Land Price Prediction with Machine Learning in Mueang Khon Kaen District
การทำนายราคาที่ดินด้วยการเรียนรู้ของเครื่องในอำเภอเมืองขอนแก่น

Yosita Sriwuttisap[1], Dusita Sungklinhom[2], Sakpod Tongleamnak[3], and Thanaphon Tangchoopong[4,*]
โยษิตา ศรีวุฒิทรัพย์[1], ดุสิตา สังข์กลิ่นหอม[2], ศักดิ์พจน์ ทองเลี่ยมนาค[3], และ ธนพล ตั้งชูพงศ์[4,*]

Received: 20 March 2024;
Revised: 18 May 2024;
Accepted: 22 May 2024;
Published: 15 August 2024;

Abstract

This research aims to develop a model of land price assessment and study of factors influencing land prices with machine learning used as a guideline for determining the estimated price to be close to the actual purchase price in Mueang Khon Kaen District from land price information traded on websites in enforcement department of 193 locations, and land price assessment information in land department of 1,500 locations in this study the factors involved include appraisal value, property type, land size, distance, and the average appraisal value of five nearby plots of land. The models used for analysis are Regression Tree, Random Forest, Gradient Boosted Trees, and Linear Regression. Which, the land price assessment from the case study by model measurement in MAE, RMSE, R-squared, Grid Search, and Cross-validation to selected model parameters and evaluate their performance. The results found that the model with the best predictive performance is Gradient Boosted Trees in R-squared at the highest of 0.80, MAE, and RMSE at the lowest of 7929.40, and 15281.33, respectively. Feature Importance in the locations with the most influence on prediction, followed by area size, average appraised value from five nearby locations, and property type.

Keywords: Land Price Assessment, Machine Learning, Important Factors.

[1] Student, Department of Computing Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand; นักศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002 ประเทศไทย; Email: yosita.sr@kkumail.com

[2] Student, Department of Computing Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand; นักศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002 ประเทศไทย; Email: dusita.su@kkumail.com

[3] Lecturer, Dr., Department of Computing Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand; อาจารย์ ดร. สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002 ประเทศไทย; Email: sakpod@kku.ac.th

[4,*] Lecturer, Department of Computing Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand; อาจารย์ สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002 ประเทศไทย; Email: thanaphon@kku.ac.th

*Corresponding authors: Thanaphon Tangchoopong (thanaphon@kku.ac.th)

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลการประเมินราคาที่ดิน และศึกษาปัจจัยที่มีอิทธิพลต่อราคาที่ดิน ด้วยการเรียนรู้ของเครื่อง ที่ใช้เป็นแนวทางของการกำหนดราคาประเมินให้ได้ใกล้เคียงกับราคาซื้อขายจริงในอำเภอเมืองขอนแก่น จากข้อมูลราคาที่ดินที่มีการซื้อขายในเว็บไซต์กรมบังคับคดี 193 แห่ง และข้อมูลราคาประเมินของกรมที่ดิน 1500 แห่ง ในการศึกษาปัจจัยที่เกี่ยวข้องได้แก่ ราคาประเมิน ประเภททรัพย์ ขนาดพื้นที่ ระยะทาง และราคาประเมินเฉลี่ยจากที่ดินห้าผืนใกล้เคียง โมเดลที่ใช้ในการวิเคราะห์ ได้แก่ ต้นไม้ตัดสินใจแบบถดถอย แรนดอมฟอเรสท์ เกรเดียนบูสทรี และการถดถอยเชิงเส้น ซึ่งการประเมินราคาที่ดินจากกรณีศึกษาโดยการวัดผลโมเดลในค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวน การค้นหาแบบกริดและการประเมินประสิทธิภาพแบบไขว้ทบเพื่อคัดเลือกพารามิเตอร์ของโมเดล และประเมินประสิทธิภาพ ผลวิจัยพบว่า โมเดลที่มีผลการทำนายข้อมูลที่ดีที่สุดคือ เกรเดียนบูสทรี ที่มีค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวนสูงสุดเท่ากับ 0.80 ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองต่ำสุด เท่ากับ 7929.40 และ 15281.33 ตามลำดับ ความสำคัญของคุณลักษณะในกลุ่มสถานที่ที่มีผลต่อการทำนายมากที่สุด รองลงมาคือขนาดพื้นที่ ราคาประเมินเฉลี่ยจากห้าตำแหน่งใกล้เคียง และประเภททรัพย์

คำสำคัญ: ประเมินราคาที่ดิน, การเรียนรู้ของเครื่อง, ปัจจัยสำคัญ

1. บทนำ (Introduction)

พื้นที่ในเขตอำเภอเมือง จังหวัดขอนแก่น มีการพัฒนาและเติบโตอย่างต่อเนื่อง เป็นศูนย์กลางในการขยายธุรกิจในภูมิภาคตะวันออกเฉียงเหนือรวมถึงธุรกิจอสังหาริมทรัพย์ ซึ่งส่งผลให้ความต้องการที่ดินเพื่อการพัฒนาเพิ่มสูงขึ้น ราคาที่ดินคือมูลค่าปัจจุบันของผลประโยชน์จากการใช้ที่ดิน และเกี่ยวข้องกับธุรกรรมทางการเงินและเศรษฐกิจ การกำหนดราคาที่ดินในประเทศไทยดำเนินการโดยทั้งหน่วยงานภาครัฐและเอกชน โดยกรมธนารักษ์ กระทรวงการคลัง เป็นหน่วยงานหลักที่รับผิดชอบการประเมินและจัดทำบัญชีราคาที่ดินเพื่อใช้เป็นฐานในการจัดเก็บภาษีและค่าธรรมเนียมต่าง ๆ อย่างไรก็ตามราคาประเมินที่ดินมักต่ำกว่าราคาตลาดจริง (Kumpu & Piyathamronchai, 2024) ด้วยเหตุนี้การพัฒนาโมเดลในการประเมินราคาที่ดินจึงมีความสำคัญอย่างยิ่งเพื่อช่วยภาครัฐ และให้ผู้ที่สนใจลงทุนในที่ดินมีข้อมูลที่ดีและแม่นยำมากยิ่งขึ้น

งานวิจัยนี้มุ่งเน้นที่การพัฒนาแบบจำลองเพื่อประเมินราคาที่ดินในเขตอำเภอเมือง จังหวัดขอนแก่นโดยใช้ข้อมูลที่มีความเกี่ยวข้อง โดยแบ่งออกเป็นสามกลุ่มปัจจัย ได้แก่ ปัจจัยด้านราคาประเมิน ปัจจัยด้านขนาดและปัจจัยทำเลที่ตั้ง และปัจจัยด้านประเภททรัพย์ (Soltani et al., 2022) ดังแสดงในกรอบแนวคิดในการวิจัย Figure 1. ซึ่งข้อมูลที่ได้ถูกนำไปพัฒนาโมเดล และอาศัยการศึกษาปัจจัยที่ส่งผลต่อการทำนาย เพื่อใช้ในการทดลองปรับปรุงประสิทธิภาพของโมเดลผ่านกระบวนการปรับตัวแปรต้น (Feature Selection) ดำเนินการเลือกใช้เทคนิคการคัดเลือกโมเดล (Model Selection) จากโมเดลต้นไม้ตัดสินใจแบบถดถอย แรนดอมฟอเรสท์ เกรเดียนบูสทรี และสมการถดถอยเชิงเส้น (Soltani et al., 2022; Burnham & Anderson, 2002) สำหรับประเมินราคาที่ดินกรณีศึกษาเขตอำเภอเมือง จังหวัดขอนแก่นและวัดผลโมเดลโดย ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง และค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวน

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

2. วัตถุประสงค์งานวิจัย (Research Objectives)

1. พัฒนาโมเดลการประเมินราคาที่ดินในเขตอำเภอเมืองจังหวัดขอนแก่น
2. เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อราคาที่ดินโดยใช้การเรียนรู้ของเครื่อง

3. กรอบแนวคิดงานวิจัย (Conceptual Framework)

งานวิจัยนี้กำหนดกรอบแนวคิดในการทำงาวิจัยและดำเนินการตามแนวคิดทฤษฎี ดัง Figure 1.

