



Original Article

Application of Decision Tree Model in Mass Land Valuation: A Case Study in Vung Tau City

Nguyen Huu Cuong*

*Ho Chi Minh City University of Natural Resources and Environment,
236B Le Van Sy, Ward 1, Tan Binh, Ho Chi Minh City, Vietnam*

Received 18 March 2020

Revised 29 August 2020; Accepted 06 September 2020

Abstract: Land valuation is one of many important tasks in land use and management. In order to be objective and scientific in land valuation, it is necessary to identify factors affecting land price and quantify the relationship between land feature factors and land value. The goal of the study is to apply the decision-tree regression model to build mass land pricing models in Vung Tau city. The decision-tree regression model is applied in land valuation by determining the relationship between a combination of land feature factors and corresponding averaged land price. Research on building the land pricing model with independent variables includes business density, area, road types, business activities and road width based on 883 survey samples. The final result of this tree model was categorized into 51 divided groups, and pruned trees included 15 divided groups. The importance of the independent variables, namely, business density, region, road level, business activities and road width are 100%, 83.48%, 78.87%, 58.78% and 10.03% respectively. The interpretation rate of the independent variables to the formation of land price model is 86%. The decision-tree regression model suggests another approach to land valuation theory.

Keywords: Decision tree, mass valuation, model, land price, Vung Tau city.

* Corresponding author.

E-mail address: nhcuong@hcmunre.edu.vn

<https://doi.org/10.25073/2588-1094/vnuces.4588>

Định giá đất hàng loạt ứng dụng mô hình cây quyết định: Trường hợp nghiên cứu thành phố Vũng Tàu

Nguyễn Hữu Cường*

*Trường Đại học Tài nguyên và Môi trường Thành phố Hồ Chí Minh,
236B Lê Văn Sỹ, Phường 1, Tân Bình, Thành phố Hồ Chí Minh, Việt Nam*

Nhận ngày 18 tháng 3 năm 2020

Chỉnh sửa ngày 29 tháng 8 năm 2020; Chấp nhận đăng ngày 06 tháng 9 năm 2020

Tóm tắt: Định giá đất là một nội dung quan trọng của công tác quản lý và sử dụng đất đai. Để công tác định giá đất được khách quan và khoa học, việc xác định các yếu tố ảnh hưởng đến giá đất và lượng hóa mối quan hệ giữa các yếu tố đặc điểm đất đai với giá trị đất đai là rất cần thiết. Mục tiêu của nghiên cứu là ứng dụng cây quyết định xây dựng mô hình định giá đất hàng loạt trên địa bàn thành phố Vũng Tàu. Mô hình cây quyết định được ứng dụng trong định giá đất bằng cách xác định mối quan hệ giữa tổ hợp các yếu tố đặc điểm đất đai và giá đất trung bình tương ứng. Nghiên cứu xây dựng mô hình giá đất với các biến độc lập là mật độ kinh doanh, khu vực, cấp đường, ngành nghề kinh doanh và độ rộng đường trên cơ sở 883 mẫu điều tra. Kết quả mô hình cây đầy đủ được tạo ra với 51 nhóm chia, cây rút gọn (tỉa) tạo ra 15 nhóm chia. Tầm quan trọng của các biến độc lập mật độ kinh doanh, khu vực, cấp đường, ngành nghề kinh doanh và độ rộng đường lần lượt bằng 100%, 83,48%, 78,87%, 58,78% và 10,03%. Tỷ lệ giải thích của các biến độc lập đến sự hình thành giá đất theo mô hình đạt 86%. Mô hình cây quyết định gợi ý cho chúng ta một cách tiếp cận khác trong lý thuyết định giá đất đai.

Từ khóa: Cây quyết định, định giá hàng loạt, mô hình, giá đất, thành phố Vũng Tàu.

1. Mở đầu

Trong công tác quản lý và sử dụng đất đai, giá đất luôn là vấn đề được quan tâm vì nó ảnh hưởng đến nguồn thu của nhà nước, nghĩa vụ về tài chính của các đối tượng sử dụng đất cũng như quyền lợi của họ khi bị nhà nước thu hồi đất. Không giống như các tài sản thông thường khác, xác định chính xác giá đất được xem là công tác khó khăn.

Thậm chí định giá hàng loạt là định giá có hệ thống một nhóm tài sản vào một thời điểm nhất định, áp dụng những phương pháp được chuẩn

hóa và kiểm tra thống kê [1]. Định giá đất hàng loạt về mặt không gian thể hiện qua vùng giá trị đất đai. Dưới góc độ giá trị tiền tệ ở nước ta, vùng giá trị đất đai được hiểu là tập hợp các đối tượng được ước tính giống nhau về giá trị của quyền sử dụng đất trên một đơn vị diện tích, hay còn gọi là vùng giá đất đai [2].

Để công tác định giá đất được khách quan và khoa học, việc xác định các yếu tố ảnh hưởng đến giá đất và lượng hóa mối quan hệ giữa các yếu tố đặc điểm đất đai với giá trị đất đai là rất cần thiết. Điều này có thể thực hiện thông qua phân tích thống kê các yếu tố ảnh hưởng đến đất

* Tác giả liên hệ.

Địa chỉ email: nhcuong@hcmunre.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4588>

đai. Đã có nhiều nghiên cứu về mô hình định giá đất trên thế giới cũng như tại Việt Nam với phương pháp truyền thống là xây dựng các hàm hồi quy hedonic giá đất. Nguyễn Quỳnh Hoa và cộng sự [1] xây dựng mô hình cộng định giá đất hàng loạt cho địa bàn quận Gò Vấp, TP.HCM. Kết quả cho thấy các yếu tố ảnh hưởng mạnh nhất đến giá đất tại quận Gò Vấp là: An ninh, lộ giới, lợi thế kinh doanh, khoảng cách đến mặt tiền; trục giao thông, trình độ dân trí và vị trí trung tâm. Phạm Sỹ Liêm và cộng sự [3] sử dụng phương pháp định giá hàng loạt từ mô hình hồi quy giá đất kết hợp với phân tích nội suy không gian để tính toán toàn bộ vùng giá trị liên tục của xã Dương Quang, huyện Gia Lâm, Hà Nội. Mô hình giá đất ở nông thôn gồm 9 biến với các yếu tố chính ảnh hưởng như loại đường tiếp giáp, khoảng cách đến đường tiếp giáp, khoảng cách đến chợ.

