

ĐỊNH GIÁ ĐẤT TRỒNG LÚA BẰNG MÔ HÌNH RANDOM FOREST VÀ XGBOOST TRÊN CƠ SỞ DỮ LIỆU GIS TẠI THÀNH PHỐ HUẾ

Lê Hữu Ngọc Thanh*, Phan Hoàng Kiều Nhung
Email: lhnthanh@hueuni.edu.vn

VALUATION OF PADDY LAND USING RANDOM FOREST AND XGBOOST MODELS BASED ON GIS DATA IN HUE CITY

ABSTRACT

This study aims to forecast and analyze paddy land prices in Thuy Thanh commune, Hue city, based on spatial data and machine learning models. The dataset includes 454 land parcels with 9 variables belonging to two groups: production-related factors and non-production factors. Two models, Random Forest and XGBoost, were used to forecast land prices. The results show that both models have good predictive capabilities, with Random Forest achieving higher accuracy with $R^2 = 0.89$, $RMSE = 9289.11$, and $MAE = 7278.27$. Spatial analysis reveals a clear differentiation in land prices, with high-priced areas concentrated along major transportation routes and in areas affected by urban development projects. The results also show that non-production factors, especially project information and compensation policies of enterprises, strongly influence paddy land prices.

Keywords: Paddy land valuation; Random Forest; XGBoost; GIS; urbanization.

TÓM TẮT

Nghiên cứu này nhằm dự báo và phân tích giá đất trồng lúa tại xã Thủy Thanh, thành phố Huế, trên cơ sở dữ liệu không gian và các mô hình học máy. Bộ dữ liệu gồm 454 thửa đất với 9 biến thuộc hai nhóm: yếu tố liên quan đến sản xuất và yếu tố phi sản xuất. Hai mô hình Random Forest và XGBoost được sử dụng để dự báo giá đất. Kết quả cho thấy cả hai mô hình đều có khả năng dự báo tốt, trong đó Random Forest đạt độ chính xác cao hơn với $R^2 = 0,89$, $RMSE = 9289,11$ và $MAE = 7278,27$. Phân tích không gian cho thấy giá đất phân hóa rõ rệt, với các khu vực giá cao tập trung dọc theo trục giao thông chính và tại những nơi chịu tác động của các dự án phát triển đô thị. Kết quả cũng cho thấy các yếu tố phi sản xuất, đặc biệt là thông tin dự án và chính sách bồi thường của doanh nghiệp, có ảnh hưởng mạnh đến giá đất trồng lúa.

Từ khóa: Định giá đất trồng lúa; thuật toán Random Forest; thuật toán XGBoost; GIS, đô thị hóa.

I. ĐẶT VẤN ĐỀ

Đất nông nghiệp, đặc biệt là đất trồng lúa, đóng vai trò quan trọng đối với an ninh lương thực, sinh kế của người dân và sự ổn định trong sử dụng tài nguyên đất. Tuy nhiên, trong bối cảnh đô thị hóa, phát triển hạ tầng và chuyển dịch cơ cấu kinh tế, giá trị đất nông nghiệp ngày càng chịu tác động đồng thời của nhiều yếu tố khác nhau, không chỉ từ điều kiện sản xuất mà còn từ vị trí, quy hoạch và các chính sách liên quan đến sử dụng đất. Quá trình chuyển đổi nông thôn gắn với mở rộng hạ tầng và liên kết nông thôn - đô thị được xem là xu hướng tất yếu, đồng thời tạo ra những thay đổi đáng kể trong không gian sử dụng đất nông nghiệp (FAO, 2017). Trong thực tế, việc xác định giá trị đất trồng lúa thường còn phụ thuộc nhiều vào kinh nghiệm hoặc khung giá hành chính, trong khi giá đất chịu ảnh hưởng của nhiều yếu tố có quan hệ phi tuyến và phân bố không đồng đều trong không gian.

Tại Việt Nam, quá trình đô thị hóa và phát triển hạ tầng đã làm gia tăng áp lực chuyển đổi mục đích sử dụng đất, đồng thời đặt ra yêu cầu cần nâng cao tính minh bạch và cơ sở định lượng trong công tác định giá

đất (World Bank, 2020). Trong bối cảnh phát triển của khoa học dữ liệu, các mô hình học máy ngày càng được ứng dụng rộng rãi trong các bài toán dự báo nhờ khả năng xử lý dữ liệu đa biến và mô hình hóa các mối quan hệ phức tạp. Trong đó, Random Forest là phương pháp dựa trên tập hợp nhiều cây quyết định nhằm cải thiện khả năng dự báo (Breiman, 2001), trong khi XGBoost là phương pháp tăng cường gradient có khả năng tối ưu hóa hiệu quả tính toán và độ chính xác (Chen & Guestrin, 2016). Khi kết hợp với hệ thống thông tin địa lý (GIS), các mô hình này không chỉ hỗ trợ dự báo mà còn cho phép phân tích và trực quan hóa sự phân bố giá đất theo không gian.

Đối với khu vực Thủy Thanh, thành phố Huế, nơi đất trồng lúa vẫn giữ vai trò quan trọng nhưng đang chịu tác động của quá trình chuyển dịch sử dụng đất, việc nghiên cứu định giá đất trồng lúa là cần thiết nhằm làm rõ ảnh hưởng của các yếu tố liên quan đến sản xuất và phi sản xuất đến giá đất, đồng thời hỗ trợ cho công tác quản lý, quy hoạch và sử dụng đất trong giai đoạn phát triển hiện nay. Từ đó, nghiên cứu này được thực hiện nhằm xây dựng mô hình dự báo giá đất trồng lúa trên cơ sở dữ liệu không gian và các yếu tố ảnh hưởng, đồng thời phân tích mức độ tác động của từng nhóm yếu tố và thành lập bản đồ phân bố giá đất phục vụ cho công tác quản lý đất đai tại địa phương.

