

ระบบประเมินราคาที่ดินจากแปลงที่ดินตัวแทนโดยใช้วิธีการเครื่องจักรเรียนรู้

Land Appraisal Price Prediction System from Reference Lots using Machine Learning Models

เจตนา สมคำนิ่ง*

ธนภัทร ชั่งกะจิตร**

บทคัดย่อ

ในปัจจุบันภารกิจด้านการประเมินราคาที่ดินรายแปลงของกรมธนารักษ์ สังกัดกระทรวงการคลัง ได้มีการพัฒนารูปแบบวิธีการให้อยู่ภายใต้แนวคิดเศรษฐกิจดิจิทัล (Digital Economy) ด้วยการบริหารจัดการฐานข้อมูลขนาดใหญ่ของรัฐบาลไทย ซึ่งเป็นการสอดคล้องกับบริบทของแผนพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม ยุทธศาสตร์ที่ 4 การปรับเปลี่ยนภาครัฐสู่การเป็นรัฐบาลดิจิทัล อย่างไรก็ดี การเปลี่ยนผ่านจากวิธีการเดิมที่ต้องอาศัยประสบการณ์ของผู้ประเมินมาใช้วิธีการเรียนรู้ของเครื่องนั้นมีขั้นตอนกระบวนการที่มาก รวมถึงต้องมีการเปลี่ยนแปลงแนวคิดในหลายๆ จุด งานฉบับนี้จึงได้นำเสนอแนวทางเบื้องต้นในการพัฒนาแบบจำลองการประเมินราคาที่ดินที่ผสมผสานแนวคิดการวิเคราะห์ทางด้านวิทยาศาสตร์ข้อมูลรวมกับความรู้และประสบการณ์ของผู้เชี่ยวชาญ โดยมีวัตถุประสงค์เพื่อช่วยส่งเสริมประสิทธิภาพการทำงานด้านการประเมินราคาที่ดินให้มีความทันสมัย อธิบายได้ ในสารนิพนธ์นี้ได้นำเสนอการประเมินราคาที่ดินผ่านแปลงที่ดินตัวแทนซึ่งพิจารณาจากแปลงที่ดินที่ตั้งอยู่ใกล้กับจุดสนใจ (Points of interest) ต่างๆ ส่งผลให้แปลงที่ดินตัวแทนที่ได้มาจากวิธีนี้มีความโดดเด่นในแง่ของสถานที่ตั้ง (Location) โดยนำวิธีการทางสถิติเชิงพื้นที่เรื่องการวิเคราะห์จุดความร้อนด้วยวิธี GetisOrd Gi* เข้ามาช่วยวิเคราะห์ความใกล้เคียงจากแต่ละแปลงที่ดินไปยังแต่ละจุดสนใจ ผลการทดลองแสดงให้เห็นว่า แบบจำลองสามารถประเมินราคาที่ดินตัวแทนได้ดีกว่าแบบจำลองมวลรวมที่ใช้กันโดยทั่วไป (MAPE 21% vs 25%) อีกทั้งยังสามารถอธิบายเหตุผลเบื้องหลังได้ดี ทำให้ไม่ส่งผลกระทบในเชิงการเปลี่ยนแปลงอย่างฉับพลันจากรูปแบบการประเมินราคาแบบเดิมมากเกินไป ดังนั้น บุคลากรที่เกี่ยวข้องสามารถนำไปใช้งานได้โดยมีต้นทุนการเรียนรู้ที่ต่ำ

* นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

** ที่ปรึกษาสารนิพนธ์หลัก

ABSTRACT

The assessment system of land appraisal price responsible by the Treasury department, Ministry of Finance, has been revolutionizing under the concept of “Digital Economy” using big data technology. This also complies the 4th strategy (Transform towards Digital Thailand) of Digital Development for National Economic and Social Development Plan. However, transformations from current practices, which rely mainly on experienced assessors, to machine learning ways of thinking need time and breakthrough paradigm shifts. . This paper proposes the preliminary but promising ideas to develop land appraisal price system through both data science and domain experts in the field in order to promote its modernization and explainability. Main idea of this work is to determine and predict appraisal price of reference-lot lands which can further use to assess all lands nearby. The reference-lot lands are considered from closeness to POI (Points of interest) which implies outstanding in terms of locations they situated. The main technique is performed by a hotspot analysis called GetisOrd Gi* to help investigate the distance to each POI. Experimental results show that our approach produced better performance in terms of MAPE (21% vs 25%) compared to domain-experts approach. Moreover, model’s interpretability and simplicity result in a shallow learning curve. Consequently, officers with little knowledge can easily follow along and understand the main theme of the system in a short period of time.

บทนำ

ด้วยพระราชบัญญัติภาษีที่ดินและสิ่งปลูกสร้าง พ.ศ. 2562 มีบทบัญญัติให้ใช้บัญชีกำหนดราคาประเมินทุนทรัพย์ของอสังหาริมทรัพย์ที่กรมธนารักษ์จัดทำขึ้นเป็นฐานในการคำนวณภาษีที่ดินและสิ่งปลูกสร้าง ซึ่งเป็นการขยายขอบเขตการปฏิบัติงานของกรมธนารักษ์ที่นอกเหนือจากการประเมินราคาทรัพย์สิน เพื่อการจดทะเบียนสิทธิและนิติกรรมตามประมวลกฎหมายที่ดิน พ.ศ. 2497 (และที่แก้ไขเพิ่มเติม) โดยปัจจุบัน ภารกิจด้านการประเมินราคาที่ดินของกรมธนารักษ์ ได้มีการพัฒนารูปแบบการบริหารจัดการภายใต้แนวคิดเศรษฐกิจดิจิทัล (Digital economy) ของรัฐบาล โดยนำการจัดการฐานข้อมูลข้อมูลขนาดใหญ่ (Big data) เข้ามานำรวมการ ซึ่งเป็นการสอดคล้องกับบริบทของแผนพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม ยุทธศาสตร์ที่ 4

ในการประเมินราคาที่ดินโดยกองประเมินราคาทรัพย์สิน กรมธนารักษ์ได้ยึดหลักวิธีเปรียบเทียบตลาด (Market comparison approach) ซึ่งเป็น 1 ใน 6 วิธีการประเมินมาตรฐาน (ที่มา: มูลนิธิประเมินค่า-นายหน้าแห่งประเทศไทย (องค์การสาธารณประโยชน์)) เนื่องจากมีความชัดเจน

