

Ứng dụng mô hình hồi quy cây quyết định trong định giá hàng loạt đất ở: Trường hợp nghiên cứu tại Quận 4, Thành phố Hồ Chí Minh năm 2025

Võ Quốc Khánh*, Nguyễn Hữu Cường, Nguyễn Văn Anh, Nguyễn Kim Huệ

Trường Đại học Tài nguyên và Môi trường TP. Hồ Chí Minh

Application of decision tree regression for mass appraisal of land: A case study in District 4, Ho Chi Minh City

Vo Quoc Khanh*, Nguyen Huu Cuong, Nguyen Van Anh, Nguyen Kim Hue

Ho Chi Minh City University of Natural Resources and Environment

*Corresponding author: khanh692002@yahoo.com

<https://doi.org/10.55250/jo.vnuf.15.3.2026.078-086>

Thông tin chung:

Ngày nhận bài: 27/11/2025

Ngày phản biện: 03/01/2026

Ngày quyết định đăng: 09/02/2026

Từ khóa:

Bảng giá đất, định giá đất hàng loạt, giá đất, khai phá dữ liệu, mô hình hồi quy cây quyết định.

Keywords:

Data mining, decision tree regression model, land price, land price table, mass land valuation.

TÓM TẮT

Bài báo nghiên cứu ứng dụng phương pháp khai phá dữ liệu với phần mềm DTREG để xây dựng mô hình giá đất ở Quận 4 (TP.HCM) dựa trên các yếu tố định lượng phản ánh điều kiện kinh tế-xã hội và không gian đô thị. Theo hướng định giá hàng loạt (mass appraisal), nghiên cứu sử dụng cây quyết định hồi quy nhằm mô hình hóa quan hệ phi tuyến giữa giá đất và các nhóm biến dự báo như: khả năng kinh doanh, cấp đường (hẻm/mặt tiền), khu vực, công trình vị thế, cùng các chỉ tiêu tiếp cận trong hẻm (rộng, dài hẻm). Kết quả đầu ra không chỉ cung cấp giá dự báo, mà còn tạo các tổ hợp điều kiện vùng giá trị (các nút lá) giúp diễn giải rõ cơ chế hình thành giá và hỗ trợ phân vùng giá đất theo dữ liệu thực nghiệm. Dữ liệu gồm 156 giá đất quan sát (mỗi quan sát là giá trị trung bình của 3 điểm khảo sát) tại 102 đoạn đường theo bảng giá đất Quận 4 và 54 hẻm, với 7 biến dự báo: Cấp đường, khu vực, mật độ/loại hình kinh doanh, công trình vị thế, rộng-dài hẻm. Kết quả cho thấy mô hình đạt $R^2 = 0,8753$, mức phù hợp cao; biến mật độ kinh doanh có mức quan trọng lớn nhất (100%), đóng vai trò nút tách chủ đạo của cây. Từ các nút lá, mô hình hình thành 24 vùng giá trị với giá trung bình khoảng 62–383,44 triệu đồng/m², phản ánh logic tiếp cận (hẻm/mặt tiền), khả năng kinh doanh, khu vực và công trình vị thế. So sánh với giá Nhà nước cho thấy giá của mô hình nhìn chung cao hơn giá trong bảng giá khoảng 1,86 lần.

ABSTRACT

The paper applies data-mining techniques using DTREG software to develop a land price model for District 4 (Ho Chi Minh City) based on quantitative factors that capture local socio-economic conditions and the urban spatial structure. Following a mass appraisal approach, the study employs a regression decision tree to model the nonlinear relationships between land prices and key predictors, including business potential, road class (alley vs. street frontage), subarea, location landmarks, and alley accessibility indicators (alley width and length). The model output not only provides predicted prices but also generates value-zone condition combinations (leaf nodes), which help clarify the pricing mechanism and support evidence-based delineation of land value zones. The dataset comprises 156 observations, including 102 road segments defined from the official price table and 54 alley locations, with seven predictors: road class, subarea, business density/type, location landmarks, and alley width/length. Results show a strong model fit ($R^2 = 0.8753$). Business density is the most influential variable (100%), serving as the primary splitting factor. The leaf nodes form 24 value zones with mean prices of approximately 62–383.44 million VND/m², reflecting access conditions (alley vs. frontage), business potential, subarea, and landmark effects. Compared with the official land price table, the model's estimates are generally higher by about 1.86 times.

