tảng cho mọi hoạt động xây dựng và phát triển. Đất đai bao gồm các loại đất được Nhà nước giao quyền sử dụng hoặc sở hữu. Theo Luật Đất đai 2013, đất đai được chia thành nhiều loại dựa trên mục đích sử dụng, bao gồm:

- Đất nông nghiệp: Phục vụ các hoạt động sản xuất nông nghiệp, trồng trọt, chăn nuôi và nuôi trồng thủy sản.
- Đất phi nông nghiệp: Bao gồm đất xây dựng nhà ở, đất sử dụng vào mục đích thương mại, dịch vụ, đất khu công nghiệp, đất xây dựng công trình công cộng (như trường học, bệnh viện).
- Đất chưa sử dụng: Là những khu vực đất chưa được đưa vào quy hoạch sử dụng, thường là tiềm năng cho các dự án phát triển trong tương lai.

Đất đai là tài sản không thể di dời, giá trị của nó phụ thuộc nhiều vào vị trí, quy hoạch sử dụng đất và các yếu tố kinh tế - xã hội của khu vực.

1.1.2.2. Nhà ở và công trình xây dựng gắn liền với đất đai

Loại hình bất động sản này bao gồm các tài sản đã được xây dựng và gắn liền với đất, không thể tách rời mà không làm thay đổi giá trị hoặc cấu trúc của chúng.

Nhà ở là công trình xây dựng được sử dụng làm nơi cư trú cho cá nhân và hộ gia đình. Các loại hình nhà ở phổ biến gồm: nhà riêng lẻ (biệt thự, nhà phố, nhà cấp 4,...), căn hộ chung cư (các tòa nhà cao tầng được thiết kế để ở, thường dành cho các gia đình tại đô thị).

Công trình xây dựng khác bao gồm các công trình phục vụ mục đích kinh doanh, sản xuất hoặc công cộng như: nhà xưởng, kho bãi, trung tâm thương mại, khách sạn, văn phòng cho thuê, các công trình giao thông như cầu, đường, cảng biển, sân bay.

1.1.2.3. Tài sản khác gắn liền với đất đai, nhà ở và công trình xây dựng

Tài sản khác gắn liền với đất đai, nhà ở và công trình xây dựng bao gồm những yếu tố gắn liền hoặc cố định với đất, nhà và công trình xây dựng, không thể di chuyển mà không làm ảnh hưởng đến tài sản chính.

- Cây lâu năm: Các loại cây trồng được xem là tài sản cố định do có chu kỳ khai thác dài và giá trị kinh tế cao, chẳng hạn như cây công nghiệp (cao su, cà phê), cây ăn quả lâu năm (xoài, mít), hoặc cây lấy gỗ.
- Hệ thống hạ tầng kỹ thuật: Bao gồm các công trình phụ trợ như hệ thống cấp thoát nước, hệ thống điện, hệ thống viễn thông gắn liền với các khu đất hoặc công trình xây dựng.
- Công trình ngầm: Như tầng hầm, bãi đỗ xe ngầm, hoặc các hệ thống kỹ thuật dưới lòng đất.

Các tài sản này thường được liệt kê và định giá cùng với bất động sản chính trong các giao dịch và quy hoạch.

1.1.2.4. Các tài sản khác theo quy định của pháp luật

Bên cạnh những loại bất động sản chính nêu trên, pháp luật còn quy định các tài sản cụ thể khác là bất động sản trong từng trường hợp. Điều này giúp đảm bảo việc điều chỉnh các loại tài sản mới phù hợp với sự phát triển của kinh tế và xã hội.

1.1.3. Vai trò của thị trường bất động sản trong nền kinh tế

Thị trường bất động sản là một trong những thành phần quan trọng, góp phần thúc đẩy sự phát triển kinh tế - xã hội của mỗi quốc gia. Với vai trò là động lực tăng trưởng kinh tế, nguồn thu ngân sách lớn, và nền tảng phát triển hạ tầng, thị trường này tác động đến hầu hết các lĩnh vực của nền kinh tế. Vai trò này được thể hiện rõ qua các khía cạnh sau:

1.1.1.1. Thúc đẩy tăng trưởng kinh tế

Thị trường bất động sản đóng góp trực tiếp vào tăng trưởng GDP của quốc gia thông qua các hoạt động xây dựng, giao dịch mua bán, phát triển dự án và đầu tư cơ sở hạ tầng. Tại Việt Nam, theo Tổng cục Thống kê, lĩnh vực bất động sản chiếm từ 8-10% GDP hàng năm. Đây là con số đáng kể, khẳng định vai trò của bất động sản như một trong những ngành kinh tế chủ chốt.

Thị trường bất động sản còn có tác động lan tỏa mạnh mẽ đến nhiều ngành công nghiệp liên quan. Chẳng hạn, lĩnh vực xây dựng chiếm khoảng 6,5% GDP vào năm 2022, được thúc đẩy mạnh mẽ nhờ nhu cầu phát triển các dự án bất động sản. Các ngành như sản xuất vật liệu xây dựng (xi măng, thép), nội thất, và dịch vụ tài chính (tín dụng bất động sản) cũng hưởng lợi trực tiếp từ sự phát triển của thị trường.

Bất động sản còn tạo ra một lượng lớn việc làm trực tiếp và gián tiếp. Theo Hiệp hội Bất động sản Việt Nam, lĩnh vực này tạo ra hơn 4 triệu việc làm mỗi năm. Các công việc phổ biến bao gồm xây dựng, thiết kế kiến trúc, môi giới bất động sản, quản lý dự án, và các

dịch vụ liên quan. Thu nhập ổn định từ các việc làm này không chỉ cải thiện chất lượng cuộc sống mà còn tăng cường sự ổn định của nền kinh tế.

1.1.1.2. Thu hút đầu tư và tăng cường nguồn lực tài chính

Thị trường bất động sản là kênh đầu tư hấp dẫn, thu hút dòng vốn lớn từ cả nhà đầu tư trong và ngoài nước. Theo báo cáo của Bộ Kế hoạch và Đầu tư, lĩnh vực bất động sản đứng thứ hai về thu hút vốn đầu tư trực tiếp nước ngoài (FDI) tại Việt Nam, với hơn 4,45 tỷ USD vốn đăng ký vào năm 2022, chiếm 18,7% tổng vốn FDI. Điều này không chỉ cho thấy tiềm năng của thị trường mà còn khẳng định vị thế của bất động sản trong việc thu hút các nguồn lực tài chính từ quốc tế.

Ngoài ra, các hoạt động giao dịch bất động sản còn đóng góp nguồn thu lớn cho ngân sách nhà nước. Các khoản thu từ thuế thu nhập doanh nghiệp, thuế thu nhập cá nhân từ chuyển nhượng bất động sản, và các loại phí liên quan như phí trước bạ đã mang lại hàng trăm nghìn tỷ đồng mỗi năm. Chẳng hạn, trong năm 2022, thu ngân sách từ đất đai và bất động sản đạt hơn 130.000 tỷ đồng, chiếm gần 15% tổng thu ngân sách quốc gia.

Những nguồn thu này không chỉ hỗ trợ chi tiêu công mà còn tạo điều kiện phát triển các dự án cơ sở hạ tầng, xây dựng đô thị và các chương trình an sinh xã hội.

1.1.1.3. Góp phần phát triển cơ sở hạ tầng và đô thị hóa

Sự phát triển của thị trường bất động sản luôn đi đôi với việc đầu tư vào cơ sở hạ tầng và thúc đẩy quá trình đô thị hóa. Các dự án bất động sản, từ nhà ở, khu thương mại đến khu công nghiệp, đều đòi hỏi sự phát triển đồng bộ về hệ thống giao thông, cấp thoát nước, và các tiện ích công cộng khác.

Tại Việt Nam, các khu đô thị lớn như Vinhomes Grand Park (TP. Hồ Chí Minh) hay Ecopark (Hưng Yên) không chỉ đáp ứng nhu cầu nhà ở mà còn góp phần thay đổi bộ mặt hạ tầng và cảnh quan đô thị của khu vực. Đồng thời, sự xuất hiện của các dự án này còn tạo ra một làn sóng di cư từ nông thôn ra đô thị, thúc đẩy tốc độ đô thị hóa của quốc gia. Theo báo cáo của Ngân hàng Thế giới, tỷ lệ đô thị hóa tại Việt Nam đạt khoảng 40% vào năm 2022, và dự kiến tăng lên 45% vào năm 2030, trong đó thị trường bất động sản đóng vai trò quan trọng.

Ngoài ra, bất động sản còn thúc đẩy sự phát triển của các hạ tầng giao thông lớn như cao tốc, sân bay, và cảng biển. Điển hình là các dự án bất động sản ven đô hoặc khu nghỉ dưỡng thường kích thích đầu tư vào các tuyến đường cao tốc và các cầu nối giao thông quan trọng, tạo điều kiện thuận lợi cho giao thương và phát triển kinh tế khu vực.

Thị trường bất động sản không chỉ đơn thuần là một lĩnh vực kinh doanh mà còn là động lực quan trọng thúc đẩy sự phát triển kinh tế, xã hội. Với vai trò góp phần tăng trưởng GDP, tạo việc làm, thu hút đầu tư, và cải thiện cơ sở hạ tầng, bất động sản ngày càng khẳng định vị thế quan trọng trong nền kinh tế Việt Nam. Tuy nhiên, để phát triển bền vững, thị trường này cần sự quản lý chặt chẽ, minh bạch thông tin và những chính sách hỗ trợ hiệu quả từ nhà nước.

1.2. Đặc điểm của thị trường bất động sản Việt Nam

1.2.1. Thực trạng thị trường bất động sản

1.2.1.1. Tổng quan tình hình cung và cầu trong thị trường bất động sản Việt Nam

Thị trường bất động sản Việt Nam đang ở giai đoạn phát triển mạnh, với sự gia tăng đáng kể về nhu cầu nhà ở, bất động sản thương mại, và công nghiệp. Theo báo cáo của Bộ Xây dựng, dân số đô thị tại Việt Nam chiếm khoảng 40% tổng dân số cả nước (năm 2022) và dự kiến sẽ tăng lên khoảng 45% vào năm 2030. Điều này dẫn đến nhu cầu nhà ở tại các khu vực đô thị lớn như Hà Nội, TP. Hồ Chí Minh, Đà Nẵng ngày càng tăng cao.

Về nguồn cung, số lượng dự án bất động sản liên tục gia tăng, đặc biệt trong các phân khúc cao cấp và trung cấp. Tuy nhiên, nguồn cung nhà ở giá rẻ vẫn còn hạn chế, chưa đáp ứng đủ nhu cầu của người lao động và các gia đình có thu nhập trung bình. Thực tế cho thấy, hơn 70% nhu cầu nhà ở tại các đô thị lớn thuộc phân khúc nhà ở giá rẻ, nhưng nguồn cung chỉ đáp ứng được khoảng 20% nhu cầu này. Đây là một trong những thách thức lớn của thị trường, đòi hỏi sự điều chỉnh từ phía các nhà phát triển dự án và chính sách hỗ trợ của nhà nước.

1.2.1.2. Các giai đoạn phát triển nổi bật của thị trường bất động sản Việt Nam

Bảng 1. 1 Các giai đoạn phát triển của thị trường bất động sản Việt Nam.

Giai đoạn	Thực trạng bất động sản trong giai đoạn này
1986 - 2000	Đây là thời kỳ khởi đầu khi Việt Nam thực hiện chính sách Đổi mới, mở cửa nền kinh tế. Thị trường bất động sản bắt đầu manh nha hình thành nhưng chủ yếu mang tính tự phát, chưa có sự quản lý chặt chẽ.
2001 - 2010	Thị trường bước vào giai đoạn phát triển bùng nổ, với sự ra đời của nhiều dự án lớn và sự tham gia mạnh mẽ của các nhà đầu tư. Tuy nhiên, bong bóng bất động sản năm 2008 đã gây ra khủng hoảng nghiêm trọng, khiến giá trị thị trường giảm mạnh và ảnh hưởng tiêu cực đến toàn bộ nền kinh tế.
2011 - nay	Đây là giai đoạn thị trường phục hồi và phát triển bền vững hơn nhờ các chính sách điều chỉnh từ Chính phủ, như thắt chặt tín dụng bất động sản và kiểm soát chặt chẽ quy hoạch đất đai. Đồng thời, giai đoạn này cũng chứng kiến sự xuất hiện của các phân khúc mới như bất động sản nghỉ dưỡng, bất động sản công nghiệp và nhà ở thông minh.

1.2.2. Các yếu tố đặc thù của thị trường bất động sản Việt Nam

1.2.2.1. Tính khu vực hóa (Đặc trưng từng vùng, khu vực)

Thị trường bất động sản Việt Nam có sự khác biệt rõ rệt giữa các vùng miền do đặc điểm kinh tế, xã hội, và hạ tầng. Tại các đô thị lớn như Hà Nội, TP. Hồ Chí Minh, Đà Nẵng, bất động sản tập trung vào phân khúc cao cấp, nhà ở thương mại và các dự án quy mô lớn. Những khu vực này là trung tâm kinh tế, chính trị và văn hóa, thu hút lượng lớn dân cư và nhà đầu tư, dẫn đến nhu cầu bất động sản cao hơn so với các tỉnh thành khác.

Ngược lại, các tỉnh ven đô như Bình Dương, Đồng Nai (phía Nam) hay Bắc Ninh, Hải Phòng (phía Bắc) chủ yếu phát triển các dự án đất nền, nhà ở giá rẻ hoặc khu công nghiệp để phục vụ nhu cầu đầu tư và sản xuất. Tính khu vực hóa này không chỉ phản ánh sự phân hóa trong nhu cầu mà còn là đặc điểm đặc trưng của thị trường bất động sản Việt Nam.

1.2.2.2. Tính không đồng đều về phát triển cơ sở hạ tầng

Một trong những yếu tố ảnh hưởng lớn đến thị trường bất động sản Việt Nam là sự phát triển không đồng đều của cơ sở hạ tầng. Các thành phố lớn như Hà Nội và TP. Hồ Chí Minh có hệ thống giao thông và hạ tầng tiện ích phát triển vượt trội, khiến giá trị bất động sản ở những khu vực này tăng cao. Theo CBRE Vietnam (2024), giá nhà ở tại các quận trung tâm TP. Hồ Chí Minh cao gấp 3-5 lần so với các khu vực ngoại thành.

Trong khi đó, nhiều khu vực khác, đặc biệt là các tỉnh vùng sâu, vùng xa, vẫn thiếu hạ tầng giao thông kết nối, làm hạn chế tiềm năng phát triển bất động sản. Tuy nhiên, với các dự án lớn như đường cao tốc Bắc – Nam, sân bay Long Thành, và các tuyến metro tại Hà Nội, TP. Hồ Chí Minh, tình hình này đang dần được cải thiện.

1.2.3. Xu hướng phát triển hiện nay

Giá bất động sản tại các đô thị lớn đang có xu hướng tăng trưởng mạnh mẽ, đặc biệt tại Hà Nội và TP. Hồ Chí Minh. Báo cáo của Savills Vietnam (2022) cho thấy, giá bán căn hộ tại TP. Hồ Chí Minh tăng trung bình 10-12%/năm trong 5 năm qua, trong khi Hà Nội tăng khoảng 8-10%. Các yếu tố thúc đẩy giá tăng bao gồm sự khan hiếm quỹ đất, chi phí xây dựng tăng cao, và sự gia tăng nhu cầu nhà ở từ tầng lớp trung lưu và thượng lưu.

Bên cạnh đó, bất động sản nghỉ dưỡng đang trở thành xu hướng đầu tư hấp dẫn nhờ sự phát triển mạnh mẽ của ngành du lịch. Các khu vực như Phú Quốc, Đà Nẵng, Nha Trang là điểm đến đầu tư lý tưởng với nhiều dự án resort, villa cao cấp. Đồng thời, bất động sản công nghiệp cũng ghi nhận sự bứt phá nhờ vào làn sóng dịch chuyển sản xuất từ Trung Quốc sang Việt Nam. Theo Bộ Kế hoạch và Đầu tư, diện tích đất khu công nghiệp tại các tỉnh như Bình Dương, Đồng Nai, Bắc Ninh được lấp đầy hơn 90%, chứng tỏ sức hút lớn của loại hình này.

1.3. Các yếu tố ảnh hưởng đến thị trường bất động sản

1.3.1. Yếu tố kinh tế

1.3.1.1. Tăng trưởng GDP

Tăng trưởng GDP là thước đo quan trọng phản ánh sức khỏe của nền kinh tế và trực tiếp ảnh hưởng đến thị trường bất động sản. Khi GDP tăng trưởng mạnh, nền kinh tế phát triển ổn định, thu nhập của người dân tăng lên, dẫn đến nhu cầu sở hữu và đầu tư bất động sản cũng tăng theo.

Tại Việt Nam, theo Tổng cục Thống kê (2024), GDP tăng trưởng ước tính 7,09% - nằm trong nhóm nước tăng trưởng cao nhất trong khu vực và trên thế giới. Sự phát triển này kéo theo nhiều dự án bất động sản lớn được triển khai, đáp ứng nhu cầu nhà ở, văn phòng và khu công nghiệp. Các thành phố lớn như Hà Nội, TP. Hồ Chí Minh, Đà Nẵng luôn là tâm điểm của sự phát triển, với nhiều dự án cao cấp và hiện đại nhằm thu hút nhà đầu tư trong nước và quốc tế.

1.3.1.2. Thu nhập bình quân đầu người

Thu nhập bình quân đầu người là một trong những yếu tố quan trọng quyết định khả năng chi trả của người dân đối với bất động sản. Khi thu nhập tăng, người dân có xu hướng chuyển từ thuê nhà sang sở hữu nhà, đồng thời tăng cường đầu tư vào bất động sản để bảo toàn và gia tăng tài sản.

Theo ước tính của Tổng cục Thống kê, thu nhập bình quân đầu người tại Việt Nam năm 2024 đạt khoảng 4.700 USD, cao hơn nhiều so với các năm trước. Sự gia tăng thu nhập này tạo điều kiện để người dân tiếp cận dễ dàng hơn với các sản phẩm bất động sản, đặc biệt là ở phân khúc trung cấp và cao cấp. Đồng thời, các khu vực đô thị lớn cũng ghi nhận sự chuyển dịch từ nhu cầu nhà ở cơ bản sang các dự án cao cấp, tiện nghi và gần trung tâm.