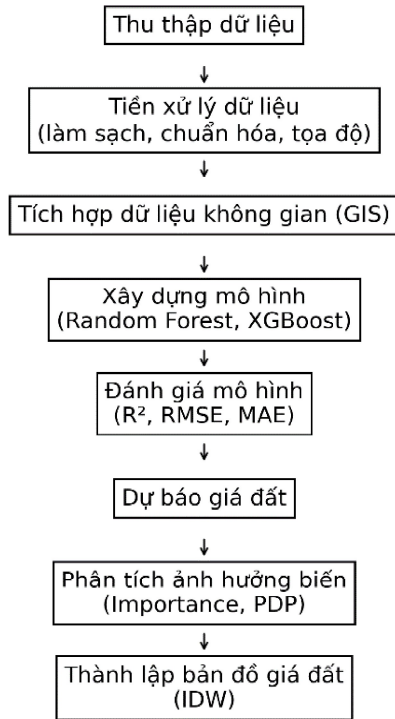
II. PHƯƠNG PHÁP NGHIÊN CỨU

1. Mô tả khu vực nghiên cứu

Xã Thủy Thanh, nay là một phần của phường Thanh Thủy, là khu vực có quỹ đất nông nghiệp lớn và sản xuất lúa giữ vai trò chủ đạo trong cơ cấu sử dụng đất. Năm 2024, trên tổng diện tích tự nhiên 849,16 ha, đất nông nghiệp chiếm 582,72 ha (68,62%), trong đó riêng đất trồng lúa đạt 536,16 ha, tương đương 62,14% diện tích tự nhiên. Điều này cho thấy đất lúa không chỉ là loại hình sử dụng đất chiếm ưu thế mà còn gắn chặt với sinh kế và không gian sản xuất nông nghiệp của địa phương. Tuy nhiên, theo định hướng đến năm 2030, diện tích đất nông nghiệp của xã dự kiến giảm còn 413,09 ha, trong đó đất trồng lúa giảm xuống còn 403,16 ha theo báo cáo quy hoạch sử dụng đất xã Thủy Thanh giai đoạn 2020-2030 tầm nhìn đến năm 2050. Xu hướng thu hẹp này phản ánh tác động của quá trình đô thị hóa, phát triển hạ tầng và chuyển đổi mục đích sử dụng đất. Trong bối cảnh đó, việc định giá đất trồng lúa tại khu vực Thủy Thanh có ý nghĩa cần thiết, không chỉ nhằm nhận diện sự phân hóa giá trị đất theo không gian mà còn cung cấp cơ sở khoa học cho công tác quản lý đất đai, bồi thường, quy hoạch và sử dụng đất hiệu quả trong quá trình chuyển dịch phát triển đô thị.

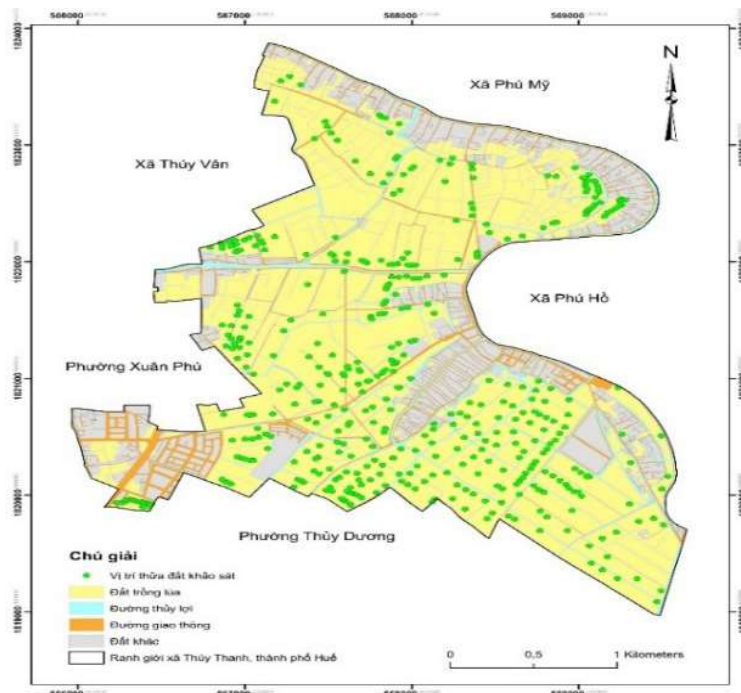
2. Quy trình nghiên cứu

Nghiên cứu được thực hiện theo quy trình tích hợp giữa dữ liệu không gian (Geographic Information System – GIS) và các mô hình học máy nhằm dự báo giá đất trồng lúa tại khu vực nghiên cứu. Quy trình gồm các bước chính: thu thập dữ liệu, tiền xử lý, tích hợp dữ liệu không gian, xây dựng mô hình, đánh giá kết quả, phân tích ảnh hưởng của các yếu tố và thành lập bản đồ giá đất. Trình tự thực hiện được thể hiện trong sơ đồ dưới đây.



Hình 1. Quy trình nghiên cứu

Dữ liệu sử dụng trong nghiên cứu được thu thập từ khảo sát thực địa kết hợp với các nguồn thông tin địa phương. Bộ dữ liệu gồm 454 thửa đất trồng lúa phân bố trên toàn khu vực nghiên cứu. Giá đất của từng thửa được xác định thông qua phỏng vấn trực tiếp người sử dụng đất, người môi giới am hiểu thị trường tại địa phương về mức giá chuyển nhượng thực tế hoặc dự kiến chuyển nhượng tại thời điểm khảo sát. Mỗi mẫu quan sát được gắn với tọa độ địa lý (kinh độ, vĩ độ) để phục vụ phân tích không gian và tích hợp trong môi trường GIS.



Hình 2. Vị trí các thửa đất

Trong nghiên cứu này, giá đất trồng lúa được sử dụng làm biến phụ thuộc. Các biến giải thích được phân thành hai nhóm chính nhằm phản ánh các khía cạnh khác nhau của giá đất. Nhóm thứ nhất là các yếu tố liên quan đến sản xuất, bao gồm diện tích thửa đất, khoảng cách đến đường giao thông chính, khoảng cách đến hệ thống tưới tiêu, độ phì nhiêu, năng suất và điều kiện hạ tầng nội đồng. Nhóm này phản ánh điều kiện canh tác và khả năng phục vụ sản xuất nông nghiệp. Nhóm thứ hai là các yếu tố phi sản xuất, bao gồm thông tin dự án và các chính sách bồi thường. Nhóm này phản ánh tác động của quy hoạch, đầu tư và các yếu tố ngoài sản xuất đến giá đất. Các biến sử dụng trong nghiên cứu được tổng hợp trong bảng dưới đây.