เพราะการซื้อขายเกิดขึ้นแล้วจริง จึงสามารถอธิบายได้โดยง่าย ในเชิงเศรษฐศาสตร์ ที่ดินถือเป็น “ทุน” ต้นน้ำในการประกอบธุรกิจต่างๆ ทำให้การนำวิธีการประเมินแบบอื่นมาใช้ก่อให้เกิดปัญหาในเชิงเปรียบเทียบ ยกตัวอย่างเช่น วิเคราะห์มูลค่าจากต้นทุนนั้นใช้การรวมต้นทุนในการได้มาซึ่งทรัพย์สินมากำหนดเป็นราคาประเมิน แต่ที่ดินทำได้แค่เพียงการเทียบเคียงกับที่ดินผืนอื่นเท่านั้น สร้างทดแทนไม่ได้ หรือ กรณีวิธีแปลงรายได้เป็นมูลค่า ซึ่งมีสมมติฐานทางธุรกิจหลากหลาย ไม่ว่าจะเป็นอัตราคิดลด (ดอกเบี้ย) การแกว่งตัวของรายได้ในแต่ละช่วงเวลา ก่อให้เกิดปัญหาความโปร่งใสในการกำหนดราคาประเมิน

นอกจากนี้ การประเมินที่ดินที่ยึดหลักการอ้างอิงข้อมูลธุรกรรมการซื้อขายที่เกิดขึ้นในตลาดนั้น ควรพิจารณาถึงการใช้ประโยชน์สูงสุดของทรัพย์สิน (Highest and best use) หมายถึง การใช้ประโยชน์ในทรัพย์สินให้ดีที่สุดเท่าที่จะเป็นไปได้ภายใต้เงื่อนไขที่เป็นไปตามตัวบทกฎหมาย ความเหมาะสมในทางกายภาพ และความเป็นไปได้ทางการตลาดและทางการเงิน ยิ่งไปกว่านั้น ข้อมูลตลาดหรือหลักฐานการซื้อขายในตลาดที่นำมาใช้ในการประเมินราคาที่ดิน ควรมีทำเลที่ตั้งอยู่ในบริเวณเดียวกันหรือใกล้เคียงกัน มีลักษณะหรือสภาพที่สามารถเปรียบเทียบกันได้ ตลอดจนมีการใช้ประโยชน์สูงสุดคล้ายคลึงกัน ซึ่งเป็นไปตามมาตรฐานวิชาชีพการประเมินมูลค่าทรัพย์สินเรื่องที่ 2 หลักเกณฑ์การประเมินเพื่อกำหนดมูลค่าตลาด

สารนิพนธ์ฉบับนี้นำเสนอการประเมินราคาที่ดินโดยผ่านแปลงที่ดินตัวแทนในบริเวณเดียวกัน โดยนาระยะทางจากแต่ละแปลงที่ดินไปยังแต่ละจุดสนใจ (POI – Points of interest) มาทำการวิเคราะห์จุดความร้อน (Hotspot analysis by GetisOrd Gi*) ทำให้แปลงตัวแทนที่ได้จากวิธีนี้จะมีมีความโดดเด่นในแง่ของการอธิบาย เช่น แปลงที่ดิน A มีราคาแพงกว่า B เนื่องจากแปลงที่ดิน A อยู่ใกล้กับห้างสรรพสินค้าขนาดใหญ่มากกว่า เป็นต้น เมื่อได้แปลงที่ดินตัวแทนแล้ว จะทำการฝึกแบบจำลองโดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้แปลงซื้อขายจดทะเบียนในบริเวณเดียวกันมาเป็นข้อมูลเทรน หลังจากนั้น จึงนำโมเดลที่ได้ไปประเมินราคาแปลงที่ดินตัวแทนแต่ละแปลงต่อไป จากผลการทดลองได้ผลลัพธ์สุดท้ายอยู่ในระดับที่น่าพอใจ โดยมีค่า Mean absolute percentage error (MAPE) อยู่ที่ 21% (เทียบระหว่างราคาประเมินที่ได้จากแบบจำลองกับ ราคาประเมินที่ได้จากกรมธนารักษ์) นอกจากนี้ หากนำไปเปรียบเทียบกับวิธีการประเมินราคาที่ดินแบบมวลรวมโดยทั่วไป (ไม่ได้ผ่านแปลงที่ดินตัวแทน) พบว่า MAPE ของวิธีที่นำเสนอสามารถเอาชนะ MAPE ของวิธีแบบมวลรวมโดยทั่วไปได้ (21% vs 25%)

วัตถุประสงค์

1. เพื่อออกแบบและพัฒนา “แบบจำลองต้นแบบ” ในการประเมินราคาที่ดินตัวแทน
2. เพื่อพัฒนา “ระบบตรวจสอบ” ความถูกต้องของราคาประเมินที่ได้มาจากแบบจำลอง

ประโยชน์ที่คาดว่าจะได้รับ

1. ช่วยลดภาระหน้าที่ของนักประเมินราคา ทำให้สามารถประเมินราคาที่ดินได้ในเวลาที่สั้นลงกว่าเดิม โดยมีแบบแผนเดียวกันทั้งประเทศและเป็นที่ยอมรับในระดับสากล
2. ช่วยให้ราคาประเมินที่ดินมีความโปร่งใส อธิบายได้ และมีเหตุผล

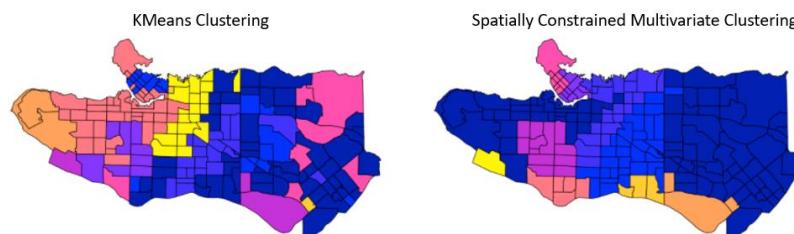
ขอบเขตงานวิจัย

1. พื้นที่ต้นแบบที่ใช้พัฒนาแบบจำลอง คือ พื้นที่เขตจตุจักร
2. ราคาประเมินที่ดินจากกรมธนารักษ์เป็นราคาที่ใช้ประกาศใช้ในปี พ.ศ. 2559
3. แปลงที่ดินที่จะใช้ในแบบจำลองจะไม่รวม แปลงที่ดินบล็อคอบลู (แปลงที่ดินใช้ประโยชน์ร่วมกันหลายโฉนด), แปลงที่ดินบล็อกฟิชซ์ (แปลงหมู่บ้านจัดสรร) และแปลงที่ดินที่ไม่มีทางเข้า-ออก (ที่ดินตาบอด)