1.3.1.3. Lãi suất vay ngân hàng và chính sách tín dụng

Lãi suất vay ngân hàng và chính sách tín dụng đóng vai trò quan trọng trong việc quyết định sức mua của thị trường bất động sản. Khi lãi suất thấp, người dân và nhà đầu tư có thể dễ dàng tiếp cận nguồn vốn vay để mua nhà hoặc đầu tư bất động sản, từ đó thúc đẩy giao dịch trên thị trường.

Ngược lại, khi lãi suất tăng cao, chi phí vay vốn lớn hơn, người dân sẽ thận trọng hơn trong việc vay tiền mua bất động sản, làm giảm thanh khoản của thị trường. Ví dụ, trong năm 2023, việc tăng lãi suất từ mức trung bình 8% lên 10% đã ảnh hưởng tiêu cực đến nhu cầu vay vốn mua nhà tại nhiều khu vực.

Ngoài ra, các chính sách tín dụng của ngân hàng, chẳng hạn như ưu đãi lãi suất cho người mua nhà lần đầu hoặc hỗ trợ vay vốn cho các dự án nhà ở xã hội, cũng đóng vai trò quan trọng trong việc điều tiết cung cầu trên thị trường.

1.3.2. Yếu tố pháp lý

1.3.2.1. Các quy định về sở hữu, chuyển nhượng và thế chấp bất động sản

Hệ thống pháp lý minh bạch và ổn định là nền tảng quan trọng để phát triển thị trường bất động sản bền vững. Các quy định về quyền sở hữu, chuyển nhượng và thế chấp bất động sản không chỉ bảo vệ quyền lợi của các bên tham gia giao dịch mà còn giúp giảm thiểu rủi ro pháp lý và tăng cường niềm tin của nhà đầu tư.

Tại Việt Nam, Luật Đất đai 2013 và Luật Nhà ở 2014 đã quy định rõ các điều kiện chuyển nhượng quyền sử dụng đất, sở hữu nhà ở và các tài sản gắn liền với đất. Tuy nhiên, trên thực tế, việc thực thi các quy định pháp luật đôi khi còn chậm trễ hoặc thiếu minh bạch, gây khó khăn cho nhà đầu tư và người dân. Ví dụ, các thủ tục cấp sổ đỏ hoặc chuyển nhượng quyền sử dụng đất tại một số địa phương vẫn mất nhiều thời gian và chi phí.

1.3.2.2. Chính sách nhà nước về đất đai và đầu tư bất động sản

Chính sách của nhà nước về quy hoạch đất đai, phát triển hạ tầng và đầu tư vào bất động sản là yếu tố quyết định sự phát triển của thị trường. Các chính sách ưu tiên phát triển nhà ở xã hội, cải cách thủ tục hành chính trong giao dịch đất đai, hoặc thúc đẩy đầu tư vào các khu công nghiệp đều góp phần định hình xu hướng của thị trường.

Một ví dụ điển hình là quy hoạch đường cao tốc Bắc – Nam, tuyến metro tại Hà Nội và TP. Hồ Chí Minh, và sân bay Long Thành. Các dự án này không chỉ nâng cao giá trị bất động sản tại các khu vực lân cận mà còn tạo ra cơ hội đầu tư mới cho thị trường.

1.3.3. Yếu tố xã hội

Việt Nam là một quốc gia đông dân với dân số hơn 100 triệu người (năm 2023), trong đó tỷ lệ đô thị hóa đạt khoảng 40%. Sự gia tăng dân số, đặc biệt tại các khu vực đô thị, đã tạo ra nhu cầu lớn về nhà ở, văn phòng và bất động sản thương mại. Bên cạnh đó, xu hướng di dân từ nông thôn ra thành phố để tìm kiếm cơ hội việc làm cũng góp phần gia tăng áp lực lên thị trường bất động sản đô thị. Các khu vực như Hà Nội, TP. Hồ Chí Minh, Đà Nẵng là điểm đến hấp dẫn nhờ sự phát triển vượt trội về hạ tầng và cơ hội kinh tế.

Nhu cầu nhà ở tại Việt Nam không chỉ xuất phát từ dân số tăng mà còn từ sự thay đổi trong tâm lý tiêu dùng. Người mua hiện nay không chỉ tìm kiếm nơi ở mà còn yêu cầu các tiện ích đi kèm như khu vui chơi, trường học, trung tâm thương mại. Điều này đã thúc đẩy sự phát triển của các dự án bất động sản tích hợp, như khu đô thị xanh, khu đô thị thông minh.

1.3.4. Yếu tố công nghệ

Công nghệ đóng vai trò ngày càng quan trọng trong việc thu thập, phân tích và quản lý thông tin bất động sản. Các công cụ như Big Data, trí tuệ nhân tạo (AI), và học máy (Machine Learning) giúp dự đoán xu hướng giá cả, phân tích cung cầu và đánh giá tiềm

năng của các khu vực bất động sản. Ví dụ, các thuật toán AI hiện nay có thể phân tích hàng triệu dữ liệu để đưa ra dự báo về biến động giá bất động sản một cách chính xác. Các nền tảng trực tuyến như Batdongsan.com.vn, Homedy, hay Propzy đã thay đổi cách thức tiếp cận và giao dịch bất động sản tại Việt Nam. Những nền tảng này không chỉ cung cấp thông tin chi tiết về giá cả, vị trí mà còn hỗ trợ người mua và nhà đầu tư so sánh, lựa chọn bất động sản phù hợp. Sự phát triển của công nghệ cũng giúp thị trường bất động sản trở nên minh bạch và hiệu quả hơn.

1.4. Ứng dụng công nghệ trong lĩnh vực bất động sản

1.4.1. Tầm quan trọng của công nghệ

Công nghệ giúp cải thiện hiệu quả trong các giao dịch bất động sản thông qua việc tối ưu hóa quy trình và cung cấp thông tin một cách minh bạch, dễ dàng tiếp cận. Trước đây, thông tin bất động sản thường bị giới hạn hoặc không minh bạch, dẫn đến các rủi ro về pháp lý và tài chính trong giao dịch. Giờ đây, các nền tảng trực tuyến cho phép người mua và người bán tiếp cận thông tin chi tiết về giá cả, vị trí, và tình trạng pháp lý của bất động sản chỉ trong vài cú nhấp chuột. Điều này giúp giảm thời gian tìm kiếm, đàm phán và thực hiện giao dịch. Ngoài ra, các công cụ trực tuyến còn cung cấp thông tin về quy hoạch, dự án liên quan, và xu hướng phát triển tại khu vực cụ thể, giúp người dùng có cái nhìn toàn diện và đáng tin cậy trước khi đưa ra quyết định.

Công nghệ, đặc biệt là các công cụ phân tích dữ liệu lớn (Big Data) và trí tuệ nhân tạo (AI), cung cấp khả năng phân tích xu hướng thị trường một cách chính xác và hiệu quả. Nhờ đó, người dùng có thể nắm bắt được các biến động giá, dự đoán xu hướng phát triển của các khu vực và đưa ra các quyết định đầu tư phù hợp. AI có thể dự đoán giá trị bất động sản dựa trên các yếu tố như vị trí, hạ tầng, và lịch sử giao dịch, giúp nhà đầu tư giảm thiểu rủi ro và tối ưu hóa lợi nhuận.

1.4.2. Các công nghệ đang được áp dụng

Big Data là một trong những công nghệ cốt lõi được áp dụng trong lĩnh vực bất động sản. Công nghệ này cho phép thu thập và xử lý khối lượng lớn dữ liệu từ nhiều nguồn khác nhau, bao gồm thông tin giao dịch, xu hướng giá, và nhu cầu thị trường. Các nền tảng bất động sản sử dụng Big Data để phân tích hành vi người dùng, từ đó đưa ra các gợi ý phù hợp và tối ưu hóa trải nghiệm người dùng.

Trí tuệ nhân tạo (AI) đang được áp dụng ngày càng rộng rãi trong việc phân tích và dự đoán giá trị bất động sản. AI có khả năng phân tích hàng triệu dữ liệu trong thời gian ngắn, từ đó đưa ra các dự đoán chính xác về biến động giá, mức độ hấp dẫn của từng khu vực và tiềm năng đầu tư. Nổi bật là các thuật toán học máy của Redfin hoặc Opendoor (Hoa Kỳ), cho phép đánh giá giá trị bất động sản dựa trên các yếu tố như vị trí, diện tích, tiện ích xung quanh, và dữ liệu lịch sử giao dịch. Tại Việt Nam, một số công ty công nghệ bất động

sản (PropTech) đã bắt đầu áp dụng AI để dự báo giá nhà và cung cấp các khuyến nghị đầu tư phù hợp.

Blockchain là một công nghệ mang tính cách mạng trong việc đảm bảo tính minh bạch và bảo mật trong giao dịch bất động sản. Các hợp đồng thông minh (Smart Contracts) dựa trên Blockchain cho phép thực hiện giao dịch một cách tự động và an toàn, loại bỏ nhu cầu về trung gian truyền thống và giảm thiểu rủi ro về pháp lý. Ngoài ra, Blockchain cũng được sử dụng để quản lý tài sản, lưu trữ thông tin quyền sở hữu, và theo dõi lịch sử giao dịch của bất động sản. Điều này không chỉ giúp tăng cường niềm tin của các bên tham gia mà còn giảm thiểu tình trạng lừa đảo trong giao dịch.

1.4.3. Thực trạng ứng dụng công nghệ tại Việt Nam

1.4.3.1. Tình hình sử dụng các nền tảng trực tuyến tại Việt Nam

Trong những năm gần đây, việc sử dụng các nền tảng trực tuyến trong lĩnh vực bất động sản tại Việt Nam đã trở nên phổ biến hơn, nhờ vào sự phát triển của công nghệ và sự gia tăng nhu cầu giao dịch qua các kênh số. Các nền tảng như Batdongsan.com.vn, Homedy, Chotot.vn, và Propzy đã trở thành những công cụ quen thuộc, cung cấp thông tin chi tiết về bất động sản như giá cả, vị trí, diện tích, tình trạng pháp lý, và tiện ích kèm theo.

Các nền tảng này không chỉ giúp người mua và người bán tiếp cận nhau một cách dễ dàng hơn mà còn cung cấp các tính năng nâng cao như:

- Công cụ so sánh giá: Cho phép người dùng so sánh giá trị bất động sản tại các khu vực khác nhau, giúp họ đưa ra quyết định hợp lý.
- Tìm kiếm theo tiêu chí cá nhân hóa: Người dùng có thể tìm kiếm bất động sản dựa trên các tiêu chí cụ thể như diện tích, giá tiền, và tiện ích xung quanh.
- Hỗ trợ pháp lý và tư vấn tài chính: Một số nền tảng cung cấp dịch vụ tư vấn pháp lý và hỗ trợ vay vốn từ ngân hàng, giúp quá trình giao dịch trở nên thuận tiện hơn.

Theo báo cáo của Batdongsan.com.vn, lượng người dùng tìm kiếm bất động sản qua các kênh trực tuyến đã tăng trung bình 15% mỗi năm trong giai đoạn 2018-2022, đặc biệt tại các thành phố lớn như Hà Nội, TP. Hồ Chí Minh, và Đà Nẵng. Sự gia tăng này phản ánh xu hướng chuyển đổi số mạnh mẽ trong lĩnh vực bất động sản, cũng như sự thay đổi hành vi của người tiêu dùng khi họ ngày càng ưu tiên sử dụng công nghệ để tiết kiệm thời gian và tối ưu hóa chi phí.

1.4.3.2. Những thách thức trong việc áp dụng công nghệ vào thị trường bất động sản

Mặc dù đã có nhiều tiến bộ trong việc áp dụng công nghệ vào lĩnh vực bất động sản, nhưng tại Việt Nam, quá trình này vẫn gặp phải không ít khó khăn và thách thức. Những hạn chế này xuất phát từ nhiều nguyên nhân, bao gồm cơ sở hạ tầng công nghệ, nhận thức của người dùng, và các yếu tố pháp lý.

Một trong những thách thức lớn nhất là tình trạng thiếu đồng bộ và minh bạch trong dữ liệu bất động sản. Hiện nay, thông tin về giá cả, tình trạng pháp lý, và quy hoạch đất đai

vẫn còn phân tán tại nhiều cơ quan khác nhau, gây khó khăn cho việc thu thập và phân tích dữ liệu. Các nền tảng trực tuyến tại Việt Nam thường phải dựa vào dữ liệu tự báo cáo từ người dùng, dẫn đến tình trạng thông tin sai lệch hoặc không đầy đủ. Ví dụ, một số tin rao bán trên các trang web có thể không chính xác về giá cả hoặc pháp lý, gây ảnh hưởng đến quyết định của người mua.

Tại nhiều khu vực nông thôn hoặc vùng sâu, vùng xa, cơ sở hạ tầng công nghệ như mạng internet tốc độ cao hoặc các thiết bị hỗ trợ vẫn còn hạn chế. Điều này làm giảm khả năng tiếp cận công nghệ của một bộ phận lớn dân cư, đặc biệt là ở các tỉnh thành chưa phát triển mạnh về kinh tế. Các nền tảng trực tuyến hiện tại chủ yếu tập trung vào đối tượng người dùng tại các thành phố lớn, trong khi các khu vực ngoại thành hoặc nông thôn ít được chú trọng.

Một số doanh nghiệp bất động sản và người dùng cá nhân vẫn chưa thực sự sẵn sàng áp dụng công nghệ trong hoạt động giao dịch hoặc đầu tư. Nhiều người vẫn duy trì thói quen tìm kiếm thông tin và giao dịch qua các kênh truyền thống như môi giới trực tiếp hoặc các văn phòng giao dịch. Sự thiếu hiểu biết về lợi ích của công nghệ, cùng với tâm lý e ngại trước các rủi ro liên quan đến giao dịch trực tuyến, cũng làm chậm quá trình ứng dụng công nghệ vào lĩnh vực bất động sản.

Việc áp dụng các công nghệ mới như Blockchain hay hợp đồng thông minh (Smart Contracts) vào giao dịch bất động sản vẫn gặp phải rào cản pháp lý. Các quy định hiện hành tại Việt Nam chưa thực sự hỗ trợ cho việc triển khai các giải pháp công nghệ tiên tiến này. Ngoài ra, các khung pháp lý liên quan đến bảo mật thông tin và quản lý dữ liệu người dùng vẫn còn hạn chế, gây khó khăn trong việc xây dựng niềm tin đối với các nền tảng trực tuyến.

Các giải pháp công nghệ hiện đại như trí tuệ nhân tạo (AI), học máy (Machine Learning), hay Blockchain đòi hỏi nguồn vốn đầu tư lớn. Điều này đặt ra thách thức đối với các doanh nghiệp vừa và nhỏ, khiến họ khó tiếp cận và triển khai các giải pháp công nghệ vào hoạt động kinh doanh.

1.4.3.3. Định hướng khắc phục và tiềm năng phát triển trong tương lai

Dù còn nhiều thách thức, việc ứng dụng công nghệ vào lĩnh vực bất động sản tại Việt Nam vẫn có tiềm năng phát triển lớn. Một số giải pháp định hướng bao gồm:

- Đầu tư vào hạ tầng công nghệ: Chính phủ và doanh nghiệp cần tập trung cải thiện cơ sở hạ tầng công nghệ, đặc biệt tại các khu vực ngoại thành và nông thôn, để tạo điều kiện thuận lợi cho việc áp dụng các giải pháp số hóa.
- Xây dựng hệ thống dữ liệu đồng bộ: Việc tích hợp và đồng bộ hóa dữ liệu bất động sản từ các cơ quan chức năng sẽ giúp tăng cường tính minh bạch và hiệu quả trong giao dịch.
- Nâng cao nhận thức và đào tạo kỹ năng số: Các chương trình đào tạo và nâng cao nhận thức về công nghệ trong bất động sản có thể giúp các bên liên quan hiểu rõ hơn về lợi ích và cách sử dụng công nghệ.

- Cải thiện khung pháp lý: Chính phủ cần điều chỉnh và cập nhật các quy định pháp luật để hỗ trợ việc triển khai các công nghệ tiên tiến, như Blockchain, hợp đồng thông minh, và trí tuệ nhân tạo.

Ứng dụng công nghệ trong lĩnh vực bất động sản tại Việt Nam đang ở giai đoạn đầu với nhiều cơ hội phát triển. Với sự đầu tư đúng đắn và quản lý hiệu quả, công nghệ có thể trở thành động lực mạnh mẽ thúc đẩy sự minh bạch, hiệu quả và bền vững của thị trường bất động sản trong tương lai.

1.5. Nhận định và vấn đề cần nghiên cứu

1.5.1. Nhận định về tiềm năng của thị trường bất động sản Việt Nam

Việt Nam là một trong những quốc gia có tốc độ tăng trưởng kinh tế nhanh nhất khu vực Đông Nam Á. GDP tăng trưởng trung bình 6-7% mỗi năm trong giai đoạn 2015-2022 (Tổng cục Thống kê, 2022). Sự phát triển này tạo điều kiện thúc đẩy nhu cầu về nhà ở, bất động sản thương mại, khu công nghiệp, và các dự án nghỉ dưỡng. Bên cạnh đó, việc đầu tư mạnh mẽ vào cơ sở hạ tầng cũng là một lợi thế lớn. Các dự án lớn như đường cao tốc Bắc – Nam, sân bay Long Thành, và hệ thống metro tại Hà Nội, TP. Hồ Chí Minh không chỉ cải thiện kết nối giao thông mà còn nâng cao giá trị bất động sản tại các khu vực lân cận. Ví dụ, giá đất tại các khu vực gần sân bay Long Thành đã tăng gấp 3 lần trong giai đoạn 2019-2022 (Savills Vietnam, 2023).

Chính phủ Việt Nam đã ban hành nhiều chính sách khuyến khích và hỗ trợ đầu tư vào thị trường bất động sản. Luật Nhà ở 2014 và Luật Đất đai 2013 đã cải thiện môi trường pháp lý, tạo điều kiện thuận lợi cho các doanh nghiệp và nhà đầu tư trong nước cũng như quốc tế. Ngoài ra, các chương trình phát triển nhà ở xã hội, khu đô thị thông minh, và khu công nghiệp đã nhận được sự quan tâm lớn từ Chính phủ. Chính sách ưu đãi thuế, đơn giản hóa thủ tục hành chính trong giao dịch bất động sản cũng giúp nâng cao niềm tin của nhà đầu tư và tạo động lực phát triển thị trường.