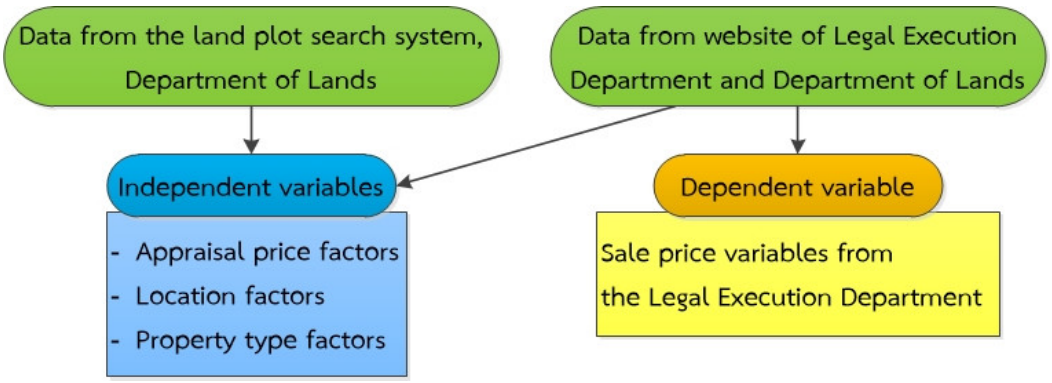

Figure 1. Conceptual Framework

4. การทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง (Literature Review)

Soltani et al. (2022) ได้นำเสนอโมเดลเรียนรู้ของเครื่องทั้งหมด 4 รูปแบบเพื่อศึกษาผลของลักษณะต่าง ๆ เช่น คุณลักษณะของทรัพย์สินและคุณภาพของย่านบริเวณที่อยู่ต่อการแปรผันของราคาที่พักอาศัยในพื้นที่ภูมิศาสตร์ที่แตกต่าง โดยใช้ชุดข้อมูลราคาที่พักอาศัยใน Metropolitan Adelaide, ออสเตรเลีย ในระยะเวลา 32 ปี (1984 - 2016) การวิจัยนี้ขึ้นอยู่กับข้อมูลการทำธุรกรรมการขายที่มีจำนวนมากถึง 428,000 รายการและตัวแปรอธิบาย 38 ตัวแปร พบว่าโมเดลที่ใช้ต้นไม้ตัดสินใจแบบถดถอย มีประสิทธิภาพมากกว่าโมเดลเชิงเส้น นอกจากนี้ เทคนิคการเรียนรู้แบบกลุ่ม เช่น แรนดอมฟอเรสท์ และเกรเดียนบูสทรี มีประสิทธิภาพมากกว่าในการทำนายราคาที่พักอาศัยในอนาคต โดยได้เพิ่มตัวแปรช่วงเวลาและพื้นที่ (Spatio-Temporal Lag) เพื่อเพิ่มความแม่นยำในการทำนายของโมเดล การวิจัยนี้ชี้ให้เห็นว่าตัวแปรช่วงเวลาและพื้นที่ (หรือตัวบ่งชี้ของพื้นที่ เวลาที่คล้ายกัน) เป็นปัจจัยที่สำคัญในการทำนายราคาที่พักอาศัยในอนาคต

Kumpu & Piyathamronchai (2024) ได้พัฒนาแบบจำลองการประเมินราคาที่ดินในอำเภอเมืองเชียงใหม่ จังหวัดเชียงใหม่ โดยใช้ข้อมูลการซื้อขายและจดทะเบียนสิทธิและนิติกรรมจากสำนักงานที่ดินจังหวัดเชียงใหม่ระหว่างเดือนพฤษภาคม พ.ศ. 2560 ถึง พฤศจิกายน พ.ศ. 2565 จำนวน 2,821 รายการ และใช้แบบสอบถามออนไลน์ (Google Forms) จำนวน 300 ราย การวิเคราะห์ข้อมูลใช้สถิติเชิงพรรณนาและสถิติเชิงอนุมาน รวมถึงการถดถอยเชิงเส้นพหุคูณแบบเป็นขั้นตอน (Stepwise Regression) พบว่าปัจจัยที่มีอิทธิพลต่อราคาประเมินที่ดินในทิศทางบวก ได้แก่ มูลค่าถนน ความกว้างของแปลงที่ดิน ระยะห่างจากแหล่งมลภาวะ และความครบถ้วนของสาธารณูปโภค ส่วนปัจจัยที่มีอิทธิพลในทิศทางลบ ได้แก่ เนื้อที่แปลงที่ดิน ผลการศึกษาพบว่าแบบจำลองสามารถอธิบายราคาที่ดินได้เพียงร้อยละ 56.9 ซึ่งชี้ให้เห็นว่ายังมีปัจจัยอื่น ๆ ที่ไม่ได้ถูกนำมาพิจารณา ซึ่งควรมีการศึกษาปัจจัยเพิ่มเติม เช่น ปัจจัยทางเศรษฐกิจและการเก็งกำไร เพื่อเพิ่มความแม่นยำของแบบจำลอง นอกจากนี้ยังมีข้อจำกัดเรื่องช่วงเวลาของข้อมูลที่อาจส่งผลต่อความแม่นยำในการวิเคราะห์และประมวลผล

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

4.1 แฮเวอร์ซีน (Haversine)

การคำนวณระยะทางวงกลม ระหว่างตัวอย่างในเซ็ต X และ Y บนพื้นผิวของทรงกลม พิกัดแรกของแต่ละจุดถือเป็นละติจูด และพิกัดที่สองถือเป็นลองจิจูด และมีหน่วยเป็นเรเดียน (Pedregosa et al., 2012)

$$D_{(x,y)} = 2\arcsin\left[\sqrt{\sin^2\left(\frac{x1-y1}{2}\right) + \cos(x_1)\cos(y_1)\sin^2\left(\frac{x_2-y_2}{2}\right)}\right] \quad (1)$$

4.2 อิทธิพลของคุณลักษณะ

เป็นเทคนิคที่ใช้ในการระบุว่าแต่ละคุณลักษณะ (Feature) มีความสำคัญต่อผลลัพธ์ของโมเดลมากน้อยแค่ไหน การประเมินความสำคัญของคุณลักษณะช่วยให้เข้าใจว่าโมเดลให้ความสำคัญกับข้อมูลส่วนใดมากที่สุด ซึ่งสามารถใช้ในการปรับปรุงโมเดลหรือการตัดสินใจทางธุรกิจ ตัวอย่างเทคนิคที่ใช้ในการประเมินความสำคัญของคุณลักษณะได้แก่ การใช้ค่าสัมประสิทธ์ (Coefficient) ของโมเดลเชิงเส้น (Linear) การใช้ค่าความสำคัญของคุณลักษณะในต้นไม้ตัดสินใจ หรือการใช้ค่าแชบ (SHAP values) (Vanderplas, 2017; Lundberg & Lee, 2017)

4.3 ปัญหาสมการถดถอย (Regression Problem)

ปัญหาการถดถอย (Na Bangchang, 2011) เป็นกระบวนการในการสร้างแบบจำลองทางสถิติหรือการเรียนรู้ของเครื่องเพื่อทำนายค่าตัวแปรตาม (Dependent Variable) ที่เป็นค่าต่อเนื่อง (Continuous Variable) จากค่าตัวแปรอิสระ (Indepe-ndent Variables) ที่เป็นตัวแปรเชิงอธิบายหรือคุณลักษ ณะต่าง ๆ (Features) โดยมีเป้าหมายในการหาเส้นหรือฟังก์ชันที่สามารถอธิบายความสัมพันธ์ระหว่างตัวแปรเหล่านี้ได้อย่างดีที่สุด ซึ่งแบบจำลองที่ใช้แก้ปัญหาประเภทนี้ ได้แก่

1. การวิเคราะห์การถดถอยเชิงเส้น

เป็นการนำเอาข้อมูลหรือตัวแปรมาหาความสัมพันธ์กันโดยเมื่อมีตัวแปรอิสระเพียงตัวเดียวเรียกว่า การถดถอยอย่างง่าย (Simple Linear Regression) เมื่อมีตัวแปรอิสระมากกว่า 1 ตัวแปรเรียกว่าการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression) (Na Bangchang, 2011) โดยมีรูปแบบสมการในการวิเคราะห์ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + e \quad (2)$$