Bảng 1. Thông tin các biến sử dụng trong nghiên cứu

Nhóm biến	Tên biến	Đơn vị	Mô tả
Biến phụ thuộc	Giá đất trồng lúa	Nghìn đồng/m ²	Giá đất thu thập từ khảo sát
Yếu tố liên quan đến sản xuất	Diện tích thửa đất	m ²	Quy mô diện tích
	Khoảng cách đến đường giao thông chính	m	Mức độ tiếp cận giao thông
	Khoảng cách đến hệ thống tưới tiêu	m	Mức độ tiếp cận nguồn nước
	Độ phì nhiêu	0–1	Chất lượng đất
	Năng suất	Tạ/ha	Hiệu quả sản xuất
	Hạ tầng nội đồng	0–1	Điều kiện phục vụ sản xuất
Yếu tố phi sản xuất	Thông tin dự án	0–1	Ảnh hưởng quy hoạch
	Bồi thường của Nhà nước	0–1	Chính sách bồi thường của Nhà nước
	Bồi thường của doanh nghiệp	0–1	Chính sách bồi thường của doanh nghiệp

Trước khi xây dựng mô hình, dữ liệu được tiền xử lý nhằm đảm bảo tính nhất quán. Các bước xử lý bao gồm kiểm tra và loại bỏ giá trị thiếu, rà soát các giá trị bất thường, chuẩn hóa định dạng số liệu, kiểm tra các biến dạng nhị phân và đồng bộ hệ tọa độ không gian. Sau đó, dữ liệu được tích hợp với các lớp bản đồ nền như giao thông, thủy lợi và ranh giới hành chính để phục vụ tính toán các biến không gian và phân tích trong môi trường GIS.

Trên cơ sở dữ liệu đã xử lý, hai mô hình học máy là Random Forest và XGBoost được sử dụng để dự báo giá đất. Các biến giải thích được sử dụng làm đầu vào của mô hình, trong khi giá đất là biến cần dự báo. Dữ liệu được chia thành tập huấn luyện và tập kiểm định theo tỷ lệ 80/20, trong đó việc chia mẫu được thực hiện ngẫu nhiên nhằm đảm bảo tính khách quan trong quá trình đánh giá. Kết quả dự báo từ các mô hình được đánh giá thông qua các chỉ tiêu thống kê gồm hệ số xác định (R^2), sai số tuyệt đối trung bình (MAE) và căn bậc hai sai số bình phương trung bình (RMSE). Các chỉ tiêu này được sử dụng để so sánh mức độ sai lệch giữa giá trị dự báo và giá trị quan sát. Sau khi đánh giá mô hình, kết quả dự báo được sử dụng để phân tích ảnh hưởng của các yếu tố đầu vào. Mức độ ảnh hưởng của từng biến được xác định thông qua chỉ số tầm quan trọng của biến (Variable Importance). Đồng thời, biểu đồ phụ thuộc bộ phận (Partial Dependence Plot – PDP) được sử dụng để phân tích xu hướng thay đổi của giá đất theo từng yếu tố trong điều kiện các yếu tố khác được giữ không đổi. Cuối cùng, kết quả dự báo được tích hợp với dữ liệu không gian để thành lập bản đồ giá đất. Phương pháp nội suy khoảng cách nghịch đảo (Inverse Distance Weighting – IDW) được sử dụng để xây dựng bề mặt giá đất liên tục từ các điểm dữ liệu. Bản đồ kết quả cho phép thể hiện phân bố không gian của giá đất trong khu vực nghiên cứu.

III. KẾT QUẢ VÀ THẢO LUẬN

1. Kết quả đánh giá mô hình học máy

Kết quả đánh giá hai mô hình Random Forest và XGBoost được trình bày trong Bảng 2. Ba chỉ tiêu được sử dụng gồm hệ số xác định, căn bậc hai sai số bình phương trung bình và sai số tuyệt đối trung bình, qua đó phản ánh mức độ phù hợp của mô hình cũng như mức sai lệch giữa giá trị dự báo và giá trị quan sát.

Bảng 2. Kết quả mô hình Xgboost

STT	Chỉ số	Mô hình Random Forest	Mô hình XgBoost
1	R2	0.89	0.86
2	Root Mean Squared Error (RMSE) (nghìn đồng/m ²)	9.289	10.547
3	Mean Absolute Error (MAE) (nghìn đồng/m ²)	7.278	7.582.

Theo kết quả thu được, cả hai mô hình đều cho khả năng dự báo tương đối tốt đối với bài toán định giá đất trồng lúa. Hệ số xác định của mô hình Random Forest đạt 0,89, trong khi mô hình XGBoost đạt 0,86. Điều này cho thấy phần lớn biến động của giá đất trong bộ dữ liệu nghiên cứu đã được giải thích thông qua các biến đầu vào được sử dụng trong mô hình. Đồng thời, giá trị hệ số xác định của Random Forest cao hơn cho thấy mô hình này có mức độ phù hợp với dữ liệu tốt hơn so với XGBoost.