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

1. Spatially Constrained Multivariate Clustering

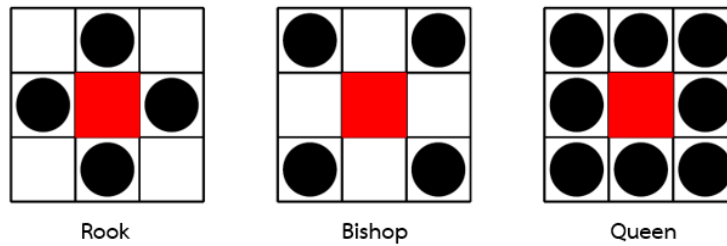
เป็นอัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) แบบหนึ่ง แต่มีวิธีการแบ่งกลุ่มที่คำนึงถึงความสัมพันธ์เชิงพื้นที่ระหว่างกัน โดยมีสมมติฐานว่า แปลงที่ดินที่อยู่ในบริเวณเดียวกันหรือใกล้เคียงกันเท่านั้น จึงจะถูกจัดให้อยู่ในกลุ่มเดียวกัน มิเช่นนั้นแล้ว จะถูกแบ่งออกเป็นอีกกลุ่มหนึ่ง



ภาพที่ 1: เปรียบเทียบผลลัพธ์การจัดกลุ่มระหว่างวิธีทั่วไป (KMeans) และวิธี Spatially Constrained Multivariate Clustering

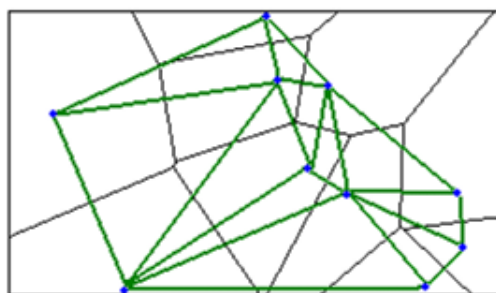
2. รูปแบบการเลือกเพื่อนบ้าน

เป็นรูปแบบเพื่อใช้กำหนดว่าแปลงที่ดินใดบ้างที่เป็นเพื่อนบ้านกับแปลงที่ดินที่สนใจ โดยรูปแบบที่เป็นที่นิยมมีอยู่ 3 แบบ ได้แก่ 1) รูปแบบความสัมพันธ์เชิงด้าน (Rook) 2) รูปแบบความสัมพันธ์เชิงมุม (Bishop) และ 3) รูปแบบความสัมพันธ์ทั้งเชิงด้านและเชิงมุม (Queen)

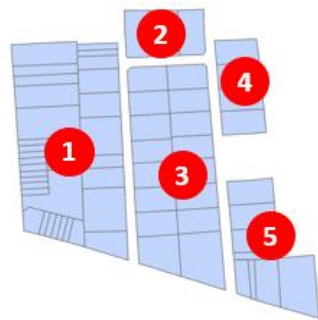


ภาพที่ 2: แสดงรูปแบบการเลือกเพื่อนบ้านทั้ง 3 แบบ (Gonzalez Canche, October 5, 2020)

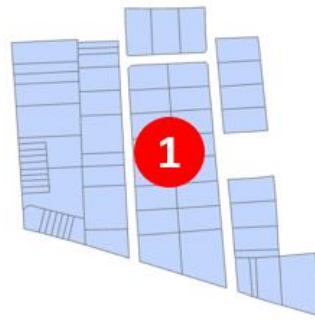
อย่างไรก็ดี รูปแบบทั้ง 3 แบบดังกล่าวข้างต้นไม่เหมาะสมกับรูปแปลงที่ดินเนื่องจากแปลงที่ดินไม่ได้ติดกันเป็นผืนเดียว มักถูกคั่นด้วยถนน, แม่น้ำ, ลำคลอง เป็นต้น ดังนั้น จึงนำรูปแบบการหาเพื่อนบ้านอีกวิธีหนึ่ง คือ Trimmed Delaunay Triangulation เข้ามาประยุกต์ใช้ โดยวิธีนี้ ใช้ลักษณะการวาดสามเหลี่ยมไวโรนอยด์จากจุด centroid ของแต่ละโพลีกอนแปลงที่ดิน หมายความว่า หากจุด centroid 3 จุด (เฉพาะจุดที่ใกล้กัน) ใดๆ ที่สามารถเชื่อมได้เป็นสามเหลี่ยม แสดงว่าแปลงที่ดินที่เป็นตัวแทนของ 3 จุดนั้น ถือว่าเป็นเพื่อนบ้านกัน



ภาพที่ 3: แสดงรูปแบบการเลือกเพื่อนบ้านแบบ Trimmed Delaunay Triangulation (ESRI)



เลือกเพื่อนบ้านแบบเชิงด้าน
จะแบ่งเป็น 5 กลุ่ม



เลือกเพื่อนบ้านแบบ Trimmed Delaunay Triangulation
จะถือเป็นกลุ่มเดียว

ภาพที่ 4: แสดงความแตกต่างของผลลัพธ์การแบ่งกลุ่มที่ได้จากการหาเพื่อนบ้านด้วย
วิธี “Rook” และ “Trimmed Delaunay Triangulation”

3. อัตราสัมพันธ์เชิงพื้นที่ระดับเฉพาะที่ GetisOrd Gi*

GetisOrd Gi* เป็นค่าสถิติเฉพาะที่ (Local) ที่วัดความสัมพันธ์เชิงพื้นที่ที่โดยพิจารณาในบริบทของหน่วยเพื่อนบ้านใกล้เคียง หากเพื่อนบ้านมีลักษณะหรือคุณสมบัติที่สนใจคล้ายๆ กัน แสดงว่าบริเวณนั้นมีความสัมพันธ์เชิงพื้นที่ต่อกัน

ภาพที่ 5 แสดงให้เห็นถึงความสัมพันธ์ระหว่างชั้นข้อมูลนำเข้า (ซ้ายสุด) โดยเมื่อคำนวณ GetisOrd Gi* จะได้ Z-scores (ค่าสถิติ GetisOrd Gi* นั้นเอง) และค่า p-value ตามลำดับ



ภาพที่ 5: แสดงผลลัพธ์จาก GetisOrd Gi* ในรูปแบบแผนที่ (ESRI)

จากภาพด้านบน (ภาพกลาง) ค่า Z-scores ได้ถูกนำมาแบ่งออกเป็น 7 bins โดยยึดจากค่าความเชื่อมั่นทางสถิติในระดับต่างๆ ดังนี้