กำหนดให้

Y	คือ ตัวแปรเกณฑ์
$X_1 X_2 \ldots X_n$	คือ ค่าของตัวแปรอิสระแต่ละตัว
β_0	คือ ส่วนตัดแกน Y
$\beta_2 \ldots \beta_n$	คือ ค่าสัมประสิทธิ์ความถดถอย
e	คือ ค่าความคลาดเคลื่อน (Error Residual)

2. ต้นไม้ตัดสินใจแบบถดถอย (Regression Trees)

เป็นการสร้างเงื่อนไข If-else ขึ้นมาจากข้อมูลในตัวแปรเพื่อแบ่งข้อมูลเป็นกลุ่มใหม่ที่อธิบายเป้าหมาย (Target) ได้ดีที่สุดในต้นไม้ตัดสินใจเราใช้ฟังก์ชันวัตถุประสงค์ (Objective Function) ที่เหมาะสมเพื่อกำหนดแยก (Split) ในแต่ละตัวแปร โดยแต่ละประเภทของ ต้นไม้ตัดสินใจมีฟังก์ชันวัตถุประสงค์ต่างกัน เช่น จีนีอิมพอริที (Gini Impurity) หรือเอนทรอพี (Entropy) สำหรับต้นไม้ตัดสินใจแบบจำแนก (Classification Tree) และค่าผลรวมของผลต่างกำลังสอง (Residual

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

sum of squares : RSS) สำหรับต้นไม้ตัดสินใจแบบถดถอย (Regression Tree) ซึ่งจะช่วยในการแบ่งข้อมูลให้อยู่ในกลุ่มที่เหมาะสมสำหรับการอธิบายตัวแปรเป้าหมาย (Target Variable) (Vanderplas, 2017)

$$e_i = y_i - \hat{y}_i \tag{3}$$

$$RSS = e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2 \tag{4}$$

Residual (Error Term, e_i) คือ ค่าความคลาดเคลื่อนหรือค่าผิดพลาด (Error) ระหว่าง y ทุก ๆ จุดในข้อมูล กับ $\hat{y}$ ที่ได้มาจากการประมาณค่าการทำนาย (Prediction) ขึ้นมาการคำนวณค่าความคลาดเคลื่อนของข้อมูลตัวที่ i

3. แรนดอมฟอเรสท์ (Random Forest)

เป็นแบบจำลองที่ถูกพัฒนาขึ้นจากต้นไม้ตัดสินใจ โดยที่แรนดอมฟอเรสท์จะทำการเพิ่มจำนวนต้น ไม้เป็นหลาย ๆ ต้นโดยใช้กระบวนการสุ่มข้อมูล (Bootstrap Sampling) และสุ่มคุณลักษณะ ในแต่ละโหนดของต้นไม้เพื่อลดความเกี่ยวข้องระหว่างต้นไม้แต่ละต้น และทำนายด้วยการโหวตหรือหาค่าเฉลี่ย (Vanderplas, 2017)

4. เกรเดียนบูสทรี (Gradient boosted Trees)

เป็นแบบจำลองที่ถูกพัฒนาขึ้นจากต้นไม้ตัดสินใจโดยสร้างต้นไม้ทีละต้นโดยทำนายค่าผลลัพธ์และคำนวณค่าความผิดพลาด ระหว่างค่าทำนายกับค่าจริงจากนั้นในทุกรอบจะทำการสร้างต้นไม้เพิ่มเติมเพื่อลดความผิดพลาดที่เหลือ (Vanderplas, 2017)

4.4 การปรับแต่งไฮเปอร์พารามิเตอร์ และการคัดเลือกโมเดล

1. การปรับแต่งไฮเปอร์พารามิเตอร์ (Hyper-Parameter Tuning)

กระบวนการการปรับแต่งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลเครื่องจักรอัลกอริทึมเพื่อให้โมเดลมีประสิทธิภาพสูงสุด หรือให้ผลลัพธ์ที่ดีที่สุดที่เป็นไปได้สำหรับงานที่กำลังดำเนินการอยู่ (Vanderplas, 2017)

2. รูปแบบของการประเมินผล (Model Evaluation)

เป็นกระบวนการที่ใช้เพื่อวัดประสิทธิภาพของโมเดลที่ได้ฝึกและทดสอบด้วยข้อมูลที่ไม่เคยเห็นในกระบวนการฝึก (Unseen Data) เพื่อประเมินความสามารถในการทำนายหรือจำแนกข้อมูลใหม่ที่ไม่เคยเห็นมาก่อน กระบวนการนี้เป็นส่วนสำคัญของการพัฒนาและใช้งานแบบจำลอง (Vanderplas, 2017)

3. การทดสอบแบบไขว้ทบ (Cross Validation)

เป็นเทคนิคที่ใช้ในการประเมินความสามารถของโมเดลและช่วยในการป้องกันปัญหาโอเวอร์ฟิต (Overfitting) โดยทั่วไปจะใช้วิธีการแบ่งข้อมูลออกเป็นหลายส่วน (Fold) ซึ่งเรียกว่าการทดสอบแบบไขว้ทบแบบ K ส่วน (K-Fold Cross-Validation) (Kohavi, 1995) วิธีการนี้จะทำการแบ่งข้อมูลเป็น K ส่วน จากนั้นทำการฝึกโมเดลด้วยข้อมูล K-1 ส่วน และใช้ส่วนที่เหลือในการทดสอบโมเดล ทำแบบนี้ K รอบ เพื่อให้มั่นใจว่าโมเดลได้รับการทดสอบกับข้อมูลทั้งหมด กระบวนการนี้ช่วยให้เราสามารถประเมินประสิทธิภาพของโมเดลได้อย่างแม่นยำยิ่งขึ้น

4. การคัดเลือกโมเดล

เป็นกระบวนการเลือกโมเดลที่ดีที่สุดจากชุดของโมเดลที่ใช้ทำนายหรือจัดกลุ่มข้อมูล กระบวนการนี้รวมถึงการเปรียบเทียบประสิทธิภาพของโมเดลต่างๆ โดยใช้ metrics เช่น ความแม่นยำ (Accuracy) ค่า F1 score ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เป็นต้น นอกจากนี้ยังรวมถึงการเลือกค่าพารามิเตอร์ที่เหมาะสม

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

สำหรับโมเดลเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด รวมไปถึงการใช้การทดสอบแบบไขว้ทบก็เป็นส่วนสำคัญในการเลือกโมเดลที่ดีที่สุด (Burnham & Anderson, 2002)

5. การค้นหาพารามิเตอร์แบบกริด (Grid Search for Hyperparameter Tuning)

การค้นหาพารามิเตอร์แบบกริด (Bergstra & Bengio, 2012) เป็นวิธีการปรับแต่งค่าพารามิเตอร์ของโมเดล โดยการค้นหาชุดค่าที่เหมาะสมที่สุดในพื้นที่ของพารามิเตอร์ที่กำหนดไว้ล่วงหน้า วิธีการนี้จะทำการทดสอบค่าพารามิเตอร์ต่างๆ ที่เป็นไปได้ทั้งหมดอย่างเป็นระบบและครอบคลุมทุกค่าในกริดที่กำหนด จากนั้นจะประเมินประสิทธิภาพของแต่ละชุดค่าพารามิเตอร์ด้วยการใช้ การทดสอบแบบไขว้ทบเพื่อหาโมเดลที่มีประสิทธิภาพสูงที่สุด กระบวนการนี้ช่วยเพิ่มความแม่นยำและความสามารถในการทำนายของโมเดล

6. กระบวนการคัดเลือกคุณลักษณะ

กระบวนการเลือกคุณลักษณะมีความสำคัญและเป็นประโยชน์ในการสร้างโมเดลจากชุดคุณลักษณะที่มีอยู่หลายๆ ตัว โดยมีเป้าหมายเพื่อเพิ่มประสิทธิภาพของโมเดล ลดความซับซ้อน และลดความเสี่ยงของการโอเวอร์ฟิต กระบวนการนี้ช่วยให้โมเดลสามารถทำงานได้เร็วขึ้นและแม่นยำมากขึ้นมีหลายวิธีที่ใช้ในการเลือกคุณลักษณะเช่น

