

ĐÁNH GIÁ CÁC YẾU TỐ ẢNH HƯỞNG TỚI QUY HOẠCH SỬ DỤNG ĐẤT BẰNG PHƯƠNG PHÁP HỌC MÁY

Bùi Ngọc Tú^{1*}, Bùi Thị Cẩm Ngọc², Nguyễn Thị Hồng Hạnh³

¹Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội;

²Trường Đại học Tài nguyên và Môi trường;

³Trường Đại học Nông lâm Thành phố Hồ Chí Minh.

*Tác giả liên hệ: buingoctu@hus.edu.vn

Nhận bài: 12/06/2025

Hoàn thành phản biện: 14/09/2025

Chấp nhận bài: 15/09/2025

TÓM TẮT

Đánh giá các yếu tố ảnh hưởng đến kết quả quy hoạch sử dụng đất (QHSDĐ) là nhiệm vụ quan trọng nhưng gặp nhiều thách thức, đặc biệt khi cần đánh giá chi tiết tới từng vị trí quy hoạch trên bản đồ. Với khả năng xử lý dữ liệu lớn, đa dạng có thể mạnh trong phân tích, phát hiện các mối quan hệ phi tuyến tính, học máy càng được ứng dụng trong nhiều lĩnh vực. Nghiên cứu này đề xuất và thử nghiệm giải pháp ứng dụng mô hình học máy để đánh giá các yếu tố ảnh hưởng đến kết quả thực hiện QHSDĐ. Dữ liệu gồm 474 mẫu thể hiện các khoảnh đất quy hoạch giai đoạn 2010 – 2020 tại 4 huyện của tỉnh Sơn La, với 19 biến ảnh hưởng. Mô hình Random Forest (RF) cho kết quả tốt hơn so với Support Vector Machine (SVM) và Decision Tree (DT) ở bốn trong năm chỉ số đánh giá hiệu suất. Trên cơ sở này, vai trò của các yếu tố được lượng hóa bằng kỹ thuật SHAP (SHapley Additive exPlanations) áp dụng cho mô hình RF. Các yếu tố ảnh hưởng lớn nhất đến kết quả thực hiện QHSDĐ gồm: khoảng cách đến khu dân cư, mục đích sử dụng đất quy hoạch, diện tích đất quy hoạch, độ cao. Nghiên cứu này đề xuất một cách tiếp cận mới trong vấn đề đánh giá QHSDĐ, giúp nâng cao tính khách quan, độ tin cậy và góp phần cải thiện hiệu quả công tác QHSDĐ.

Từ khóa: Học máy, Quy hoạch sử dụng đất, Yếu tố ảnh hưởng

A MACHINE LEARNING APPROACH TO ASSESSING THE FACTORS INFLUENCING LAND USE PLANNING IMPLEMENTATION

Bui Ngoc Tu^{1*}, Bui Thi Cam Ngoc², Nguyen Thi Hong Hanh³

¹University of Science, Vietnam National University;

²Hanoi University of Natural Resources and Environment;

³Nong Lam University – Ho Chi Minh city.

Received: 12/06/2025

Revised: 14/09/2025

Accepted: 15/09/2025

ABSTRACT

Evaluating the factors that influence land use planning implementation is a crucial challenging task, particularly when detailed assessments of each planned location on the map are required. Machine learning, with its ability to process large, diverse datasets and detect complex nonlinear relationships, has proven effective across various fields. This study proposes a procedure using the Random Forest machine learning model to evaluate the factors affecting land use planning implementation. The dataset consists of 474 samples representing planned location for the 2010–2020 period in four districts of Son La Province, with 19 influencing variables. Random Forest (RF) outperformed Support Vector Machine (SVM) and Decision Tree (DT) on four of five performance metrics. Accordingly, we employed SHAP (SHapley Additive exPlanations) on the RF model to quantify variable contributions. The most significant factors affecting the success of land use planning implementation include: proximity to residential area, land use planning type, the planned land area, and elevation. This study presents a novel approach to land use planning assessment, enhancing objectivity, reliability, and ultimately improving the effectiveness of land use planning practices.

Keywords: Machine learning, Land use planning, Influencing factors

1. MỞ ĐẦU

Quy hoạch sử dụng đất (QHSDĐ) đóng vai trò quan trọng trong công tác quản lý nhà nước về đất đai. Công tác này giúp định hướng việc phân bổ và sử dụng hợp lý nguồn tài nguyên đất đai, bảo đảm hiệu quả trong việc phát triển kinh tế - xã hội và bảo vệ môi trường. Đánh giá các yếu tố ảnh hưởng tới QHSDĐ giúp nhà quản lý và các chuyên gia quy hoạch hiểu rõ hơn các thuận lợi và vấn đề tồn tại khi triển khai các phương án quy hoạch (Vũ Khắc Hùng và cs., 2022). Tuy vậy, đây không phải là nhiệm vụ dễ dàng vì bài toán này cần xét tới nhiều yếu tố như: kinh tế - xã hội, môi trường, chính sách quản lý,... (Feitelson và cs, 2017; Vũ Khắc Hùng và cs., 2022; Hailu và cs, 2023). Các nguồn dữ liệu này thường phong phú, đa nguồn, thuộc nhiều định dạng. Điều này làm cho việc xử lý, tích hợp các yếu tố vào mô hình đánh giá trở nên phức tạp hơn. Khó khăn này càng đáng kể khi yêu cầu bài toán cần đánh giá chi tiết tới từng khoanh đất quy hoạch với số lượng lớn, đa dạng về đặc điểm không gian và thuộc tính. Vấn đề này đặt ra thách thức với cách tiếp cận truyền thống. Do vậy, cần thiết phải áp dụng phương pháp phân tích hiện đại, có khả năng xử lý khối lượng dữ liệu lớn và đa dạng, kết hợp xử lý được cả dữ liệu không gian và phi không gian.