- Những vấn đề còn tồn tại:

Một trong những thách thức lớn nhất đối với thị trường bất động sản Việt Nam là sự thiếu minh bạch trong thông tin và dữ liệu. Hiện nay, thông tin về giá cả, quy hoạch đất đai, và tình trạng pháp lý của các dự án bất động sản thường không được công khai đầy đủ hoặc không đồng bộ giữa các cơ quan chức năng. Điều này không chỉ gây khó khăn cho các bên tham gia giao dịch mà còn làm tăng rủi ro pháp lý và tài chính. Theo Hiệp hội Bất động sản Việt Nam (2022), khoảng 60% giao dịch bất động sản tại Việt Nam gặp phải vấn đề liên quan đến thiếu minh bạch thông tin, đặc biệt ở các tỉnh thành nhỏ và khu vực nông thôn.

Dù công nghệ đang dần thay đổi cách thức hoạt động của thị trường bất động sản nhưng việc áp dụng các công nghệ hiện đại như Big Data, trí tuệ nhân tạo (AI), hay Blockchain vẫn còn hạn chế ở nhiều khu vực, đặc biệt là vùng sâu, vùng xa. Bên cạnh đó, sự thiếu đồng bộ về hạ tầng công nghệ, internet tốc độ cao, và nhận thức về lợi ích của công

nghệ trong cộng đồng cũng là rào cản lớn đối với việc triển khai các giải pháp công nghệ toàn diện trong lĩnh vực bất động sản.

1.5.2 Hướng giải quyết và mục tiêu nghiên cứu của đề tài

Một trong những giải pháp hiệu quả nhất để giải quyết các vấn đề hiện tại của thị trường bất động sản là áp dụng công nghệ hiện đại vào phân tích và dự báo thị trường. Các công cụ Big Data, AI, và học máy (Machine Learning) có thể giúp thu thập và xử lý dữ liệu lớn từ nhiều nguồn khác nhau, từ đó cung cấp thông tin chính xác và dự báo xu hướng thị trường một cách hiệu quả.

Ví dụ, các nền tảng bất động sản tiên tiến có thể sử dụng Big Data để phân tích giá trị bất động sản theo thời gian thực, hoặc AI để đánh giá tiềm năng của từng khu vực, từ đó giúp nhà đầu tư đưa ra các quyết định chính xác hơn.

Đề tài hướng đến việc xây dựng một hệ thống hỗ trợ trực tuyến, cung cấp các công cụ và dịch vụ cần thiết cho người dùng và nhà đầu tư. Các tính năng chính của hệ thống có thể bao gồm:

- Công cụ tra cứu thông tin minh bạch: Cho phép người dùng tra cứu giá cả, quy hoạch, và tình trạng pháp lý của bất động sản một cách chính xác và nhanh chóng.
- Hệ thống dự báo xu hướng giá cả và cung cầu: Sử dụng AI và học máy để đưa ra các dự báo về xu hướng giá cả, nhu cầu tại các khu vực tiềm năng.
- Tích hợp Blockchain để đảm bảo an toàn giao dịch: Blockchain giúp lưu trữ và xác minh thông tin giao dịch một cách an toàn, minh bạch và không thể sửa đổi.
- Hỗ trợ tư vấn và kết nối: Hệ thống có thể kết nối người mua, người bán và nhà đầu tư với các chuyên gia pháp lý, tài chính để hỗ trợ trong quá trình giao dịch.

1.6 Dữ liệu và lý thuyết về các mô hình học máy

Trên cơ sở phân tích thị trường bất động sản bao gồm các khái niệm, đặc điểm, yếu tố tác động đồng thời dựa vào hướng giải quyết được đề cập ở phần trên nhóm quyết định lựa chọn dữ liệu và các công nghệ học máy phù hợp để đưa vào thực nghiệm việc dự đoán bất động sản.

1.6.1 Dữ liệu sử dụng

Dữ liệu nhóm sử dụng để thử nghiệm mô hình học máy được lấy từ trang BatDongSan.com và các nguồn uy tín trong giới máy học. Dưới đây là giải thích 10 yếu tố dữ liệu chính được đưa vào mô hình:

- Diện tích bất động sản, được đo bằng mét vuông (m^2), là một trong những yếu tố quan trọng nhất ảnh hưởng đến giá trị bất động sản. Thông thường, diện tích lớn hơn sẽ mang lại giá trị cao hơn do khả năng sử dụng đa dạng, bao gồm việc xây dựng,

mở rộng hoặc kinh doanh. Đây là yếu tố trực tiếp phản ánh quy mô và tiềm năng khai thác của bất động sản.

- Mặt tiền của bất động sản, đo bằng chiều rộng tính từ phía trước, đóng vai trò quan trọng trong việc định giá. Bất động sản có mặt tiền lớn thường có giá trị cao hơn, đặc biệt ở khu vực đô thị hoặc ven đường lớn. Điều này xuất phát từ tính thẩm mỹ, tiềm năng kinh doanh, và sự thuận tiện trong tiếp cận từ mặt đường.
- Độ rộng của đường tiếp cận bất động sản, thường tính bằng mét, ảnh hưởng lớn đến giá trị bất động sản. Đường tiếp cận rộng rãi, đặc biệt là những tuyến đường chính, tạo điều kiện thuận lợi cho giao thông và tăng tính hấp dẫn cho bất động sản, đặc biệt đối với các mục đích thương mại.
- Số tầng của bất động sản thể hiện quy mô xây dựng theo chiều cao và ảnh hưởng đến giá trị tổng thể. Bất động sản có nhiều tầng thường được định giá cao hơn do cung cấp không gian sử dụng lớn hơn, phù hợp với nhu cầu ở, làm việc hoặc kinh doanh.
- Số lượng phòng ngủ đại diện cho khả năng đáp ứng nhu cầu sinh hoạt của một gia đình hoặc nhiều người. Bất động sản với nhiều phòng ngủ thường có giá trị cao hơn, đặc biệt ở các khu vực dân cư đông đúc hoặc dành cho các gia đình lớn.
- Số phòng tắm là một yếu tố quan trọng phản ánh tiện nghi của bất động sản. Bất động sản có nhiều phòng tắm sẽ hấp dẫn hơn, đặc biệt đối với các gia đình lớn hoặc bất động sản cao cấp, qua đó làm tăng giá trị của tài sản.
- Giá bất động sản là biến mục tiêu trong mô hình dự đoán, thể hiện giá trị tiền tệ mà thị trường sẵn sàng trả cho bất động sản đó. Giá chịu ảnh hưởng từ tất cả các yếu tố đầu vào, bao gồm diện tích, vị trí, tiện nghi và các đặc điểm khác.
- Xã hoặc phường là cấp hành chính nhỏ nhất xác định vị trí của bất động sản. Giá trị bất động sản thường dao động đáng kể giữa các xã/phường trong cùng một huyện, tùy thuộc vào mức độ phát triển và tiềm năng kinh tế của khu vực.
- Quận hoặc huyện là yếu tố định vị quan trọng quyết định giá trị bất động sản. Các khu vực thuộc trung tâm thành phố hoặc có tốc độ đô thị hóa nhanh thường có giá trị cao hơn so với các huyện ngoại thành hoặc khu vực kém phát triển.
- Tỉnh hoặc thành phố nơi bất động sản tọa lạc có ảnh hưởng lớn đến giá cả, do đây là yếu tố quyết định mức độ phát triển kinh tế, hạ tầng, và thị trường của khu vực. Bất động sản tại các tỉnh, thành phố lớn như Hà Nội, TP. Hồ Chí Minh thường có giá trị cao hơn các khu vực khác.

1.6.2 Hồi quy và ứng dụng trong bất động sản

Bài toán hồi quy là một dạng bài toán trong học máy, nơi mục tiêu là dự đoán một giá trị liên tục (số thực) từ các đặc trưng đầu vào. Bài toán này thường xuất hiện khi các biến đầu vào có mối quan hệ định lượng với biến mục tiêu và có thể được biểu diễn dưới dạng hàm toán học. Trong hồi quy, mục tiêu là xây dựng một mô hình sao cho giá trị dự đoán của biến mục tiêu càng gần với giá trị thực càng tốt. Các mô hình hồi quy có thể bao gồm

những mô hình đơn giản như Hồi quy tuyến tính (Linear Regression) hoặc những mô hình phức tạp hơn như Random Forest Regressor và Gradient Boosting Regressor. Các ứng dụng phổ biến của bài toán hồi quy bao gồm dự đoán giá cả, doanh thu, mức độ tiêu thụ, hoặc các chỉ số khác có thể đo lường liên tục.

Trong lĩnh vực bất động sản, bài toán hồi quy đóng vai trò quan trọng trong việc dự đoán giá trị của các bất động sản dựa trên các yếu tố như diện tích, vị trí, số phòng, và các đặc trưng khác. Đây là một bài toán hồi quy vì giá trị của bất động sản là một biến liên tục có thể thay đổi tùy thuộc vào nhiều yếu tố ảnh hưởng. Các mô hình hồi quy, đặc biệt là Random Forest Regressor và Gradient Boosting Regressor, thường được sử dụng để xây dựng các mô hình dự đoán giá nhà vì khả năng xử lý dữ liệu phi tuyến tính và tính tương tác giữa các đặc trưng.

Các yếu tố như diện tích (Area), số phòng ngủ (Bedrooms), số phòng tắm (Bathrooms), vị trí (District, Province), và mặt tiền (Frontage) là những yếu tố quan trọng trong việc xác định giá trị của bất động sản. Việc xây dựng mô hình dự đoán giá bất động sản dựa trên các đặc trưng này giúp các nhà đầu tư, người mua, và các công ty bất động sản có thể đưa ra quyết định chính xác hơn trong việc mua bán và đầu tư. Các phương pháp hồi quy giúp cải thiện khả năng dự đoán khi có đủ dữ liệu và cung cấp một công cụ hữu ích cho thị trường bất động sản.

1.6.3 Lý thuyết các mô hình học máy

1.6.3.1 RandomForestRegressor

Nguyên lý:

RandomForestRegressor là một mô hình học máy thuộc nhóm Ensemble Learning, được xây dựng trên nguyên lý Bagging (Bootstrap Aggregating). Mô hình này tạo ra một tập hợp các cây quyết định (Decision Trees), mỗi cây được huấn luyện trên một mẫu con ngẫu nhiên của tập dữ liệu gốc. Sau khi các cây được huấn luyện, dự đoán cuối cùng được đưa ra bằng cách tính trung bình hoặc bỏ phiếu (cho bài toán phân loại). Việc huấn luyện mỗi cây trên một mẫu con ngẫu nhiên và việc lấy trung bình dự đoán từ nhiều cây quyết định giúp giảm thiểu overfitting và nâng cao độ chính xác của mô hình.

Ứng dụng:

RandomForestRegressor có khả năng làm việc hiệu quả trong các bài toán hồi quy phức tạp, đặc biệt là khi dữ liệu có các mối quan hệ phi tuyến tính và các đặc trưng có sự tương tác phức tạp. Mô hình này thường được sử dụng trong các ứng dụng như:

- Dự đoán giá bất động sản: Mô hình có thể dự đoán giá trị bất động sản dựa trên các yếu tố như diện tích, vị trí, số phòng, mặt tiền, v.v.
- Dự đoán doanh thu bán hàng: Dự đoán doanh thu trong các ngành công nghiệp dựa trên các yếu tố như chiến dịch marketing, mùa vụ, v.v.
- Dự đoán sản lượng nông nghiệp: Mô hình giúp dự đoán sản lượng thu hoạch dựa trên các yếu tố như khí hậu, giống cây trồng, và các yếu tố môi trường khác.

1.6.3.2 LinearRegression

Nguyên lý:

LinearRegression là một mô hình hồi quy tuyến tính, trong đó mối quan hệ giữa các đặc trưng (biến độc lập) và biến mục tiêu (biến phụ thuộc) được giả định là tuyến tính. Phương trình của mô hình có dạng:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Trong đó:

- y là biến mục tiêu (dự đoán).
- x_1, x_2, \dots, x_n là các đặc trưng (biến độc lập).
- $\beta_0, \beta_1, \dots, \beta_n$ là các hệ số cần được học từ dữ liệu. Mô hình tối thiểu hóa sai số giữa giá trị thực tế và giá trị dự đoán bằng cách sử dụng phương pháp **Least Squares**.

Ứng dụng:

LinearRegression được sử dụng trong các bài toán hồi quy khi có mối quan hệ tuyến tính rõ ràng giữa các đặc trưng và biến mục tiêu. Các ứng dụng phổ biến bao gồm:

- Dự đoán giá trị bất động sản: Dự đoán giá nhà dựa trên các yếu tố như diện tích, số phòng ngủ, v.v., trong trường hợp mối quan hệ giữa các yếu tố này và giá trị nhà có tính tuyến tính.
- Dự đoán tiêu thụ năng lượng: Dự đoán mức tiêu thụ năng lượng của các hộ gia đình hoặc các công ty dựa trên các yếu tố như diện tích sử dụng, số lượng thiết bị tiêu thụ điện, v.v.
- Dự đoán hiệu suất học tập: Dự đoán điểm số của học sinh dựa trên các yếu tố như thời gian học, số buổi học, v.v.

Tiểu kết chương 1

Trong chương 1, nhóm đã tập trung phân tích cơ sở lý thuyết liên quan đến thị trường bất động sản Việt Nam và các mô hình học máy trong dự báo giá bất động sản. Thị trường bất động sản Việt Nam được đánh giá thông qua các yếu tố đặc trưng như vị trí địa lý, hạ tầng, pháp lý, cũng như sự tác động của các yếu tố kinh tế vĩ mô. Đây là những cơ sở quan trọng giúp định hình các biến đầu vào trong các bài toán dự báo giá.

Nhóm cũng đã đi sâu phân tích mô hình hồi quy tuyến tính (Linear Regression) và rừng ngẫu nhiên (RandomForest Regressor) – hai mô hình phổ biến trong dự báo giá bất động sản. Linear Regression, với khả năng đơn giản và dễ giải thích, thích hợp cho các bài toán có quan hệ tuyến tính rõ ràng giữa các biến. Trong khi đó, RandomForest Regressor, với sự kết hợp của nhiều cây quyết định, cho phép dự đoán hiệu quả trên các tập dữ liệu phi tuyến tính và có tính biến động cao, vốn là đặc điểm của dữ liệu bất động sản.

Những đặc điểm, ưu điểm và hạn chế của hai mô hình này đã được phân tích, nhằm hỗ trợ cho việc lựa chọn phương pháp phù hợp với bài toán dự báo giá bất động sản. Đây là nền tảng lý thuyết quan trọng để triển khai các bước thực nghiệm trong các chương tiếp theo.

CHƯƠNG II. CHUẨN BỊ VÀ XÂY DỰNG MÔ HÌNH DỰ BÁO

2.1 Chuẩn bị dữ liệu

2.1.1 Nguồn dữ liệu

Nguồn dữ liệu được thu thập từ website Batdongsan.com.vn tại chuyên mục Nhà đất bán (<https://batdongsan.com.vn/nha-dat-ban>). Đây là một trong những trang web hàng đầu tại Việt Nam chuyên cung cấp thông tin về bất động sản, bao gồm các tin đăng mua bán nhà đất, căn hộ, và các loại bất động sản khác.

2.1.2 Thu thập dữ liệu

2.1.2.1 Chuẩn bị môi trường

Để có thể thu thập dữ liệu từ website Batdongsan.com.vn thì cần chuẩn bị một số công cụ và thư viện cần thiết :

- Python: Ngôn ngữ lập trình chính.
 - Python là một ngôn ngữ lập trình bậc cao, đa năng, được sử dụng rộng rãi trong nhiều lĩnh vực, từ phát triển web, khoa học dữ liệu, trí tuệ nhân tạo đến tự động hóa.
 - Phiên bản Python được sử dụng là Python 3.11
- Selenium: Thư viện để tự động hóa trình duyệt.
 - Selenium là một thư viện tự động hóa trình duyệt mạnh mẽ, thường được sử dụng để thu thập dữ liệu từ các trang web động.
- WebDriver: Phiên bản phù hợp với trình duyệt
 - Phiên bản webdriver được sử dụng là ChromeDriver 130.0.6723.91
- Database : PostgreSQL
 - PostgreSQL là hệ cơ sở dữ liệu miễn phí được sử dụng phổ biến trên thế giới
- Thư viện và công cụ:
 - psycopg2: Kết nối và chèn dữ liệu vào cơ sở dữ liệu PostgreSQL.
 - User-Agent giả lập: Sử dụng User-Agent để trình duyệt tự động hóa có thể giả lập giống người dùng thật, tránh bị phát hiện là bot.

2.1.2.2. Cách thức thực hiện

Giải thích nguyên lý hoạt động:

- Sử dụng thư viện Selenium để truy cập vào website Batdongsan.com.vn tại chuyên mục Nhà đất bán (<https://batdongsan.com.vn/nha-dat-ban>).
- Sau đó truy cập tuần tự vào các thẻ html của trang web để thực hiện việc thu thập dữ liệu .
- Sau khi thu thập dữ liệu tuần tự từ các trang thì tiến hành thêm vào cơ sở dữ liệu một cách tự động.

Các bước thực hiện:

Bước 1: Cài đặt các thư viện cần thiết

```
import time
import random
```

```
import psycopg2
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from urllib.parse import urljoin
from selenium.webdriver.chrome.options import Options
from selenium.common.exceptions import NoSuchElementException
```

Hình 2. 1 Cài đặt thư viện thu thập dữ liệu

Giải thích

1. time

- Dùng để tạm dừng chương trình một thời gian ngắn, giúp mô phỏng hành vi của người dùng thật khi lướt web.
- Ví dụ: `time.sleep(2)` sẽ tạm dừng chương trình trong 2 giây.

2. random

- Dùng để tạo các khoảng thời gian chờ ngẫu nhiên, tránh bị phát hiện là bot tự động.
- Ví dụ: `random.uniform(1, 3)` sẽ tạo ra một số ngẫu nhiên từ 1 đến 3 giây.