ฟิวเตอร์เมธอด (Filter Methods) วิธีการนี้จะเลือกคุณลักษณะโดยพิจารณาจากคุณสมบัติทางสถิติ เช่น การใช้ค่าสหสัมพันธ์ (Correlation) ค่าสถิติไคสแคว์ (Chi-Square Test) หรือ สารสนเทศร่วม (Mutual Information) โดยวิธีนี้ไม่ต้องใช้โมเดลในการเลือกคุณลักษณะ

วิธีแรปเปอร (Wrapper Methods) วิธีการนี้จะเลือกคุณลักษณะโดยใช้โมเดลเป็นส่วนหนึ่งของกระบวนการเลือก เช่น การกำจัดคุณลักษณะด้วยกระบวน
การเวียนเกิด (Recursive Feature Elimination: RFE) ที่ทำการลบคุณลักษณะที่มีความสำคัญน้อยที่สุดทีละตัว จนกว่าจะได้ชุดคุณลักษณะที่ดีที่สุด

วิธีฝังตัว (Embedded Methods) วิธีการนี้จะเลือกคุณลักษณะในขณะที่สร้างโมเดล เช่น การใช้สมการถดถอยแบบลาสโซ่ (Lasso Regression) ที่สามารถทำให้ค่าของสัมประสิทธิ์ของคุณลักษณะที่ไม่สำคัญเป็นศูนย์ (Vanderplas, 2017) หรือการใช้คุณลักษณะสำคัญของ ของกระบวนการต้นไม้ตัดสินใจแบบถดถอย และแรนดอมฟอเรสท์

การเลือกคุณลักษณะที่เหมาะสม (Guyon & Elisseeff, 2003; Chandrashekar & Sahin, 2014; Choompol, 2019) ช่วยเพิ่มประ สิทธิภาพของโมเดล ลดเวลาในการประมวลผล และทำให้การวิเคราะห์ผลลัพธ์ของโมเดลง่ายขึ้น

5. วิธีดำเนินงานวิจัย (Research Methodology)

งานวิจัยนี้ได้ทำการการรวบรวมข้อมูล (Data Collection) ข้อมูลรายงานผลการขายทอดตลาด ข้อมูลค้นหาทรัพย์ประกาศขายทอดตลาด และข้อมูลระบบค้นหารูปแปลงที่ดิน ระบบค้นหารูปแปลงที่ดินของกรมที่ดินจากนั้นทำการสร้างคุณลักษณะใหม่ (Feature Construction) ตัวแปรระยะทางจากสถานที่สำคัญ 14 สถานที่หน่วยเมตร (แฮเวอร์ซีน) ตัวแปรราคาประเมินจากค่าเฉลี่ย 5 สถานที่ใกล้เคียง ตัวแปรขนาดพื้นที่หน่วยตารางวาและประเภททรัพย์

ต่อมาทำการวิเคราะห์ข้อมูลเชิงการตรวจสอบและกระบวนการเตรียมข้อมูล (EDA & Data Pre Processing) โดยทำการนำเสนอแผนภาพข้อมูล (Data Visualization) การจัดการข้อมูลที่หายไป (Missing Value Handling) การตรวจสอบค่าผิดปกติ (Outlier Detection) และการเข้ารหัสข้อมูลแบบหมวดหมู่ (Encoding Categorical Data) จากนั้นนำข้อมูลมาทำการเรียนรู้ของเครื่องโดยใช้ 4 โมเดลดังนี้ ต้นไม้ตัดสินใจ แรนดอมฟอเรสท์ เกรเดียนบูสทรีและการถดถอยเชิงเส้นโดยมีการใช้การค้นหาแบบกริด เป็นการค้นหาค่าพารามิเตอร์โดยการกำหนดค่าที่เป็นไปได้ทั้งหมดของแต่ละพารามิเตอร์ที่สนใจในรูปแบบของรายการที่กำหนดล่วงหน้า และสุดท้ายรูปแบบของการประเมินผล (Models Evaluation)

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

โดยใช้ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง และค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวนในการวัดและทดสอบประสิทธิภาพของแบบจำลอง

5.1 การเก็บรวบรวมข้อมูล (Data Collection)

1. พื้นที่กรณีศึกษาราคาขายจริง 193 แห่ง

ดำเนินการเก็บรวบรวมข้อมูลจากเว็บไซต์กรมบังคับคดีและกรมที่ดินโดยที่ดินจากกรมบังคับคดีในพื้นที่เขตอำเภอเมืองจังหวัดขอนแก่นระหว่างปี 2560 ถึง 2566 ข้อมูลใน 1 ระเบียนแสดงข้อมูลตัวอย่างตาม Table 1. โดยข้อจำกัดของชุดข้อมูลจากเว็บไซต์กรมบังคับคดีเมื่อทำการขายเสร็จสิ้นข้อมูลจะถูกลบออกไปบางส่วน ซึ่งการเก็บข้อมูลย้อนหลังอาจทำให้ข้อมูลการจำนองติดไปของที่ดินอาจสูญหายโดยในงานวิจัยนี้ไม่ได้นำข้อมูลการติดจำนองมาเป็นตัวแปรต้น

Table 1. The case study of land detail of 193 locations.

Data	Example
Case number	ผบ.1003
Property type	Land with Structures
Rai (ไร่)	-
Ngan (งาน)	-
Square Wah (ตารางวา)	73.4
Appraised Value	3,196,000.00
Sub-district	Banped
District	Mueang Khon Kaen
Province	Khon Kaen
Title Deed Land	254492
Achievable Selling Price/ Highest Bid	3,350,000
Cadastral Map	5541 I 6414-05 (1000)

Table 2. The color setting of case study of land detail of 193 locations.

	The selling price range per square wah		Average	Color	Number
Group	lower bound	upper bound	selling price		
1	150000	300000	205675	White	7
2	100000	149999	131298	Red	4
3	50000	99999	63107	Blue	9
4	10000	49999	25579	Green	66
5	5000	9999	7114	Orange	21
6	1000	4999	2425	Purple	54
7	300	999	589	Pink	28
8	1	299	180	Brown	4

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

จาก Table 2. กำหนดสีและสัญลักษณ์สำหรับกลุ่มพล็อตกระจายเพื่อดูการแบ่งกลุ่มตามราคาขายต่อตารางวา โดยกำหนดกลุ่มของราคาประเมินเป็น 8 กลุ่มตามตัวเลขเริ่มต้นและตัวเลขสิ้นสุดของราคา

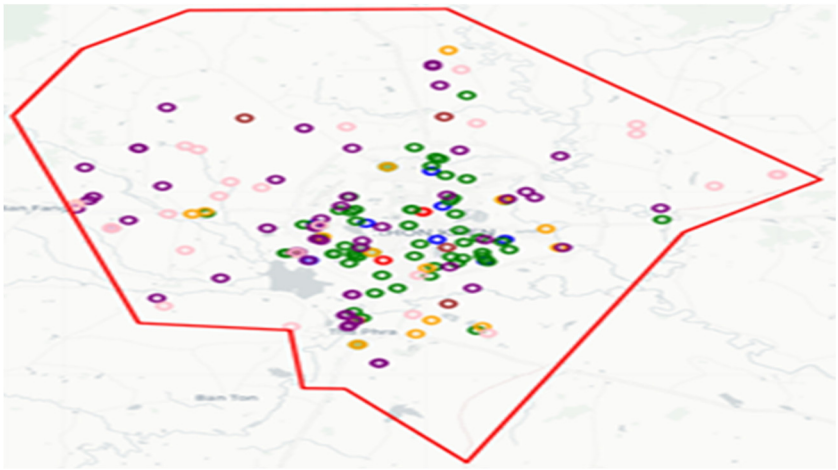

Figure 2. ตำแหน่งการกระจายตัว ของข้อมูลราคาขายต่อตารางวา ซึ่งเป็นข้อมูลราคาขายจริงจากเว็บไซต์กรมบังคับคดี

จาก Figure 2. แสดงข้อมูลตำแหน่งบนแผนที่เพื่อแสดงข้อมูลที่คัดแยกตามกลุ่มราคาขายต่อตารางวาที่กำหนดกลุ่มของราคาประเมินเป็น 8 กลุ่มตามตัวเลขเริ่มต้นและตัวเลขสิ้นสุดของราคา และเส้นสีแดงแสดงถึงกรอบบริเวณอำเภอเมืองขอนแก่น

