ระบบประเมินราคาที่ดินจากแปลงที่ดินตัวแทนโดยใช้วิธีการเครื่องจักรเรียนรู้

เจตนา สมคำนึง

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
พ.ศ. 2564

LAND APPRAISAL PRICE PREDICTION SYSTEM FROM REFERENCE LOTS USING MACHINE LEARNING MODELS

JETANA SOMKAMNUENG

A Thematic Paper Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Engineering

Department of Big Data Engineering

College of Innovative Technology and Engineering

Dhurakij Pundit University

2021

หัวข้อสารนิพนธ์	ระบบประเมินราคาที่ดินจากแปลงที่ดินตัวแทนโดยใช้วิธีการเครื่องจักรเรียนรู้
ชื่อผู้เขียน	เจตนา สมคำนึง
อาจารย์ที่ปรึกษา	ดร.ธนภัทร ฆังคะจิตร
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2563

บทคัดย่อ

ในปัจจุบันภารกิจด้านการประเมินราคาที่ดินรายแปลงของกรมธนารักษ์ สังกัดกระทรวงการคลัง ได้มีการพัฒนารูปแบบวิธีการให้อยู่ภายใต้แนวคิดเศรษฐกิจดิจิทัล (Digital Economy) ด้วยการบริหารจัดการฐานข้อมูลขนาดใหญ่ของรัฐบาลไทย ซึ่งเป็นการสอดคล้องกับบริบทของแผนพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม ยุทธศาสตร์ที่ 4 การปรับเปลี่ยนภาครัฐสู่การเป็นรัฐบาลดิจิทัล อย่างไรก็ดี การเปลี่ยนผ่านจากวิธีการเดิมที่ต้องอาศัยประสบการณ์ของผู้ประเมินมาใช้วิธีการเรียนรู้ของเครื่องนั้นมีขั้นตอนกระบวนการที่มาก รวมถึงต้องมีการเปลี่ยนแปลงแนวคิดในหลายๆ จุด งานฉบับนี้จึงได้นำเสนอแนวทางเบื้องต้นในการพัฒนาแบบจำลองการประเมินราคาที่ดินที่ผสมผสานแนวคิดการวิเคราะห์ทางด้านวิทยาศาสตร์ข้อมูลร่วมกับความรู้และประสบการณ์ของผู้เชี่ยวชาญ โดยมีวัตถุประสงค์เพื่อช่วยส่งเสริมประสิทธิภาพการทำงานทางด้านการประเมินราคาที่ดินให้มีความทันสมัยอธิบายได้ ในสารนิพนธ์นี้นำเสนอการประเมินราคาที่ดินผ่านแปลงที่ดินตัวแทนซึ่งพิจารณาจากแปลงที่ดินที่ตั้งอยู่ใกล้กับจุดสนใจ (Points of interest) ต่างๆ ส่งผลให้แปลงที่ดินตัวแทนที่ได้มาจากวิธีนี้มีความโดดเด่นในแง่ของสถานที่ตั้ง (Location) โดยนำวิธีการทางสถิติเชิงพื้นที่เรื่องการวิเคราะห์จุดความร้อนด้วยวิธี GetisOrd Gi* เข้ามาช่วยวิเคราะห์ความใกล้-ไกลจากแต่ละแปลงที่ดินไปยังแต่ละจุดสนใจ ผลการทดลองแสดงให้เห็นว่า แบบจำลองสามารถประเมินราคาที่ดินตัวแทนได้ดีกว่าแบบจำลองมวลรวมที่ใช้กันโดยทั่วไป (MAPE 21% vs 25%) อีกทั้งยังสามารถอธิบายเหตุผลเบื้องหลังได้ดี ทำให้ไม่ส่งผลกระทบในเชิงการเปลี่ยนแปลงอย่างฉับพลันจากรูปแบบการประเมินราคาแบบเดิมมากจนเกินไป ดังนั้น บุคลากรที่เกี่ยวข้องจึงสามารถนำแบบจำลองไปประยุกต์ใช้งานได้ในระยะเวลาอันสั้น

Thematic Paper Title	LAND APPRAISAL PRICE PREDICTION SYSTEM FROM REFERENCE LOTS USING MACHINE LEARNING MODELS
Author	Jetana Somkamnueng
Thematic Paper Advisor	Dr.Thanapat Kangkachit
Department	Big Data Engineering
Academic Year	2020

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 explicability. 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 can easily follow along and apply the main theme of the system in a short period of time.

กิตติกรรมประกาศ

สารนิพนธ์เรื่อง “ระบบประเมินราคาที่ดินสำหรับแปลงที่ดินตัวแทนโดยใช้วิธีการเครื่องจักรเรียนรู้” สำเร็จลุล่วงตามวัตถุประสงค์ที่ตั้งไว้ด้วยดี ผู้เขียนขอกราบขอบพระคุณอาจารย์ที่ปรึกษา ดร.ธนภัทร ฆังคะจิตร เป็นอย่างยิ่งในการเคี่ยวเข็ญและเสียสละเวลาอันมีค่าของท่านในการให้คำแนะนำในทุกขั้นตอน

ผู้เขียนขอขอบพระคุณ ผศ.ดร.ธนาวินท์ รักธรรมานนท์ อาจารย์ประจำคณะวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์สำหรับคำแนะนำในการเรื่องงานวิจัยที่เกี่ยวข้อง ขอขอบพระคุณ คุณทวีศักดิ์ อภิวัฒน์พงค์ ในฐานะผู้เชี่ยวชาญพิเศษเกี่ยวกับการประเมินราคาทรัพย์สิน โดยเฉพาะอย่างยิ่งคำแนะนำและการอธิบายเกี่ยวกับการวิเคราะห์งานทางด้านภูมิศาสตร์ รวมทั้ง ดร.ดิษย์ มณีพิทักษ์ และเจ้าหน้าที่ที่กองประเมินราคาทรัพย์สิน กรมธนารักษ์ กระทรวงการคลังที่อำนวยความสะดวกในการเข้าถึงข้อมูลที่เกี่ยวข้องเป็นอย่างดี

นอกจากนี้ ขอกราบขอบพระคุณ ดร.สรรพฤทธิ์ มฤคทัต เป็นอย่างสูงที่กรุณาให้เกียรติมาเป็นประธานและให้คำแนะนำจนสารนิพนธ์นี้สมบูรณ์ยิ่งขึ้น ขอขอบพระคุณ ดร.เอกสิทธิ์ พัชรวงศ์ศักดา และ ผศ.ดร. ดวงใจ จิตต์คงชื่น ที่ประสิทธิ์ประสาทวิชาต่างๆ จนมีความรู้เพียงพอในการทำให้สารนิพนธ์นี้สำเร็จเสร็จสิ้น ตลอดจน นางสาวกุลธิดา รอดบุญ รวมถึงเจ้าหน้าที่ที่เกี่ยวข้องทุกท่านในด้านการอำนวยความสะดวกจนเสร็จการ

สุดท้ายนี้ ผู้เขียนขอกราบขอบพระคุณบิดา มารดา และภรรยาที่เป็นกำลังใจและเตือนสติในยามท้อถอย ประโยชน์อันใดที่พึงมีจากการศึกษาครั้งนี้ขอมอบแด่ผู้มีพระคุณทุกๆ ท่าน

เจตนา สมคำนึง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ฆ
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ	จ
สารบัญตาราง	ซ
สารบัญภาพ	ฌ
บทที่	
1. บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	3
1.3 ประโยชน์ที่คาดว่าจะได้รับ	3
1.4 ขอบเขตงานวิจัย	3
1.5 นิยามศัพท์	4
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	7
2.1 ทฤษฎีที่เกี่ยวข้อง	7
2.2 งานวิจัยที่เกี่ยวข้อง	16
3. ระเบียบวิธีวิจัย	30
3.1 ข้อมูลนำเข้า	32
3.2 อุปกรณ์และซอฟต์แวร์ที่ใช้ในการศึกษาวิจัย	32
3.3 วิธีการศึกษาวิจัย	32
4. ผลการศึกษาวิจัย	50
4.1 ลักษณะทางกายภาพของพื้นที่ต้นแบบ (เขตจตุจักร กรุงเทพมหานคร)	50
4.2 ผลจากการสำรวจข้อมูล	52

สารบัญ (ต่อ)

บทที่		หน้า
	4.3 ผลจากการทำ One-hot-encoding ข้อมูลเชิงประเภท	53
	4.4 รูปแปลงที่ดินหลังจากทำความสะอาดข้อมูล	53
	4.5 ผลจากการวิเคราะห์ตัวแปรเพื่อหาตัวแปรที่มีความสำคัญ โดยวิธี Exploratory Regression	54
	4.6 ผลจากการวิเคราะห์ตัวแปรเพื่อหาตัวแปรที่มีความสำคัญ โดยวิธี Forest-based Regression	55
	4.7 ผลการจัดกลุ่มแปลงที่ดิน	57
	4.8 ผลการคัดเลือกแปลงที่ดินตัวแทนของแต่ละหน่วยถนน	57
	4.9 ผลจากการค้นหาตัวแปรที่สำคัญสำหรับแบบจำลองประเมินราคาแปลงที่ดินตัวแทน	59
	4.10 ผลการประเมินราคาแปลงที่ดินตัวแทนจากแบบจำลอง	61
5.	บทสรุปและข้อเสนอแนะ	69
	บรรณานุกรม	72
	ภาคผนวก	77
	ภาคผนวก ก. รายละเอียดชั้นข้อมูลนำเข้า	78
	ภาคผนวก ข. รายละเอียดการสำรวจข้อมูลและทำความสะอาดข้อมูล	83
	ภาคผนวก ค. รายละเอียดการจัดกลุ่มแปลงที่ดิน	105
	ประวัติผู้เขียน	111

สารบัญตาราง

ตารางที่		หน้า
2.1	สรุปผลกระทบของแต่ละปัจจัยต่อราคาที่ดินประเภทที่อยู่อาศัย.....	27
2.2	สรุปผลกระทบของแต่ละปัจจัย.....	29
3.1	แสดงรายการข้อมูลนำเข้าที่ใช้ในสารนิพนธ์นี้.....	32
4.1	แสดงผลการประเมินราคาของแปลงที่ดินตัวแทนในภาพรวม ตามแนวทางที่ 1.....	63
4.2	แสดงผลการประเมินราคาของแปลงที่ดินตัวแทนในภาพรวม ตามแนวทางที่ 2.....	64

สารบัญภาพ

ภาพที่		หน้า
2.1	แสดงรูปแบบการกระจายตัวเชิงพื้นที่แบบต่างๆ	9
2.2	แสดงผลลัพธ์จาก GetisOrd Gi* ในรูปแบบแผนที่	10
2.3	แสดงสถิติเชิงพื้นที่	11
2.4	แสดงผลลัพธ์ก่อนและหลังจากการจัดกลุ่มด้วย SCMC (ที่มา: ESRI)	12
2.5	แสดงรูปแบบความสัมพันธ์เชิงด้าน	12
2.6	แสดงรูปแบบความสัมพันธ์เชิงมุม	13
2.7	แสดงการเลือกเพื่อนบ้านแบบ Trimmed Delaunay Triangulation	13
2.8	แสดงรูปแบบการวัดระยะห่างแบบ Euclidean และ Manhattan	15
2.9	แสดงจำนวนแปลงเพื่อนบ้านที่ถูกกำหนดเป็นค่า K เริ่มต้นของ KNN ในทางภูมิศาสตร์	16
2.10	แสดงแบบจำลองที่นำเสนอโดยงานวิจัยของ Sukran Yalpir และ Fatma Bunyan Unel, 2017	17
2.11	แสดงปัจจัยต่างๆ ที่ใช้ฝึกแบบจำลองนำเสนอโดย Sukran Yalpir และ Fatma Bunyan Unel, 2017	18
2.12	แสดงกระบวนการตามแบบจำลอง iLVM ที่นำเสนอโดย Jannet C. Bencure, 2019	20
2.13	แสดงปัจจัยนำเข้าเพื่อใช้ในแบบจำลอง iLVM ที่นำเสนอโดย Jannet C. Bencure, 2019	20
2.14	แสดงค่า Quasi-neighbor ที่นำเสนอในงานวิจัยของ Waights, 2018	22
2.15	แสดงราคาที่ดินของ 105 หัวเมืองในจีน ในปี 2008, 2011, และ 2014 (Yang et al., 2017)	24
3.1	แสดงฮิสโทแกรมของตัวแปร “ความกว้างของถนน หรือ STREET_WID”	33
3.2	Count plot ของตัวแปร “การใช้ประโยชน์ที่ดิน หรือ LAND_USE”	33

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.3	ขั้นตอนการทำความสะอาดข้อมูลพร้อมแสดงจำนวนแปลงที่ดินที่เหลือจากแต่ละขั้นตอน	34
3.4	แสดงข้อผิดพลาด	37
3.5	การคัดกรองแปลงที่ดินที่ไม่เป็นรูป 4 เหลี่ยมเนื่องจากมีผลต่างพื้นที่ของแปลงที่ดินกับ MBR เกิน 5%	37
3.6	การเลือกช่วงอัตราส่วน DEPTH_WR ของแปลงส่วนใหญ่ในกลุ่ม	38
3.7	แสดงการเลือกช่วงพื้นที่ (Area) ของแปลงส่วนใหญ่ในกลุ่มที่สอดคล้องกับอัตราส่วน DEPTH_WR และ DEPTH_RW	39
3.8	แสดงขั้นตอนการคัดเลือกแปลงที่ดินตัวแทนด้วยการวิเคราะห์จุดความร้อน	40
3.9	แสดงตัวอย่างการคำนวณค่าเฉลี่ยความใกล้-ไกลถ่วงน้ำหนักสำหรับแต่ละแปลงที่ดิน	41
3.10	แสดงตัวอย่างการคำนวณค่าเฉลี่ยความใกล้-ไกลถ่วงน้ำหนักสำหรับแต่ละแปลงที่ดิน	42
3.11	แสดงตัวอย่างการวัดระยะทาง (ระยะขจัด) จากแปลงที่ดินไปยังแต่ละ POI เฉพาะตัวที่ใกล้ที่สุด	43
3.12	แสดงผลลัพธ์จากการรัน GetisOrd Gi* สำหรับคลัสเตอร์ที่ 1	44
3.13	แสดงการหาช่วงของข้อมูลที่มีความถี่สูงสุดของ 2 ตัวแปรเพื่อนำมากำหนดเงื่อนไขในการเลือกแปลงที่ดินตัวแทน	45
3.14	แสดงการเลือกค่า K โดยอัตโนมัติด้วยวิธี Elbow และ Silhouette Score ตามลำดับ	47
3.15	แสดงการ buffer จากแปลงที่ดินตัวแทนไปหาแปลงซื้อ-ขายจดทะเบียนจำนวน 5 แปลง	48

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3.16	แสดงหน่วยถนนที่ “มี” แปลงซื้อ-ขายจดทะเบียน.....	49
4.1	แสดงรูปแปลงที่ดินเขตจตุจักร กรุงเทพมหานคร.....	51
4.2	แสดงผลจากการทำ One-hot-encoding ข้อมูลเชิงประเภทเพื่อให้พร้อมสำหรับการวิเคราะห์.....	53
4.3	รูปแปลงที่ดินหลังจากทำความสะอาดข้อมูล.....	53
4.4	แสดงผลการรัน Exploratory Regression เพื่อค้นหาตัวแปรที่สำคัญ.....	54
4.5	แสดงผลการรัน Forest-based Classification and Regression เพื่อค้นหาตัวแปรที่สำคัญ.....	55
4.6	แสดงผลจากการแบ่งกลุ่มแปลงที่ดินโดยใช้วิธี Spatially Constrained Multivariate Clustering.....	56
4.7	แสดงตัวอย่างแปลงที่ดินตัวแทนที่ค้นหาได้จากองค์ความรู้ของผู้เชี่ยวชาญและลักษณะข้อมูล.....	57
4.8	แสดงตัวอย่างแปลงที่ดินตัวแทนที่ค้นหาได้จากวิธี GetisOrd Gi*.....	58
4.9	แสดงตัวอย่างแปลงที่ดินตัวแทนที่ค้นหาได้จากวิธีเปรียบเทียบเชิงกายภาพ.	59
4.10	แสดง “ตัวแปรสำคัญ” ที่ได้จาก Random Forest Regressor.....	60
4.11	แสดง “ตัวแปรสำคัญ” ที่ได้จาก Extra Trees Regressor.....	60
4.12	แสดง “ตัวแปรสำคัญ” ที่ได้จาก Gradient Boosting Regressor.....	61
4.13	แสดงตัวอย่างการประเมินราคาแปลงที่ดินตัวแทนด้วย distanced-based KNN (K=3) ตามแนวทางที่ 1.....	62

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
4.14	แสดงตัวอย่างผลการประเมินราคาแปลงที่ดินตัวแทนด้วย distanced-based KNN (K=3) ตามแนวทางที่ 2.....	64
5.1	แสดงตัวอย่าง Scatter Plot ระหว่าง “LAND_VALUE” และ “Percentage Error” (แปลงที่ดินตัวแทนได้จากวิธี GetisOrd Gi*).....	70
5.2	แสดงตัวอย่าง Scatter Plot ระหว่าง “LAND_VALUE” และ “Percentage Error” (แปลงที่ดินตัวแทนได้จากวิธีผู้เชี่ยวชาญ).....	71

บทที่ 1

บทนำ

1.1. ที่มาและความสำคัญของปัญหา

ด้วยพระราชบัญญัติภาษีที่ดินและสิ่งปลูกสร้าง พ.ศ. 2562 มีบทบัญญัติให้ใช้บัญชีกำหนดราคาประเมินทุนทรัพย์ของอสังหาริมทรัพย์ที่กรมธนารักษ์จัดทำขึ้นเป็นฐานในการคำนวณภาษีที่ดินและสิ่งปลูกสร้าง ซึ่งเป็นการขยายขอบเขตการปฏิบัติงานของกรมธนารักษ์ที่นอกเหนือจากการประเมินราคาทรัพย์สิน เพื่อการจดทะเบียนสิทธิและนิติกรรมตามประมวลกฎหมายที่ดิน พ.ศ. 2497 (และที่แก้ไขเพิ่มเติม) โดยปัจจุบันภารกิจด้านการประเมินราคาที่ดินของกรมธนารักษ์ ได้มีการพัฒนารูปแบบการบริหารจัดการภายใต้แนวคิดเศรษฐกิจดิจิทัล (Digital economy) ของรัฐบาลโดยนำการจัดการฐานข้อมูลข้อมูลขนาดใหญ่ (Big data) เข้ามาบูรณาการ ซึ่งเป็นการสอดคล้องกับบริบทของแผนพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม ยุทธศาสตร์ที่ 4

แนวคิดการประเมินราคาทรัพย์สินตามมาตรฐานสากลนั้นมีอยู่ 6 วิธี (ที่มา: มูลนิธิประเมินค่า-นายหน้าแห่งประเทศไทย (องค์การสาธารณประโยชน์)) ได้แก่

1) วิธีวิเคราะห์มูลค่าจากต้นทุน (Cost approach to value)
2) วิธีเปรียบเทียบตลาด (Market comparison approach)
3) วิธีแปลงรายได้เป็นมูลค่า (Income approach)
4) วิธีการตั้งสมมติฐานในการพัฒนา (Hypothetical development approach or residual technique)
5) วิธีวิเคราะห์กระแสเงินสด (Discounted cashflow analysis)
6) วิธีประเมินโดยใช้แบบจำลอง (Computer-assisted mass appraisal)

ในปัจจุบัน กองประเมินราคาทรัพย์สิน กรมธนารักษ์ได้ยึดหลักวิธีเปรียบเทียบตลาดในการประเมินราคาที่ดิน เนื่องจากมีความชัดเจนเพราะการซื้อ-ขายเกิดขึ้นแล้วจริง จึงสามารถอธิบายได้โดยง่าย ในเชิงเศรษฐศาสตร์ ที่ดินถือเป็น “ทุน” ต้นน้ำในการประกอบธุรกิจต่างๆ ทำให้การนำวิธีการประเมินแบบอื่นมาใช้ก่อให้เกิดปัญหาในเชิงเปรียบเทียบ ยกตัวอย่างเช่น วิธีวิเคราะห์มูลค่าจากต้นทุนนั้นใช้การรวมต้นทุนในการได้มาซึ่งทรัพย์สินมากำหนดเป็นราคาประเมิน แต่ที่ดินทำได้แค่เพียงการเทียบเคียงกับที่ดินผืนอื่นเท่านั้นสร้าง

ทดแทนไม่ได้ หรือ กรณีวิธีแปลงรายได้เป็นมูลค่า ซึ่งมีสมมติฐานทางธุรกิจหลากหลาย ไม่ว่าจะเป็นอัตราคิดลด (ดอกเบี้ย) การแกว่งตัวของรายได้ในแต่ละช่วงเวลา ก่อให้เกิดปัญหาความโปร่งใสในการกำหนดราคาประเมิน

นอกจากนี้ การประเมินที่ดินที่ยึดหลักการอ้างอิงข้อมูลธุรกรรมการซื้อ-ขายที่เกิดขึ้นในตลาดนั้น ควรพิจารณาถึงการใช้ประโยชน์สูงสุดของทรัพย์สิน (Highest and best use) หมายถึง การใช้ประโยชน์ในทรัพย์สินให้ดีที่สุดเท่าที่จะเป็นไปได้ภายใต้เงื่อนไขที่เป็นไปตามตัวบทกฎหมาย ความเหมาะสมในทางกายภาพ และความเป็นไปได้ทางการตลาดและทางการเงิน ยิ่งไปกว่านั้น ข้อมูลตลาดหรือหลักฐานการซื้อ-ขายในตลาดที่นำมาใช้ในการประเมินราคาที่ดิน ควรมีทำเลที่ตั้งอยู่ในบริเวณเดียวกันหรือใกล้เคียงกัน มีลักษณะหรือสภาพที่สามารถเปรียบเทียบกันได้ ตลอดจนมีการใช้ประโยชน์สูงสุดคล้ายคลึงกัน (Geospatial characteristics) ซึ่งเป็นไปตามมาตรฐานวิชาชีพการประเมินมูลค่าทรัพย์สินเรื่องที่ 2 หลักเกณฑ์การประเมินเพื่อกำหนดมูลค่าตลาด

ปัญหาสำคัญอย่างหนึ่งในการได้มาซึ่งข้อมูลธุรกรรมการซื้อ-ขายที่ดินในประเทศไทย คือ ไม่ทราบราคาซื้อ-ขายจดทะเบียนที่แท้จริง เนื่องจากไม่มีข้อกฎหมายบังคับ ดังนั้น ผู้ซื้อและผู้ขายจึงมีแนวโน้มที่จะแจ้งราคาซื้อ-ขายในระดับที่ต่ำกว่าความเป็นจริง (ส่งผลให้ราคาประเมินที่ดินที่ประกาศใช้โดยกรมธนารักษ์ต่ำตามไปด้วย) เพื่อประโยชน์ในการชำระภาษีที่เกี่ยวข้องให้น้อยลง ส่งผลให้รัฐบาลเก็บภาษีได้ไม่เต็มเม็ดเต็มหน่วย ดังนั้น จึงมีแนวคิดที่จะนำข้อมูลตลาดแบบอื่นมาใช้ เช่น ราคาเสนอซื้อ หรือ ราคาเสนอขายที่เกิดขึ้นบนเว็บไซต์ หรือ ป้ายโฆษณาต่างๆ เป็นต้น อย่างไรก็ดี แนวคิดดังกล่าวยังอยู่ในระยะเบื้องต้น ยังจำเป็นต้องหาข้อสรุปรวมทั้งหารือกับหน่วยงานต่างๆ ที่เกี่ยวข้องอีกหลายขั้นตอน สารนิพนธ์ฉบับนี้เพียงแค่นำเสนอแนวทางเบื้องต้นในการพัฒนาแบบจำลองการประเมินราคาที่ดิน เพื่อศึกษาความเป็นไปได้และเพื่อสะท้อนให้เห็นถึงความแตกต่างที่เกิดขึ้นระหว่างวิธีแบบเก่าและวิธีแบบใหม่ ซึ่งจะส่งผลต่อการพัฒนาโครงการที่เกี่ยวข้องในอนาคตต่อไป

การประเมินราคาทรัพย์สินประเภทที่ดินแบ่งออกได้หลายกรณี เช่น การประเมินราคาที่ดินประเภทที่อยู่อาศัย, การประเมินราคาที่ดินประเภทพาณิชยกรรม, การประเมินราคาที่ดินประเภทอุตสาหกรรม, การประเมินราคาที่ดินประเภทเกษตรกรรม, การประเมินราคาที่ดินประเภทหมู่บ้านจัดสรรหรือคอนโดมิเนียม, การประเมินราคาที่ดินประเภทที่รกร้างว่างเปล่า เป็นต้น เนื่องจากในสารนิพนธ์ฉบับนี้มุ่งเน้นการประเมินราคาที่ดินในเขตเมืองทำให้ขอบเขตส่วนใหญ่ตกอยู่ในที่ดินประเภทที่อยู่อาศัย โดยมีที่ดินประเภทพาณิชยกรรมบ้างถือเป็นส่วนน้อย

สารนิพนธ์ฉบับนี้ได้นำเทคนิคการเรียนรู้ด้วยเครื่องเข้ามาทำแบบจำลองหลายเทคนิค ได้แก่ เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) สำหรับการแบ่งกลุ่มแปลงที่ดินออกเป็นคลัสเตอร์เชิงพื้นที่

และเทคนิคการเรียนรู้แบบมีผู้สอน (Supervised learning) เพื่อประเมินราคาแปลงที่ดินตัวแทน โดยผลการทดลองได้ผลลัพธ์สุดท้ายอยู่ในระดับที่น่าพอใจ โดยมีค่า Mean absolute percentage error (MAPE) อยู่ที่ 21%

1.2. วัตถุประสงค์ของการวิจัย

1.2.1. เพื่อออกแบบและพัฒนา “แบบจำลองต้นแบบ” ในการประเมินราคาที่ดินตัวแทน

1.2.2. เพื่อพัฒนา “ระบบตรวจสอบ” ความถูกต้องของราคาประเมินที่ได้มาจากแบบจำลอง

1.3. ประโยชน์ที่คาดว่าจะได้รับ

1.3.1. ช่วยลดภาระหน้าที่ของนักประเมินราคา ทำให้สามารถประเมินราคาที่ดินได้ในเวลาที่สั้นลงกว่าเดิม โดยมีแบบแผนเดียวกันทั้งประเทศและเป็นที่ยอมรับในสากล

1.3.2. ช่วยให้ราคาประเมินที่ดินมีความโปร่งใส อธิบายได้ และมีเหตุมีผล

1.4. ขอบเขตงานวิจัย

1.4.1. ศึกษาแนวคิดจากงานวิจัยทั้งในและต่างประเทศ เพื่อนำมาใช้พัฒนารูปแบบความเป็นไปได้ในการประเมินราคาที่ดินแปลงตัวแทนด้วยวิธีการเรียนรู้ของเครื่อง (Machine learning)

1.4.2. จัดเตรียมข้อมูลปัจจัย ตัวแปร และจัดทำโครงสร้างปัจจัยตัวแปรที่สำคัญ เพื่อให้สามารถนำมาใช้ในการพัฒนาแบบจำลองต้นแบบได้อย่างเหมาะสม

1.4.3. ศึกษาและกำหนดว่าปัจจัยใดที่มีความสำคัญในการวิเคราะห์ข้อมูลด้วยวิธีการทางวิทยาศาสตร์ข้อมูล รวมทั้งปรึกษาหารือกับกรมธนารักษ์ในการสร้างรูปแบบแนวทางในการพัฒนาแบบจำลองให้มีมาตรฐานสากล

1.4.4. พัฒนาแอพพลิเคชันเบื้องต้นในการค้นหาแปลงที่ดินตัวแทนและการกำหนดราคาแปลงที่ดินตัวแทน

1.4.5. พื้นที่ต้นแบบที่ใช้พัฒนาแบบจำลอง คือ พื้นที่เขตจตุจักร กรุงเทพมหานคร ซึ่งอยู่ในเขตเมือง โดยส่วนใหญ่เป็นแปลงที่ดินประเภทที่อยู่อาศัย

1.4.6. แฟ้มข้อมูลตั้งต้น เช่น แปลงที่ดิน, แปลงถนน, แปลงซื้อ-ขายจดทะเบียน, แปลงบล็อกฟิกซ์, แปลงบล๊อคบลู เป็นต้น จะเป็นไฟล์นามสกุล .shp (shape file) ซึ่งเป็นไฟล์สำหรับงาน GIS โดยเฉพาะ

1.4.7. ราคาประเมินที่ดินจากกรมธนารักษ์ (เพื่อเปรียบเทียบผลลัพธ์กับแบบจำลอง) เป็นราคาที่ประกาศใช้ตั้งแต่ปี พ.ศ. 2559 จนถึงปัจจุบัน (คาดว่าน่าจะกำลังประกาศใช้ราคาตัวใหม่ภายในปี พ.ศ. 2564 หรือ 2565)

1.4.8. ราคาซื้อ-ขายจดทะเบียนจากกรมที่ดิน เป็นราคาซื้อ-ขายในช่วงปี พ.ศ. 2555 จนถึง พ.ศ. 2558 เพื่อให้สอดคล้องกับราคาประเมินที่ประกาศใช้

1.4.9. แปลงที่ดินที่ประเมินจะไม่รวม แปลงที่ดินบล๊อคบลู, แปลงที่ดินบล็อกฟิกซ์ และแปลงที่ดินที่ไม่มีทางเข้า-ออก (ที่ตาบอด)

1.4.10. ใช้ภาษา python และโปรแกรม ArcGIS Pro ในการพัฒนาแบบจำลอง

1.5. นิยามศัพท์

1.5.1. **Shape File** หมายถึง ไฟล์ที่มีนามสกุล .shp ซึ่งใช้ในระบบสารสนเทศภูมิศาสตร์ (GIS) โดยใน 1 ชุดไฟล์จะประกอบไปด้วยไฟล์นามสกุลอื่นอีก ได้แก่ .cpg, .prj, .dbf, .shx, .sbn

1.5.2. **Coordinate System** หมายถึง ระบบพิกัดภูมิศาสตร์ที่ใช้อ้างอิงกันทั่วโลก มี 2 แบบ คือ Geographic coordinate system ซึ่งมีหน่วยเป็นองศาละติจูดและองศาลองกิจูด และ Projected Coordinate system ซึ่งอ่านค่าเป็นเมตร โดยไฟล์ตั้งต้นทุกไฟล์ที่ใช้ในเอกสารฉบับนี้จะเป็นพิกัดแบบหลัง และกำหนดค่าเป็น UTM 24047

1.5.3. **แปลงที่ดินบล๊อคบลู (Block blue)** หมายถึง ชุดแปลงที่ดิน (เลขที่โฉนดต่างกัน) ที่ใช้ประโยชน์ร่วมกัน เช่น แปลงที่ดินบริเวณมหาวิทยาลัยเกษตรศาสตร์

1.5.4. **แปลงที่ดินบล็อกฟิกซ์ (Block fix)** หมายถึง ชุดแปลงที่ดินที่ถูกพัฒนาไปเป็น “หมู่บ้านจัดสรร” หรือ “คอนโดมิเนียม”

1.5.5. **แปลงที่ดินตัวแทน (Reference lots)** หมายถึง แปลงที่ดินที่มีคุณสมบัติที่ดีตามข้อกำหนดเพื่อเป็นตัวแทนของแปลงที่ดินโดยรอบ (เช่น บนหน่วยถนนเดียวกัน) ซึ่งจะถูกนำไปประเมินราคาแปลงที่ดินรายแปลงต่อไป

1.5.6. **หน่วยถนน** หมายถึง การระบุชื่อถนน ซอย ทาง แม่น้ำ ลำคลองฯลฯ เพื่อให้ทราบถึงตำแหน่งที่ตั้งของแปลงที่ดิน ซึ่งต้องตรงกับชื่อที่ปรากฏในระวางแผนที่ (นิยามจากกรมธนารักษ์)

1.5.7. **การใช้ประโยชน์ที่ดิน** หมายถึง การใช้ประโยชน์ในภาพรวมของแปลงที่ดินหรือหน่วยถนนนั้น เช่น พาณิชยกรรม ที่อยู่อาศัย สถานที่ราชการ เกษตรกรรม หรืออุตสาหกรรม เป็นต้น

1.5.8. **ผิวจราจร** หมายถึง สภาพพื้นผิวจราจรของแต่ละแปลงที่ดินหรือหน่วยถนนนั้น เช่น คอนกรีต หน้าแอสฟัลต์ คอนกรีต ลาดยาง หินคลุก ลูกรัง ดิน ไม้ น้ำ หรือไม่มีสภาพ เป็นต้น

1.5.9. **FAR หรือ Floor Area Ratio** หมายถึง อัตราส่วนพื้นที่อาคารรวมต่อพื้นที่ดิน เป็นตัวกำหนดว่าจะสามารถสร้างตึกหรืออาคารได้ขนาดเท่าไร ยกตัวอย่างเช่น ที่ดิน 100 ตร.ว. (หรือ 400 ตร.ม.) หากกำหนดค่า FAR = 6 แสดงว่า พื้นที่อาคารสูงสุดที่สร้างได้ คือ 6x400 = 2,400 ตร.ม.