3. psycopg2

- Dùng để kết nối và làm việc với cơ sở dữ liệu PostgreSQL.
- Các tác vụ chính:
 - Kết nối cơ sở dữ liệu.
 - Chèn dữ liệu (INSERT) từ web vào bảng.

4. selenium.webdriver

- Thư viện chính để điều khiển trình duyệt web (như Chrome) tự động.

Các thành phần chính:

- **webdriver.Chrome:** Dùng để mở trình duyệt Chrome.
- **By:** Xác định cách tìm các phần tử trên trang web, ví dụ:
 - `By.XPATH`: Tìm phần tử theo đường dẫn XPATH.
 - `By.CLASS_NAME`: Tìm phần tử theo tên lớp.
- **Service:** Chỉ đường dẫn đến tệp `chromedriver.exe`.

- **Options:** Cấu hình trình duyệt Chrome, ví dụ:
 - Giả lập User-Agent để trình duyệt giống như người dùng thật.
- **NoSuchElementException:** Bắt lỗi khi không tìm thấy phần tử trên trang web, giúp chương trình không bị dừng.

5. urllib.parse.urljoin

- Dùng để kết hợp URL chính với các liên kết phụ (relative link) để tạo URL đầy đủ.
- Ví dụ: Kết hợp `https://batdongsan.com.vn` với `/nha-dat-ban` thành URL hoàn chỉnh.

Bước 2: Thiết lập những yếu tố cần thiết để thực hiện việc thu thập dữ liệu

- Thiết lập User Agent để giả làm người dùng thật để vượt qua hệ thống ngăn chặn của trang web.

chrome_options = Options()

chrome_options.add_argument("user-agent=Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36")

- Khởi chạy webdriver để sử dụng trình duyệt:

```
service = Service('chromedriver.exe')
driver = webdriver.Chrome(service=service,
options=chrome_options)
```

- Kết nối đến cơ sở dữ liệu PostgreSQL trên local Host:

```
def connect_db():
    try:
        connection = psycopg2.connect(
            dbname="postgres", # Tên cơ sở dữ liệu trong
LocalHost
            user="postgres", # Tên người dùng trong cơ sở
dữ liệu
            password="123", # Mật khẩu để trung cập cơ sở
dữ liệu
            host="localhost", # Host
            port="5432" # Cổng mặc định của PostgreSQL
        )
        return connection
    except Exception as e:
        print(f"Không thể kết nối đến cơ sở dữ liệu: {e}")
        return None
```

- Thiết lập việc chèn dữ liệu thu thập được vào cơ sở dữ liệu:

```
def insert_property_data(data):
    conn = connect_db()
    if conn:
        try:
            cursor = conn.cursor()
            insert_query = """
                INSERT INTO thông_tin_bất_động_sản (
                    tiêu_de, địa_chi, diện_tích, giá, số_phòng_ngủ,
                    số_phòng_vs, nơi_thất,
                    pháp_ly, hướng_nhà, hướng_ban_công, số_tầng,
                    mat_tien, duong_vao, loại_bds
                ) VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s,
                    %s, %s, %s)
            """
            cursor.executemany(insert_query, data)
            conn.commit()
            print(f"Đã chèn {len(data)} bất động sản vào cơ sở dữ liệu.")
        except Exception as e:
            print(f"Lỗi khi chèn dữ liệu: {e}")
        finally:
            cursor.close()
            conn.close()
```

Hình 2. 2 Thiết lập việc chèn dữ liệu thu thập được vào cơ sở dữ liệu

- Khởi tạo hàm thu thập thông tin chi tiết bất động sản: Hàm `title = driver.find_element` : dùng để tìm các element trong html có dạng `"h1[@class='re__pr-title pr-title js__pr-title']"` và `"//span[@class='re__pr-short-description js__pr-address']"`. Đây là hai thẻ lưu trữ các thông tin cơ bản của bất động sản.

```
# Hàm thu thập thông tin chi tiết bất động sản
def crawl_property_specs(url, loai_bds):
    driver.get(url)
    driver.implicitly_wait(5)

    property_data = []

    try:
        title = driver.find_element(By.XPATH,
            '//h1[@class="re__pr-title pr-title js__pr-title"]').text
    except NoSuchElementException:
        title = 'Không có tiêu đề'

    try:
        short_description = driver.find_element(By.XPATH,
            '//span[@class="re__pr-short-description js__pr-address"]').text
    except NoSuchElementException:
        short_description = 'Không có mô tả ngắn'
```

Hình 2. 3 Khởi tạo hàm thu thập thông tin chi tiết

- Duyệt qua toàn bộ các thẻ HTML để thực hiện việc lấy dữ liệu:

```
• spec_items = driver.find_elements(By.XPATH, '//div[@class="re__pr-
spec-specs-content-item"]')
•
• for item in spec_items:
•     try:
•         spec_title = item.find_element(By.XPATH,
'./span[@class="re__pr-specs-content-item-title"]').text
•         spec_value = item.find_element(By.XPATH,
'./span[@class="re__pr-specs-content-item-value"]').text
•
•         # Cập nhật thông tin vào từ điển
•         if spec_title in property_specs:
•             property_specs[spec_title] = spec_value
•
•     except NoSuchElementException:
•         continue
•
• # Lấy giá trị từ từ điển property_specs
• dien_tich = property_specs["Diện tích"]
• gia = property_specs["Mức giá"]
• huong_nha = property_specs["Hướng nhà"]
• huong_ban_cong = property_specs["Hướng ban công"]
• so_tang = property_specs["Số tầng"]
```

- `so_phong_ngu = property_specs["Số phòng ngủ"]`
- `so_phong_vs = property_specs["Số toilet"]`
- `phap_ly = property_specs["Pháp lý"]`
- `noi_that = property_specs["Nội thất"]`
- `mat_tien = property_specs["Mặt tiền"]`
- `duong_vao = property_specs["Đường vào"]`

- Bắt đầu duyệt qua các trang con trong danh sách các bất động sản.

```
# Hàm thu thập dữ liệu từ trang danh sách
def crawl_data(driver, loai_bds):
    scroll_pause_time = random.uniform(1, 3) # Thêm thời gian
    chờ ngẫu nhiên giữa các lần cuộn
    driver.execute_script("window.scrollTo(0,
document.body.scrollHeight);")
    time.sleep(scroll_pause_time)

    elems = driver.find_elements(By.XPATH,
'//a[@class="js__product-link-for-product-id"]')

    page_data = []

    for elem in elems:
        href = elem.get_attribute('href')
        if href:
            full_link = urljoin('https://batdongsan.com.vn',
href)
            page_data.append(full_link)

    return page_data
```

Hình 2. 4 Duyệt các trang con trong danh sách

- Tiến hành việc thu thập dữ liệu
 - Cấu trúc của một liên kết.
 - Sau khi thu thập hết dữ liệu trong một trang. Sẽ tự động mở trang mới và tiếp tục lặp lại quy trình như trên.

```

def main():
    n = 2 # Bắt đầu từ trang 2
    try:
        while True:
            url =
f"https://batdongsan.com.vn/nha-dat-ban/p{n}" # Đặt URL mặc
định

            loai_bds = 1 # Đánh dấu là mua bán
            if "cho-thue" in url: # Nếu URL thuộc loại cho
thuê

                loai_bds = 2

            print(f"Opening URL: {url}")
            driver.get(url)
            driver.implicitly_wait(5)

            # Crawl data from the current page
            page_data = crawl_data(driver, loai_bds)

            # Thu thập dữ liệu từ từng liên kết bất động sản
            for link in page_data:
                print(f"Crawling property specs from:
{link}")

                property_data = crawl_property_specs(link,
loai_bds)

                if property_data:
                    insert_property_data(property_data) #
Chèn dữ liệu vào cơ sở dữ liệu

                # Thời gian chờ trước khi chuyển sang trang tiếp
theo

                time.sleep(1)

                # Tăng số trang
                n += 1

                # Dừng vòng lặp sau một số trang nhất định (tùy
chọn)

                if n > 100000000: # Điều chỉnh theo nhu cầu
                    break

    except Exception as e:
        print(f"An error occurred: {e}")
    finally:
        # Đóng trình duyệt
        driver.quit()

```

Hình 2. 5 Tiến hành thu thập dữ liệu

2.2 Lưu trữ dữ liệu

Nhóm lựa chọn hệ cơ sở dữ liệu PostgreSQL trong sản phẩm này dựa trên một số tiêu chí sau:

PostgreSQL là một hệ quản trị cơ sở dữ liệu quan hệ mã nguồn mở mạnh mẽ, được thiết kế để hỗ trợ các ứng dụng yêu cầu xử lý dữ liệu phức tạp và hiệu năng cao. Đây là một trong những hệ quản trị cơ sở dữ liệu phổ biến nhất hiện nay với các tính năng tiên tiến và độ tin cậy cao.

Lịch sử và sự phát triển

PostgreSQL được phát triển từ dự án POSTGRES tại Đại học California, Berkeley vào năm 1986, do Giáo sư Michael Stonebraker khởi xướng. Năm 1996, hệ quản trị này chính thức được đổi tên thành PostgreSQL, đánh dấu sự hỗ trợ đầy đủ cho ngôn ngữ SQL.

Lý do lựa chọn

- Tính năng mạnh mẽ và linh hoạt
 - PostgreSQL là một hệ quản trị cơ sở dữ liệu quan hệ nguồn mở tiên tiến, hỗ trợ đầy đủ các tính năng của SQL và các tính năng nâng cao như lưu trữ JSON, xử lý dữ liệu không gian (PostGIS), và các hàm do người dùng tự định nghĩa.
 - Khả năng lưu trữ cả dữ liệu có cấu trúc (như bảng bất động sản) và bán cấu trúc (như thông tin chi tiết mô tả hoặc cào từ web).
- Khả năng mở rộng
 - PostgreSQL cho phép mở rộng theo chiều ngang, phù hợp với các dự án lớn khi dữ liệu ngày càng tăng.
 - Nó hỗ trợ các chỉ số (indexes) nâng cao như GIN và BRIN, giúp tăng tốc các truy vấn tìm kiếm trên tập dữ liệu lớn.
- Khả năng tích hợp tốt với Python
 - Thư viện psycopg2 hỗ trợ kết nối và thao tác dữ liệu giữa Python và PostgreSQL một cách dễ dàng, cho phép nhóm xử lý, phân tích, và trích xuất dữ liệu một cách linh hoạt.
- Miễn phí và mã nguồn mở
- Các thành viên trong nhóm đã có kinh nghiệm sử dụng PostgreSQL làm việc trong môi trường thực tế ở nơi thực hiện thực tập/kiến tập.

Kết luận

Với những tính năng mạnh mẽ, hiệu suất cao và khả năng mở rộng, PostgreSQL đã đáp ứng tốt các yêu cầu của dự án về quản lý dữ liệu, đặc biệt trong bài toán dự đoán giá trị bất động sản. Việc sử dụng PostgreSQL giúp đảm bảo tính chính xác, hiệu quả và khả năng phát triển bền vững của hệ thống trong tương lai.

The diagram illustrates the data schema for the experiment. It features a central box labeled 'tablename'. Surrounding this are several other boxes, each representing a different data source and its attributes:

- thong_tin_bat_dong_san** (top left):
 - lat
 - new_usd
 - dia_chi
 - dien_tich
 - gia
 - so_phong_ngu
 - so_phong_ho
 - noi_tro_thuat
 - phat_ly
 - huong_hoa
 - huong_ban_cong
 - so_tang
 - mat_tien
 - duong_vao
 - bai_bds
 - ten_bds
 - latitude
 - longitude
- danh_dau_cac_thong_tin** (top middle):
 - id
 - new_usd
 - dia_chi
 - dien_tich
 - gia
 - so_phong_ngu
 - so_phong_ho
 - noi_tro_thuat
 - phat_ly
 - huong_hoa
 - huong_ban_cong
 - so_tang
 - mat_tien
 - duong_vao
 - bai_bds
 - new_usd
 - so_tang_danh_ghi
 - quan_huyen
 - phong_hoa
 - chi_tiet
 - ten_duong
- danh_dau_cac_thong_tin_test** (top right):
 - dia_chi
 - gia
 - new_bds
 - ten_thanh_pho
 - quan_huyen
 - phong_hoa
 - chi_tiet
 - ten_duong
 - dien_tich_chuan
 - so_phong_ngu_danh_dau
 - so_phong_ho_danh_dau
 - noi_tro_thuat_danh_dau
 - phat_ly_danh_dau
 - huong_ban_cong_danh_dau
 - so_tang_danh_dau
 - mat_tien_danh_dau
 - duong_vao_danh_dau
 - ma_to
 - ma_ph
 - ma_gn
- danh_dau_cac_thong_tin_test_merged** (bottom right):
 - dia_chi
 - gia
 - new_bds
 - ten_thanh_pho
 - quan_huyen
 - phong_hoa
 - chi_tiet
 - ten_duong
 - dien_tich_chuan
 - so_phong_ngu_danh_dau
 - so_phong_ho_danh_dau
 - noi_tro_thuat_danh_dau
 - phat_ly_danh_dau
 - huong_ban_cong_danh_dau
 - so_tang_danh_dau
 - mat_tien_danh_dau
 - duong_vao_danh_dau
 - ma_to
 - ma_ph
 - ma_gn

2.3 Xử lý dữ liệu

Az phap_ly	Az huong_nha	Az huong_ban_cong	Az so_tang	Az mat_tien	Az duong_vao	Az loi_bad	Az toa_do	123 latitude	123 longitude
Số đồ.	Không có thông tin	Không có thông tin	4 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đồ/số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Hợp đồng mua bán	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đồ/số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin		Đông - Nam	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin	Không có thông tin	Không có thông tin	2 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin	Không có thông tin	Không có thông tin	5 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số hồng	Không có thông tin	Không có thông tin	4 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đồ/số hồng	Tây	Đông - Bắc	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đồ/số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]

Xử lý thông qua View: `danh_dau_cac_thong_tin_test`. Để loại bỏ những từ không cần thiết/ lọc địa chỉ của bất động sản. Dưới đây là code xử lý:

```
-- Lấy tất cả các cột từ bảng thông tin bất động sản, chỉ gọi các cột cần thiết
ttbds.dia_chi,
ttbds.gia,
-- Phân loại kiểu bất động sản
```

WHEN *ttbds.dia_chi* **LIKE** '%Dự án%' **THEN** 'Căn hộ/Chung cư/Dự án'
ELSE 'Khác'

END AS kieu bds,


```

-- Tỉnh/Thành phố (Lấy phần sau dấu phẩy cuối cùng)
TRIM(split_part(ttbd.ds.dia_chi, ',', array_length(regexp_split_to_array(ttbd.ds.dia_chi,
','), 1))) AS tinh_thanh_pho,
-- Quận/Huyện (Lấy phần từ dấu phẩy thứ 2 từ cuối)
CASE
    WHEN array_length(regexp_split_to_array(ttbd.ds.dia_chi, ',', 1) > 1 THEN
        TRIM(split_part(ttbd.ds.dia_chi,                                ','
array_length(regexp_split_to_array(ttbd.ds.dia_chi, ','), 1) - 1))
    ELSE
        NULL
END AS quan_huyen,
-- Xã/Phường (Lấy phần từ dấu phẩy thứ 3 từ cuối)
CASE
    WHEN array_length(regexp_split_to_array(ttbd.ds.dia_chi, ',', 1) > 2 THEN
        TRIM(split_part(ttbd.ds.dia_chi,                                ','
array_length(regexp_split_to_array(ttbd.ds.dia_chi, ','), 1) - 2))
    ELSE
        NULL
END AS phuong_xa,
-- Tách toàn bộ thông tin trước dấu phẩy đầu tiên (Số nhà và tên đường)
TRIM(split_part(ttbd.ds.dia_chi, ',', 1)) AS chi_tiet,
-- Tách Đường
TRIM(split_part(ttbd.ds.dia_chi, ',', 2)) AS ten_duong,
-- Xử lý diện tích, chỉ lấy số
CAST(REGEXP_REPLACE(ttbd.ds.dien_tich, '[^0-9.]', '', 'g') AS DECIMAL(10, 2)) AS
dien_tich_chuan,
-- Xử lý số phòng ngủ
CASE
    WHEN ttbd.ds.so_phong_ngu = 'Không có thông tin' THEN NULL
    ELSE ttbd.ds.so_phong_ngu
END AS so_phong_ngu_danh_dau,
-- Xử lý số phòng vệ sinh
CASE
    WHEN ttbd.ds.so_phong_vs = 'Không có thông tin' THEN NULL
    ELSE ttbd.ds.so_phong_vs
END AS so_phong_vs_danh_dau,
-- Xử lý nội thất (Cơ bản, đầy đủ, không có thông tin)
CASE

```

```

WHEN ttbds.noi_that = 'Cơ bản' THEN 1
WHEN ttbds.noi_that = 'Đầy đủ' THEN 2
WHEN ttbds.noi_that = 'Không có thông tin' THEN NULL
ELSE 0
END AS noi_that_danh_dau,
-- Mã hóa thông tin pháp lý (Số đỏ/Số hồng, Đang chờ số, Không có thông tin)
CASE
WHEN ttbds.phap_ly = 'Số đỏ/Số hồng' THEN 0
WHEN ttbds.phap_ly = 'Đang chờ số' THEN 1
WHEN ttbds.phap_ly = 'Số đỏ' THEN 2
WHEN ttbds.phap_ly = 'Số hồng' THEN 3
WHEN ttbds.phap_ly = 'Không có thông tin' THEN NULL
ELSE 0
END AS phap_ly_danh_dau,
-- Mã hóa hướng nhà (Thêm nhiều hướng nhà chi tiết hơn)
CASE
WHEN ttbds.huong_nha = 'Tây - Bắc' THEN 1
WHEN ttbds.huong_nha = 'Đông - Nam' THEN 2
WHEN ttbds.huong_nha = 'Đông - Bắc' THEN 3
WHEN ttbds.huong_nha = 'Tây - Nam' THEN 4
WHEN ttbds.huong_nha = 'Bắc' THEN 5
WHEN ttbds.huong_nha = 'Nam' THEN 6
WHEN ttbds.huong_nha = 'Đông' THEN 7
WHEN ttbds.huong_nha = 'Tây' THEN 8
WHEN ttbds.huong_nha = 'Không có thông tin' THEN NULL
ELSE 0
END AS huong_nha_danh_dau,
-- Mã hóa hướng ban công
CASE
WHEN ttbds.huong_ban_cong = 'Tây - Bắc' THEN 1
WHEN ttbds.huong_ban_cong = 'Đông - Nam' THEN 2
WHEN ttbds.huong_ban_cong = 'Đông - Bắc' THEN 3
WHEN ttbds.huong_ban_cong = 'Tây - Nam' THEN 4
WHEN ttbds.huong_ban_cong = 'Bắc' THEN 5
WHEN ttbds.huong_ban_cong = 'Nam' THEN 6
WHEN ttbds.huong_ban_cong = 'Đông' THEN 7
WHEN ttbds.huong_ban_cong = 'Tây' THEN 8