Nghiên cứu của Gang-Zhi Fan và cộng sự [4] cho rằng phương pháp hồi quy hedonic đã được sử dụng rộng rãi để điều tra mối quan hệ giữa giá bất động sản và đặc điểm của nó. Tuy nhiên, cách tiếp cận này phải chịu những chỉ trích phát sinh từ các vấn đề tiềm ẩn liên quan đến các giả định và ước lượng mô hình như việc xác định cung và cầu, mất cân bằng thị trường, phân khúc thị trường, lựa chọn các biến độc lập, lựa chọn hình thức chức năng của phương trình hedonic. Đồng thời, nghiên cứu này giới thiệu một cách tiếp cận khác, cách tiếp cận cây quyết định, là một công cụ nhận dạng mẫu thống kê quan trọng. Cây quyết định với ưu điểm là mô hình không tham số, có thể xử lý các biến định lượng và định tính dễ dàng nên khắc phục các nhược điểm của mô hình hàm hồi quy. Sử dụng dữ liệu thị trường nhà ở tại Singapore, nghiên cứu cho thấy sự hữu ích của kỹ thuật này trong việc kiểm tra mối quan hệ giữa giá nhà và đặc điểm nhà ở, xác định các yếu tố quan trọng quyết định giá nhà đất và dự đoán giá nhà đất.

Theo hướng nghiên cứu trên, Claudio Acciani và cộng sự [5] đã sử dụng mô hình cây xác định các yếu tố tác động đến giá nông trại tại Ý. Các yếu tố khoảng cách đến thị trấn, diện tích, sự thuận lợi về giao thông, điện nước được xây dựng để đưa vào mô hình. Mingchin Chen và

cộng sự [6] đã sử dụng cây quyết định để đánh giá mức độ quan trọng của các yếu tố tác động đến giá nhà tại Đài Loan. Bốn yếu tố khu vực, diện tích khu đất, diện tích nhà, diện tích bãi đỗ xe tác động lớn nhất đến giá nhà ở. Tại Việt Nam, Nguyễn Thị Mỹ Linh [7] đã ứng dụng mô hình cây quyết định trong định giá đất hàng loạt ở thành phố Hồ Chí Minh. Công trình sử dụng kỹ thuật phân tích hồi quy trên cơ sở thiết lập mối quan hệ giữa giá trị của đất đai với các đặc tính riêng của đất. Trần Thanh Hùng và cộng sự [8] với nghiên cứu xây dựng hệ thống thông tin hỗ trợ định giá đất ứng dụng công nghệ WebGIS mã nguồn mở đã xây dựng mô hình giá đất hàng loạt trên địa bàn thành phố Bạc Liêu ứng dụng cây quyết định. Theo mô hình cây quyết định giá đất xác định được 21 tổ hợp các yếu tố đặc điểm vị thế đất đai tương ứng với 21 vùng giá trị. Mục tiêu của nghiên cứu là ứng dụng cây hồi quy quyết định thành lập mô hình định giá đất hàng loạt tại thành phố Vũng Tàu.

2. Phương pháp và quy trình nghiên cứu

2.1. Phương pháp nghiên cứu

2.1.1. Phương pháp phân tích cây quyết định

Mô hình cây quyết định (decision tree) là một công cụ mạnh và hiệu quả trong việc phân lớp và dự báo sử dụng trong khai phá dữ liệu lớn (big data mining). Thuật toán cây quyết định được xây dựng trên mô hình một cây mục đích thể hiện cấu trúc của một hệ thống ra quyết định, hay nói cách khác, cách con người tư duy logic ra sao để đi đến quyết định cuối cùng [8]. Trong lĩnh vực dữ liệu, cây quyết định thể hiện mối quan hệ giữa các yếu tố và sự tác động của chúng đến biến mục tiêu.

Cây quyết định gồm có gốc cây (root node) là điểm chứa giá trị của biến đầu tiên được dùng để phân nhánh; thân cây (internal node) là các biến chứa các thuộc tính, giá trị dữ liệu được dùng để xét cho các phân nhánh tiếp theo; lá cây (leaf node) chứa giá trị của biến phân loại sau cùng. Các nút (node) được nối với nhau bởi nhánh (branch) là quy luật phân nhánh, thể hiện

mối quan hệ giữa giá trị của biến độc lập và giá trị của biến mục tiêu. Một dự đoán được tạo bằng cách căn cứ vào cây từ gốc, theo nhánh dựa vào giá trị biến dự báo cho đến khi tới lá. Mỗi lá chỉ ra giá trị có khả năng phù hợp nhất cho biến mục tiêu đã cho bởi giá trị dự báo dẫn đến lá.

Độ lợi thông tin (Information Gain) và chỉ số Gini (Gini index) được dùng cho mô hình cây quyết định phân lớp (classification tree) có biến kết quả là biến rời rạc. Đối với mô hình cây quyết định hồi quy (regression tree) với biến kết quả là biến liên tục thì việc xây dựng cây quyết định lại dựa vào phương pháp bình phương bé nhất (least squares). Phép tách được thực hiện sao cho tổng phương sai của các dòng trong mỗi node so với giá trị trung bình mỗi node là bé nhất. Trong nghiên cứu, cây quyết định hồi quy được sử dụng với giá trị biến kết quả (giá đất) là biến liên tục.