1.5.10. **OSR หรือ Open Space Ratio** หมายถึง อัตราส่วนของที่ว่างปราศจากสิ่งปกคลุมต่อพื้นที่อาคารรวม ยกตัวอย่างเช่น ที่ดิน 100 ตร.ว. (หรือ 400 ตร.ม.) สร้างอาคารมีพื้นที่ 2,400 ตร.ม. เอาไว้ หากกำหนดค่า OSR = 5 แสดงว่า ต้องเหลือผืนดินโล่งๆ ที่ปราศจากสิ่งปกคลุมหรือหลังคาอย่างน้อยเท่ากับ 120 ตารางเมตร (2,400x5/100)

1.5.11. **Points of Interest หรือ POI** หมายถึง ตำแหน่งหรือจุดที่น่าสนใจโดยมีลักษณะ geometry เป็น point type เช่น POI ห้างเซ็นทรัลลาดพร้าว, POI มหาวิทยาลัยเกษตรศาสตร์, POI 7-11 เป็นต้น

1.5.12. **ArcGIS Pro** หมายถึง โปรแกรมสำเร็จรูปที่สร้างขึ้นโดย Esri (Environmental systems research institute) มีสำนักงานใหญ่อยู่ที่รัฐแคลิฟอร์เนีย ประเทศสหรัฐอเมริกา เป็นโปรแกรมที่ใช้ประมวลผลและวิเคราะห์ข้อมูลเชิงพื้นที่โดยเฉพาะ

1.5.13. **GetisOrd Gi*** หมายถึง เครื่องมือทางสถิติเชิงพื้นที่แบบเฉพาะที่ (Local) ใช้สำหรับวิเคราะห์ค่าจุดความร้อนเพื่อหาบริเวณที่มีนัยสำคัญ

1.5.14. **MAPE หรือ Mean Absolute Percentage Error** หมายถึง ค่าเฉลี่ยร้อยละความผิดพลาดสัมบูรณ์ เป็นวิธีวัดความแม่นยำของแบบจำลองประเภทรีเกรสชัน (Regression) โดยคำนวณเป็นเปอร์เซ็นต์ความผิดพลาดในการพยากรณ์ โดยไม่คำนึงถึงเครื่องหมาย

1.5.15. **R^2 หรือ R-squared** หมายถึง ค่าสถิติที่ใช้ชี้วัดว่าตัวแปรต้นที่ใช้ในแบบจำลองสามารถอธิบายความผันแปรของตัวแปรตามได้มากหรือน้อยเพียงใด มีขอบเขตอยู่ในช่วงตั้งแต่ 0 ถึง 1 โดยค่าที่เข้าใกล้ 1 บ่งบอกว่าตัวแปรต้นสามารถอธิบายตัวแปรตามได้ดี

1.5.16. **K-Means** หมายถึง อัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) ที่ใช้สำหรับจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันไว้ด้วยกันโดยอาศัยหลักการ “ค่าเฉลี่ย (Means)”

1.5.17. **K-Medoids** หมายถึง อัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) ที่ใช้สำหรับจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันไว้ด้วยกัน แต่แทนที่จะคำนวณตำแหน่ง centroid ของแต่ละกลุ่มด้วยค่าเฉลี่ยแบบวิธี K-Means วิธีนี้จะเลือกตำแหน่ง centroid ด้วยค่าจริงๆ ที่อยู่ในกลุ่มตัวอย่างแทน

1.5.18. **K-Prototypes** หมายถึง อัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) ที่ใช้สำหรับจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันไว้ด้วยกัน เป็นอัลกอริทึมที่ผสมผสานวิธี K-Means และวิธี K-Modes เข้าด้วยกัน เหมาะสำหรับชุดข้อมูลที่มีแอตทริบิวต์ทั้งแบบตัวเลข (Numerical) และแบบเชิงประเภท (Categorical) รวมอยู่ด้วยกัน มีวิธีการทำงานคล้าย K-Means แตกต่างกันตรงวิธีการปรับตำแหน่งของ centroid ในแต่ละกลุ่ม ซึ่งจะถูกแยกคิด โดยถ้าเป็นแอตทริบิวต์เชิงตัวเลขยังคงใช้ค่าเฉลี่ยเหมือนเดิม แต่ถ้าเป็นแอตทริบิวต์เชิงประเภทจะเปลี่ยนไปใช้ค่าฐานนิยม (Mode) แทน

1.5.19. **Spatially Constrained Multivariate Clustering หรือ SCMC** หมายถึง อัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) เช่นเดียวกัน แต่เป็นวิธีการแบ่งกลุ่มที่คำนึงถึงความสัมพันธ์เชิงพื้นที่ระหว่างกัน โดยมีสมมติฐานว่า แปลงที่ดินที่อยู่ในบริเวณเดียวกันหรือใกล้เคียงกันเท่านั้น จึงจะถูกจัดให้อยู่ในกลุ่มเดียวกัน มิเช่นนั้นแล้ว จะถูกแบ่งออกเป็นอีกกลุ่มหนึ่ง

1.5.20. **Linear Regression** หมายถึง อัลกอริทึมแบบมีผู้สอน (Supervised learning) ที่ใช้สำหรับวิเคราะห์การถดถอย โดยศึกษาความสัมพันธ์ระหว่างชุดตัวแปรต้น ($X_1, X_2, \ldots, X_n$) กับ ตัวแปรตาม (y)

1.5.21. **KNN Regressor** หมายถึง อัลกอริทึมแบบมีผู้สอน (Supervised learning) ใช้สำหรับพยากรณ์ตัวเลขแบบต่อเนื่อง โดยอาศัยหลักการเพื่อนบ้านใกล้สุด ในเอกสารฉบับนี้จะแบ่งออกไปอีก 2 กรณี คือ KNN-D และ KNN-U

1.5.22. **KNN-D** หมายถึง KNN Regressor ที่นำค่าระยะห่าง (D = distance) เข้ามาคำนวณค่าเฉลี่ยถ่วงน้ำหนักกับเพื่อนบ้านใกล้เคียงด้วย

1.5.23. **KNN-U** หมายถึง KNN Regressor ที่มีสมมติฐานว่าเพื่อนบ้านใกล้เคียงทุกตัวมีน้ำหนักเท่ากันหมด (U = uniform)

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. ระบบสารสนเทศเชิงภูมิศาสตร์ (Geographic Information System – GIS)

ระบบสารสนเทศภูมิศาสตร์ คือ ระบบการทำงานที่เกี่ยวข้องกับข้อมูลในเชิงพื้นที่ โดยใช้คอมพิวเตอร์เข้ามาช่วยรวบรวม วิเคราะห์ และประมวลผล ข้อมูลหรือสารสนเทศเหล่านี้มักมีความสัมพันธ์ในเชิงตำแหน่งพื้นที่ (Spatial data) เช่น โฉนดที่ดินเลขที่ 182 ตั้งอยู่ในซอยพหลโยธิน 45 เขตบางซื่อ กรุงเทพมหานคร ซึ่งรูปแบบและความสัมพันธ์ของข้อมูลเชิงพื้นที่จะสามารถนำมาวิเคราะห์ด้วย GIS และทำให้สื่อความหมายได้โดยง่ายเมื่อแสดงอยู่บนแผนที่ นอกจากนี้ ระบบสารสนเทศเชิงภูมิศาสตร์ยังสามารถนำมาวิเคราะห์สัมพันธ์กับเวลาที่เปลี่ยนไปได้ ยกตัวอย่างเช่น การเฝ้าระวังค่าฝุ่น pm 2.5 ในแต่ละช่วงเวลาของวัน ของเดือน หรือของปี, อุณหภูมิน้ำทะเลในแต่ละปีเพื่อพยากรณ์ถึงปรากฏการณ์เอลนิโญ่หรือลานีญ่า เป็นต้น

2.1.2. การวิเคราะห์สถิติเชิงพื้นที่ (Geospatial statistics)

ในทางทฤษฎี การวิเคราะห์สถิติเชิงพื้นที่แบ่งออกเป็น 2 ระดับ ได้แก่

<u>สถิติเชิงพื้นที่ระดับคลอบคลุม (Global)</u> เป็นการวิเคราะห์ในภาพรวม มักใช้ในการวิเคราะห์อัตสหสัมพันธ์เชิงพื้นที่ (Spatial autocorrelation) เพื่อดูว่า ในบริเวณพื้นที่ที่พิจารณาทั้งหมดปัจจัยใดปัจจัยหนึ่งมีความสัมพันธ์ในเชิงพื้นที่เป็นแบบใด ได้แก่ เชิงสุ่ม เชิงเกาะกลุ่ม หรือกระจายตัว ตัวสถิติสำคัญๆ เช่น อัตสหสัมพันธ์เชิงพื้นที่มอแรน (Moran's I) และ ค่าสัมประสิทธิ์เกียรี่ (Geary's coefficient)

สถิติเชิงพื้นที่ระดับเฉพาะที่ (Local) เป็นการวิเคราะห์ที่เฉพาะเจาะจงลงไปในบริเวณที่สนใจ มักทำหลังจากได้ผลลัพธ์จากการทำสถิติเชิงพื้นที่ระดับคลอบคลุมแล้วเพื่อให้สามารถวิเคราะห์ได้ลึกยิ่งขึ้นในบางมุมมอง ตัวสถิติสำคัญ เช่น GetisOrd Gi* และ Anselin Local Moran's I (LISA)

การวิเคราะห์อัตสหสัมพันธ์เชิงพื้นที่ (Spatial autocorrelation) นั้น มีพื้นฐานมาจากแนวคิดสำคัญทางภูมิศาสตร์ของ Waldo R. Tobler (1969) ที่ระบุว่า "Everything is related to everything else. But near things are more related than distant things." แปลความในมุมมองของงานฉบับนี้ได้ว่า แปลงที่ดินที่ตั้งอยู่ใกล้เคียงกันควรมีความสัมพันธ์ หรือ มีอิทธิพลต่อกัน มากกว่าแปลงที่ดินแปลงที่อยู่ห่างไกลออกไป ซึ่งเป็นสมมติฐานหลักของแนวคิดการประเมินราคาแปลงที่ดินโดยอาศัยแปลงซื้อ-ขายจดทะเบียนที่อยู่ใกล้ๆ กัน ยกตัวอย่างการวิเคราะห์อัตสหสัมพันธ์เชิงพื้นที่ที่นิยมใช้ ดังนี้

1) อัตสหสัมพันธ์เชิงพื้นที่ระดับครอบคลุมมอแรนส์ไอ (Global Moran's I)

เป็นตัวชี้วัดความสัมพันธ์เชิงพื้นที่อัตโนมัติที่พัฒนาโดย Patrick Alfred Pierce Moran การคำนวณค่าสถิติใช้ทั้งค่า "ตำแหน่ง (Locations)" และ "คุณสมบัติ (Attribute values)" ตามสมการด้านล่าง

$$Moran's\ I = \frac{n}{W}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

โดยที่

$\bar{X}$ *คือ ค่าเฉลี่ยในแต่ละแอตทริบิวต์*

x_i, x_j *คือ ค่าตัวอย่างแถวที่* i *และแถวที่* j *ของแอตทริบิวต์*

w *คือ ค่าน้ำหนักเชิงพื้นที่ระหว่าง* i *และ* j

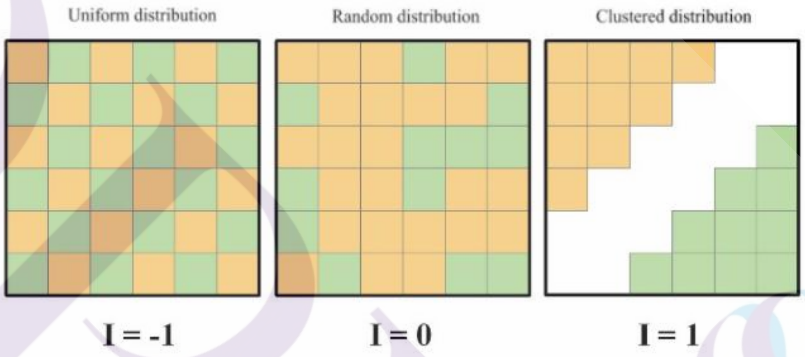
n *คือ จำนวนตัวอย่างทั้งหมด*

W *คือ ผลรวมของค่าน้ำหนักเชิงพื้นที่ทั้งหมด*

เนื่องจากสถิติมอแรนส์ไอประเมินรูปแบบและแนวโน้มของข้อมูลแบบทั่วทั้งหมด หมายความว่า หากรูปแบบหรือแพทเทิร์นเชิงพื้นที่มีความคงเส้นคงวาไปทั่วทุกบริเวณ ค่าสถิติตัวนี้จะเป็นตัวแทนที่ดีและบ่งชี้ความสัมพันธ์เชิงพื้นที่ได้อย่างมีนัยสำคัญ

ค่าสถิติมอแรนส์ไอจะมีค่าอยู่ระหว่าง -1 ถึง +1 โดยที่

- -1 หมายถึง ชุดข้อมูลมีการกระจายตัวแบบสม่ำเสมอ (Perfect dispersion)
- 0 หมายถึง ชุดข้อมูลมีการกระจายตัวแบบสุ่มไม่มีแพทเทิร์นที่แน่นอน (Perfect randomness)
- 1 หมายถึง ชุดข้อมูลมีการกระจายตัวเป็นกลุ่มอย่างชัดเจน (Perfect clustering)



ภาพที่ 2.1: แสดงรูปแบบการกระจายตัวเชิงพื้นที่แบบต่างๆ

2) อัตสหสัมพันธ์เชิงพื้นที่ระดับเฉพาะที่ GetisOrd Gi*

GetisOrd Gi* เป็นค่าสถิติเฉพาะที่ (Local) ที่วัดความสัมพันธ์เชิงพื้นที่โดยพิจารณาในบริบทของหน่วยเพื่อนบ้านใกล้เคียง หากเพื่อนบ้านมีลักษณะหรือคุณสมบัติที่สนใจคล้ายๆ กัน แสดงว่าบริเวณนั้นมีความสัมพันธ์เชิงพื้นที่ต่อกัน คำนวณได้จากสมการด้านล่าง

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \bar{x}\sum_{j=1}^{n} w_{ij}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_i j\right)^2}{n-1}}}$$

โดยที่

G_i^* *คือ ค่าคะแนนมาตรฐานความสัมพันธ์แบบ GetisOrd ที่ตำแหน่งใดๆ*

x_j *คือ ค่าคุณลักษณะของแอตทริบิวต์ที่ตำแหน่ง j*

$\bar{x}$ *คือ ค่าเฉลี่ยของแอตทริบิวต์*

w_{ij} *คือ ค่าน้ำหนักของตำแหน่ง i และ j*

n *คือ จำนวนตัวอย่างทั้งหมด (แถว)*

S *คือ ค่าส่วนเบี่ยงเบนมาตรฐาน คำนวณจาก* $S = \sqrt{\frac{\sum_{j=1}^{n} x_j}{n} - (\bar{x})^2}$

นอกจากนี้ ค่าสถิติอื่นๆ ที่สำคัญที่มักจะคำนวณออกมาพร้อมกันด้วย คือ ค่า p-value และค่าระดับความเชื่อมัน Gi_Bin ภาพที่ 2 แสดงให้เห็นถึงความสัมพันธ์ระหว่างชั้นข้อมูลนำเข้า (ซ้ายสุด) โดยเมื่อคำนวณ GetisOrd Gi* จะได้ Z-scores (ค่า G_i^* นั่นเอง) และค่า p-value ตามลำดับ

ภาพที่ 2.2: *แสดงผลลัพธ์จาก* GetisOrd Gi* *ในรูปแบบแผนที่* (*ที่มา*: ESRI)

จากภาพด้านบน (ภาพกลาง) ค่า Z-scores ได้ถูกนำมาแบ่งออกเป็น 7 bins โดยยึดจากค่าความเชื่อมั่นทางสถิติในระดับสำคัญๆ 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

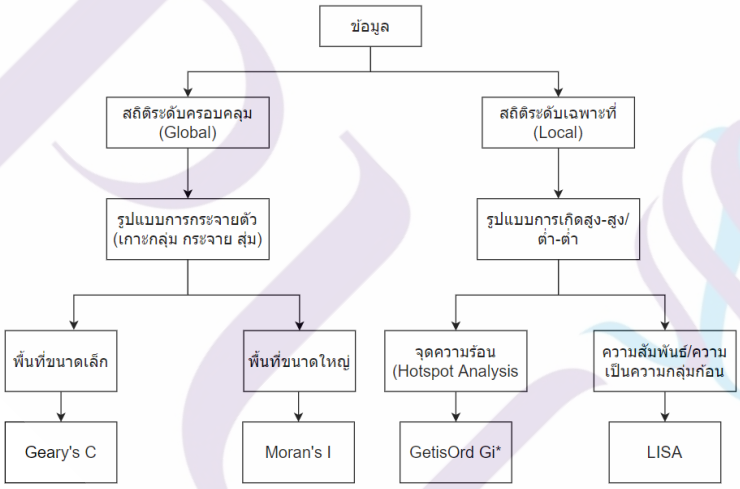
Bin +1: Hot Spot – 90% Confidence แสดงด้วยสีแดงอ่อน

Bin +2: Hot Spot – 95% Confidence แสดงด้วยสีแดงปานกลาง

Bin +3: Hot Spot – 99% Confidence แสดงด้วยสีแดงเข้ม

ตัวอย่างการแปรความ เช่น แปลงที่ดินที่มี Gi_Bin เป็น +3 หรือ สีแดงเข้ม หมายความว่า แปลงที่ดินนั้นอยู่ใกล้กับ POIs ทุกประเภทในภาพรวมมาก ถือเป็นแปลงที่ดินที่น่าสนใจ มีนัยสำคัญเชิงบวก ราคาประเมินควรจะสูงในเชิงเปรียบเทียบ เป็นต้น

ในการศึกษาครั้งนี้ ได้เลือกใช้การวิเคราะห์จุดความร้อนด้วยเทคนิค GetisOrd Gi* (ซึ่งเป็นระดับเฉพาะที่) เนื่องจากสามารถระบุบริเวณที่มีค่าความหนาแน่นของกลุ่มแปลงที่ดินที่อยู่ใกล้กับ POIs (Point of interests) ได้ โดยแยกทำในระดับเฉพาะที่ คือ ทำในแต่ละคลัสเตอร์ที่เป็นผลลัพธ์จาก Spatially Constrained Multivariate Clustering นอกจากนี้ ไม่ได้นำวิธี LISA (สถิติเชิงพื้นที่เฉพาะที่อีกตัวหนึ่ง) มาใช้เนื่องจากการศึกษาครั้งนี้ไม่ได้สนใจศึกษาความสัมพันธ์ระหว่างพื้นที่ที่มีความหนาแน่นของกลุ่มแปลงที่ดินที่อยู่ใกล้ POIs กับพื้นที่อื่นโดยรอบ

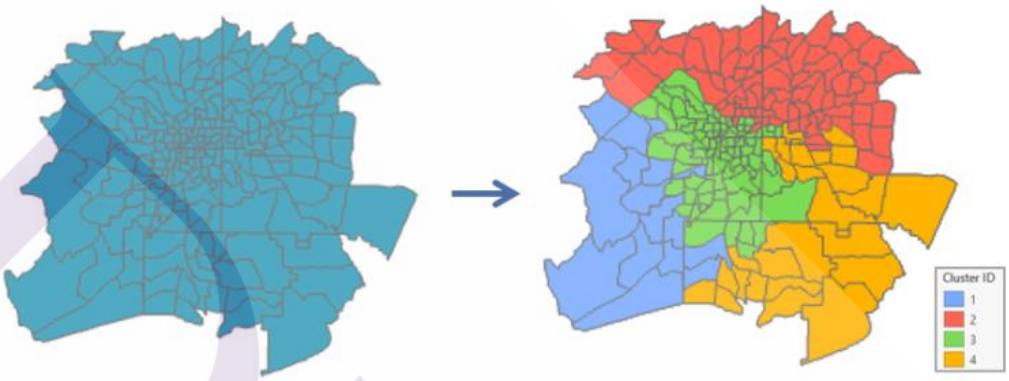


ภาพที่ 2.3: *แสดงสถิติเชิงพื้นที่ (ที่มา*: *ธีรไนย ศรีธรรมรงค์*, 2559)

2.1.3. Spatially Constrained Multivariate Clustering (SCMC)

เป็นอัลกอริทึมแบบไม่มีผู้สอน (Unsupervised learning) แบบหนึ่ง แต่มีวิธีการแบ่งกลุ่มที่คำนึงถึงความสัมพันธ์เชิงพื้นที่ระหว่างกัน โดยมีสมมติฐานว่า แปลงที่ดินที่อยู่ในบริเวณเดียวกันหรือ

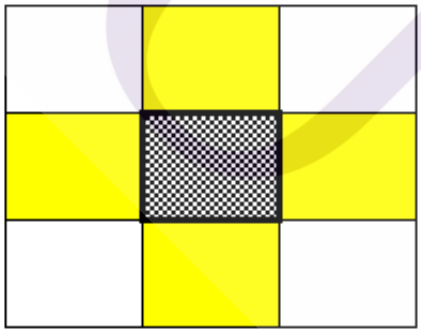
ใกล้เคียงกันเท่านั้น จึงจะถูกจัดให้อยู่ในกลุ่มเดียวกัน มิเช่นนั้นแล้ว จะถูกแบ่งออกเป็นอีกกลุ่มหนึ่ง ภาพที่ 4 แสดงผลลัพธ์ก่อนและหลังจากการจัดกลุ่มด้วยวิธีนี้



ภาพที่ 2.4: *แสดงผลลัพธ์ก่อนและหลังจากการจัดกลุ่มด้วย* SCMC (*ที่มา*: ESRI)

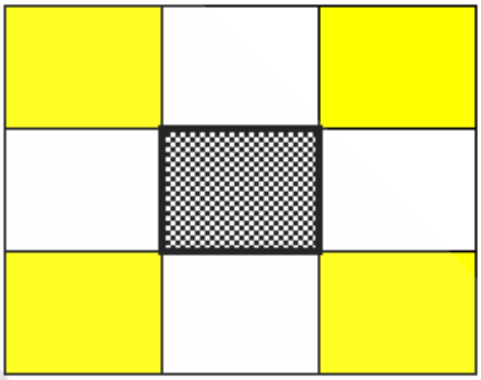
สิ่งสำคัญที่ต้องพิจารณาเมื่อใช้อัลกอริทึมนี้ คือ รูปแบบการจัดกลุ่มเพื่อนบ้าน ซึ่งมีหลายกรณีดังนี้

- เลือกเฉพาะเพื่อนบ้านที่ติดกันทางด้านข้างเท่านั้น (Contiguity edges only หรือ Rook's case)



ภาพที่ 2.5: แสดงรูปแบบความสัมพันธ์เชิงด้าน

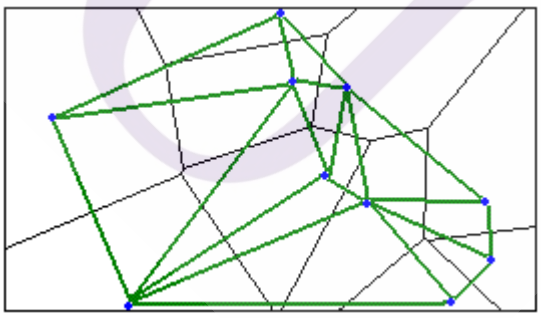
- เลือกเฉพาะเพื่อนบ้านที่ติดกันเชิงมุมเท่านั้น (Contiguity edges corners หรือ Bishop's case)



ภาพที่ 2.6: แสดงรูปแบบความสัมพันธ์เชิงมุม

- เลือกเพื่อนบ้านแบบ Trimmed Delaunay Triangulation

วิธีนี้ใช้หลักการหาเพื่อนบ้านด้วยการวาดสามเหลี่ยมโวโรนอยด์ (Voronoi) โดยเชื่อมกันระหว่างจุด centroid ของแต่ละแปลงที่ดิน (feature) หากสามารถเชื่อมจุดได้เป็นสามเหลี่ยมแสดงว่าแปลงที่ดินเหล่านั้นเป็นเพื่อนบ้านกัน



ภาพที่ 2.7: แสดงการเลือกเพื่อนบ้านแบบ Trimmed Delaunay Triangulation (ที่มา: ESRI)

การใช้วิธีนี้ค่อนข้างเหมาะสมกับงานฉบับนี้เนื่องจากลักษณะแปลงที่ดินมีทั้งแบบติดกันและไม่ติดกัน (มีถนนหรือซอยหรือคลองคั่นระหว่างกัน) ทำให้การใช้ 2 วิธีแรกมีปัญหาที่มักจะเกิดการแตกกลุ่มย่อยเป็นจำนวนมาก วิธี Trimmed Delaunay Triangulation มักจะหาเพื่อนบ้านได้เสมอแม้รูปแบบของแปลงที่ดินเป็นลักษณะเกาะ (Island) หรือเป็นลักษณะแปลกแยก (Outliers)

2.1.4. Linear Regression

เป็นอัลกอริทึมแบบมีผู้สอน (Supervised learning) ใช้สำหรับศึกษาความสัมพันธ์ระหว่างตัวแปรต้นหรือตัวแปรอิสระ (Independent variable) และตัวแปรตาม (Dependent variable) ที่มีลักษณะเป็นตัวเลขต่อเนื่อง มักถูกเรียกอีกแบบว่า การวิเคราะห์การถดถอยเชิงเส้นตรง โดยถ้ามีตัวแปรต้นหรือตัวแปรอิสระเพียงตัวเดียวจะเรียกว่า การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression) แต่ถ้ามีตัวแปรต้นหรือตัวแปรอิสระหลายตัว จะถูกเรียกว่า การวิเคราะห์การถดถอยพหุคูณ (Multiple linear regression) มีสมการในรูปแบบทั่วไปตามด้านล่าง

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots . + \beta_n x_n + \varepsilon$$

โดยที่

y *คือ ค่าของตัวแปรตาม*

x_i *คือ ค่าของตัวแปรต้นหรือตัวแปรอิสระแต่ละตัว*

β_0 *คือ ค่าคงที่ของสมการถดถอย*

β_i *คือ ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ* x_i

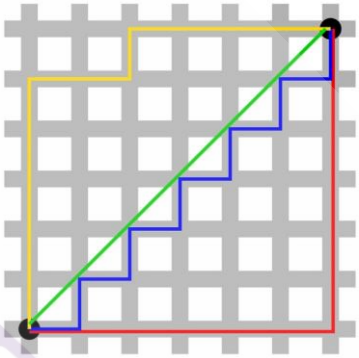
n *คือ จำนวนตัวแปรต้นหรือตัวแปรอิสระในสมการถดถอย*

ε *คือ ค่าความคลาดเคลื่อน หรือเรียกอีกอย่างว่า residual*

ในทางปฏิบัติ มักจะไม่พบลักษณะของตัวแปรต้นหรือตัวแปรอิสระเพียงตัวเดียวที่จะใช้ทำนายตัวแปรตามได้ดี หมายความว่า การวิเคราะห์การถดถอยเชิงพหุคูณจึงมีความสำคัญและถูกใช้ในงานฉบับนี้

2.1.5. K-Nearest Neighbors Regressor หรือ KNN

เป็นอัลกอริทึมแบบมีผู้สอน (Supervised learning) ที่ใช้พยากรณ์ตัวแปรตามแบบตัวเลข ต่อเนื่องโดยอาศัยหลักการ “ความเหมือน” ระหว่างตัวอย่างในชุดข้อมูล การวัดความเหมือน (Similarity measure) จะใช้การวัดระยะห่าง โดยมีฟังก์ชันที่นิยมใช้ คือ Euclidean (วัดแบบระยะขจัด) และ Manhattan (วัดแบบซิตี้บล็อก)



ภาพที่ 2.8: แสดงรูปแบบการวัดระยะห่างแบบ Euclidean และ Manhattan

ที่มา: FC-42 - Distance Operations | GIS&T Body of Knowledge (ucgis.org))

ตามภาพที่ 2.8 ด้านบน เส้นสีเขียวคือการวัดระยะห่างแบบ Euclidean เป็นลักษณะการลากเส้นตรงระหว่างจุด ส่วนเส้นสีเหลือง สีน้ำเงิน และสีแดงคือการวัดระยะห่างแบบ Manhattan ซึ่งมีได้หลายกรณี แต่ทุกกรณีวัดระยะห่างได้เท่ากันทั้งหมด

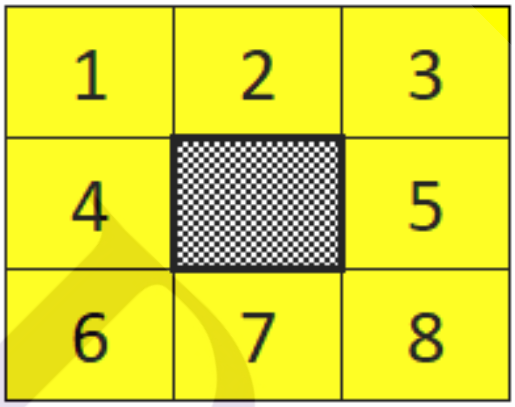
ส่วนด้านล่างเป็นสมการในรูปแบบทั่วไปของ Euclidean และ Manhattan ตามลำดับ

$$D(Euclidean) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

$$D(Manhattan) = \sum_{i=1}^{n}|x_i - y_i|$$

อัลกอริทึม KNN มีพารามิเตอร์สำคัญ คือ ค่า K ซึ่งหมายถึงจำนวนเพื่อนบ้านที่จะใช้เพื่อวัดระยะห่างและพยากรณ์ ในกรณีทั่วไป ค่า K อาจจะมาจากคำแนะนำของผู้เชี่ยวชาญ (Domain expert) หรือในกรณีทางเทคนิค อาจหาค่า K ที่เหมาะสมด้วยวิธี Elbow หรือวัดคะแนน Silhouette แล้วแต่กรณี