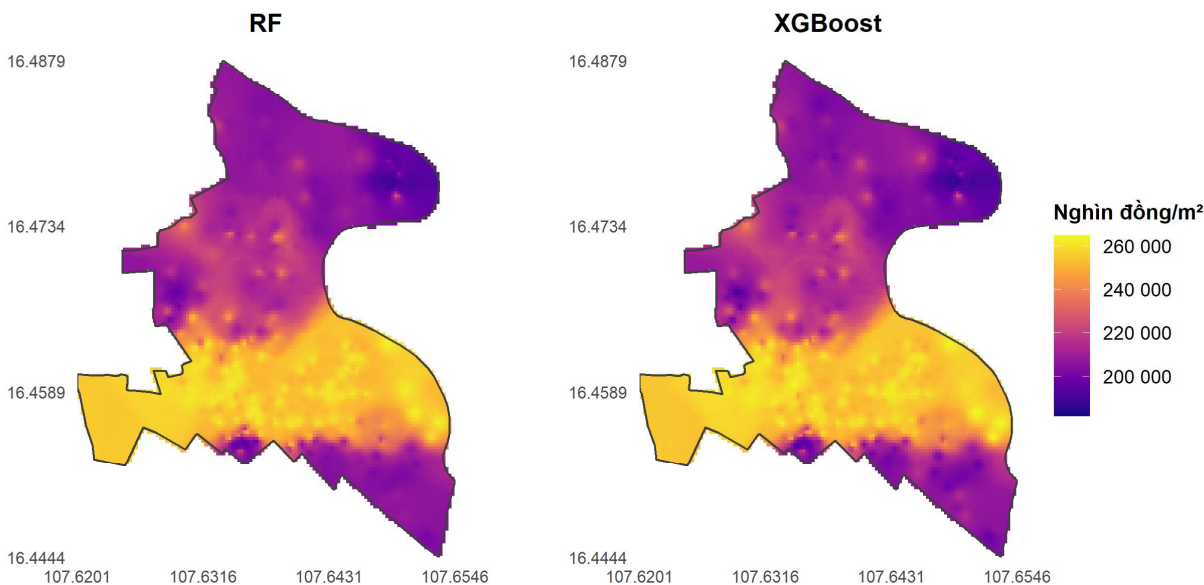
Xét theo các chỉ tiêu sai số, mô hình Random Forest tiếp tục cho kết quả khả quan hơn. Cụ thể, căn bậc hai sai số bình phương trung bình của Random Forest đạt 9.289, thấp hơn so với 10.547 của XGBoost. Tương tự, sai số tuyệt đối trung bình của Random Forest là 7.278, trong khi ở XGBoost là 7.582. Kết quả này cho thấy khoảng cách giữa giá trị dự báo và giá trị thực tế từ mô hình Random Forest nhỏ hơn, đồng nghĩa với việc mô hình có khả năng dự báo ổn định hơn trên bộ dữ liệu nghiên cứu.

So sánh tổng thể giữa hai mô hình cho thấy Random Forest đạt kết quả tốt hơn XGBoost ở cả ba chỉ tiêu đánh giá. Mặc dù mức chênh lệch không quá lớn, xu hướng này vẫn thể hiện nhất quán ở cả hệ số xác định và các đại lượng sai số, cho thấy Random Forest phù hợp hơn trong việc mô tả mối quan hệ giữa giá đất với các yếu tố đầu vào trong nghiên cứu này. Kết quả đó cũng cho thấy bộ dữ liệu nghiên cứu có thể phù hợp với cách tiếp cận dựa trên tổ hợp nhiều cây quyết định của Random Forest, trong khi XGBoost có thể cần được tối ưu tham số sâu hơn để phát huy đầy đủ hiệu quả dự báo.

Nhìn chung, cả hai mô hình đều có thể sử dụng trong bài toán dự báo giá đất trồng lúa, tuy nhiên Random Forest cho kết quả tốt hơn về độ phù hợp và mức sai số. Trên cơ sở đó, kết quả từ hai mô hình tiếp tục được sử dụng để phân tích sự phân bố không gian của giá đất và đánh giá vai trò của các yếu tố đầu vào trong các nội dung tiếp theo.

3. Kết quả dự báo và phân bố không gian giá đất

Kết quả dự báo giá đất trồng lúa theo hai mô hình Random Forest và XGBoost được thể hiện dưới dạng bản đồ không gian trong Hình 3, với giá trị được biểu diễn theo thang màu từ thấp đến cao.



Hình 3. Phân bố không gian giá đất dự báo theo hai mô hình

Kết quả cho thấy giá đất có sự phân bố không đồng đều trong khu vực nghiên cứu. Các vùng có giá trị cao tập trung chủ yếu ở khu vực phía Nam và Tây Nam, trong khi các khu vực phía Bắc và Đông Bắc có xu hướng có giá thấp hơn. Sự phân hóa này phản ánh sự khác biệt về điều kiện vị trí và khả năng tiếp cận hạ tầng giữa các khu vực.