1. ĐẶT VẤN ĐỀ

Một trong những thách thức nổi bật trong quản lý đất đai đô thị hiện nay là khoảng cách đáng kể giữa bảng giá đất do Nhà nước ban hành và giá giao dịch thực tế trên thị trường [1]. Khoảng cách này không chỉ làm suy giảm vai trò điều tiết của công cụ giá đất mà còn tạo ra những méo mó trong phân bổ nguồn lực, phát sinh khiếu nại, khiếu kiện và tiềm ẩn rủi ro thất thu ngân sách [2]. Ở cấp độ vi mô, giá trị mỗi thửa đất chịu tác động đồng thời của nhiều yếu tố khác nhau như cấp đường, cấu trúc hẻm, khu vực không gian, mật độ và loại hình kinh doanh, cũng như sự hiện diện của các công trình vị thế (chợ, công viên, khu tòa nhà, trung tâm hành chính...). Khi các yếu tố này chưa được lượng hóa một cách hệ thống trong mô hình định giá, kết quả thường phụ thuộc nhiều vào nhận định chủ quan, dễ sai lệch so với giá trị thị trường và thiếu công bằng trong thực hiện nghĩa vụ tài chính về đất đai [1].

Trong bối cảnh dữ liệu giao dịch và thông tin không gian ngày càng phong phú, các kỹ thuật khai phá dữ liệu và học máy mở ra hướng tiếp cận mới cho định giá đất hàng loạt (mass appraisal), cho phép xây dựng mô hình vừa đạt độ chính xác chấp nhận được, vừa đảm bảo tính minh bạch và khả năng giải thích [3]. Các nghiên cứu gần đây về thẩm định giá hàng loạt đã khai thác nhiều thuật toán học máy khác nhau như hồi quy đa biến, rừng ngẫu nhiên, mạng nơ-ron và các mô hình lai để nâng cao hiệu quả dự báo giá đất và giá bất động sản [4]. Trong nhóm đó, mô hình hồi quy cây quyết định có ưu thế ở khả năng xử lý quan hệ phi tuyến, tương tác phức tạp giữa các biến dự báo và biểu diễn kết quả dưới dạng cấu trúc phân nhánh trực quan [5].

Bài báo này ứng dụng mô hình hồi quy cây quyết định, triển khai trên phần mềm DTREG, để định giá đất hàng loạt trên địa bàn Quận 4, Thành phố Hồ Chí Minh, kế thừa và phát triển các nghiên cứu trước đây về định giá đất hàng loạt ở Việt Nam [6]. Nghiên cứu sử dụng bộ dữ

liệu điều tra giá đất và các yếu tố cấu thành giá trị, xây dựng hệ thống biến dự báo phù hợp với bối cảnh đô thị hẻm đặc thù, từ đó mô hình hóa mối quan hệ giữa các yếu tố vị trí, kinh doanh và công trình vị thế với giá đất thị trường [7]. Trên cơ sở đó, mô hình hình thành các vùng giá trị tương đối đồng nhất và đề xuất hàm ý ứng dụng vào công tác xây dựng bảng giá đất tiệm cận hơn với giá thị trường [8].

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Phương pháp khai phá dữ liệu (data mining) và mô hình hồi quy cây quyết định (Decision tree regression model)

Trong nghiên cứu này, nhóm tác giả sử dụng khai phá dữ liệu (data mining) như một quy trình khoa học nhằm trích xuất thông tin từ bộ dữ liệu giá đất, qua đó nhận diện các yếu tố ảnh hưởng, mức độ ảnh hưởng và mối quan hệ của chúng để phục vụ định giá đất hàng loạt [9]. Cách tiếp cận này đặc biệt phù hợp khi dữ liệu bao gồm đồng thời biến định lượng và định tính, cho phép chuyển hóa dữ liệu quan sát thành các quy tắc phân loại/ước lượng có thể diễn giải [10]. Trên nền tảng đó, nghiên cứu xây dựng mô hình hồi quy cây quyết định bằng phần mềm DTREG để mô tả quan hệ giữa các biến dự báo và giá đất thị trường. Thuật toán bắt đầu từ nút gốc chứa toàn bộ mẫu, sau đó chia tách tuần tự theo biến và ngưỡng tối ưu nhằm giảm sai số dự báo (thường đánh giá bằng MSE) và dừng lại khi đạt các điều kiện về độ sâu cây hoặc kích thước mẫu tối thiểu [7]. Tại các nút lá, giá đất dự báo được xác định bằng giá trị trung bình của các quan sát thuộc nút, từ đó hình thành các mức giá tương đối đồng nhất. Ưu điểm của phương pháp là không đòi hỏi giả định tuyến tính hay phân phối chuẩn, mô tả tốt quan hệ phi tuyến và tương tác giữa các biến, đồng thời dễ giải thích để hỗ trợ xây dựng bảng giá đất phục vụ công tác quản lý nhà nước về đất đai [8].

2.2. Số lượng mẫu điều tra

Thông tin về giá đất và các yếu tố ảnh hưởng đến giá trị đất được thu thập bằng phương

pháp phỏng vấn trực tiếp. Đối tượng phỏng vấn là những người am hiểu nhiều về giá cả đất đai trên thị trường Quận 4 như những chuyên gia về thị trường bất động sản, các cán bộ địa chính và những người môi giới hay công ty môi giới bất động sản. Vị trí điều tra thông tin về giá đất là các tuyến, đoạn đường trong bảng giá đất của địa bàn Quận 4 và các vị trí đất trong hẻm.

