

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

# ĐỒ ÁN TỐT NGHIỆP

## Applying MLOps for Enhanced AI Prediction in Real Estate

Lê Thành Long

long.lt194099@sis.hust.edu.vn

Ngành: Khoa học máy tính

Giảng viên hướng dẫn: TS. Trần Văn Đặng

Chữ kí GVHD

Khóa: 64

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2024

# LỜI CẢM ƠN

Em xin bày tỏ cảm ơn sâu sắc đến thầy Tiến sĩ Trần Văn Đăng và sự hướng dẫn tỉ mỉ của thầy không những ở khía cạnh kiến thức chuyên môn mà còn các kỹ năng mềm trong quá trình làm đồ án tại Trường Công nghệ thông tin và Truyền Thông, Đại Học Bách Khoa Hà Nội. Bên cạnh em xin cảm ơn các thầy cô như những người lái đò giúp bản thân em phát triển mạnh mẽ hơn và có được những kiến thức cho chặng đường sắp tới. Em đã có được những sự trưởng thành, mạnh mẽ và chắc chắn hơn và sự quyết tâm hơn cho giai đoạn tiếp theo của sự nghiệp.

Em xin cảm ơn bạn bè, người thân, những vấp ngã trong các cuộc thi độ sát ở Đại học Bách Khoa Hà Nội đã khiến em phát triển hoàn thiện hơn. Những sự ủng hộ động viên không ngừng nghỉ đó đã giúp em biết mình phải làm gì, chuẩn bị và sắp xếp thời gian của mình sao cho hợp lý. Những kiến thức không chỉ nằm vồn vện trên ghế nhà trường mà còn là những kinh nghiệm quý báu trong cuộc sống em đã được những thầy cô tâm huyết chỉ bảo. Đó là những món quà mà em rất hãnh diện có được cho đến tận thời điểm này, em đã chuẩn bị hành trang đầy đủ cho sau này.

Cuối cùng em xin cảm ơn ngôi nhà chung Đại Học Bách Khoa Hà Nội nói chung và ngôi nhà riêng khoa học máy tính 06 đã là nơi để em sẽ trở về và ôn lại những kỷ niệm, những bài thi, những cái ôm, những cái tên mà em sẽ nhớ mãi.

# Lời Cam Đoan

Họ và tên sinh viên: Lê Thành Long  
MSSV: 20194099  
Điện thoại liên lạc: 0969973012  
Email: long.lt194099@sis.hust.edu.vn  
Lớp: IT1-06-K64  
Chương trình đào tạo: Khoa Học Máy Tính

Tôi - Lê Thành Long - cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của TS. Trần Văn Đặng. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày      tháng      năm

Tác giả ĐATN

*(Chữ ký và tên đầy đủ)*

# TÓM TẮT NỘI DUNG ĐỒ ÁN

Việc phát triển và triển khai các mô hình AI trong thế giới thực đặt ra những thách thức đáng kể do hạn chế về thời gian và nguồn lực. Độ tin cậy là tối quan trọng, đòi hỏi độ trễ thấp, thời gian hoạt động cao và khả năng chịu lỗi mạnh mẽ. Ngoài ra, việc cập nhật liên tục với kiến thức mới và triển khai tự động là rất quan trọng để duy trì độ chính xác dự đoán với dữ liệu đang phát triển. Các quy trình tích hợp truyền thống thường thiếu sự gắn kết, bao gồm các giai đoạn rời rạc từ chuẩn bị dữ liệu đến giám sát hiệu quả mô hình và tự động hóa. Sự xuất hiện gần đây của MLOps (Quy trình học máy) cung cấp một phương pháp tiếp cận chuẩn hóa cho vòng đời học máy, giải quyết những hạn chế này. Trong nghiên cứu này, tôi sẽ đề xuất một quy trình MLOPS để nâng cao tính tự động hóa và hiệu quả của hệ thống AI. Cùng với quy trình MLOps, tôi tập trung giải quyết vấn đề dự đoán giá bất động sản tại thị trường Việt Nam.