ในทางภูมิศาสตร์ KNN มักถูกนำมาใช้จัดกลุ่มหาแปลงหรือโพลีกอนเพื่อนบ้านที่อยู่ติดกัน ซึ่งในกรณีโดยทั่วไปมักกำหนดค่า K เริ่มต้นที่ 8 ด้วยเหตุผลตามภาพที่ 9



ภาพที่ 2.9: แสดงจำนวนแปลงเพื่อนบ้านที่ถูกกำหนดเป็นค่า K เริ่มต้นของ KNN ในทางภูมิศาสตร์

2.2 งานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 3 หมวดหมู่หลัก ได้แก่ 1) แนวคิดแบบจำลองที่ใช้ในการประเมินราคาที่ดินแบบมวลรวม (Mass appraisal), 2) ปัจจัยที่มีผลต่อราคาที่ดินประเภทที่อยู่อาศัย, และ 3) ปัจจัยที่มีผลต่อราคาที่ดินประเภทประเภทพาณิชยกรรม

2.2.1 แนวคิดแบบจำลองที่ใช้ในการประเมินราคาที่ดินแบบมวลรวม (Mass appraisal)

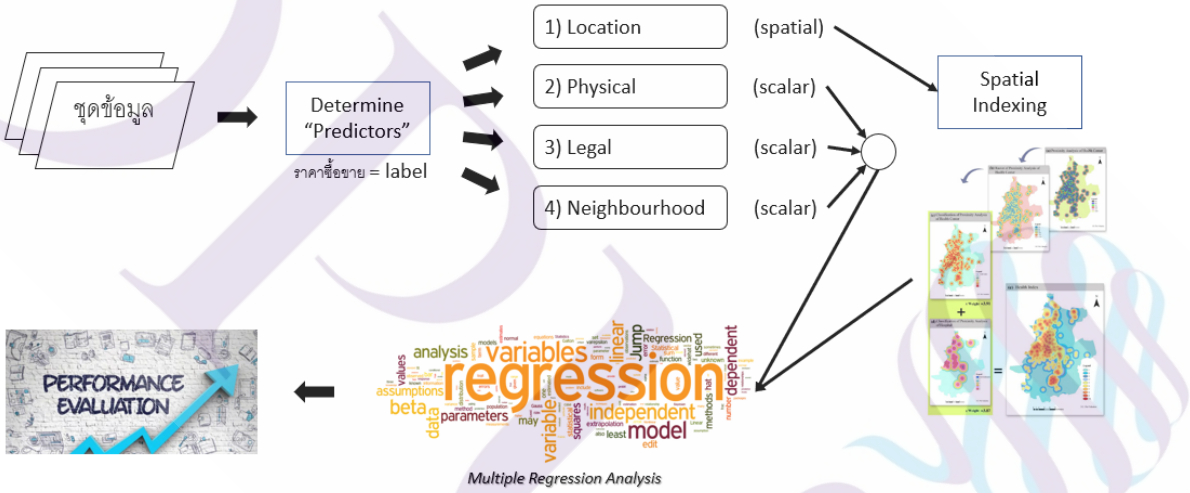
งานวิจัยของ Sukran Yalpir และ Fatma Bunyan Unel ปี 2017

งานวิจัยนี้นำเสนอแบบจำลองประเมินราคาที่ดินแบบมวลรวมในเขตจังหวัดคอนย่า (Konya) ประเทศตุรกี โดยมีวัตถุประสงค์หลักเพื่อ 1) ศึกษาปัจจัยที่ส่งผลต่อราคาที่ดินในเขตนี้ 2) ลดความสูญเสียที่ไม่จำเป็นลง เช่น จำนวนเอกสาร, จำนวนเวลาที่ใช้ในการประเมิน และ 3) เพิ่มความ

ถูกต้องในการประเมินราคา ซึ่งมีความใกล้เคียงกับสิ่งที่ทำในเอกสารฉบับนี้ แต่ข้อได้เปรียบของงานวิจัยนี้ คือ ราคาซื้อ-ขายที่ดินเป็นราคาแท้จริง ซึ่งไม่เหมือนกับกรณีของประเทศไทย

นอกจากนี้ งานวิจัยนี้ได้นำเสนอการทำดัชนีเชิงพื้นที่ (Spatial indexing) สำหรับแต่ละปัจจัย โดยอาศัยปัจจัยย่อยที่เกี่ยวข้อง ยกตัวอย่างเช่น ดัชนีเชิงพื้นที่ของปัจจัยการเดินทาง จะได้มาจากการนำระยะทางไปยังสถานีรถไฟ, ระยะทางไปยังป้ายรถประจำทาง, ระยะทางไปยังท่าเรือ, ระยะทางไปยังสถานีท่ารถตู้ เป็นต้น มากระทำการทางคณิตศาสตร์เชิงพื้นที่บางอย่างโดยได้ผลลัพธ์ออกมาค่าเดียว

ภาพที่ 2.10 แสดงแนวทางการทำแบบจำลองที่งานวิจัยนี้นำเสนอ โดยผู้วิจัยใช้แบบจำลอง Multiple Linear Regression มาฝึกชุดข้อมูลซื้อ-ขายจดทะเบียนด้วยปัจจัยต่างๆ ตามภาพที่ 11 แล้วนำไปประเมินราคาที่ดินรายแปลงต่อไป ค่า R-squared ของชุดฝึกได้เท่ากับ 0.85



ภาพที่ 2.10: แสดงแบบจำลองที่นำเสนอโดยงานวิจัยของ Sukran Yalpir และ Fatma Bunyan Unel, 2017

NO	A. LEGAL FEATURES	Q38	Proximity to Educational Institutions	Q78	Proximity to underpass/ overpass
Q1	Property Conditions	Q39	Proximity to Pre-schools	Q79	Proximity to Unsanitary Areas
Q2	Full Ownership	Q40	Proximity to High Schools	Q80	Proximity to was disposal areas
Q3	Shared Ownership	Q41	Proximity to Higher Education Institutions	Q81	Proximity to treatment facilities
Q4	Zoning Status	Q42	Proximity to courses	Q82	Proximity to natural gas and tube filling facilities
Q5	The Gross Floor Area	Q43	Proximity to Public Institutions	Q83	Proximity to petrol stations
Q6	Total Construction Area	Q44	Proximity to governorships	Q84	Proximity to base stations
Q7	The number of floors≥10	Q45	Proximity to Municipalities	Q85	Proximity to energy transmission lines
Q8	The number of floors<10	Q46	Proximity to Courthouse	Q86	Proximity to underdeveloped areas
Q9	Detached Building	Q47	Proximity to Jailhouse	Q87	Proximity to marsh areas
Q10	Attached Buildings	Q48	Proximity to Security Units	Q88	Proximity to natural disaster areas
Q11	Legal Restraints	Q49	Proximity to Police Stations	Q89	Proximity to not improved river areas
Q12	Right of Mortgage	Q50	Proximity to Military Zones	Q90	Proximity to Industrial Zones
Q13	Easement	Q51	Proximity to Fire Departments/ 112 Emergency	Q91	Proximity to Graveyards
Q14	Annotation of Lease	Q52	Proximity to Attraction Centers	Q92	Proximity to Worship Places
Q15	*Plot Area*	Q53	Proximity to Shopping Centers	Q93	Proximity to Business Centers
NO	B. PHYSICAL FEATURES	Q54	Proximity to Hypermarkets	Q94	Proximity to Parking Areas
Q16	*The location of the plot*	Q55	Proximity to mini-markets	Q95	The View From The Plot
Q17	Corner parcel	Q56	Proximity to open/closed bazaars	Q96	Mountain, valley, etc. views
Q18	Intermediate parcel	Q57	Proximity to commercial enterprises	Q97	Lake, river, stream, etc. view
Q19	Geometric Structure	Q58	Proximity to Cultural Centers	Q98	City view
Q20	Length of the Frontage	Q59	Proximity to cinemas/theaters	NO	D. NEIGHBOURHOOD FEATURES
Q21	The number of frontage	Q60	Proximity to historical sites and touristic attractions	Q99	Population density
Q22	Geometric shape	Q61	Proximity to Entertainment Centers	Q100	Education Level
Q23	Technical Infrastructure Services	Q62	Proximity to fairs, concert areas, etc.	Q101	Level of income
Q24	Water supply	Q63	Proximity to sport facilities	Q102	Immigrant receiving
Q25	Electricity, sewer, natural gas, and telephone	Q64	Proximity to stadium/hippodrome	Q103	Criminal Rate
Q26	Solid waste collection service	Q65	Proximity to entertainment venues	Q104	Neighborliness Relations
Q27	Storm drainage	Q66	Proximity to Green Areas	Q105	Homeowner/tenant
Q28	Unpaved road	Q67	Proximity to forest/copses	Q106	The Surrounding Environment
Q29	Asphalt road	Q68	Proximity to recreation areas	Q107	The favorite neighborhood
Q30	The Road Condition	Q69	Proximity to parks	Q108	Residential Density
Q31	The Periphery Road	Q70	Proximity to playgrounds	Q109	Development potential
Q32	Road width≥10	Q71	Proximity to Public Transportation Points	Q110	Purchasing and selling mobility of real estate
Q33	Road width<10 metre	Q72	Proximity to airports	Q111	Underground, soil, and aboveground features
Q34	The Slope of The Plot	Q73	Proximity to railway stations	Q112	Slope of the neighborhood
NO	C. LOCATIONAL FEATURES	Q74	Proximity to coach station	Q113	Geological condition
Q35	Proximity to Health Facilities	Q75	Proximity to tramway, subway and metrobus stations	Q114	Climate Condition
Q36	Proximity to health center, dispensary, etc.	Q76	Proximity to bus stops	Q115	Air Pollution
Q37	Proximity to State/Private Hospitals	Q77	Proximity to shared taxi routes	Q116	Noise Pollution

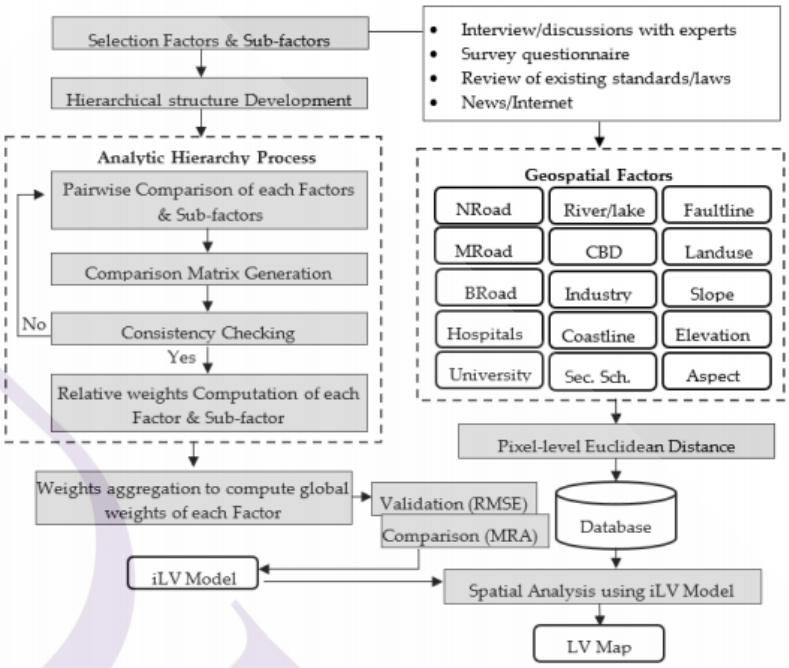
ภาพที่ 2.11: แสดงปัจจัยต่างๆ ที่ใช้ฝึกแบบจำลองนำเสนอโดย Sukran Yalpir และ Fatma Bunyan Unel, 2017

งานวิจัยของ Jannet C. Bencure และคณะ ในปี 2019, พื้นที่ที่ใช้ทำแบบจำลองในงานวิจัยชิ้นนี้ คือ เมืองเบย์เบย์ (Baybay City) ประเทศฟิลิปปินส์ ซึ่งเป็นพื้นที่ชานเมืองติดทะเลแต่มีระดับการเติบโตที่สูงมาก

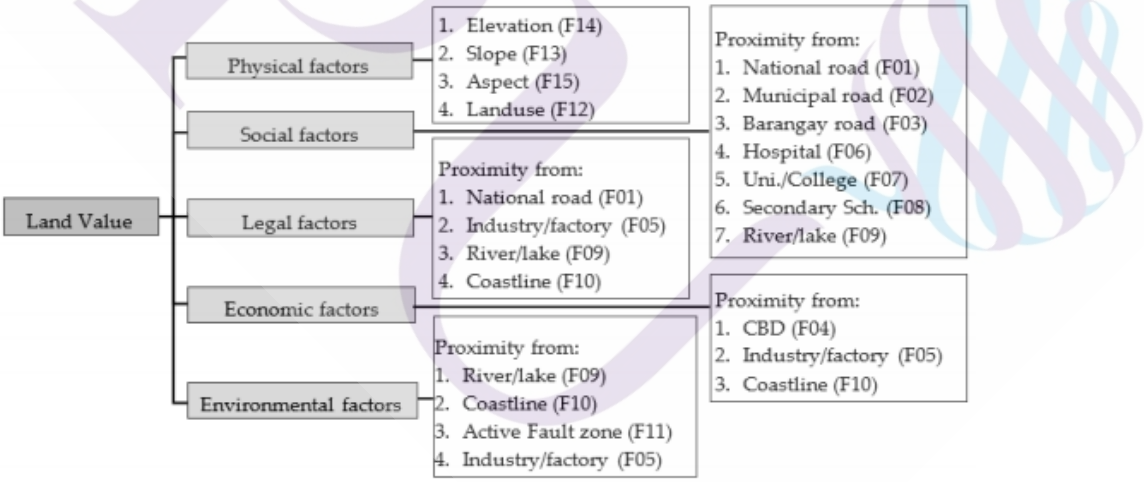
งานวิจัยนี้นำเสนอแบบจำลองใหม่ชื่อ Innovative Land Valuation Model (iLVM) โดยโจมตีวิธีการประเมินราคาแบบ Regression-based models ว่าไม่สมเหตุสมผลเพราะว่าอันที่จริงแล้วแบบจำลองที่ได้นั้นก็มีลักษณะตามปัจจัยที่นำเข้าไปอยู่ดี (data-dependent) นอกจากนี้ การขาดแคลนข้อมูลธุรกรรมการซื้อ-ขาย รวมถึง ความไม่น่าเชื่อถือของธุรกรรมดังกล่าวทำให้วิธีแบบเดิมๆ ใช้ไม่ได้ผล ซึ่งมักทำให้ผู้ประเมินต้องหาวิธีทางเลือก เช่น ไปหยิบธุรกรรมซื้อ-ขายที่เกิดขึ้นในเขตตำบลอื่นมาใช้ ซึ่งก็ยิ่งทำให้การประเมินราคาเกิดปัญหาไม่คงเส้นคงวา

ผู้วิจัยจึงนำเสนอวิธีใหม่ คือ iLVM โดยจะอาศัย AHP (Analytic hierarchy process) เป็นตัวขับโดยพยายามแปลงปัจจัยทางด้านนามธรรมต่างๆ เช่น ปัจจัยทางกายภาพ, ปัจจัยทางสังคม, ปัจจัยทางด้านสิ่งแวดล้อม, ปัจจัยทางเศรษฐกิจ รวมถึงปัจจัยทางกฎหมายให้กลายเป็นปัจจัยเชิงรูปธรรมด้วยการให้ค่าน้ำหนักเป็นตัวเลขโดยผู้เชี่ยวชาญ นอกจากนี้ การพิจารณาทั้งปัจจัยที่ส่งผลเชิงบวกและปัจจัยที่ส่งผลเชิงลบไปพร้อมกันก็ส่งผลลัพธ์ที่ดีกว่าอีกด้วย อย่างไรก็ตาม ผู้วิจัยยอมรับว่า ถึงแม้วิธีที่นำเสนอนี้อาจจะมีปัญหาในแง่นามธรรม (Subjective) อยู่บ้าง แต่ก็มีความโปร่งใส, ยืดหยุ่น, และเป็นเหตุเป็นผลเพียงพอที่จะนำไปใช้ประเมินราคาที่ดินแบบมวลรวมได้ ภาพที่ 12 แสดงกระบวนทั้งหมดของวิธี iLVM และภาพที่ 2.13 แสดงปัจจัยนำเข้าที่ใช้ทำแบบจำลอง

ผลการทดลองด้วยแบบจำลอง iLVM พบว่าได้ผลลัพธ์ที่ดีกว่าแบบจำลองถดถอยเชิงเส้นธรรมดา โดย RMSE มีค่าเท่ากับ 0.526 (vs 1.953) และมีค่า R-squared เท่ากับ 0.673



ภาพที่ 2.12: แสดงกระบวนการตามแบบจำลอง iLVM ที่นำเสนอโดย Jannet C. Bencure, 2019



ภาพที่ 2.13: แสดงปัจจัยนำเข้าเพื่อใช้ในแบบจำลอง iLVM ที่นำเสนอโดย Jannet C. Bencure, 2019

งานวิจัยของ MIE OAK CHAE และ INHYUK KWON ในปี 2018 ผู้วิจัยทั้งสองท่านเป็นสมาชิกคนสำคัญของ Korea Appraisal Board ที่เป็นหน่วยงานหลักในการออกกฎเกณฑ์การประเมินราคาที่ดินในประเทศเกาหลีใต้ งานวิจัยนี้ถูกนำไปใช้เป็นหลักการโดยทั่วไปสำหรับการประเมินราคาที่ดินในเขตเมืองของประเทศเกาหลีใต้

งานวิจัยนำเสนอหลักการสำคัญ 3 ประการในการประเมินราคาที่ดิน คือ

1) การสำรวจเพื่อค้นหาแปลงที่ดินตัวแทนที่มีศักยภาพ โดยพิจารณาแปลงที่ดินที่มีคุณสมบัติดังนี้

- เป็นแปลงที่ดินที่สามารถสะท้อนโครงสร้างของราคาที่ดินในบริเวณใกล้เคียงได้ดี
- เป็นแปลงที่ดินที่สามารถสะท้อนลักษณะทางกายภาพของที่ดินอื่นในบริเวณนั้นได้
- เป็นแปลงที่ดินที่มีโอกาสในการเปลี่ยนแปลงลักษณะการใช้ประโยชน์ในระยะยาวต่ำ เช่น โอกาสในการเปลี่ยนจากแปลงที่อยู่อาศัยไปเป็นแปลงประเภทอื่นมีน้อย
- เป็นแปลงที่ดินที่มีความโดดเด่น สามารถเข้าถึงหรือทำการสำรวจได้ง่าย

2) ประเมินราคาแปลงที่ดินตัวแทน เพื่อใช้เป็นฐานในการประเมินราคาที่ดินรายแปลง

3) ตารางการปรับปรุงราคาที่ดินแตกแปลง ซึ่งเป็นการเปรียบเทียบแปลงที่ดินที่ต้องการประเมินราคากับแปลงที่ดินตัวแทน แล้วทำการปรับราคาขึ้น-ลงตามหลักเกณฑ์ที่กำหนด เช่น หากแปลงที่ดินที่กำลังพิจารณามีลักษณะเป็นรูปธง ให้ทำการทอนราคาลง 11% จากแปลงที่ดินตัวแทนในบริเวณดังกล่าว

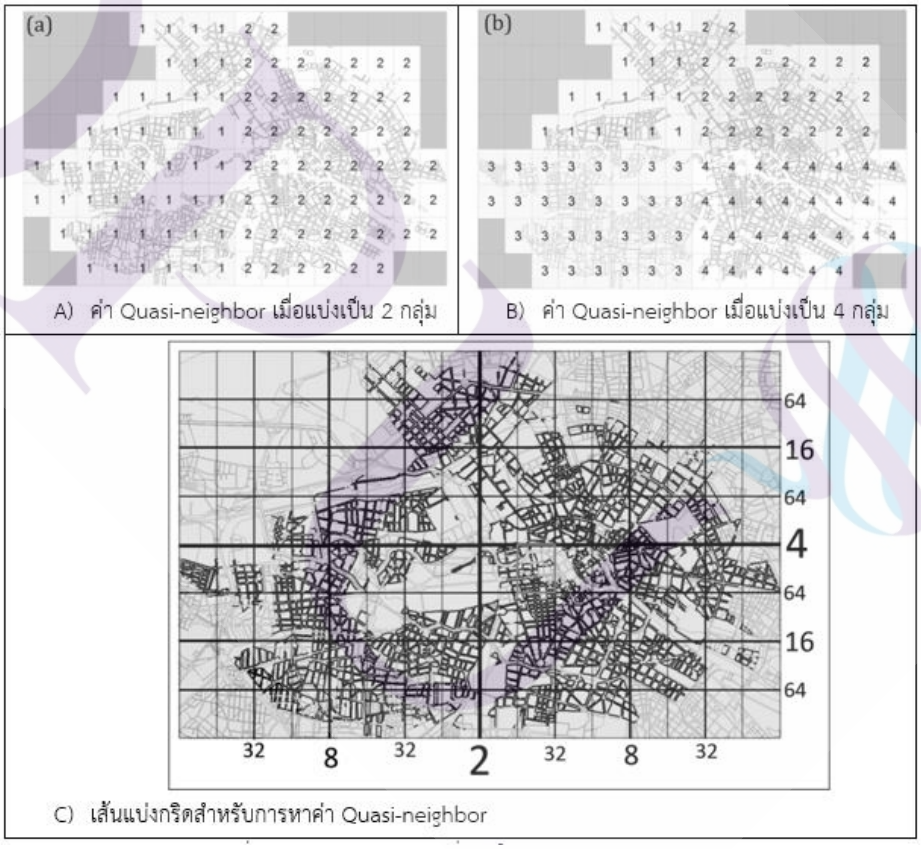
งานวิจัยฉบับนี้ค่อนข้างมีความแตกต่างจากงานวิจัยอื่นๆ ทั้งในแง่ของหลักการและแนวคิดในการประเมินราคาที่ดิน งานวิจัยอื่นๆ มักจะมุ่งเน้นหาปัจจัยต่างๆ ที่มีอิทธิพลต่อราคาซื้อ-ขายแล้วฝึกแบบจำลองด้วยแปลงที่ดินที่มีธุรกรรมการซื้อ-ขายจริง จากนั้นจึงนำแบบจำลองไปประเมินราคาที่ดินแปลงอื่นๆ ในขณะที่งานวิจัยนี้ ใช้หลักการหาแปลงที่ดินตัวแทน แล้วทำการประเมินแปลงที่ดินตัวแทนแล้วนำแปลงที่ดินตัวแทนไปเทียบเคียงเพื่อปรับราคาแตกแปลงต่อไป

2.1.6. กฎราคาเดียว (Law of one price)

งานวิจัย (Waights, 2018) อธิบายถึงปรากฏการณ์ที่น่าสนใจที่เกี่ยวกับราคาที่ดิน คือ กฎราคาเดียว ซึ่งกล่าวว่า แม้ว่าราคาที่ดินจะมีความแตกต่างกันในปัจจุบัน แต่ในระยะยาวราคาทั้งหมดจะลู่เข้าสู่ราคาเดียวเท่านั้น ซึ่งการอธิบายในลักษณะนี้จะหมายถึงกฎราคาเดียวแบบสัมบูรณ์ แต่ในความเป็นจริง

แล้วกฎราคาเดียวอาจเกิดขึ้นแบบสัมพัทธ์ก็ได้ ซึ่งหมายถึงส่วนต่างของราคาระหว่างที่ดินที่อยู่คนละพื้นที่กันนั้นจะคงที่ หมายความว่าส่วนต่างของราคาจะลู่เข้าสู่ค่าศูนย์สำหรับกฎแบบสัมบูรณ์และลู่เข้าสู่ค่าคงที่สำหรับกฎแบบสัมพัทธ์

ตัวแปรส่วนหนึ่งที่ใช้ในงานวิจัยนี้ ได้แก่ ความหนาแน่นของสถานีรถไฟ, อายุของละแวกบ้านใกล้เคียง, ระยะทางไปยัง CBD ที่ใกล้ที่สุด, ระยะทางไปยังสถานที่สำคัญของเมือง, ระยะห่างจากสถานที่พึงประสงค์ (Amenities), ระยะห่างจากสถานที่ไม่พึงประสงค์ (Disamenities), ระยะห่างไปยังพื้นที่สีเขียว, ระยะห่างไปยังแหล่งน้ำ, ระยะห่างไปยังรางรถไฟบนดิน, และค่า Quasi-neighbor ซึ่งเป็นการจัดกลุ่มที่ดินให้มีความไดนามิกส์มากขึ้นโดยที่ดินที่อยู่ใกล้กันจะอยู่ในกลุ่มเดียวกัน ดังแสดงในภาพที่ 2.14



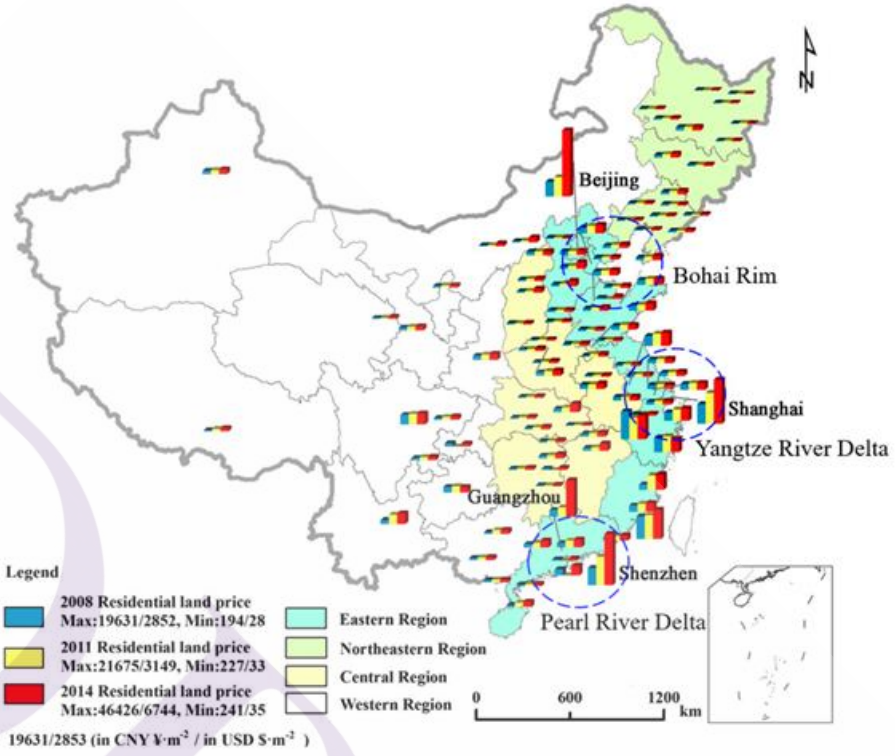
ภาพที่ 2.14: แสดงค่า Quasi-neighbor ที่นำเสนอในงานวิจัยของ Waights, 2018

2.1.7. ปัจจัยที่มีผลต่อราคาที่ดินประเภทที่อยู่อาศัย

งานวิจัยของ Yang, Hu, Li, Zhang, และ Torres ในปี 2017งานวิจัยนี้ได้ทำการศึกษาปัจจัยที่มีผลกระทบต่อราคาที่ดินประเภทที่อยู่อาศัยโดยศึกษาการเปลี่ยนแปลงของราคาที่ดินจาก 105 หัวเมืองใหญ่ในประเทศจีน โดยใช้ราคาในปี 2008, 2011, และ 2014 จากการศึกษาพบว่ามีปัจจัยที่สำคัญ 3 ประการที่มีผลต่อราคาที่ดิน ได้แก่ จำนวนแรงงานต่างถิ่น (Immigrant population), GDP (Gross domestic product) หรือผลิตภัณฑ์มวลรวมภูมิภาค, และการลงทุนในอสังหาริมทรัพย์ ข้อมูลราคาที่ดินที่ใช้ในงานวิจัยนี้ถูกแสดงในภาพที่ 2.15

งานวิจัยแสดงให้เห็นว่า GDP เป็นปัจจัยที่มีผลกระทบสูงสุด โดยส่งผลในเชิงบวกต่อราคาที่อยู่อาศัยที่เปลี่ยนแปลงไปตามเวลา กล่าวคือ หัวเมืองที่มีค่า GDP สูงจะมีแนวโน้มที่มีราคาที่ดินประเภทที่อยู่อาศัยสูงตามไปด้วย ส่วนปัจจัยในแง่ของจำนวนแรงงานต่างถิ่นส่งผลในเชิงลบต่อราคาที่ดินประเภทที่อยู่อาศัยในภาพรวม อย่างไรก็ดี ผลกระทบของปัจจัยนี้กลับลดลงเมื่อปีเพิ่มขึ้น สำหรับปัจจัยสุดท้าย คือ ปัจจัยการลงทุนในอสังหาริมทรัพย์นั้นพบว่าปัจจัยนี้ส่งผลในเชิงบวกต่อราคาที่ดินจริง แต่ผลเชิงบวกนี้กลับมีแนวโน้มลดลงในแต่ละช่วงเวลา คาดว่าน่าจะเกิดขึ้นเนื่องจากเมื่อมีการลงทุนในอสังหาริมทรัพย์ที่มากขึ้นกลับก่อให้เกิดอุปสงค์เทียมของความต้องการเป็นเจ้าของที่ดินประเภทที่อยู่อาศัย (Fake land demand) อย่างไรก็ตาม งานวิจัยนี้มิได้ให้น้ำหนักกับปัจจัยที่อาจมีความสำคัญอื่นๆ เช่น นโยบายการซื้อ-ขายที่ดิน เป็นต้น

งานวิจัยนี้ใช้แบบจำลอง Geographically Weighted Regression (GWR) ในการประเมินราคาที่ดินประเภทที่อยู่อาศัย ความแม่นยำของแบบจำลองเมื่อวัดด้วยค่า R-squared สำหรับปี 2008, 2011, และ 2014 มีค่าเท่ากับ 0.65, 0.74, และ 0.79 ตามลำดับ ส่วนค่า Adjusted R-squared มีค่าเท่ากับ 0.54, 0.66, และ 0.73 ตามลำดับ



ภาพที่ 2.15: แสดงราคาที่ดินของ 105 หัวเมืองในจีน ในปี 2008, 2011, และ 2014 (Yang et al., 2017)

งานวิจัยของ Yang, Song, และ Choi ในปี 2016 งานวิจัยนี้ศึกษาปัจจัยที่มีผลกระทบต่อราคาที่ดินประเภทที่อยู่อาศัยในกรุงโซล ประเทศเกาหลีใต้ ซึ่งมีความหนาแน่นของประชากรต่อ 1 ตารางกิโลเมตร คือ 17,000 คน สูงกว่าหลายๆ หัวเมืองสำคัญในโลก เช่น นิวยอร์ค 10,500 คน/ตร.กม., ปารีส 8,500 คน/ตร.กม. หรือ กรุงเทพมหานคร ที่มีความหนาแน่นของประชากรที่ 5,295 คน/ตร.กม. เท่านั้น