Dữ liệu gồm 156 quan sát, mỗi quan sát là giá trị trung bình của 3 điểm khảo sát tại địa bàn Quận 4 được lập bảng tổng hợp, xử lý và đưa vào chạy phần mềm DTREG để xây dựng mô hình cây quyết định.

2.3. Xây dựng các biến cho mô hình cây quyết định

Trong nghiên cứu này, mô hình định giá đất được xây dựng trên cơ sở lựa chọn các biến phản ánh đặc trưng vị trí, khả năng kinh doanh và không gian đô thị bám sát cơ sở pháp lý về các yếu tố hình thành giá đất từ nghị định 71/2024/NĐ-CP [11], đồng thời tham khảo các nghiên cứu định giá đất hàng loạt trước đây ở Việt Nam [6] và đề xuất thêm các vị trí trong hẻm. Các biến được mã hóa và phân cấp theo Bảng 1.

Bảng 1. Các biến dự báo của mô hình giá đất tại Quận 4

TT	Tên trường thông tin	Chú thích
1	TENDUONG	Tên tuyến đường
2	TU	Điểm bắt đầu
3	DEN	Điểm kết thúc
4	GIA DAT THI TRUONG	Giá đất thị trường (triệu đồng/m ²)
5	KHU VUC	KV1, KV2, KV3
6	CAP DUONG	Duongphochinh, Duongpho, Duongnhanh, Hem
7	MAT DO KINH DOANH	0 – 20; 21 – 40; 41 – 60; 61 – 80; 81 – 100 (%)
8	NGANH NGHE KINH DOANH	VP-Khách sạn; Tạp hóa – cà phê, Ẩm thực; Không
9	CONG TRINH VI THE	Chợ, TT hành chính; Công viên; Tòa nhà; Không
10	Ronghem	<3 m; 3-5 m; >5 m
11	Daihem	<100 m; 100-500 m; >500 m

- **Khu vực:** Phân loại không gian nghiên cứu theo ba khu vực để phản ánh khác biệt về vị thế đô thị và mức độ phát triển kinh tế-xã hội. Khu vực 1 (KV1) nằm giáp Quận 1, có chất lượng không gian đô thị cao, gần sông Bến Nghé và các cụm hành chính-tòa nhà; phạm vi được giới hạn theo trục từ Hoàng Diệu đến dọc Bến Vân Đồn hướng về Quận 1. Khu vực 3 (KV3) giáp cầu Tân Thuận qua Quận 7, xa trung tâm, mật độ dân cư thấp hơn và hoạt động kinh doanh kém sôi động; gồm đoạn Nguyễn Tất Thành từ Xóm Chiếu đến cầu Tân Thuận và Tôn Thất Thuyết. Khu vực 2 (KV2) là phần còn lại ở giữa quận (Vinh Khánh, Tôn Đản, các đường số...).

Các biến trên được sử dụng làm đầu vào cho mô hình cây quyết định nhằm lượng hóa mức

độ ảnh hưởng của từng yếu tố đến giá đất thị trường trên địa bàn nghiên cứu, đồng thời kiểm định tính phù hợp của hệ thống biến dự báo trong bối cảnh đô thị có mạng lưới hẻm dày đặc như Quận 4.

3. KẾT QUẢ VÀ THẢO LUẬN

Năm 2024, sau khi Luật đất đai số 31/2024 có hiệu lực thi hành từ ngày 01/8/2024, TP. Hồ Chí Minh đã ban hành Quyết định 79/2024/QĐ-UBND [12], theo đó, giá đất các đoạn đường/tuyến đường ở vị trí 1 là tiếp giáp với mặt tiền đường tăng từ 4 đến 38 lần so với bảng giá trước đó ban hành theo quyết định số 02/2020/QĐ-UBND [13]. Giá đất ở Quận 4 gồm 81 tuyến đường được chuẩn hóa và chia thành 102 đoạn vị trí 1 tiếp giáp mặt đường. Các đường như Nguyễn Tất Thành, Bến Vân Đồn,

Đoàn Văn Bơ có thể chia nhiều đoạn để phản ánh chênh lệch theo từng đoạn. Giá dao động từ 77,5 triệu đồng/m² (một số “đường số”) đến 246,9 triệu đồng/m² (một số đoạn Nguyễn Tất Thành/khu vực Cầu Khánh Hội). Các vị trí trong hẻm (vị trí 2–4) quy đổi theo hệ số của quyết định. Các thửa không mặt tiền quy đổi theo hệ số: vị trí 2 (hẻm ≥5 m) = 0,5 vị trí 1; vị trí 3 (hẻm 3–<5 m) = 0,8 vị trí 2 và vị trí 4 = 0,8 vị trí 3 cho những vị trí còn lại. Độ sâu >100 m thì giảm 10% cho từng vị trí.

