

DOI:10.22144/ctu.jvn.2018.043

ỨNG DỤNG KỸ THUẬT KHAI PHÁ DỮ LIỆU ĐÁNH GIÁ THÍCH NGHI ĐẤT ĐAI CÂY CAO SU TRÊN ĐỊA BÀN HUYỆN PHÚ GIÁO, TỈNH BÌNH DƯƠNG

Nguyễn Hữu Cường*

Khoa Quản lý đất đai, Trường Đại học Tài nguyên và Môi trường Thành phố Hồ Chí Minh

*Người chịu trách nhiệm về bài viết: Nguyễn Hữu Cường (email: nhcuong@hcmunre.edu.vn)

Thông tin chung:

Ngày nhận bài: 11/08/2017

Ngày nhận bài sửa: 22/11/2017

Ngày duyệt đăng: 26/04/2018

Title:

Application of data mining technique in land evaluation for rubber trees in Phu Giao district, Binh Duong province

Từ khóa:

Cây cao su, cây quyết định, đánh giá đất đai, khai phá dữ liệu

Keywords:

Data mining, decision tree, land evaluation, rubber tree

ABSTRACT

This study is aimed to evaluate the applicability of data mining technique by using decision tree in land evaluation. It can be used to determine the land characteristic factors affecting the agricultural land-use potential and quantify the relationship between land characteristic factors and plant productivity in order to improve land evaluation methods that support the foundation of land use planning. Regression decision tree model in this study includes two kinds of variables. The target variable is the productivity (t/ha) and the predictor variables consist of soil types, soil depth, slope, irrigation and texture. The analytical result of survey data shows several factor combinations according to plant average productivity. Based on productivity can evaluate the adaptation level for each correlative factor combination. This study is applied for rubber trees and conducted in Phu Giao district, Binh Duong province. The study shows that the interpretation level of the predictive variables is 96.49%. The area of highly suitable (S1) is 474.67 hectares, suitable (S2) is 53,597.70 hectares. This result is different from the Analytic Hierarchy Process (AHP) method.

TÓM TẮT

Nghiên cứu đánh giá khả năng ứng dụng kỹ thuật khai phá dữ liệu bằng cây quyết định trong đánh giá đất đai nhằm xác định các yếu tố đặc điểm đất đai ảnh hưởng đến tiềm năng sử dụng đất nông nghiệp, lượng hóa mối quan hệ giữa các yếu tố đặc điểm đất đai với năng suất cây trồng nhằm hoàn thiện phương pháp đánh giá đất đai có khả năng cung cấp căn cứ lập quy hoạch sử dụng đất đai. Mô hình hồi quy cây quyết định được thực hiện với biến mục tiêu (target) là năng suất (tấn/ha), các biến dự báo (predictor) là: loại đất, độ dày tầng đất, độ dốc, khả năng tưới và thành phần cơ giới. Từ kết quả phân tích dữ liệu điều tra theo mô hình cây quyết định ta rút ra những tổ hợp các yếu tố theo năng suất trung bình của cây trồng. Dựa vào năng suất để đánh giá mức độ thích nghi cho từng tổ hợp yếu tố ảnh hưởng. Nghiên cứu áp dụng trên địa bàn huyện Phú Giáo, tỉnh Bình Dương cho cây cao su. Kết quả nghiên cứu cho thấy mức độ giải thích các biến dự báo là 96,49%. Cấp thích nghi cao chiếm 474,67 ha, cấp thích nghi trung bình chiếm 53.597,70 ha. Kết quả có sự sai lệch so với phương pháp phân tích thứ bậc (AHP).

Trích dẫn: Nguyễn Hữu Cường, 2018. Ứng dụng kỹ thuật khai phá dữ liệu đánh giá thích nghi đất đai cây cao su trên địa bàn huyện Phú Giáo, tỉnh Bình Dương. Tạp chí Khoa học Trường Đại học Cần Thơ. 54(3B): 84-93.

1 GIỚI THIỆU

Đánh giá thích nghi đất đai nhằm mục đích cung cấp những thông tin về sự thuận lợi và khó khăn cho việc sử dụng đất đai, làm căn cứ cho việc đưa ra những quyết định về việc sử dụng và quản lý đất đai một cách hợp lý. Hiện nay, phương pháp phổ biến thực hiện đánh giá thích nghi đất đai về mặt tự nhiên là kết hợp theo điều kiện hạn chế - được áp dụng với giả thiết là các yếu tố chất lượng đất có tầm quan trọng như nhau và không có sự tương tác với nhau. Hạn chế của phương pháp này là không tính đến sự tương tác bù trừ qua lại của các yếu tố chất lượng đất đai. Trong thực tế, sự thiếu hụt về lượng của yếu tố đặc điểm này có thể được thay thế bằng lượng của các yếu tố khác trong tổ hợp các đặc điểm chất lượng của đất đai (Nguyễn Ánh Nga, 2012). Chính vì vậy, một số nghiên cứu khác đề xuất sử dụng phương pháp toán học để tính toán chỉ tiêu tổng hợp thích nghi đất đai, cụ thể là ứng dụng phương pháp phân tích thứ bậc Analytic Hierachy Process (AHP) để tính toán chỉ số thích nghi đất đai, trong đó có tính đến mối quan hệ giữa các yếu tố đặc điểm đơn tính của đất đai. Phương pháp này mang tính định tính hoặc bán định lượng, có sự tham gia ý kiến của chuyên gia.

Mức độ chính xác của việc đánh giá phân hạng thích nghi đất đai không chỉ phụ thuộc vào việc xác định số lượng thích nghi và loại yếu tố đặc điểm đất đai, mà còn phụ thuộc vào việc định lượng mối quan hệ giữa các yếu tố đặc điểm này với năng suất cây trồng. Việc định lượng này phải không mang tính chủ quan, áp đặt của con người mà dựa vào những giá trị điều tra thực tế được lượng hóa thành.