Gần đây, vấn đề giá bất động sản được quan tâm nhiều hơn khi có nhiều hiện tượng lạm phát giá. Do đó, định giá bất động sản đã trở thành một trong những câu hỏi mà người dùng quan tâm. Nhận thấy tiềm năng của thị trường, nhiều chuyên gia trong lĩnh vực thống kê và trí tuệ nhân tạo đã tích cực tham gia và đề xuất vô số giải pháp cho các nền tảng tin tức và đơn vị cung cấp dịch vụ bất động sản. Dữ liệu bất động sản được cập nhật hàng ngày đòi hỏi các dịch vụ trí tuệ nhân tạo cũng phải được cập nhật tương ứng. Nhiều dịch vụ AI trong lĩnh vực bất động sản vẫn chưa nhận ra tầm quan trọng của vấn đề này và đã ảnh hưởng đến trải nghiệm của người dùng về độ tin cậy của dự đoán giá bất động sản. Ngoài ra, quá trình xây dựng và triển khai các mô hình AI trong nhiều sản phẩm không đảm bảo tính toàn vẹn và độ tin cậy và ảnh hưởng trực tiếp đến người dùng cuối như: các mô hình triển khai được người xây dựng mô hình đánh giá chủ quan, quá trình đào tạo không được giám sát đúng cách. Do đó, việc các mô hình AI đưa ra kết quả dự đoán không đáng tin cậy, ảnh hưởng trực tiếp đến người dùng là điều không thể tránh khỏi. Hơn nữa, việc cập nhật kiến thức về giá bất động sản ở thời điểm hiện tại là cần thiết trong việc định giá nhà cho người dùng và đơn vị cung cấp dịch vụ.

Để giải quyết những thách thức này, nghiên cứu này đề xuất một quy trình tự động bao gồm chuẩn bị dữ liệu, phát triển mô hình, xử lý hậu kỳ, đánh giá tự động và giám sát. Quy trình này nhằm: (i) tăng cường tự động hóa quy trình trong tích hợp mô hình AI trong quá trình phát triển phần mềm mà không ảnh hưởng đến trải nghiệm của người dùng, (ii) triển khai các cơ chế giám sát và đánh giá hiệu quả trong suốt vòng đời sản phẩm, (ii) đảm bảo độ tin cậy và tính toàn vẹn của hệ thống

trong các ứng dụng thực tế. Bằng cách tận dụng các nguyên tắc MLOps, nghiên cứu này tìm cách tối ưu hóa các hệ thống dự đoán bất động sản do AI thúc đẩy, điều chỉnh chúng theo động lực thị trường đang thay đổi và kỳ vọng của người dùng. Cuối cùng nghiên cứu xây dựng hệ thống BKPrice thành công và chứng thực bằng các thông số thực nghiệm và hoàn thành tốt những yêu cầu đã đặt ra.

Sinh viên thực hiện  
(Ký và ghi rõ họ tên)

# ABSTRACT

The development and deployment of real-world AI models pose significant challenges due to time and resource constraints. Reliability is paramount, demanding low latency, high uptime, and robust fault tolerance. Additionally, continuous updates with new knowledge and automated deployments are crucial for maintaining predictive accuracy with evolving data. Traditional integration processes often lack cohesion, encompassing fragmented stages from data preparation to model effectiveness monitoring and automation. The recent emergence of MLOps (Machine Learning Operations) offers a standardized approach to the machine learning lifecycle, addressing these limitations. In this research, I will propose an MLOps process to improve the automation and efficiency of the AI system. Together with the MLOps process, I focus on solving the problem of real estate quotation in the Vietnamese market.

Referencing this standard process with existing solutions, we can immediately see the major shortcomings in manual AI application systems. In this study, I will propose an MLOPS process to improve the automation and efficiency of the AI system. Along with the MLOps process, I focus on solving the problem of predicting real estate prices in the Vietnamese market. Recently, the issue of real estate prices has received more attention when there have been many price inflation phenomena. Therefore, real estate valuation has become one of the questions that users are interested in. Realizing the potential of the market, many experts in the fields of statistics and artificial intelligence have actively participated and proposed countless solutions for news platforms and real estate service providers. Real estate data is updated daily requires artificial intelligence services to be updated accordingly. Many AI services in the real estate sector have not yet recognized the importance of this issue and have affected the user experience in terms of the reliability of real estate price predictions. In addition, the process of building and deploying AI models in many products does not ensure integrity and reliability and directly affects end users such as: deployed models are subjectively evaluated by the model builder, the training process is not properly monitored. Therefore, it is inevitable that AI models give unreliable prediction results, directly affecting users. Moreover, updating knowledge about real estate prices at the present time is necessary in pricing houses for users and service providers. To address these challenges, this study proposes an automated process encompassing data preparation, model development, post-processing, automatic evaluation, and monitoring. This MLOps-driven approach aims to: (i) increase process automation in AI model integration within software