Bin -3: Cold Spot – 99% Confidence แสดงด้วยสีน้ำเงินเข้ม

Bin -2: Cold Spot – 95% Confidence แสดงด้วยสีน้ำเงินปานกลาง

Bin -1: Cold Spot – 90% Confidence แสดงด้วยสีน้ำเงินอ่อน

Bin 0: Not Significant แสดงด้วยสีขาว หรือ สีเหลืองอ่อนๆ

Bin +1: Hot Spot – 90% Confidence แสดงด้วยสีแดงอ่อน

Bin +2: Hot Spot – 95% Confidence แสดงด้วยสีแดงปานกลาง

Bin +3: Hot Spot – 99% Confidence แสดงด้วยสีแดงเข้ม

ตัวอย่างการแปรความ เช่น แปลงที่ดินที่มี Gi_Bin เป็น +3 หรือ สีแดงเข้ม หมายความว่า แปลงที่ดินนั้นอยู่ใกล้กับจุดสนใจ (Points of interest) ในภาพรวมมาก ถือเป็นแปลงที่ดินที่น่าสนใจ มีนัยสำคัญเชิงบวก ราคาประเมินควรจะสูงในเชิงเปรียบเทียบ เป็นต้น

4. งานวิจัยการประเมินราคาที่ดินแบบมวลรวม (Mass appraisal)

4.1 งานวิจัยของ Sukran Yalpir และ Fatma Bunyan Unel ปี 2017

งานวิจัยฉบับนี้นำเสนอแบบจำลองประเมินราคาที่ดินแบบมวลรวมในเขตจังหวัดคอนยา (Konya) ประเทศตุรกี โดยแบ่งปัจจัยสำคัญออกเป็น 4 ด้าน ได้แก่ ปัจจัยทางด้านทำเลที่ตั้ง, ปัจจัยเชิงกายภาพ, ปัจจัยทางด้านตัวบทกฎหมาย, และปัจจัยเกี่ยวกับสิ่งแวดล้อม ในขั้นตอนการเตรียมข้อมูล ผู้วิจัยได้นำเสนอการทำดัชนีเชิงพื้นที่ (Spatial indexing) สำหรับแต่ละปัจจัยโดยเจาะไปยังปัจจัยย่อย ๆ ที่เกี่ยวข้อง ยกตัวอย่างเช่น ดัชนีเชิงพื้นที่ของปัจจัยการเดินทาง จะได้มาจากการนำระยะทางไปยังสถานีรถไฟ, ระยะทางไปยังป้ายรถประจำทาง, ระยะทางไปยังท่าเรือ, ระยะทางไปยังสถานีท่ารถตู้ เป็นต้น มากระทำทางคณิตศาสตร์เชิงพื้นที่บางอย่างโดยได้ผลลัพธ์ออกมาค่าเดียว หลังจากนั้น นำอัลกอริทึม Multiple Linear Regression มาฝึกชุดข้อมูลซื้อ-ขายจดทะเบียนแล้วนำไปประเมินราคาที่ดินรายแปลง ค่า R-squared ของชุดฝึกได้ค่าสูงถึง 0.85

อย่างไรก็ดี ข้อได้เปรียบของงานวิจัยฉบับนี้ คือ ข้อมูลราคาซื้อ-ขายจดทะเบียนค่อนข้างจะใกล้เคียงความเป็นจริง (อาจจะเป็นเพราะมีกฎหมายบังคับซื้อขาย) ซึ่งแตกต่างจากสถานการณ์ในประเทศไทย

4.2 งานวิจัยของ MIE OAK CHAE และ INHYUK KWON ปี 2018

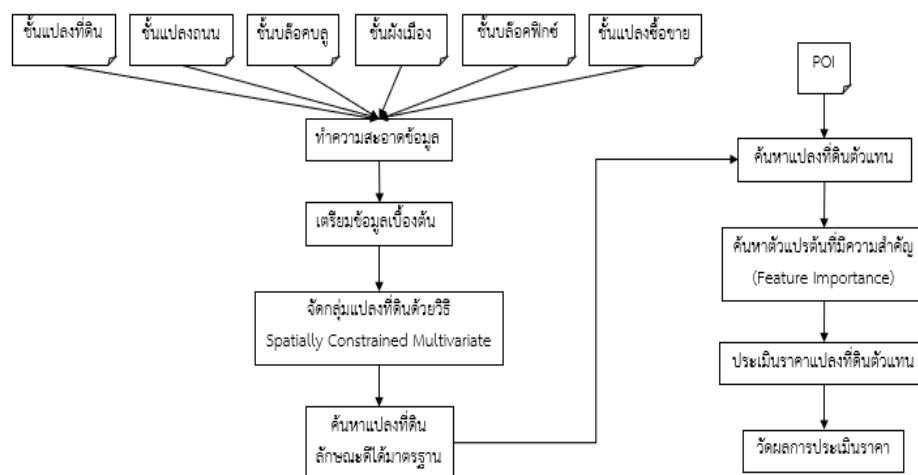
งานวิจัยฉบับนี้ถูกนำไปใช้เป็นหลักการโดยทั่วไปสำหรับการประเมินราคาที่ดินในเขตเมืองของประเทศเกาหลีใต้ โดยผู้วิจัยนำเสนอหลักการสำคัญ 3 ประการในการประเมินราคาที่ดิน คือ 1) การสำรวจเพื่อค้นหาแปลงที่ดินตัวแทนที่มีศักยภาพ เช่น เป็นแปลงที่ดินที่สามารถสะท้อน

ลักษณะทางกายภาพของที่ดินอื่นในบริเวณนั้นได้, เป็นแปลงที่ดินที่มีโอกาสในการเปลี่ยนแปลง ลักษณะการใช้ประโยชน์ในระยะยาวต่ำ, เป็นแปลงที่ดินที่สามารถเข้าถึงหรือทำการสำรวจได้ง่าย เป็นต้น 2) ประเมินราคาแปลงที่ดินตัวแทน และ 3) ตารางการปรับปรุงราคาที่ดินแตกแปลง ซึ่งเป็นการเปรียบเทียบแปลงที่ดินที่ต้องการประเมินราคากับแปลงที่ดินตัวแทน แล้วทำการปรับราคาขึ้น-ลงตามหลักเกณฑ์ที่กำหนด เช่น หากแปลงที่ดินที่กำลังพิจารณามีลักษณะเป็นรูปทรง ให้ทำการทอนราคาลง 11% จากแปลงที่ดินตัวแทนในบริเวณดังกล่าว เป็นต้น