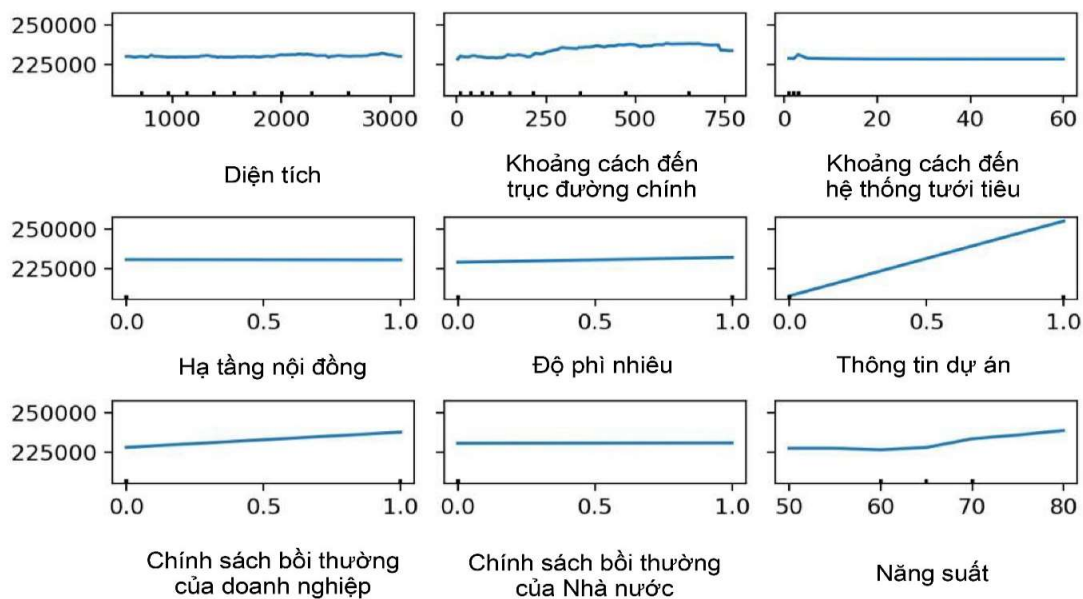
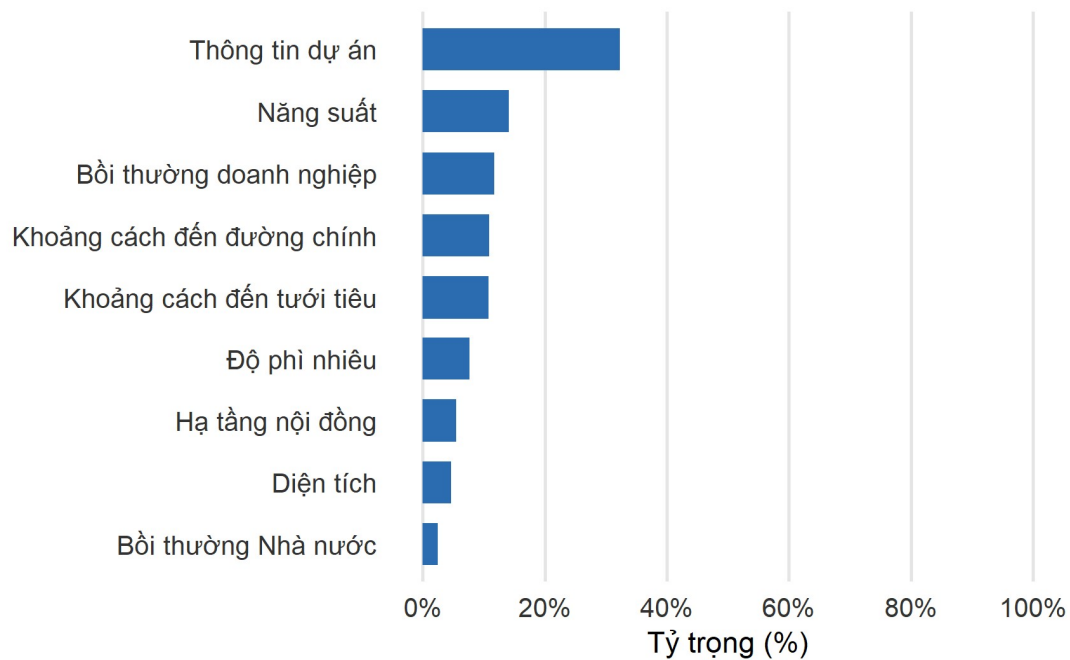
Đáng chú ý, các khu vực có giá đất cao phân bố dọc theo trục giao thông chính kết nối trung tâm đô thị với xã Thủy Thanh, cụ thể là tuyến tỉnh lộ 1. Đây là khu vực có mức độ tiếp cận thuận lợi và chịu tác động của các hoạt động phát triển đô thị, bao gồm các dự án giáo dục, y tế và khu đô thị đang và sẽ được triển khai. Điều này cho thấy yếu tố vị trí và hạ tầng đóng vai trò quan trọng trong sự hình thành giá đất tại khu vực nghiên cứu.

So sánh giữa hai mô hình cho thấy kết quả dự báo về mặt không gian là tương đối tương đồng. Cả hai mô hình đều xác định được các vùng giá cao và giá thấp với vị trí gần như trùng khớp, cho thấy tính nhất quán trong việc phản ánh cấu trúc không gian của giá đất. Sự khác biệt giữa hai mô hình chủ yếu thể hiện ở mức độ chi tiết và sự chuyển tiếp giá trị giữa các khu vực, tuy nhiên không làm thay đổi xu hướng phân bố tổng thể.

Nhìn chung, kết quả cho thấy các mô hình học máy không chỉ có khả năng dự báo giá đất mà còn phản ánh rõ ràng cấu trúc phân bố không gian của giá đất trống lúa trong khu vực nghiên cứu.

3. Ảnh hưởng của các yếu tố đến giá đất

Ảnh hưởng của các yếu tố đầu vào đến giá đất được phân tích thông qua các biểu đồ mức độ ảnh hưởng và mối tương quan giữa các yếu tố với giá đất trống lúa (Partial Dependence Plot – PDP) theo hai mô hình Random Forest và XGBoost. Kết quả được trình bày trong Hình 4 và Hình 5.