งานวิจัยนี้เกิดขึ้นเนื่องจากความไม่สอดคล้องกันของหลายงานวิจัยก่อนหน้า เช่น งานวิจัยของ Li และ Brown (1980) ซึ่งสรุปว่า ราคาที่อยู่อาศัยจะมีแนวโน้มสูงหากอยู่ใกล้กับย่านธุรกิจ แต่ในขณะที่งานวิจัยของ Mahan, Polasky, และ Adams (2000) ที่กลับสรุปในทิศทางตรงกันข้าม หรือแม้แต่งานของ Frew และ Jude (2003) ที่แสดงให้เห็นถึงความสัมพันธ์ในเชิงลบระหว่างราคาที่อยู่อาศัยกับกิจกรรมทางด้านธุรกิจ (Commercial activities) ในพื้นที่ใกล้เคียง แต่ในขณะเดียวกันงานของ

Cervero และ Duncan (2004) กลับอธิบายความสัมพันธ์ในทางตรงกันข้ามว่ากิจกรรมทางด้านธุรกิจมีผลในเชิงบวกต่อราคาที่ดินประเภทที่อยู่อาศัย

ดังนั้น Yang, Song และ Choi (2016) จึงสนใจการศึกษาแนวโน้มของราคาที่ดินประเภทที่อยู่อาศัยกับแปลงเพื่อนบ้านในบริเวณใกล้เคียง โดยงานวิจัยนี้ได้แบ่งระยะห่างออกเป็น 4 ระดับ ได้แก่ 100 เมตร, 250 เมตร, 500 เมตร และ 750 เมตร มีค่าเฉลี่ยของจำนวนหน่วยธุรกิจในแต่ละระยะเท่ากับ 103, 586, 2265, และ 5019 แห่ง ตามลำดับ

นอกจากแอททริบิวต์ที่แสดงถึงความหนาแน่นของธุรกิจในพื้นที่ใกล้เคียงนั้น งานวิจัยนี้ยังใช้แอททริบิวต์ อื่นๆ อีก ได้แก่ ระยะห่างจากศูนย์กลางธุรกิจ ทั้งสามย่านหลักของโซล ได้แก่ Central Business District (CBD), Gangnam Business District (GBD) และ Yeouido Business District (YBD), โซนของพื้นที่, สัดส่วนการจ้างงานต่อหลังคาเรือน (Job-Housing Ratio), ระยะทางไปยังสถานีรถไฟที่ใกล้ที่สุด, ระยะทางไปยังโรงเรียนประถมที่ใกล้ที่สุด และ ระยะทางไปยังถนนหลัก อีกด้วย

จากการศึกษาพบว่าในกรุงโซล ปัจจัยระยะทางจากแปลงที่ดินประเภทที่อยู่อาศัยไปยังพื้นที่ทางด้านธุรกิจนั้นส่งผลในเชิงลบ คือ พื้นที่ธุรกิจนั้นถูกจัดว่าเป็นสถานที่ไม่พึงประสงค์ (Disamenity) และพบว่าผลกระทบนี้จะลดลงไป เมื่อระยะห่างมีมากกว่า 500 เมตร โดยระยะที่มีผลกระทบสูงสุดคือระยะในช่วงสั้นๆ คือ 250 เมตร ส่วนปัจจัยสัดส่วนการจ้างงานแสดงให้เห็นถึงโอกาสที่จะหางานได้ส่งผลในเชิงบวกอย่างมากต่อราคาที่ดิน ปัจจัยระยะห่างจากศูนย์กลางธุรกิจส่งผลในเชิงบวกต่อราคาที่ดิน ทั้ง CBD และ GBD แต่ไม่มีผลสำหรับ YBD ปัจจัยระยะทางที่ใกล้กับสถานีรถไฟส่งผลในเชิงบวกอย่างมากต่อราคาที่ดิน ส่วนการอยู่ใกล้กับโรงเรียนระดับประถมและการอยู่ใกล้กับถนนหลักนั้นส่งผลในเชิงบวกต่อราคาที่ดินเช่นกันแต่ไม่มากเท่าแอตทริบิวต์อื่นๆ

งานวิจัยนี้ใช้แบบจำลอง Quadratic Regression จำนวน 4 แบบจำลองในการประเมินราคาที่ดินประเภทที่อยู่อาศัย ได้ค่า R-squared และ Adjusted R-squared อยู่ในช่วง 0.43-0.44 สำหรับทุกแบบจำลอง

งานวิจัยของ Kilic, Baser, และ Gulser ในปี 2019, งานวิจัยนี้ศึกษาปัจจัยที่มีผลต่อราคาที่อยู่อาศัยในพื้นที่เขตเมือง (urban) ในจังหวัดซัมซุน (Samsun) ประเทศตุรกี พื้นที่เกษตรกรรมได้ถูกปรับเปลี่ยนมาเป็นพื้นที่เมืองในช่วง 20 ปีที่ผ่านมาเพราะการเพิ่มขึ้นของประชากรและสาธารณูปโภคต่างๆ ผลของการศึกษาพบว่าปัจจัยที่สำคัญที่ส่งผลต่อราคาที่ดิน เช่น ระยะห่างจากชายฝั่ง, สัดส่วนพื้นที่

อาคารต่อพื้นที่ที่ดิน (Floor area ratio), ขนาดของแปลง, ลักษณะของรูปแปลง, ระยะทางไปยังชายฝั่ง, ระยะทางไปยังสถานีรถไฟ, ระยะห่างจากดาวน์ทาวน์และศูนย์กลางธุรกิจ, ระยะทางไปยังที่ทำการไปรษณีย์, และระยะทางไปยังป้ายรถประจำทาง เป็นต้น

เมืองซัมซุนนับเป็นเมืองที่มีการก่อสร้างเป็นจำนวนมากและความเจริญกำลังหลั่งไหลเข้าสู่เมืองทำให้การประเมินราคาที่ดินที่แท้จริงนั้นทำได้ยาก การที่ความเจริญเข้าไปสู่เมืองอย่างรวดเร็วทั้งทางด้านสาธารณูปโภคต่างๆ ตลอดจนการก่อสร้างตึกสูง ทำให้เกิดปัญหาในหลายด้านทั้งในเชิงเศรษฐกิจและสังคม เช่น ปัญหามลพิษ งานวิจัยนี้จึงช่วยหาปัจจัยที่มีผลต่อราคาที่ดินที่จะเป็นประโยชน์กับทั้งฝั่งรัฐบาลและฝั่งเอกชนในการหาราคาที่เหมาะสมกับเมืองที่กำลังพัฒนานี้

นอกจากนี้ งานวิจัยนี้ประสบปัญหาการเก็บข้อมูลคล้ายกับกรณีของประเทศไทยเนื่องด้วยการเก็บข้อมูลราคาซื้อ-ขายเป็นไปอย่างยากลำบาก การสอบถามราคาขายกับเจ้าของโดยตรงหรือแม้แต่การสอบถามข้อมูลจากสำนักงานที่ดิน (Real estate office) อาจไม่ได้ข้อมูลที่เป็นความจริงเนื่องด้วยความไม่เต็มใจที่จะให้ข้อมูลเหล่านี้เพราะเหตุบางประการ เช่น การเสียภาษีซื้อ-ขายที่ดิน (Land transfer tax) เป็นต้น

ปัจจัยที่น่าสนใจอื่นๆ

ยังมีงานวิจัยอีกหลายงานที่กล่าวถึงปัจจัยที่สำคัญต่อราคาที่ดินโดยรวม ได้แก่ (Giussani et al. 1992; Pashardes, & Savva, 2009; Wu et al., 2014) ได้บ่งชี้คล้ายกันว่าราคาที่ดินในภาพรวมนั้นมีความสัมพันธ์ที่ใกล้ชิดต่ออัตราการเติบโตของชุมชนเมือง (Urban growth) ซึ่งเป็นผลมาจากปัจจัยทางเศรษฐศาสตร์ ปัจจัยที่ส่งผลต่อราคาของที่ดินยังมีอีกมากมาย อาทิเช่น Ai (2005) กล่าวว่าปัจจัยพื้นฐานที่สำคัญปัจจัยหนึ่งคือทำเลที่ตั้ง, ขณะที่ Downing (1973) ระบุว่าขนาดของแปลงและโซนตามผังเมือง (Planning zone) เป็นสิ่งสำคัญเช่นกัน, ระยะห่างจากศูนย์กลางธุรกิจ (CBD) ก็เป็นปัจจัยสำคัญดังที่ Tsolacos, McGough และ Thompson (2005) ได้สรุปไว้ในงานวิจัย ซึ่งสอดคล้องกับงานวิจัยของ Gwamna, Wan Yusoff & Ismail (2015) ที่สรุปไว้ว่าการเพิ่มขึ้นของระยะห่างจากศูนย์กลางธุรกิจทำให้ราคาของที่ดินโดยรวมลดต่ำลง นอกจากนี้ Copenheaver และคณะ (2014) ได้ค้นพบข้อสังเกตว่าที่ดินที่มีพื้นที่ขนาดใหญ่มักจะมีราคาต่อหน่วยต่ำกว่าที่ดินที่มีพื้นที่ขนาดเล็กกว่า และแปลงที่ดินที่มีลักษณะปกตินั้นจะมีราคาที่สูงกว่าแปลงที่ดินที่มีรูปแปลงไม่ปกติ (บิดเบี้ยว) อีกด้วย อย่างไรก็ตาม การค้นพบเหล่านี้ย่อมมีความแตกต่างกันไปตามรูปแบบการดำเนินชีวิต, สังคม, เศรษฐกิจ, และการเมือง ซึ่งงานวิจัยส่วนมากมิได้นำมารวมไว้ในสมมติฐานด้วย

สรุปปัจจัยที่มีผลกระทบต่อราคาที่ดินประเภทที่อยู่อาศัย

ตารางที่ 2.1: สรุปผลกระทบของแต่ละปัจจัยต่อราคาที่ดินประเภทที่อยู่อาศัย

ปัจจัย	คำอธิบาย	ผลกระทบต่อราคาที่ดิน
สัดส่วนการจ้างงาน	-	++++
GDP	ผลิตภัณฑ์มวลรวมในประเทศ/ภูมิภาค	+++
ระยะทางไปยัง CBD	CBD หรือศูนย์กลางธุรกิจอาจมีมากกว่า 1 แห่ง	+++
ขนาดของแปลงที่ดิน	ใหญ่หรือเล็กกว่า 650 ตารางเมตร	++
รูปแปลงที่ดิน	ปกติ หรือ ไม่ปกติ (บิดเบี้ยว, สามเหลี่ยม, แปลงธง เป็นต้น)	++
สัดส่วน FAR	พื้นที่ใช้สอยต่อพื้นที่ทั้งหมด	++
การมองเห็นทะเล	เห็น หรือ ไม่เห็น	++
ระยะทางไปยังชายฝั่ง	ระยะทางใกล้สุด	++
ระยะทางไปยังที่ทำการไปรษณีย์	ระยะทางใกล้สุด	++
ระยะทางไปยังป้ายรถประจำทาง	ระยะทางใกล้สุด	++
ระยะทางไปยังโรงพยาบาล	ระยะทางใกล้สุด	++
ระยะทางไปยังธนาคาร	ระยะทางใกล้สุด	++
ระยะทางไปยังสถานที่สำคัญทางศาสนา	ระยะทางใกล้สุด	++
ระยะทางไปยังสถานีรถไฟ	ระยะทางใกล้สุด	+
ระยะทางไปยังโรงเรียน	ระยะทางใกล้สุด	+
ระยะทางไปยังถนนหลัก	ระยะทางใกล้สุด	+
แรงงานต่างถิ่น	-	--
การลงทุนในอสังหาริมทรัพย์	-	--

2.1.8. ปัจจัยที่มีผลต่อราคาที่ดินประเภทพาณิชยกรรม

งานวิจัยของ Boon, Foo, และ Higgins ในปี 2007 งานวิจัยนี้ศึกษาปัจจัยที่ส่งผลต่อราคาที่ดินประเภทพาณิชยกรรม โดยศึกษาและสร้างแบบจำลองที่เป็นข้อมูลการซื้อ-ขายและเช่าอสังหาริมทรัพย์ของที่ดินต่างๆ ทั่วโลก โดยรวบรวมข้อมูลจากตลาดสำนักงานที่ประเทศสิงคโปร์ (Singapore office market) โดยใช้ข้อมูลของตลาดอสังหาริมทรัพย์นี้ตั้งแต่ปี 1992 ถึง 2005

ผู้วิจัยได้ทดสอบโดยการเปรียบเทียบความแม่นยำบนข้อมูลราคาค่าเช่าสำนักงาน โดยตัวแปรที่มีผลต่อราคาค่าเช่าได้แก่ จีดีพี (GDP), อัตราการว่างงาน, อัตราการไล่ออก, ความต้องการของตลาดงาน, ราคาของหุ้นอสังหาริมทรัพย์, อัตราการสร้างสิ่งก่อสร้าง, และอัตราการว่างงาน โดยมีความแม่นยำของแบบจำลองอยู่ที่ 72%

นอกจากนี้งานวิจัยที่เกี่ยวข้องอื่นๆ ยังมีการพยากรณ์เกี่ยวกับตลาดการซื้อ-ขายพื้นที่ว่าง (Space market) เช่นกัน ซึ่งมีการใช้ปัจจัยเพิ่มเติมนอกเหนือจากงานนี้ เช่น จำนวนประชากรในพื้นที่, การเติบโตของการจ้างงาน, ปริมาณการค้าปลีกในพื้นที่, ความมั่นใจของผู้บริโภค, และลักษณะพื้นที่ เช่น ระดับของการลาดเอียงของพื้นที่ อีกด้วย ส่วนปัจจัยที่มีผลต่อตลาดอสังหาริมทรัพย์ ได้แก่ ดัชนีราคาบ้าน, หุ้นสำนักงาน, ค่าใช้จ่ายในการก่อสร้าง, ระดับของสถานที่ไม่พึงประสงค์ เป็นต้น และปัจจัยที่มีผลต่อตลาดทุน ได้แก่ อัตราดอกเบี้ยเงินฝาก, อัตราดอกเบี้ยเงินกู้ เป็นต้น

งานวิจัยของ Ping, Jemes, Fung, Yin, Maidin, และ Abdul Rahman ในปี 2019 งานวิจัยนี้ได้ศึกษาปัจจัยที่มีผลต่อราคาของทรัพย์สินเชิงธุรกิจ (commercial property) หมายถึงทรัพย์สินที่สามารถก่อให้เกิดผลกำไรในเชิงธุรกิจได้ เช่น ร้านค้า, ห้างสรรพสินค้า, อาคารสำนักงาน, อาคารแสดงสินค้า, หรือแม้แต่สวนสาธารณะของภาคเอกชน เป็นต้น โดยแบ่งปัจจัยที่มีผลต่อราคาของทรัพย์สินออกเป็น 2 ประเภทใหญ่ๆ ได้แก่ ปัจจัยเชิงมหภาค (Macro factors) และ ปัจจัยเชิงจุลภาค (Micro factors)

ปัจจัยเชิงมหภาคที่มีผลต่อราคาของทรัพย์สิน ได้แก่ 1) ปัจจัยทางเศรษฐกิจ เช่น GDP, ค่าดัชนีสำคัญทางเศรษฐกิจต่างๆ, อัตราการจ้างงาน เป็นต้น 2) ตัวบทกฎหมายหรือนโยบายภาครัฐ เช่น ข้อกำหนดการซื้อ-ขายที่ดิน, เงื่อนไขการปลูกสิ่งก่อสร้าง เป็นต้น 3) ความหนาแน่นของประชากรวัยทำงานในพื้นที่ 4) การใช้ประโยชน์ของพื้นที่โดยรอบ เช่น หากทรัพย์สินอยู่ใกล้กับสถานที่บางประเภท เช่น มหาวิทยาลัย, โรงพยาบาล, และสถานีตำรวจ เป็นต้น ราคาทรัพย์สินมีแนวโน้มสูงขึ้น 5) โครงการใกล้เคียง ซึ่งมีผลกระทบทั้งเชิงบวกและเชิงลบ เช่น โครงการก่อสร้างคอมมูนิตี้มอลล์ส่งผลเชิงบวกต่อราคา ในขณะที่โครงการก่อสร้างเขตโรงงานส่งผลเชิงลบต่อราคาทรัพย์สิน

ปัจจัยเชิงจุลภาคที่มีผลต่อราคาของทรัพย์สิน ได้แก่ 1) ปริมาณการจราจร หากจราจรมีความหนาแน่นในพื้นที่ใกล้เคียง จะมีผลกระทบเชิงบวกต่อราคา 2) สถานที่อำนวยความสะดวกทางสังคม เช่น สถานบริการด้านสุขภาพ หรือ สถาบันการศึกษา มักให้ผลเชิงบวกต่อราคา 3) ความสะดวกในการ

เดินทาง 4) คุณภาพของเพื่อนบ้าน 5) อายุของทรัพย์สิน โดยมีผลกระทบที่แปรผกผันกับจำนวนปีที่เพิ่มขึ้น เนื่องจากค่าใช้จ่ายในการซ่อมแซมที่สูงเพิ่มขึ้น

สรุปปัจจัยที่มีผลกระทบต่อราคาที่ดินประเภทพาณิชยกรรม

ตารางที่ 2.2: สรุปผลกระทบของแต่ละปัจจัย

ปัจจัย	คำอธิบาย	ผลกระทบต่อราคาที่ดิน
GDP	ผลิตภัณฑ์มวลรวมของพื้นที่	+ หรือ -
อัตราการว่างงาน	-	-
ตัวบทกฎหมาย/นโยบายภาครัฐ	-	-
จำนวนประชากรวัยทำงาน	-	+
อุปสงค์ของการตลาดแรงงาน	-	+
อัตราการก่อสร้างสิ่งก่อสร้าง	-	+
ปริมาณร้านค้าปลีก	-	+
ความมั่นใจของผู้บริโภค	-	+
ความหนาแน่นของแรงงาน	-	+
ปริมาณการจราจร	-	+
สถานที่อำนวยความสะดวก	คุณภาพที่ดีกว่ามีผลกระทบมากกว่า	+
ความสะดวกในการเดินทาง	คุณภาพของการจราจรมีผลต่อราคา	+
คุณภาพของเพื่อนบ้าน	-	+
อายุของทรัพย์สิน	-	-
ความลาดเอียงของพื้นที่	-	-
การใช้ประโยชน์: ที่อยู่อาศัย	การใช้ประโยชน์ของบริเวณใกล้เคียง	+
การใช้ประโยชน์: อุตสาหกรรม	การใช้ประโยชน์ของบริเวณใกล้เคียง	-
การใช้ประโยชน์: ที่ว่าง	การใช้ประโยชน์ของบริเวณใกล้เคียง	-
การใช้ประโยชน์: สำนักงาน	การใช้ประโยชน์ของบริเวณใกล้เคียง	+

บทที่ 3
ระเบียบวิธีวิจัย

ผลจากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า งานวิจัยส่วนใหญ่ใช้วิธีการเรียนรู้จากชุดข้อมูลแปลงซื้อ-ขายจดทะเบียนเพราะเป็นธุรกรรมที่เกิดขึ้นแล้ว แต่ปัญหาหลักในหลายๆ ประเทศ โดยเฉพาะประเทศกำลังพัฒนา พบปัญหาราคาซื้อ-ขายจดทะเบียนที่ได้มานั้นมักจะต่ำกว่าความเป็นจริง (เพื่อประโยชน์ทางภาษี) ทำให้แบบจำลองมีปัญหาในการนำไปใช้งานจริง ดังนั้น ผู้วิจัยจึงมักคิดค้นและนำปัจจัยใหม่ๆ โดยเฉพาะอย่างยิ่งแอตทริบิวต์ที่เกี่ยวกับ POI เช่น ระยะห่างจากแปลงที่ดินไปยัง POI แต่ละประเภท เป็นต้น เข้ามาในแบบจำลองด้วยหวังว่าจะทำให้แบบจำลองแม่นยำขึ้น หลังจากได้แบบจำลองที่ดีแล้ว จึงค่อยนำแบบจำลองไปทำการประเมินราคารายแปลงต่อไป ซึ่งจากการทดลองนำแนวทางนี้ไปทำนายราคาแปลงที่ดินตัวแทนในพื้นที่ต้นแบบจตุจักร พบว่าค่า MAPE ที่ดีที่สุดนั้นอยู่ที่ 25%

แทบจะไม่มีงานวิจัยเชิงประจักษ์ที่พูดถึงการประเมินราคาที่ดินรายแปลงผ่าน “แปลงที่ดินตัวแทน (Reference lots)” ยกเว้นงานวิจัยจากประเทศเกาหลีใต้โดย MIE OAK CHAE และ INHYUK KWON ในปี 2018 ซึ่งใช้หลักการค้นหาแปลงที่ดินตัวแทนก่อนแล้วจึงค่อยประเมินราคาที่ดินรายแปลง สารนิพนธ์ฉบับนี้จึงมีแนวคิดหลักมาจากงานวิจัยชิ้นนี้ อีกทั้งยังมีความสอดคล้องกับกระบวนการในปัจจุบันของกรมธนารักษ์ โดยเพิ่มเติมการคิดค้นวิธีการค้นหาและวิธีประเมินราคาแปลงที่ดินตัวแทน ซึ่งกระบวนการต่างๆ มีดังต่อไปนี้

1) กระบวนการเริ่มต้นจากการสำรวจข้อมูลของชั้นข้อมูล shape files ต่างๆ ได้แก่ ชั้นข้อมูลแปลงที่ดิน, ชั้นข้อมูลแปลงถนน, ชั้นข้อมูลบล๊อคบลู, ชั้นข้อมูลบล็อกฟิกซ์, ชั้นข้อมูลความลึก, ชั้นข้อมูล POI, ชั้นข้อมูลผังเมือง, และชั้นข้อมูลแปลงซื้อ-ขายจดทะเบียน

2) การทำความสะอาดข้อมูล ได้แก่ การเติมข้อมูลสูญหายด้วยวิธีเทียบเคียงเพื่อนบ้าน, การนำแปลงถนนออก, การนำแปลงบล๊อคบลูและบล็อกฟิกซ์ออกไป รวมถึงการแปลงข้อมูลเชิงประเภทให้กลายเป็นข้อมูลเชิงตัวเลขด้วยเทคนิค One-hot-encoding
3) การค้นหาตัวแปรต้นหรือตัวแปรอิสระที่มีความสำคัญ (Feature importance)
4) การจัดกลุ่มแปลงที่ดิน
5) ทำการคัดเลือกแปลงที่ดินตัวแทนของแต่ละหน่วยถนนในแต่ละกลุ่มแปลงที่ดิน ซึ่งจะมี 3 วิธีย่อย ได้แก่ วิธีผู้เชี่ยวชาญ, วิธีวิเคราะห์จุดความร้อนด้วย GetisOrd Gi*, และวิธีเปรียบเทียบเชิงกายภาพ
6) ทำการค้นหาตัวแปรต้นหรือตัวแปรอิสระที่มีความสำคัญ (Feature importance) อีกครั้งหนึ่ง เนื่องจากมีการสร้างตัวแปรใหม่ๆ เป็นจำนวนมากจากขั้นตอนที่ 5)
7) ประเมินราคาแปลงที่ดินตัวแทน ซึ่งจะมี 2 วิธีย่อย ได้แก่ วิธีการกวาดออกเป็นวงกลมจนเจอแปลงซื้อ-ขายจดทะเบียนครบตามที่กำหนด และ วิธีการหาแปลงซื้อ-ขายจดทะเบียนบนหน่วยถนนเดียวกัน

3.1 ข้อมูลนำเข้า

ชั้นข้อมูลนำเข้าในรูปแบบ shape files มีรายละเอียดตามตารางที่ 3 และภาคผนวก ก โดยทุกไฟล์ต้องปรับ coordinate system ให้เป็น UTM 24047 เหมือนกัน

ตารางที่ 3.1 : แสดงรายการข้อมูลนำเข้าที่ใช้ในสารนิพนธ์นี้

ชั้นข้อมูล	รายละเอียดข้อมูล	แหล่งข้อมูล/ปี
แปลงที่ดินเขตจตุจักร	เป็นแปลงที่ดินรวมทุกประเภท	กรมธนารักษ์/2564
แปลงถนนเขตจตุจักร	แปลงถนน/ซอย/ทางต่างๆ	กรมธนารักษ์/2564
แปลงบล๊อคบลูเขตจตุจักร	แปลงที่ดินที่ใช้ประโยชน์ร่วมกัน	กรมธนารักษ์/2564
แปลงบล็อกฟิกซ์เขตจตุจักร	แปลงที่ดินหมู่บ้านจัดสรร/คอนโด	กรมธนารักษ์/2564
ชั้นความลึกเขตจตุจักร	ข้อมูลความลึกรายแปลง	กรมธนารักษ์/2564
ข้อมูล POI	ชั้นข้อมูล POI ต่างๆ เช่น มหาวิทยาลัย, โรงพยาบาล, ห้างสรรพสินค้า, 7-11 เป็นต้น	กรมธนารักษ์/2564
ผังเมืองเขตจตุจักร	การแบ่งเขตตามกฎหมายผังเมือง	กรมธนารักษ์/2564
ข้อมูลซื้อ-ขายจดทะเบียน	เป็นข้อมูลราคาซื้อ-ขายจดทะเบียน ช่วงปี 2555-2558	กรมที่ดิน/2558

3.2 อุปกรณ์และซอฟต์แวร์ที่ใช้ในการศึกษาวิจัย

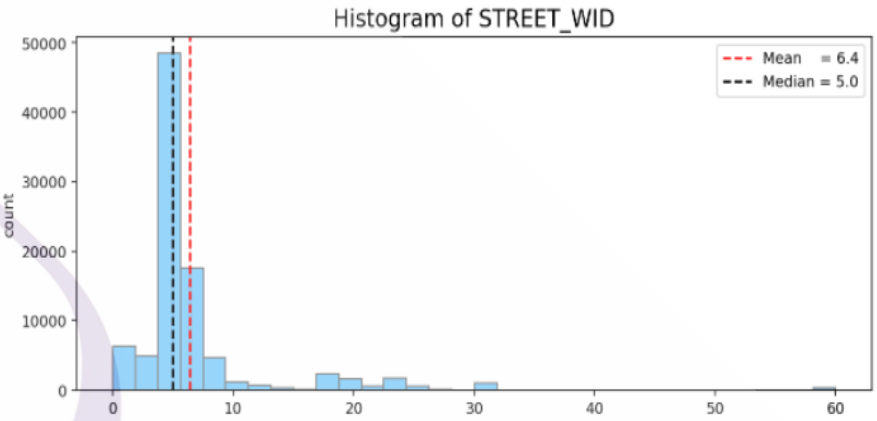
ประกอบไปด้วย 1) ซอฟต์แวร์ระบบสารสนเทศทางภูมิศาสตร์ ArcGIS Pro เวอร์ชั่น 2.7 2) ภาษาไพทอน เวอร์ชั่น 3.7 โดยผ่าน IDE Jupyter Notebook และ Visual Studio Code และ 3) คอมพิวเตอร์ระบบปฏิบัติการ Windows 10 Pro version 20H2 OS build 19042.928

3.3 วิธีการศึกษาวิจัย

3.3.1 การสำรวจข้อมูล

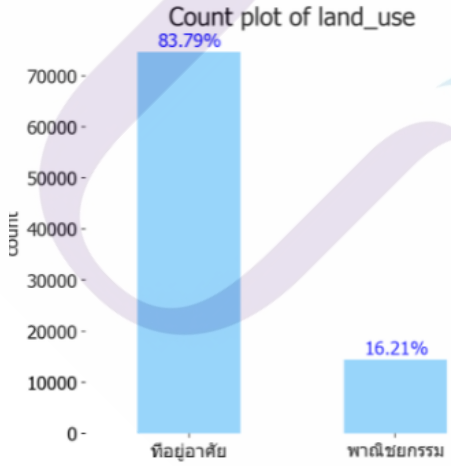
ทำการพิจารณาหาตัวแปรที่สำคัญต่อราคาประเมินที่ดิน โดยการทำ Univariate analysis เพื่อสำรวจตัวแปรต่าง ๆ โดยแบ่งเป็น 2 ประเภทดังนี้

1) ข้อมูลเชิงตัวเลข (Numerical data): ทำการหาจำนวนข้อมูลที่เป็นค่าว่าง, ค่าพิสัย (Range), ค่าเฉลี่ย, ค่ามัธยฐาน และการกระจายตัวของข้อมูล ตัวอย่างเช่น



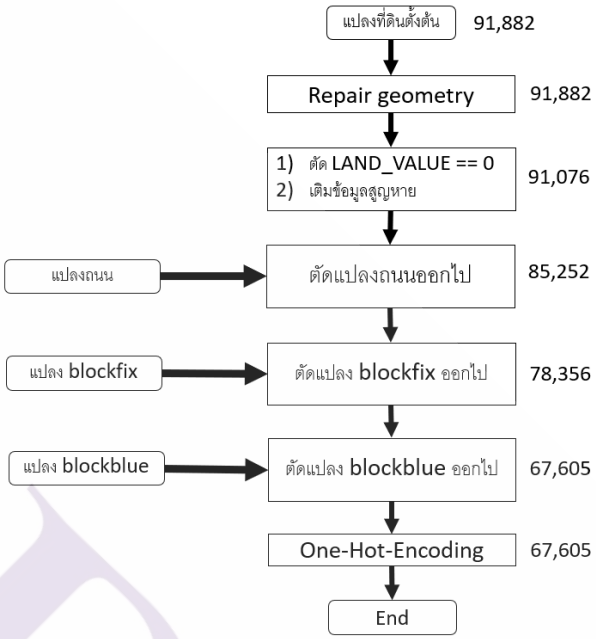
ภาพที่ 3.1 : แสดงฮิสโทแกรมของตัวแปร “ความกว้างของถนน หรือ STREET_WID”

2) ข้อมูลเชิงประเภท (Categorical data): ทำการนับความถี่ของค่าที่เป็นไปได้ของข้อมูลเชิงประเภทแต่ละตัว ดังตัวอย่างด้านล่าง



ภาพที่ 3.2: Count plot ของตัวแปร “การใช้ประโยชน์ที่ดิน หรือ LAND_USE”

3.3.2 การทำความสะอาดข้อมูล



ภาพที่ 3.3: ขั้นตอนการทำความสะอาดข้อมูลพร้อมแสดงจำนวนแปลงที่ดินที่เหลือจากแต่ละขั้นตอน

จากกระบวนการในภาพที่ 18 เริ่มจากขั้นตอนการซ่อมแซม geometry ที่มีปัญหา Topology ด้วยโปรแกรม ArcGIS Pro แล้วทำการตัดแปลงที่ดินที่ไม่มีราคาประเมินออกไป พร้อมทั้งเติมค่าสูญหายด้วยวิธีเทียบเคียงกับแปลงที่ดินเพื่อนบ้าน หลังจากนั้น ทำการนำ “แปลงถนน”, “แปลงบล๊อคฟิกซ์” และ “แปลงบล๊อคบลู” ออกไปด้วยหลักการพื้นที่ซ้อนทับ โดยในขั้นตอนสุดท้าย ตัวแปร/ปัจจัยที่เกี่ยวข้องจะถูกนำมาเข้ากระบวนการ Discretization เพื่อเปลี่ยนข้อมูลที่มีลักษณะเป็นตัวเลขให้อยู่ในลักษณะที่เป็นช่วงและ One-hot-encoding กับข้อมูลเชิงประเภทที่เกี่ยวข้อง