```

```

WHEN tbds.huong_ban_cong = 'Không có thông tin' THEN NULL
ELSE 0
END AS huong_ban_cong_danh_dau,
-- Chuyển đổi số tầng
CASE
WHEN tbds.so_tang = 'Không có thông tin' THEN NULL
ELSE tbds.so_tang
END AS so_tang_danh_dau,
-- Chuyển đổi mặt tiền (có thể có giá trị hoặc là NULL)
CASE
WHEN tbds.mat_tien = 'Không có thông tin' THEN NULL
ELSE tbds.mat_tien
END AS mat_tien_danh_dau,
-- Chuyển đổi đường vào (có thể có giá trị hoặc là NULL)
CASE
WHEN tbds.duong_vao = 'Không có thông tin' THEN NULL
ELSE tbds.duong_vao
END AS duong_vao_danh_dau
FROM thong_tin_bat_dong_san tbds;

```

Kết quả:

Từ cột dia_chi sẽ tách ra được các thông tin như dưới đây để tiến hành lưu trữ thông tin liên quan đến địa chỉ:

	A-Z gia	A-Z kieu_bds	A-Z tinh_thanh_pho	A-Z quan_huyen	A-Z phuong_xa	A-Z chi_tiet	A-Z ten_duong	123 dien_tich_chuan
1	Không có thông tin	Khác	Không có mô tả ngắn	[NULL]	[NULL]	Không có mô tả ngắn		[NULL]
2	Thỏa thuận	Khác	Hà Nội	Hoài Đức	Xã An Khánh	Xã An Khánh	Hoài Đức	50
3	16,5 tỷ	Khác	Hồ Chí Minh	Quận 10	Phường 12	134	Đường Thành Thái	68
4	3,98 tỷ	Khác	Hà Nội	Hoàng Mai	Phường Thanh Trì	Đường Nguyễn Khoái	Phường Thanh Trì	30
5	Không có thông tin	Khác	Không có mô tả ngắn	[NULL]	[NULL]	Không có mô tả ngắn		[NULL]
6	3,98 tỷ	Khác	Hà Nội	Hoàng Mai	Phường Thanh Trì	Đường Nguyễn Khoái	Phường Thanh Trì	30
7	20,5 tỷ	Khác	Hồ Chí Minh	Quận 10	Phường 14	666	Đường 3/2	97
8	ha Tranj 1,65 tỷ	Căn hộ/Chung cư/Dự án	Khánh Hòa	Nha Trang	Phường Lộc Thọ	Dự án Vinpearl Beach Front Condotel	Đường Trần Phú	40
9	20,5 tỷ	Khác	Hồ Chí Minh	Quận 10	Phường 14	666	Đường 3/2	97
10	Đ. Chí Mi 11 tỷ	Căn hộ/Chung cư/Dự án	Hồ Chí Minh	Bình Thạnh	Phường 22	Dự án Sunwah Pearl	Đường Nguyễn Hữu Cánh	106

Hình 2. 8 Kết quả

Những dữ liệu khác sẽ được mã hóa để phân tích.

A-Z phap_ly	A-Z huong_nha	A-Z huong_ban_cong	A-Z so_tang	A-Z mat_tien	A-Z duong_vao	A-Z loai_bds	A-Z toa_do	123 latitude	123 longitude
Số đo.	Không có thông tin	Không có thông tin	4 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đo/ Số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Hợp đồng mua bán	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đo/ Số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin	Đông - Nam	Tây - Bắc	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin	Không có thông tin	Không có thông tin	2 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Không có thông tin	Không có thông tin	Không có thông tin	5 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số hồng	Không có thông tin	Không có thông tin	4 tầng	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đo/ Số hồng	Tây	Đông - Bắc	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]
Số đo/ Số hồng	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	Không có thông tin	1	[NULL]	[NULL]	[N]

Xử lý để mã hóa địa chỉ. Được xử lý bằng view danh_dau_thong_tin_test_merged
 Chủ yếu xử lý ở đây là việc kết hợp giữa tất cả các thông tin đã được xử lý và mã hóa địa chỉ để thực hiện phân tích.

CREATE VIEW danh_dau_cac_thong_tin_test_merged **AS**
SELECT

-- Các cột từ view ban đầu

v.dia_chi,

v.gia,

v.kieu_bds,

v.tinh_thanh_pho,

v.quan_huyen,

v.phuong_xa,

v.chi_tiet,

v.ten_duong,

v.dien_tich_chuan,

v.so_phong_ngu_danh_dau,

v.so_phong_vs_danh_dau,

v.noi_that_danh_dau,

v.phap_ly_danh_dau,

v.huong_nha_danh_dau,

v.huong_ban_cong_danh_dau,

v.so_tang_danh_dau,

v.mat_tien_danh_dau,

v.duong_vao_danh_dau,

-- Các cột mã từ bảng table_name

t.ma_tp,

t.ma_qh,

t.ma_px

FROM

danh_dau_cac_thong_tin_test v -- VIEW đã có

JOIN

tablename t -- Bảng chứa mã tỉnh, quận, xã

ON

v.tinh_thanh_pho = t.tinh_thanh_pho

AND v.quan_huyen = t.quan_huyen

AND v.phuong_xa = t.phuong_xa;

UPDATE tablename

-- Chuẩn hóa khoảng trắng: Xóa khoảng trắng thừa ở đầu/cuối và thay thế nhiều khoảng trắng thành một khoảng trắng duy nhất

UPDATE tablename

```

SET quan_huyen = TRIM(BOTH '' FROM REGEXP_REPLACE(quan_huyen, '\s+',
'', 'g'));
-- Xóa các từ không mong muốn: 'Quận', 'Huyện', 'Thị xã'
UPDATE tablename
SET      tinh_thanh_pho      =      TRIM(BOTH      '      '      FROM
REPLACE(REPLACE(REPLACE(tinh_thanh_pho, 'Tỉnh', ''), 'Thành phố', ''), 'Thị
xã', ''));
WITH CTE AS (
  SELECT
    *,
    ROW_NUMBER() OVER (
      PARTITION BY tieu_de, dia_chi
      ORDER BY id
    ) AS row_num
  FROM thông_tin_bat_dong_san ttbd
)
DELETE FROM thông_tin_bat_dong_san
WHERE id IN (
  SELECT id
  FROM CTE
  WHERE row_num > 1
);
select count(*) from thông_tin_bat_dong_san ttbd

```

2.3.2 Xử lý dữ liệu bên ngoài cơ sở dữ liệu bằng Python để lưu trữ

Việc thu thập dữ liệu từ website Batdongsan.com.vn có một điểm yếu là thiếu thông tin về tọa độ của bất động sản bị thiếu. Nên cần xử lý để thêm tọa độ vào.

- Cài đặt thư viện cần thiết:
 - plotly.express : Dùng để tạo các biểu đồ trực quan và tương tác, như biểu đồ cột, biểu đồ đường, bản đồ.
 - geopy.geocoders: Dùng để chuyển đổi địa chỉ thành tọa độ (vĩ độ, kinh độ).

```

import pandas as pd
import plotly.express as px
from geopy.geocoders import Nominatim
import psycopg2

```

Hình 2. 9 Cài đặt thư viện xử lý dữ liệu

- Kết nối vào cơ sở dữ liệu:

```

1. connect_db = psycopg2.connect (
2.     dbname="postgres",
3.     user="postgres",
4.     password="123",
5.     host="localhost",
6.     port="5432"|
7. )
8. cursor = connect_db.cursor()

```

Hình 2. 10 Kết nối cơ sở dữ liệu

- Thêm cột latitude và longitude vào bảng `thong_tin_bat_dong_san` để lưu tọa độ sau khi thêm.

```

try:
    cursor.execute("""
        ALTER TABLE thong_tin_bat_dong_san
        ADD COLUMN IF NOT EXISTS latitude DOUBLE PRECISION,
        ADD COLUMN IF NOT EXISTS longitude DOUBLE PRECISION;
    """)|
    connect_db.commit()
    print("Đã khởi tạo cột latitude và longitude.")
except Exception as e:
    print(f"Lỗi khi thêm cột: {e}")
    connect_db.rollback()

```

Hình 2. 11 Thêm cột tọa độ vào csdl

- Đánh tọa độ cho các bất động sản:

```

geolocator = Nominatim(user_agent="geoapi")
for row in rows:
    bds_id, address = row[0], row[1].replace("Dự án", "").strip()
    try:
        location = geolocator.geocode(address)
        lat, lon = (location.latitude, location.longitude) if location
    else: (None, None)

    # Cập nhật tọa độ vào cơ sở dữ liệu ngay
    cursor.execute("""
        UPDATE thong_tin_bat_dong_san
        SET latitude = %s, longitude = %s
        WHERE id = %s;
    """, (lat, lon, bds_id))
    connect_db.commit()
    print(f"Cập nhật tọa độ cho ID {bds_id}: {lat}, {lon}")
except Exception as e:
    print(f"Không thể lấy tọa độ cho địa chỉ: {address}. Lỗi: {e}")
    connect_db.rollback()

# 5. Lấy các dòng có tọa độ hợp lệ để trực quan hóa
query = "SELECT dia_chi, latitude, longitude FROM
thong_tin_bat_dong_san WHERE latitude IS NOT NULL AND longitude IS NOT
NULL;"
cursor.execute(query)

```

- Trực quan hóa những bất động sản đã được đánh dấu tọa độ trên bản đồ:

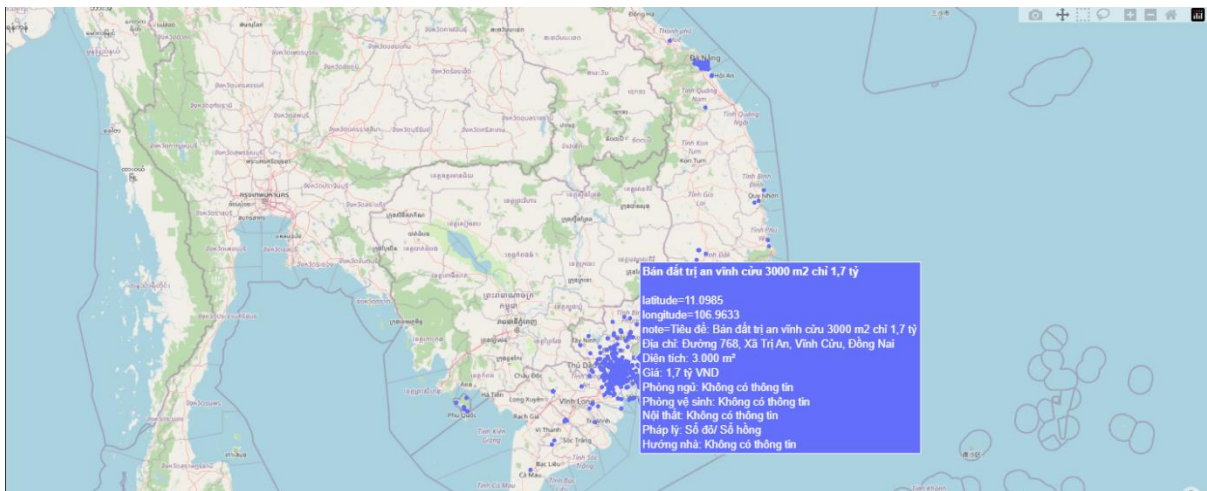
```
# 7. Trực quan hóa bản đồ với Plotly
if not df.empty:
    fig = px.scatter_mapbox(
        df,
        lat="latitude",
        lon="longitude",
        hover_name="title", # Hiển thị địa chỉ khi di chuột
        zoom=10,
        height=600
    )

    # Cài đặt kiểu bản đồ
    fig.update_layout(
        mapbox_style="open-street-map",
        title="Bản đồ bất động sản",
        margin={"r": 0, "t": 0, "l": 0, "b": 0}
    )

    # Hiển thị bản đồ
    fig.show()
else:
    print("Không có tọa độ nào để hiển thị trên bản đồ.")
```

Hình 2. 12 Trực quan hóa bất động sản

Kết quả thu được sau khi đánh tọa độ:



Hình 2. 13 Kết quả thu được khi đánh tọa độ

2.4. Chuẩn bị mô hình

2.4.1 Xử lý dữ liệu đầu vào của mô hình

Cài đặt các thư viện cần thiết

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

file_path = './TinhHuyenXa2021.xlsx'
df = pd.read_excel(file_path)

columns_to_drop = ['Tên Tiếng Anh', 'Cấp']

df = df.drop(columns_to_drop, axis=1)
```

Hình 2. 14 Cài đặt các thư viện cần thiết

File “TinhHuyenXa2021.xlsx ” Chứa tất cả các thông tin về tỉnh, thành phố , quận huyện phường, xã ở Việt Nam và đã bao gồm đánh số thứ tự như hình:

Province	Ma_TP	District	Ma_QH	Ward/Commune	Ma_PX	Cấp
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Phúc Xá	00001	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Trúc Bạch	00004	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Vĩnh Phúc	00006	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Cống Vị	00007	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Liễu Giai	00008	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Nguyễn Trung Trực	00010	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Quán Thánh	00013	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Ngọc Hà	00016	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Điện Biên	00019	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Đội Cấn	00022	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Ngọc Khánh	00025	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Kim Mã	00028	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Giảng Võ	00031	Phường
Thành phố Hà Nội	01	Quận Ba Đình	001	Phường Thành Công	00034	Phường
Thành phố Hà Nội	01	Quận Hoàn Kiếm	002	Phường Phúc Tân	00037	Phường
Thành phố Hà Nội	01	Quận Hoàn Kiếm	002	Phường Đồng Xuân	00040	Phường

Hình 2. 15 Dữ liệu Địa phương

File “Data_raw.csv” Chứa dữ liệu trong cơ sở dữ liệu được xuất ra để thực hiện việc sử dụng học máy

Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	84					4			Have certificate	8.6
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	60					5				7.5
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	90	6	13	Đông - Bắc	Đông - Bắc	5			Sale contract	8.9
Đường Nguyễn Văn Khối, Phường 11, Gò Vấp, Hồ Chí Minh	54		3.5	Tây - Nam	Tây - Nam	2	2	3	Have certi Full	5.35
Đường Quang Trung, Phường 8, Gò Vấp, Hồ Chí Minh	92			Đông - Na	Đông - Na	2	4	4	Have certi Full	6.9
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	91	7	13	Tây - Bắc					Have certificate	9.8
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	64	4				5				7.2
Dự án Him Lam Thường Tín, Huyện Thường Tín, Hà Nội	74	5	18	Nam	Nam	5	4	5	Have certificate	9.9
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	48					5	6		Basic	5.7
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	91	7				5			Have certificate	9.5
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	80	5	13	Đông - Nam		5	6	6	Have certificate	10
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	70	6	13	Đông - Nam		5	6			9.8
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	60					5				8.9
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	91								Have certificate	8.6
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	96	6	13	Tây - Bắc		4			Sale contract	9.6
Đường Phạm Văn Chiêu, Phường 14, Gò Vấp, Hồ Chí Minh	30			Đông - Nam		2	2	2	Have certi Full	3.79
Dự án Khu đô thị Phương Đông, Xã Đông Xá, Văn Điển	300	15	32	Đông	Nam	3	6	3	Have certi Full	9.4
Dự án The Crown - Vinhomes Ocean Park 3, Xã Nghĩa	91	7	17	Đông - Na	Đông - Na				Basic	10
Dự án The Empire - Vinhomes Ocean Park 2, Xã Long	65	5	20	Đông - Na	Đông - Na	5	5	4	Sale contract	7.8

Hình 2. 16 File dữ liệu gốc

Tiến hành xử lý dữ liệu của file “TinhHuyenXa2021.xlsx ”

- Loại bỏ thông tin không cần thiết

```
df['Ward/Commune'] = df['Ward/Commune'].str.replace(r"^(Xã |Phường
|Thị trấn )", "", regex=True)
df['District'] = df['District'].str.replace(r"^(Quận |Huyện |Thị
xã |Thành phố )", "", regex=True)
df['Province'] = df['Province'].str.replace(r"^(Tỉnh |Thành phố
)", "", regex=True)|
```

Hình 2. 17 Tối giản dữ liệu

- Thêm vào df và kết hợp với file “Data_raw.csv” để thực hiện việc xử lý dữ liệu:

```
file_name = "TinhHuyenXa.csv"
df.to_csv(file_name, index=False, encoding='utf-8-sig')
housing_data = pd.read_csv('Data_raw.csv')
location_data = pd.read_csv('TinhHuyenXa.csv')

merged_data = pd.merge(housing_data, location_data, how='left',
                        on=['Province', 'District',
'Ward/Commune'])
```

Hình 2. 18 Kết hợp dữ liệu

- Tiến hành xử lý:

```
categorical_columns = df.select_dtypes(include=[object]).columns
for col in categorical_columns:
    df[col].fillna(df[col].mode()[0], inplace=True)
```

```
columns_to_drop = ['Ward/Commune', 'District', 'Province']

df = df.drop(columns_to_drop, axis=1)
```

Hình 2. 19 Tiến hành xử lý

2.4.2 Thực hiện mô hình học máy

Trong báo cáo này, mô hình học máy được sử dụng để phân tích và dự đoán giá trị thẩm định bất động sản dựa trên các yếu tố như diện tích, số phòng ngủ, số tầng và các đặc điểm khác. Sau khi tìm hiểu thì nhóm tìm ra được hai mô hình có thể được sử dụng bao gồm : RandomForest , LinearRegression.

2.4.3 Lựa chọn mô hình

Để lựa chọn mô hình phù hợp ta sử dụng phương pháp MAPE. Đây là chỉ số dùng để đo lường độ sai lệch giữa giá trị dự đoán và giá trị thực tế dưới dạng tỷ lệ phần trăm. Cụ thể, nó tính toán sai số tuyệt đối trung bình giữa giá trị thực tế và giá trị dự đoán, sau đó chia cho giá trị thực tế, rồi nhân với 100 để chuyển thành phần trăm.

- MAPE RandomForest cho ra kết quả 0.2029006450221886 trên tập dữ liệu.