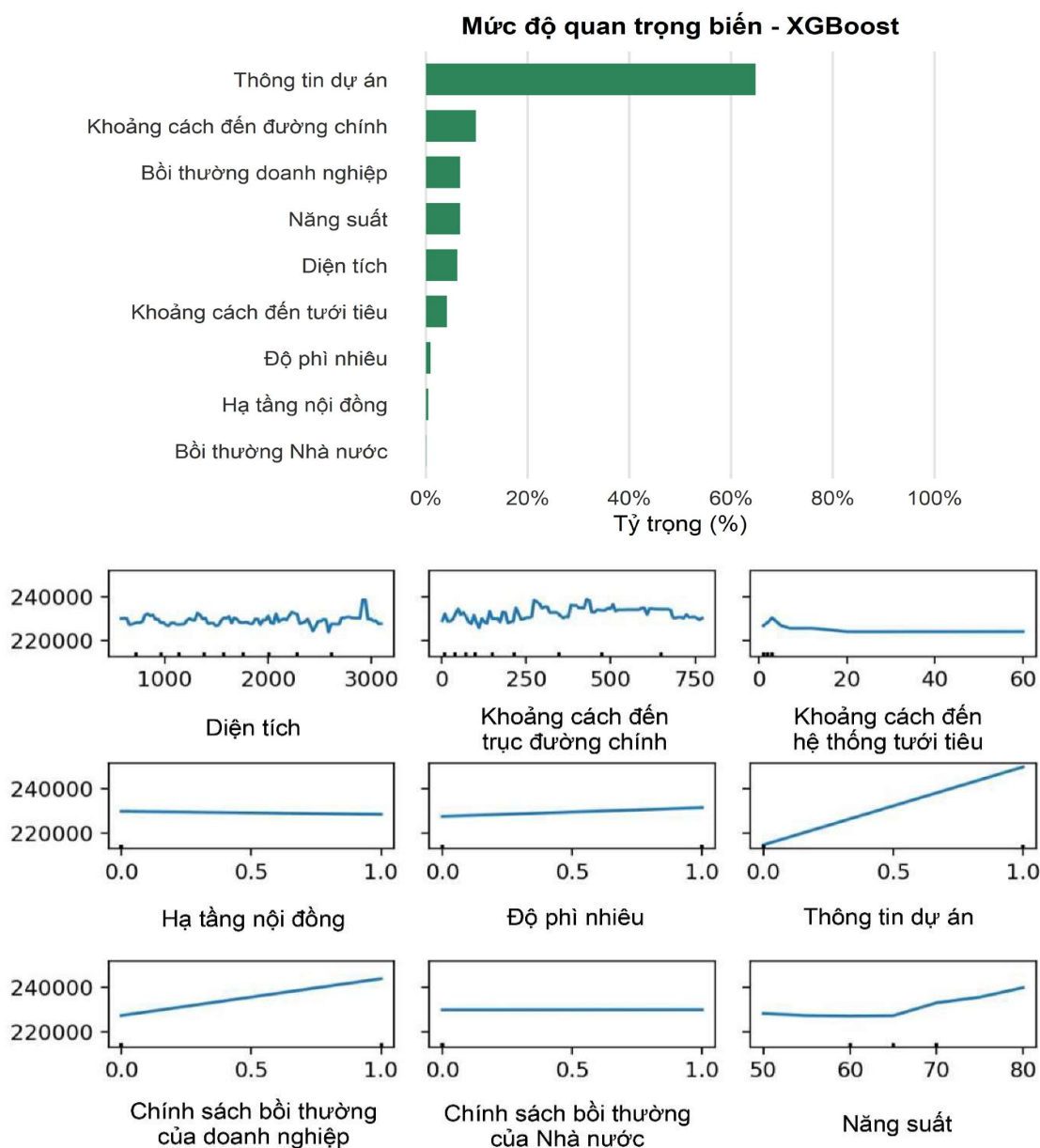


Hình 4. Biểu đồ đánh giá các yếu tố theo mô hình Random Forest

Kết quả từ mô hình Random Forest cho thấy giá đất trồng lúa chịu tác động bởi cả hai nhóm yếu tố, gồm yếu tố liên quan đến sản xuất và yếu tố phi sản xuất. Trong đó, các yếu tố sản xuất như diện tích, năng suất, độ phì nhiêu và các điều kiện tiếp cận hạ tầng phục vụ sản xuất có ảnh hưởng đến giá đất, nhưng mức độ biến thiên tương đối ổn định, phản ánh vai trò nền tảng của giá trị sử dụng đất trong hoạt động canh tác.

Ngược lại, yếu tố thông tin dự án thể hiện vai trò nổi bật trong việc tạo ra sự khác biệt về giá đất. Kết quả cho thấy các khu vực có sự xuất hiện của dự án có mức giá đất cao hơn rõ rệt so với các khu vực không có dự án. Điều này cho thấy giá đất không chỉ phản ánh giá trị hiện tại của hoạt động sản xuất, mà còn phản ánh kỳ vọng về khả năng chuyển đổi sử dụng đất và tiềm năng phát triển trong tương lai.

Bên cạnh đó, yếu tố chính sách bồi thường, đặc biệt là bồi thường từ doanh nghiệp, tiếp tục làm gia tăng giá đất tại các khu vực có dự án. Mức giá cao hơn trong các trường hợp này cho thấy tác động trực tiếp của dòng tiền bồi thường, đồng thời củng cố vai trò của yếu tố dự án trong việc hình thành giá đất.



Hình 5. Biểu đồ đánh giá các yếu tố theo mô hình Xgboost

Kết quả từ mô hình XGBoost cho thấy xu hướng tương tự, trong đó thông tin dự án tiếp tục là yếu tố có ảnh hưởng rõ rệt nhất đến giá đất. So với Random Forest, mô hình XGBoost thể hiện mức độ phân hóa mạnh hơn giữa các khu vực có và không có dự án, cho thấy tác động của yếu tố này được thể hiện rõ hơn về mặt giá trị.

Các yếu tố sản xuất vẫn có ảnh hưởng nhất định, nhưng không tạo ra sự thay đổi lớn về giá đất giữa các khu vực. Ngược lại, các yếu tố phi sản xuất, đặc biệt là thông tin dự án và bồi thường doanh nghiệp, cho thấy sự khác biệt rõ ràng về giá đất, phản ánh vai trò của các yếu tố ngoài sản xuất trong việc quyết định giá trị đất.

So sánh giữa hai mô hình cho thấy sự nhất quán trong việc xác định thông tin dự án là yếu tố chi phối chính đối với giá đất trồng lúa. Cả Random Forest và XGBoost đều cho thấy giá đất tại các khu vực có dự án cao hơn đáng kể, đồng thời phản ánh rõ vai trò của yếu tố này trong việc tạo ra sự phân hóa giá đất theo không gian.

Bên cạnh đó, cả hai mô hình đều ghi nhận chính sách bồi thường của doanh nghiệp có tác động mạnh hơn so với chính sách của Nhà nước, cho thấy vai trò của cơ chế thị trường và dòng vốn đầu tư trong việc định hình giá đất. Yếu tố này không chỉ làm tăng giá đất trực tiếp mà còn góp phần củng cố kỳ vọng giá trị đất tại các khu vực có dự án.

Điểm khác biệt giữa hai mô hình nằm ở mức độ thể hiện ảnh hưởng. Trong khi Random Forest phản ánh xu hướng tổng thể, thì XGBoost thể hiện rõ hơn sự chênh lệch giá đất giữa các nhóm khu vực, đặc biệt là giữa các khu vực có và không có dự án. Tuy nhiên, sự khác biệt này không làm thay đổi kết luận chung.