Nhiều phương pháp khai phá dữ liệu (data mining) đã được áp dụng rộng rãi trong đánh giá đất đai (Tian *et al.*, 2009) nhằm khắc phục những yếu tố mang tính chủ quan. Cây quyết định (decision tree) là một trong những thuật toán phân loại phổ biến nhất hiện nay trong khai phá dữ liệu (Kumar *et al.*, 2013). Đã có nhiều nghiên cứu áp dụng phương pháp này trong đánh giá đất đai. Lanen *et al.* (1992) trong nghiên cứu đánh giá đất đai hỗn hợp định tính và định lượng đã tiến hành với cây khoai tây tại Hà Lan. Kết quả cho thấy khoảng 65% diện tích đất có khả năng phù hợp. Bouma *et al.* (1993) đã nghiên cứu đánh giá đất đai cho cây ngô ở cấp độ nông trại tại New York. Tian *et al.* (2009) so sánh mức độ chính xác trong đánh giá đất đai ứng dụng khai phá dữ liệu với 3 kỹ thuật: cây quyết định, mạng nơ-ron và hồi quy. Kết quả cho thấy mô hình cây quyết định là mô hình tốt nhất. Yang *et al.* (2010) trong nghiên cứu của mình về cây quyết định trong đánh giá đất đai tại tỉnh Quảng Đông, Trung Quốc đã chứng minh rằng phương pháp này thuận tiện để trích xuất

các quy tắc phân loại với tỷ lệ chính xác 86,67%. Cây quyết định cũng được sử dụng để thực hiện từ dữ liệu khảo sát đất đai vùng Maharashtra, Ấn Độ bởi Kumar *et al.* (2013). Việc kiểm tra chéo 10 lần cung cấp độ chính xác 100%. Tại Việt Nam, Nguyễn Ánh Nga (2012) nghiên cứu khả năng ứng dụng khai phá dữ liệu trong đánh giá đất đai với kỹ thuật cây quyết định đối với cây điều và cây xoài trên địa bàn huyện Định Quán, Đồng Nai.

Mục tiêu của nghiên cứu là áp dụng phương pháp khai phá dữ liệu trong đánh giá thích nghi đất đai tự nhiên và so sánh kết quả với phương pháp khác đối với cây cao su trên địa bàn huyện Phú Giáo, tỉnh Bình Dương.

2 PHƯƠNG PHÁP NGHIÊN CỨU VÀ QUY TRÌNH THỰC HIỆN

2.1 Phương pháp nghiên cứu

2.1.1 Khai phá dữ liệu và mô hình cây quyết định

Khai phá dữ liệu là quá trình trích xuất các thông tin có giá trị tiềm ẩn bên trong lượng lớn dữ liệu được lưu trữ trong các kho dữ liệu (Han and Kamper, 2006). Để đạt được những tri thức từ cơ sở dữ liệu hiện có, nhiều kỹ thuật khai phá dữ liệu khác nhau ra đời như: phân lớp dữ liệu, phân cụm dữ liệu, khai phá luật kết hợp, hồi quy, giải thuật di truyền, mạng nơ-ron, cây quyết định. Trong đó, kỹ thuật cây quyết định (decision tree) là một công cụ mạnh và hiệu quả trong việc phân lớp và dự báo.

Cây quyết định là cấu trúc biểu diễn dưới dạng cây. Trong đó, mỗi nút trong (internal node) biểu diễn một thuộc tính, nhánh (branch) biểu diễn giá trị có thể có của thuộc tính, mỗi lá (leaf node) biểu diễn các lớp quyết định và đỉnh trên cùng của cây gọi là gốc (root). Cây quyết định có thể được dùng để phân lớp bằng cách xuất phát từ gốc của cây và di chuyển theo các nhánh cho đến khi gặp nút lá (Nguyễn Anh Nga, 2012).

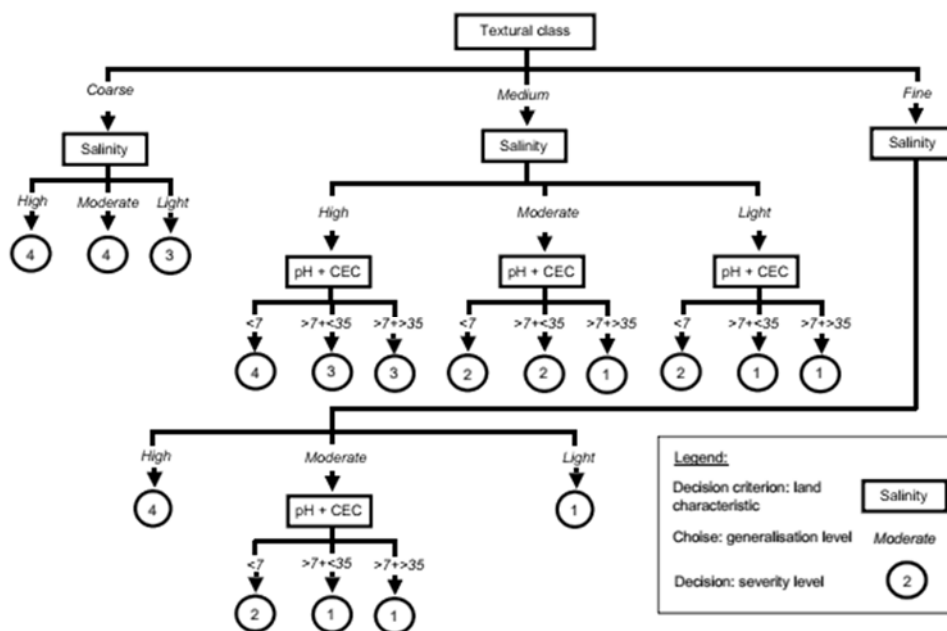
Cây quyết định được sử dụng để chia liên tiếp một tập dữ liệu lớn thành các tập con nhỏ bằng cách áp dụng một chuỗi các thuật toán. Với mỗi phép chia liên tiếp, các tập con thu được trong tập kết quả sẽ ngày càng giống nhau.

Đối với cây quyết định, tại mỗi nút, một thuộc tính sẽ được chọn ra để phân tách tập mẫu thành những lớp khác nhau nhiều nhất có thể. Các thuộc tính tham gia vào quá trình phân lớp thông thường có giá trị kiểu liên tục hay còn gọi là kiểu số và kiểu rời rạc hay còn gọi là kiểu phân loại.