งานวิจัยฉบับนี้ค่อนข้างมีความแตกต่างจากงานวิจัยอื่นๆ ในแง่ของการประเมินราคาที่ดินผ่าน “แปลงที่ดินตัวแทน” ในขณะที่งานอื่นๆ ใช้การทำแบบจำลองบนราคาซื้อ-ขายจดทะเบียน

ระเบียบวิธีวิจัย

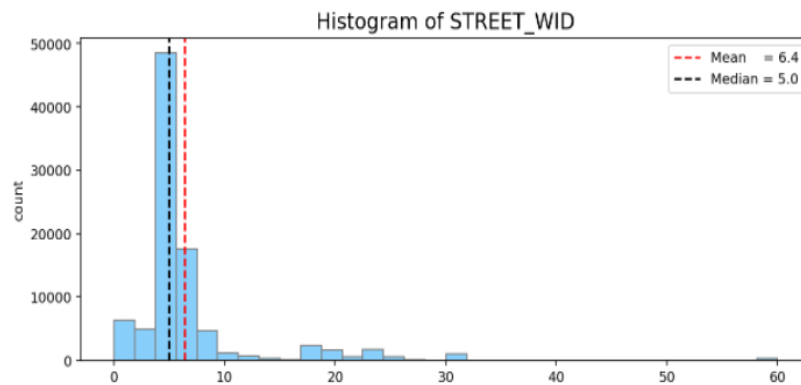
การดำเนินการวิจัยอาศัยแนวคิด “แปลงที่ดินตัวแทน” เป็นหลัก ซึ่งสอดคล้องกับงานวิจัยของ MIE OAK CHAE และ INHYUK KWON โดยกระบวนการในภาพรวมแสดงได้ดังภาพที่ 6 นอกจากนี้ ได้นำผลลัพธ์ที่ได้ตามแนวคิดนี้ไปเปรียบเทียบกับผลลัพธ์ตามแนวคิดการประเมินราคาแบบมวลรวมโดยทั่วไปด้วย เพื่อพิจารณาความเหมาะสมของแนวคิดแปลงตัวแทน



ภาพที่ 6: กระบวนการประเมินราคาแปลงตัวแทน

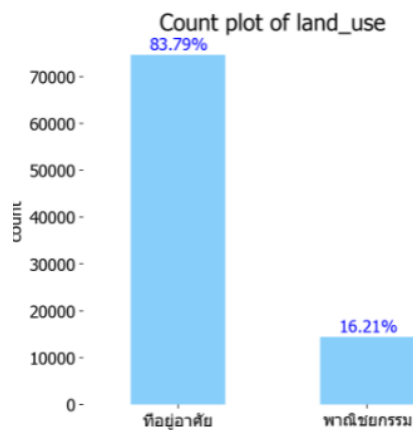
1) ตำรวจข้อมูลของชั้นข้อมูล shape files ต่างๆ ได้แก่ ชั้นข้อมูลแปลงที่ดิน, ชั้นข้อมูลแปลงถนน, ชั้นข้อมูลบล็อคนับ, ชั้นข้อมูลบล็อกฟิสิกซ์, ชั้นข้อมูล POI, ชั้นข้อมูลผังเมือง, และชั้นข้อมูลแปลงชื่อ-ขายจดทะเบียน ทั้งนี้ การสำรวจข้อมูลเบื้องต้นถูกโดยแบ่งเป็น 2 ประเภท ได้แก่

1.1) ข้อมูลเชิงตัวเลข (Numerical data) ทำการหาจำนวนข้อมูลที่เป็นค่าว่าง, ค่าพิสัย, ค่าเฉลี่ย, ค่ามัธยฐาน และการกระจายตัวของข้อมูล ตัวอย่างเช่น



ภาพที่ 7: แสดงฮิสโทแกรมของตัวแปร “ความกว้างของถนน หรือ STREET_WID”

1.2) ข้อมูลเชิงประเภท (Categorical data) ทำการนับความถี่ของค่าที่เป็นไปได้ของข้อมูลเชิงประเภทแต่ละตัว



ภาพที่ 8: Count plot ของตัวแปร “การใช้ประโยชน์ที่ดิน หรือ LAND_USE”

2) การทำความสะอาดข้อมูล

กระบวนการเริ่มจากการซ่อมแซม geometry ที่มีปัญหา Topology แล้วทำการตัดแปลงที่ดินที่ไม่มีราคาประเมินออกไป พร้อมทั้งเติมค่าสูญหายด้วยวิธีเทียบเคียงกับแปลงที่ดินเพื่อนบ้าน หลังจากนั้น ทำการตัด “แปลงถนน”, “แปลงบล็อกฟิซ” และ “แปลงบล็อกชล” ออกไปด้วยหลักการพื้นที่ซ้อนทับ โดยในขั้นตอนสุดท้าย ตัวแปร/ปัจจัยที่เกี่ยวข้องจะถูกนำมาเข้ากระบวนการ Discretization เพื่อเปลี่ยนข้อมูลที่มีลักษณะเป็นตัวเลขให้อยู่ในลักษณะที่เป็นช่วงและ One-hot-encoding กับข้อมูลเชิงประเภทที่เกี่ยวข้อง

3) การจัดกลุ่มแปลงที่ดิน

วัตถุประสงค์เพื่อจัดกลุ่มแปลงที่ดินที่มีคุณลักษณะใกล้เคียงกันไว้ด้วยกัน โดยมีสมมติฐานหลัก คือ แปลงที่ดินที่อยู่ใกล้กันและ/หรือมีคุณลักษณะบางประการใกล้เคียงกัน ควรจะมีราคาประเมินไม่แตกต่างกัน ซึ่งกลุ่มแปลงที่ดินที่หาได้จะนำไปค้นหาแปลงที่ดินตัวแทนต่อไป ผลจากการแบ่งกลุ่มพบว่าวิธี Spatially Constrained Multivariate Clustering ให้ผลลัพธ์ที่ดีที่สุด โดยไม่ก่อให้เกิดปัญหาการกระจายของแปลงที่ดินในแต่ละกลุ่ม (กรณี Clustering โดยทั่วไป เช่น KMeans จะพบปัญหานี้)