3.3.3 การค้นหาตัวแปรที่มีความสำคัญ (Feature Importance) ต่อการจัดกลุ่มแปลงที่ดิน

จากขั้นตอนการสำรวจข้อมูล พบว่ามีตัวแปรที่เข้าข่ายจะนำไปใช้ในการทำนายราคาประเมิน ได้แก่

- “STREET_WID” หรือ “ความกว้างของถนน”
- “LAND_USE” หรือ “ลักษณะการใช้ประโยชน์”

- “ZONE_NEW” หรือ “โซนผังเมือง”
- “STREET_TYP” หรือ “ชนิดผิวจราจร”
- “FAR” หรือ “อัตราส่วนพื้นที่อาคารรวมต่อพื้นที่ดิน”
- “OSR” หรือ “อัตราส่วนพื้นที่ว่างปราศจากสิ่งปกคลุมต่อพื้นที่อาคารรวม”

วัตถุประสงค์: พิจารณาว่าตัวแปรตัวใดที่มีอิทธิพลต่อ “ราคาประเมินกรมธนารักษ์ (LAND_VALUE)”

สมมติฐาน: พิจารณาความสัมพันธ์ระหว่างตัวแปรกับ “ราคาประเมิน” โดยทำการคัดออกเฉพาะตัวแปรที่มีความสัมพันธ์น้อยกับ “ราคาประเมิน” และคัดออกเฉพาะตัวแปรที่มีความสัมพันธ์กันเอง (Collinearity) ซึ่งก่อให้เกิดปัญหาความโน้มเอียงในตอนทำแบบจำลอง

ขั้นตอนการดำเนินการ:

ในการอธิบายความสัมพันธ์ระหว่างแต่ละตัวแปรกับ “ราคาประเมิน” และความสัมพันธ์ระหว่างตัวแปรด้วยกันเอง ใช้กระบวนการดังต่อไปนี้ผ่านเครื่องมือของ ArcGIS Pro

- กระบวนการ Exploratory Regression
- กระบวนการ Forest-based Regression

ผลลัพธ์จากกระบวนการนี้คือ จะได้ตัวแปรที่มีความสำคัญต่อราคาประเมินซึ่งจะถูกใช้ในขั้นตอนถัดไป

3.3.4 การจัดกลุ่มแปลงที่ดิน

วัตถุประสงค์: จัดกลุ่มแปลงที่ดินที่มีคุณลักษณะใกล้เคียงกันไว้ด้วยกัน โดยจะถูกนำมาใช้ในการหา แปลงที่ดินตัวแทนของแต่ละหน่วยถนนต่อไป

สมมติฐาน: แปลงที่ดินที่อยู่ใกล้กัน (Spatial characteristics) และ/หรือ มีคุณลักษณะบางประการใกล้เคียงกัน ควรจะมีราคาประเมินใกล้เคียงกัน

ขั้นตอนการดำเนินการ:

- แบ่งกลุ่มโดยใช้เฉพาะคุณลักษณะของแปลงที่ดิน โดยไม่พิจารณาตำแหน่งของแปลงที่ดิน โดยใช้เทคนิค K-Means, K-Medoids และ K-Prototypes

- แบ่งกลุ่มโดยใช้เฉพาะคุณลักษณะของแปลงที่ดิน พร้อมกับพิจารณาตำแหน่งของแปลงที่ดิน (Spatially Constrained Multivariate Clustering)
- ผลการจากแบ่งกลุ่มพบว่าวิธี Spatially Constrained Multivariate Clustering ให้ผลลัพธ์ที่ดีที่สุด

3.3.5 การคัดเลือกแปลงที่ดินตัวแทน (Reference lots) ของแต่ละหน่วยถนนในแต่ละกลุ่มแปลงที่ดิน

สำหรับแนวคิดในการคัดเลือกแปลงที่ดินตัวแทนของแต่ละหน่วยถนนสามารถแบ่งได้เป็น 3 แนวทางดังนี้คือ 1) คัดเลือกจากองค์ความรู้ของผู้เชี่ยวชาญและลักษณะข้อมูล 2) คัดเลือกตามการวิเคราะห์จุดความร้อน (Hot spot analysis) ภายในแต่ละกลุ่มแปลงที่ดิน และ 3) คัดเลือกจากการเปรียบเทียบแปลงที่ดินเชิงกายภาพบนหน่วยถนนเดียวกัน ดังมีรายละเอียดของแต่ละแนวทางดังนี้

แนวทางที่ 1: คัดเลือกแปลงที่ดินตัวแทนจากองค์ความรู้ของผู้เชี่ยวชาญและลักษณะข้อมูล

สมมติฐาน: แปลงที่ดินตัวแทนจะต้องเป็นแปลงที่มีลักษณะที่ดีที่สามารถเป็นตัวแทนของแปลงที่ดินที่อยู่ในแต่ละหน่วยถนนได้ โดยจะมีคุณลักษณะดังนี้

- มีลักษณะรูปแปลงเป็น 4 เหลี่ยม
- มีสัดส่วนความกว้างต่อความลึก อยู่ในช่วงเดียวกับแปลงส่วนใหญ่ของกลุ่ม
- มีพื้นที่ที่สอดคล้องกับ อัตราส่วนความกว้างต่อความลึก
- อยู่ใกล้ศูนย์กลางทางธุรกิจ หรือเข้าถึง สาธารณูปโภคง่าย
- เป็นแปลงที่มีลักษณะ highest and best use

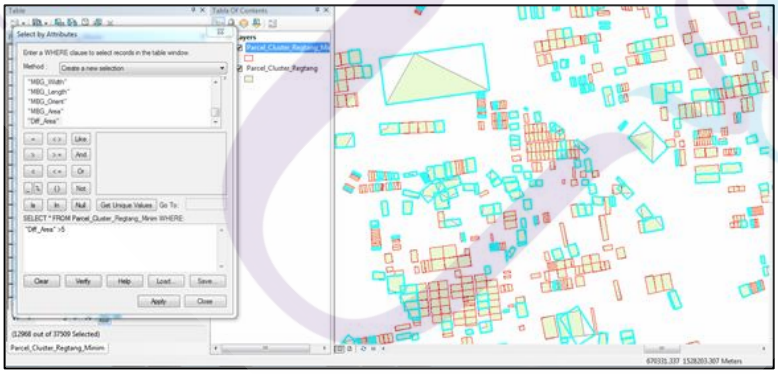
ขั้นตอนการดำเนินการ:

1. การคัดเลือกแปลงที่ดินที่มีลักษณะเป็น 4 เหลี่ยม
   1.1. เป็นรูปแปลงที่ดินที่มี 4 polylines ต่อ 1 polygon (แปลงที่ดิน) ซึ่งพบว่ายังมีบางแปลงที่ดินที่มีลักษณะไม่เป็น 4 เหลี่ยมดังในภาพที่ 3.3 ทำให้ต้องใช้กระบวนการในข้อ 1.2 มาประยุกต์ใช้เพิ่มเติม

ภาพที่ 3.4: แสดงข้อผิดพลาด (ลูกศรสีแดง)

จากการคัดเลือกรูปแปลงที่มีลักษณะเป็น 4 เหลี่ยมโดยพิจารณาจาก 4 polylines ต่อ 1 polygon

1.2. <u>เป็นรูปแปลงที่ดินที่มีผลต่างของพื้นที่แปลงกับ MBR (Minimum Boundary Rectangle) ไม่เกิน 5%</u>



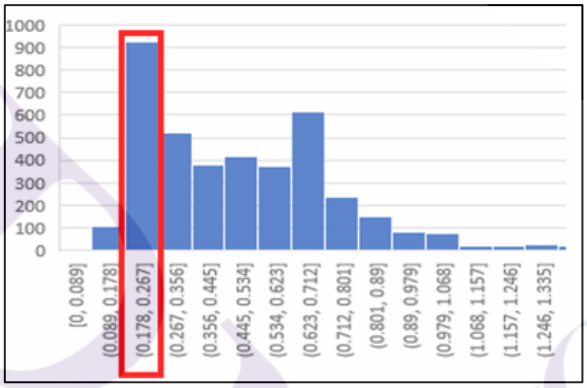
ภาพที่ 3.5: การคัดกรองแปลงที่ดินที่ไม่เป็นรูป 4 เหลี่ยม (กรอบสีฟ้า)
เนื่องจากมีผลต่างพื้นที่ของแปลงที่ดินกับ MBR เกิน 5%

2. <u>คัดเลือกแปลงที่ดินที่มีอัตราส่วนความกว้างต่อความลึกอยู่ในช่วงเดียวกับแปลงส่วนใหญ่ของกลุ่ม</u>

   2.1. คำนวณหน้ากว้างของแปลงที่ดินในส่วนที่ติดถนน

   2.2. คำนวณหาอัตราส่วนระหว่างความกว้างต่อความลึก (DEPTH_WR) และอัตราส่วนของความลึกต่อความกว้าง (DEPTH_RW)

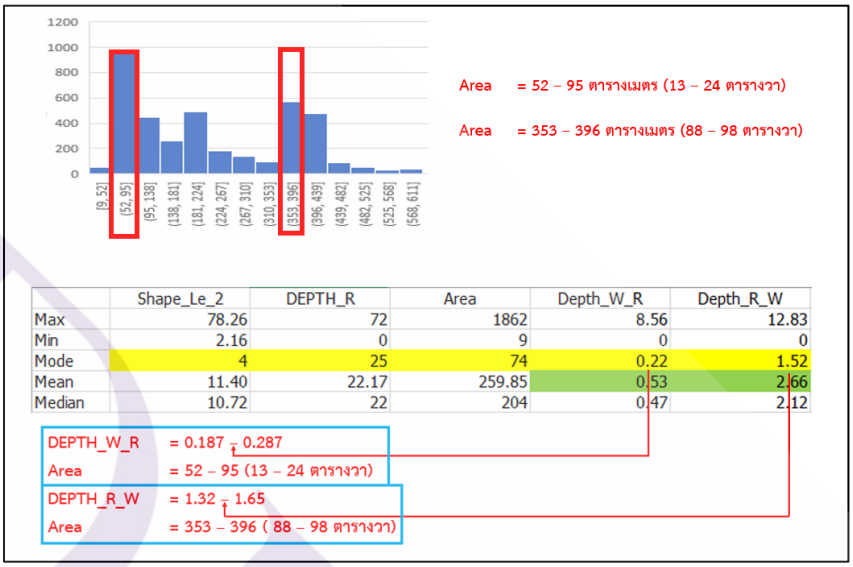
   2.3. ทำการสำรวจข้อมูลเพื่อหาอัตราส่วน DEPTH_WR และ DEPTH_RW ของแปลงส่วนใหญ่ในกลุ่ม โดยการหาช่วงการกระจายตัวของอัตราส่วนที่มีความถี่สูงสุด ดังตัวอย่างในภาพที่ 3.6



ภาพที่ 3.6 : การเลือกช่วงอัตราส่วน DEPTH_WR ของแปลงส่วนใหญ่ในกลุ่ม (กรอบสี่เหลี่ยมสีแดง)

   2.4. คัดเลือกแปลงที่ดินที่มีอัตราส่วน DEPTH_WR และ DEPTH_RW ที่อยู่ในช่วงอัตราส่วนของแปลงที่ดินส่วนใหญ่ (จากข้อ 2.3)

3. คัดเลือกแปลงที่ดินที่มี พื้นที่ ที่สอดคล้องกับ อัตราส่วนความกว้างต่อความลึก



ภาพที่ 3.7: แสดงการเลือกช่วงพื้นที่ (Area) ของแปลงส่วนใหญ่ในกลุ่มที่สอดคล้องกับอัตราส่วน DEPTH_WR และ DEPTH_RW (กรอบสี่เหลี่ยมสีแดง)

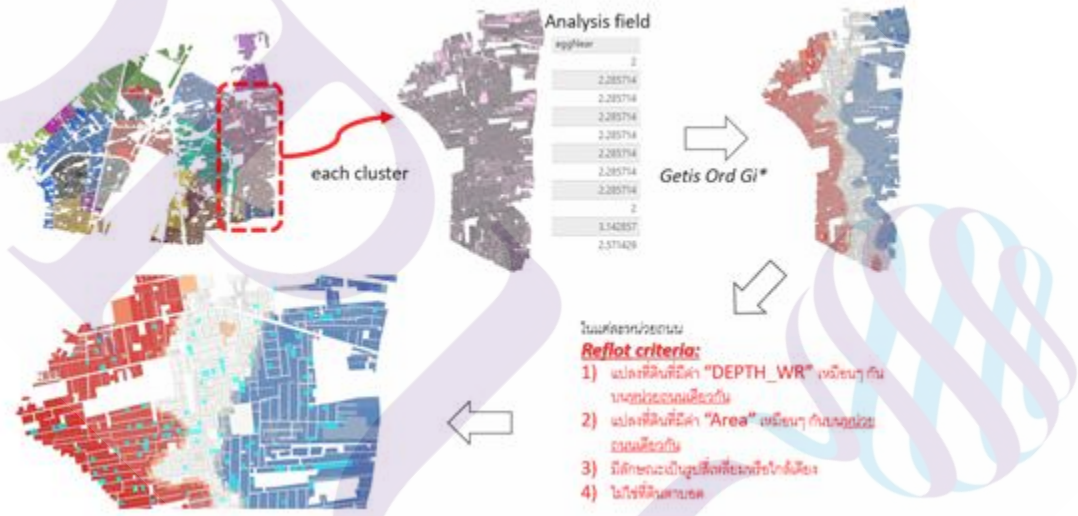
แนวทางที่ 2: คัดเลือกแปลงที่ดินตัวแทนจากการวิเคราะห์จุดความร้อนภายในแต่ละกลุ่มแปลงที่ดิน

แนวคิด: กรมธนารักษ์ได้ยึดหลักหน่วยถนน และแปลงที่ดินตัวแทน ในการประเมินราคาแปลงที่ดินรายแปลง ด้วยเหตุผลดังกล่าว ขั้นตอนนี้จึงนำเสนอวิธีการคัดเลือกแปลงที่ดินตัวแทน เพื่อให้สอดคล้องกับกรอบการทำงานที่กรมธนารักษ์ใช้อยู่ โดยในที่นี้ จะนำชุดข้อมูลจุดสนใจ (POIs – Point of interests) และนำการวิเคราะห์จุดความร้อน (Hot spot analysis) ด้วยวิธีสถิติเชิงพื้นที่ GetisOrd Gi* เข้ามาร่วมวิเคราะห์

ขั้นตอนการดำเนินงาน: ภาพที่ 23 แสดงการคัดเลือกแปลงที่ดินตัวแทนตามแนวทางที่ 2

1. เตรียมข้อมูล POIs ซึ่งจะคัดเลือกเอาเฉพาะตัวที่สำคัญในแต่ละประเภท ดังนี้
   - โรงพยาบาล
   - สถานีรถไฟฟ้าทั้ง BTS และ MRT
   - ห้างสรรพสินค้าใหญ่ๆ เช่น ห้างเซ็นทรัล, ห้างเดอะมอลล์ เป็นต้น

- ไฮเปอร์มาร์ท เช่น โลตัส บิ๊กซี ท๊อปส์ เป็นต้น
- มินิมาร์ท เช่น 7-11, โลตัสเอ็กซ์เพรส, บิ๊กซีมินิ เป็นต้น
- มหาวิทยาลัย เช่น มหาวิทยาลัยเกษตรศาสตร์ เป็นต้น
- โรงเรียน
- ธนาคาร
- ร้านขายยา
- ร้านกาแฟ (Coffee shop)
- ปั๊มน้ำมัน, ปั๊มแก๊ส - ซึ่งเป็น POI ตัวเดียวที่มีลักษณะในเชิงลบ



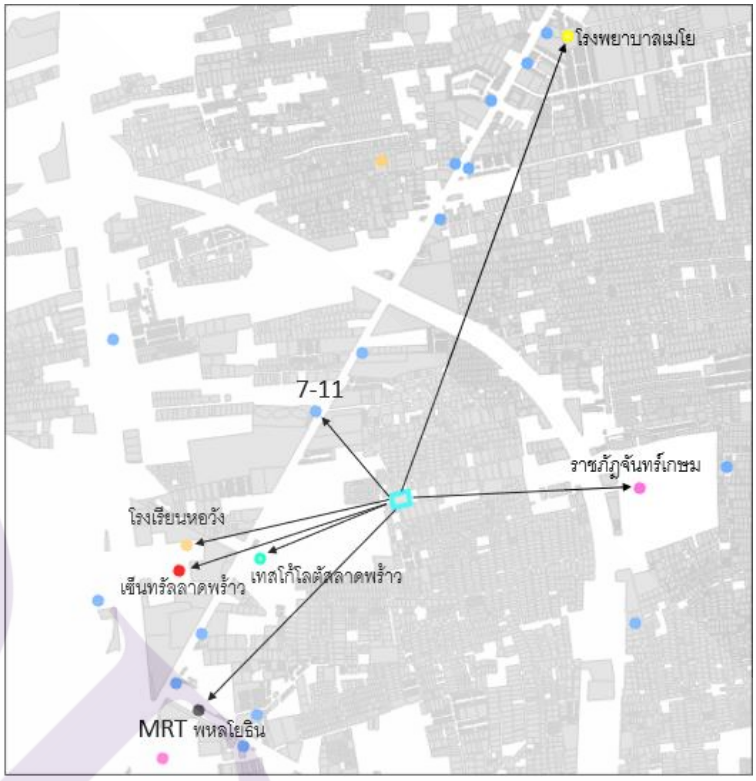
ภาพที่ 3.8: แสดงขั้นตอนการคัดเลือกแปลงที่ดินตัวแทนด้วยการวิเคราะห์จุดความร้อน

2. หาความสัมพันธ์ระหว่าง POIs และแปลงที่ดินแต่ละแปลง โดยวัดระยะทาง (ระยะขจัด) จากแต่ละแปลงที่ดิน ไปยังแต่ละ POI แล้วเก็บค่าระยะทางที่ใกล้ที่สุดสำหรับแต่ละ POI ไว้

- ให้คะแนนความใกล้-ไกลสำหรับแต่ละแปลงที่ดิน โดยมีค่าตั้งแต่ 1 ถึง 6 ค่าที่มากบ่งบอกถึงแปลงที่ดินอยู่ใกล้กับ POI มาก
    - [0, 250] ให้คะแนน 6
    - (250, 500] ให้คะแนน 5
    - (500, 750] ให้คะแนน 4
    - (750, 1000] ให้คะแนน 3
    - (1000, 1250] ให้คะแนน 2
    - (1250, เป็นต้นไป] ให้คะแนน 1
- คำนวณค่าเฉลี่ยความใกล้-ไกลถ่วงน้ำหนักสำหรับแต่ละแปลงที่ดิน โดยค่าน้ำหนักจะแปรผันไปตามความสำคัญของ POI นั้นๆ เช่น ห้างเซ็นทรัลลาดพร้าวมีค่าน้ำหนักมากกว่ายูเนียนมอลล์, ปั๊มแก๊สแอลพีจี มีค่าน้ำหนักในเชิงลบมากกว่าปั๊มน้ำมัน เป็นต้น

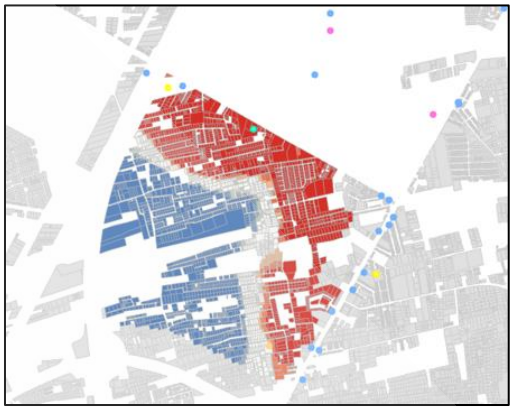
STREET_NAM	near_hospi	near_bts_m	near_deptS	near_hyper	near_minim	near_unive	near_schoo	near_pharm	near_gas_s	near_coffe	near_bank	aggNear
ซอยแยกจากซอยพหลโยธิน35	891.85459	2125.15938	496.00863	1651.66025	493.09874	1350.07698	337.12396	358.83139	414.92851	197.78874	541.91038	2.52381
categorize	3	1	5	1	5	1	5	5	5	6	4	Weighted Avg.
weight	3	3	2	3	1	3	2	1	-1	1	1	
categorize x weight	9	3	10	3	5	3	10	5	-5	6	4	

ภาพที่ 3.9: แสดงตัวอย่างการคำนวณค่าเฉลี่ยความใกล้-ไกลถ่วงน้ำหนักสำหรับแต่ละแปลงที่ดิน



ภาพที่ 3.10: แสดงตัวอย่างการวัดระยะทาง (ระยะขจัด) จากแปลงที่ดิน (กรอบสีฟ้า) ไปยังแต่ละ POI เฉพาะตัวที่ใกล้ที่สุด

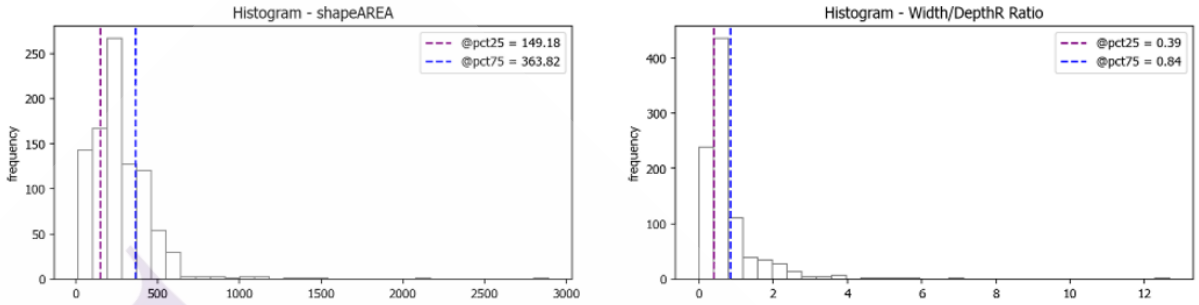
3. ทำการวิเคราะห์จุดความร้อน (Hot spot analysis) โดยวิธี GetisOrd Gi* ใน ArcGIS Pro เพื่อค้นหากลุ่มของแปลงที่ดินที่มีคุณลักษณะคล้ายๆ กัน โดยผลจากวิเคราะห์จุดความร้อนสามารถแบ่งได้เป็น 3 กลุ่มใหญ่ได้แก่
    - กลุ่มแปลงที่ดินที่อยู่ใกล้ POIs ในภาพรวม ซึ่งมีนัยสำคัญในเชิงบวก (แปลงสีแดง)
    - กลุ่มแปลงที่ดินที่อยู่ไม่ใกล้หรือไม่ไกลจาก POIs ในภาพรวม ซึ่งไม่มีนัยสำคัญ (แปลงสีขาว)
    - กลุ่มแปลงที่ดินที่อยู่ไกลจาก POIs ในภาพรวม ซึ่งมีนัยสำคัญในเชิงลบ (แปลงสีน้ำเงิน)



SOURCE_ID	aggNear	GiZScore	GiPValue	NNeighbors	Gi_Bin
0	2.42857	-4.86433	0.00000	156	-3
1	2.42857	-6.52868	0.00000	155	-3
2	2.42857	-4.67677	0.00000	158	-3

ภาพที่ 3.11 : แสดงตัวอย่างผลลัพธ์จากการรัน GetisOrd Gi* สำหรับคลัสเตอร์ที่ 1

4. ทำการค้นหาแปลงที่ดินตัวแทนในแต่ละหน่วยถนน เบื้องต้นพิจารณาจากแปลงที่ดินที่มีคุณลักษณะซ้ำๆ กันในกลุ่ม โดยใช้ช่วงของข้อมูลที่มีความถี่อยู่ระหว่างเปอร์เซ็นต์ไทล์ที่ 25 และ เปอร์เซ็นต์ไทล์ที่ 75 ของตัวแปรอัตราส่วนของความกว้างต่อความลึก (DEPTH_WR) และพื้นที่ (AREA) ดังในภาพที่ 27 หลังจากนั้น หากจำนวนแปลงที่ดินตัวแทนมีจำนวนมากเกินกว่าที่กำหนด เช่น 10% ของจำนวนแปลงที่ดินทั้งหมดบนหน่วยถนนนั้นๆ จะทำการคัดแปลงที่ดินที่มีคะแนน GiZScore ต่ำออกไปจนเหลือตามจำนวนที่ต้องการ



ภาพที่ 3.12: แสดงการหาช่วงของข้อมูลที่มีความถี่สูงสุดของ 2 ตัวแปรเพื่อนำมากำหนดเงื่อนไขในการเลือกแปลงที่ดินตัวแทน

จากภาพที่ 3.12 แปลงที่ดินตัวแทนจะถูกค้นหาตามเงื่อนไขข้างต้น คือ เป็นแปลงที่ดินที่มีค่า DEPTH_WR อยู่ในช่วง 0.39-0.84 และมีค่า AREA อยู่ในช่วง 149-363 ตร.ม. เป็นต้น

อย่างไรก็ดี เพื่อให้ได้แปลงที่ดินตัวแทนที่มีรูปทรงได้มาตรฐาน จึงได้นำผลลัพธ์จากแนวทางที่ 1 เข้ามาช่วยคัดเลือกแปลงที่ดินตัวแทนอีกชั้นหนึ่งด้วย

<u>แนวทางที่ 3: คัดเลือกแปลงที่ดินตัวแทนจากการเปรียบเทียบแปลงที่ดินเชิงกายภาพบนหน่วยถนนเดียวกัน</u>

<u>สมมติฐาน:</u> แปลงที่ดินตัวแทนจะต้องเป็นแปลงที่มีรูปลักษณะหรือคุณสมบัติที่เหมาะสมและสามารถเป็นตัวแทนของกลุ่มแปลงที่ดิน “ที่มีลักษณะเชิงกายภาพคล้ายๆ กัน” ที่อยู่บนแต่ละหน่วยถนนได้

<u>ขั้นตอนการดำเนินงาน:</u>

1. รวบรวมแปลงที่ดินทั้งหมดที่ติดอยู่กับแต่ละหน่วยถนน
2. ในกรณีที่จำนวนแปลงที่ดินมีจำนวนน้อยกว่า 10 แปลง ให้ทำการรวมกับหน่วยถนนอื่น โดยมีหลักเกณฑ์ คือ

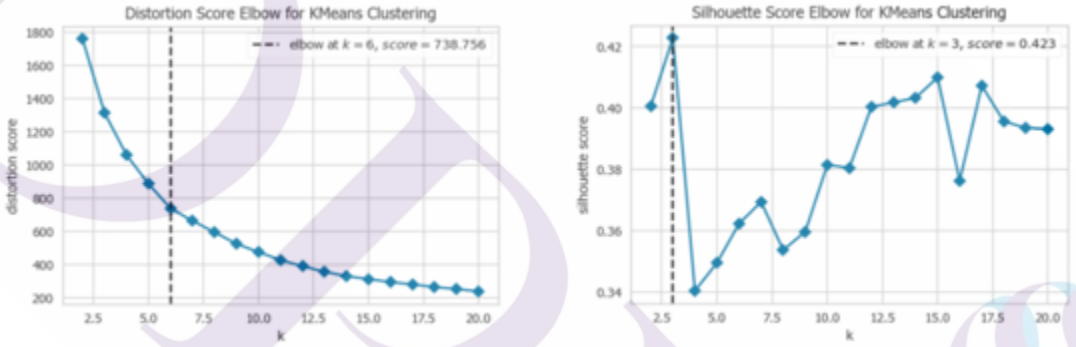
- อยู่ในบริเวณใกล้เคียงกับหน่วยถนนเดิม
- มี “ลักษณะการใช้งานแปลงที่ดิน” เป็นประเภทเดียวกัน
- มี “ความกว้างของถนน” เท่ากัน
- มี “ชนิดผิวจราจร” เหมือนกัน

3. ทำการตัดแปลงที่ดินที่เป็น Outliers ออกด้วยวิธี Interquartile range (+/- 1.5 เท่าของ IQR) โดยคำนวณจากตัวแปร DEPTH_R

4. รัน K-Means ด้วยค่า K ที่ได้มาจากการเปรียบเทียบวิธี Elbow และ Silhouette score โดยใช้ตัวแปรเชิงกายภาพ ได้แก่ “LAND_USE”, “DEPTH_W”, “DEPTH_R” และ “AREA”

5. การคัดเลือกแปลงที่ดินตัวแทนให้พิจารณาดังนี้

- 1 แปลงที่ดินตัวแทน ต่อ 1 คลัสเตอร์ โดยคัดจากแปลงที่ดินที่มีค่า AREA ที่เป็น median
- จำนวนคลัสเตอร์ที่นำมาพิจารณา ต้องครอบคลุมอย่างน้อย 70% ของจำนวนแปลงที่ดินทั้งหมดบนหน่วยถนนนั้น



ภาพที่ 3.13: แสดงการเลือกค่า K โดยอัตโนมัติด้วยวิธี Elbow และ Silhouette Score ตามลำดับ

3.3.6 <u>การค้นหาตัวแปรที่มีความสำคัญ (Feature importance) อีกครั้งหนึ่ง เนื่องจากมีการสร้างตัวแปรใหม่ๆ เป็นจำนวนมาก</u>

การค้นหาแปลงที่ดินตัวแทนทำให้เกิดตัวแปรใหม่ๆ ขึ้นอีกหลายเท่า โดยเฉพาะตัวแปรที่เกิดจากการประมวลผลเกี่ยวกับ POI ไม่ว่าจะเป็น “ระยะทาง” หรือ “จำนวน” ที่เกี่ยวข้องกับ POI แต่ละประเภท นอกจากนี้ ยังมีตัวแปรที่เกิดขึ้นจากการให้คะแนน ranking รวมทั้ง ตัวแปรที่เกิดจากการคำนวณสถิติ GetisOrd Gi*

ด้วยจำนวนตัวแปรใหม่ที่เพิ่มขึ้นอย่างมาก จึงมีความจำเป็นที่ต้องเข้าสู่กระบวนการค้นหาตัวแปรที่สำคัญอีกครั้งหนึ่ง เพื่อลดปัญหาความเอนเอียงและปัญหาการเรียนรู้มากเกินไป (overfitting) กับแบบจำลอง

ในครั้งนี้ จะทำการค้นหาตัวแปรสำคัญและเปรียบเทียบผลลัพธ์จาก 3 อัลกอริทึมตระกูล “ต้นไม้ตัดสินใจ” ได้แก่ Random Forest Regressor, Extra Trees Regressor, และ Gradient Boosting Regressor

3.3.7 การประเมินราคาของแปลงที่ดินตัวแทน

วัตถุประสงค์: ประเมินราคาของแปลงที่ดินตัวแทนโดยพิจารณาจากแปลงซื้อ-ขายจดทะเบียนที่อยู่ใกล้เคียงกัน

สมมติฐาน: แปลงซื้อ-ขายจดทะเบียนที่อยู่ใกล้เคียงสามารถใช้ในการประเมินราคาแปลงที่ดินตัวแทนได้

วิธีการประเมินราคา: แบ่งออกเป็น 2 แนวทาง ได้แก่ 1) ประเมินราคาโดยการ buffer เป็นวงกลมออกไปหาแปลงซื้อ-ขายจดทะเบียนจนครบ K (=5) แปลง และ 2) ประเมินราคาโดยใช้แปลงซื้อ-ขายจดทะเบียนที่อยู่บนหน่วยถนนเดียวกัน