```
> from sklearn.ensemble import RandomForestRegressor
> from sklearn.metrics import mean_absolute_percentage_error

> model_RFR = RandomForestRegressor(n_estimators=10)
> model_RFR.fit(X_train, Y_train)
> Y_pred = model_RFR.predict(X_valid)

> mean_absolute_percentage_error(Y_valid, Y_pred)

[5]
... 0.2029006450221886
```

Hình 2. 20 MAPE RandomForestRegressor

- MAPE LinearRegression cho ra kết quả 0.38123136698275756 trên tập dữ liệu.

```
> from sklearn.linear_model import LinearRegression

> model_LR = LinearRegression()
> model_LR.fit(X_train, Y_train)
> Y_pred = model_LR.predict(X_valid)

> print(mean_absolute_percentage_error(Y_valid, Y_pred))

[5]
... 0.38123136698275756
```

Hình 2. 21 MAPE LinearRegression

Chỉ số MAPE giúp phản ánh mức độ sai lệch của mô hình theo tỷ lệ phần trăm, từ đó dễ dàng so sánh và đánh giá hiệu quả của các mô hình khác nhau.

Với mô hình Random Forest Regressor: Kết quả MAPE là 20.29%, cho thấy mô hình có khả năng dự đoán tốt trên tập dữ liệu, với độ sai lệch chỉ hơn 20%.

Với mô hình Linear Regression: Kết quả MAPE là 38.12%, cao hơn đáng kể so với Random Forest, điều này cho thấy mô hình tuyến tính đơn giản không phù hợp bằng Random Forest trong việc dự đoán dữ liệu phức tạp.

Từ các kết quả trên, nhóm nhận thấy rằng mô hình Random Forest Regressor có hiệu năng tốt hơn so với Linear Regression khi xét trên chỉ số MAPE, do đó mô hình này được ưu tiên lựa chọn để tiếp tục cải thiện và tối ưu hóa.

Tiểu kết chương 2

Quá trình nghiên cứu và thực hiện xây dựng mô hình dự đoán giá trị thẩm định bất động sản được tiến hành qua các giai đoạn chặt chẽ từ thu thập dữ liệu, tiền xử lý, đến xây dựng và đánh giá mô hình học máy, đồng thời tích hợp kết quả vào thiết kế website hỗ trợ người dùng.

Thu thập dữ liệu

Dữ liệu bất động sản được thu thập từ các nguồn trực tuyến, sử dụng công cụ Selenium để tự động hóa quá trình trích xuất. Sau khi cào thành công, dữ liệu được tổ chức và lưu trữ trong cơ sở dữ liệu PostgreSQL nhằm đảm bảo tính toàn vẹn, đồng bộ và tiện lợi khi sử dụng cho các bước tiếp theo.

Tiền xử lý dữ liệu

Các bước tiền xử lý dữ liệu được thực hiện bao gồm:

- **Xử lý giá trị thiếu:** Điền giá trị trung vị hoặc các giá trị thay thế hợp lý vào những cột bị thiếu.
- **Biến đổi dữ liệu dạng logarit:** Áp dụng log-transform để giảm độ lệch phân phối, giúp cải thiện độ chính xác của mô hình.
- **Mã hóa các biến danh mục:** Sử dụng phương pháp mã hóa one-hot hoặc label encoding để biến các cột danh mục thành dạng số, tương thích với thuật toán học máy.

Xây dựng mô hình

Nhóm đã thử nghiệm và so sánh hiệu năng của các mô hình học máy như **Random Forest Regressor** và **Linear Regression**.

Hiệu năng của các mô hình được đánh giá dựa trên hai chỉ số:

- **MAPE (Mean Absolute Percentage Error):** Đánh giá độ sai lệch trung bình giữa giá trị thực tế và dự đoán.
- **R² (R-squared):** Đo lường mức độ phù hợp của mô hình với dữ liệu.

Kết quả:

Mô hình Random Forest Regressor với các tham số tối ưu đạt được độ chính xác cao nhất với **MAPE = 0.2029** và **R² = 0.6789**, vượt trội so với Linear Regression.

Random Forest tỏ ra hiệu quả hơn nhờ khả năng xử lý tốt dữ liệu dạng số và danh mục, đồng thời giảm thiểu tình trạng overfitting nhờ cơ chế lấy mẫu ngẫu nhiên và xây dựng nhiều cây quyết định.

CHƯƠNG III. TRIỂN KHAI MÔ HÌNH TỐI ƯU VÀ XÂY DỰNG WEBSITE

3.1 Tối ưu mô hình RandomForestRegressor

- Import thư viện máy học

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import r2_score
```

Hình 3. 1 Chương 3 – import thư viện

- Chuyển đổi logarit dữ liệu

```
df['Price'] = np.log1p(df['Price'])
df['Frontage'] = np.log1p(df['Frontage'])
```

Hình 3. 2 Chương 3 – chuyển đổi logarit dữ liệu

Chuyển đổi logarit của cột Price và Frontage để giảm độ lệch (skewness) trong dữ liệu.

- Chuyển đổi dữ liệu dạng chuỗi

```
categorical_features = ['Ma_TP', 'Ma_QH', 'Ma_PX']
df[categorical_features] = df[categorical_features].astype(str)
```

Hình 3. 3 Chương 3 – Chuyển đổi dữ liệu

- Tách dữ liệu thành đầu vào và đầu ra

```
X = df.drop(['Price'], axis=1)
Y = df['Price']
```

Hình 3. 4 Chương 3 – tách dữ liệu

- Chia dữ liệu huấn luyện và kiểm tra

```
X_train, X_valid, Y_train, Y_valid = train_test_split(
    X, Y, train_size=0.8, test_size=0.2, random_state=42
)
```

Hình 3. 5 Chương 3 – Xử lý dữ liệu

- Định nghĩa tham số và mô hình RandomForest.

```
param_grid_rf = {'n_estimators': [100, 200, 300, 400]}
rf_model = RandomForestRegressor(random_state=42, max_depth=20, min_samples_split=2, min_s
```

Hình 3. 6 Chương 3 – Định nghĩa tham số và mô hình

- Tìm kiếm tham số tốt nhất (GridSearch)

```
grid_search_rfclass = GridSearchCV(
    estimator=rf_model,
    param_grid=param_grid_rf,
    cv=5,
    scoring='neg_mean_squared_error',
    verbose=2,
    n_jobs=-1
)
```

Hình 3. 7 Chương 3 – tìm kiếm tham số tối ưu

- Huấn luyện mô hình

```
grid_search_rfclass.fit(X_train, Y_train)
```

Hình 3. 8 Chương 3 – huấn luyện

- Đánh giá mô hình:

```
preds_rf = rf_model_load.predict(X_valid)
rf_r2_score = r2_score(Y_valid, preds_rf)
print(f"RandomForest R2 Score: {rf_r2_score}")
```

Hình 3. 9 Chương 3 – đánh giá mô hình

Sử dụng các tham số tối ưu để chạy mô hình dự đoán Kết quả mô hình Random Forest đạt giá trị R^2 là **75.5%**, cho thấy mô hình có khả năng giải thích được **75.5%** phương sai của biến mục tiêu (giá bất động sản) dựa trên các biến đầu vào. Đây là một kết quả đáng khích lệ trong bối cảnh các bài toán dự đoán giá bất động sản vốn có nhiều yếu tố ảnh hưởng phức tạp.

3.2 Xây dựng website phân tích bất động sản

3.2.1 Mục tiêu thiết kế

Mục tiêu chính của website là trở thành một công cụ đáng tin cậy, giúp người dùng đưa ra các quyết định đầu tư và giao dịch bất động sản thông minh. Điều này được thực hiện thông qua việc cung cấp thông tin minh bạch, phân tích chuyên sâu và các công cụ hỗ trợ cá nhân hóa. Từ giao diện thân thiện, tốc độ tải trang nhanh đến các công cụ hỗ trợ đa dạng, mọi yếu tố đều được tối ưu hóa để đáp ứng nhu cầu của người dùng. Cuối cùng, Website hướng đến việc ứng dụng các công nghệ tiên tiến học máy (Machine Learning), và phân tích dữ liệu lớn (Big Data). Các công nghệ này giúp cải thiện khả năng dự đoán, phân tích và cá nhân hóa thông tin cho từng người dùng.

3.2.2 Yêu cầu thiết kế Website

Giao diện của website cần được thiết kế đơn giản, trực quan và dễ sử dụng để đảm bảo rằng mọi người, bao gồm cả những người không am hiểu về công nghệ, đều có thể truy cập và thao tác dễ dàng. Các yếu tố cần chú trọng gồm:

- **Bố cục hợp lý:** Các thông tin chính như giá bất động sản, xu hướng thị trường, và công cụ phân tích cần được bố trí ở vị trí dễ tìm, giúp người dùng nhanh chóng tiếp cận.
- **Thiết kế đáp ứng (Responsive Design):** Website phải tương thích với nhiều thiết bị khác nhau, bao gồm máy tính, điện thoại thông minh và máy tính bảng, để tối ưu hóa trải nghiệm người dùng.
- **Tốc độ tải trang nhanh:** Thời gian tải trang ngắn giúp người dùng không mất kiên nhẫn.

Website cần cung cấp các công cụ phân tích mạnh mẽ và chính xác, đáp ứng nhu cầu tìm kiếm và so sánh của người dùng. Một số tính năng cần thiết bao gồm:

- **Phân tích giá cả:** Hiển thị giá bất động sản tại từng khu vực, so sánh giá theo khu vực và dự báo xu hướng giá cả trong tương lai.
- **Phân tích vị trí:** Cung cấp bản đồ tương tác, cho phép người dùng xem vị trí bất động sản cùng các tiện ích xung quanh như trường học, bệnh viện, giao thông công cộng.
- **Phân tích dựa trên tiêu chí cá nhân hóa:** Người dùng có thể nhập các tiêu chí cụ thể (diện tích, giá tiền, khu vực), từ đó hệ thống sẽ đưa ra các gợi ý phù hợp nhất.

3.2.3 Công nghệ sử dụng

3.2.3.1 Về ngôn ngữ lập trình và framework sử dụng chính

Bảng 3. 1 Các ngôn ngữ sử dụng trong website

STT	Ngôn ngữ/framework	Giải thích
1	Python	Python là một ngôn ngữ lập trình bậc cao, dễ học và sử dụng, nổi bật với cú pháp rõ ràng và thư viện phong phú. Nó thường được sử dụng trong các lĩnh vực như phân tích dữ liệu, trí tuệ nhân tạo, phát triển web, và tự động hóa. Python hỗ trợ lập trình hướng đối tượng, lập trình hàm, và lập trình thủ tục.
2	HTML	HTML (HyperText Markup Language) là ngôn ngữ đánh dấu cơ bản dùng để tạo ra cấu trúc và nội dung của trang web. HTML sử dụng các thẻ để xác định các phần tử như tiêu đề,

		đoạn văn, hình ảnh, liên kết, và các thành phần khác trên một trang web.
3	CSS	CSS (Cascading Style Sheets) là ngôn ngữ dùng để định dạng và trang trí cho các trang web. CSS cho phép điều chỉnh màu sắc, phông chữ, bố cục, và các hiệu ứng trực quan khác, giúp trang web trở nên hấp dẫn và dễ nhìn hơn.
4	JavaScript	JavaScript là một ngôn ngữ lập trình kịch bản chủ yếu được sử dụng trong phát triển web để tạo ra các tính năng tương tác trên trang web. Nó cho phép cập nhật động nội dung của trang mà không cần phải tải lại toàn bộ trang, như tạo ra các hiệu ứng động, xử lý sự kiện, và giao tiếp với server.
5	Streamlit	Streamlit là một framework Python mã nguồn mở, cho phép xây dựng ứng dụng web tương tác dễ dàng và nhanh chóng. Nó chủ yếu được sử dụng để tạo các ứng dụng phân tích dữ liệu, trực quan hóa dữ liệu, và triển khai mô hình học máy mà không cần nhiều kiến thức về phát triển web. Các ứng dụng Streamlit có thể được phát triển bằng cách viết mã Python đơn giản mà không cần phải lo lắng về HTML, CSS hay JavaScript.

3.2.3.2 Thư viện sử dụng

Bảng 3. 2 Các thư viện sử dụng trong website

STT	Thư viện	Giải thích
1	NumPy	Thư viện Python hỗ trợ tính toán số học, mảng đa chiều và các hàm toán học hiệu quả.
2	Pandas	Thư viện Python giúp xử lý và phân tích dữ liệu dạng bảng (DataFrame) dễ dàng.
3	Plotly	Thư viện Python tạo đồ họa tương tác và trực quan hóa dữ liệu mạnh mẽ.
4	Requests	Thư viện Python đơn giản hóa việc gửi yêu cầu HTTP và xử lý phản hồi.
5	Folium	Thư viện Python giúp tạo bản đồ tương tác với dữ liệu địa lý.
6	Openpyxl	Thư viện Python dùng để đọc và ghi tệp Excel (xlsx).

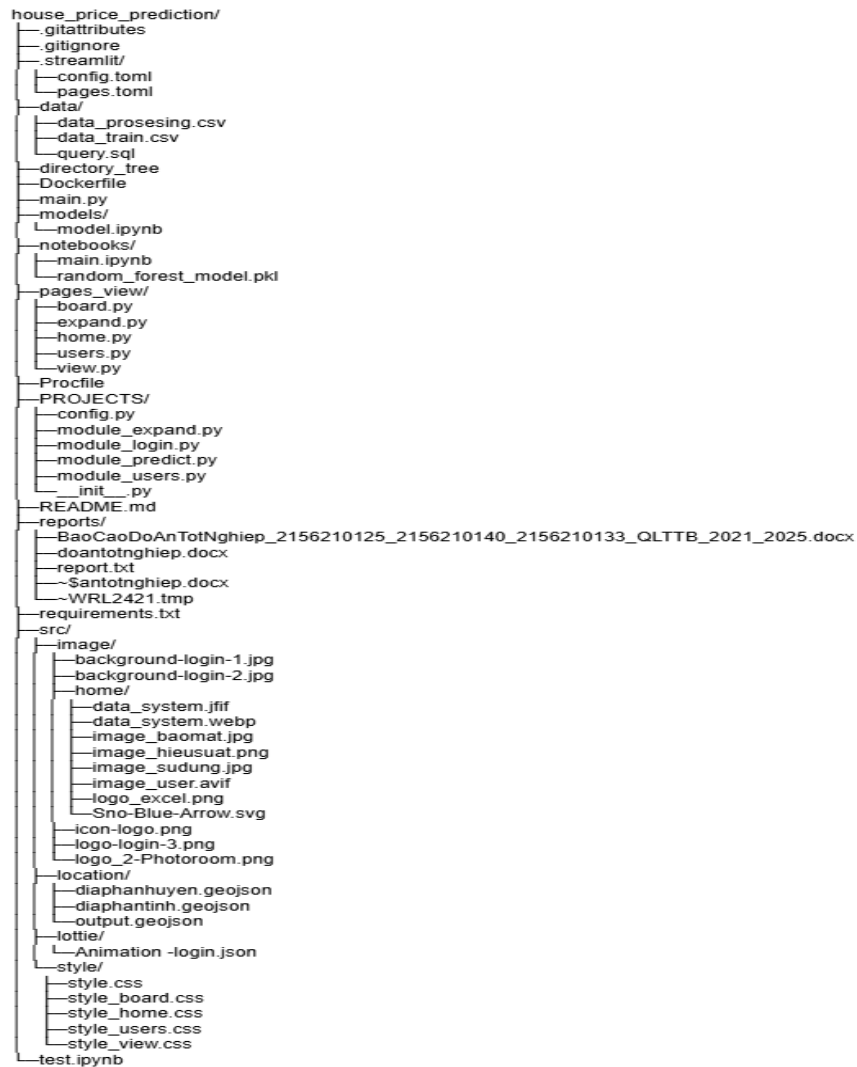
7	JSON	Định dạng dữ liệu nhẹ và thư viện Python hỗ trợ làm việc với dữ liệu JSON.
---	------	--

3.2.4. Phương pháp triển khai và quy trình thực hiện

Sau khi triển khai và thử nghiệm mô hình, nhận thấy mô hình RandomForestRegressor có hiệu quả cao trong việc dự đoán giá bất động sản nhóm tiếp tục vào bước xây dựng website để tạo giao diện cho người dùng thao tác. Quy trình đó gồm các bước xây dựng như sau:

3.2.4.1. Tải dữ liệu và các module liên quan vào dự án

Trước khi bắt tay vào việc xây dựng website, nhóm tiến hành tạo cấu trúc thư mục dự án để thuận lợi cho việc triển khai về sau. Tiếp đó, nhóm sẽ tải các dữ liệu dạng json về vị trí địa lý, tọa độ cũng như là mẫu bản đồ từ nguồn uy tín để làm cơ sở thiết kế các bản đồ giá cả.



Hình 3. 10 Cấu trúc thư mục dự án

3.2.4.2 Xây dựng các module

- Module view: bao gồm các file trong thư mục page_view như board.py, expand.py, home.py, users.py, view.py
- Module model: bao gồm các file trong thư mục PROJECTS như config.py, module_expand.py, module_login.py, module_predict.py, module_users.py
- Module cho framework streamlit: bao gồm các file trong thư mục .streamlit là config.toml và pages.toml.

3.2.4.3 Xây dựng chức năng website

Ở phần này, nhóm tiến hành sử dụng kết hợp cơ sở dữ liệu và framework streamlit để tiến hành xây dựng backend và frontend cho website. Các module chức năng bao gồm :

- Đăng nhập và đăng ký tài khoản.

- Xem dữ liệu bất động sản theo khu vực.
- Dự đoán giá bất động sản và gợi ý các bất động sản liên quan tương ứng.
- Thống kê mô tả và so sánh bất động sản theo các yếu tố
- Quản lý tài khoản người dùng.
- Gửi báo cáo về biến động giá bất động sản qua email.