Để đánh giá mô hình cây quyết định cần phân tích các thông số:

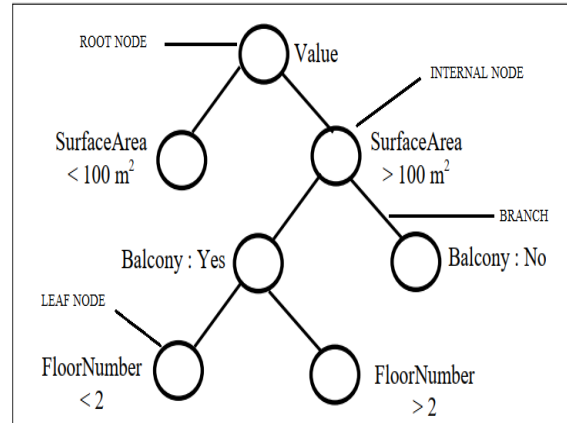
- Phương sai trong toàn bộ dữ liệu học trước khi thực hiện bất cứ phép tách nào (Variance in initial data sample).

- Phương sai tồn tại sau khi cây được ứng dụng với dữ liệu dự báo giá trị mục tiêu (Residual variance after tree fitting). Nếu cây dự báo hoàn hảo tập dữ liệu thì phương sai phần dư sẽ là 0,0.

- Tỷ lệ phản ánh mức độ giải thích phương sai ban đầu bởi cây quyết định (Proportion of variance explained). Giá trị càng lớn thì độ thích hợp của cây càng lớn và càng có thể giải thích dữ liệu. Nếu cây phù hợp dữ liệu và dự báo chính xác giá trị mục tiêu cho mỗi dòng, tỷ lệ phương sai được giải thích là 1,0 (100%) [10].

Ứng dụng cây quyết định trong định giá đất hàng loạt bằng việc xác định những tổ hợp các yếu tố ảnh hưởng đến giá đất (Khu vực, loại đường, ngành nghề kinh doanh, mật độ kinh doanh,...) và giá đất trung bình tương ứng [8]. Theo sơ đồ cây quyết định thì mỗi vùng giá trị đất là một tổ hợp các yếu tố đặc điểm đất đai được xác định theo các nhánh từ nút gốc đến từng nút lá.

Nghiên cứu sử dụng phần mềm DTREG để xây dựng mô hình cây quyết định.



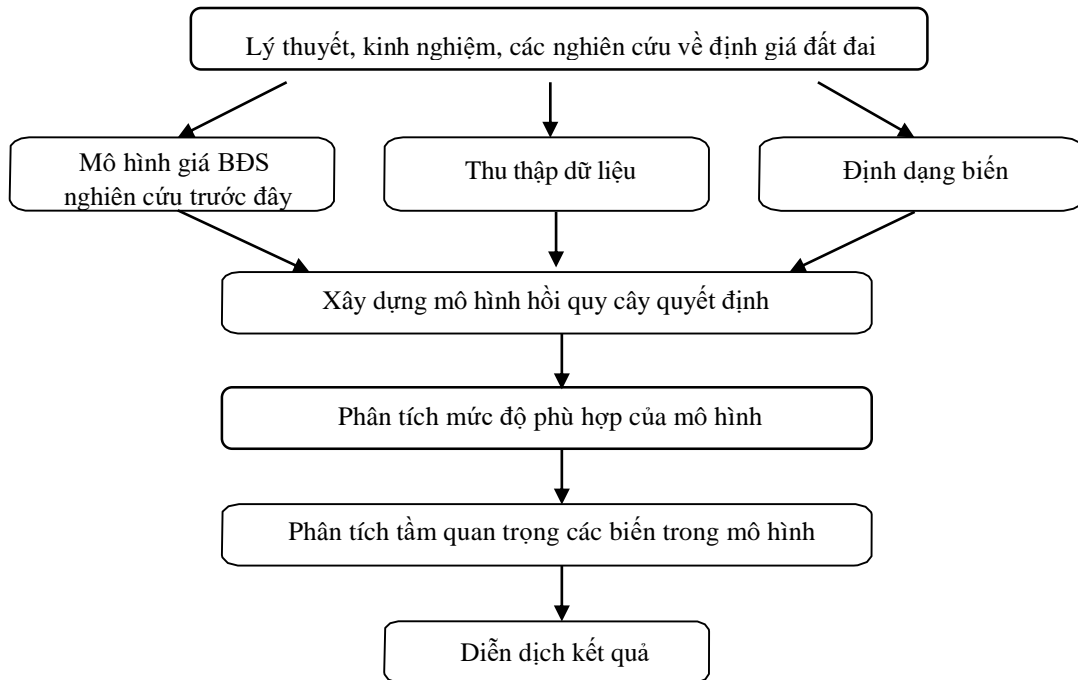
Hình 1. Ví dụ mô hình cây quyết định trong định giá bất động sản [9].

2.1.2. Phương pháp điều tra, khảo sát

Bảng 1. Tổng hợp phiếu điều tra giá đất

STT	Điểm điều tra	Số phiếu	%
1	Phường 1	65	7,36
2	Phường 2	64	7,25
3	Phường 3	62	7,02
4	Phường 4	55	6,23
5	Phường 5	60	6,80
6	Phường 7	53	6,00
7	Phường 8	42	4,76
8	Phường 9	54	6,12
9	Phường 10	65	7,36
10	Phường 11	45	5,10
11	Phường 12	50	5,66
12	Phường Thăng Nhất	61	6,91
13	Phường Thăng Nhì	54	6,12
14	Phường Thăng Tam	47	5,32
15	Phường Nguyễn An Ninh	57	6,46
16	Phường Rạch Dừa	49	5,55
Tổng		883	100,00

Dùng phiếu điều tra để thu thập các thông tin liên quan đến giá đất đai đối với các thửa đất có phát sinh biến động chuyển nhượng trong năm 2019. Nghiên cứu sử dụng phiếu điều tra giá đất trên địa bàn thành phố Vũng Tàu của dự án “Xây dựng cơ sở dữ liệu về giá đất và lập bản đồ giá đất tỉnh Bà Rịa - Vũng Tàu” kết hợp điều tra ngoại nghiệp với 883 mẫu.