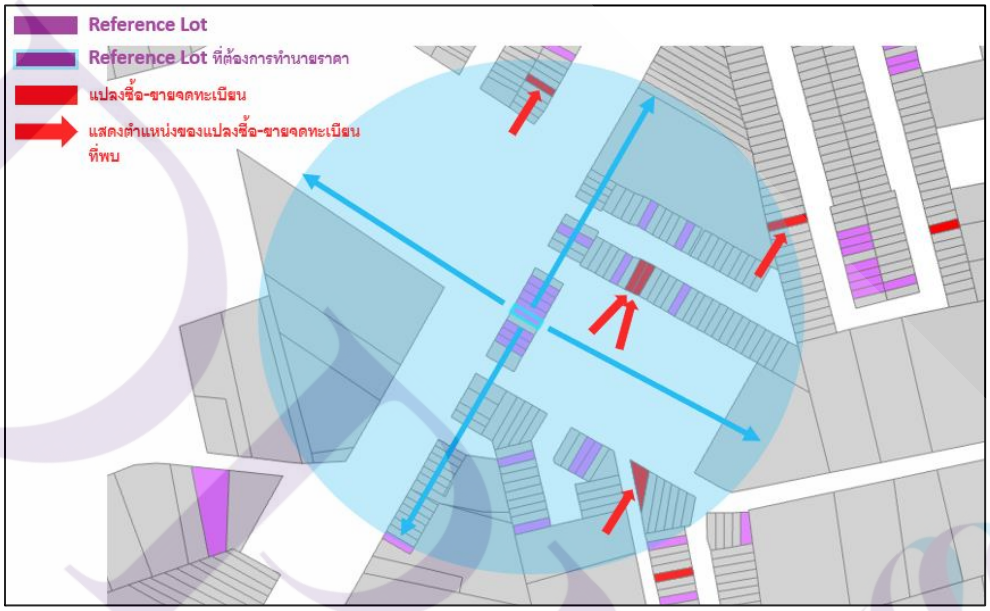
อัลกอริทึมและพารามิเตอร์: **ใช้สำหรับทุกแนวทาง*

- ชั้นแปลงที่ดินที่ใช้ ได้แก่ ชั้นแปลงที่ดินสุดท้ายหลังจากได้แปลงที่ดินตัวแทนในแต่ละวิธี (ซึ่งจะมี “isGoodShp” กำกับอยู่) และชั้นแปลงซื้อ-ขายจดทะเบียนซึ่งกรองเฉพาะวันซื้อ-ขายจดทะเบียนในช่วงปีพ.ศ. 2555-2558 เท่านั้น ทั้งนี้ เพื่อให้สอดคล้องกับราคาประเมินของกรมธนารักษ์ซึ่งประกาศใช้ในปี พ.ศ. 2559
- อัลกอริทึมที่ใช้ ได้แก่ Linear Regression, distance-based KNN และ uniformed-based KNN
- ค่า K ของ KNN เท่ากับ 3 (เพื่อให้สอดคล้องกับจำนวนตัวอย่างที่มีไม่มากในแต่ละหน่วยถนน)
- ตัวแปรต้นที่ใช้ ได้แก่ “การใช้ประโยชน์ที่ดิน”, “ความกว้างของถนน”, “จำนวนร้านกาแฟในระยะ 1500 เมตร”, และ“ระยะทางสั้นที่สุดไปยังมหาวิทยาลัย” ที่ได้จากขั้นตอนการหาตัวแปรที่สำคัญก่อนหน้า
- เมตริกซ์ที่ใช้วัดผล: Mean Absolute Percentage Error (MAPE) เนื่องจากวัดค่าผิดพลาด (error) เป็นเปอร์เซ็นต์ทำให้ง่ายต่อการทำความเข้าใจ
- Standardization/Normalization: ใช้ MinMaxScaler เนื่องจากให้ผลลัพธ์ที่ดีที่สุด

<u>แนวทางที่ 1:</u> ประเมินราคาโดยการ buffer เป็นวงกลมออกไปหาแปลงซื้อ-ขายจดทะเบียน

<u>ขั้นตอนการดำเนินการ:</u>

1. ทำการ buffer จากแปลงที่ดินตัวแทนไปจนกว่าจะเจอแปลงซื้อ-ขายจดทะเบียน จำนวน K (=5) แปลง ดังในภาพที่ 3.13



ภาพที่ 3.14 : แสดงการ buffer จากแปลงที่ดินตัวแทนไปหาแปลงซื้อ-ขายจดทะเบียน จำนวน 5 แปลง

2. ทำการสร้าง 1 แบบจำลองสำหรับทำนายราคาประเมินของ 1 แปลงที่ดินตัวแทน
3. ทำการประเมินผลระหว่างราคาประเมินที่ได้จากแบบจำลองกับราคาประเมินที่ประกาศใช้ของกรมธนารักษ์
4. เปรียบเทียบผลลัพธ์กับวิธี baseline ที่ใช้การเรียนรู้แบบจำลองจากแปลงซื้อ-ขายจดทะเบียนทั้งหมด

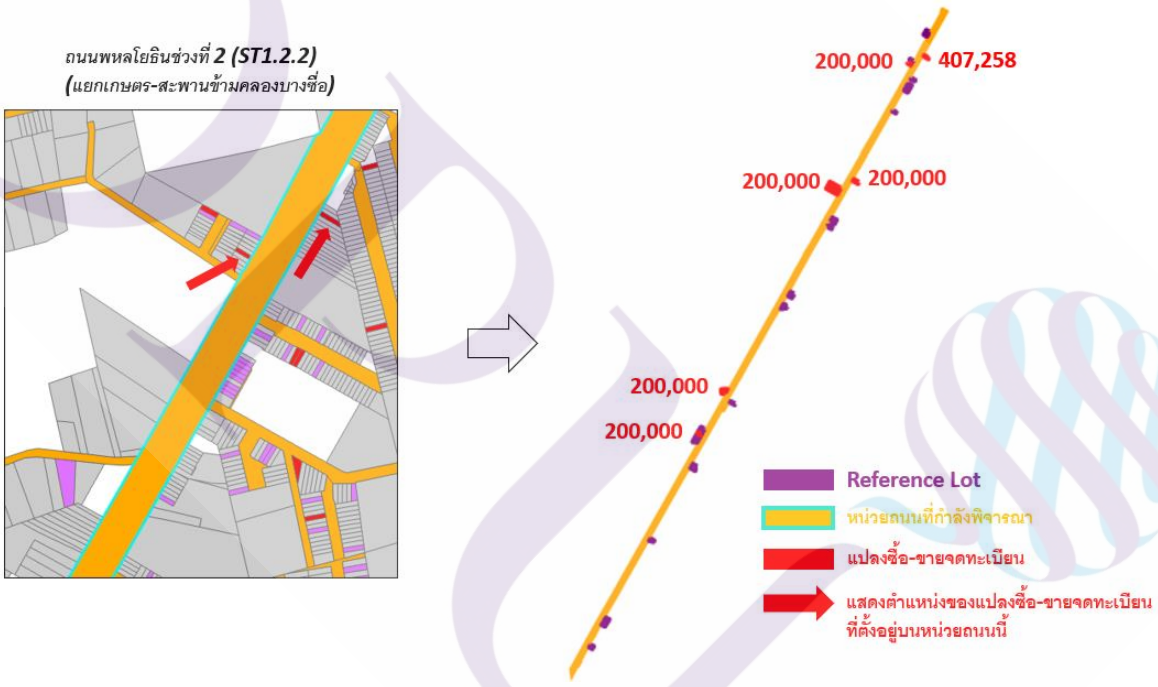
<u>แนวทางที่ 2: ประเมินราคาโดยใช้แปลงซื้อ-ขายจดทะเบียนที่อยู่บนหน่วยถนนเดียวกัน</u>

<u>ขั้นตอนการดำเนินการ:</u>

1. ทำการค้นหาแปลงซื้อ-ขายจดทะเบียนที่ตั้งอยู่บนหน่วยถนนเดียวกัน ซึ่งจะแบ่งได้ออกเป็น 2 กรณี คือ

   <u>กรณีปกติ: “มี” แปลงซื้อ-ขายจดทะเบียนที่ตั้งอยู่บนหน่วยถนนเดียวกัน</u>

   - ทำการสร้าง 1 แบบจำลองสำหรับแต่ละหน่วยถนน
   - ทำการประเมินผลระหว่างราคาประเมินที่ได้จากแบบจำลองกับราคาประเมินที่ประกาศใช้ของกรมธนารักษ์
   - เปรียบเทียบผลลัพธ์กับวิธี baseline ที่ใช้การเรียนรู้แบบจำลองจากแปลงซื้อ-ขายจดทะเบียนทั้งหมด



ภาพที่ 3.15: แสดงหน่วยถนนที่ “มี” แปลงซื้อ-ขายจดทะเบียน

<u>กรณีพิเศษ: “ไม่มี” แปลงซื้อ-ขายจดทะเบียนที่ตั้งอยู่บนหน่วยถนนเดียวกัน</u>

- ทำการค้นหาหน่วยถนนใกล้เคียงที่มีลักษณะกายภาพหรือมีสภาวะแวดล้อมคล้ายกัน โดยอาจพิจารณาจากปัจจัยต่างๆ เช่น เป็นถนน/ซอยเหมือนกัน, ความกว้างของถนนใกล้เคียงกัน, ชนิดผิวจราจรแบบเดียวกัน หรือ อยู่ในโซน POI เดียวกัน เป็นต้น
- นำแปลงซื้อ-ขายจดทะเบียนที่อยู่บนหน่วยถนนดังกล่าวมาใช้ทำแบบจำลอง

ภาพที่ 3.16: แสดงหน่วยถนนที่ “ไม่มี” แปลงซื้อ-ขายจดทะเบียน

2. ทำการสร้าง 1 แบบจำลองสำหรับแต่ละหน่วยถนน
3. ทำการประเมินผลระหว่างราคาประเมินที่ได้จากแบบจำลองกับราคาประเมินที่ประกาศใช้ของกรมธนารักษ์
4. เปรียบเทียบผลลัพธ์กับวิธี baseline ที่ใช้การเรียนรู้แบบจำลองจากแปลงซื้อ-ขายจดทะเบียนทั้งหมด

บทที่ 4
ผลการศึกษาวิจัย

4.1 ลักษณะทางกายภาพของพื้นที่ต้นแบบ (เขตจตุจักร กรุงเทพมหานคร)

เขตจตุจักร ตั้งอยู่ใจกลางและถือเป็นเขตเศรษฐกิจที่สำคัญเขตหนึ่งในกรุงเทพมหานคร พื้นที่โดยทั่วไปเป็นที่ราบ มีคูคลองสายเล็กสายน้อยหลายสาย โดยมีทิศเหนือติดต่อกับเขตหลักสี่ มีคลองบางเขนเป็นเส้นแบ่งเขต ทิศตะวันออกติดต่อกับเขตบางเขนและเขตลาดพร้าว มีคลองบางบัวและคลองลาดพร้าวเป็นเส้นแบ่งเขต ทิศตะวันตกติดต่อกับเขตบางซื่อและอำเภอเมืองนนทบุรี มีทางรถไฟสายเหนือและคลองประปาเป็นเส้นแบ่งเขต ส่วนทิศใต้ติดกับเขตพญาไท เขตดินแดงและเขตห้วยขวาง มีคลองน้ำแก้ว คลองพระยาเวิก และคลองบางซื่อเป็นเส้นแบ่งเขต มีพื้นที่โดยประมาณ 32,908 ตารางกิโลเมตร (แหล่งข้อมูล: สำนักงานเขตจตุจักร)

เขตจตุจักรแบ่งเขตการปกครองย่อยออกเป็น 5 แขวง ได้แก่ 1) ลาดยาว 2) เสนานิคม 3) จันทรเกษม 4) จอมพล และ 5) จตุจักร จากรายงานทางสถิติ สำนักงานสถิติแห่งชาติ ระบุว่ามีจำนวนประชากรอยู่อาศัยรวมทั้งสิ้น 160,948 คน (ชาย 75,600 คน, หญิง 85,348 คน) ณ เดือนธันวาคม พ.ศ. 2556 (แหล่งข้อมูล: สำนักงานเขตจตุจักร)

แปลงที่ดินของเขตจตุจักร (จากกรมธนารักษ์) เพื่อใช้ในงานนี้ มีลักษณะสำคัญดังนี้

- มีจำนวนแปลงที่ดินทั้งสิ้น 91,882 แปลง
- มีถนนหลัก ถนนรอง ซอย/ทาง รวม 1,336 สาย (ก่อนแตกหน่วยถนน)
- ความกว้างถนนมีตั้งแต่ 0 เมตร ไปจนถึง 60 เมตร (ถนนรัชดาภิเษก ช่วงห้าแยกลาดพร้าว-ดินแดง)
- มีผังเมืองที่บังคับใช้อยู่ภายในพื้นที่ 10 แบบ ได้แก่ ย.5, ย.7, ย.8, ย.9, ย.10, อ.2, พ.1, พ.3, พ.4, ส.
- แปลงที่ดินส่วนใหญ่มีหน้าแปลงติดกับถนนคอนกรีต (85%)

- มีห้างสรรพสินค้าใหญ่ที่เป็นที่รู้จักอยู่ 1 ห้าง คือ ห้างเซ็นทรัลลาดพร้าว
- มีมหาวิทยาลัยสำคัญ ได้แก่ มหาวิทยาลัยเกษตรศาสตร์
- สวนจตุจักร ถือเป็นพื้นที่สีเขียวที่สำคัญในเขตนี้
- ในปัจจุบัน (2564) รัฐบาลกำลังย้ายสถานีรถไฟหลักของกรุงเทพมหานคร จากสถานีรถไฟหัวลำโพง มาตั้งอยู่ในเขตบางซื่อ เรียกว่า สถานีกลางบางซื่อ ทำให้เขตจตุจักรกำลังเป็นศูนย์กลางการเดินทางที่สำคัญในอนาคต

ภาพที่ 4.1: แสดงรูปแปลงที่ดินเขตจตุจักร กรุงเทพมหานคร (ที่มา: กรมธนารักษ์)

4.2 ผลจากการสำรวจข้อมูล **รายละเอียดอยู่ในภาคผนวก ข*

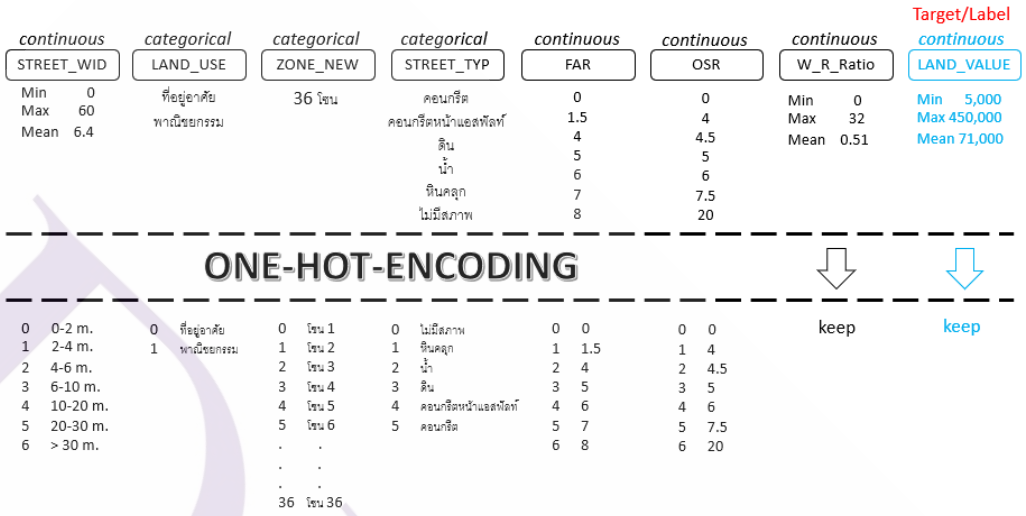
4.2.1 ผลจากการสำรวจข้อมูล โดยวิธี Univariate analysis สำหรับข้อมูลเชิงตัวเลข

ชื่อตัวแปร	พิสัย (Range)	ค่าเฉลี่ย	ค่ามัธยฐาน	ค่าสูญหาย
STREET_VAL (มูลค่าถนน)	0 ถึง 450,000 บาท/ตร.ว.	77,336.90 บาท/ตร.ว.	60,000 บาท/ตร.ว.	803 ค่า
LAND_VALUE (ราคาประเมิน) **ตัวแปรตาม (target variable)*	0 ถึง 450,000 บาท/ตร.ว.	70,885.80 บาท/ตร.ว.	60,000 บาท/ตร.ว.	806 ค่า
STREET_WID (ความกว้างถนน)	0 ถึง 60 เมตร	6.4 เมตร	5.0 เมตร	ไม่มี
WIDTH (หน้ากว้างของแปลง)	0.09 ถึง 273.9 เมตร	10.2 เมตร	7.6 เมตร	37,544 ค่า
DEPTH_R (ความลึกของแปลง)	0 ถึง 1,647 เมตร	33.8 เมตร	19.0 เมตร	ไม่มี

4.2.2 ผลจากการสำรวจข้อมูล โดยวิธี Univariate analysis สำหรับข้อมูลเชิงประเภท

ชื่อตัวแปร	จำนวนค่าที่แตกต่างกัน	ค่าที่มีความถี่สูงสุด	ค่าสูญหาย
LAND_USE	2 ค่า: ที่อยู่อาศัย และพาณิชยกรรม	ที่อยู่อาศัย (83.79%)	2,794 ค่า
STREET_COD (หน่วยถนน)	1,335 หน่วยถนน	-	803 ค่า
STREET_NAM (ชื่อหน่วยถนน)	1,335 หน่วยถนน	-	803 ค่า
STREET_SMG(รหัสถนนย่อย)	29 ถนนย่อย	2R41	37,089 ค่า
STREET_TYP (ชนิดผิวจราจร)	6 ค่า: คอนกรีต, คอนกรีตหน้าแอสฟัลต์, หินคลุก, ดิน, น้ำ, และไม่มีสภาพ	คอนกรีต (93.57%)	2,794 ค่า
ZONE_NEW (ผังเมือง)	36 โซน	ผังเมืองประเภท “ย”	351 ค่า
PARCEL_TYP (ชนิด/ลักษณะของแปลงที่ดิน)	6 ค่า: 1, 6, 7, 13, 14, และ 99	1 - แปลงปกติ (96.98%)	ไม่มี
FAR (อัตราส่วนพื้นที่อาคารรวมต่อพื้นที่ดิน)	7 ค่า: 0.0, 1.5, 4.0, 5.0, 6.0, 7.0, และ 8.0	ค่า 4.0 (39.19%)	314 ค่า
OSR (อัตราส่วนพื้นที่ว่างปราศจากสิ่งปกคลุมต่อพื้นที่อาคารรวม)	7 ค่า: 0.0, 4.0, 4.5, 5.0, 6.0, 7.5, และ 20.0	ค่า 7.5 (39.19%)	314 ค่า

4.3 ผลจากการทำ One-hot-encoding ข้อมูลเชิงประเภท ดังภาพที่ 4.2



ภาพที่ 4.2: แสดงผลจากการทำ One-hot-encoding ข้อมูลเชิงประเภทเพื่อให้พร้อมสำหรับการวิเคราะห์

4.4 รูปแปลงที่ดินหลังจากทำความสะอาดข้อมูล ดังภาพที่ 4.3

ภาพที่ 4.3: รูปแปลงที่ดินหลังจากทำความสะอาดข้อมูล

4.5 ผลจากการวิเคราะห์ตัวแปรเพื่อหาตัวแปรที่มีความสำคัญ โดยวิธี Exploratory Regression

ผลลัพธ์ที่ได้ชี้ให้เห็นว่า "LAND_USE", "ZONE_NEW", "STREET_WID", และ "STREET_TYP" เป็นแอตทริบิวต์ที่มีความสำคัญมาก โดยมี %Significant อยู่ที่ 100% ส่วน "FAR" และ "OSR" มีระดับ Collinearity ต่อกันสูง ซึ่งในทางปฏิบัติ ต้องเลือกตัดตัวแปรตัวใดตัวหนึ่งออกไป แต่เนื่องจากว่าทั้ง 2 แอตทริบิวต์ มีระดับความสำคัญน้อยกว่า 4 แอตทริบิวต์ข้างต้น จึงไม่ได้เลือกตัวใดตัวหนึ่ง แต่ตัดออกไปทั้งคู่ ดังในภาพที่ 4.4

```
************* Exploratory Regression Global Summary (LAND_VALUE) *************
Percentage of Search Criteria Passed
                    Search Criterion Cutoff Trials # Passed % Passed
             Min Adjusted R-Squared > 0.50     62       47    75.81
            Max Coefficient p-value < 0.05     62       56    90.32
                      Max VIF Value < 7.50     62       47    75.81
            Min Jarque-Bera p-value > 0.10     62        0     0.00
Min Spatial Autocorrelation p-value > 0.10     17        0     0.00
--------------------------------------------------------------------------------
Summary of Variable Significance
Variable   % Significant % Negative % Positive
LAND_USE          100.00       0.00     100.00
STREET_TYP        100.00     100.00       0.00
ZONE_NEW          100.00      87.10      12.90
STREET_WID        100.00       0.00     100.00
FAR                93.55       0.00     100.00
OSR                87.10      87.10      12.90
--------------------------------------------------------------------------------
Summary of Multicollinearity
Variable     VIF Violations Covariates
LAND_USE    1.71     0      --------
STREET_TYP  1.01     0      --------
ZONE_NEW    2.10     0      --------
FAR        11.84     15     OSR (93.75)
OSR        12.45     15     FAR (93.75)
STREET_WID  1.65     0      --------
```

ภาพที่ 4.4: แสดงผลการรัน Exploratory Regression เพื่อค้นหาตัวแปรที่สำคัญ

4.6 ผลจากการวิเคราะห์ตัวแปรเพื่อหาตัวแปรที่มีความสำคัญ โดยวิธี Forest-based Regression

อัลกอริทึมในตระกูล “ต้นไม้ตัดสินใจ” ไม่ว่าจะเป็น Decision Tree / Random Forest / Boosted Tree ล้วนแล้วแต่มีข้อดีในแง่ของการ ranking ความสำคัญของแต่ละตัวแปรออกมาให้ด้วยทุกครั้งที่รันอัลกอริทึม ซึ่งใน ArcGIS Pro ก็มีอัลกอริทึมที่มีลักษณะเดียวกับ Random Forest ที่เรียกว่า Forest-based Classification/Regression ผลจากการรัน พบว่า “LAND_USE” และ “STREET_WID” เป็นเพียงแอตทริบิวต์ 2 ตัวที่มีความสำคัญสูงสุด โดยตัวแปรทั้ง 2 ตัวนี้ได้นำไปใช้ในการจัดกลุ่มแปลงที่ดินต่อไป

```
-------------------Model Characteristics--------------------
Number of Trees                                        1000
Leaf Size                                                 5
Tree Depth                                               10
% of Training Available per Tree                        100
Number of Randomly Sampled Variables                      2
% of Training Data Excluded for Validation               10

-----------Training Data: Regression Diagnostics------------
metricName                        score
R-Squared                          0.91
Mean Squared Error          233103257.78
Root Mean Squared Error         15267.72
Mean Absolute Error             10282.84
Explained Variance         2175907319.28
*Predictions for the data used to train the model compared to the observed categories for those features

----------Validation Data: Regression Diagnostics-----------
metricName                        score
R-Squared                          0.91
Mean Squared Error          229492857.62
Root Mean Squared Error         15149.02
Mean Absolute Error             10056.31
Explained Variance         2001564969.97
*Predictions for the validation data (excluded from model training) compared to the observed values for those test features

-----------------Top Variable of Importance-----------------
Variable                                          Importance Percentage
street_typ                                                0.01
OSR                                                       0.02
FAR                                                       0.02
zone_new                                                  0.08
land_use                                                  0.30
STREET_WID                                                0.58
```

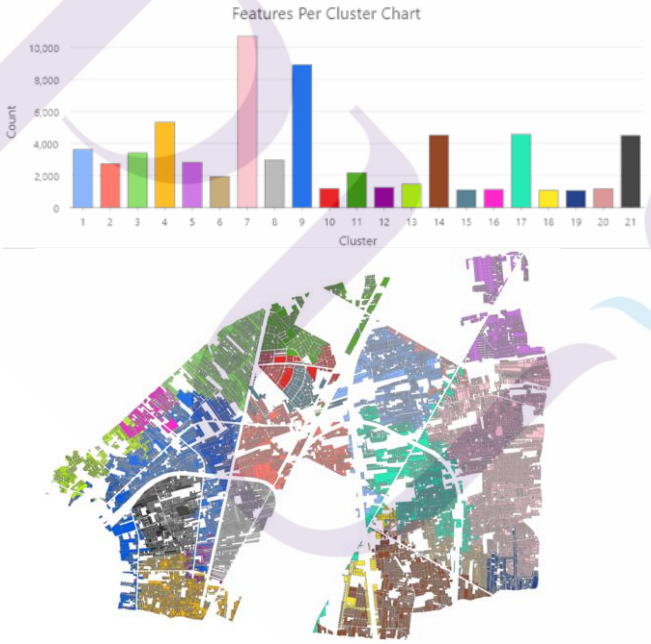
ภาพที่ 4.5: แสดงผลการรัน Forest-based Classification and Regression เพื่อค้นหาตัวแปรที่สำคัญ

4.7 ผลการจัดกลุ่มแปลงที่ดิน **รายละเอียดอยู่ในภาคผนวก ค*

ใช้วิธี Spatially Constrained Multivariate Clustering ซึ่งเป็นวิธีการแบ่งกลุ่มที่คำนึงถึงความสัมพันธ์เชิงพื้นที่ระหว่างกัน โดยมีสมมติฐานว่า แปลงที่ดินที่อยู่ในบริเวณเดียวกันหรือใกล้เคียงกันเท่านั้น จึงจะถูกจัดอยู่ในกลุ่มเดียวกัน มิเช่นนั้นแล้ว จะถูกแบ่งออกเป็นอีกกลุ่มหนึ่ง การรันอัลกอริทึมได้อาศัยเครื่องมือที่ ArcGIS Pro จัดเตรียมไว้ให้แล้ว

แปลงที่ดินถูกแบ่งออกเป็น 21 กลุ่มโดยอาศัยตัวแปรสำคัญที่ได้จากขั้นตอนก่อนหน้า (“LAND_USE”, “STREET_WID”) อย่างไรก็ดี ได้นำตัวแปร “ZONE_NEW” เข้ามาด้วยเนื่องจากเป็นข้อกำหนดจากกฎหมายผังเมือง ซึ่งจากการทดลองพบว่าตัวแปร “STREET_WID” และ “ZONE_NEW” ให้ผลลัพธ์ออกมาค่อนข้างดีจนแทบไม่มีปัญหาความกระจัดกระจายของกลุ่มแปลงที่ดิน ดังแสดงในภาพที่ 4.6

โดยผลการจัดกลุ่มหน่วยถนนด้วยอัลกอริทึม K-Means, K-Medoids และ K-Prototypes แสดงในภาคผนวก ค



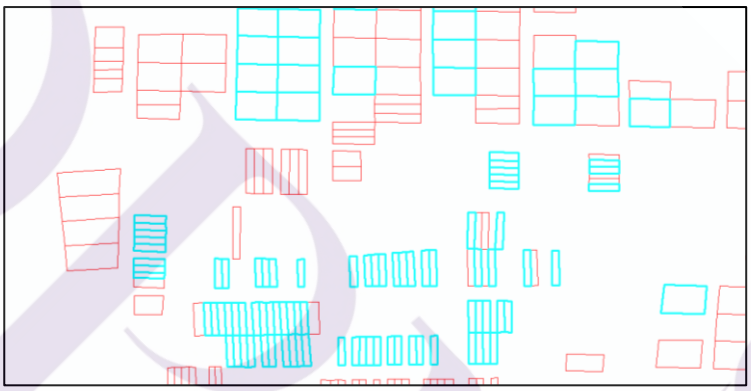
ภาพที่ 4.6: แสดงผลจากการแบ่งกลุ่มแปลงที่ดินโดยใช้วิธี Spatially Constrained Multivariate Clustering

4.8 ผลการคัดเลือกแปลงที่ดินตัวแทนของแต่ละหน่วยถนน

ผลการคัดเลือกแปลงที่ดินตัวแทนตามแนวทางที่ 1: คัดเลือกแปลงที่ดินตัวแทนจากองค์ความรู้ของผู้เชี่ยวชาญและลักษณะข้อมูล

ตัวอย่างกฎที่ใช้ในการคัดเลือกแปลงที่ดินตัวแทน ได้แก่

"DEPTH_WR" >= 0.187 AND "DEPTH_WR" <= 0.287 AND "Area" >= 53 AND "Area" <= 95 OR "DEPTH_RW" >= 1.32 AND "DEPTH_RW" <=1.65 AND "Area" >= 353 AND "Area" <= 396 ดังตัวอย่างในภาพที่ 4.7



ภาพที่ 4.7: แสดงตัวอย่างแปลงที่ดินตัวแทนที่ค้นหาได้จากองค์ความรู้ของผู้เชี่ยวชาญและลักษณะ ข้อมูล

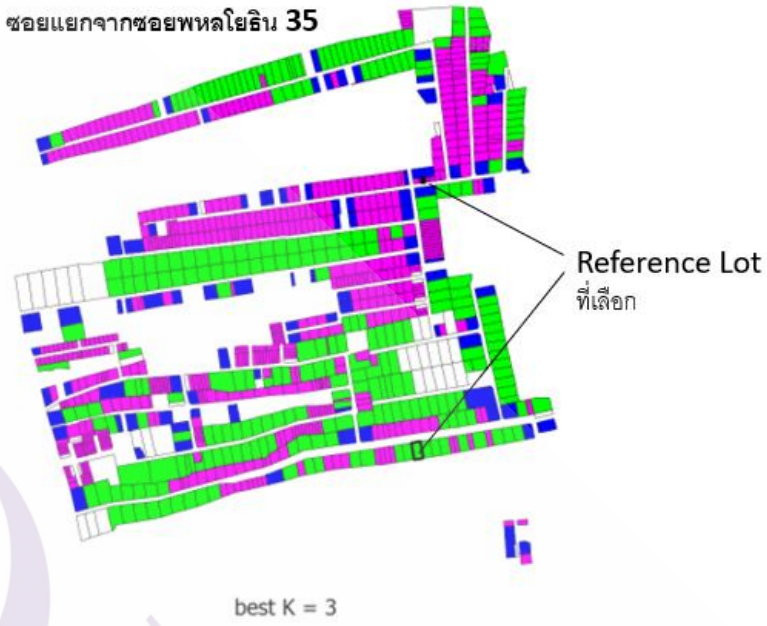
ผลการเลือกแปลงที่ดินตัวแทนตามแนวทางที่ 2: คัดเลือกแปลงที่ดินตัวแทนจากแต่ละหน่วยถนน ตามการวิเคราะห์จุดความร้อน (Hot spot analysis) ภายในแต่ละกลุ่มแปลงที่ดิน

ตัวอย่างแปลงที่ดินตัวแทนบนหน่วยถนน “ซอยแยกจากซอยพหลโยธิน 35” โดยกำหนดจากแปลงที่ดินที่มีค่า DEPTH_WR หรือ W_R_Ratio อยู่ในช่วง 0.39-0.84, มีค่า AREA อยู่ในช่วง 149-363 ตร.ม. และมีจำนวนไม่เกิน 10% ของจำนวนแปลงที่ดินทั้งหมดบนหน่วยถนนนี้ โดยแปลงส่วนเกินถูกคัดออกด้วยคะแนน GiZScore ที่ต่ำ เป็นต้น