Sau đây là kết quả thực hiện định giá đất hàng loạt theo mô hình cây quyết định, các mẫu khảo sát giá đất bám theo các đoạn/tuyến đường trong bảng giá đất theo quyết định số 79/2024/QĐ-UBND của TP. Hồ Chí Minh.

3.1. Sự phù hợp của mô hình

Mô hình cây quyết định được xây dựng từ bộ dữ liệu từ 156 quan sát từ 102 đoạn đường có tên trong bảng giá đất và bổ sung 54 quan sát trong hẻm với 7 biến dự báo để hiệu chỉnh đặc trưng không gian – tiếp cận. Kết quả ở Bảng 2 cho thấy mô hình giải thích 87,53% phương sai ($R^2 = 0,8753$), phản ánh mức phù hợp cao trong bối cảnh định giá hàng loạt (mass appraisal), vốn thường đạt R^2 từ trung bình đến cao tùy cấu trúc thị trường và chất lượng dữ liệu [7]. Sai số dự báo tại các nút lá nằm trong ngưỡng chấp nhận, không ghi nhận chênh hệ thống theo vùng giá trị; do đó mô hình có thể sử dụng như công cụ hỗ trợ xây dựng bảng giá đất.

Bảng 2. Phương sai của mô hình cây quyết định

TT	Chỉ tiêu	Giá trị
1	Phương sai dữ liệu (Variance in initial data sample)	9.321,779
2	Phương sai sau khi chạy mô hình (Residual (unexplained) variance after tree fitting)	1.161,700
3	Tỷ lệ phương sai được giải thích (Proportion of variance explained)	0,8753

3.2. Mức độ quan trọng của từng yếu tố

Mức độ quan trọng của các biến dự báo (Bảng 3) cho thấy mật độ kinh doanh là biến chi phối mạnh nhất (100%), cho thấy giá đất ở Quận 4 bị ảnh hưởng nhiều theo khả năng tạo ra dòng tiền của các tuyến đường. Các biến còn lại có vai trò bổ trợ: cấp đường (5,28) phản ánh chênh lệch mặt tiền–hẻm; ngành nghề kinh doanh (3,76) thể hiện chất lượng cầu và mức

sẵn sàng chi trả khác nhau; khu vực (3,20) đại diện cho vị thế đô thị và lợi thế hạ tầng. Công trình vị thế (2,24) tạo lợi thế cục bộ (công viên, trung tâm hành chính...). Rộng hẻm (1,89) và dài hẻm (0,44) có tác động nhỏ hơn do chỉ ảnh hưởng một phần mẫu (nhóm hẻm) và thường bị chi phối bởi các yếu tố thương mại–vị trí lớn hơn.

Bảng 3. Mức độ quan trọng của các biến dự báo

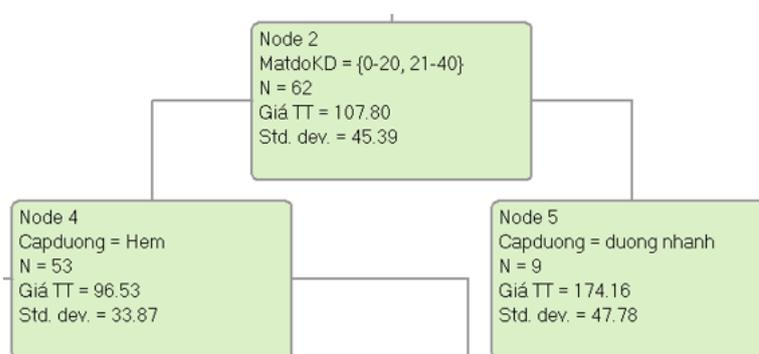
TT	Biến dự báo	Mức độ quan trọng
1	Mật độ kinh doanh	100
2	Cấp đường	5,28
3	Ngành nghề kinh doanh	3,76
4	Khu vực	3,20
5	Công trình vị thế	2,24
6	Rộng hẻm	1,89
7	Dài hẻm	0,44

3.3. Phân nhánh của mô hình cây quyết định

Cây quyết định cho thấy cấu trúc phân tách khá hợp lý với thị trường. Nút gốc tách trước theo mật độ kinh doanh (MatdoKD), xác nhận đây là yếu tố chi phối mạnh nhất mặt bằng giá: nhóm mật độ thấp (0–20; 21–40) kéo giá xuống, nhóm mật độ cao (41–60; 61–80; 81–100) đẩy giá lên rõ rệt. Ở nhánh giá thấp, cây tiếp tục tách theo cấp đường (hẻm/đường nhánh) rồi tinh chỉnh bằng rộng–dài hẻm, khu vực, công trình vị thế, phản ánh điều kiện tiếp cận. Ở nhánh mật độ kinh doanh cao, cây tiếp

tục phân nhánh theo khu vực (KV1–KV3). Tổng thể, mô hình tạo các vùng giá trị khác biệt rõ ràng.