Hình 2. Quy trình xây dựng mô hình định giá ứng dụng cây quyết định.

Bảng 2. Tóm tắt kích thước mô hình cây quyết định

STT	Chỉ tiêu	Giá trị
1	Độ sâu tối đa của cây (Maximum depth of the tree)	10
2	Số nhóm chia (Total number of group splits)	50
3	Số lá cây đầy đủ (The full tree has 51 terminal (leaf) nodes)	51
4	Số lá cây tía (With 0.500 S.E. allowance)	15

Thông tin của phiếu điều tra gồm: giá đất/giá bất động sản chuyển nhượng; vị trí, diện tích, mục đích thửa đất; khoảng cách từ thửa đất đến các trung tâm, hiện trạng hạ tầng giao thông (bề rộng, lớp phủ mặt đường),... Ngoài ra, còn điều tra ngoại nghiệp bổ sung để xác định ngành nghề, mật độ kinh doanh tại mặt tiền các tuyến đường.

2.2. Quy trình nghiên cứu

Quy trình thực hiện bắt đầu từ việc nghiên cứu các lý thuyết về kinh tế và định giá đất đai,

các mô hình giá bất động sản của các nghiên cứu trước đây để xây dựng khung lý thuyết, định dạng các biến cho mô hình và thu thập dữ liệu. Mô hình cây quyết định sau khi được thành lập sẽ tiến hành phân tích mức độ phù hợp, tầm quan trọng các biến của mô hình và diễn dịch kết quả (Hình 2).

3. Kết quả và thảo luận

3.1. Kích thước mô hình cây quyết định

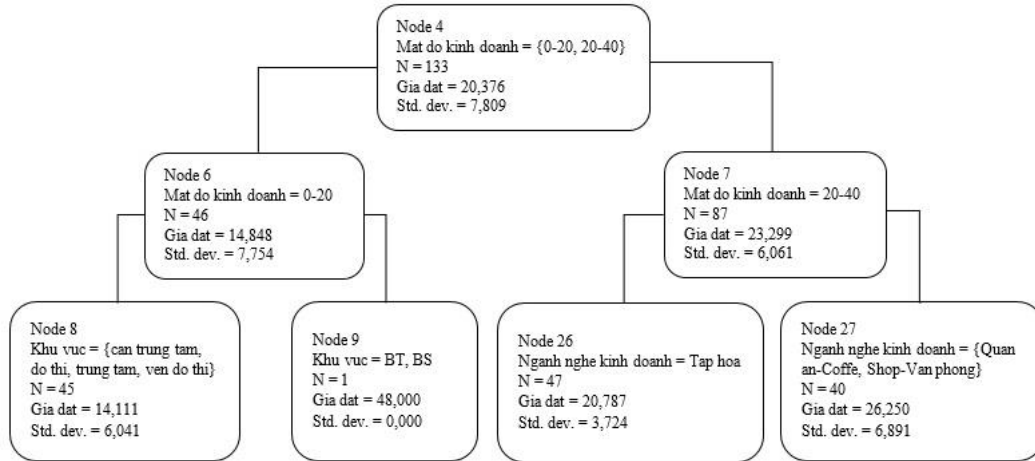
Khác với định giá cá biệt, định giá hàng loạt chỉ tính đến các yếu tố quan trọng nhất tác động đến giá trị bất động sản cũng như các quy luật khách quan của sự thay đổi giá trị [1]. Dựa trên cơ sở lý thuyết và nghiên cứu thực nghiệm tại địa bàn khảo sát, mô hình cây quyết định trong nghiên cứu được sử dụng với:

- Biến mục tiêu (target): giá đất (biến liên tục, tr/m²).
- Các biến dự báo (predictors): mật độ kinh doanh (biến rời rạc), khu vực (biến rời rạc), cấp

đường (biến rời rạc), ngành nghề kinh doanh (biến rời rạc) và độ rộng đường (biến liên tục, m).

Với bộ dữ liệu điều tra, mô hình đã xây dựng cây quyết định đầy đủ với 10 tầng, 50 nhánh

(group split), 51 lá (leaf). Với sai số chuẩn cho phép là 0,5 để giảm kích thước, cây quyết định cho ra mô hình tối ưu với 15 lá.



Hình 3. Một “nhánh” kết quả mô hình cây quyết định.

3.2. Phân tích phương sai và độ lợi mô hình

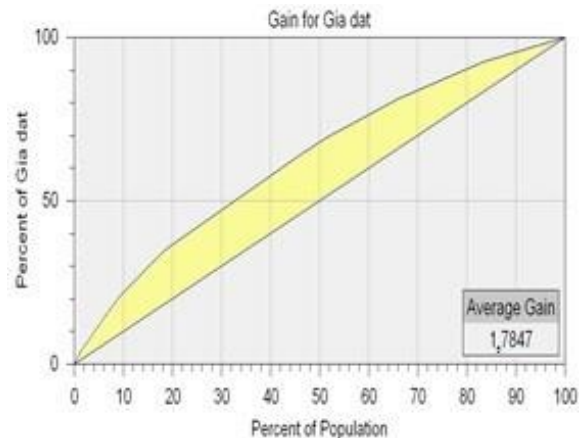
Bảng 3. Phân tích phương sai

STT	Chỉ tiêu	Giá trị
1	Phương sai dữ liệu (Variance in initial data sample)	306,56
2	Phương sai sau khi chạy mô hình (Residual (unexplained) variance after tree fitting)	43,82
3	Tỷ lệ phương sai được giải thích (Proportion of variance explained)	0,86