development without compromising user experience, (ii) implement effective monitoring and evaluation mechanisms throughout the product lifecycle, (ii) ensure system reliability and integrity in real-world applications. By leveraging MLOps principles, this research seeks to optimize AI-driven real estate prediction systems, aligning them with evolving market dynamics and user expectations.

## MỤC LỤC

<b>CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....</b>	<b>1</b>
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	3
1.3 Định hướng giải pháp.....	4
1.4 Đóng góp của đồ án .....	5
1.5 Bố cục của đồ án .....	6
1.6 Các giải pháp hiện tại và hạn chế .....	7
1.6.1 Xây dựng mô hình không tin cậy .....	7
1.6.2 Quy trình xây dựng mô hình thủ công .....	8
1.7 Mục tiêu và định hướng giải pháp .....	8
1.7.1 Mục tiêu.....	8
1.7.2 Định hướng giải pháp .....	8
<b>CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH TỔNG QUAN.....</b>	<b>10</b>
2.1 Khảo sát .....	10
2.2 Phân tích chức năng tổng quan hệ thống .....	12
2.2.1 Tổng quan hệ thống .....	12
2.2.2 Các thành phần chính của hệ thống.....	13
<b>CHƯƠNG 3. CƠ SỞ LÝ THUYẾT .....</b>	<b>15</b>
3.1 Giải thuật và tính hiệu quả của hệ thống.....	15
3.2 Framework và tính tự động của hệ thống trong lưu trữ và xây dựng mô hình .....	21
<b>CHƯƠNG 4. PHƯƠNG PHÁP ĐỀ XUẤT.....</b>	<b>30</b>
4.1 Tổng quan giải pháp.....	30



4.2 Hệ thống thu thập dữ liệu tự động .....	30
4.2.1 Data Components.....	31
4.2.2 Data Pipeline .....	32
4.2.3 Xử lý và lưu trữ dữ liệu .....	35
4.3 Hệ thống dự đoán giá bất động sản tin cậy .....	37
4.3.1 Giải thuật định giá bất động sản .....	37
4.3.2 Hệ thống huấn luyện và triển khai mô hình tự động.....	49
4.3.3 Triển khai quá trình MLOPs bằng mô hình ngôn lớn .....	58
<b>CHƯƠNG 5. ĐÁNH GIÁ VÀ THỰC NGHIỆM .....</b>	<b>61</b>
5.1 Các tham số đánh giá .....	61
5.1.1 Tập tham số đánh giá hiệu năng mô hình dự đoán .....	61
5.1.2 Tập tham số đánh giá quy trình MLOps.....	62
5.2 Phương pháp thí nghiệm.....	62
5.2.1 Cấu hình thiết bị sử dụng .....	62
5.2.2 Môi trường lập trình thử nghiệm.....	62
5.2.3 Cấu hình siêu tham số .....	63
5.2.4 Cấu hình tập thuộc tính.....	64
5.2.5 Tiến hành thí nghiệm.....	65
<b>CHƯƠNG 6. GIẢI PHÁP VÀ ĐÓNG GÓP NỔI BẬT .....</b>	<b>76</b>
6.0.1 Tính mới và tính sáng tạo.....	76
6.0.2 Tính module hóa và tái sử dụng .....	76
<b>CHƯƠNG 7. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI .....</b>	<b>78</b>
7.1 Kết luận .....	78
7.2 Hướng phát triển trong tương lai .....	79
<b>CHƯƠNG 8. PHỤ LỤC.....</b>	<b>81</b>

**REFERENCE ..... 84**