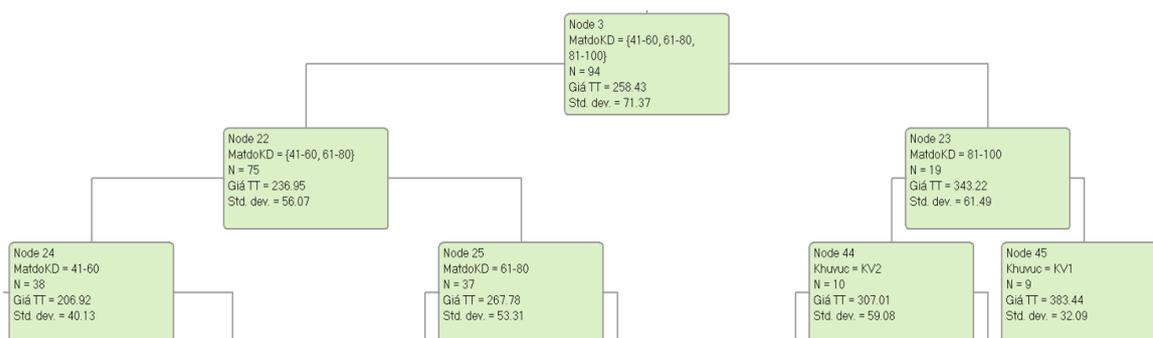
Hình 1 cho thấy nhóm mật độ kinh doanh thấp (0–20%, 21–40%) chỉ có đường nhánh và đường hẻm, mô hình tách tiếp theo cấp đường. Nhóm đường hẻm có giá trung bình thấp hơn (~96,53 triệu đồng/m²) do hạn chế tiếp cận và khả năng khai thác thương mại. Nhóm đường nhánh cho giá đất cao hơn (~174,16) nhờ khả năng tiếp cận và giao dịch tốt hơn.



Hình 1. Một nhánh phân chia từ nút 2

Hình 2 cho thấy trong nhóm mật độ kinh doanh từ 41–100%, cây lại tiếp tục tách theo mật độ kinh doanh, nhóm 41-80% và nhóm 81-100%. Nhóm 81–100% (Nút 23) có giá TB 343,22 triệu đồng/m² cao hơn rõ so với nhóm 41–80% (Nút 22: 236,95) vì khả năng tạo doanh

thu và thanh khoản tốt hơn. Tiếp theo, trong 81–100%, mô hình tách theo yếu tố khu: KV1 đạt 383,44 triệu đồng/m² cao hơn KV2 (307,01), điều này phù hợp với lợi thế về tiện ích tại khu vực giáp ranh với Quận 1.



Hình 2. Một nhánh phân chia từ nút 3

3.4. Kết quả phân tích mô hình cây quyết định

Bảng 4 cho thấy mô hình cây quyết định (DTREG) đã khai phá chuỗi giá trị trong không

gian giá đất ở Quận 4 thành các vùng giá trị theo tổ hợp điều kiện, qua đó phản ánh rõ quy luật hình thành giá.