Ứng dụng cây quyết định trong đánh giá thích nghi đất đai bằng việc xác định những *tổ hợp các yếu tố đặc điểm đất đai* (độ dày tầng đất, độ dốc địa

hình, thành phần cơ giới, khả năng tưới,...) và mức sản lượng cây trồng tương ứng. Quy trình đánh giá

đất đai theo các tiêu chí cây quyết định là dễ tiếp cận và minh bạch (Bouma *et al.*, 1993).



Hình 1: Mô hình cây quyết định trong phân lớp đất đai

(Nguồn: Rosa and Diepen, 2002)

2.1.2 Phương pháp điều tra, khảo sát

Nghiên cứu thực hiện điều tra khảo sát nông hộ phân bố trên các đơn vị đất đai khác nhau. Thông tin điều tra gồm các đặc điểm tự nhiên đất đai (loại đất, độ dày tầng đất, độ dốc địa hình, thành phần cơ giới, khả năng tưới) và năng suất mùa tươi cây cao su. Số phiếu được sử dụng để chạy mô hình là 98 phiếu.

2.1.3 Phương pháp ứng dụng GIS

Nghiên cứu sử dụng phần mềm GIS (MapInfo) xây dựng các bản đồ đơn tính và bản đồ thích nghi đất đai.

2.1.4 Phương pháp ứng dụng phần mềm phân tích thống kê

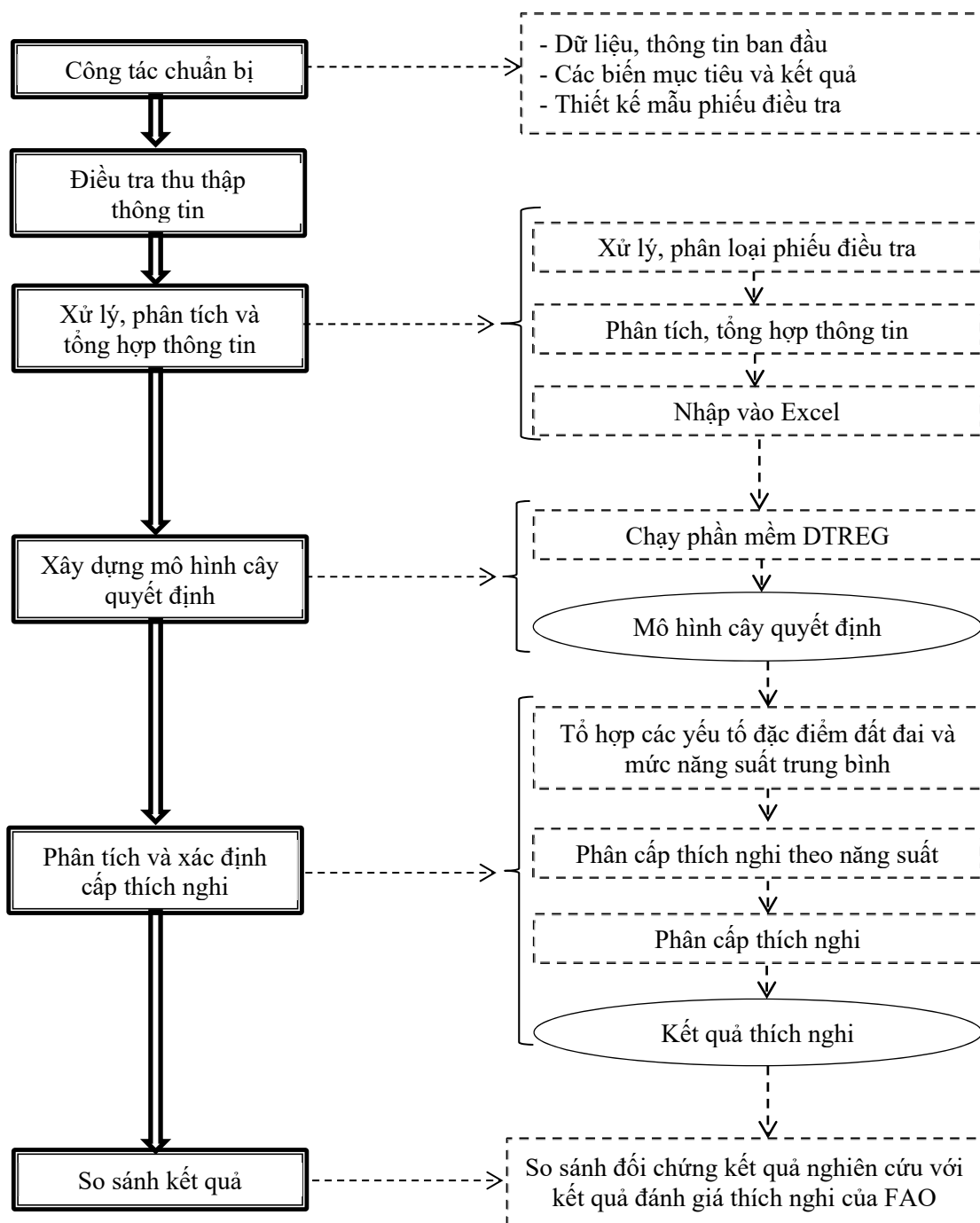
Nghiên cứu sử dụng phần mềm DTREG - phần mềm phân tích thống kê mạnh mẽ, có khả năng xây dựng cây quyết định phân lớp, hồi quy và máy vector hỗ trợ (SVM) để mô tả mối quan hệ dữ liệu.

DTREG chấp nhận tập hợp dữ liệu có nhiều dòng với một cột cho mỗi biến. Một trong các biến

là biến mục tiêu, giá trị của nó được mô hình hóa và được dự đoán là một hàm của biến dự báo. DTREG phân tích giá trị và cho ra một mô hình chỉ cách tốt nhất để dự đoán giá trị của biến kết quả dựa trên giá trị của biến dự báo. Đặc biệt DTREG có khả năng nhận biết các giá trị định tính cho các biến (ví dụ: “Có tưới”, “không tưới”, “tưới bổ sung”,...) và xác định tầm quan trọng (mức độ ảnh hưởng) mỗi biến dự báo đến biến kết quả. Ngoài việc xây dựng mô hình dự báo, DTREG còn đo chất lượng mô hình.