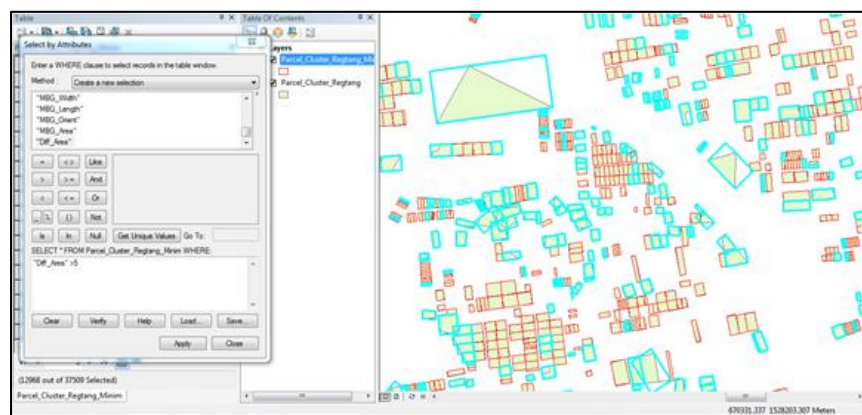


ภาพที่ 9: แสดงผลลัพธ์จากการจัดกลุ่มด้วย Spatially Constrained Multivariate Clustering

4) ค้นหาแปลงที่ดินลักษณะดีได้มาตรฐาน

แปลงที่ดินที่มีลักษณะดีได้มาตรฐานในที่นี้ หมายถึง แปลงที่ดินที่มีลักษณะเป็นสี่เหลี่ยม โดยไม่มีรูปทรงที่แปลกประหลาด เช่น รูปแปลงขงขาลีบ, รูปแปลงสามเหลี่ยม, รูปแปลงรีว หรือรูปแปลงที่มีลักษณะแปลกอื่นๆ

หลักการในการค้นหาจะใช้เทคนิค Minimum Boundary Rectangle (MBR) เข้ามาช่วย โดยหากแปลงที่ดินใดที่มีผลต่างพื้นที่เทียบกับ MBR ไม่เกิน 5% จะถือเป็นแปลงลักษณะดีได้มาตรฐาน



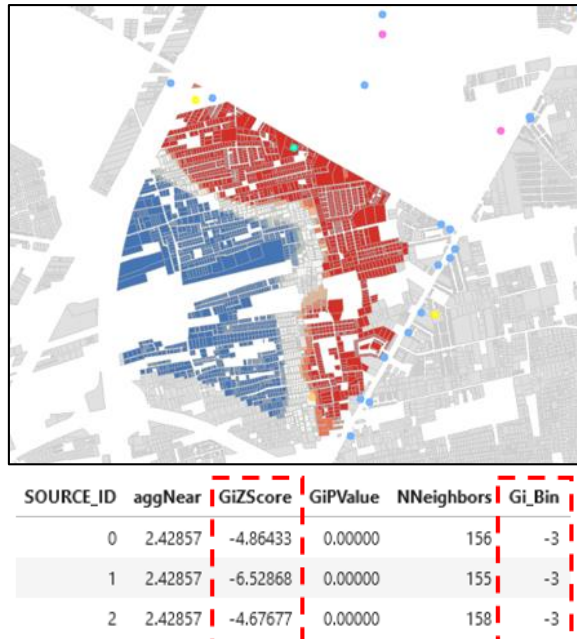
ภาพที่ 10: แสดงเทคนิคการใช้ MBR เพื่อเปรียบเทียบผลต่างพื้นที่

5) ทำการคัดเลือกแปลงที่ดินตัวแทนของแต่ละหน่วยถนนในแต่ละกลุ่มแปลงที่ดิน

ในขั้นตอนนี้ การคัดเลือกและค้นหาแปลงที่ดินตัวแทนจะใช้หลักการ “หน่วยถนนเดียวกัน” หมายความว่า แปลงที่ดินตัวแทนที่ค้นหาได้จะเป็น “ตัวแทน” ของทุกๆ แปลงที่ดินบนหน่วยถนนเดียวกันนั้น การค้นหาแปลงที่ดินตัวแทนนำเสนอ 3 วิธี ได้แก่ 1) วิธีผู้เชี่ยวชาญ, 2) วิธีวิเคราะห์จุดความร้อนด้วย GetisOrd Gi*, และ 3) วิธีเปรียบเทียบเชิงกายภาพ ในที่นี้ จะเน้นวิธีที่ 2) เนื่องจากให้ผลลัพธ์ที่ดีที่สุด

วิธีวิเคราะห์จุดความร้อนด้วย GetisOrd Gi*

กระบวนการเริ่มจากแต่ละกลุ่มคลัสเตอร์ที่ได้จาก Spatially Constrained Multivariate Clustering



ภาพที่ 12: แสดงตัวอย่างผลลัพธ์จากการรัน GetisOrd Gi* สำหรับคลัสเตอร์ที่ 1

5) ทำการค้นหาแปลงที่ดินตัวแทนในแต่ละหน่วยถนน เบื้องต้นพิจารณาจากแปลงที่ดินที่มีคุณลักษณะซ้ำๆ กันในกลุ่ม โดยใช้ช่วงของข้อมูลที่มีความถี่อยู่ระหว่างเปอร์เซ็นต์ไทล์ที่ 25 และเปอร์เซ็นต์ไทล์ที่ 75 ของตัวแปรอัตราส่วนของความกว้างต่อความลึก (DEPTH_WR) และ พื้นที่ (AREA) หลังจากนั้น หากจำนวนแปลงที่ดินตัวแทนมีจำนวนมากเกินกว่าที่กำหนด เช่น 10% ของจำนวนแปลงที่ดินทั้งหมดบนหน่วยถนนนั้นๆ จะทำการคัดแปลงที่ดินที่มีคะแนน GiZScore ต่ำออกไปจนเหลือตามจำนวนที่ต้องการ



ภาพที่ 13: แสดงแปลงที่ดินตัวแทน (ภาพกลาง) ที่ค้นหาได้จากวิธี GetisOrd Gi*

6) ทำการค้นหาตัวแปรต้นที่มีความสำคัญ (Feature importance)