Trong những năm gần đây, công nghệ học máy (Machine Learning), một nhánh của trí tuệ nhân tạo (AI - Artificial Intelligence), đã được nhiều học giả trên thế giới nghiên cứu và ứng dụng trong các vấn đề liên quan đến QHSDĐ như: dự báo biến động sử dụng đất (Yue và cs., 2024), lựa chọn vị trí tối ưu cho công trình quy hoạch (Al-Ruzouq và Abdallah, 2021; Al Awadh và Mallick, 2024). Khi kết hợp với Hệ thống thông tin địa lý (GIS), học máy phát huy thế mạnh trong việc phân tích dữ liệu không gian, tăng tính khách quan, nâng cao hiệu

suất, và cải thiện độ chính xác của kết quả đầu ra so với cách tiếp cận truyền thống (Yilmaz và cs., 2023; Bilgili và cs., 2024). Mặc dù có nhiều ưu điểm, các mô hình ML truyền thống thường vận hành như một “hộp đen”, gây khó khăn cho việc giải thích quá trình ra quyết định. Do đó, gần đây nhiều nghiên cứu đã áp dụng các kỹ thuật diễn giải mô hình học máy để làm rõ các mối quan hệ tiềm ẩn và định lượng vai trò của các yếu tố ảnh hưởng trong bài toán xây dựng phương án QHSDĐ (Al Awadh và Mallick, 2024; Bilgili và cs., 2024).

Tại Việt Nam, mặc dù còn hạn chế, nhưng một số nghiên cứu đã bước đầu ứng dụng học máy trong công tác quản lý đất đai. Tuy vậy, những nghiên cứu liên quan tới ứng dụng học máy hỗ trợ công tác QHSDĐ vẫn còn khá hạn chế và chưa thật sự phát huy được tiềm năng của công nghệ này. Trong những năm qua, nhiều nghiên cứu trong nước về đánh giá các yếu tố ảnh hưởng tới QHSDĐ đã đạt được nhiều kết quả nhất định. Các nghiên cứu này sử dụng phương pháp khảo sát và các mô hình phân tích thống kê truyền thống để xác định các yếu tố tác động đến QHSDĐ, cung cấp góc nhìn tổng thể đối với công tác QHSDĐ (Nguyễn Minh Thông và Phan Trung Hiền, 2021; Vũ Khắc Hùng và cs., 2022; Vũ Kỳ Long, 2024). Mặc dù vậy, cách tiếp cận này chủ yếu phân tích dựa theo ý kiến từ các bên liên quan mà chưa xem xét kỹ vai trò của dữ liệu, đặc biệt là dữ liệu bản đồ. Điều này có thể phần nào ảnh hưởng tới tính khách quan hoặc thiếu đi các phân tích từ yếu tố không gian. Ngoài ra, cách tiếp cận này cũng gặp khó khăn trong việc xử lý, tích hợp dữ liệu đa nguồn, trong đó có dữ liệu không gian, và đặc biệt là khi cần đánh giá chi tiết từng khoanh đất quy hoạch với khối lượng dữ liệu lớn. Vì vậy, nghiên cứu này hướng tới mục tiêu đề xuất giải pháp đánh giá các yếu tố ảnh hưởng tới QHSDĐ bằng phương

pháp học máy. Kết quả nghiên cứu sẽ góp phần gợi mở cách tiếp cận mới trong vấn đề đánh giá các yếu tố ảnh hưởng đến QHSDĐ, làm phong phú các nghiên cứu về ứng dụng học máy trong công tác quản lý đất đai tại Việt Nam.

2. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Để đánh giá các yếu tố ảnh hưởng tới quy hoạch sử dụng đất (QHSDĐ), nghiên cứu này sử dụng 02 tài liệu chính là bản đồ QHSDĐ, bản đồ hiện trạng sử dụng đất (HTSDĐ) năm cuối của kỳ quy hoạch. Mục đích là để xác định kết quả thực hiện quy hoạch (đã thực hiện hoặc chưa thực hiện quy hoạch) của các khoanh đất quy hoạch khi chồng xếp 2 lớp bản đồ này. Nghiên cứu này thu thập bản đồ QHSDĐ giai đoạn 2010 – 2020, bản đồ HTSDĐ năm 2020 của 04 địa bàn thử nghiệm tại tỉnh Sơn La. Bên cạnh đó, các yếu tố chuyên đề được thu thập và phân tích từ các nguồn tài liệu như: cơ sở dữ liệu nền địa lý tỉnh Sơn La, tỷ lệ 1:50.000 và 1:25.000; các báo cáo có liên quan. Phạm vi khoa học giới hạn đánh giá cho 5 loại hình sử dụng đất là đất ở, đất xây dựng cơ sở y tế, đất xây dựng cơ sở giáo dục và đào tạo, đất khu công nghiệp, cụm công nghiệp (KCN/CCN), đất thương mại, dịch vụ. Dữ liệu không gian được phân tích và xử lý trong phần mềm ArcGIS. Mô hình học máy được đào tạo và huấn luyện bằng ngôn ngữ Python và sử dụng bộ thư viện Scikit-learn. Để đảm tính đa dạng về mặt dữ liệu và giúp mô hình có thể “học” các đặc trưng của các khoanh đất quy hoạch tại các khu vực khác nhau, nghiên cứu này thử nghiệm ở 04 huyện nằm ở phía Đông Nam của tỉnh Sơn La gồm: huyện Mộc Châu, huyện Vân Hồ, huyện Phù Yên và huyện Yên Châu (Hình 2a). Phạm vi nghiên cứu được xác định theo ranh giới hành chính trước khi áp dụng Nghị quyết 1681/NQ-UBTVQH15 năm 2025 về

việc sắp xếp đơn vị hành chính của tỉnh Sơn La. Công tác quản lý đất đai tại địa bàn trong những năm gần đây có nhiều chuyển biến tích cực, góp phần phục vụ phát triển kinh tế - xã hội. Tuy nhiên, do đặc điểm địa hình phức tạp, đặc điểm dân cư, trình độ phát triển giữa các vùng có sự chênh lệch lớn nên công tác quản lý đất đai vẫn còn gặp nhiều khó khăn.