ภาพที่ 4.8: แสดงตัวอย่างแปลงที่ดินตัวแทน (สีม่วง-ภาพกลาง) ที่ค้นหาได้จากวิธี GetisOrd Gi*

<u>ผลการเลือกแปลงที่ดินตัวแทนตามแนวทางที่ 3: คัดเลือกแปลงที่ดินตัวแทนจากการเปรียบเทียบแปลงที่ดินเชิงกายภาพบนหน่วยถนนเดียวกัน</u>

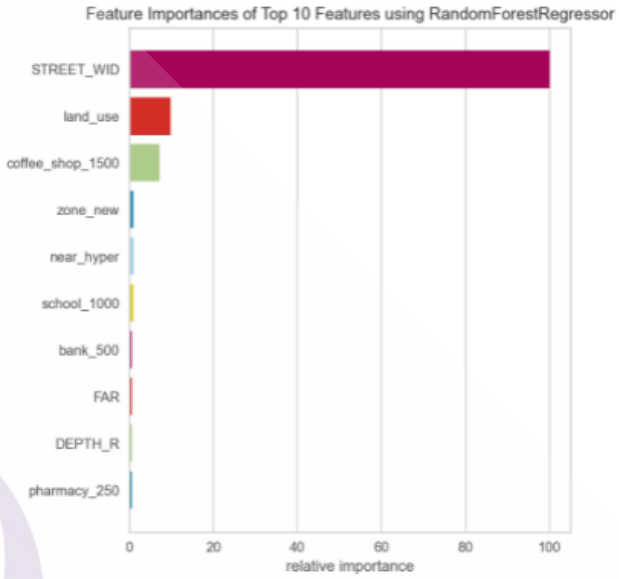
ตัวอย่างแปลงที่ดินตัวแทนบนหน่วยถนน “ซอยแยกจากซอยพหลโยธิน 35” โดยพบว่าค่า K ที่เหมาะสม คือ 3 และจำนวนคลัสเตอร์ที่ครอบคลุม 70% ของจำนวนแปลงที่ดินทั้งหมด คือ 2 คลัสเตอร์ ได้แก่ คลัสเตอร์สีม่วง และ คลัสเตอร์สีเขียว



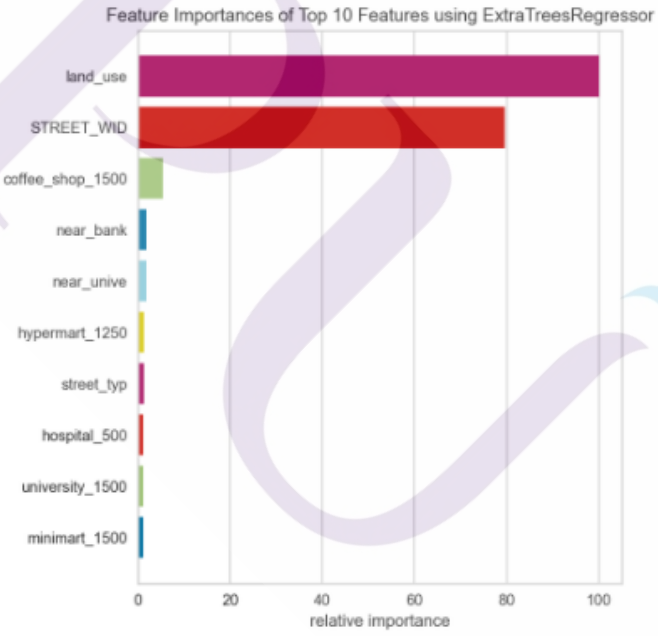
ภาพที่ 4.9: แสดงตัวอย่างแปลงที่ดินตัวแทนที่ค้นหาได้จากวิธีเปรียบเทียบเชิงกายภาพ

4.9 ผลจากการค้นหาตัวแปรที่สำคัญสำหรับแบบจำลองประเมินราคาแปลงที่ดินตัวแทน

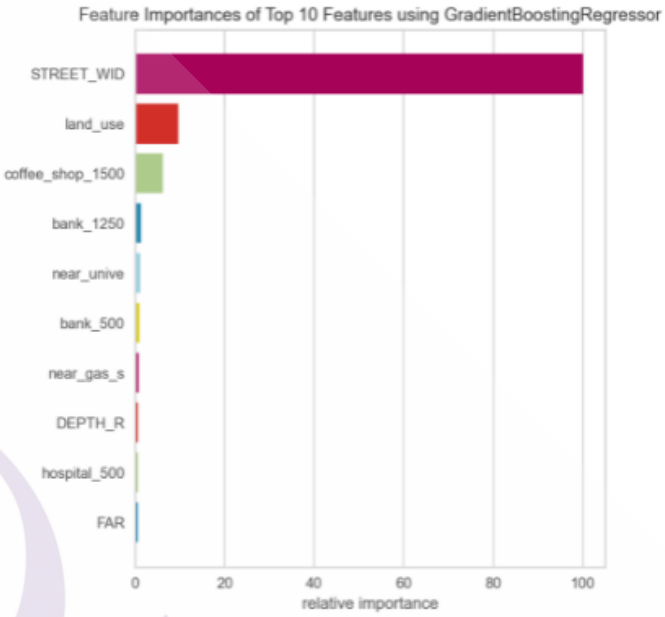
เนื่องจากจำนวนตัวแปรมีการเพิ่มขึ้นจากขั้นตอนการค้นหาแปลงที่ดินตัวแทน ทำให้ต้องทำการค้นหาตัวแปรที่สำคัญอีกครั้งหนึ่ง ซึ่งผลจากการรันอัลกอริทึม Random Forest Regressor, Extra Trees Regressor, และ Gradient Boosting Regressor สรุปว่าตัวแปรที่สำคัญสำหรับแบบจำลองประเมินราคาแปลงที่ดินตัวแทน ได้แก่ “การใช้ประโยชน์ที่ดิน”, “ความกว้างของถนน”, “จำนวนร้านกาแฟในระยะ 1500 เมตร”, และ“ระยะทางสั้นที่สุดไปยังมหาวิทยาลัย”



ภาพที่ 4.10: แสดง “ตัวแปรสำคัญ” ที่ได้จาก Random Forest Regressor



ภาพที่ 4.11: แสดง “ตัวแปรสำคัญ” ที่ได้จาก Extra Trees Regressor

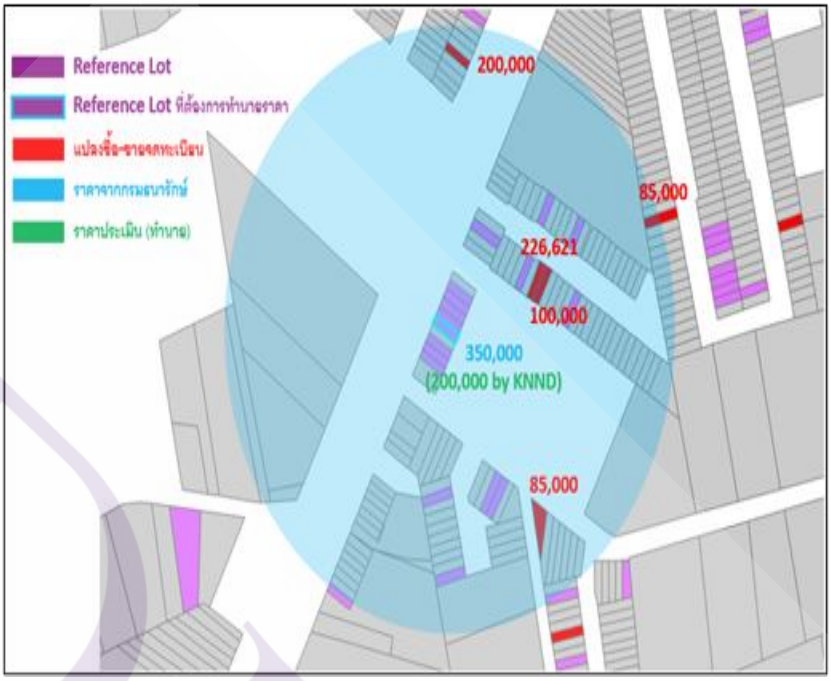


ภาพที่ 4.12: แสดง “ตัวแปรสำคัญ” ที่ได้จาก Gradient Boosting Regressor

4.10 ผลการประเมินราคาแปลงที่ดินตัวแทนจากแบบจำลอง

ผลการประเมินราคาแปลงที่ดินตัวแทน ตามแนวทางที่ 1: ประเมินราคาโดยการ buffer ออกไปหาแปลงซื้อ-ขายจดทะเบียนจนครบ K (=5) แปลง

ภาพที่ 4.13 แสดงผลการประเมินราคาแปลงที่ดินตัวแทนแปลงหนึ่งบนหน่วยถนน “ซอยแยกจากซอยพหลโยธิน 35” ตามแนวทางที่ 1 ด้วยอัลกอริทึม distance-based KNN



ภาพที่ 4.13: แสดงตัวอย่างการประเมินราคาแปลงที่ดินตัวแทนด้วย distanced-based KNN (K=3) ตามแนวทางที่ 1

ตารางที่ 4.1: แสดงผลการประเมินราคาของแปลงที่ดินตัวแทนในภาพรวม ตามแนวทางที่ 1

	แปลงที่ดินตัวแทนจากผู้เชี่ยวชาญ	แปลงที่ดินตัวแทนจาก GetisOrd Gi*	แปลงที่ดินตัวแทนจากการเปรียบเทียบเชิงกายภาพ
แปลงที่อยู่อาศัย			
▪ MAPE - Linear Regression	336%	277%	320%
▪ MAPE - KNN-U	31%	26%	28%
▪ MAPE - KNN-D	31%	26%	28%
▪ จำนวนแปลงที่ดินตัวแทน	1,303	5,182	1,688
แปลงพาณิชยกรรม			
▪ MAPE - Linear Regression	194%	130%	187%
▪ MAPE - KNN-U	56%	37%	55%
▪ MAPE - KNN-D	56%	37%	55%
▪ จำนวนแปลงที่ดินตัวแทน	127	895	129
แปลงที่อยู่อาศัยและพาณิชยกรรม			
▪ MAPE - Linear Regression	345%	261%	325%
▪ MAPE - KNN-U	38%	31%	33%
▪ MAPE - KNN-D	38%	31%	33%
▪ จำนวนแปลงที่ดินตัวแทน	1,430	6,077	1,817
Baseline - แปลงที่อยู่อาศัยและพาณิชยกรรม			
▪ MAPE - Linear Regression	27%	25%	26%

ผลการประเมินราคาแปลงที่ดินตัวแทน ตามแนวทางที่ 2: ประเมินราคาโดยใช้แปลงซื้อ-ขายจดทะเบียนที่อยู่บนหน่วยถนนเดียวกัน

ภาพที่ 4.14 แสดงผลการประเมินราคาแปลงที่ดินตัวแทนแปลงหนึ่งบนหน่วยถนน “ซอยแยกจากซอยพหลโยธิน 35” ตามแนวทางที่ 2 ด้วยอัลกอริทึม distance-based KNN

ภาพที่ 4.14: แสดงตัวอย่างผลการประเมินราคาแปลงที่ดินตัวแทนด้วย distanced-based KNN (K=3) ตามแนวทางที่ 2

ตารางที่ 4.2: แสดงผลการประเมินราคาของแปลงที่ดินตัวแทนในภาพรวม ตามแนวทางที่ 2

	แปลงที่ดินตัวแทนจากผู้เชี่ยวชาญ	แปลงที่ดินตัวแทนจาก GetisOrd Gi*	แปลงที่ดินตัวแทนจากการเปรียบเทียบเชิงกายภาพ
แปลงที่อยู่อาศัย			
วิธีปกติ			
▪ MAPE - Linear Regression	25%	25%	27%
▪ MAPE - KNN-U	18%	19%	19%
▪ MAPE - KNN-D	18%	19%	19%
▪ จำนวนแปลงที่ดินตัวแทน	640	3,987	882
วิธีพิเศษ			
▪ MAPE - Linear Regression	36%	34%	38%
▪ MAPE - KNN-U	27%	26%	28%
▪ MAPE - KNN-D	27%	26%	28%
▪ จำนวนแปลงที่ดินตัวแทน	663	1,195	805
ทั้งหมด - วิธีปกติ + วิธีพิเศษ			

ตารางที่ 4.2: (ต่อ)

▪ MAPE - Linear Regression	30%	27%	32%
▪ MAPE - KNN-U	23%	21%	23%
▪ MAPE - KNN-D	22%	21%	23%
▪ จำนวนแปลงที่ดินตัวแทน	1,303	5,182	1,687
แปลงพาณิชยกรรม			
วิธีปกติ			
▪ MAPE - Linear Regression	31%	23%	25%
▪ MAPE - KNN-U	26%	21%	20%
▪ MAPE - KNN-D	26%	21%	20%
▪ จำนวนแปลงที่ดินตัวแทน	85	765	89
วิธีพิเศษ			
▪ MAPE - Linear Regression	32%	32%	35%
▪ MAPE - KNN-U	31%	32%	35%
▪ MAPE - KNN-D	31%	32%	35%
▪ จำนวนแปลงที่ดินตัวแทน	42	130	40
ทั้งหมด - วิธีปกติ + วิธีพิเศษ			
▪ MAPE - Linear Regression	31%	24%	28%
▪ MAPE - KNN-U	27%	23%	25%
▪ MAPE - KNN-D	27%	23%	25%
▪ จำนวนแปลงที่ดินตัวแทน	127	895	129
แปลงที่อยู่อาศัยและพาณิชยกรรม			
วิธีปกติ			
▪ MAPE - Linear Regression	26%	24%	27%
▪ MAPE - KNN-U	19%	20%	19%
▪ MAPE - KNN-D	19%	20%	19%
▪ จำนวนแปลงที่ดินตัวแทน	725	4,753	971
วิธีพิเศษ			
▪ MAPE - Linear Regression	35%	34%	38%
▪ MAPE - KNN-U	27%	26%	28%

ตารางที่ 4.2: (ต่อ)

▪ MAPE - KNN-D	27%	26%	28%
▪ จำนวนแปลงที่ดินตัวแทน	705	1,324	846
ทั้งหมด - วิธีปกติ + วิธีพิเศษ			
▪ MAPE - Linear Regression	31%	27%	32%
▪ MAPE - KNN-U	23%	21%	23%
▪ MAPE - KNN-D	23%	21%	23%
▪ จำนวนแปลงที่ดินตัวแทน	1,430	6,077	1,817
Baseline - แปลงที่อยู่อาศัยและพาณิชยกรรม			
▪ MAPE - Linear Regression	27%	25%	26%

ผลการทดลองในมุมมองของ “แนวทาง”

วิธีการประเมินราคาตามแนวทางที่ 2 ดีกว่าแนวทางที่ 1 ไม่ว่าแปลงที่ดินตัวแทนจะได้มาจากวิธีผู้เชี่ยวชาญ, วิธี GetisOrd Gi*, หรือวิธีเปรียบเทียบเชิงกายภาพ และ ไม่ว่าจะเป็นอัลกอริทึม Linear Regression, KNN-D, หรือ KNN-U *โดยค่า MAPE ที่ต่ำที่สุดอยู่ที่ 21% ด้วยอัลกอริทึม KNN และแปลงที่ดินตัวแทนได้มาจากวิธี GetisOrd Gi* ซึ่งจะเห็นได้ว่า “ดีกว่า” วิธีแบบ baseline ที่ใช้การเรียนรู้จากแปลงซื้อ-ขายจดทะเบียนทั้งหมดด้วย Linear Regression (ค่าที่ดีที่สุด คือ 25%)*

การที่วิธีการประเมินราคาตามแนวทางที่ 2 ให้ผลที่ดีกว่าแนวทางที่ 1 นั้น น่าจะมีผลมาจากการใช้แปลงซื้อ-ขายจดทะเบียนที่อยู่บนหน่วยถนนเดียวกันเป็นชุดฝึกตอนทำแบบจำลอง ซึ่งหลักการนี้สอดคล้องกับวิธีการประเมินราคาของผู้เชี่ยวชาญจากกรมธนารักษ์ในปัจจุบัน สำหรับแนวทางที่ 1 นั้นใช้ลักษณะการกวาดเป็นวงกลมโดยใช้แปลงที่ดินตัวแทนเป็นศูนย์กลาง หมายความว่า แปลงซื้อ-ขายจดทะเบียนที่นำมาใช้เป็นชุดฝึกมีได้หลายรูปแบบโดยไม่ได้คำนึงถึงลักษณะเชิงกายภาพเลย (เช่น แปลงซื้อ-ขายจดทะเบียนแต่ละแปลงอาจมีหน้ากว้างไม่เท่ากัน, ลักษณะการใช้ประโยชน์ไม่เหมือนกัน เป็นต้น) ถึงแม้ว่าตัวแปรที่ใช้ในการฝึกแบบจำลองจะได้ใส่ตัวแปรเชิงกายภาพไว้แล้วก็ตาม ได้แก่ “การใช้ประโยชน์ที่ดิน” และ “ความกว้างของถนน” ซึ่งปัญหาน่าจะมาจากจำนวนชุดฝึกที่ค่อนข้างจำกัดในแต่ละแบบจำลอง (ในแต่ละหน่วยถนนนั่นเอง) ทำให้แบบจำลองเรียนรู้ไม่พอ

ผลการทดลองในมุมมองของ “อัลกอริทึม”

โดยส่วนใหญ่ ผลจาก Linear Regression พยากรณ์ได้แย่กว่า KNN-D และ KNN-U โดยเฉพาะแนวทางที่ 1 นั้น MAPE ของ Linear Regression เกินกว่า 200% ชี้ให้เห็นถึงปัญหา Overfitting อย่างชัดเจน เป็นการย้ำว่า การกวาดหาแปลงซื้อ-ขายจดทะเบียนเป็นวงกลมมีปัญหาในแง่ของความหลากหลายของชุดฝึกที่มากเกินไป ในขณะที่จำนวนชุดฝึกกลับมีน้อย ทำให้แบบจำลองไม่ได้เรียนรู้อย่างเพียงพอ ส่วนผลการพยากรณ์ของ KNN-D และ KNN-U ให้ผลงานที่ดีพอๆ กัน

ผลการทดลองในแง่มุมของ “จำนวนแปลงที่ดินตัวแทน”

จำนวนแปลงที่ดินตัวแทนที่มากส่งผลให้ความแม่นยำของแบบจำลองดีขึ้น อ้างอิงตามแนวทางที่ 2 (กรณีปกติ + กรณีพิเศษ) จะพบว่า MAPE ของวิธี KNN (ไม่ว่าจะเป็น KNN-D หรือ KNN-U) สำหรับแปลงที่ดินตัวแทนที่ได้จากวิธี GetisOrd Gi* (จำนวนแปลงที่ดินตัวแทนเท่ากับ 6,077 แปลง) มีค่าเท่ากับ 21% ซึ่งดีกว่ากรณีแปลงที่ดินตัวแทนที่ได้มาจากวิธีผู้เชี่ยวชาญ (จำนวนแปลงที่ดินตัวแทนเท่ากับ 1,430 แปลง) และวิธีเปรียบเทียบเชิงกายภาพ (จำนวนแปลงที่ดินตัวแทนเท่ากับ 1,817 แปลง) ซึ่งมี MAPE เท่ากันอยู่ที่ 23%

ผลการทดลองในแง่มุมของ “การใช้ประโยชน์ที่ดิน”

ในแง่มุมเรื่องการใช้ประโยชน์ที่ดิน พบว่า แปลงที่ดินตัวแทนที่เป็นประเภทพาณิชยกรรม มีค่า MAPE ที่มากกว่าแปลงประเภทที่อยู่อาศัย ซึ่งน่าจะเป็นเพราะว่าแปลงที่ดินส่วนใหญ่ในเขตจตุจักรเป็นประเภทที่อยู่อาศัย (83.79%) ทำให้แบบจำลองมีโอกาสเรียนรู้แปลงที่ดินประเภทพาณิชยกรรมได้จำกัด

ผลการทดลองในแง่มุมของ “วิธีปกติ และ วิธีพิเศษ” ตามแนวทางที่ 2

สำหรับแนวทางที่ 2 ถึงแม้ว่าในภาพรวมจะประเมินราคาได้ดีระดับหนึ่ง แต่แบบจำลอง “วิธีพิเศษ” ให้ผลที่แย่กว่า “วิธีปกติ” อยู่พอสมควร ทั้งๆ ที่ได้พยายามค้นหาหน่วยถนนที่มีลักษณะเชิงกายภาพใกล้เคียงกันแล้วก็ตาม แสดงว่าอาจจะมีปัจจัยแฝงอื่นๆ ที่มีอิทธิพลกับราคาประเมินในแต่ละหน่วยถนนเป็นการเฉพาะ อย่างไรก็ดี แนวทางที่ 2 นั้นได้ค้นหาหน่วยถนนใกล้เคียงเพียงแค่ 1 หน่วยถนนเท่านั้น ซึ่งอาจจะไม่สามารถสะท้อนปัจจัยแฝงเหล่านั้นได้ดีก็เป็นได้ จำเป็นต้องทำการปรับปรุงต่อไป

บทที่ 5

บทสรุปและข้อเสนอแนะ

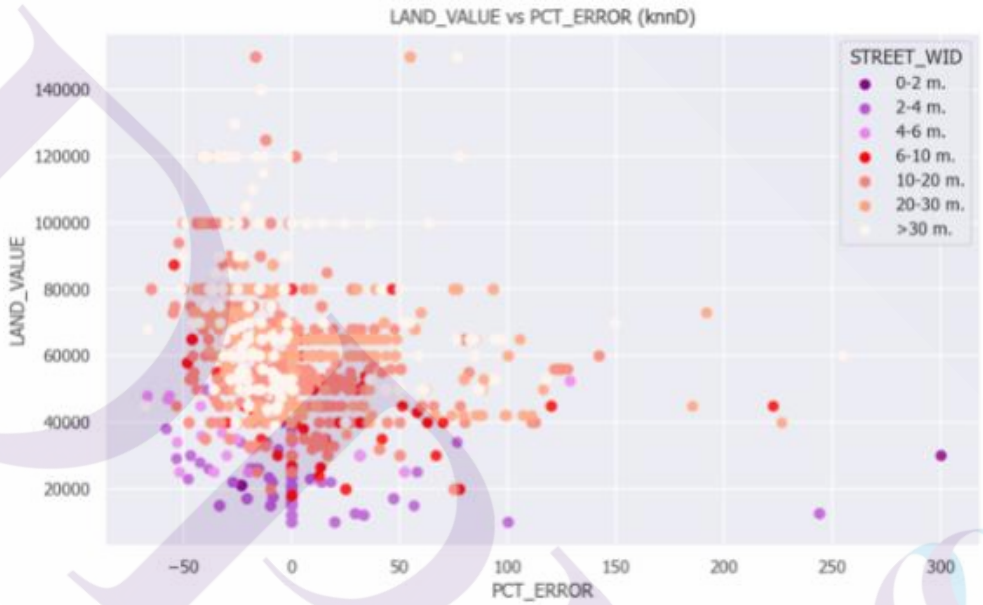
สารนิพนธ์ฉบับนี้ได้นำเสนอ 3 แนวทางในการค้นหาแปลงที่ดินตัวแทนและเสนอ 2 แนวทางในการประเมินราคาที่ดินตัวแทน โดยนำผลลัพธ์จากการพยากรณ์ไปเปรียบเทียบกับราคาประเมินที่ประกาศใช้โดยกรมธนารักษ์อยู่ในปัจจุบันของเขตจตุจักร กรุงเทพมหานคร (ประกาศใช้มาตั้งแต่ปี พ.ศ. 2559) และชี้วัดผลด้วยเมตริกซ์ MAPE (Mean absolute percentage error) หลังจากนั้น นำผลลัพธ์นี้มาเปรียบเทียบกับผลลัพธ์ที่ได้จากแบบจำลองที่เป็นวิธี Market practice

ที่มาของราคาประเมินของกรมธนารักษ์นั้นมาจากแนวคิดหลัก คือ ประเมินราคาหน่วยถนนก่อน แล้วจึงนำไปประเมินราคาที่ดินรายแปลง ก่อนจะทำการปรับปรุงค่าสุดท้ายด้วยผู้เชี่ยวชาญ ดังนั้น การนำวิธีการดั้งเดิมของกรมธนารักษ์มาเปรียบเทียบกันตรงๆ กับแนวคิดในสารนิพนธ์นี้ จึงอาจมีข้อถกเถียงในแง่มุมของความสมเหตุสมผล หมายความว่า การนำแค่ค่าความผิดพลาดมาตัดสินใจตัดแนวคิดใหม่ๆ ออกไปอาจจะไม่ได้ประโยชน์ในระยะยาว (ถึงแม้ว่ายังไม่มีวิธีอื่นใดที่ดีกว่านี้ในปัจจุบัน) แนวคิดหรือแบบจำลองที่อธิบายให้เข้าใจได้โดยง่ายก็เป็นอีกคุณสมบัติหนึ่งที่สำคัญ

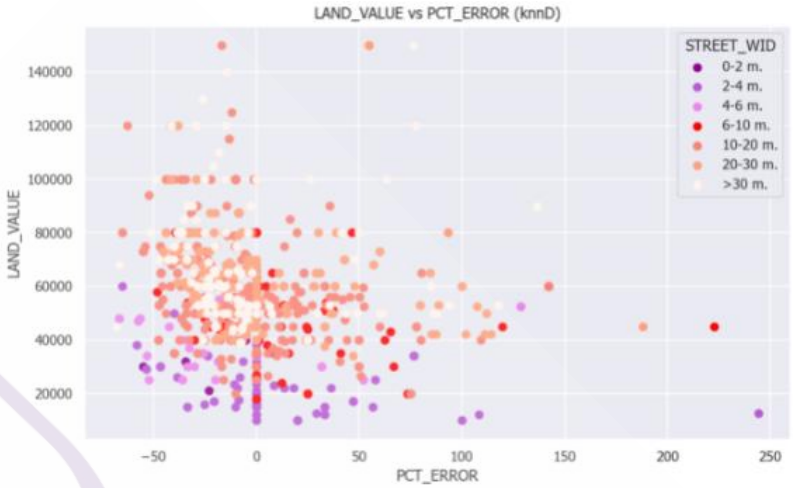
จากผลการทดลอง สรุปได้ว่า แนวคิดที่ให้ผลลัพธ์ในแง่ของ MAPE ที่ดีที่สุด คือ แนวคิดการค้นหาแปลงที่ดินตัวแทนด้วยวิธี GetisOrd Gi ผนวกกับแนวคิดการประเมินราคาแปลงที่ดินตัวแทนด้วยหลักการหน่วยถนนเดียวกัน ซึ่งได้ค่า MAPE อยู่ที่ 21% (ด้วยอัลกอริทึมประเภท KNN Regressor) ในเบื้องต้น นับว่าค่า MAPE ในระดับนี้ถือเป็นจุดเริ่มต้นที่ดี เนื่องจากสามารถเอาชนะ MAPE ที่ได้จากวิธี Market practice ซึ่งมีค่าเท่ากับ 25%*

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

อย่างไรก็ตาม หากวิเคราะห์ลึกลงไปในรายละเอียดจะพบว่า MAPE ในทุกๆ อัลกอริทึม แกว่งตัวค่อนข้างกว้าง ยกตัวอย่างดังภาพที่ 46 ด้านล่าง ซึ่งเป็น Scatter plot ระหว่าง “ราคาประเมินกรมธนารักษ์ (LAND_VALUE)” และ “Percentage Error” ที่ได้จาก KNN-D (แปลงที่ดินตัวแทนได้จากวิธี GetisOrd Gi*) จะเห็นได้ว่า มีข้อมูลหลายค่าที่ทำนายผิดพลาดมากกว่า 50%



ภาพที่ 5.1 : แสดงตัวอย่าง Scatter Plot ระหว่าง “LAND_VALUE” และ “Percentage Error” (แปลงที่ดินตัวแทนได้จากวิธี GetisOrd Gi*)



ภาพที่ 5.2 : แสดงตัวอย่าง Scatter Plot ระหว่าง “LAND_VALUE” และ “Percentage Error” (แปลงที่ดินตัวแทนได้จากวิธีผู้เชี่ยวชาญ)

การหาทางปรับปรุงค่า MAPE ให้ดีกว่า (ต่ำกว่า) 21% ถือเป็นเป้าหมายสำคัญเนื่องจากส่งผลต่อความแม่นยำในการประเมินราคาที่ดินรายแปลงต่อไป โดยมีข้อแนะนำดังต่อไปนี้

1. จากการวิเคราะห์เบื้องต้นเกี่ยวกับแปลงที่ดินตัวแทนที่มีค่าผิดพลาด MAPE สูงๆ พบว่าส่วนหนึ่งเกิดจากการพยายามปรับรอยต่อของราคาให้เรียบด้วยมือ (Manual interpolation) ยกตัวอย่างเช่น ราคาที่ประเมินมีความแตกต่างกับราคาโดยรอบอย่างมาก ทางผู้ประเมินของกรมธนารักษ์จึงจำเป็นต้องทำการปรับลดราคาลงมาด้วยมือ ดังนั้น “ทิศทางราคา” จึงเป็นสิ่งที่ควรให้ความสำคัญในอนาคต

2. จำนวนแปลงที่ดินตัวแทนที่เหมาะสมก็มีความสำคัญไม่แพ้กัน ผลจากการทดลองชี้ให้เห็นว่า จำนวนที่น้อยเกินไปส่งผลโดยตรงต่อความแม่นยำของแบบจำลองที่ลดลง

3. รูปแบบหน่วยถนนมีความซับซ้อนไม่เหมือนกัน ทำให้เกิดปัญหาการหยิบราคาซื้อ-ขายจดทะเบียนมาใช้ ราคาซื้อ-ขายที่แตกต่างกันมาก มักเกิดจากสิ่งแวดล้อมที่ไม่เหมือนกัน เช่น บางหน่วยถนนมีลักษณะเป็นรูปตัวไอยาวหลายกิโลเมตร ทำให้ราคาหัวถนนกับราคาท้ายถนนอาจมีความแตกต่างกันมาก การตัดแบ่งหน่วยถนนใหม่อาจให้ผลลัพธ์ที่ดีกว่า

4. การพิจารณานำแหล่งข้อมูลเสนอซื้อ เสนอขายเข้ามาปรับใช้ เพื่อช่วยแก้ปัญหาราคาซื้อ-ขายจดทะเบียนที่มีจำนวนน้อยและคุณภาพต่ำ

บรรณานุกรม

บรรณานุกรม

สำนักงานเขตจตุจักร. **ข้อมูลทั่วไปเขตจตุจักร**, สืบคืนเมื่อวันที่ 28 เม.ย. 2564 จาก http://www.bangkok.go.th/chatuchak

ธีรไนย ศรีธรรมรงค์ 2559. **การประเมินพื้นที่เสี่ยงการเกิดอาชญากรรมด้วยการวิเคราะห์สถิติเชิงพื้นที่.** วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยศิลปากร.

ว่าที่ร้อยตรีหญิงรัชนีพร จันทร์สา 2560. **แบบจำลองวิเคราะห์ถดถอยเชิงพื้นที่เพื่อพยากรณ์ผลการเลือกตั้งของไทย.** วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์), สถาบันบัณฑิตพัฒนบริหารศาสตร์.