2. พื้นที่กรณีศึกษาราคาประเมิน 1500 แห่ง

ผู้วิจัยได้ทำการสุ่มข้อมูลราคาประเมินที่ดินตามกลุ่มราคาที่แสดงใน Table 3. จากระบบค้นหารูปแปลงที่ดิน (Landsmaps) กรมที่ดินจำนวน 1500 แห่ง ซึ่งใน 1 ระเบียน ประกอบด้วย

(1) ละติจูดและลองติจูดจากที่ดินโดยการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Sampling) โดยกำหนดกลุ่มของราคาประเมินเป็น 6 กลุ่มตามตัวเลขเริ่มต้นและตัวเลขสิ้นสุด (Range) ของราคา

(2) ข้อมูลราคาประเมินบาทต่อตารางวาจาก Table 3. กำหนดสีและสัญลักษณ์สำหรับกลุ่มพล็อตกระจายเพื่อดูการแบ่งกลุ่มตามราคาต่อตารางวาจากกรณีศึกษาที่ดิน 1,500 แห่งที่มีการสุ่มตัวอย่างแบบชั้นภูมิโดยกำหนดกลุ่มของราคาประเมินเป็น 6 กลุ่มตามตัวเลขเริ่มต้นและตัวเลขสิ้นสุดของราคา

Table 3. The color setting of case study of land detail of 1,500 locations.

Group	The selling price range per square wah		Average selling price	Color	Number
	lower bound	upper bound			
1	100000	150000	122736	Pink	250
2	50000	99999	63088	Red	250
3	10000	49999	23168	Blue	250
4	5000	9999	6940	Green	250
5	1000	4999	2086	Orange	250
6	300	999	579	Purple	250

Citation: Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

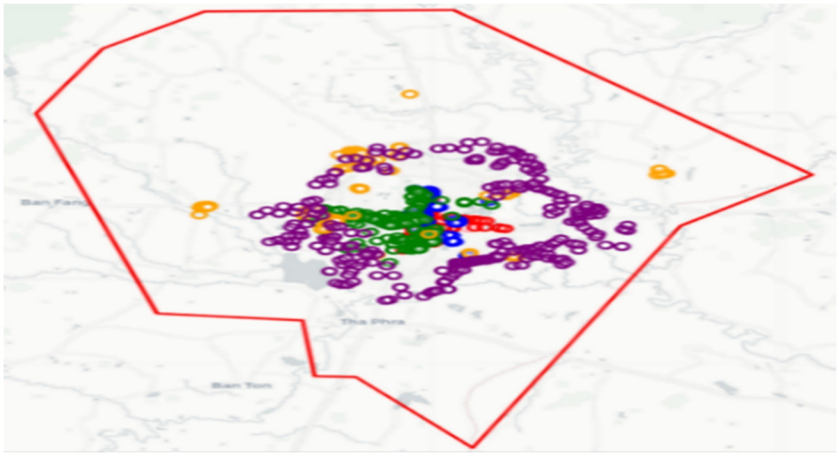

Figure 3. Scatter plot to observe clustering based on the price per square wah (appraisal price) in a case study of 1,500 locations.

จาก Figure 3. แสดงข้อมูลตำแหน่งบนแผนที่เพื่อแสดงข้อมูลที่คัดแยกตามกลุ่มราคาต่อตารางวาที่กำหนดกลุ่มของราคาประเมินเป็น 6 กลุ่ม และเส้นสีแดงล้อมบริเวณอำเภอเมืองขอนแก่น

5.2 การสร้างคุณลักษณะ (Feature Construction)

1. สกัดคุณลักษณะสำคัญจากระยะทาง โดยใช้แฮเวอร์ซีน ซึ่งเป็นระยะห่างเชิงมุมระหว่างจุดสองจุดบนพื้นผิวของทรงกลมละติจูดลองติจูด จากที่ดินที่สนใจไปหาสถานที่สำคัญ 14 สถานที่ โดยแบ่งเป็น สวนสาธารณะ โรงแรม ห้างสรรพสินค้า พิพิธภัณฑ์ สนามบิน โรงพยาบาล สถาบันการศึกษา ตลาด และศาสนสถานใช้หน่วยระยะทางเป็นกิโลเมตร

2. สกัดราคาประเมินจากค่าเฉลี่ย 5 สถานที่ใกล้เคียงผู้จัดทำคำนวณระยะห่างระหว่างพิกัดที่ดินของกรณีศึกษาที่ดิน 1,500 แห่ง ในอำเภอเมือง จังหวัดขอนแก่นแต่ละแถวกับพิกัดเป้าหมายของกรณีศึกษาที่ดิน 193 แห่ง ในอำเภอเมือง จังหวัดขอนแก่น เรียงลำดับตามระยะห่างเพื่อหาตำแหน่งที่ใกล้ที่สุด 5 ตำแหน่งแล้วนำ 5 ตำแหน่งที่หามานำมาคำนวนหาค่าเฉลี่ยของราคาประเมินที่ดิน

3. นำคอลัมน์ไร่ งาน ตารางวา มาคำนวนเปลี่ยนเป็นคอลัมน์ขนาดพื้นที่ต่อตารางวาโดย 1 ไร่ เท่ากับ 4 งาน และ 1 งาน เท่ากับ 100 ตารางวา

4. นำคอลัมน์ราคาประเมินและราคาขายมาหารด้วยขนาดพื้นที่ เปลี่ยนเป็นคอลัมน์ราคาประเมินต่อตารางวาและราคาขายต่อตารางวา

5.3 กระบวนการเตรียมข้อมูลและการสำรวจข้อมูลเบื้องต้น

1. กระบวนการเตรียมข้อมูล มุ่งเน้นการเตรียมข้อมูลให้อยู่ในสภาพพร้อมใช้ในการวิเคราะห์หรือการเรียนรู้เชิงข้อมูล โดยเปลี่ยนแปลงและปรับปรุงข้อมูลตามความจำเป็น เพื่อเพิ่มคุณภาพและความน่าเชื่อถือของข้อมูล และเพิ่มประสิทธิภาพในกระบวนการเรียนรู้ของเครื่อง

(1) ลบคอลัมน์ประเภททรัพย์ที่เป็นห้องชุดออก

(2) ลบแถวที่มีค่าราคาขายได้/ราคาเสนอสูงสุดน้อยกว่า 0

(3) เปลี่ยนประเภทของคอลัมน์ ราคาประเมิน ราคาขายได้/ราคาเสนอสูงสุด ไร่ งาน ตารางวา

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

(4) นำคอลัมน์ประเภททรัพย์มาทำการเข้ารหัสแบบวัน-ฮอท (One-Hot Encoding) วิธีนี้ทำให้แต่ละค่าของข้อมูลประเภทกลายเป็นตัวแปรใหม่ที่มีค่า 0 หรือ 1 เพื่อระบุว่าข้อมูลอยู่ในหมวดหมู่ใดโดยในคอลัมน์ประเภททรัพย์จะมีอยู่ 2 ประเภทคือที่ดินพร้อมสิ่งปลูกสร้างและที่ดินว่างเปล่า

2. การสำรวจข้อมูลเบื้องต้น ดำเนินการศึกษาการกระจายตัวของราคาขายต่อตารางวา โดยพบว่าว่าข้อมูลส่วนมากจะอยู่ในช่วง 1,812.95 บาทถึง 25,416.67 บาทและมีค่าข้อมูลผิดปกติ (Outliers) อยู่ 12 จุดในชุดข้อมูลที่มีค่ามากกว่า 60,355.03 บาทของขอบเขตสูงสุด (Upper Fence) จากนั้นทำการเปรียบเทียบจำนวนของประเภทที่ดิน โดยประเภททรัพย์ที่มี 2 ประเภทคือที่ดินพร้อมสิ่งปลูกสร้างและที่ดินว่างเปล่า ข้อมูลทั้งหมด 193 รายการมีที่ดินพร้อมสิ่งปลูกสร้าง 127 รายการมากกว่าที่ดินว่างเปล่าที่มีเพียง 66 รายการ นอกจากนี้เมื่อศึกษาความสัมพันธ์ระหว่างขนาดพื้นที่ที่แสดงบนแกน X และราคาขายได้/ราคาเสนอสูงสุดที่แสดงบนแกน Y ดัง Figure 4.