Phương sai được giải thích bởi cây được tạo ra là thước đo tốt nhất về mức độ cây phù hợp với dữ liệu [10]. Phương sai của tập dữ liệu trước khi xây dựng cây quyết định là 306,56. Phương sai sau khi mô hình được ứng dụng vào tập dữ liệu để dự báo biến mục tiêu là 43,82. Giá trị này cho thấy mức độ cải thiện phương sai đáng kể, cũng như tính thích hợp của mô hình cây quyết định được đưa ra. Mô hình càng hoàn hảo khi phương sai cây được ứng dụng tiến đến 0. Tỷ lệ biến thiên của biến kết quả được mô tả bởi cây quyết định là 86%. Cụ thể, các yếu tố mật độ kinh doanh, khu vực, cấp đường, ngành nghề

kinh doanh và độ rộng đường giải thích được 86% sự hình thành giá đất địa bàn nghiên cứu.

Độ lợi (Gain) được xác định là sự khác biệt giữa thông tin gốc (thông tin khi chưa phân lớp) và thông tin mới (thông tin sau khi đã phân lớp) [10]. Độ lợi trung bình của mô hình là 1,7847. Hình 4 cho thấy đường cong biểu diễn độ lợi khi có mô hình “nằm trên” đường chéo, biểu diễn khi không có mô hình. Do đó mô hình được thực hiện có ý nghĩa.



Hình 4. Độ lợi của mô hình.

3.3. Mức độ quan trọng các biến dự báo trong mô hình

Mức độ quan trọng các biến dự báo trong mô hình được tính bằng cách sử dụng thông tin về cách các biến được sử dụng làm bộ tách chính và cũng như bộ chia thay thế. Bằng cách xem xét phân chia thay thế, tầm quan trọng được tính toán bởi mô hình đưa ra thước đo chính xác hơn về giá trị thực tế và tiềm năng của một biến dự báo [10]. Thang đo tầm quan trọng biến dự báo được chia theo tỷ lệ phần trăm (quan trọng nhất là 100%).

Bảng 4. Mức độ quan trọng của các biến dự báo

STT	Biến dự báo (variable)	Mức độ quan trọng (importance)
1	Mật độ kinh doanh	100,00
2	Khu vực	83,48
3	Cấp đường	78,87
4	Ngành nghề kinh doanh	58,78
5	Độ rộng đường	10,03

Kết quả mô hình cho thấy tầm quan trọng của các yếu tố ảnh hưởng đến giá đất ở thành phố Vũng Tàu như sau: yếu tố Mật độ kinh doanh ảnh hưởng đến giá nhiều nhất (100%), sau đó là Khu vực (82,58%), Cấp đường (82,54%), Ngành nghề kinh doanh (59,68%) và thấp nhất là Độ rộng đường (10,03%).

Từ thực tế điều tra cho thấy, yếu tố Mật độ kinh doanh có ảnh hưởng cao đến giá đất, nơi nào có mật độ kinh doanh càng cao thì giá càng cao. Vũng Tàu là một thành phố du lịch, có mức sống cũng như phân bố và phát triển dân cư hầu như đồng đều, sự khác biệt rõ rệt về phân vùng ở đây là không đáng kể. Ngoài ra, đường dọc biển dài là lợi thế rất lớn cho việc phát triển các khu vực chuyên kinh doanh du lịch nghỉ dưỡng. Do đó, giữa các khu vực sự chênh lệch này không nhiều hoặc không rõ rệt để có thể làm rõ sự phân chia khu vực giá trong cây giá trị. Ngược lại, vì là thành phố du lịch nên mật độ kinh doanh lại thể hiện rõ sự chênh lệch về giá đất giữa các tuyến đường với nhau hơn. Vì vậy cây giá trị

chọn biến Mật độ kinh doanh là biến vị thế ảnh hưởng lớn nhất đến giá đất.

Thực tế cho thấy sự phân bố các ngành nghề kinh doanh tốt cũng như mật độ kinh doanh ở thành phố Vũng Tàu, hầu hết đều ở mức độ cao và rất cao khi ở các cấp đường lớn và quan trọng. Các cấp đường nhỏ hơn thường chỉ diễn ra hoạt động buôn bán nhỏ lẻ để phục vụ trong khu vực gần đó. Do đó, biến Cấp đường có tầm quan trọng cao hơn biến Ngành nghề kinh doanh.

Yếu tố Độ rộng đường theo kết quả phân tích hồi quy không có sự tác động đáng kể đến giá đất. Thực tế độ rộng đường tác động không nhất quán trên toàn bộ địa bàn nghiên cứu. Ở những khu vực trung tâm, những đường được hình thành lâu đời lại có giá cao hơn so với những đường mới có độ rộng lớn hơn. Điều này trái với khu vực ven đô, ngoại ô.

3.4. Nguyên tắc phân tích cây quyết định

Quá trình phân nhánh là quá trình chia bộ dữ liệu điều tra, được thực hiện trên cơ sở các mẫu có các thuộc tính giống nhau sẽ được chia vào cùng một nhóm. Khi nhóm mẫu có nhiều thuộc tính khác nhau thì thuộc tính có mức độ ảnh hưởng cao sẽ được lựa chọn để chia tách nhóm. Nhóm có giá trị trung bình cao sẽ tương ứng với yếu tố lựa chọn phân nhánh thuận lợi và ngược lại. Các nhóm lại tiếp tục được chia nhỏ hơn theo nguyên tắc này đến khi không còn các thuộc tính khác nhau (hoặc đến lớp phân nhánh cao nhất). Khi đó, giá trị của nút sẽ đại diện cho nhóm và mang các đặc điểm chung của vùng đất.