Nhìn chung, kết quả cho thấy giá đất trồng lúa tại khu vực nghiên cứu chịu chi phối chủ yếu bởi các yếu tố phi sản xuất, trong đó thông tin dự án đóng vai trò trung tâm, còn yếu tố sản xuất đóng vai trò bổ trợ. Kết quả này cũng phù hợp với phân bố không gian giá đất đã được trình bày ở mục trước, khi các khu vực có giá cao tập trung dọc theo các trục phát triển và khu vực có hoạt động đầu tư.

4. Thảo luận

Kết quả nghiên cứu cho thấy giá đất trồng lúa tại khu vực nghiên cứu chịu tác động của cả hai nhóm yếu tố, gồm yếu tố liên quan đến sản xuất và yếu tố phi sản xuất. Tuy nhiên, trong bối cảnh khu vực đang đô thị hóa mạnh, nhóm yếu tố phi sản xuất, đặc biệt là thông tin dự án và chính sách bồi thường, thể hiện vai trò chi phối rõ hơn, trong khi các yếu tố sản xuất đóng vai trò bổ trợ. Kết quả này vừa kế thừa, vừa mở rộng các nghiên cứu trước đây về định giá đất nông nghiệp.

Các nghiên cứu trước đây chủ yếu nhấn mạnh vai trò của các yếu tố sản xuất trong việc hình thành giá đất. Chẳng hạn, chất lượng đất được xác định là yếu tố làm tăng giá trị sử dụng và giá thuê đất thông qua việc nâng cao năng suất, tuy nhiên tác động này có thể bị hạn chế bởi các yếu tố khác như vị trí và khả năng tiếp cận thị trường (Qi et al., 2024). Tương tự, hệ thống thủy lợi và khả năng cung cấp nước được xem là yếu tố quan trọng trong việc duy trì sản xuất lúa và ảnh hưởng trực tiếp đến giá trị đất (Yang & Wang, 2022). Ngoài ra, các nghiên cứu về chất lượng đất cũng cho thấy sự khác biệt về điều kiện đất dẫn đến sự khác biệt về năng suất và giá trị sử dụng đất (Jumiun et al., 2024).

Bên cạnh đó, yếu tố vị trí và khả năng tiếp cận thị trường cũng được xác định là yếu tố quan trọng trong nhiều nghiên cứu. Các nghiên cứu cho thấy khoảng cách đến trục giao thông có ảnh hưởng đáng kể đến giá đất thông qua chi phí vận chuyển và khả năng tiêu thụ sản phẩm (Qi et al., 2024; Srisuksai & Vanitcharearnthum, 2022). Như vậy, phần lớn các nghiên cứu trước tiếp cận giá đất nông nghiệp chủ yếu dưới góc độ giá trị sản xuất và hiệu quả kinh tế trực tiếp.

Tuy nhiên, trong bối cảnh các khu vực ven đô hoặc khu vực có tốc độ đô thị hóa nhanh, cách tiếp cận này chưa phản ánh đầy đủ cơ chế hình thành giá đất. Các nghiên cứu về chuyển đổi sử dụng đất cho thấy đất nông nghiệp có xu hướng tăng giá và chuyển đổi sang mục đích phi nông nghiệp do lợi nhuận cao hơn (Ayub et al., 2022; Barchia et al., 2021). Đồng thời, sự thay đổi quyền sử dụng đất cũng gắn liền với sự thay đổi giá trị kinh tế của đất trong bối cảnh phát triển (Nurjannah, 2022). Tuy nhiên, các nghiên cứu này chủ yếu xem xét kết quả sau chuyển đổi, chưa làm rõ giai đoạn kỳ vọng trước chuyển đổi.

Kết quả nghiên cứu này bổ sung vào khoảng trống đó khi cho thấy thông tin dự án là yếu tố trung tâm chi phối giá đất trồng lúa. Giá đất có xu hướng tăng ngay từ giai đoạn xuất hiện thông tin quy hoạch hoặc dự án, cho thấy giá đất không chỉ phản ánh giá trị hiện tại mà còn phản ánh kỳ vọng phát triển trong tương lai. Kết quả này phù hợp với xu hướng chuyển dịch giá trị đất trong các khu vực chịu tác động của đô thị hóa. Bên cạnh đó, kết quả nghiên cứu cũng chỉ ra rằng chính sách bồi thường của doanh nghiệp có tác động mạnh hơn so với chính sách của Nhà nước. Trong khi các nghiên cứu trước đây thường đề cập đến vai trò của chính sách Nhà nước trong điều tiết thị trường đất đai (Husodo et al., 2021), nghiên cứu này cho thấy sự tham gia của khu vực doanh nghiệp tư nhân có thể làm gia tăng giá đất thông qua cơ chế chi trả theo thị trường. Điều này cho thấy thông tin dự án tạo ra kỳ vọng, trong khi bồi thường doanh nghiệp đóng vai trò nâng cao giá trị đất thông qua việc gia tăng khoản tiền bồi thường.

Xét về mặt phương pháp, nghiên cứu này cũng cho thấy sự khác biệt so với các cách tiếp cận truyền thống. Mô hình Hedonic vẫn là phương pháp phổ biến trong định giá đất do khả năng phân tích giá trị theo từng thuộc tính (Ciardullo & Gennari, 2025). Tuy nhiên, phương pháp này có thể gặp sai lệch nếu không xử lý tốt yếu tố không gian. Các mô hình kinh tế lượng không gian đã được phát triển để khắc phục hạn chế này bằng cách xem xét mối liên hệ giữa các thửa đất lân cận (Donald, 2024).