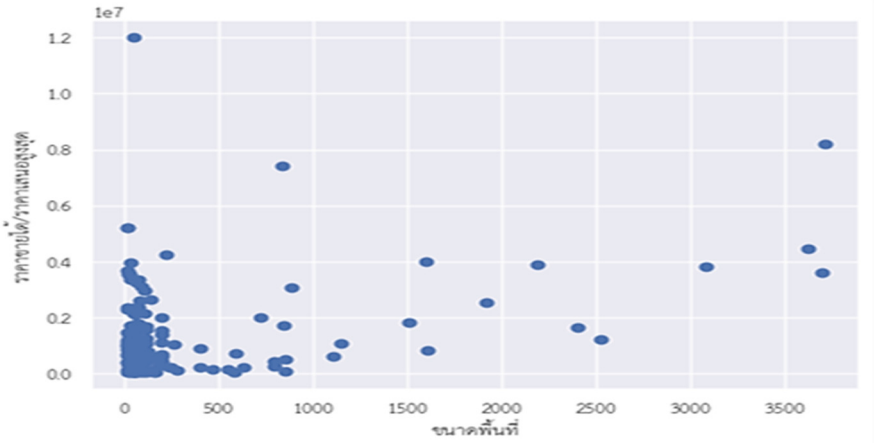

Figure 4. Relationship between area size and achievable selling price/ highest bid

จาก Figure 4. แสดงความสัมพันธ์ระหว่างขนาดพื้นที่ที่แสดงบนแกน X และราคาขายได้/ราคาเสนอสูงสุดที่แสดงบนแกน Y ในแผนภูมิจุดแบบกระจัดกระจายนี้ข้อมูลกรณีศึกษาที่ดิน 193 แห่งที่ขายได้ส่วนมากจะมีขนาดพื้นที่ประมาณไม่เกิน 400 ตารางวา มีราคาขายได้/ราคาเสนอสูงสุดราคา 276 บาทถึงราคา 12,000,000 บาทและขนาดพื้นที่มากกว่า 400 ตารางวาถึงขนาดพื้นที่ไม่เกิน 4,000ตารางวาจะมีราคาขายได้/ราคาเสนอสูงสุดไม่เกิน ตารางวาจะมีราคาขายได้/ราคาเสนอสูงสุดไม่เกิน 9,000,000 บาท

5.4 การเรียนรู้ของเครื่อง

ในการวิจัยนี้ได้พัฒนาตัวแบบด้วยการใช้ภาษาไพทอนและไลบรารีหลักเช่น Scikit Learn เพื่อช่วยในการประเมินและพัฒนาประสิทธิภาพของตัวแบบ ข้อมูลถูกแบ่งออกเป็นสองส่วนโดย 80% ใช้สำหรับการฝึกสอนและ 20% เป็นข้อมูลทดสอบ สำหรับแบบจำลองที่ใช้ในการทดลอง ได้แก่ การถดถอยเชิงเส้น, ต้นไม้ถดถอย, แรนดอมฟอเรสท์ และเกรเดียนบูสทรี งานวิจัยนี้ได้ใช้วิธีการทดสอบแบบไขว้ทบ แบบ 5 โฟล (5-Fold Cross Validation) จากชุดข้อมูลฝึกสอน เพื่อค้นหาแบบจำลองที่มีประ สิทธิภาพสูงสุด แบบจำลองเหล่านี้จะถูกปรับแต่งพารามิ เตอร์โดยการใช้การค้นหาแบบกริด ซึ่งทำการทดลองกับทุกค่าพารามิเตอร์ที่กำหนดไว้ล่วงหน้า ดังแสดงใน Table 4. จากนั้นโมเดลที่ได้จากแต่ละอัลกอริทึมจะถูกนำไปทดสอบกับชุดข้อมูลทดสอบ 20% เพื่อประเมินประสิทธิ ภาพของตัวแบบในการทำนายข้อมูลจริง

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

Table 4. Parameters of the model using grid search.

Models	Parameters
Regression Trees	max_depth: [None, 10, 20, 30], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4]
Random Forest	n_estimators: [100, 200, 300], max_depth: [None, 10, 20, 30], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4]
Gradient boosted Trees	n_estimators: [100, 200, 300], learning_rate: [0.01, 0.1, 0.2] max_depth: [3, 4, 5], min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4]

ในการศึกษาปัจจัยที่ส่งผลต่อ ราคาขายต่อตารางวา งานวิจัยนี้ได้ทำการทดลองโดย ออกแบบการทดลองด้วยการจัดกลุ่มตัวแปรต้นเป็น 5 แบบการทดลอง ตามวงรอบของการปรับปรุงประสิทธิภาพเพื่อทำนายตัวแปรตาม ราคาขายต่อตารางวาโดยตัวแปรต้นในแต่ละการทดลองได้แสดงไว้ใน Table 5. ราคาประเมินต่อตารางวา ขนาดพื้นที่ ประเภททรัพย์ราคาประเมินจากห้าตำแหน่ง และกลุ่มสถานที่สำคัญซึ่งจะมี 14 สถานที่โดยแบ่งเป็น สวนสาธารณะ โรงแรม ห้างสรรพ สินค้า พิพิธภัณฑ์ สนามบิน โรงพยาบาล สถาบันการศึกษา ตลาด และศาสนสถานใช้หน่วยระยะทางเป็นกิโลเมตร

Table 5. The description of independent variable for each experimental.

Experiment	The appraisal price per square wah	Area size	Property type	Important location group	The appraisal price from 5 locations
1	✓	✓	✓	✓	
2	✓				
3	✓	✓	✓	✓	✓
4		✓	✓	✓	✓
5		✓	✓	✓	

5.5 การประเมินประสิทธิภาพ

เนื่องจากค่าข้อผิดพลาดที่ได้จะมีทั้งค่าบวกและค่าลบตรงจุดนี้เอง จึงทำให้ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง และค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวน เข้ามามีบทบาทสำคัญ เนื่องจากทั้ง 3 วิธีจะมีการทำให้ค่าข้อผิดพลาดกลายเป็นค่าบวกก่อนเสมอ ก่อนที่จะนำค่าข้อผิดพลาดมารวมกันและหาค่าเฉลี่ยเพื่อประเมินประสิทธิภาพของโมเดลได้

6. ผลการวิจัย (Results)

จากวิธีการดำเนินงานข้างต้นสามารถสรุปผลการดำเนินงานในแต่ละการทดลองได้ดังนี้

Citation: Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

Table 6. Results of the experiment.

Experiment/ Best Model	Hyper parameter	R^2	MAE	RMSE
Experiment 1 Gradient Boosted Trees	LR: 0.01 MD: 3 MSL: 2 MSS: 10 NE: 200	0.58	10911	22252
Experiment 2 Regression Tree	MD:None MSL: 2 MSS: 2	0.73	8908	17822
Experiment 3 Regression Tree	MD: 10 MSL: 1 MSS: 2	0.65	10482	20207
Experiment 4 Gradient Boosted Trees	LR: 0.2 MD: 3 MSL: 1 MSS: 10 NE: 100	0.80	7929	15281
Experiment 5 Gradient Boosted Trees	LR: 0.01 MD: 3 MSL: 1 MSS: 10 NE: 300	0.74	10807	17400