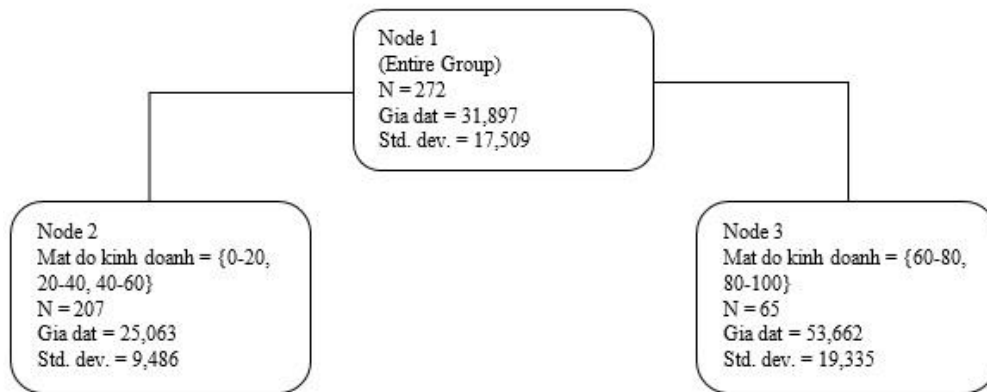
Từ nút (node) ban đầu, mức giá trung bình đất đai của thành phố Vũng Tàu được xác định là 31,897 (ba mươi một triệu, tám trăm chín mươi bảy ngàn đồng) và độ lệch chuẩn là 17,509. Điều này là hợp lý, vì một bộ dữ liệu lớn bao gồm toàn bộ thành phố thì sự biến thiên của tập dữ liệu là rất lớn. Từ mức giá trung bình chung của toàn thành phố, mô hình tiếp tục phân tích và chia tách đến lá cuối cùng.

Ở tại node chia đầu tiên, với mức độ ảnh hưởng tuyệt đối (100%), mật độ kinh doanh là yếu tố đầu tiên mô hình lựa chọn để phân tách cây: một nhóm có mật độ kinh doanh 0-20, 20-

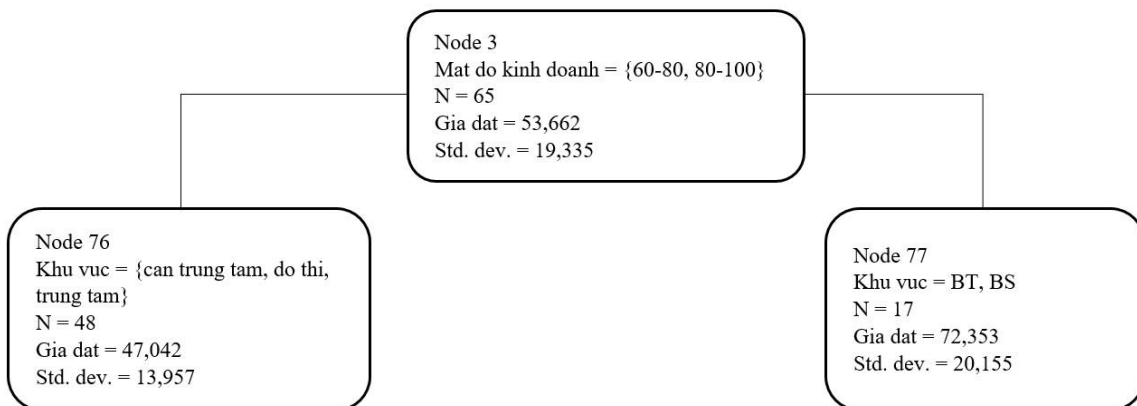
40, 40-60(%) và một nhóm có mật độ: 60-80, 80-100(%) (Hình 5). Quá trình phân cấp thể hiện sự khách quan và khoa học, các vị trí có khả năng kinh doanh rất thuận lợi được tách thành nhánh có giá trị trung bình cao, các vị trí còn lại được tách thành nhánh có giá trị trung bình thấp.

Tương tự, ở lần phân nhánh thứ 2, các nhánh tiếp tục được phân chia theo yếu tố có mức ảnh

hưởng cao nhất. Tại node 3, yếu tố được lựa chọn phân cấp là khu vực (Hình 6). Các khu vực Bãi Trước, Bãi Sau có giá trị trung bình cao hơn các khu vực còn lại. Đúng với thực tế, Vũng Tàu là thành phố biển và có 2 bãi có thể tắm biển. Các hoạt động du lịch tập trung xung quanh 2 bãi biển này và khu vực xung quanh 2 bãi này có giá đất cao nhất.



Hình 5. Chia nhánh tại node 1.

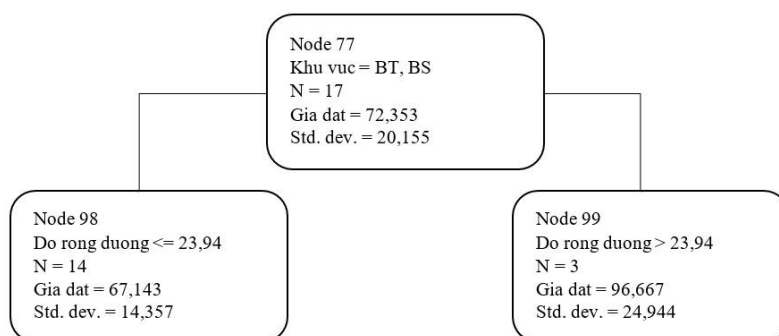


Hình 6. Chia nhánh tại node 3.