2.2. Phương pháp nghiên cứu

Phương pháp nghiên cứu được khái quát hóa và thể hiện theo bước như sau:

Bước 1: Điều tra, thu thập dữ liệu

Dữ liệu được thu thập từ nhiều nguồn khác nhau, bao gồm bản đồ QHSDĐ giai đoạn 2010 – 2020, bản đồ HTSDĐ năm 2020 (năm của kỳ quy hoạch), báo cáo thuyết minh QHSDĐ, báo cáo tình hình phát triển kinh tế - xã hội tại khu vực nghiên cứu, kết quả khảo sát thực tế, các bài báo trong và ngoài nước có liên quan tới đề tài,...

Bước 2: Đánh giá, phân tích, và chuẩn hóa dữ liệu

Trong bước này, các vấn đề như dữ liệu thiếu, sai lệch hoặc không nhất quán cần được phát hiện và xử lý. Đồng thời, dữ liệu cũng phải được chuẩn hóa, nhất là tài liệu là bản đồ. Dữ liệu bản đồ HTSDĐ và QHSDĐ được thu thập ở định dạng *.dgn. Dữ liệu sẽ được chuyển đổi định dạng cơ sở dữ liệu *.gdb, được chuẩn hóa quan hệ không gian (topology) và thiết lập hệ tọa độ VN2000 trong phần mềm ArcGIS. Nhân mã mục đích sử dụng đất được rà soát, gán đúng theo thông tư Thông tư 08/2024/TT-BTNMT. Ngoài ra, các trường thuộc tính khác sẽ được trích xuất từ các tài liệu liên quan và kết quả phân tích không gian.

Bước 3: Đề xuất danh mục yếu tố ảnh hưởng tới triển khai quy hoạch sử dụng đất

Nhóm nghiên cứu tiến hành tổng hợp, xác định các yếu tố ảnh hưởng đến

QHSDĐ. Các yếu tố này bao gồm các yếu tố như: tự nhiên (như độ cao, độ dốc), yếu tố xã hội (như mật độ dân số, trình độ dân trí, tỷ lệ lao động qua đào tạo), yếu tố kinh tế (như tốc độ tăng trưởng kinh tế, thu nhập), yếu tố về khoảng cách không gian (khoảng cách tới khu dân cư, khoảng cách tới đường giao thông chính,...).

Bước 4: Xác định tình trạng triển khai quy hoạch sử dụng đất

Vị trí các khoanh đất quy hoạch (gọi là mẫu) sẽ được xác định trong bản đồ QHSDĐ. Để đảm bảo tính sẵn sàng của dữ liệu, nghiên cứu này sử dụng bản đồ QHSDĐ giai đoạn 2010 – 2020 để thử nghiệm. Tình trạng triển khai của các mẫu này được xác định bằng cách chồng xếp và so sánh với mã mục đích sử dụng đất ở vị trí tương ứng trong bản đồ hiện trạng sử dụng đất năm 2020. Các mẫu này được phân loại thành 2 nhóm là “Đã thực hiện quy hoạch” hoặc “Chưa thực hiện quy hoạch”.

Bước 5: Tính toán, xác định giá trị các yếu tố ảnh hưởng

Đối với từng mẫu, bên cạnh việc xác định tình trạng triển khai quy hoạch (như tại bước 4), cần phải tính toán, xác định giá trị cụ thể của các yếu tố ảnh hưởng (có thể gọi là đặc trưng). Các giá trị này được tính toán thông qua phân tích không gian trong GIS hoặc trích xuất từ các báo cáo. Các yếu tố về khoảng cách không gian được phân tích khoảng cách Euclid (khoảng cách theo đường chim bay) bằng công cụ *Euclidean Distance* của phần mềm ArcGIS. Thuộc tính giá trị khoảng cách của từng mẫu sẽ được gán bằng trung bình giá trị của các pixel nằm trong mẫu bằng công cụ *Zonal statistic as table*. Yếu tố độ cao, độ dốc được nội suy từ điểm độ cao và được gán giá trị tương tự như đối với thuộc tính khoảng cách không gian. Do khó xác định được chi tiết theo đặc điểm của từng vị trí quy hoạch nên các yếu tố kinh tế - xã hội

trong nghiên cứu này được coi như yếu tố nền và được cụ thể theo cấp đơn vị hành chính (trước thời điểm áp dụng mô hình địa phương 2 cấp). Trong đó, một số yếu tố được xác định chi tiết tới cấp xã (ví dụ: mật độ dân số, quy mô diện tích xã), một số yếu tố được xác định theo cấp huyện (ví dụ: trình độ dân trí, nhu nhập,...). Giá trị cụ thể được đánh giá theo thang đo Likert 3 cấp độ (3 : cao, 2: trung bình, 1: thấp), trên cơ sở tổng hợp các báo cáo liên quan và tham vấn chuyên gia địa phương.

Bước 6: Tiền xử lý dữ liệu phục vụ xây dựng mô hình học máy

Các mẫu sau khi đã xác định đầy đủ tình trạng quy hoạch trong giai đoạn 2010 – 2020 và giá trị các yếu tố ảnh hưởng trong phần mềm ArcGIS sẽ được kết xuất thành file excel. Bước tiền xử lý dữ liệu sẽ tiến hành kiểm tra những giá trị thiếu, trùng lặp của các mẫu để cập nhật giá trị hoặc sàng lọc loại bỏ. Các biến có kiểu dữ liệu dạng phân loại sẽ được mã hóa dưới dạng số. Sau đó, dữ liệu sẽ được chia thành 2 tập là tập huấn luyện và tập kiểm tra với tỷ lệ tương ứng là 80% và 20%.