Table 6. แสดงผลลัพธ์ที่ได้จากการทำโมเดลซีเลคชั่น ของการค้นหาค่าของไฮเปอร์พารามิเตอร์ของโมเดลแบบกริด ซึ่งจาก Table 6. สามารถอธิบายตัวย่อของตัวแปรต่าง ๆ ได้ดังนี้ (Pedregosa et al., 2012) พารามิเตอร์ (Hyperparameter) มีLR, MD, MSL, MSS, NE คือ อัตราการเรียนรู้ (Learning Rate, LR) กำหนดค่าที่ใช้ในกระบวนการฝึกสอนของเกรเดียนบูสทรี ค่าที่เลือกจะมีผลต่อการ อัปเดตค่าของแต่ละต้นไม้ในเอนเซมเบิล (Ensemble) ในแต่ละรอบ ค่าความลึกของต้นไม้ (Max Depth, MD) กำหนดความลึกสูงสุดของต้นไม้ค่ามากมักทำให้ต้นไม้ซับซ้อนมากขึ้น แต่หากมีค่ามากเกินไปอาจเป็นการโอเวอฟิตติ้ง มินแซมเพิลลีฟ (Min Samples Leaf, MSL) กำหนดจำนวนตัวอย่างขั้นต่ำที่ต้องอยู่ในใบ (Leaf Node) ค่านี้ช่วยลดความซับซ้อนของต้นไม้และป้องกันการโอเวอฟิตติ้ง มินแซมเพิลสปริท (Min Samples Split, MSS) จำนวนตัวอย่างขั้นต่ำในการแยกและ เอ็นเอสติเมท (N Estimators, NE) จำนวนต้นไม้ ตามลำดับ

การทดลองที่ 1 ได้ผลการคัดเลือกโมเดลโดยพบว่า เกรเดียนบูสทรีเป็นโมเดลที่มีประสิทธิภาพดีที่สุด และจากการวิเคราะห์ความสำคัญของคุณลักษณะโดยใช้ค่าสารสนเทศ (Information Gain) แสดงว่าคุณลักษณะ “ราคาประเมินต่อตารางวา” มีผลต่อการทำนายมากที่สุด รองลงมาจะเป็นกลุ่มสถานที่สำคัญขนาดพื้นที่ และประเภททรัพย์ที่ไม่มีผลต่อการทำนาย

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

การทดลองที่ 2 ได้เล็งเห็นถึงความสำคัญของคุณลักษณะ “ราคาประเมินต่อตารางวา” ที่มีผลต่อการทำนายมากที่สุดจากการวิเคราะห์ผลการทดลองที่ 1 ดังนั้นจึงนำเฉพาะส่วนราคาประเมินต่อตารางวามาทดสอบเพิ่มเติม โดยผลการคัดเลือกโมเดลพบว่า โมเดลต้นไม้ตัดสินใจแบบถดถอย เป็นโมเดลที่มีประสิทธิภาพดีที่สุด ซึ่งค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวนของโมเดลที่มีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองมีค่าที่น้อยลงเมื่อเทียบกับการทดลองที่ 1

การทดลองที่ 3 ได้เพิ่มตัวแปรราคาประเมินจากพื้นที่ใกล้เคียงห้าตำแหน่งเข้าไปในการทดลองที่ 1 เพื่อตรวจสอบว่าราคาประเมินจากห้าตำแหน่งมีผลต่อการทำนายมากขึ้นหรือไม่ ซึ่งผลการคัดเลือกโมเดลพบว่า ต้นไม้ตัดสินใจแบบถดถอยเป็นโมเดลที่มีประสิทธิภาพดีที่สุด และจากการวิเคราะห์ความสำคัญของคุณลักษณะ พบว่าว่าโมเดลให้เห็นว่า “ราคาประเมินต่อตารางวา” ยังคงมีผลต่อการทำนายมากที่สุด รองลงมาคือกลุ่มสถานที่สำคัญ และขนาดพื้นที่ ในขณะที่ราคาประเมินจากพื้นที่ใกล้เคียงห้าตำแหน่งและประเภททรัพย์ไม่มีผลต่อการทำนาย เมื่อเทียบกับการทดลองที่ 1 ผลลัพธ์ของการทดลองที่ 3 มีประสิทธิภาพดีกว่า แต่ยังคงด้อยกว่าโมเดลที่ได้จากการทดลองที่ 2

การทดลองครั้งที่ 4 และ 5 เป็นการนำราคาประเมินต่อตารางวาออกจากการทดลองที่ 3 และ 1 ตามลำดับเพื่อทดสอบว่าถ้าไม่ทราบราคาประเมินต่อตารางวาของที่ดินผืนนั้น ๆ จะสามารถทำการประเมินราคาที่ใกล้เคียงราคาขายจริงได้อยู่หรือไม่

การทดลองที่ 4 ได้โมเดลเกรเดียนบูสทรีเป็นโมเดลที่มีประสิทธิภาพดีที่สุด และจากการวิเคราะห์ความสำคัญของคุณลักษณะ พบว่าโมเดลให้ความสำ คัญกับคุณลักษณะกลุ่มสถานที่สำคัญมีผลต่อการทำนายมากที่สุด รองลงมาคือขนาดพื้นที่ ราคาประเมินจากห้าตำแหน่ง และประเภทที่ดิน นอกจากนี้สามารถศึกษาค่าประสิทธิภาพของโมเดลต่าง ๆ ในการคัดเลือกโมเดลที่ดีที่สุด ผลการวิจัย ดัง Table 7.

Table 7. Performance metrics of various models for selecting the best model from Experiment 4.

โมเดล	R^2	MAE	RMSE
Linear Regression	0.23	17378.70	30199.13
Regression Tree	0.21	13539.32	30596.87
Random Forest	0.72	10371.91	18180.34
GradientBoosting	0.74	10807.34	17400.46

จาก Table 7. พบว่า ค่าประสิทธิภาพของโมเดลต่าง ๆ ที่ได้จากการทำการค้นหาพารามิเตอร์แบบกริดของการทดลองที่ 4 โดยจากตารางพบว่า ค่าสัมประสิทธิ์การอธิบายความแปรปรวน (R^2) ของเกรเดียนบูสทรีสูงสุดที่ 0.7445 ซึ่งแสดงว่าโมเดลสามารถอธิบายความแปรปรวนในข้อมูลได้ประมาณ 74% ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ ของเกรเดียนบูสทรีอยู่ที่ 10807.34 ซึ่งสูงกว่าค่าของแรนดอมฟอเรสท์แต่ยังค่อนข้างต่ำ นอกจากนี้ความคลาดเคลื่อนเฉลี่ยรากที่สอง เกรเดียนบูสทรีอยู่ที่ 17400.46 ซึ่งต่ำที่สุดในบรรดาโมเดลทั้งหมด แสดงว่าโมเดลมีข้อผิดพลาดในการทำนายต่ำที่สุด

การทดลองที่ 5 พบว่าเกรเดียนบูสทรีเป็นโมเดลที่มีประสิทธิภาพดีที่สุด เมื่อทำการตัดคุณลักษณะด้านราคาประเมินออกจากตัวแปรต้น โมเดลที่ได้มีประสิทธิภาพเป็นอันดับ 2 รองจากโมเดลที่ได้จากการทดลองที่ 4 จากการวิเคราะห์ความสำคัญของคุณลักษณะพบว่าโมเดลให้ความสำคัญกับกลุ่มสถานที่สำคัญมากที่สุด รองลงมาคือขนาดพื้นที่ และประเภทที่ดิน แสดงให้เห็นว่าสามารถพยากรณ์ราคาขายต่อตารางวาได้ใกล้เคียงกับราคาขายจริงหากทราบว่าที่ดินที่ต้องการประเมินนั้นอยู่ห่างจากสถานที่สำคัญเท่าใดเมื่อเทียบกับ การประเมินโดยทราบว่าขนาดของพื้นที่ หรือประเภททรัพย์เป็นอะไร

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

ข้อสังเกตที่น่าสนใจจากการเปรียบเทียบผลการทดลองที่ 3 และผลการทดลองที่ 4 ที่ตัดตัวแปรราคาประเมินต่อตารางวาของที่ดินผืนนั้น ๆ ออกพบว่าได้โมเดลที่มีประสิทธิภาพสูงขึ้น ซึ่งอาจจะเป็นผลมาจากการโอเวอร์ฟิตของโมเดลที่ให้น้ำหนักไปในการจดจำราคาประเมินต่อตารางวาเป็นหลัก หรือ เมื่อฟีเจอร์ที่มีอิทธิพลถูกลบออก โมเดลจะพึ่งพาฟีเจอร์อื่น ๆ ที่อาจมีข้อมูลที่สำคัญมากขึ้น ซึ่งทำให้โมเดลสามารถจับโครงสร้างที่แท้จริงของข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น

การนำโมเดลมาทดลองใช้ในเว็บไซต์

เมื่อได้โมเดลเป็นที่เรียบร้อยแล้ว ผู้วิจัยได้ทำการออกแบบและพัฒนาเว็บไซต์สำหรับการทำนายราคาขายต่อตารางวาโดยใช้โมเดลเกรเดียนบูสทรีของการทดลองที่ 4 ที่มีค่าผลการทดลองที่ดีที่สุดในการทำนายข้อมูล เมื่อทดลองกรอกข้อมูลค่าละติจูด ลองติจูด ไร่ งาน ตารางวา และเลือกประเภทที่ดินว่างเปล่าโดยที่ดินที่เลือกมาจากระบบค้นหารูปแปลงที่ดินของกรมที่ดิน ที่เป็นจังหวัดขอนแก่น อำเภอเมืองขอนแก่น ตำบลในเมือง พบผลการวิจัย ดัง Figure 5.