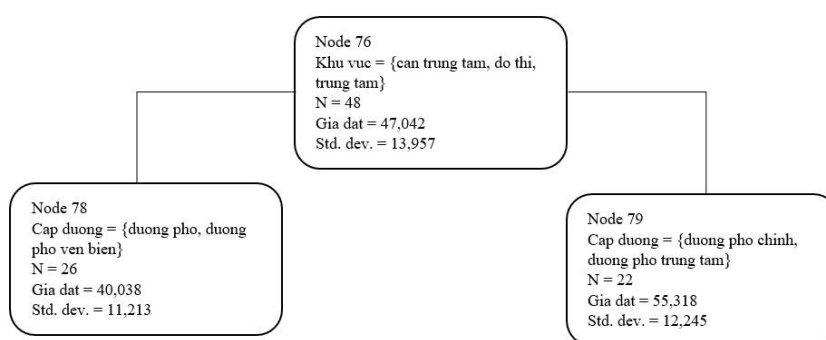
Mô hình cây quyết định cũng cho thấy trong cùng một lớp phân nhánh, yếu tố phân nhánh của các nút khác nhau có thể khác nhau. Trong khi lần phân nhánh thứ 3, độ rộng đường là yếu tố được lựa chọn để phân cấp tại node 77 còn tại node 76 là cấp đường. Điều này được giải thích vì yếu tố được lựa chọn để phân nhánh là yếu tố có mức độ ảnh hưởng cao nhất của nhóm đó nên

yếu tố được lựa chọn để phân nhánh không phải là duy nhất trong một lớp phân nhánh.

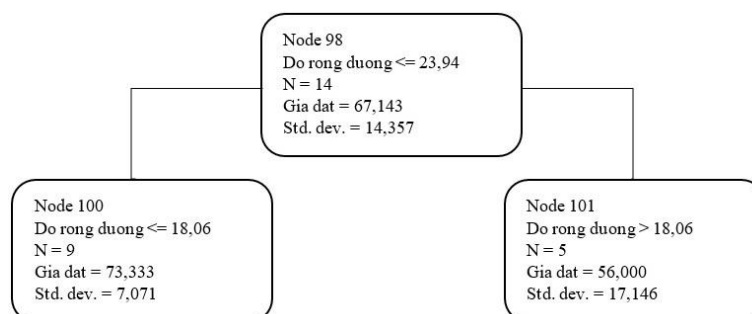
Tại node 77, độ rộng đường được chia thành 2 nhóm là lớn hơn và bé hơn 23,94 m với mức giá tương ứng là 96,67 triệu và 67,14 triệu (Hình 7). Tại node 76, yếu tố cấp đường được chia thành 2 nhóm tương ứng với 2 mức giá 40,04 và 55,32 triệu (Hình 8).



Hình 7. Chia nhánh tại node 77.



Hình 8. Chia nhánh tại node 76.



Hình 9. Chia nhánh tại node 98.

Tương tự như vậy, tại node 79 tiếp tục chia theo mật độ kinh doanh thành 2 nhóm 60-80% (node 90) và 80-100% (node 91). Tại node 91 tiếp tục chia theo yếu tố khu vực thành node 94 và 95. Như vậy nếu ta đi theo 1 “nhánh” 1-3-76-79-91-94 tương ứng với bộ tính chất mật độ kinh doanh 80-100 %, khu vực đô thị, cấp đường là đường phố chính hoặc đường phố trung tâm thì có giá đất trung bình là 40,00 triệu/m².

Trong mô hình cây vừa xây dựng có node 98 cần phân tích kỹ. Tại node này, yếu tố độ rộng đường được chia thành 2 nhóm, tuy nhiên tại nhóm có độ rộng đường bé hơn lại có mức giá cao hơn so với nhóm có độ rộng đường lớn hơn (Hình 9). Điều này được giải thích tại khu vực này, những tuyến đường được hình thành lâu đời, buôn bán sầm uất có giá cao hơn so với các tuyến đường mới.

Như vậy, nguyên tắc phân nhánh của cây quyết định phụ thuộc vào các yếu tố sau: i) Biên dự báo có mức ảnh hưởng cao nhất đến giá đất; ii) Quá trình phân nhánh diễn ra đến khi không

còn sự khác biệt của các biên dự báo trong nhánh đó; và iii) Nhánh có yếu tố thuận lợi hơn trong phân hành sẽ có giá trị trung bình cao hơn nhánh còn lại.

Bảng 5. Tổ hợp các yếu tố vùng và mức giá đất đai tương ứng

S T T	Số vùng	Tổ hợp các yếu tố đất đai					Mức giá trung bình (tr/m ²)	Độ lệch chuẩn	Tỷ lệ lệch
		Mật độ kinh doanh (%)	Ngành nghề kinh doanh	Cấp đường	Khu vực	Độ rộng đường (m)			
1	10	0-20			Đô thị, Ven đô thị		12,294	4,416	0,359
2	11	0-20			Cận trung tâm, Trung tâm		19,727	6,877	0,349
3	26	20-40	Tạp hóa				20,787	3,724	0,179
4	27	20-40	Quán ăn – Coffe, Shop – Văn phòng				26,250	6,891	0,263
5	80	60-80, 80-100		Đường phố, Đường phố ven biển	Trung tâm		31,428	2,382	0,076
6	50	40-60				≤23,44	31,928	3,977	0,125
7	51	40-60				>23,44	38,333	7,071	0,184
8	94	80-100		Đường phố chính, Đường phố trung tâm	Đô thị		40,000	0,000	0,000
9	9	0-20			Bãi Trước, Bãi Sau		48,000	0,000	0,000
10	90	60-80		Đường phố chính, Đường phố trung tâm	Trung tâm, Cận trung tâm, Đô thị		48,090	8,017	0,167
11	81	60-80, 80-100		Đường phố, Đường phố ven biển	Cận trung tâm		50,083	8,855	0,177
12	101	60-80, 80-100			Bãi Trước, Bãi Sau	≤23,94 và >18,06	56,000	17,146	0,306
14	100	60-80, 80-100			Bãi Trước, Bãi Sau	≤18,06	73,333	7,071	0,096
15	99	60-80, 80-100			Bãi Trước, Bãi Sau	>23,94	96,666	24,944	0,258

3.5. Kết quả phân nhóm cây quyết định

Giá trị của các nút lá là giá trị trung bình đại diện cho các vùng đất có các đặc điểm chung được thể hiện từ nút gốc đến nút lá. Cây quyết định dưới dạng tĩa (pruned tree) tối ưu tạo ra 15 vùng giá trị tương ứng với tổ hợp các yếu tố tạo ra giá đất (Bảng 5).