Bước 7: Đào tạo và tối ưu hóa siêu tham số cho mô hình học máy

Nghiên cứu này sử dụng ba thuật toán học máy phổ biến gồm: Decision Tree (DT), Random Forest (RF) và Support Vector Machine (SVM). Decision Tree (DT) là thuật toán học có giám sát dùng cho cả phân loại và hồi quy. Mô hình biểu diễn dưới dạng cây: mỗi nút là thuộc tính, các nhánh là quy tắc quyết định và các lá là dự đoán cuối cùng. Mô hình DT được đánh giá là dễ hiểu và trực quan hóa nhờ nhờ logic Boolean (IBM, 2025). Phương pháp RF hoạt động dựa trên nguyên lý kết hợp nhiều cây quyết định, trong đó mỗi cây được xây dựng từ một mẫu dữ liệu ngẫu nhiên, chỉ sử dụng một tập con ngẫu nhiên các biến đầu vào và chia nhánh tại mỗi nút. Quá trình này

được lặp lại nhiều lần để tạo ra một “rừng” gồm nhiều cây, từ đó cho phép mô hình có khả năng khái quát hóa tốt và hạn chế hiện tượng quá khớp (Breiman, 2001). Support Vector Machine (SVM) là thuật toán học có giám sát dựa trên lý thuyết học thống kê. Mục tiêu của SVM là xây dựng siêu phẳng phân tách hai lớp mẫu với biên tối đa để nâng cao khả năng tổng quát hóa (Cervantes và cs., 2020). Trong nghiên cứu này, các thuật toán được sử dụng để xây dựng mô hình phân loại nhị phân nhằm đánh giá các yếu tố ảnh hưởng đến QHSDĐ đất tại khu vực nghiên cứu trong giai đoạn 2010 - 2020. Với nhãn mục tiêu (y) là tình trạng quy hoạch với hai giá trị: “Đã thực hiện quy hoạch” (y = 1) và “Chưa thực hiện quy hoạch” (y = 0), trong khi các biến đầu vào (X) bao gồm các yếu tố ảnh hưởng. Hiệu suất mô hình phụ thuộc nhiều vào tối ưu siêu tham số. Nghiên cứu sử dụng kỹ thuật *Tìm kiếm Lưới* (GridSearchCV) kết hợp *Xác thực chéo* (Cross-Validation) 5-fold. Kết quả chỉ số đo hiệu suất được tính toán bằng giá trị trung bình của 5 lần lặp lại, đảm bảo đánh giá mô hình một cách khách quan, đảm bảo độ tin cậy. Cách tiếp cận này cũng tương đồng với một số nghiên cứu tương tự của Al-Ruzouq và cs. (2021), Bilgili và cs. (2024).

Bước 8: Đánh giá và so sánh hiệu suất mô hình học máy

Trong đó: TP (True Positive) biểu thị tổng số mẫu được phân loại đúng nhãn “Đã thực hiện quy hoạch”; TN (True Negative) biểu thị tổng số mẫu được phân loại đúng nhãn “Chưa thực hiện quy hoạch”; FP

$$TPR = TP / (TP + FN), FPR = FP / (FP + TN).$$

Bước 9: Diễn giải mô hình và phân tích các yếu tố ảnh hưởng

Đánh giá và so sánh hiệu suất là bước then chốt trong lựa chọn mô hình. Mô hình được coi là tối ưu khi tổng thể các chỉ số hiệu suất tốt hơn so với các mô hình còn lại, và sẽ được sử dụng để diễn giải, phân tích các yếu tố ảnh hưởng tới QHSDĐ. Trong nghiên cứu này, năm chỉ số được sử dụng gồm: *Accuracy*, *Recall*, *Precision*, *F1 Score* và *AUC*. Trong đó, *Accuracy* (độ chính xác) phản ánh tỷ lệ dự đoán đúng trên tổng số mẫu; *Precision* đo lường tỷ lệ mẫu được dự đoán là dương và thực sự thuộc lớp dương; *Recall* cho biết tỷ lệ phát hiện đúng các trường hợp thuộc lớp dương; *F1 Score* là trung bình điều hòa giữa *Precision* và *Recall*, đặc biệt hữu ích khi dữ liệu không cân bằng giữa hai lớp; và chỉ số *AUC* (diện tích dưới đường cong) được sử dụng để đánh giá khả năng phân biệt giữa hai lớp của mô hình, với giá trị *AUC* càng gần 1 cho thấy mô hình càng hiệu quả. Công thức tính như sau (Awadh và Mallick, 2024):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1\ Score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4)$$

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (5)$$

(False Positive) là tổng số mẫu được phân loại sai vào nhãn “Đã thực hiện quy hoạch”; FN (False Negative) là tổng số mẫu phân loại sai vào nhãn “Chưa thực hiện quy hoạch”.

Việc diễn giải mô hình và phân tích các yếu tố ảnh hưởng tới QHSDĐ được thực hiện bằng việc áp dụng kỹ thuật SHAP (SHapley Additive exPlanations). Đây một