2.2 Quy trình thực hiện đánh giá đất đai ứng dụng mô hình cây quyết định

Quy trình nghiên cứu bắt đầu từ việc xác định các biến trong mô hình, gồm biến dự báo và biến kết quả. Biến dự báo được đề xuất dựa trên đặc điểm tự nhiên đất đai của địa phương và yêu cầu sử dụng đất của cây trồng. Biến kết quả là năng suất thực tế cây trồng trên địa bàn nghiên cứu. Mô hình nghiên cứu được xây dựng dựa trên các thông tin được thu thập thông qua điều tra nông hộ. Nhóm nghiên cứu đề xuất quy trình thực hiện theo Hình 2.



Hình 2: Sơ đồ các bước tiến hành đánh giá đất đai ứng dụng cây quyết định

3 KẾT QUẢ VÀ THẢO LUẬN

3.1 Các đặc điểm đất đai trên địa bàn huyện Phú Giáo

Dựa vào dữ liệu thu thập được ở địa bàn nghiên cứu và đặc tính tự nhiên cây cao su, các đơn vị đất đai được xây dựng trên cơ sở mối quan hệ giữa tính chất thổ nhưỡng, độ dày tầng đất hữu hiệu, độ dốc, khả năng tưới và thành phần cơ giới. Đây cũng là

những yếu tố được lựa chọn là biến dự báo trong mô hình cây quyết định.

Loại hình thổ nhưỡng: Địa bàn huyện Phú Giáo bao gồm đất phù sa (P, Pf, Pg), đất xám gley và đất dốc tụ (Xg), đất xám trên phù sa cổ và đất xám nâu vàng (X, Fp).

Độ dày tầng đất: Được chia 5 cấp: > 100 cm, 70 – 100 cm, 50 – 70 cm, 30 – 50 cm và < 30 cm.

Độ dốc địa hình: Được phân chia như sau: $0^0 - 3^0$, $3^0 - 8^0$, $8^0 - 15^0$.

Thành phần cơ giới: Phổ biến là thịt trung bình và cát pha, thịt nhẹ.

Khả năng tưới: Được chia làm tưới mặt và tưới ngầm.

Chồng xếp các bản đồ đơn tính trên địa bàn huyện Phú Giáo tạo nên 15 đơn vị đất đai.

3.2 Mô hình cây quyết định phân nhóm các đặc điểm đất đai và năng suất tương ứng cây cao su trên địa bàn huyện Phú Giáo

Mô hình hồi quy cây quyết định đánh giá thích nghi đất đai cây cao su trên địa bàn huyện Phú Giáo được thực hiện với các biến:

– Biến mục tiêu (target): Năng suất (tấn/ha) (trong nghiên cứu thu thập năng suất mủ cao su tươi).

– Các biến dự báo (predictor): Loại đất (loại đất), độ dày tầng đất (tầng dày (cm)), độ dốc (do dốc (do)), khả năng tưới (kha nang tưới) và thành phần cơ giới (tpcg).

Mô hình cây quyết định sau khi chạy được xây dựng gồm 7 tầng với số nhóm phân chia là 10, tổng số nút (node) là 25. Kết quả “phân tích phương sai” và “tầm quan trọng các biến” được sử dụng để đánh giá mô hình.

Phương sai của tập dữ liệu trước khi xây dựng cây quyết định là 0,81. Phương sai sau khi cây được ứng dụng vào tập dữ liệu để dự báo biến mục tiêu là 0,03. Kết quả cho thấy một mức độ cải thiện phương

sai đáng kể, cũng cho thấy tính thích hợp của mô hình cây quyết định được đưa ra. Khả năng được giải thích của biến mục tiêu bởi cây quyết định là 96,49%, còn lại 3,51% không thể giải thích được do chịu ảnh hưởng của các yếu tố khác. Cụ thể, các yếu tố loại đất, độ dày tầng đất, độ dốc, khả năng tưới, thành phần cơ giới giải thích được 96,49% sự hình thành năng suất cây trồng. Như vậy có thể nói, mô hình cây quyết định được xây dựng có mức độ thích hợp và khả năng dự báo là khá cao.

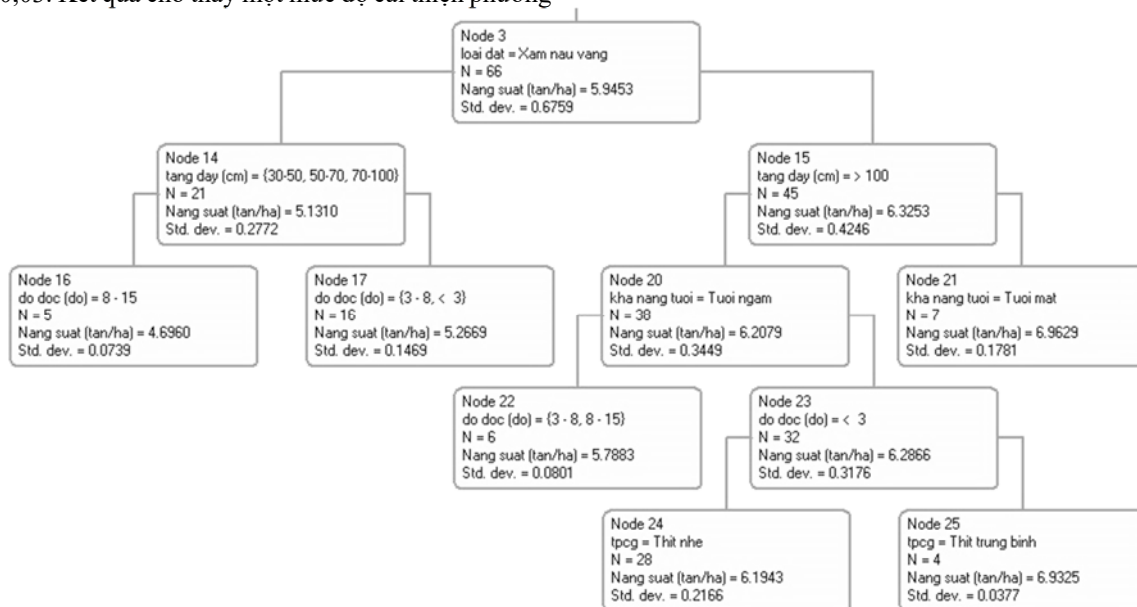
Bảng 1: Kết quả phân tích phương sai của mô hình