Dựa vào bảng có thể nhận thấy các vùng giá trị của mô hình cây quyết định có thể chia làm 3 nhóm chính: nhóm tổ hợp thuộc khu vực Bãi Trước, Bãi Sau với mật độ kinh doanh cao và ngành nghề kinh doanh nhà hàng, khách sạn với điều kiện kinh doanh thuận lợi có mức giá trung bình cao nhất. Tiếp đến là tổ hợp thuộc khu vực trung tâm, cận trung tâm với ngành nghề kinh doanh quán ăn, coffe. Thấp nhất là nhóm khu vực ven đô thị, với điều kiện kinh doanh không thuận lợi.

Nếu lấy mức độ lệch ở mức cho phép là 25% thì có 5 tổ hợp có mức độ lệch cao hơn, cụ thể ở các tổ hợp số 1, 2, 4, 12 và 15. Trong đó hai tổ hợp 4 và 15 có mức độ chênh lệch vượt chuẩn không đáng kể lần lượt là 26,3% và 25,8%. Do đó, nếu sử dụng cây dạng tĩa cần điều tra điều chỉnh và bổ sung thông tin các yếu tố vị thế đất đai của các tuyến và đoạn đường trong vùng này. Các tổ hợp còn lại có độ lệch trong tỉ lệ cho phép hoặc không đáng kể.

4. Kết luận

Nghiên cứu đã xây dựng mô hình định giá đất hàng loạt ứng dụng cây quyết định bằng cách xác định những tổ hợp các yếu tố ảnh hưởng đến giá đất và giá đất trung bình tương ứng tổ hợp đất. Trái với sự phức tạp của các mô hình hàm hồi quy và một số mô hình khác thì cây quyết định có thể tự động xử lý sự tương tác giữa các biến, hiển thị trực quan dễ hiểu. Mô hình cây quyết định bổ sung cho lý thuyết định giá đất một phương pháp mới. Ứng dụng trên địa bàn thành phố Vũng Tàu, mô hình cây quyết định chia làm 3 nhóm chính: nhóm tổ hợp thuộc khu vực Bãi Trước, Bãi Sau với ngành nghề kinh doanh nhà hàng, khách sạn và điều kiện kinh doanh thuận lợi có mức giá trung bình cao nhất (96,67 triệu

đồng/m²). Tiếp đến là tổ hợp thuộc khu vực trung tâm, cận trung tâm với ngành nghề kinh doanh quán ăn, coffe. Thấp nhất là nhóm khu vực ven đô thị, với điều kiện kinh doanh không thuận lợi (12,29 triệu đồng/m²).

Tài liệu tham khảo

- [1] N. Q. Hoa, N. H. Duong, Building Mass Land Valuation Model in Go Vap District, Ho Chi Minh City, Journal of Economic Development, Vol. 26, No. 3, 2015, pp. 82-103 (in Vietnamese).
- [2] T. H. Lien, Textbook of Land Valuation, Science and Technics Publishing House, Hanoi, 2014 (in Vietnamese).
- [3] P. S. Liem, T. V. Tuan, Building Land Price Zone in Rural Areas: A Case Study of Duong Quang Commune, Gia Lam District, Hanoi City, Vietnam Journal of Agricultural Sciences, Vol. 16, No. 6, 2019, pp. 601-612 (in Vietnamese).
- [4] G. Z. Fan, S. E. Ong, H. C. Koh, Determinants of House Price: A Decision Tree Approach, Urban Studies, Vol. 43, No. 12, 2006, pp. 2301-2315, <https://doi.org/10.1080/00420980600990928>.
- [5] C. Acciani, V. Fucilli, R. Sardaro, Data Mining in Real Estate Appraisal: A Model Tree and Multivariate Adaptive Regression Spline Approach, Aestimium, Vol. 58, 2011, pp. 27-45, <https://doi.org/10.13128/Aestimium-9560>.
- [6] M. Chen, P. D. Wang, A Roadmap to Determine the Important Factors of the House Value: A Case Study by Using Actual Price Registration Data of Taipei Housing Transactions, Independent Journal of Management & Production, Vol. 9, No. 1, 2018, pp. 245-261, <https://doi.org/10.14807/ijmp.v9i1.682>.
- [7] N. T. M. Linh, Application of Decision Tree Model in Mass Land Valuation in Vietnam, Journal of Finance, Vol. 8, 2011, pp. 43-46 (in Vietnamese).
- [8] Ministry of Natural Resources and Environment, Research on Building Land Valuation Support Systems Using Open Source WebGIS Technology, Reports of Scientific Research, 2016(in Vietnamese).
- [9] T. Güneş, Ü. Yıldız, Mass Valuation Techniques Used in Land Registry and Cadastre Modernization Project of Republic of Turkey, FIG Working Week: From the Wisdom of the Ages to the Challenges of the Modern World, 2015.
- [10] DTREG Predictive Modeling Software, Manual for DTREG, https://www.dtreg.com/uploaded/downloadfile/DownloadFile_5.pdf, 2020 (accessed on: March 3rd, 2020).