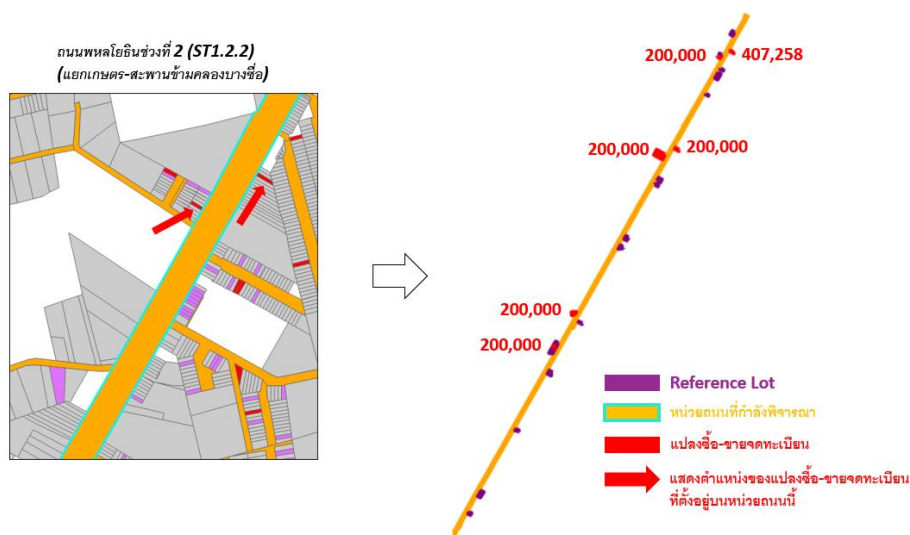
เนื่องจากจำนวนตัวแปรต้นที่สกัดขึ้นในหลายขั้นตอนก่อนหน้านี้มีจำนวนมากเกินกว่า 100 ตัวแปร ในขั้นตอนนี้จึงต้องการค้นหาตัวแปรต้นเฉพาะตัวที่สำคัญเพื่อนำไปทำแบบจำลอง ประเมินราคาแปลงตัวแทนต่อไป ทั้งนี้ ใช้การค้นหาผ่านอัลกอริทึมตระกูลต้นไม้ตัดสินใจ ซึ่งมักจะมีฟีเจอร์ที่สรุปเกี่ยวกับ Feature importance ไว้ สุดท้ายแล้ว ตัวแปรต้นที่สำคัญได้แก่ “การใช้ประโยชน์ที่ดิน หรือ LAND_USE”, “ความกว้างถนน หรือ STREET_WID”, “จำนวนร้านค้าภายในระยะ 1500 เมตร”, และ “ระยะทางสั้นที่สุดไปยังมหาวิทยาลัย”

7) ประเมินราคาแปลงที่ดินตัวแทน

นำเสนอการประเมินราคาแปลงที่ดินตัวแทนออกเป็น 2 วิธี ได้แก่ วิธีการ buffer ออกไปหาแปลงซื้อ-ขายจดทะเบียนจนครบตามจำนวนที่กำหนด และ วิธีการใช้แปลงซื้อ-ขายจดทะเบียนบนหน่วยถนนเดียวกัน ซึ่งวิธีหลังให้ผลลัพธ์ที่ดีกว่า

วิธีการใช้แปลงซื้อ-ขายจดทะเบียนบนหน่วยถนนเดียวกัน

ทำการค้นหาแปลงซื้อ-ขายจดทะเบียนที่ตั้งอยู่บนหน่วยถนนเดียวกัน ยกตัวอย่างเช่นตามภาพที่ 14 แปลงสีแดง คือ แปลงซื้อ-ขายจดทะเบียนที่จะนำไปทำแบบจำลอง โดยใช้อัลกอริทึม Linear Regression, Distance-based KNN, และ Uniform-based KNN



ภาพที่ 14: แสดงการใช้แปลงซื้อ-ขายจดทะเบียนบนหน่วยถนนเดียวกันมาทำแบบจำลอง

ผลการวิจัย

ตารางที่ 1: แสดงค่า MAPE ของการประเมินราคาแปลงที่ดินตัวแทนแต่ละวิธี

แบบจำลองที่ได้จากการ buffer ออกไป หาแปลงซื้อ-ขายจดทะเบียน				แบบจำลองที่ได้จากการใช้แปลงซื้อ-ขาย จดทะเบียนบนหน่วยถนนเดียวกัน			
	แปลงที่ดินตัวแทน จากผู้เชี่ยวชาญ	แปลงที่ดินตัวแทน จาก GetisOrd Gi*	แปลงที่ดินตัวแทน จากการเปรียบเทียบ เชิงกายภาพ		แปลงที่ดินตัวแทน จากผู้เชี่ยวชาญ	แปลงที่ดินตัวแทน จาก GetisOrd Gi*	แปลงที่ดินตัวแทน จากการเปรียบเทียบ เชิงกายภาพ
แปลงที่อยู่อาศัย				แปลงที่อยู่อาศัย			
MAPE - Linear Regression	336%	277%	320%	MAPE - Linear Regression	30%	27%	32%
MAPE - KNN-U	31%	26%	28%	MAPE - KNN-U	23%	21%	23%
MAPE - KNN-D	31%	26%	28%	MAPE - KNN-D	22%	21%	23%
จำนวนแปลงที่ดินตัวแทน	1,303	5,182	1,688	จำนวนแปลงที่ดินตัวแทน	1,303	5,182	1,687
แปลงพาณิชยกรรม				แปลงพาณิชยกรรม			
MAPE - Linear Regression	194%	130%	187%	MAPE - Linear Regression	31%	24%	28%
MAPE - KNN-U	56%	37%	55%	MAPE - KNN-U	27%	23%	25%
MAPE - KNN-D	56%	37%	55%	MAPE - KNN-D	27%	23%	25%
จำนวนแปลงที่ดินตัวแทน	127	895	129	จำนวนแปลงที่ดินตัวแทน	127	895	129
แปลงที่อยู่อาศัยและพาณิชยกรรม				แปลงที่อยู่อาศัยและพาณิชยกรรม			
MAPE - Linear Regression	345%	261%	325%	MAPE - Linear Regression	31%	27%	32%
MAPE - KNN-U	38%	31%	33%	MAPE - KNN-U	23%	21%	23%
MAPE - KNN-D	38%	31%	33%	MAPE - KNN-D	23%	21%	23%
จำนวนแปลงที่ดินตัวแทน	1,430	6,077	1,817	จำนวนแปลงที่ดินตัวแทน	1,430	6,077	1,817
Baseline - แปลงที่อยู่อาศัยและพาณิชยกรรม				Baseline - แปลงที่อยู่อาศัยและพาณิชยกรรม			
MAPE - Linear Regression	27%	25%	26%	MAPE - Linear Regression	27%	25%	26%