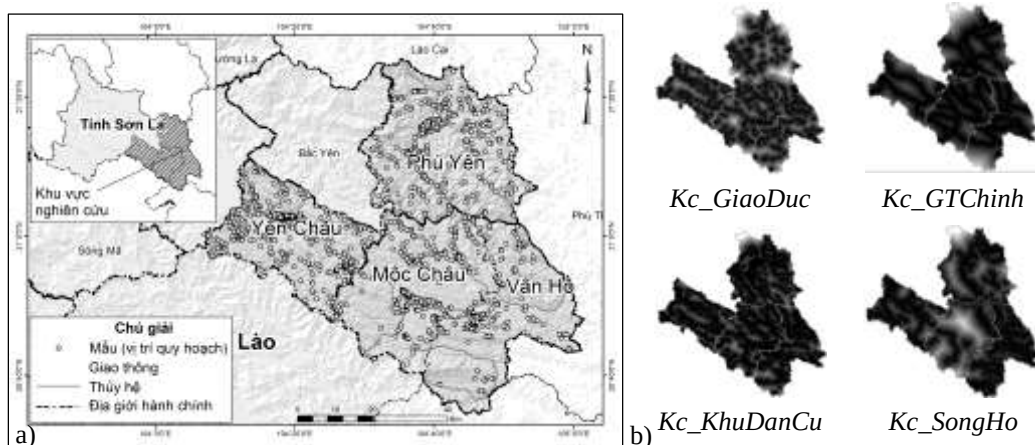
kỹ thuật hiện đại để diễn giải mô hình học máy, được nhiều nghiên cứu công nhận (Bilgili và cs., 2024). Kỹ thuật SHAP dựa trên lý thuyết trò chơi, gán cho mỗi đặc trưng một giá trị quan trọng thể hiện đóng góp của nó vào dự đoán (Lundberg và Lee, 2017). Trong nghiên cứu này, kỹ thuật SHAP giúp xác định vai trò, mức độ quan trọng của các yếu tố đầu vào (là danh mục yếu tố ảnh hưởng) với nhãn đầu ra (tình trạng thực hiện QHSDD), trực quan hóa kết quả diễn giải bằng biểu đồ SHAP Bar Plot và SHAP Beeswarm.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Chuẩn hóa và xử lý dữ liệu

Các bản đồ hiện trạng sử dụng đất và quy hoạch sử dụng đất các huyện được chuyển từ định dạng *.dgn sang *.shp thông

qua phần mềm FME Workbench. Các lớp dữ liệu được kiểm tra và chỉnh sửa lỗi topology theo quy tắc *Must Not Overlap* (không có lỗi chồng đè) và *Must Not Have Gaps* (không có lỗi khoảng trống). Các lớp chuyên đề như: khu dân cư, cơ sở giáo dục, độ cao,... được bóc tách từ bản đồ HTSDD, cơ sở dữ liệu nền địa lý. Các lớp này được chuyển đổi từ dạng vector sang raster (độ phân giải 10m), tính toán khoảng cách Euclid bằng công cụ *Euclidean Distance* trong phần mềm ArcGIS. Các khoanh đất quy hoạch được kết xuất từ bản đồ QHSDD. Sau khi sàng lọc theo 05 loại hình sử dụng đất thử nghiệm, tổng số mẫu giữ lại là 474 mẫu. Tình trạng triển khai quy hoạch của các mẫu được xác định theo bước 4 nội dung trình bày tại mục 2.2. Một số kết quả xử lý dữ liệu được thể hiện như Hình 1.



Hình 1. Sơ đồ khu vực nghiên cứu (a) và kết quả xử lý một số lớp dữ liệu chuyên đề (b)

3.2. Đề xuất danh mục các yếu tố ảnh hưởng

Trên cơ sở tổng quan tài liệu, ý kiến chuyên gia và thực tiễn tại địa phương, nghiên cứu này đề xuất bộ danh mục gồm

19 biến thuộc 4 nhóm yếu tố, trình bày tại Bảng 1 như sau:

Bảng 1. Danh mục các yếu tố ảnh hưởng

Nhóm yếu tố	Tên biến	Mô tả	Phương pháp đánh giá Nguồn dữ liệu
Đặc điểm khoanh đất quy hoạch	DienTich	Diện tích khoanh đất quy hoạch	Calculate Geometry Bản đồ QHSDD
	MDSD_QH	Mục đích sử dụng đất theo quy hoạch (1: đất ở, 2: đất xây dựng cơ sở giáo dục và đào tạo, 3: xây dựng cơ sở y tế, 4: đất KCN/CCN, 5: đất thương mại dịch vụ)	Field Calculator Bản đồ QHSDD
Khoảng cách không gian	Kc_KhuDanCu	Khoảng cách tới khu dân cư	Euclidean distance Bản đồ QHSDD, bản đồ HTSDD, cơ sở dữ liệu nền địa lý
	Kc_GTTuong	Khoảng cách tới giao thông thường	
	Kc_GTChinh	Khoảng cách tới giao thông chính	
	Kc_GiaoDuc	Khoảng cách tới cơ sở giáo dục	
	Kc_YTe	Khoảng cách tới cơ sở y tế	
	Kc_SongHo	Khoảng cách tới sông, hồ chính	
	Kc_BaiRac	Khoảng cách tới bãi rác	
Địa hình	Kc_NghiaTrang	Khoảng cách tới nghĩa trang	Spatial Interpolation Cơ sở dữ liệu nền địa lý Calculate Geometry Bản đồ QHSDD Field Calculator Bản đồ QHSDD
	DoCao	Độ cao của khoanh đất quy hoạch	
	DoDoc	Độ dốc của khoanh đất quy hoạch	
	QuyMoDTXa	Quy mô diện tích xã	
Kinh tế - xã hội	MatDoDanSo	Mật độ dân số xã	Field Calculator Bản đồ QHSDD, báo cáo thuyết minh QHSDD, báo cáo tổng hợp quy hoạch tỉnh, khảo sát cán bộ quản lý
	DoThiHoa	Tốc độ đô thị hóa	
	HoNgheo	Tỷ lệ hộ làm hộ nghèo	
	DanTri	Trình độ dân trí	
	LDDaoTao	Tỷ lệ lao động qua đào tạo	
	ThuNhap	Thu nhập trung bình của hộ gia đình	

QHSDD: Quy hoạch sử dụng đất; HTSDD: Hiện trạng sử dụng đất

3.3. Đào tạo và đánh giá hiệu suất mô hình học máy

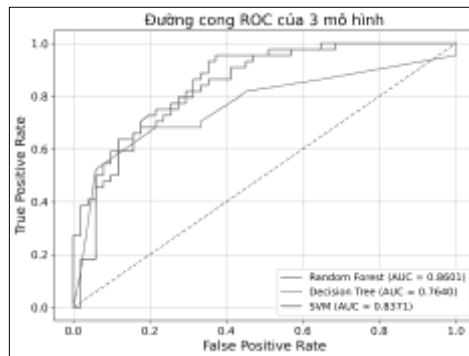
Hiệu suất của ba mô hình học máy sau khi đào tạo và tối ưu hóa siêu tham số như sau:

Bảng 2. Kết quả hiệu suất của ba mô hình học máy

Mô hình	Accuracy	Precision	Recall	F1-score	AUC
DT	0,7368	0,7317	0,6818	0,7059	0,7640
RF	0.7579	0,7234	0,7727	0,7473	0,8601
SVM	0,7564	0,7400	0,7400	0,7400	0,8371

Kết quả được thể hiện ở Bảng 2 cho thấy, mô hình RF có kết quả tốt nhất trong hầu hết các chỉ số đánh giá. Cụ thể, các chỉ số Accuracy đạt 75,79%, Recall 77,27%, F1 74,73%, cho thấy đồng thời mức độ dự báo chính xác, khả năng phát hiện đúng các trường hợp dương và sự cân bằng hợp lý giữa Precision–Recall, tốt hơn so với SVM và DT. Mô hình SVM có chỉ số Precision

cao nhất (74,00%) nhưng các chỉ số còn lại kém hơn RF, trong khi DT xếp sau cùng. Biểu đồ thể hiện chỉ số AUC (Hình 2) cũng khẳng định ưu thế của RF khi đạt 0,8601, nhỉnh hơn hai mô hình còn lại. Do đó, RF được xem là mô hình tốt nhất trong nghiên cứu này và được lựa chọn cho các phân tích, diễn giải các yếu tố ảnh hưởng tới QHSDD.

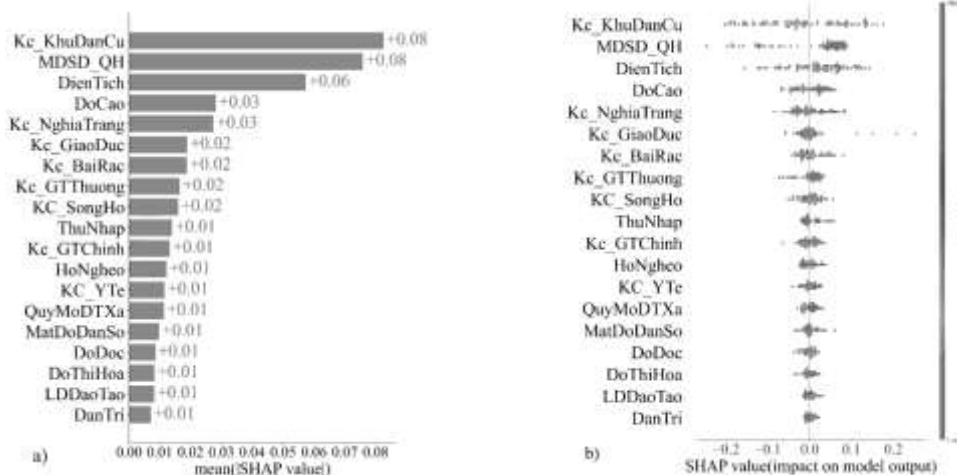


Hình 2. Biểu đồ thể hiện chỉ số AUC của ba mô hình

3.5. Đánh giá các yếu tố ảnh hưởng

Biểu đồ SHAP Bar Plot (Hình 3a) và SHAP Beeswarm (3b) cung cấp góc nhìn trực quan về các yếu tố ảnh hưởng đến QHSDĐ. Biểu đồ SHAP Bar Plot chỉ ra rằng các yếu tố quan trọng nhất là *Kc_KhuDanCu* và *MDSD_QH*, cả hai đều có giá trị SHAP cao (+0.08), cho thấy ảnh

hưởng mạnh mẽ của chúng đến việc quyết định tình trạng triển khai quy hoạch. Các yếu tố như *Kc_NghiaTrang*, *DoCao*, và *Kc_GiaoDuc* có mức độ ảnh hưởng trung bình (+0.02 đến +0.03), trong khi các yếu tố kinh tế - xã hội như *ThuNhap*, *HoNgheo*, *DanTri* có ảnh hưởng không nhiều tới kết quả thực hiện QHSDĐ.



Hình 3. Biểu đồ SHAP Bar Plot (a) và biểu đồ (b) SHAP Beeswarm

Biểu đồ SHAP Beeswarm hiển thị mức độ ảnh hưởng của từng biến đầu vào. Trục hoành thể hiện giá trị SHAP của từng mẫu, với mỗi điểm phản ánh tác động cụ thể của biến lên kết quả dự đoán. Màu sắc của điểm, từ đỏ (giá trị cao) đến xanh (giá trị thấp), cho thấy sự phân bố của các giá trị SHAP và tác động của chúng lên kết quả. Có thể thấy, yếu tố *Kc_KhuDanCu* là yếu tố quan trọng nhất, với giá trị SHAP trải rộng cả hai phía âm và dương. Các điểm màu xanh (gần khu dân cư) chủ yếu nằm ở phía

SHAP dương, cho thấy các khu vực này có khả năng được thực hiện quy hoạch cao hơn, phù hợp với thực tế khi các khu vực gần dân cư thuận lợi cho phát triển hạ tầng và dịch vụ, đáp ứng nhu cầu dân sinh và phát triển cộng đồng.

Yếu tố *MDSD_QH* cũng có ảnh hưởng mạnh đến kết quả dự đoán. Các điểm có giá trị thấp (màu xanh), tương ứng với các loại đất: đất ở (1), đất xây dựng cơ sở giáo dục và đào tạo (2), đất y tế (3), chủ yếu phân bố phía SHAP dương, cho thấy các

loại đất này có xu hướng được quy hoạch tốt hơn so với các loại đất KCN/CCN (4) và đất thương mại dịch vụ (5). Điều này có thể phản ánh xu hướng ưu tiên hơn việc triển khai quy hoạch đối với các loại đất ở và đất phục vụ cơ sở hạ tầng xã hội (trường học, bệnh viện) trong giai đoạn 2010 – 2020.