STT	Thông số	Kết quả
1	Phương sai mẫu dữ liệu đầu	0,81
2	Phương sai sau khi tạo cây	0,03
3	Tỷ lệ phương sai được giải thích	0,9649 (96,49%)

Bảng 2: Kết quả phân tích tầm quan trọng các biến theo mô hình

STT	Biến trong mô hình	Tầm quan trọng
1	Loại đất	100,00
2	Tầng dày	54,83
3	Khả năng tưới	7,72
4	Độ dốc	7,53
5	Thành phần cơ giới	4,37

Kết quả mô hình còn cho thấy tầm quan trọng (mức độ ảnh hưởng) của mỗi biến dự báo (loại đất, độ dày tầng đất, độ dốc địa hình, thành phần cơ giới, khả năng tưới) đến biến kết quả (năng suất cây cao su) là khác nhau.



Hình 3: Kết quả một “nhánh” mô hình cây quyết định

Năng suất cây cao su bị ảnh hưởng mạnh mẽ nhất bởi yếu tố loại đất. Tiếp theo, độ dày tầng đất cũng là chỉ tiêu quan trọng trong việc đánh giá (yếu tố quyết định đến khả năng giữ nước và cung cấp chất dinh dưỡng cho cây cao su), mức độ ảnh hưởng của tầng dày đến năng suất bằng 54,83% so với loại đất ảnh hưởng đến năng suất. Tương tự, mức độ ảnh hưởng theo thứ tự tiếp theo là khả năng tưới, độ dốc và thành phần cơ giới với các giá trị lần lượt là 7,72%, 7,53% và 4,37% so với loại đất.

Mô hình cây quyết định đã được tạo ra với nút phân chia đầu tiên là biến Loại đất.

Bảng 3: Phân cấp thích nghi của FAO theo năng suất cây trồng

STT	Cấp thích nghi	Hướng dẫn phân cấp của FAO	Phân cấp trong nghiên cứu (tấn/ha)
1	S1 (Thích nghi cao)	> 80%	> 6,4
2	S2 (Thích nghi trung bình)	40% – 80%	3,2 – 6,4
3	S3 (Thích nghi kém)	20 – 40%	1,6 – 3,2
4	N (Không thích nghi)	< 20%	< 1,6

Bảng 4: Kết quả tổ hợp các yếu tố ảnh hưởng đến năng suất cây cao su theo mô hình cây quyết định

STT	Tổ hợp các yếu tố					Năng suất trung bình (tấn/ha)	Mức độ thích nghi
	Loại đất	Tầng dày (cm)	Độ dốc (độ)	Khả năng tưới	TPCG		
1	Phù sa, Xám gley	30-50, 50-70	< 3	Tưới mặt	Thịt nhẹ	3,87 ($\pm 0,09$)	S2
2	Phù sa	70-100, > 100	< 3, 3 - 8	Tưới mặt, Tưới ngầm	Thịt trung bình, Thịt nhẹ	4,29 ($\pm 0,18$)	S2
3	Xám gley	70-100, > 100	8 - 15	Tưới mặt	Thịt trung bình, Thịt nhẹ	4,21 ($\pm 0,34$)	S2
4	Xám gley	70-100, > 100	< 3, 3 - 8	Tưới mặt, Tưới ngầm	Thịt trung bình, Thịt nhẹ	4,51 ($\pm 0,06$)	S2
5	Xám nâu vàng	30-50, 50-70, 70-100	8 - 15	Tưới ngầm	Thịt trung bình, Thịt nhẹ	5,01 ($\pm 0,07$)	S2
6	Xám nâu vàng	30-50, 50-70, 70-100	< 3, 3 - 8	Tưới ngầm	Thịt nhẹ	4,70 ($\pm 0,15$)	S2
7	Xám nâu vàng	> 100	3 - 8, 8 - 15	Tưới ngầm	Thịt trung bình, Thịt nhẹ	5,79 ($\pm 0,08$)	S2
8	Xám nâu vàng	> 100	< 3	Tưới ngầm	Thịt nhẹ	6,19 ($\pm 0,22$)	S2
9	Xám nâu vàng	> 100	< 3	Tưới ngầm	Thịt trung bình	6,93 ($\pm 0,04$)	S1
10	Xám nâu vàng	> 100	< 3	Tưới mặt	Thịt trung bình, Thịt nhẹ	6,96 ($\pm 0,18$)	S1

Kết quả Bảng 4 cho thấy mức độ thích nghi cây cao su theo các tổ hợp tính chất đất đai chịu sự ảnh hưởng lớn nhất của loại đất và độ dày tầng đất. Có những nhận xét như sau:

– Các đơn vị đất đai có loại đất giống nhau là xám nâu vàng có năng suất dao động từ 4,79 đến 6,96 tấn/ha. Cùng một loại đất và tầng dày thì độ dốc là yếu tố tiếp theo quyết định cấp thích nghi và sau

Dựa trên kết quả mô hình cây quyết định, tiến hành đi theo từng phân nhánh mô hình để xác định được tổ hợp các yếu tố đặc điểm đất đai và mức năng suất trung bình của cây cao su tương ứng với tổ hợp đấy. Cấp thích nghi được phân chia theo gợi ý của FAO dựa trên tỷ lệ năng suất thực tế với năng suất tối hảo cây trồng (với năng suất tối hảo thu thập được trong nghiên cứu 8 tấn/ha).