Dưới đây là các hàm sử dụng trong quá trình xây dựng các chức năng website được xếp theo module:

```
File: pages_view\board.py
- Function: load_data_from_csv
- Function: load_region_data_json
- Function: map_vietnam_chart_price
- Function: filter_data
- Function: Pie_chart_by_params
- Function: box_plot_by_params
- Function: house_board_predict
- Function: run_board
File: pages_view\home.py
- Function: main
- Function: arrow_down_design
- Function: arrow_up_design
- Function: create_sidebar_toc
- Function: sidebar_design
- Function: main_container_design
File: pages_view\users.py
- Function: check_password
- Function: ui_info
- Function: sidebar_ui
- Function: change_password_user
- Function: display_user_change
- Function: account_delete
- Function: user
- Function: load_data_for_user
File: pages_view\view.py
- Function: create_map
- Function: predict_house_price
- Function: ui_predict
- Function: ui_info_detail
- Function: bar_chart_compare_price
- Function: bar_chart_compare_area
- Function: bar_chart_compare_bedrooms_and_bathrooms
- Function: bar_chart_compare_fronstage_and_accessRoad
- Function: map_chart_show_params
- Function: sidebar
- Function: run_view
```

Hình 3. 11 Các hàm trong module view

```

File: PROJECTS\config.py
- Function: create_db_pool
- Function: connect_to_mysql
- Function: check_connection
- Function: reconnect_if_needed
- Function: is_connection_alive
- Function: show_confirmation_dialog
- Function: get_relative_file_path
- Function: social_media_show
- Function: add_sidebar_footer
- Function: show_expander_sidebar
- Function: format_number
- Function: query_to_dataframe
- Function: show_province
File: PROJECTS\module_login.py
- Function: load_lottiefile
- Function: load_lottieurl
- Function: check_user_access
- Function: select_info_user
- Function: update_user_by_users
- Function: update_user_by_admin
- Function: login
File: PROJECTS\module_predict.py
- Function: query_to_dataframe
- Function: load_data
- Function: load_data_from_csv
- Function: load_region_data_json
- Function: load_model
File: PROJECTS\module_users.py
- Function: load_data_for_user
- Function: change_password
- Function: change_profile
- Function: delete_user

```

Hình 3. 12 Các hàm trong module model

3.2.5 Chức năng website

3.2.5.1 . Đăng nhập và đăng ký tài khoản

Chức năng:

- Cho phép người dùng tạo tài khoản mới bằng cách cung cấp thông tin cá nhân (như tên, email, mật khẩu).
- Người dùng đã có tài khoản có thể đăng nhập để sử dụng các chức năng nâng cao của website.

Mục đích:

- Bảo mật và phân quyền cho từng người dùng.
- Lưu thông tin cá nhân và lịch sử sử dụng của người dùng.

Tính năng liên quan:

- Reset mật khẩu khi quên.
- Xác thực email hoặc số điện thoại.

3.2.5.2 Dự đoán giá bất động sản và gợi ý các bất động sản liên quan tương ứng

Chức năng:

- Sử dụng mô hình học máy (machine learning) để dự đoán giá bất động sản dựa trên các yếu tố như vị trí, diện tích, tiện ích, thời điểm.
- Gợi ý các bất động sản tương tự trong khu vực hoặc phù hợp với sở thích người dùng.

Mục đích:

- Giúp người dùng hiểu được giá trị thực tế của bất động sản trước khi giao dịch.
- Tăng khả năng khám phá và lựa chọn bất động sản liên quan.

Tính năng liên quan:

- Phân tích dữ liệu lịch sử giá để cải thiện độ chính xác của dự đoán.
- Cung cấp biểu đồ giá dự đoán và giá thị trường.

3.2.5.3 Xem dữ liệu bất động sản theo khu vực

Chức năng:

- Hiển thị thông tin bất động sản (giá cả, diện tích, vị trí, tiện ích...) theo các khu vực như tỉnh, thành phố, quận/huyện.
- Tích hợp bộ lọc giúp người dùng tìm kiếm theo tiêu chí (giá, loại hình bất động sản, diện tích...).

Mục đích:

- Cung cấp thông tin trực quan để người dùng dễ dàng tìm kiếm bất động sản phù hợp.
- Hỗ trợ người dùng ra quyết định mua bán, đầu tư.

Tính năng liên quan:

- Tích hợp bản đồ để xem vị trí trực tiếp.
- Lọc nâng cao (theo số phòng, ..).

3.2.5.4 Thống kê mô tả và so sánh bất động sản theo các yếu tố

Chức năng:

- Cung cấp thống kê về các yếu tố như giá trung bình, diện tích trung bình, xu hướng tăng/giảm giá.
- So sánh bất động sản theo các yếu tố khác nhau (khu vực, loại hình, tiện ích...).

Mục đích:

- Hỗ trợ người dùng trong việc phân tích dữ liệu để đưa ra quyết định đầu tư chính xác.
- Cung cấp bức tranh tổng quan về thị trường bất động sản.

Tính năng liên quan:

- Biểu đồ (cột, đường, tròn) để minh họa trực quan.
- So sánh nhiều bất động sản cạnh nhau trên cùng giao diện.

3.2.5.5 Quản lý tài khoản người dùng

Chức năng:

- Cho phép người dùng cập nhật thông tin cá nhân, đổi mật khẩu, quản lý lịch sử giao dịch, và tùy chỉnh thông báo.
- Admin có thể quản lý danh sách người dùng, xóa tài khoản hoặc khóa tài khoản vi phạm.

Mục đích:

- Tăng tính cá nhân hóa trải nghiệm của người dùng.
- Đảm bảo quản trị hệ thống hiệu quả và an toàn.

Tính năng liên quan:

- Cài đặt thông báo qua email/SMS.
- Quản lý quyền hạn người dùng.

3.2.5.6 Gửi báo cáo về biến động giá bất động sản qua email

Chức năng:

- Tự động gửi báo cáo định kỳ qua email về các biến động giá bất động sản trong khu vực người dùng quan tâm.
- Báo cáo có thể bao gồm thông tin về xu hướng giá, bất động sản nổi bật hoặc dự báo giá trong tương lai.

Mục đích:

- Cập nhật thông tin thị trường nhanh chóng và tiện lợi.
- Giúp người dùng nắm bắt cơ hội đầu tư kịp thời.

Tính năng liên quan:

- Tùy chỉnh tần suất gửi báo cáo (hàng ngày, hàng tuần, hàng tháng).
- Báo cáo dưới dạng PDF hoặc file Excel đính kèm.

Tiểu kết chương 3

Trong Chương 3, nhóm đã tập trung vào hai nhiệm vụ chính: xây dựng mô hình tối ưu RandomForestRegressor và phát triển website dự đoán giá bất động sản.

Đầu tiên, nhóm đã tiến hành xây dựng mô hình dự đoán giá bất động sản sử dụng thuật toán RandomForestRegressor. Thông qua việc tinh chỉnh các siêu tham số bằng GridSearchCV, mô hình đã đạt được hiệu suất tốt với $R^2 = 75.5\%$, phản ánh khả năng mô hình trong việc giải thích mối quan hệ giữa các đặc trưng và giá bất động sản. Bên cạnh đó, các bước tiền xử lý dữ liệu như điền giá trị khuyết, biến đổi log-transform, và xử lý đặc trưng phân loại đã được thực hiện để đảm bảo tính ổn định và hiệu quả của mô hình.

Tiếp theo, nhóm đã phát triển một website dự đoán giá bất động sản dựa trên mô hình đã được xây dựng. Website này cung cấp các chức năng chính, bao gồm nhập thông tin đặc trưng của bất động sản, hiển thị giá dự đoán, và gợi ý các bất động sản liên quan. Đồng thời, nhóm đã đảm bảo tích hợp đầy đủ các thành phần cần thiết, từ giao diện người dùng (frontend) đến xử lý dữ liệu và logic dự đoán (backend).

Kết quả đạt được ở Chương 3 không chỉ góp phần xây dựng một mô hình dự đoán giá chính xác mà còn tạo nền tảng ứng dụng thực tiễn thông qua website, hỗ trợ người dùng dễ dàng tiếp cận và khai thác thông tin về thị trường bất động sản. Kết quả này sẽ là cơ sở để đánh giá, kiểm nghiệm và hoàn thiện hệ thống trong các chương tiếp theo.

KẾT LUẬN

1. Kết quả đạt được

Thông qua quá trình thực hiện đồ án, nhóm đã đạt được một số kết quả quan trọng. Về cơ sở lý thuyết, nhóm đã phân tích sâu các đặc điểm của thị trường bất động sản Việt Nam, tập trung vào các yếu tố như vị trí địa lý, hạ tầng, pháp lý và các tác động kinh tế vĩ mô. Đồng thời, nhóm cũng đã nghiên cứu hai mô hình học máy phổ biến là Linear Regression và RandomForestRegressor, làm nền tảng lý thuyết để lựa chọn phương pháp dự đoán phù hợp.

Trong quá trình thực nghiệm, nhóm đã tiến hành thu thập dữ liệu bất động sản từ các nguồn trực tuyến, sử dụng công cụ Selenium để tự động hóa quá trình trích xuất và lưu trữ dữ liệu vào cơ sở dữ liệu PostgreSQL. Các bước tiền xử lý dữ liệu đã được thực hiện đầy đủ, bao gồm xử lý giá trị thiếu, biến đổi logarit và mã hóa các biến danh mục. Mô hình RandomForestRegressor được đánh giá là hiệu quả nhất với các chỉ số $MAPE = 0.2029$ và $R^2 = 0.6789$, vượt trội hơn so với Linear Regression.

Về ứng dụng thực tiễn, nhóm đã xây dựng một mô hình dự đoán giá bất động sản với thuật toán RandomForestRegressor được tinh chỉnh, đạt hiệu suất $R^2 = 75.5\%$. Dựa trên mô hình này, nhóm phát triển một website hỗ trợ người dùng, cho phép nhập thông tin bất động sản, hiển thị giá dự đoán và gợi ý các bất động sản liên quan.

2. Ưu điểm và hạn chế

Về ưu điểm:

Đồ án có một số ưu điểm nổi bật. Thứ nhất, cơ sở lý thuyết được xây dựng vững chắc, bao gồm phân tích đặc điểm thị trường bất động sản và các thuật toán học máy, đảm bảo tính định hướng cho toàn bộ quá trình thực nghiệm. Thứ hai, mô hình RandomForestRegressor được tinh chỉnh kỹ lưỡng, cho thấy khả năng dự đoán chính xác, đặc biệt phù hợp với dữ liệu bất động sản có tính phi tuyến và biến động cao. Thứ ba, việc phát triển website dự đoán giúp chuyển giao kết quả đồ án thành ứng dụng thực tế, tạo điều kiện để người dùng dễ dàng tiếp cận thông tin về giá bất động sản.

Về hạn chế :

Bên cạnh những kết quả đạt được, nghiên cứu vẫn tồn tại một số hạn chế. Dữ liệu bất động sản thu thập được chủ yếu từ các nguồn trực tuyến, dẫn đến việc thiếu thông tin chi tiết về một số yếu tố quan trọng như tình trạng pháp lý hay chất lượng hạ tầng. Mặc dù mô hình RandomForestRegressor đạt hiệu suất khá tốt, nhưng giá trị $R^2 = 75.5\%$ vẫn chưa thực sự cao, cho thấy khả năng mô hình chưa giải thích hoàn toàn các yếu tố ảnh hưởng đến giá bất động sản. Ngoài ra, website hiện tại mới tập trung vào chức năng dự đoán giá mà chưa tích hợp nhiều tính năng hỗ trợ phân tích chuyên sâu hoặc cập nhật dữ liệu theo thời gian thực.

3. Hướng phát triển

Để nâng cao chất lượng và khả năng ứng dụng thực tiễn của dự án, nhóm sẽ tập trung vào các hướng phát triển sau:

Ứng dụng thời gian thực:

Nhằm tăng tính tương tác và tiện ích của hệ thống, nhóm sẽ tích hợp cơ chế cập nhật dữ liệu bất động sản theo thời gian thực. Điều này yêu cầu xây dựng một pipeline xử lý dữ liệu trực tiếp từ các nguồn đáng tin cậy, như các trang web bất động sản và cơ sở dữ liệu quy hoạch. Ngoài ra, hệ thống sẽ được tối ưu để phản hồi nhanh chóng các truy vấn từ người dùng, đảm bảo thông tin dự báo luôn được cập nhật và chính xác.

Ứng dụng Big Data:

Việc áp dụng các công nghệ Big Data sẽ giúp hệ thống mở rộng khả năng xử lý và phân tích dữ liệu từ các nguồn phong phú, bao gồm dữ liệu lịch sử, thông tin quy hoạch và các yếu tố kinh tế vĩ mô. Sử dụng các nền tảng như Apache Spark hoặc Hadoop sẽ hỗ trợ xử lý khối lượng lớn dữ liệu, đồng thời cải thiện hiệu suất dự đoán của mô hình bằng cách khai thác thêm các mối quan hệ tiềm ẩn trong dữ liệu.

Tối ưu hóa chi phí và kinh phí dự án:

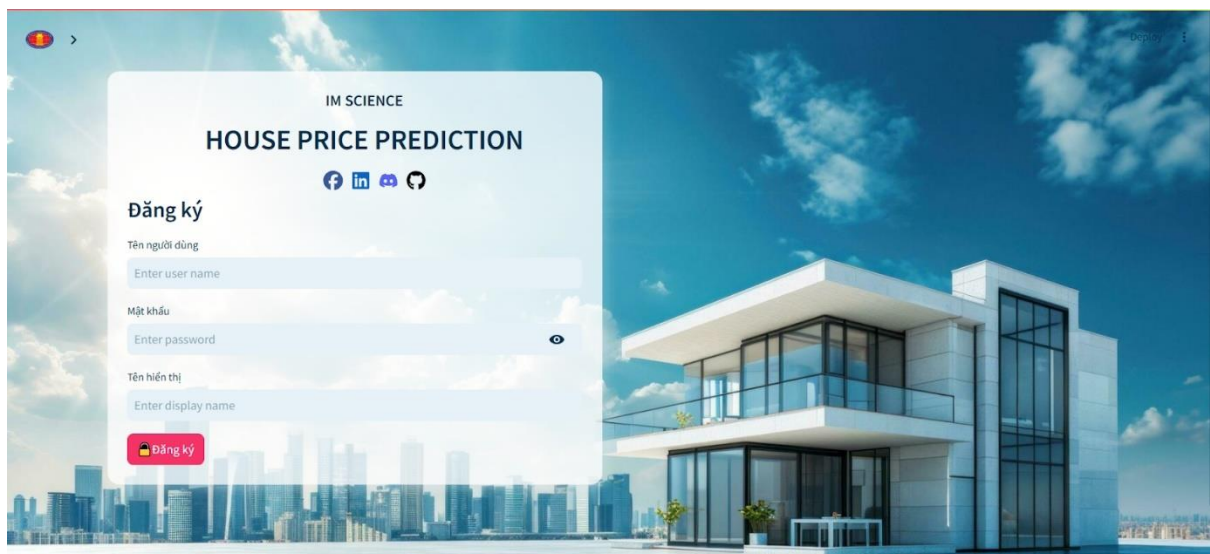
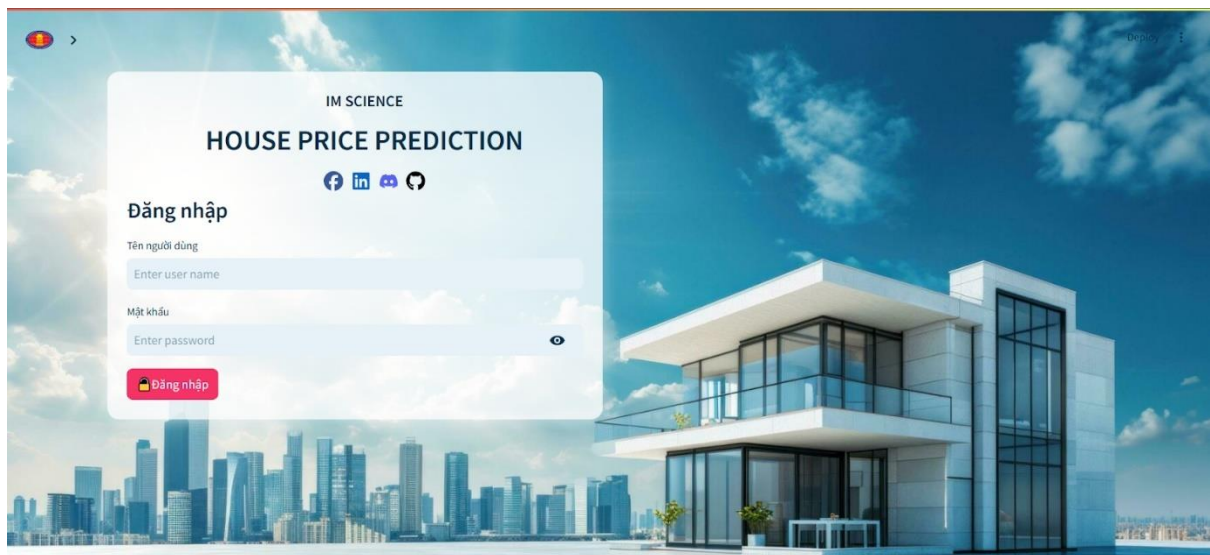
Do các công nghệ như Big Data và ứng dụng thời gian thực có thể đòi hỏi kinh phí cao, nhóm sẽ xây dựng kế hoạch triển khai chi tiết, tập trung vào việc sử dụng các dịch vụ đám mây linh hoạt như AWS, Google Cloud hoặc Azure để giảm thiểu chi phí ban đầu. Ngoài ra, việc hợp tác với các tổ chức hoặc doanh nghiệp trong lĩnh vực bất động sản có thể mang lại nguồn kinh phí tài trợ và hỗ trợ kỹ thuật. Đồng thời, việc triển khai hệ thống theo từng giai đoạn sẽ giúp kiểm soát chi phí hiệu quả và đảm bảo nguồn lực được phân bổ hợp lý.