Yếu tố *DienTich* có ảnh hưởng đáng kể đến kết quả triển khai QHSDĐ. Các khoanh đất quy hoạch có diện tích nhỏ có xu hướng có kết quả triển khai quy hoạch không tốt so với các khoanh đất có diện tích lớn. Nguyên nhân có thể do các thửa có diện tích lớn chủ yếu là khu dân cư mới, khu đô thị, KCN/CCN,... được địa phương chú trọng hơn.

Độ cao và độ dốc có đóng góp nhất định trong mô hình, trong đó độ cao có vai trò nổi bật hơn trong kết quả thực hiện QHSDĐ. Từ biểu đồ cho thấy, các khu vực có độ cao lớn (màu đỏ, nằm phía trục SHAP dương) và độ dốc thấp (màu xanh, nằm phía trục SHAP dương) xu hướng có kết quả thực hiện quy hoạch tốt hơn. Có thể là do các khu vực có độ cao lớn, độ dốc vừa phải thường ít bị ảnh hưởng bởi ngập lụt, thuận lợi cho xây dựng cơ sở hạ tầng.

Khoảng cách tới nghĩa trang và bãi rác thường có ảnh hưởng tới việc triển khai QHSDĐ do đây là yếu tố có tính nhạy cảm. Mặc dù chưa có sự biểu diễn rõ ràng so với một số yếu tố khác nhưng từ biểu đồ SHAP, ta có thể thấy các khu vực xa nghĩa trang, bãi rác (màu đỏ - thường nằm ở phía SHAP dương), có khả năng thực hiện quy hoạch cao hơn. Thực tế, quy hoạch yêu cầu nghĩa trang cách xa khu dân cư và công trình xã hội để bảo vệ môi trường, hạn chế ảnh hưởng đến người dân. Ngoài ra, các biến như khoảng cách tới cơ sở giáo dục, đường giao thông thường, cơ sở y tế, sông/hồ,... không thể hiện rõ sự tác động tới tình trạng QHSDĐ từ kết quả phân tích của các biểu đồ.

Trong nhóm yếu tố kinh tế - xã hội, *ThuNhap* là biến có ảnh hưởng nhiều nhất đến QHSDĐ. Trên biểu đồ SHAP, giá trị âm của *ThuNhap* cho thấy mức thu nhập cao gắn với xu hướng xác suất triển khai QHSDĐ thấp hơn. Giả thuyết có thể đặt ra là: các khu vực có thu nhập cao thường là các khu vực thành thị, QHSDĐ tại khu vực này thường đối mặt với các vấn đề về quỹ đất, giải phóng mặt bằng,... Nhận định này phù hợp khi liên hệ với tình hình QHSDĐ tại huyện Mộc Châu. Đây là huyện có lợi thế về điều kiện phát triển kinh tế tốt nhất so với 3 huyện còn lại là: Yên Châu, Phù Yên và Vân Hồ. Tuy nhiên, tỷ lệ số khoanh đất đã thực hiện quy hoạch chỉ đứng thứ hai ($67/128 \approx 52\%$), xếp sau huyện Vân Hồ ($84/142 \approx 59\%$). Các yếu tố còn lại như: *HoNgheo*, *MatDoDanSo*, *QuyMoDTXa*, *DoThiHoa*, *LDDaoTao* và *DanTri* có ảnh hưởng tới kết quả thực hiện QHSDĐ không nổi bật như yếu tố *ThuNhap*.

Phân tích trong nghiên cứu này tương đồng với một số nghiên cứu trước đây (Nguyễn Minh Thông và Phan Trung Hiền, 2021; Vũ Khắc Hùng và cs., 2022), khi các yếu tố liên quan tới đặc điểm tự nhiên đóng vai trò quan trọng. Nghiên cứu cũng đã bổ sung và đánh giá thêm các yếu tố về khoảng cách không gian và đặc điểm khoanh đất quy hoạch. Mặc dù trong thực tế các yếu tố kinh tế - xã hội có tác động đáng kể tới QHSDĐ, tuy vậy vai trò của các yếu tố đó chưa thực sự được nổi bật trong nghiên cứu này. Điều này có thể lý giải do dữ liệu chưa đủ chi tiết và có sự phân biệt rõ ràng tới từng mẫu. Mặt khác, nghiên cứu chưa đề cập tới thông tin pháp lý – một yếu tố quan trọng và ảnh hưởng đáng kể tới QHSDĐ. Do vậy, trong giai đoạn tiếp theo, nghiên cứu sẽ tích hợp yếu tố này vào mô hình học máy để kết quả đánh giá có tính đa chiều hơn.

4. KẾT LUẬN

Nghiên cứu đã đề xuất và thử nghiệm giải pháp ứng dụng học máy để đánh giá các yếu tố ảnh hưởng đến kết quả thực hiện quy hoạch sử dụng đất (QHSDĐ). Giải pháp được thử nghiệm tại 4 huyện trên địa bàn tỉnh Sơn La và bước đầu cho kết quả tích cực. So sánh ba mô hình học máy (RF, SVM, DT) cho thấy RF có chỉ số hiệu suất tốt hơn một chút so với hai mô hình còn lại. Áp dụng kỹ thuật SHAP để đánh giá yếu tố ảnh hưởng tới QHSDĐ cho thấy, khoảng cách đến khu dân cư, mục đích sử dụng đất theo quy hoạch, diện tích khoanh đất là ba yếu tố có ảnh hưởng lớn nhất. Các yếu tố địa hình và khoảng cách không gian cũng đóng vai trò quan trọng. Mặc dù việc đánh giá thực hiện trong giai đoạn QHSDĐ 2010 – 2020 nhưng kết quả này vẫn có thể giúp các chuyên gia và các nhà quản lý có thêm góc nhìn ngoài các báo cáo hành chính, góp phần nâng cao hiệu quả công tác QHSDĐ tại địa phương. Đồng thời, giải pháp này có thể được tái áp dụng ở những địa bàn khác khi có thêm dữ liệu. Việc mở rộng và đa dạng hóa dữ liệu sẽ giúp mô hình phân biệt rõ hơn các đặc trưng riêng của từng khu vực.