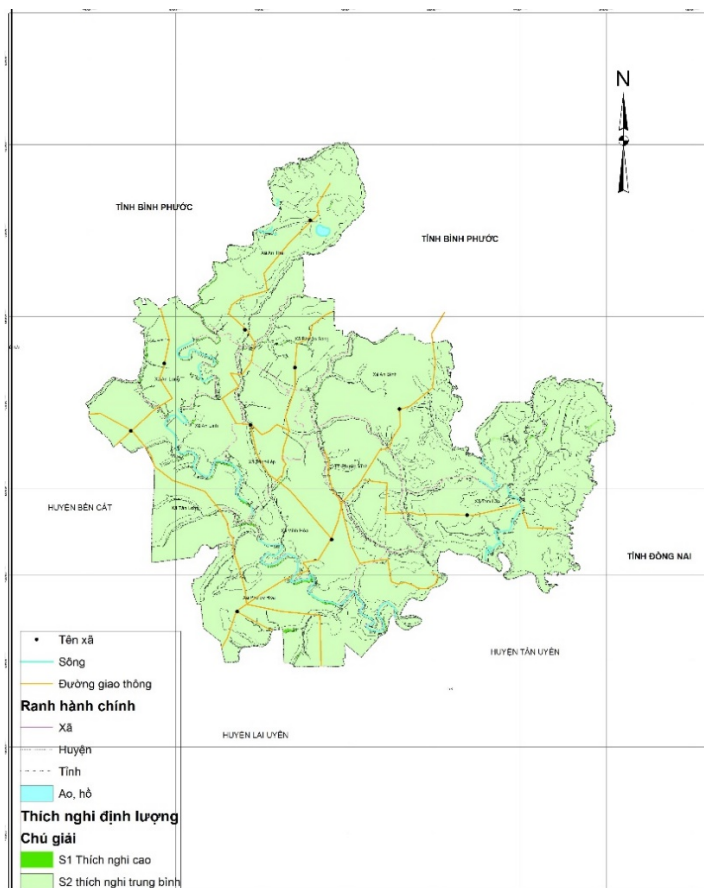
Từ đấy xác định mức độ thích nghi cho từng tổ hợp dựa vào năng suất trung bình tương ứng của tổ hợp. Kết quả được thể hiện tại Bảng 4.

đó đến khả năng tưới. Thành phần cơ giới hầu như không ảnh hưởng lớn tới năng suất cây cao su.

– Đối với những đơn vị đất đai có loại đất là xám gley có năng suất dao động từ 3,87 đến 4,51 tấn/ha, các đơn vị này có chung tầng dày, thì tiếp tục độ dốc là một yếu tố tạo ra sự khác biệt về năng suất. Tiếp đến là khả năng tưới và thành phần cơ giới, 2 yếu tố này không ảnh hưởng nhiều.

– Những đơn vị đất đai có loại đất phù sa có năng suất từ 3,87 đến 4,29 tấn/ha, độ dày tầng đất là yếu tố tiếp theo ảnh hưởng làm thay đổi năng suất cây trồng. Các yếu tố còn lại có mức ảnh hưởng rất ít.

Dựa vào kết quả phân cấp thích nghi cây cao su trên địa bàn huyện Phú Giáo ứng dụng cây quyết định ở Bảng 4 cùng với bản đồ đơn vị đất đai ta xây dựng được bản đồ mức độ thích nghi đối với cây cao su.



Hình 4: Bản đồ thích nghi cây cao su theo mô hình cây quyết định

Diện tích các cấp thích nghi (Bảng 5) được thống kê từ bản đồ thích nghi. Kết quả cho thấy diện tích đất ở huyện Phú Giáo có khả năng thích nghi trung bình và thích nghi cao đối với cây cao su. Thích nghi trung bình chiếm phần lớn với 99,12% diện tích, còn lại là thích nghi cao với tỷ lệ là 0,88%.

Bảng 5: Thống kê diện tích thích nghi theo mô hình cây quyết định

Phân cấp thích nghi	Diện tích (ha)	Tỷ lệ (%)
S1	474,67	0,88
S2	53.597,70	99,12
Tổng diện tích	54.072,37	100,00

3.3 So sánh kết quả với đánh giá đất đai sử dụng phương pháp AHP

Để có cơ sở kết luận, nghiên cứu đồng thời thực hiện đánh giá thích nghi cây cao su trên cùng địa bàn

sử dụng phương pháp phân tích thứ bậc Analytic Hierachy Process (AHP).

3.3.1 Đánh giá thích nghi cây cao su áp dụng phương pháp AHP

Xác định trọng số các yếu tố đặc điểm đất đai

Trọng số của các yếu tố chính là mức độ ảnh hưởng của yếu tố đó đến loại hình sử dụng đất. Trên cơ sở tham khảo ý kiến của 5 chuyên gia, nghiên cứu tiến hành tính toán trọng số cho từng yếu tố ảnh hưởng theo phương pháp phân tích thứ bậc 9 cấp độ. Thực hiện so sánh từng cặp các yếu tố với sự tham gia của các chuyên gia.

Tiếp theo, xác định ma trận so sánh tổng hợp các chuyên gia theo công thức $A_{ij} = (\prod_{k=1}^5 a_{ijk})^{1/5}$, trên cơ sở đó, tính trọng số các yếu tố theo phương pháp chuẩn hóa ma trận, kết quả được thể hiện ở Bảng 7.

Bảng 6: Giá trị so sánh cặp các yếu tố của các chuyên gia

So sánh		Kết quả đánh giá của chuyên gia thứ					A _{ij}
i	j	1	2	3	4	5	
Tầng dày	Độ dốc	3	1	1	3/2	1	1,35
Tầng dày	TPCG	5	3	7/2	3	3	3,43
Tầng dày	Loại đất	1/3	1/3	1/3	1/3	1/4	0,31
Tầng dày	Khả năng tưới	5	3	7/2	3	3	3,43
Độ dốc	TPCG	3	3	7/2	3	3	3,09
Độ dốc	Loại đất	1/5	1/5	1/5	1/3	1/5	0,22
Độ dốc	Khả năng tưới	3	1	1	3/2	1	1,35
TPCG	Loại đất	1/7	1/7	1/7	1/5	1/7	0,15
TPCG	Khả năng tưới	1/3	1/3	1/3	1/3	1/4	0,31
Loại đất	Khả năng tưới	5	3	7/2	3	3	3,42

Bảng 7: Ma trận so sánh tổng hợp các yếu tố và trọng số các yếu tố tổng hợp

Chỉ tiêu	Tầng dày	Độ dốc	TPCG	Loại đất	Khả năng tưới	Trọng số
Tầng dày	1	1,35	3,43	0,31	3,43	0,21
Độ dốc	0,74	1	3,09	0,22	1,35	0,14
TPCG	0,29	0,32	1	0,15	0,31	0,05
Loại đất	3,23	4,55	6,67	1	3,42	0,48
Khả năng tưới	0,29	0,74	3,23	0,26	1	0,12

Từ đó, tính toán các thông số theo AHP, kết quả được thể hiện ở Bảng 8.