Trong khi đó, các mô hình học máy ngày càng được sử dụng rộng rãi trong dự báo giá đất nhờ khả năng xử lý mối quan hệ phi tuyến và dữ liệu đa biến (Agosta et al., 2025). Trong nghiên cứu này, kết quả cho thấy cả Random Forest và XGBoost đều có khả năng phản ánh các yếu tố ảnh hưởng đến giá đất, tuy nhiên Random Forest cho kết quả tốt hơn về độ chính xác dự báo, thể hiện qua các chỉ tiêu đánh giá mô hình. Điều này cho thấy Random Forest phù hợp hơn với cấu trúc dữ liệu của nghiên cứu, trong khi XGBoost thể hiện rõ hơn mức độ phân hóa ảnh hưởng của các yếu tố.

Nhìn chung, kết quả nghiên cứu khẳng định rằng trong bối cảnh đô thị hóa, giá đất nông nghiệp không còn được quyết định chủ yếu bởi các yếu tố sản xuất mà ngày càng chịu chi phối bởi các yếu tố phi sản xuất, đặc biệt là thông tin dự án và cơ chế bồi thường của doanh nghiệp. Đồng thời, nghiên cứu cũng cho thấy việc kết hợp mô hình học máy với GIS là một hướng tiếp cận hiệu quả trong phân tích giá đất tại các khu vực có sự phân hóa không gian mạnh.

IV. KẾT LUẬN

Nghiên cứu đã ứng dụng các mô hình học máy kết hợp với hệ thống thông tin địa lý (GIS) để dự báo và phân tích giá đất trồng lúa tại xã Thủy Thanh, thành phố Huế. Trong đó cả 2 mô hình Random Forest và XGBoost đều cho thấy khả năng dự báo tốt, tuy nhiên Random Forest đạt hiệu quả cao hơn về độ chính xác theo các chỉ tiêu đánh giá. Kết quả phân tích cho thấy giá đất trồng lúa không chỉ phụ thuộc vào các yếu tố sản xuất như năng suất, độ phì nhiêu hay điều kiện hạ tầng phục vụ canh tác, mà còn chịu tác động mạnh từ các yếu tố phi sản xuất. Đặc biệt, thông tin dự án được xác định là yếu tố có vai trò chi phối trong việc hình thành và phân hóa giá đất, phản ánh rõ cơ chế kỳ vọng phát triển trong tương lai, trong khi chính sách bồi thường từ doanh nghiệp đóng vai trò nâng cao giá trị đất thông qua tác động tài chính trực tiếp. Trong bối cảnh đô thị hóa, giá đất trồng lúa không chỉ phản ánh thuần túy giá trị sản xuất mà ngày càng gắn với tiềm năng chuyển đổi sử dụng đất và các hoạt động đầu tư. Kết quả nghiên cứu góp phần làm rõ vai trò ngày càng quan trọng của các yếu tố phi sản xuất trong định giá đất trồng lúa, đồng thời cho thấy việc kết hợp mô hình học máy với GIS là một hướng tiếp cận hiệu quả trong phân tích và dự báo giá đất tại các khu vực chịu tác động lớn của quá trình đô thị hóa.

TÀI LIỆU THAM KHẢO

Ủy ban nhân dân xã Thủy Thanh (2024), Quy hoạch sử dụng đất xã Thủy Thanh giai đoạn 2020-2030 tầm nhìn đến năm 2050.

Agosta, M., Schimmenti, E., Franco, C. P. D., & Asciuto, A. (2025). *Agricultural land markets: A systematic literature review on the factors affecting land prices*. *Land*. <https://doi.org/10.3390/land14050978>

Ayub, A., Noorachmat, B. P., & Purwanto, M. Y. J. (2022). Faktor-faktor pendorong konversi lahan sawah di Kabupaten Bantul. *Enviroscientiae*, 18(2), 66. <https://doi.org/10.20527/es.v18i2.14244>

Barchia, M. F., Utama, S. P., Novanda, R. R., & Ishak, A. (2021). *Future uncertainty of sustainable paddy fields in Bengkulu, Indonesia*. <https://doi.org/10.2991/ABSR.K.210609.039>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Ciardullo, V., & Gennari, A. (2025). *Agricultural land valuation—Hedonic pricing and geostatistical advances: A state-of-the-art review*. Retrieved from https://www.scielo.org.ar/scielo.php?pid=S1853-86652025000200224&script=sci_arttext
- Donald, L. J. (2024). *Bayesian spatial analysis of US agricultural land values*. <https://doi.org/10.6084/m9.figshare.25766242>
- FAO. (2017). *The State of Food and Agriculture 2017: Leveraging food systems for inclusive rural transformation*. Rome: FAO.
- Husodo, T., Wulandari, I., Abdoellah, O. S., Cahyandito, M. F., & Shanida, S. S. (2021). Impact of agricultural land changes on farmers' income in Cirasea Watershed, Bandung Regency, West Java. *International Journal of Ecological and Environmental Sciences*, 5(3), 95–104. <https://doi.org/10.26554/IJEMS.2021.5.3.95-104>
- Jumiun, S., Darwis, D., Namriah, N., Ginting, S., Leomo, S., & Zulfikar, Z. (2024). Analisis indeks kualitas tanah pada tipe penggunaan lahan berbeda. *AGRONU Jurnal Agroteknologi*, 3(1), 1–11. <https://doi.org/10.53863/agronu.v3i01.564>
- Nurjannah, A. (2022). Legal due to the transfer of rights to agricultural land to non agriculture. *Yurisdiksi*, 17(4), 367–378. <https://doi.org/10.55173/yurisdiksi.v17i4.122>
- Qi, Y., et al. (2024). Impact of farmland quality on farmland rents and its constraints. *Land Use Policy*. <https://doi.org/10.1016/j.landusepol.2024.107075>
- Srisuksai, P., & Vanitcharearnthum, V. (2022). The pricing model of rice: Evidence from Thailand. *WSEAS Transactions on Business and Economics*, 19, 1245–1254. <https://doi.org/10.37394/23207.2022.19.110>
- World Bank. (2020). *Vietnam's urbanization at a crossroads: Embedding institutions for resilient and inclusive growth*. Washington, DC: World Bank.
- Yang, A., & Wang, Y. (2022). Spatiotemporal variations in farmland rents and its drivers in rural China. *Land*, 11(2), 229. <https://doi.org/10.3390/land11020229>