Bảng 4. Tổ hợp các điều kiện vùng giá trị đất đai và giá đất

TT	Nút số	Giá đất TB (triệu đồng/m ²)	Điều kiện vùng giá trị (tổ hợp biến dự báo)
1	14	62,00	Cấp đường: hẻm; Rộng hẻm: < 3 m; Dài hẻm: > 500 m; Khu vực: KV2 hoặc KV3; Mật độ KD: 21–40%; Công trình vị thế: Không
2	12	62,43	Cấp đường: hẻm; Rộng hẻm: < 3 m; Khu vực: KV2 hoặc KV3; Mật độ KD: 0–20%; Công trình vị thế: Không
3	16	68,5	Cấp đường: hẻm; Rộng hẻm: < 3 m; Dài hẻm: 100–500 m hoặc < 100 m; Khu vực: KV2 hoặc KV3; Mật độ KD: 21–40%; Ngành nghề KD: tạp hóa/cà phê; Công trình vị thế: Không
4	18	78,75	Cấp đường: hẻm; Rộng hẻm: < 3 m; Dài hẻm: 100–500 m hoặc < 100 m; Khu vực: KV3; Mật độ KD: 21–40%; Ngành nghề KD: Không; Công trình vị thế: Không
5	19	84,36	Cấp đường: hẻm; Rộng hẻm: < 3 m; Dài hẻm: 100–500 m hoặc < 100 m; Khu vực: KV2; Mật độ KD: 21–40%; Ngành nghề KD: Không; Công trình vị thế: Không
6	9	107,75	Cấp đường: hẻm; Rộng hẻm: < 3 m; Khu vực: KV1; Mật độ KD: 0–20% hoặc 21–40%
7	20	111,83	Cấp đường: hẻm; Rộng hẻm: 3–5 m hoặc > 5 m; Dài hẻm: < 100 m hoặc > 500 m; Mật độ KD: 0–20% hoặc 21–40%
8	11	118,67	Cấp đường: hẻm; Rộng hẻm: < 3 m; Khu vực: KV2 hoặc KV3; Mật độ KD: 0–20% hoặc 21–40%; Công trình vị thế: Chợ
9	21	147,57	Cấp đường: hẻm; Rộng hẻm: 3–5 m hoặc > 5 m; Dài hẻm: 100–500 m; Mật độ KD: 0–20% hoặc 21–40%
10	30	170,00	Cấp đường: đường nhánh hoặc đường phố; Khu vực: KV1; Mật độ KD: 41–60%; Ngành nghề KD: Không
11	5	174,16	Cấp đường: đường nhánh; Mật độ KD: 0–20% hoặc 21–40%
12	36	193,05	Mật độ KD: 61–80%; Ngành nghề KD: Không hoặc tạp hóa/cà phê hoặc khu ẩm thực; Công trình vị thế: Khu tòa nhà
13	31	203,76	Cấp đường: đường nhánh hoặc đường phố; Khu vực: KV2; Mật độ KD: 41–60%; Ngành nghề KD: Không
14	32	208,06	Cấp đường: đường nhánh hoặc đường phố; Khu vực: KV2; Mật độ KD: 41–60%; Ngành nghề KD: Tạp hóa/cà phê
15	33	208,88	Cấp đường: đường nhánh hoặc đường phố; Khu vực: KV1; Mật độ KD: 41–60%; Ngành nghề KD: Tạp hóa/cà phê
16	40	232,46	Cấp đường: đường nhánh hoặc đường phố; Mật độ KD: 61–80%; Ngành nghề KD: Khu ẩm thực; Công trình vị thế: Công viên hoặc Không hoặc Chợ
17	27	233,72	Cấp đường: đường phố chính; Mật độ KD: 41–60%
18	42	260,52	Cấp đường: đường nhánh hoặc đường phố; Mật độ KD: 61–80%; Ngành nghề KD: Tạp hóa/cà phê; Công trình vị thế: Không
19	43	260,77	Cấp đường: đường nhánh hoặc đường phố; Mật độ KD: 61–80%; Ngành nghề KD: Tạp hóa/cà phê; Công trình vị thế: Công viên
20	46	285,64	Khu vực: KV2; Mật độ KD: 81–100%; Công trình vị thế: Khu tòa nhà hoặc Chợ
21	39	290,67	Cấp đường: đường phố chính; Mật độ KD: 61–80%; Ngành nghề KD: Không hoặc Tạp hóa/cà phê hoặc Khu ẩm thực; Công trình vị thế: Công viên hoặc Không hoặc Chợ
22	47	356,87	Khu vực: KV2; Mật độ KD: 81–100%; Công trình vị thế: Công viên
23	35	357,5	Mật độ KD: 61–80%; Ngành nghề KD: VP–KS
24	45	383,44	Khu vực: KV1; Mật độ KD: 81–100%

Giá đất ở Quận 4 biến thiên theo một logic về địa tô đô thị: khả năng tiếp cận và cường độ hoạt động kinh tế quyết định mặt bằng giá. Nhóm thấp nhất (khoảng 62–84 triệu đồng/m²) tập trung ở hẻm <3 m thuộc KV2–KV3, mật độ kinh doanh thấp và/hoặc hẻm dài (giảm tính thuận tiện và thanh khoản). Khi đưa yếu tố chợ vào cùng điều kiện hẻm hẹp, giá tăng lên (118,67 triệu đồng/m²), phản ánh mức tiêu dùng làm tăng doanh thu kỳ vọng cho hộ kinh doanh nhỏ. Khi mở rộng tiếp cận (hẻm 3–5 m hoặc >5 m; hẻm ngắn/trung bình), giá nhích lên (111,83–147,57 triệu đồng/m²) do xe cộ ra vào thuận lợi và khả năng giao dịch tốt hơn.

Ở nhóm đường nhánh/đường phố, chỉ cần mật độ KD 41–60% đã đạt 170–208 triệu đồng/m², và tăng mạnh ở 61–80% (232–261 triệu đồng/m²) cho thấy Quận 4 có cấu trúc kinh tế dịch vụ–thương mại rõ rệt, giá trị đất cao hơn những nơi có mức độ kinh doanh cao.

Đỉnh giá xuất hiện khi mật độ KD 81–100% kết hợp KV1 (383,44 triệu đồng/m²) hoặc các tiện ích/không gian công cộng (công viên), phản ánh lợi thế trung tâm, tiện ích đô thị và khả năng khai thác thương mại cao.

3.5. So sánh với giá trong bảng giá đất Nhà nước theo Quyết định 79/2024/QĐ-UBND của TP. Hồ Chí Minh

Theo Quyết định 79/2024/QĐ-UBND [12], giá đất địa bàn Quận 4 đã tăng lên từ 4 đến 38 lần so với bảng giá đất cũ, tuy nhiên, trên toàn bộ 156 quan sát tại địa bàn Quận 4, giá đất của mô hình nhìn chung cao hơn giá đất trong bảng giá 1,86 lần với số trung vị là 1,76 lần. Điều này cho thấy sự chênh lệch về giá đất thực tế ngoài thị trường và giá đất trong bảng giá vẫn còn, nhưng mức chênh lệch chủ yếu là dưới 2 lần.