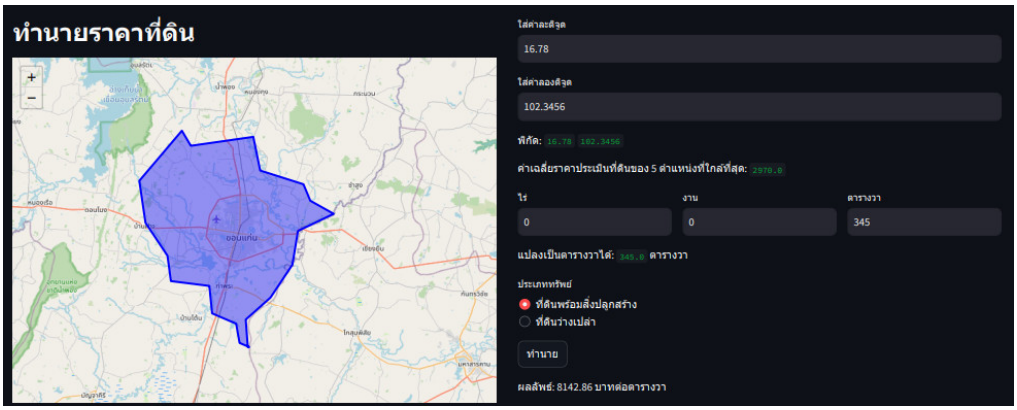

Figure 5. Entering data to predict the price per square meter.

จาก Figure 5. พบว่า ในเนื้อที่ 0 ไร่ 0 งาน 49.8 ตารางวา ราคาประเมินที่ดินกรมธนารักษ์ 6,500 บาทต่อตารางวา ค่าพิกัดแปลง 16.44714910,102.82108872 ผลลัพธ์ที่ได้จากทำนายราคาขายต่อตารางวาของแปลงที่ดินนี้คือ 7290.63 บาทต่อตารางวา

7. สรุปผลการวิจัย (Conclusion)

การวิจัยนี้ได้ทำการพัฒนาอัลกอริทึมการเรียนรู้ของเครื่องเพื่อประเมินราคาที่ดินในเขตอำเภอเมือง จังหวัดขอนแก่น โดยใช้ข้อมูลจากกรมบังคับคดีและกรมที่ดิน จากการทดลองปรับชุดกลุ่มตัวแปรต้นและการทำการคัดเลือกตัวแบบด้วยกระบวนการค้นหาพามิเตอร์แบบกริด และการทำการทดสอบแบบไขว้ทบ ผลการวิจัยพบว่าโมเดลเกรเดียนบูสทรีให้ผลประสิทธิภาพการทำนายราคาที่ดินจากข้อมูลการซื้อขายจากข้อมูลกรมบังคับคดีได้ดีที่สุด โดยมีค่าสัมประสิทธิ์กำลังสองของการอธิบายความแปรปรวนสูงที่สุดมีค่า 0.80 ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองมีน้อยที่สุดโดยมีค่า 7929.40 และ 15281.33 ตามลำดับ ซึ่งจากโมเดลที่ได้ อนุมาณความสำคัญของคุณลักษณะ จากมากไปน้อยได้ดังต่อไปนี้ กลุ่มสถานที่สำคัญขนาดพื้นที่ ราคาประเมินจากห้าตำแหน่ง และประเภททรัพย์ ข้อสังเกตที่น่าสนใจอีกประการคือ ถึงแม้ว่าราคาประเมินของ ที่ผืนนั้น ๆ จะมีความสำคัญของคุณลักษณะที่สูงที่สุดจากทุก ๆ ตัวแปร แต่เมื่อตัดออกทำให้ได้โมเดลที่มีประสิทธิภาพสูงขึ้น ซึ่งอาจเกิดจากการโอเวอร์ฟิตของตัวแบบที่อาศัยการจดจำราคาประเมิน หรือ

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

การตัดตัวแปรที่มีอิธิพลส่งผลให้โมเดลพึ่งพาคุณลักษณะอื่น ๆ ที่มีข้อมูลสำคัญมากขึ้น ทำให้โมเดลสามารถระบุและหาความสัมพันธ์ที่แท้จริงของข้อมูลได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

8. ข้อเสนอแนะงานวิจัย (Recommendation)

สำหรับการพัฒนาต่อยอดงานวิจัยนี้ สามารถที่จะพัฒนาในส่วนของการเก็บฐานข้อมูลเพิ่มเติมในส่วนของข้อมูลการติดจำนองซึ่งยังมีข้อจำกัดคือ ข้อมูลจากกรมบังคับคดีจะถูกลบออกไปเมื่อที่ดินแปลงนั้นขายได้ทำให้ไม่สามารถเก็บข้อมูลในส่วนนี้ย้อนหลังได้ และสามารถเพิ่มคุณลักษณะตัวแปรต้นในการนำข้อมูลรูปถ่ายของเมืองตามระยะเวลามาทำการจำแนกรูปภาพเชิงความหมาย (Semantic Segmentation) เพื่อสกัดคุณลักษณะของพื้นที่ ที่มีการเปลี่ยนแปลงเชิงเวลาเพิ่มเติมสำหรับโมเดลในการพยากรณ์เป็นแนวทางในการพัฒนาต่อไป

9. กิตติกรรมประกาศ (Acknowledgement)

งานวิจัยชิ้นนี้การได้รับการสนับสนุนทรัพยากร คอมพิวเตอร์จากบริษัทอินเตอร์เน็ตประเทศไทย ซึ่งเป็นส่วนสำคัญในการช่วยเหลือและสนับสนุนสำหรับการวิจัย

10. เอกสารอ้างอิง (References)

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research, 13*(10), 281–305.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. Springer Science & Business Media.

Chandrashekar, G., & Sahin, F. (2014). A Survey on Feature Selection Methods. *Computers & Electrical Engineering, 40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

Choompol, A. (2019). *Feature Selection and Redundant Feature Elimination for Opinion Classification on Social Network*. [Doctoral dissertation, Mahasarakham University]. Mahasarakham University Intellectual Repository. http://202.28.34.124/dspace/handle/123456789/537. (In Thai).

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research, 3*(7–8), 1157–1182.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1137–1143.

Kumpu, P., & Piyathamronchai, K. (2024). Application of the Geographic Information System to Devolop Land Valuation Models, Case Study: Muang Chiangmai District, Chiangmai Province. *The Journal of Spatial Innovation Development, 5*(2), 42–64. (In Thai)

Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems*. 4765-4774. https://doi.org/10.48550/ARXIV.1705.07874.

Citation: Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.

Na Bangchang, K. (2011). *A Variable Selection in Multiple Linear Regression Models Based on Tabu Search* [National Institute of Development Administration]. https://doi.org/10.14457/NIDA.the.2011.17. (In Thai).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. https://doi.org/10.48550/ARXIV.1201.0490

Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. J. (2022). Housing Price Prediction Incorporating Spatio-Temporal Dependency into Machine Learning Algorithms. *Cities, 131*, 103941. https://doi.org/10.1016/j.cities.2022.103941.

Vanderplas, J. (2017). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly.

Citation:
Sriwuttisap, Y., Sungklinhom, D., Tongleamnak, S., & Tangchoopong, T. (2024). Land Price Prediction with Machine Learning in Mueang Khon Kaen District. *Journal of Computer and Creative Technology, 2*(2), 71-86. https://doi.org/10.14456/jcct.2024.8.