Mặc dù nghiên cứu đã đạt được một số kết quả tích cực nhưng bài báo này còn một số hạn chế nhất định. Do đặc điểm của dữ liệu hiện có nên kết quả phân tích tập yếu vào các yếu tố định lượng mà chủ yếu liên quan tới yếu tố không gian, trong khi số lượng biến nhóm yếu tố kinh tế - xã hội còn chưa được phong phú, chưa đề cập tới yếu tố mang tính chính sách và pháp luật trong mô hình học máy. Các nghiên cứu tiếp theo sẽ xây dựng giải pháp để đánh giá rõ hơn tác động của các yếu tố này đối với kết quả thực hiện QHSDĐ, có thể đánh giá kết quả thực hiện QHSDĐ trong thời gian thực hiện QHSDĐ giai đoạn 2021 – 2030. Ngoài ra, việc thử nghiệm thêm các mô hình học máy

khác sẽ được nghiên cứu tiếp theo thực hiện để so sánh và lựa chọn mô hình phù hợp với bối cảnh dữ liệu đa dạng hơn.

TÀI LIỆU THAM KHẢO

1. Tài liệu tiếng Việt

- Bộ Tài nguyên và Môi trường. (2024). Thông tư số 08/2024/TT-BTNMT ngày 31 tháng 7 năm 2024 quy định về thống kê, kiểm kê đất đai và lập bản đồ hiện trạng sử dụng đất, Hà Nội.
- Vũ Khắc Hùng, Trần Văn Tuấn và Phạm Thuý Trang. (2022). Đánh giá những yếu tố ảnh hưởng đến kết quả thực hiện quy hoạch sử dụng đất thành phố Quy Nhơn thời kỳ 2011-2020, *Tạp chí Khoa học Đo đạc và Bản đồ*, số 54-12/2022.
- Vũ Kỳ Long. (2024). Yếu tố ảnh hưởng đến quản lý quy hoạch sử dụng đất của thành phố Hà Nội, *Tạp chí quản lý nhà nước*, đăng ngày 25/7/2024.
- Nguyễn Minh Thông và Phan Trung Hiền. (2021). Xác định yếu tố ảnh hưởng đến quy hoạch sử dụng đất đai trên địa bàn thành phố Cần Thơ theo góc nhìn người sử dụng đất, *Tạp chí Khoa học Trường Đại học Cần Thơ*, 57(3), 23-32.
- Ủy ban nhân dân tỉnh Sơn La. (2024). Báo cáo tổng hợp quy hoạch tỉnh Sơn La thời kỳ 2021-2030, tầm nhìn đến 2050, Sở Kế hoạch và Đầu tư tỉnh Sơn La.
- Ủy ban nhân dân huyện Mộc Châu. (2022). Báo cáo tổng hợp thuyết minh quy hoạch sử dụng đất thời kỳ 2021 – 2030, tầm nhìn đến năm 2050 và kế hoạch sử dụng đất năm 2022 huyện Mộc Châu, tỉnh Sơn La.
- Ủy ban nhân dân huyện Yên Châu. (2022). Báo cáo tổng hợp thuyết minh quy hoạch sử dụng đất thời kỳ 2021 – 2030, tầm nhìn đến năm 2050 và kế hoạch sử dụng đất năm 2022 huyện Yên Châu, tỉnh Sơn La.
- Ủy ban nhân dân huyện Vân Hồ. (2022). Báo cáo tổng hợp thuyết minh quy hoạch sử dụng đất thời kỳ 2021 – 2030, tầm nhìn đến năm 2050 và kế hoạch sử dụng đất năm 2022 huyện Vân Hồ, tỉnh Sơn La.
- Ủy ban nhân dân huyện Phù Yên. (2022). Báo cáo tổng hợp thuyết minh quy hoạch sử dụng đất thời kỳ 2021 – 2030, tầm nhìn đến năm 2050 và kế hoạch sử dụng đất năm 2022 huyện Phù Yên, tỉnh Sơn La.

2. Tài liệu nước ngoài

- Al Awadh, M., & Mallick, J. (2024). A decision-making framework for landfill site selection

- in Saudi Arabia using explainable artificial intelligence and multi-criteria analysis, *Environmental Technology & Innovation*, Volume 33, 103464.
- Al-Ruzouq, R., & Abdallah. (2021). Dam Site Suitability Mapping and Analysis Using an Integrated GIS and Machine Learning Approach, *Water*, 11, 1880.
- Bilgili, A., Arda, T., & Kilic, B. (2024). Explainability in wind farm planning: A machine learning framework for automatic site selection of wind farms, *Energy Conversion and Management*, 309, 118441.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Cervantes, J., García-Lamont, F., Rodríguez-Mazahua, L., & López, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215.
- Feitelson, E., Felsenstein, D., Razin, E., & Stern, E. (2017). Assessing land use plan implementation: Bridging the performance-conformance divide, *Land Use Policy*, 61, 251-264.
- IBM. (2025, April 18). What is a decision tree? In IBM Topics. Retrieved August 17, 2025, from <https://www.ibm.com/topics/decision-trees>
- Hailu, T.; Assefa, E.; & Zeleke, T. (2023). Land use planning implementation and its effect on the ecosystem in Addis Ababa, Ethiopia, *Environmental Challenges*, 13, 100798.
- Lundberg, S., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4766-4777.
- Yilmaz, I.; Adem, A., & Dağdeviren, M. (2023). A machine learning-integrated multi-criteria decision-making approach based on consensus for selection of energy storage locations, *Journal of Energy Storage*, 69, 107941.
- Yue, W., Qin, C., Su, M., Teng, Y., & Xu, C. (2024). Simulation and prediction of land use change in Dongguan of China based on ANN cellular automata - Markov chain model. *Environmental and Sustainability Indicators*, 22, 100355.