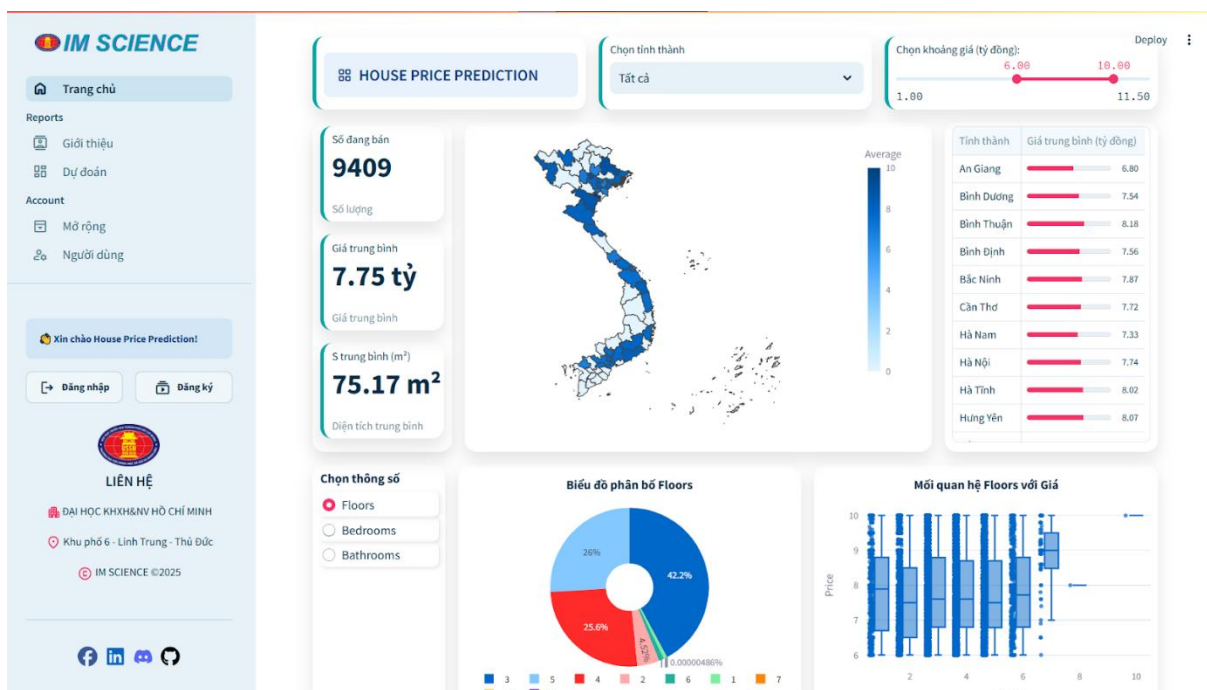
DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Tổng cục Thống kê. (2023). Báo cáo tình hình kinh tế - xã hội năm 2023. Truy cập từ <https://www.gso.gov.vn>
- [2] Hiệp hội Bất động sản Việt Nam. (2022). *Tác động của bất động sản đến nền kinh tế Việt Nam*. Truy cập từ <https://vrea.vn>
- [3] Bộ Kế hoạch và Đầu tư. (2022). *Tổng quan đầu tư trực tiếp nước ngoài tại Việt Nam năm 2022*. Truy cập từ <https://www.mpi.gov.vn>
- [4] Ngân hàng Thế giới. (2023). *Đô thị hóa và phát triển bền vững tại Việt Nam*. Truy cập từ <https://www.worldbank.org>
- [5] Báo cáo của Ngân hàng Thế giới. (2024). *Tốc độ đô thị hóa và phát triển kinh tế tại Việt Nam*. Truy cập từ <https://www.worldbank.org>
- [6] Báo cáo của Bộ Tài chính. (2022). *Thu ngân sách từ đất đai và bất động sản năm 2022*. Truy cập từ <https://www.mof.gov.vn>
- [7] Savills Vietnam. (2024). *Thị trường bất động sản Việt Nam: Tổng quan và dự báo*. Truy cập từ <https://www.savills.com.vn>
- [8] CBRE Vietnam. (2022). *Vietnam Real Estate Market Outlook 2022*. Truy cập từ <https://www.cbrevietnam.com>
- [9] Hiệp hội Bất động sản Việt Nam. (2022). *Những xu hướng mới trên thị trường bất động sản Việt Nam*. Truy cập từ <https://vrea.vn>
- [10] Hiệp hội Bất động sản Việt Nam. (2023). *Ứng dụng công nghệ trong bất động sản Việt Nam: Thực trạng và xu hướng*. Truy cập từ <https://vrea.vn>
- [11] Propzy Việt Nam. (2022). *Tác động của nền tảng PropTech tới thị trường bất động sản Việt Nam*. Truy cập từ <https://propzy.vn>
- [12] B Insights. (2022). *Blockchain in Real Estate: Current Uses and Future Potential*. Truy cập từ <https://www.cbinsights.com>
- [13] Bộ Thông tin và Truyền thông. (2022). *Báo cáo về tình hình phát triển công nghệ số tại Việt Nam*. Truy cập từ <https://mic.gov.vn>

PHỤ LỤC

Phụ lục 1. Giao diện trang web





IM SCIENCE

Trang chủ

Reports

- Giới thiệu
- Dự đoán

Account

- Mở rộng
- Người dùng

Xin chào House Price Prediction!

[Đăng nhập](#) [Đăng ký](#)

LIÊN HỆ

ĐẠI HỌC KHXH&NV HỒ CHÍ MINH

Khu phố 6 - Linh Trung - Thủ Đức

IM SCIENCE ©2025

[f](#) [in](#) [m](#) [w](#)

Để thoát khỏi chế độ toàn màn hình, hãy nhấn và giữ Thoát

IM SCIENCE

House Price Prediction >

Viet Nam in 2025 >

Khám phá và phân tích hiệu suất giá nhà tại Việt Nam **trong năm 2024** được tạo bởi **nhóm developers** tại **Thành phố Hồ Chí Minh**.

[f](#) [in](#) [m](#) [w](#)

▼

Tại sao bạn nên sử dụng IM SCIENCE?

Lợi ích **chính** được trình bày dưới dạng thẻ hoặc hình ảnh

- Hệ thống thông minh**
- Dễ dàng sử dụng hơn!**
- Phân tích hiệu suất**
- Bảo mật thông tin**

IM SCIENCE

Trang chủ

Reports

Giới thiệu

Dự đoán

Account

Mở rộng

Người dùng

Dự đoán

Thông kê chi tiết

Xin chào House Price Prediction!

Đăng nhập

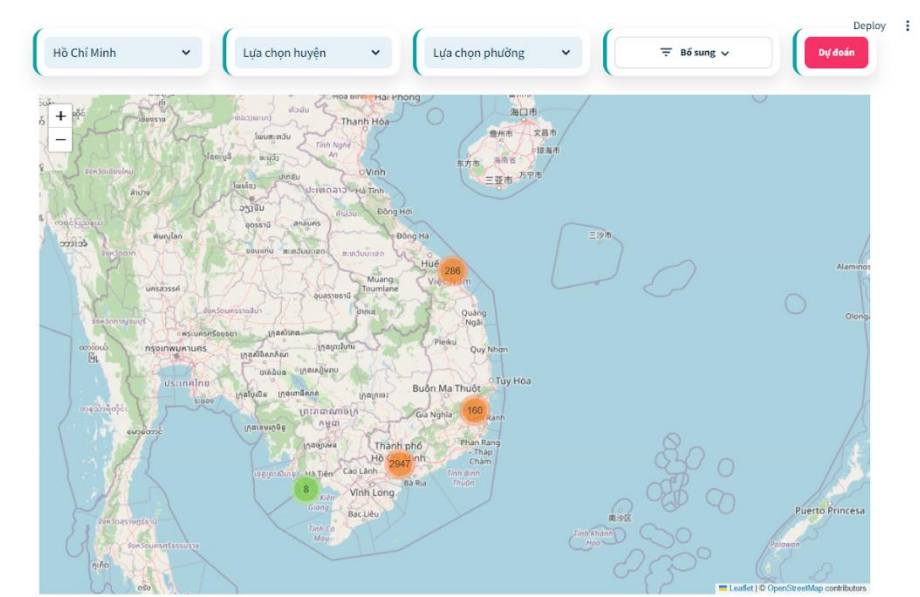
Đăng ký

LIÊN HỆ

ĐẠI HỌC KHCH&NV HỒ CHÍ MINH

Khu phố 6 - Linh Trung - Thủ Đức

IM SCIENCE ©2025



IM SCIENCE

Trang chủ

Reports

Giới thiệu

Dự đoán

Account

Mở rộng

Người dùng

Dự đoán

Thông kê chi tiết

Xin chào House Price Prediction!

Đăng nhập

Đăng ký

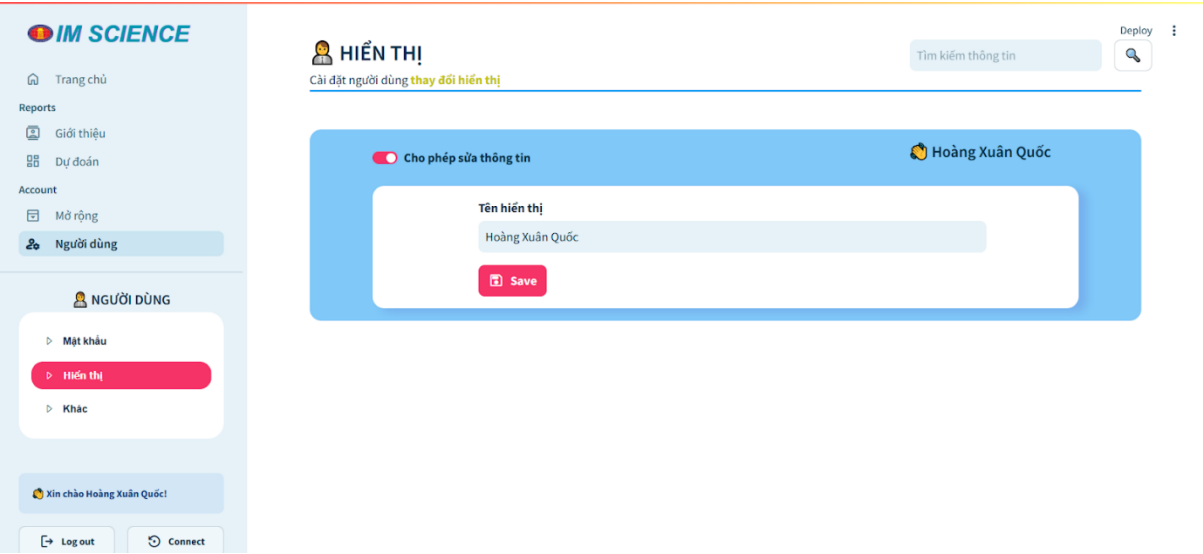
LIÊN HỆ

ĐẠI HỌC KHCH&NV HỒ CHÍ MINH

Khu phố 6 - Linh Trung - Thủ Đức

IM SCIENCE ©2025





PHIẾU ĐĂNG KÝ ĐỀ TÀI ĐỒ ÁN TỐT NGHIỆP

1. Tên đồ án

Website Phân tích thị trường bất động sản Việt Nam

2. Sinh viên thực hiện

(1)Hoàng Xuân Quốc MSSV:2156210125 Lớp:QLTTB.....

(2)Hứa Gia Tường MSSV:2156210133 Lớp: QLTTB.....

(3)Võ Châu Trường Thịnh MSSV:2156210140 Lớp: QLTTB.....

Thông tin về giảng viên hướng dẫn được đề xuất

- Họ và tên: Trần Đình Anh Huy

- Email: tdahuy@hcmussh.edu.vn

Số điện thoại:

3. Mô tả đồ án tốt nghiệp

- Lý do lựa chọn đề tài:Trong quá trình làm việc tại đơn vị thực tập , nhận thấy việc định giá bất động sản trong việc thẩm định giá trị của bất động sản ở đơn vị thực tập còn trải qua nhiều thời gian . Khi hoàn thành định giá chưa ra được một báo trực quan và còn nhiều hạn chế, và hiện nay trên thị trường cũng chưa có một sản phẩm đáp ứng được việc đưa ra những báo cáo và thông tin nhanh và chuẩn xác đáp ứng nhu cầu trên. Đó là lý do để nhóm lựa chọn thực hiện đề tài.....

- Mục tiêu, nhiệm vụ của đồ án: Tạo ra một website cung cấp cho người sử dụng thông tin về giá cả, số lượng , xu hướng, thị hiếucủa thị trường bất động sản một cách nhanh và chính xác nhất

- Phương pháp thực hiện: Thực hiện việc thu thập , lưu trữ và xử lý dữ liệu. Phân loại và tính toán . Tạo website bằng các công cụ . Tạo mô hình phân tích dựa trên kiến thức đã học

- Ý nghĩa của đồ án: Tạo ra một sản phẩm giúp cung cấp cho người sử dụng những thông tin nhanh và chính xác về giá trị , xu hướng , thị hiếu của thị trường bất động sản ở Việt Nam

- Nội dung, kết quả dự kiến: Một website cung cấp thông tin chung tổng quát, các báo cáo ,
phân tích về bất động sản ở Việt Nam.....

.....
.....

Thành phố Hồ Chí Minh, ngày ... tháng ... năm

Sinh viên thực hiện

(Ký và ghi rõ họ tên từng sinh viên)

KẾ HOẠCH THỰC HIỆN ĐỒ ÁN TỐT NGHIỆP

1. Tên đồ án: Website phân tích thị trường bất động sản Việt Nam

2. Sinh viên thực hiện:

Họ & Tên	Lớp	MSSV
1. Võ Châu Trường Thịnh	QLTT B	2156210140
2. Hoàng Xuân Quốc	QLTT B	2156210125
3. Hứa Gia Tường	QLTT B	2156210133

3. Thông tin về giảng viên hướng dẫn được đề xuất

- Họ và tên: Trần Đình Anh Huy
- Email:
- Số điện thoại:

4. Mô tả đồ án tốt nghiệp

- Lý do lựa chọn đề tài:

Thị trường bất động sản Việt Nam có vai trò quan trọng trong nền kinh tế nhưng còn nhiều hạn chế về minh bạch thông tin và khả năng phân tích dữ liệu. Đề tài tập trung ứng dụng công nghệ để cải thiện các vấn đề này, hỗ trợ người dùng và nhà đầu tư ra quyết định chính xác hơn.

- Mục tiêu, nhiệm vụ của đồ án:
 - Xây dựng website hỗ trợ tra cứu thông tin, phân tích và dự báo thị trường bất động sản.
 - Ứng dụng công nghệ Big Data, AI để nâng cao hiệu quả và minh bạch trong giao dịch.
- Phương pháp thực hiện:
 - Thu thập dữ liệu từ các nguồn uy tín và xử lý bằng Big Data, AI.
 - Kết hợp phương pháp phân tích định tính và định lượng để đánh giá hiệu quả hệ thống.
- Ý nghĩa đồ án:
 - Góp phần minh bạch hóa thông tin thị trường bất động sản.
 - Tạo công cụ hiện đại hỗ trợ phân tích và ra quyết định.
 - Thúc đẩy ứng dụng công nghệ trong quản lý và phát triển bất động sản.
- Nội dung, kết quả dự kiến:
 - Xây dựng website trực quan, hỗ trợ phân tích giá cả và xu hướng thị trường.
 - Đề xuất giải pháp công nghệ và báo cáo hiệu quả của hệ thống.

5. Kế hoạch thực hiện

STT	Nội dung công việc cần thực hiện	Thời gian bắt đầu	Thời gian kết thúc
1	Tìm hiểu, thu thập tài liệu và dữ liệu	10/12/2024	17/12/2024
2	Phân tích và tổng hợp các khái niệm lý thuyết	18/12/2024	24/12/2024
3	Xây dựng cơ sở dữ liệu và thiết kế website	25/12/2024	02/01/2025
4	Lập trình và phát triển tính năng website	03/01/2025	09/01/2025
5	Kiểm thử hệ thống và điều chỉnh	10/01/2025	14/01/2025
6	Viết báo cáo đồ án	12/01/2025	16/01/2025
7	Nộp và bảo vệ đồ án	17/01/2025	15/02/2025

Xác nhận GVHD đồ án Thành phố Hồ Chí Minh, ngày tháng năm
Sinh viên thực hiện

PHIẾU THEO DÕI TIẾN ĐỘ
THỰC HIỆN ĐỒ ÁN TỐT NGHIỆP

- 1. Tên đồ án:** WEBSITE PHÂN TÍCH THỊ TRƯỜNG BẤT ĐỘNG SẢN VIỆT NAM
- 2. Giảng viên hướng dẫn:** Th.S Trần Đình Anh Huy
- 3. Sinh viên thực hiện:**
- (1) Hoàng Xuân Quốc – 2156210125 – QLTTB
- (2) Võ Châu Trường Thịnh – 2156210140 – QLTTB
- (3) Hứa Gia Tường – 2156210133 - QLTTB

Tháng	Nội dung	Nhận xét của GVHD
11	Nghiên cứu tài liệu về học máy, dự báo giá bất động sản và phát triển web, xác định yêu cầu hệ thống. Thu thập dữ liệu bất động sản từ các nguồn uy tín (giá nhà, diện tích, vị trí, tiện ích, v.v.), xử lý và làm sạch dữ liệu để loại bỏ giá trị không hợp lệ, trùng lặp và chuẩn hóa dữ liệu.	
12	Xây dựng mô hình học máy sử dụng các thuật toán phù hợp để dự báo giá bất động sản và chọn mô hình tối ưu dựa trên hiệu suất. Phát triển website hiển thị thông tin dự báo, bao gồm biểu đồ trực quan và chức năng nhập liệu. Tích hợp mô hình dự báo vào website và kiểm tra hiệu suất.	

Tháng	Nội dung	Nhận xét của GVHD
1	Thử nghiệm hệ thống với dữ liệu thực tế, thu thập phản hồi và tối ưu hóa mô hình cùng website. Chuẩn bị báo cáo chi tiết và luyện tập thuyết trình, hoàn thiện tài liệu để bảo vệ đồ án, trình bày toàn bộ kết quả và đề xuất cải tiến.	

Thành phố Hồ Chí Minh, ngày ... tháng ... năm

Sinh viên thực hiện

(Ký và ghi rõ họ tên từng sinh viên)