Bảng 8: Các thông số theo AHP

Thông số	Kết quả
Lamdamax (λ_{\max})	5,17
Chỉ số nhất quán (CI)	0,04
Chỉ số ngẫu nhiên (RI)	1,12
Tỷ số nhất quán (CR)	0,04

Như vậy, tỷ số nhất quán CR đạt yêu cầu ($< 0,1$),

Bảng 9: Bảng giá trị tiêu chuẩn X_i đối với cây cao su

Chỉ tiêu (X_i)	Mã hóa			
	9	7	5	1
Độ dốc ($^\circ$)	$< 8^\circ$	$8^\circ - 15^\circ$	$15^\circ - 20^\circ$	$> 20^\circ$
Khả năng tưới	Tưới mặt	Tưới ngầm	Tưới ít	Không tưới
Tầng dày(cm)	$> 100\text{cm}$	70-100cm	50-70cm	$< 50\text{cm}$
Loại đất	Ft, Fk, Fu, Fv, Fn, Fd	Fe, Fj, Fs, Fp, X, R	Fa, Fq, Xa, Glu, CM	Các đất khác
Thành phần cơ giới	Thịt nặng	Thịt trung bình	Thịt nhẹ	

Bảng 10: Phân cấp chỉ số thích nghi

Giá trị S_i	Hạng thích nghi
8-9	S1
6-8	S2
4-6	S3
< 4	N

Tiến hành xác định chỉ số thích nghi S_i cho từng đơn vị đất đai bằng công thức $S_i = \sum_i^n W_i * X_i$. Chỉ

nên bộ trọng số trên được chấp nhận và đưa vào tính toán chỉ số thích nghi kết hợp xây dựng bản đồ thích nghi cho cây cao su trên địa bàn huyện Phú Giáo.

Tích hợp chỉ số thích nghi và đánh giá thích nghi

Ứng dụng thang phân loại tầm quan trọng của Saaty (1997, 1980, 1994), tham khảo ý kiến chuyên gia và kết hợp với thực tiễn của huyện Phú Giáo để thiết lập bảng phân cấp chỉ số thích nghi (X_i) của các yếu tố ảnh hưởng đến khả năng thích nghi của cây cao su được thể hiện ở Bảng 9.

số thích nghi dao động trong khoảng từ 1 – 9 được phân cấp theo cấu trúc phân hạng thích nghi đất đai của FAO được thể hiện ở Bảng 10; từ đó xác định hạng thích nghi cho từng đơn vị đất đai.

3.3.2 So sánh kết quả đánh giá đất đai bằng hai phương pháp

Tiến hành so sánh kết quả đánh giá thích nghi cho cây cao su bằng phương pháp cây quyết định và phân tích thứ bậc.

Bảng 11: Kết quả đánh giá đất đai bằng phương pháp AHP và cây quyết định

STT	Tổ hợp các yếu tố					Phương pháp AHP		Phương pháp cây quyết định	
	Độ dốc (độ)	Khả năng tưới	Tầng dày (cm)	Loại đất	TPCG	Chỉ số thích nghi S _i	Phân cấp thích nghi	Năng suất trung bình	Phân cấp thích nghi
1	< 3	Tưới ngầm	30-50	Xám nâu vàng	Thịt nhẹ	5,92	S3	4,70	S2
2	< 3	Tưới mặt	> 100	Phù sa	Thịt nhẹ	4,96	S3	4,29	S2
3	< 3	Tưới mặt	> 100	Xám gley	Thịt trung bình	5,06	S3	4,51	S2
4	< 3	Tưới mặt	> 100	Xám gley	Thịt nhẹ	4,96	S3	4,51	S2
5	8 – 15	Tưới ngầm	> 100	Xám nâu vàng	Thịt trung bình	7,42	S2	5,79	S2
6	< 3	Tưới mặt	> 100	Phù sa	Thịt trung bình	5,06	S3	4,29	S2
7	8 – 15	Tưới ngầm	50-70	Xám nâu vàng	Thịt nhẹ	6,48	S2	5,01	S2
8	< 3	Tưới ngầm	50-70	Xám nâu vàng	Thịt nhẹ	6,76	S2	4,70	S2
9	< 3	Tưới ngầm	> 100	Xám nâu vàng	Thịt trung bình	7,70	S2	6,93	S1
10	< 3	Tưới ngầm	> 100	Xám nâu vàng	Thịt nhẹ	7,60	S2	6,19	S2
11	< 3	Tưới ngầm	70-100	Xám nâu vàng	Thịt nhẹ	7,18	S2	4,70	S2
12	3 – 8	Tưới ngầm	> 100	Xám nâu vàng	Thịt trung bình	7,70	S2	5,79	S2
13	3 – 8	Tưới ngầm	> 100	Xám nâu vàng	Thịt nhẹ	7,60	S2	5,79	S2
14	3 – 8	Tưới ngầm	70-100	Xám nâu vàng	Thịt nhẹ	7,18	S2	4,70	S2
15	3 – 8	Tưới ngầm	50-70	Xám nâu vàng	Thịt nhẹ	6,76	S2	4,70	S2

Từ bảng tổng hợp trên cho thấy kết quả đánh giá thích nghi đất đai bằng 2 phương pháp có sự sai khác tại một số đơn vị đất đai, cụ thể như sau:

– Tại đơn vị đất đai thứ nhất, có tổ hợp loại đất là xám trên phù sa cổ và nâu vàng trên phù sa cổ, tầng dày là 30-50 cm, độ dốc < 3°, có tưới ngầm và thành phần cơ giới là thịt nhẹ, đánh giá thích nghi theo AHP là S3 còn theo phương pháp theo cây quyết định là S2. Năng suất trung bình thực tế là 4,70 tấn/ha, thuộc cấp thích nghi S2 theo đề nghị của FAO. Có thể nhận thấy hạn chế về tầng dày của đất được khắc phục bằng đặc tính lý hóa của loại đất này. Cả hai loại đất đều tơi xốp, thuận lợi cho cơ giới hóa và thích hợp cho nhiều loại cây trồng cạn.