Để so sánh giữa giá Nhà nước (Giá NN) và giá mô hình (Giá MH), nghiên cứu lựa chọn các mẫu minh họa theo phương pháp phân tầng theo khu vực (KV1, KV2, KV3) và tiếp tục phân theo cấp đường (đường phố chính, đường phố, đường nhánh) nhằm bảo đảm tính đại diện về không gian và hạ tầng. Trong từng nhóm, các quan sát đại diện được chọn theo các phân vị của Giá NN (ví dụ P25–P50–P75), tương ứng với các mức thấp – trung bình – cao của bảng giá Nhà nước. Cách chọn này giúp mẫu trích phân bố đều theo cấu trúc bảng giá, qua đó làm rõ mức chênh lệch của giá mô hình so với giá Nhà nước.

Bảng 5. So sánh giá đất trong bảng giá nhà nước và giá mô hình (trích)

ĐVT: triệu đồng/m²

TT	Tên đường	Từ	Đến	Giá NN	Giá MH	Tỷ lệ giá MH/NN
1	Bến Vân Đồn	Cầu Nguyễn Kiệu	Nguyễn Khoái	164,7	383,44	2,33
2	Đoàn Như Hải	Trọn đường		176,8	383,44	2,17
3	Đoàn Văn Bơ	Hoàng Diệu	Tôn Đản	172,2	290,67	1,69
4	Đường 10C	Trọn đường		117,3	208,88	1,78
5	Đường Phụng 1	Trọn đường		87,3	232,46	2,66
6	Đường số 5	Trọn đường		123,1	203,76	1,66
7	Đường 15	Trọn đường		123,1	174,16	1,41
8	Đường 50	Trọn Đường		82	208,88	2,55
9	Hoàng Diệu	Nguyễn Tất Thành	Đoàn Văn Bơ	246,9	383,44	1,55
10	Lê Quốc Hưng	Hoàng Diệu	Lê Văn Linh	236,2	285,64	1,21
11	Nguyễn Hữu hào	Bến Vân Đồn	Hoàng Diệu	124,4	232,46	1,87
12	Nguyễn Tất Thành	Cầu Khánh Hội	Lê Văn Linh	246,9	383,44	1,55
13	Nguyễn Thân Hiến	Trọn đường		110,8	233,72	2,11

TT	Tên đường	Từ	Đến	Giá NN	Giá MH	Tỷ lệ giá MH/NN
14	Nguyễn Trường Tộ	Trộn đường		194,7	383,44	1,97
15	Tôn Đản	Trộn đường		130,4	285,64	2,19
16	Tôn Thất Thuyết	Nguyễn Tất Thành	Nguyễn Thân Hiến	104,3	233,72	2,24
17	Trương Đình Hối	Trộn đường		149,7	174,16	1,16
18	Vĩnh Hội	Tôn Đản	Khánh hội	172,8	356,87	2,07
19	Vĩnh Khánh	Bến Vân Đồn	Hoàng Diệu	130,9	232,46	1,78
20	Xóm Chiếu	Trộn đường		105,3	232,46	2,21

Qua Bảng 5 cho thấy giá MH luôn cao hơn giá NN, với tỷ lệ giá MH/NN dao động 1,16–2,66 cho thấy mức chênh lệch hệ thống giữa bảng giá hành chính nhà nước và mức giá mô hình theo thị trường.

Điểm đáng chú ý là độ vênh không đồng nhất giữa các tuyến: các tuyến có giá NN thấp như đường Phường 1 (87,3 triệu đồng/m²) hay đường 50 (82 triệu đồng/m²) lại có tỷ lệ chênh lệch giữa giá MH và giá NN rất cao (2,66 và 2,55), phản ánh khả năng bảng giá NN “chưa bắt kịp” giá trị sử dụng/khả năng sinh lợi tại một số vị trí. Ngược lại, các tuyến có giá NN cao như Hoàng Diệu/Nguyễn Tất Thành (246,9 triệu đồng/m²) hoặc Lê Quốc Hưng (236,2 triệu đồng/m²) có tỷ lệ chênh lệch thấp hơn, khoảng 1,21 đến 1,55. Điều này cho thấy tại các tuyến đường quan trọng, khung giá nhà nước đã tiệm cận hơn so với mức định giá của mô hình.