การวัดผลการวิจัยเลือกใช้เมตริกซ์ Mean Absolute Percentage Error (MAPE) เนื่องจากค่าที่ได้มีลักษณะเป็นเปอร์เซ็นต์ทำให้ง่ายต่อการทำความเข้าใจ เช่น ราคากรรมสิทธิ์ 10,000 บาท แต่ราคาประเมิน 11,000 บาท แสดงว่า ประเมินราคาสูงไป 10% เป็นต้น ตัวอย่างในตาราง เช่น ค่า MAPE ของอัลกอริทึม Distance-based KNN (KNN-D) ภายใต้ “แปลงที่อยู่อาศัยและพาณิชยกรรม” โดยแปลงที่ดินตัวแทนหาจาก GetisOrd Gi* และใช้การประเมินราคาแบบใช้แปลงซื้อ-ขายจดทะเบียนบนหน่วยถนนเดียวกันมาสร้างแบบจำลอง มีค่าประมาณเท่ากับ 21% หมายความว่า โดยเฉลี่ยแล้ว แบบจำลองที่ได้จาก KNN-D นี้ประเมินราคาแปลงที่ดินตัวแทนทั้งหมดผิดพลาดไปจากราคาประเมินของกรรมสิทธิ์อยู่ที่ 21%

ส่วน Baseline ที่เห็นในตาราง หมายถึง MAPE ที่ได้จาก Linear Regression ซึ่งเป็นกรณีใช้ “การประเมินแบบมวลรวมโดยทั่วไป” เข้ามาร่วมเปรียบเทียบด้วย

สรุปและอภิปรายผลการวิจัย

จากผลการวิจัย สรุปได้ว่า แนวคิดที่ให้ผลลัพธ์ในแง่ของ MAPE ที่ดีที่สุด คือ แนวคิดการค้นหาแปลงที่ดินตัวแทนด้วยวิธี GetisOrd Gi* ผนวกกับแนวคิดการประเมินราคาแปลงที่ดินตัวแทนด้วยหลักการหน่วยถนนเดียวกัน ซึ่งได้ค่า MAPE อยู่ที่ 21% (ด้วยอัลกอริทึมประเภท KNN Regressor) นับว่าค่า MAPE ในระดับนี้ถือเป็นจุดเริ่มต้นที่ดี เนื่องจากสามารถเอาชนะ MAPE ที่ได้จากวิธีประเมินแบบมวลรวมโดยทั่วไป ซึ่งมีค่าเท่ากับ 25%

วิธีค้นหาแปลงที่ดินตัวแทนด้วย GetisOrd Gi* มีข้อดีตรงที่ว่า 1) ค้นหาแปลงที่ดินตัวแทนโดยมองหาความโดดเด่นของแปลงด้วยระยะที่ใกล้กับจุดสนใจต่างๆ (POI) เช่น แปลงที่ดินที่ใกล้กับห้างสรรพสินค้า หรือ ใกล้กับจุดขึ้น-ลงรถไฟฟ้า ควรเป็นตัวแทนที่ดีกว่าแปลงอื่นๆ ที่ไกลออกไป และ 2) จำนวนแปลงตัวแทนอ้างอิงที่มากกว่าวิธีอื่นทำให้แบบจำลองมีโอกาสเรียนรู้รูปแบบแพทเทิร์นได้ดีกว่า

ข้อเสนอแนะ

จากการวิเคราะห์เบื้องต้นเกี่ยวกับแปลงที่ดินตัวแทนที่มีค่าผิดพลาด MAPE สูงๆ พบว่าส่วนหนึ่งเกิดจากการพยายามปรับรอยต่อของราคาให้เรียบด้วยมือ (Manual interpolation) ยกตัวอย่างเช่น ราคาที่ประเมินมีความแตกต่างกับราคาโดยรอบอย่างมาก ทางผู้ประเมินของกรมธนารักษ์จึงจำเป็นต้องทำการปรับลดราคาลงมาด้วยมือ ดังนั้น “ทิศทางการราคา” จึงเป็นสิ่งที่ควรให้ความสำคัญในอนาคต นอกจากนี้ การพิจารณานำแหล่งข้อมูลเสนอซื้อ-เสนอขายในตลาดเข้ามาร่วมพิจารณาน่าจะช่วยแก้ปัญหาราคาซื้อ-ขายจดทะเบียนที่มีจำนวนน้อยและคุณภาพต่ำได้ดีขึ้น

บรรณานุกรม

- ธีรไนย ศรีธรรมรงค์. (2559). การประเมินพื้นที่เสี่ยงการเกิดอาชญากรรมด้วยการวิเคราะห์สถิติเชิงพื้นที่. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต : มหาวิทยาลัยศิลปากร.
- Ai, P. (2005). **Residential Land Value Modelling; Case Study of Hankou, China**. International Institute for geo-information and earth observation.
- Bao, H., Glascock, J., Zhou, .S, & Feng, L. (2014). **Land value determination in an emerging market: empirical evidence from China**. International Journal of Managerial Finance, 10(2), 180-199.
- ESRI. How Hot Spot Analysis (Getis-Ord Gi*) works. Retrieved May 6, 2021, from ArcGIS Pro website
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm>
- ESRI. Spatial Autocorrelation (Global Moran's I) (Spatial Statistics). Retrieved May 6, 2021, from ArcGIS Pro website
<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm>
- Ho, W. K. O., Tang, B. S., & Wong, S. W. (2020). **Predicting property prices with machine learning algorithms**. Journal of Property Research, 38(1) , 48– 70.
<https://doi.org/10.1080/09599916.2020.1832558>
- Kim, B., & Kim, T. (2016). **A Study on Estimation of Land Value Using Spatial Statistics: Focusing on Real Transaction Land Prices in Korea**. Sustainability, 8(3) , 203.
<https://doi.org/10.3390/su8030203>
- Lee, C. (2021). **Predicting Land Prices and Measuring Uncertainty by Combining Supervised and Unsupervised Learning**. International Journal of Strategic Property Management, 25(2), 169–178. <https://doi.org/10.3846/ijsp.2021.14293>
- Mie Oak Chae, Inhyuk Kwon. (March 19-23, 2018). **Korea's Mass Assessment System of Land**

Pricing for Taxation, Utilizing ICT, paper prepared for presentation at the “2018 WORLD BANK CONFERENCE ON LAND AND POVERTY” The World Bank – Washington DC.

Noa Ohana-Levi, Idan Bahat, Aviva Peeters, Alexandra Shtein, Yishai Netzer, Yafit Cohen, Alon Ben-Gal. (2019). **A weighted multivariate spatial clustering model to determine irrigation management zones**. Computers and Electronics in Agriculture 162 p.719-731.

Sukran Yalpir, Fatma Bunyan Unel. (2017). **Use of Spatial Analysis Methods in Land Appraisal; Konya Example**. 5th International Symposium on Innovative Technologies in Engineering and Science (ISITES2017 Baku - Azerbaijan).