– Tại các đơn vị đất đai có số thứ tự 2 và 6 có loại đất phù sa, có tầng dày > 100 cm, độ dốc < 3°, có tưới mặt và thành phần cơ giới là thịt nhẹ và thịt trung bình, đánh giá thích nghi theo AHP là S3 còn theo phương pháp cây quyết định cho kết quả S2. Năng suất trung bình thực tế là 4,29 tấn/ha, vẫn thuộc cấp thích nghi S2 theo đề nghị của FAO. Ở đây, tuy loại đất là một điều kiện hạn chế nhưng lại có thuận lợi về địa hình, chế độ tưới cũng như tầng dày.

– Tại 2 đơn vị đất đai có số thứ tự 3 và 4 có loại đất là xám gley, tầng dày > 100 cm, độ dốc < 3°, có tưới mặt và thành phần cơ giới là thịt nhẹ, trung bình, đánh giá thích nghi theo AHP là S3 còn theo phương pháp cây quyết định cho kết quả S2. Năng suất trung bình thực tế là 4,51 tấn/ha, thuộc cấp thích nghi S2 theo đề nghị của FAO. Đất xám tuy có độ chua, ít dinh dưỡng nhưng được khắc phục bởi tầng dày, bằng phẳng cùng chế độ tưới khá thuận lợi nên đã nâng cao năng suất.

– Đơn vị đất đai số 9 có tổ hợp loại đất là xám trên phù sa cổ và nâu vàng trên phù sa cổ, tầng dày > 100 cm, độ dốc < 3°, có tưới ngầm và thành phần cơ giới là thịt nhẹ, đánh giá thích nghi theo AHP là S2 còn theo phương pháp định lượng cây quyết định là S1. Năng suất trung bình thực tế là 6,93 tấn/ha, thuộc cấp thích nghi S1 theo đề nghị của FAO. Loại đất ở đây không là loại tối ưu thích hợp tuyệt đối để cây cho năng suất cao nhất. Tuy vậy, với các yếu tố thuận lợi về tầng dày và độ dốc đã nâng khả năng thích nghi lên hạng S1.

Thực hiện so sánh diện tích theo mức độ thích nghi của 2 phương pháp theo AHP và theo sử dụng cây quyết định, ta thấy có sự sai khác.

Bảng 12: So sánh diện tích thích nghi đất đai sử dụng phương pháp AHP và cây quyết định

Mức độ thích nghi	Diện tích (ha)		Tỉ lệ (%)	
	AHP	Cây quyết định	AHP	Cây quyết định
S1 (Thích nghi cao)	0,00	474,67	0,00	0,88
S2 (Thích nghi trung bình)	48.408,40	53.597,70	89,53	99,12
S3 (Ít thích nghi)	5.663,97	0,00	10,47	0,00

Với cấp thích nghi cao, theo phương pháp AHP không tồn tại, theo phương pháp định lượng ứng

dụng cây quyết định là 474,67 ha, chiếm 0,88% diện tích. Cấp thích nghi trung bình đối với phương pháp đánh giá theo AHP là 48.408,40 ha chiếm 89,53%

còn với cây quyết định cho kết quả 53.597,70 ha chiếm 99,12% diện tích. Cấp ít thích nghi của phương pháp AHP cho kết quả 5.663,97 ha chiếm 10,47% diện tích còn theo cây quyết định cho kết quả không tồn tại.

4 KẾT LUẬN

Phương pháp khai phá dữ liệu bằng cây quyết định cho thấy sự tương quan hỗ trợ lẫn nhau giữa các yếu tố đơn tính của đất để đạt mục tiêu cuối cùng là năng suất. Phương pháp đánh giá theo cây quyết định bổ sung căn cứ định lượng cho phân cấp thích nghi, nhưng không hoàn toàn thay thế phương pháp luận đánh giá đất đai dựa trên phân cấp thứ bậc hay theo FAO. Để hoàn thiện tính ứng dụng của mô hình này, người điều tra cần hoàn thiện hơn cấu trúc và nội dung mẫu phiếu cũng như việc điều tra nên được thực hiện ở quy mô lớn hơn.

TÀI LIỆU THAM KHẢO

- Bouma, J., Wagenet, R. J., Hoosbeek, M. R., Hutson, J. L., 1993. Using expert systems and simulation modelling for land evaluation at farm level: a case study from New York State. *Soil Use and Management*. 9(4): 131–139.
- Rosa, D., Diepen, C.A., 2002. Qualitative and Quantitative Land Evaluation. In Willy H. Verheye. *Land Use, Land Cover and Soil Sciences - Volume II: Land Evaluation*. EOLSS, pp. 59-77.
- Han, J., Kamper, M., 2006. *Data Mining: Concepts and Techniques*, Second Edition. Morgan Kaufmann Publishers, Elsevier Inc, 772 pages.
- Tian, J., Hu, Y., Liu, J., Zhao, Y., Wang, C., 2009. The comparative analysis of various classification models on land evaluation. *Proc. SPIE 7492, International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, 74921A*, 15 October 2009, Wuhan, China.
- Kumar, N., Obi Reddy, G. P., Chatterji, S., 2013. Evaluation of best first decision tree on categorical soil survey data for land capability classification. *International Journal of Computer Applications*. 72(4): 5-8.
- Nguyễn Ánh Nga, 2012. Ứng dụng kỹ thuật khai phá dữ liệu cho việc định lượng trong đánh giá đất đai trên địa bàn huyện Định Quán, tỉnh Đồng Nai. Luận văn thạc sĩ Khoa học nông nghiệp. Trường Đại học Nông Lâm Thành phố Hồ Chí Minh.
- Lanen, H.A.J., Hack-ten Broeke, M.J.D., Bouma, J., de Groot, W.J.M., 1992. A mixed qualitative/quantitative physical land evaluation methodology. *Geoderma*. 55(1-2): 37-54.
- Yang, J., Li, T., Chen, Z., 2010. Land evaluation method based on decision tree produced by C4.5 and fuzzy decision. *Agricultural Science & Technology – Hunan*. 11(3): 1-3.