Kết quả so sánh tại Quận 4 cho thấy định giá hàng loạt bằng cây quyết định (DTREG) tiệm cận mặt bằng giá thị trường hơn so với bảng giá Nhà nước hiện hành. Giá mô hình thường cao hơn giá Nhà nước, với tỷ số điển hình khoảng 1,7–2,0 lần; mức chênh lệch biến thiên theo khu vực (KV1–KV3) và điều kiện tiếp cận (mặt tiền/hẻm), cho thấy bảng giá đất hiện hành còn độ trễ và chưa phản ánh đầy đủ các yếu tố thị trường. Do đó, mô hình có thể được sử dụng như công cụ hỗ trợ quản lý để rà soát theo tuyến/đoạn đường, phân vùng những khu vực có mức định giá thấp hơn thực tế và đề xuất hệ số điều chỉnh theo không gian

hạ tầng, góp phần tăng tính nhất quán, minh bạch và công bằng trong định giá đất.

4. KẾT LUẬN

Kết quả nghiên cứu cho thấy mô hình hồi quy cây quyết định triển khai bằng DTREG đáp ứng tốt mục tiêu định giá đất hàng loạt tại Quận 4, với mức phù hợp cao ($R^2 = 0,8753$). Trên cơ sở 156 quan sát và 7 biến dự báo, mô hình không chỉ ước tính giá mà còn khai phá không gian giá đất thành 24 vùng giá trị tương đối đồng nhất, với mức giá trung bình khoảng 62–383,44 triệu đồng/m², phản ánh thực tế giá đất theo khả năng kinh doanh, điều kiện tiếp cận (mặt tiền/hẻm), khu vực và công trình vị thế. Trong đó, mật độ kinh doanh là biến có mức độ quan trọng cao nhất (100) nhấn mạnh vai trò của mức độ kinh doanh, khả năng tạo dòng tiền trong hình thành giá đất tại khu vực nghiên cứu. So sánh với bảng giá đất theo Quyết định 79/2024/QĐ-UBND [12] thì giá mô hình cao hơn trung bình 1,86 lần, điều này cho thấy bảng giá đất hiện hành vẫn còn độ trễ so với mức giá quan sát. Trên cơ sở kết quả nghiên cứu, khuyến nghị áp dụng mô hình cây quyết định cho định giá đất hàng loạt với các biến dự báo được hiệu chỉnh theo đặc thù từng địa phương. Bên cạnh đó, nên bổ sung chỉ tiêu định lượng về khả năng kinh doanh vào hệ thống yếu tố hiệu chỉnh bảng giá đất để nâng cao mức độ tiệm cận giá đất thị trường.

TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Quỳnh Hoa (2015). Ứng dụng phương pháp thẩm định giá hàng loạt trong quản lý nhà nước về đất đai. Tạp chí Phát triển và Hội nhập. 22: 32-40.

- [2]. Nguyễn Quỳnh Hoa & Nguyễn Thạch (2013). Định giá đất hàng loạt bằng mô hình hồi quy. Tạp chí Phát triển Kinh tế. 269: 11-19.
- [3]. Wang D. & Li V. J. (2019). Mass appraisal models of real estate in the 21st century: A systematic literature Review Sustainability. 11(24): 7006.
- [4]. Bilgilioglu S. S. & Yilmaz H. M. (2023). Comparison of different machine learning models for mass appraisal of real estate. Survey Review. 55(388): 32-43.
- [5]. Yilmazer S. & Kocaman S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. Land Use Policy. 99.
- [6]. Nguyễn Hữu Cường (2022). Định giá đất hàng loạt ứng dụng mô hình cây quyết định: Trường hợp nghiên cứu thành phố Vũng Tàu. Tạp chí Khoa học Đại học Quốc gia Hà Nội: Các Khoa học Trái đất và Môi trường. 38(1): 1-11.
- [7]. Bui Ngoc Tu, Tran Quoc Binh & Bui Thi Cam Ngoc (2024). Analyzing mass appraisal of urban residential land with machine learning – A case study in Hanoi, Vietnam. Geoinformatics for Spatial-Infrastructure Development in Earth and Allied Sciences. Springer, Cham. 297-314.
- [8]. Le Hoang Nam Thanh (2024). Application of hedonic model to determine residential land valuation in suburban areas: The case of Hue city, Vietnam. Real Estate Management and Valuation. 32(2): 70-79.
- [9]. Han J., Kamber M. & Pei J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann, Waltham.
- [10]. Reyes-Bueno F., García-Samaniego J. M. & Sánchez-Rodríguez A. (2018). Large-scale simultaneous market segment definition and mass appraisal using decision tree learning for fiscal purposes. Land Use Policy. 79: 116-122.
- [11]. Chính phủ (2024). Nghị định số 71/2024/NĐ-CP ngày 27/6/2024 quy định về giá đất.
- [12]. UBND TP. Hồ Chí Minh (2024). Quyết định 79/2024/QĐ-UBND ngày 21/10/2024 về bảng giá đất TP. HCM (sửa đổi, bổ sung QĐ 02/2020/QĐ-UBND).
- [13]. UBND TP. Hồ Chí Minh (2020). Quyết định 02/2020/QĐ-UBND ngày 16/01/2020 về bảng giá đất TP. HCM giai đoạn 2020-2024.