อนุสรณ์ ภูมิผล 2559**. การวิเคราะห์คุณภาพน้ำของแม่น้ำน่าน ในเขตอำเภอเมือง จังหวัดพิษณุโลก ด้วยเทคนิคอัตสหสัมพันธ์เชิงพื้นที่แบบอัตราส่วนเพิ่ม.** วิทยานิพนธ์ปริญญาวิทยาศาสตรบัณฑิต, มหาวิทยาลัยนเรศวร.

Ai, P. (2005). **Residential Land Value Modelling; Case Study of Hankou, China.** International Institute for geo-information and earth observation.

Bencure, J.C.; Tripathi, N.K.; Miyazaki, H.; Ninsawat, S.; Kim, S.M. (2019). **Development of an Innovative Land Valuation Model (iLVM) for Mass Appraisal Application in Sub-Urban Areas Using AHP: An Integration of Theoretical and Practical Approaches.** Sustainability 2019, 11, 3731. https://doi.org/10.3390/su11133731

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.

Boon Foo, N. G., & Higgins, D. (2007). **Modelling the Commercial Property Market: An Empirical Study of the Singapore Office Market.** Pacific Rim Property Research Journal, 13(2), 176–193. https://doi.org/10.1080/14445921.2007.11104229

Cervero, R., & Duncan, M. (2004). **Neighbourhood composition and residential land prices: Does exclusion raise or lower values?** Urban Study, 41, 299–315.

Copenheaver, C., Kidd, K., Shockey, M., & Stephens, B. (2014). **Environmental and Social Factors Influencing the Price of Land in Southwestern Virginia, USA, 1786–1830.** Mountain Research and Development, 34(4), 386-395.

Frew, J., Jud, G.D. (2003). **Estimating the value of apartment buildings.** J. Real Estate Res. 2003, 25, 77–86.

Giussani, B., Hsia, M. & Tsolocas, S. (1992). **A Comparative Analysis of the Major Determinants of Office Rental values in Europe.** Journal of Property Valuation and Investment, 11, 157-173.

Gwamna, E., Wan Yusoff, W., & Ismail, M. (2015). **Determinants of Land Use and Property Value.** Advanced Science Letters, 21(5), 1150-1153.

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

Kilic, O., Baser, U., & Gulser, C. (2019). **Factors Explaining Urban Land Value Variability: A Case Study in Atakum District, Samsun-Turkey.** New Medit, 18(4). https://doi.org/10.30682/nm1904f

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/ijspm.2021.14293

Li, M.M., & Brown, H.J. (1980). **Micro-neighborhood externalities and hedonic housing prices.** Land Econ, 56, 125–141.

Mahan, B. L., Polasky, S., Adams, R. M. (2000). **Valuing urban wetlands: A property price approach.** Land Econ, 76, 100–113.

Mie Oak Chae, Inhyuk Kwon (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, March 19-23, 2018.

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(2019) 719-731. https://doi.org/10.1016/j.compag.2019.05.012

Pashardes, P. & Savva, C. S. (2009). **Factors Affecting House Prices in Cyprus: 1988-2008.** Cyprus Economic Policy Review, 3(1), 3-25.

Ping, H. L., Jemes, J., Fung, L. K., Yin, N. P., Maidin, N. A., & Abdul Rahman, M. S. (2019). **Factors Affecting Commercial Property Value.** International Journal Of Scientific & Technology Research, 8(12). 840-846.

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 29-30 September 2017 (ISITES2017 Baku - Azerbaijan)

Tsolacos, S., McGough, T., & Thompson, B. (2005). **Affordability and performance in the industrial property market.** Journal of Property Investment & Finance, 23(4), 311-328.

Waights, S. (2018). **Does the law of one price hold for hedonic prices?** Urban Studies, 55(15), 3299–3317. https://doi.org/10.1177/0042098017749403

Wu, Y., Zhang, X., Skitmore, M., Song, Y. & Hui, E. (2014). **Industrial land price and its impact on urban growth: A Chinese case study.** Land Use Policy, 36. 199-209.

Yang, H., Song, J., & Choi, M. (2016). **Measuring the Externality Effects of Commercial Land Use on Residential Land Value: A Case Study of Seoul.** Sustainability, 8(5), 432. https://doi.org/10.3390/su8050432

Yang, S., Hu, S., Li, W., Zhang, C., & Torres, J. (2017). **Spatiotemporal Effects of Main Impact Factors on Residential Land Price in Major Cities of China.** Sustainability, 9(11), 2050. https://doi.org/10.3390/su9112050

ภาคผนวก

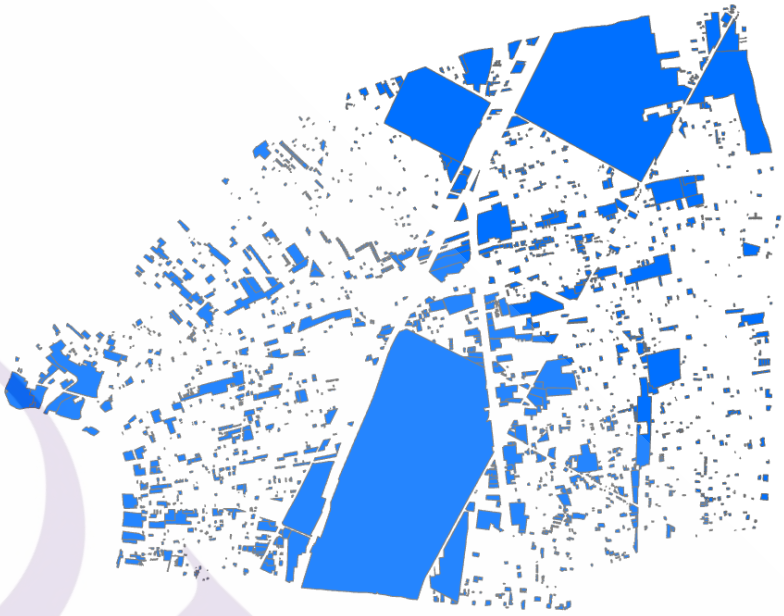
ภาคผนวก ก

รายละเอียดชั้นข้อมูลนำเข้า

1. **รูปแปลงที่ดินเขตจตุจักร กรุงเทพมหานคร**

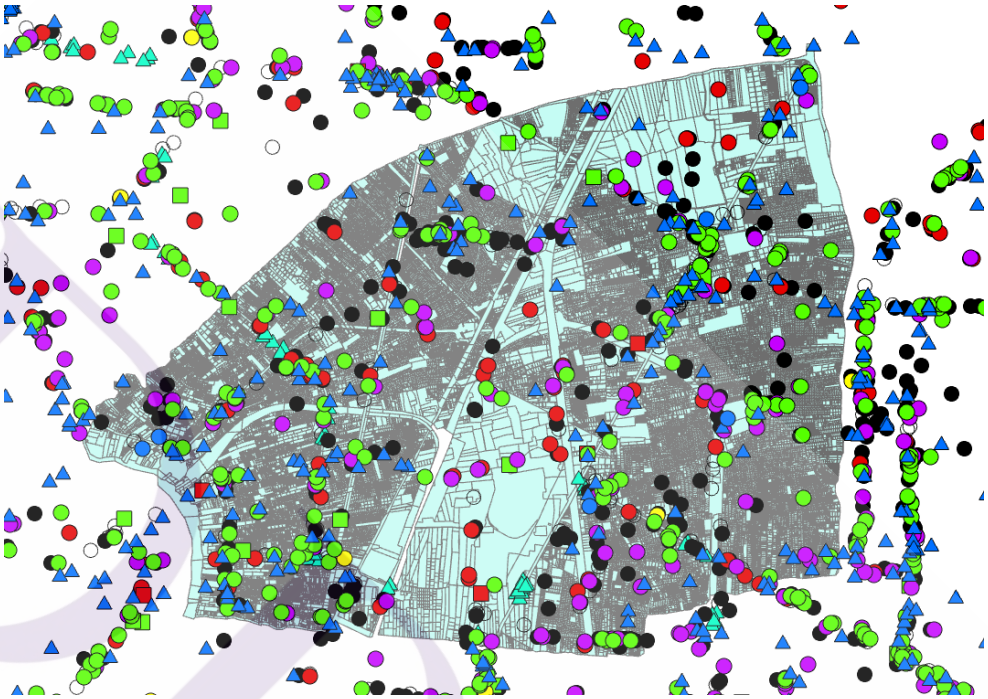
2. **รูปแปลงถนนเขตจตุจักร กรุงเทพมหานคร**

3. **รูปแปลงบล๊อคบลูเขตจตุจักร กรุงเทพมหานคร**



4. **รูปแปลงบล็อกฟิกซ์เขตจตุจักร กรุงเทพมหานคร**

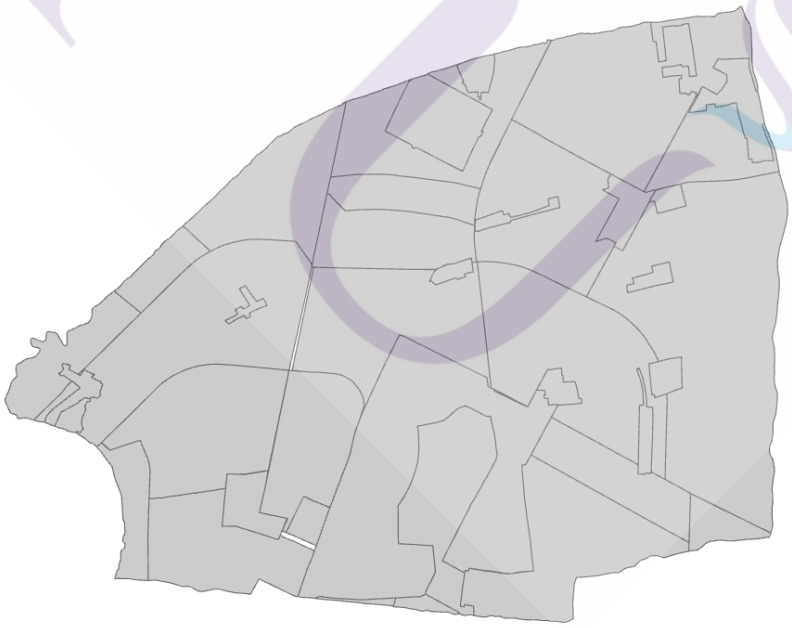
5. ข้อมูล Point of Interest (POI)



6. รูปแปลงซื้อ-ขายจดทะเบียนจากกรมที่ดิน (ปีพ.ศ. 2555 ถึง 2558)

7. ข้อมูลความลึกของแปลงที่ดินเขตจตุจักร กรุงเทพมหานคร

8. ข้อมูลการแบ่งเขตตามกฎหมายผังเมืองเขตจตุจักร กรุงเทพมหานคร



ภาคผนวก ข

รายละเอียดการสำรวจข้อมูลและทำความสะอาดข้อมูล

1. การสำรวจข้อมูลเบื้องต้น

จากการสำรวจข้อมูลแปลงที่ดินเบื้องต้น พบสิ่งที่น่าสนใจ รวมถึงปัญหาสำคัญๆ ดังนี้

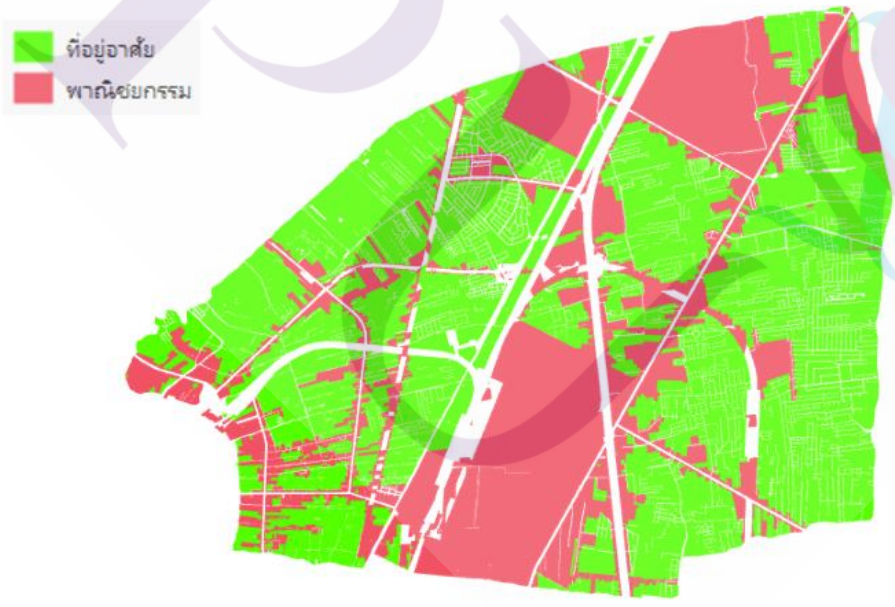
- จากภาพพบว่า “โซนผังเมือง” เป็นตัวแทนในการแยกโซนต่างๆ ได้ดีตัวหนึ่ง

ภาพแสดงแปลงที่ดินพล๊อตตามการแบ่งเขตผังเมือง

- “ความกว้างถนน” มีตั้งแต่ 0 เมตร ไปจนถึง 60 เมตร โดยกระจายเป็นถนนรองและซอยต่างๆ ที่แตกแขนงออกไปจากถนนหลักนั่นเอง

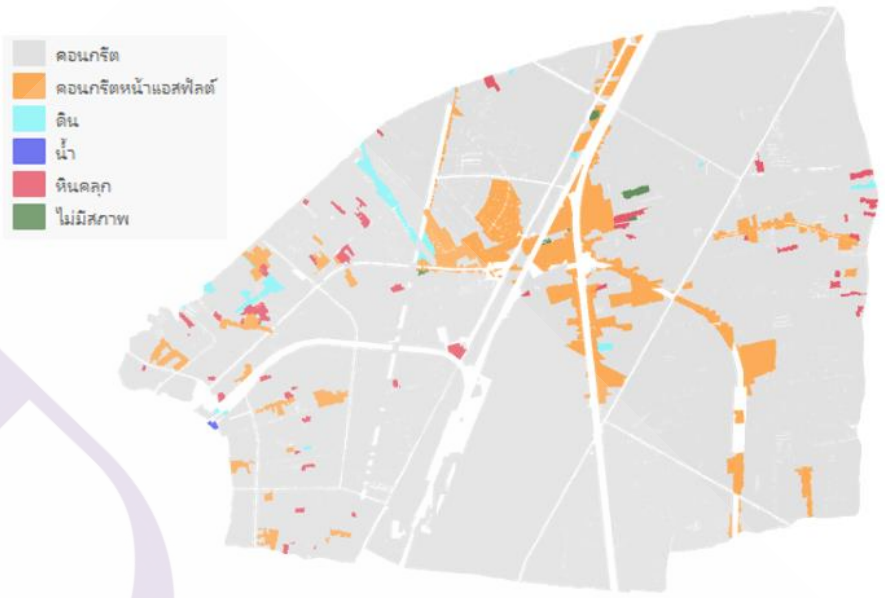
ภาพแสดงแปลงที่ดินพล๊อตตามความกว้างถนน

- ลักษณะการใช้ประโยชน์ที่ดิน มี 2 ประเภท ได้แก่ ที่อยู่อาศัย และ พาณิชยกรรม โดยแปลงที่ดินพาณิชยกรรมมักติดอยู่กับถนนหลัก รวมถึงแปลงบล๊อคบลูบางแปลง



ภาพแสดงแปลงที่ดินพล๊อตตามลักษณะการใช้ประโยชน์ที่ดิน

- ชนิดผิวจราจรส่วนใหญ่ในพื้นที่เขตจตุจักรเป็น “คอนกรีต”



ภาพแสดงแปลงที่ดินพล๊อตตามชนิดผิวจราจร

- แปลงที่ดินที่มีราคาประเมินสูง ส่วนใหญ่เป็นแปลงที่ติดอยู่กับถนนหลัก และมักเป็นแปลงพาณิชยกรรม

แปลงที่ดินพล๊อตตามราคาประเมิน โดยแยกตาม percentiles

- มีค่าสูญหายจำนวนมากในหลายๆ ตัวแปร ได้แก่
  - “STREET_COD” มีค่าสูญหายจำนวน 803 แปลง
  - “STREET_SMG” มีค่าสูญหายจำนวน 37,089 แปลง
  - “STREET_NAM” มีค่าสูญหายจำนวน 803 แปลง
  - “AMPHUR_NAM” มีค่าสูญหายจำนวน 11,741 แปลง
  - “TUMBON_NAME” มีค่าสูญหายจำนวน 11,741 แปลง
  - “WIDTH” มีค่าสูญหายจำนวน 37,544 แปลง
  - “LAND_USE” มีค่าสูญหายจำนวน 2,794 แปลง
  - “STREET_TYP” มีค่าสูญหายจำนวน 2,794 แปลง
  - “ZONE_NEW” มีค่าสูญหายจำนวน 351 แปลง
  - “FAR” มีค่าสูญหายจำนวน 314 แปลง
  - “OSR” มีค่าสูญหายจำนวน 314 แปลง
- มีแปลงถนนต่างๆ รวมอยู่กับแปลงที่ดิน ซึ่งจำเป็นต้องตัดออก

กรอบสีฟ้าแสดงแปลงถนนและซอย/ทางที่รวมกันอยู่กับแปลงที่ดิน

- มีแปลงที่ดินหลายแปลงที่ไม่มีราคาประเมิน (LAND_VALUE = 0) จากการตรวจสอบเบื้องต้น พบว่าส่วนใหญ่เป็นแปลงถนนต่างๆ

กรอบสีฟ้าแสดงแปลงที่ดินที่ไม่มีราคาประเมิน

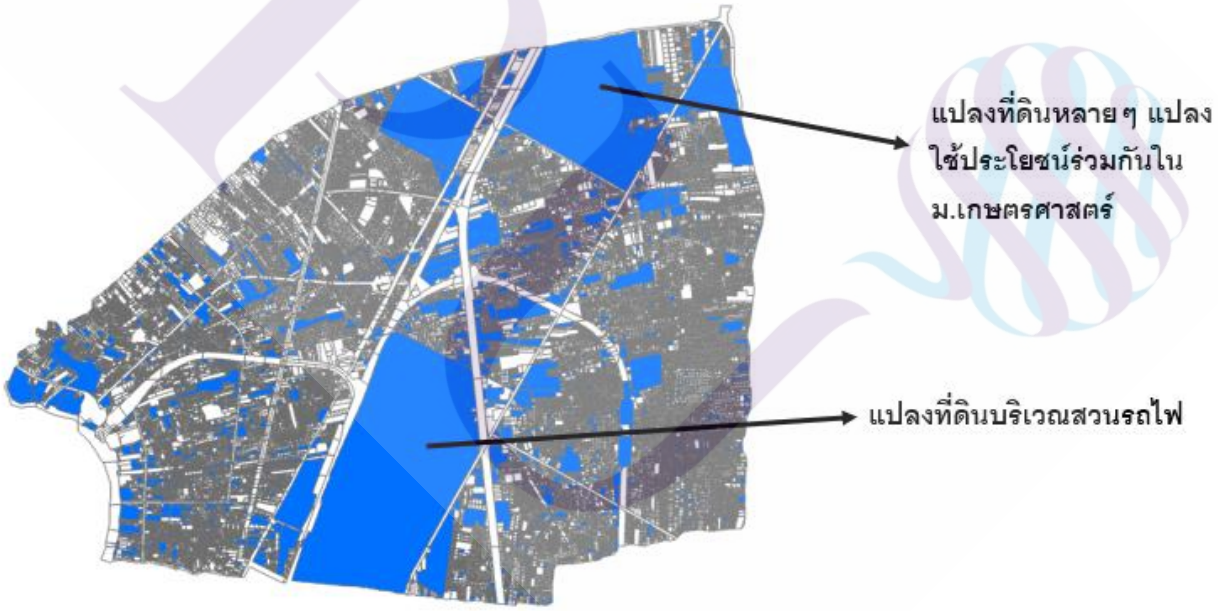
- มีจำนวนตัวแปรที่เป็นเชิงประเภท categorical อยู่ 16 ตัวแปร ซึ่งจำเป็นต้องทำการแปลงเป็นตัวเลขด้วยเทคนิคต่างๆ เช่น ONE-HOT-ENCODING ก่อนนำไปใช้งาน
- รูปแปลง geometry มีความเสียหายในเชิง topology อยู่หลายแปลง เช่น เส้นขอบระหว่างแปลงทับกันไม่พอดี เกิด void ระหว่างแปลง, รูปแปลงเป็นสี่เหลี่ยมปกติ แต่กลับมีหลาย polygon เป็นต้น

นอกจากนี้ มีรูปแบบแปลงที่ดินบางประเภทที่จำเป็นต้องตัดออกไปก่อน เนื่องจากทำการประเมินราคาด้วยวิธีพิเศษ ได้แก่

1) แปลงที่ดินบล๊อคฟิกซ์ หมายถึง แปลงที่ดินที่ต่อมาถูกรวมเข้ากันและพัฒนาไปเป็น “หมู่บ้านจัดสรร” หรือ “คอนโดมิเนียม”

สีแดงแสดงแปลงบล๊อคฟิกซ์ในเขตจตุจักร

2) แปลงที่ดินบล๊อคบลู หมายถึง แปลงที่ดินที่ใช้ประโยชน์ร่วมกัน เช่น แปลงที่ดินบริเวณมหาวิทยาลัยเกษตรศาสตร์ ซึ่งเกิดจากแปลงที่ดินหลายๆ แปลงมารวมใช้งานด้วยกัน

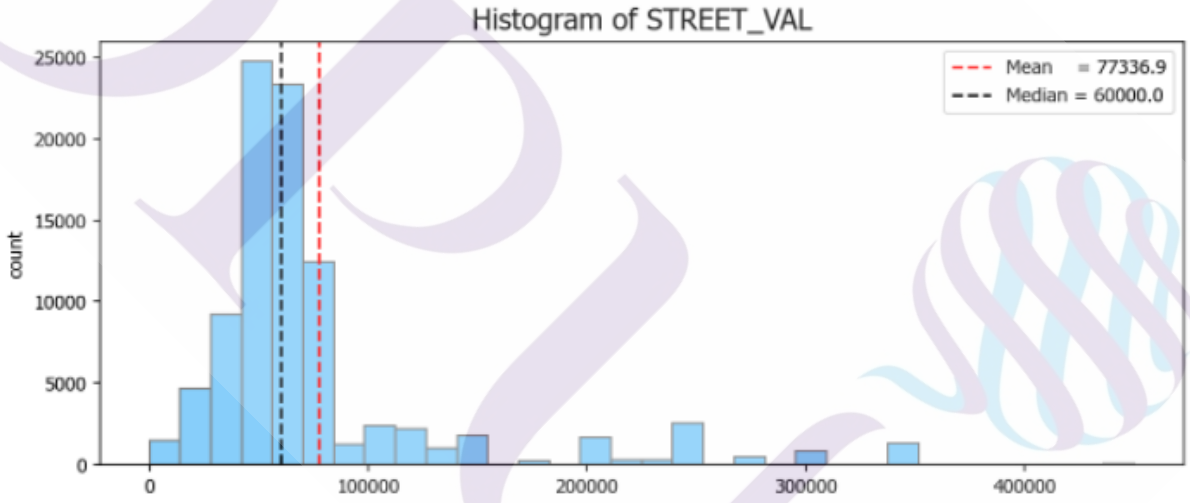


สีน้ำเงินแสดงแปลงบล๊อคบลูในเขตจตุจักร

2. การสำรวจข้อมูลเชิงลึก

2.1 ผลการสำรวจข้อมูลโดยวิธี Univariate Analysis สำหรับข้อมูลเชิง “ตัวเลข” สำหรับแปลงที่ดินตั้งต้น (91,882 แปลง) มีรายละเอียดดังต่อไปนี้

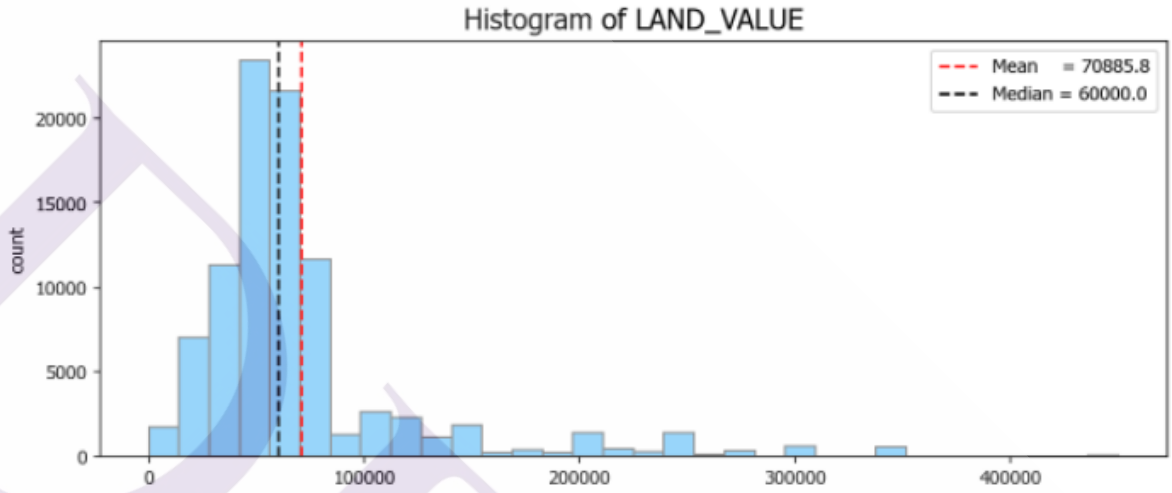
- “STREET_VAL” หรือ “มูลค่าถนน”
  - มีพิสัยตั้งแต่ 0 ถึง 450,000 บาท ต่อ ตารางวา
  - มีค่าเฉลี่ย (simple average) อยู่ที่ 77,336.90 บาท ต่อ ตารางวา
  - มีค่ามัธยฐาน (median) อยู่ที่ 60,000.00 บาท ต่อ ตารางวา
  - ค่า STREET_VAL = 0 บ่งบอกถึงค่าสูญหาย (missing values)
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 803 ค่า (0.87%)



ฮิสโทแกรมของตัวแปร “STREET_VAL”

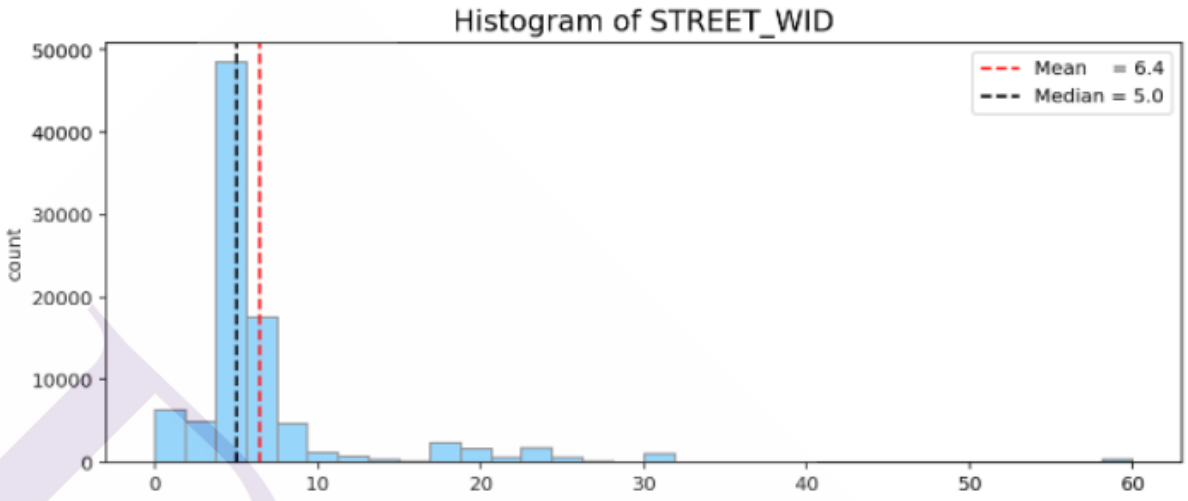
- “LAND_VALUE” หรือ “ราคาประเมิน”
  - เป็นราคาประเมินของกรมธนารักษ์ประกาศใช้เมื่อปีพ.ศ. 2559
  - ***ถือเป็นตัวแปรเป้าหมาย (target) ที่ใช้เปรียบเทียบผลลัพธ์ในงานวิจัยนี้***
  - มีพิสัยตั้งแต่ 0 ถึง 450,000 บาท ต่อ ตารางวา
  - มีค่าเฉลี่ย (simple average) อยู่ที่ 70,885.8 บาท ต่อ ตารางวา

- มีค่ามัธยฐาน (median) อยู่ที่ 60,000.00 บาท ต่อ ตารางวา
- ค่า LAND_VALUE = 0 บ่งบอกถึงค่าสูญหาย (missing values)
- จำนวนค่าสูญหาย (missing values) อยู่ที่ 806 ค่า (0.88%)



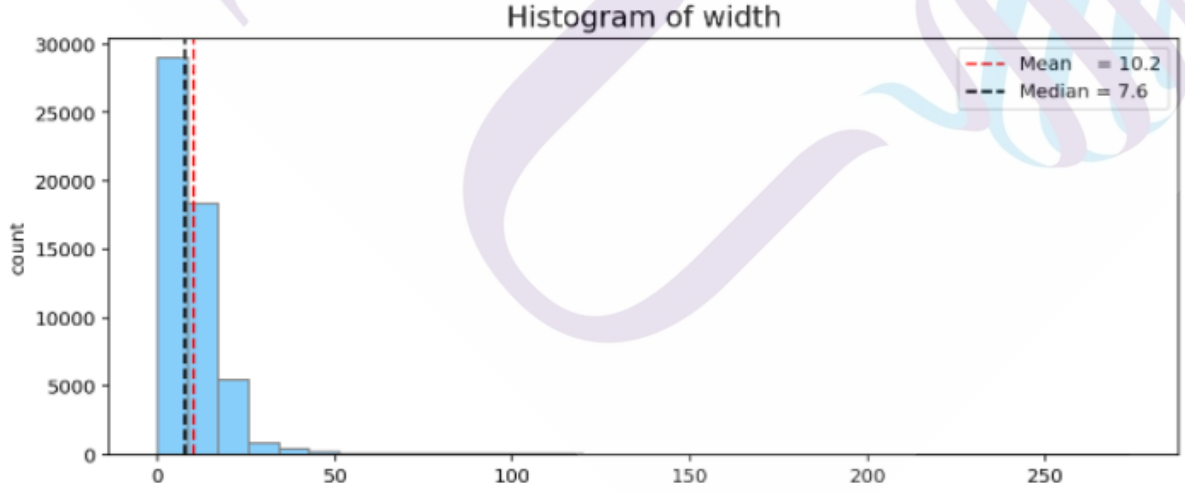
ฮิสโทแกรมของตัวแปร "STREET_VAL"

- "STREET_WID" หรือ ความกว้างของถนน
  - มีพิสัยตั้งแต่ 0 ถึง 60 เมตร
  - มีค่าเฉลี่ย (simple average) อยู่ที่ 6.4 เมตร
  - มีค่ามัธยฐาน (median) อยู่ที่ 5.0 เมตร
  - ไม่พบค่าสูญหาย (missing values)



ฮิสโทแกรมของตัวแปร "STREET_WID"

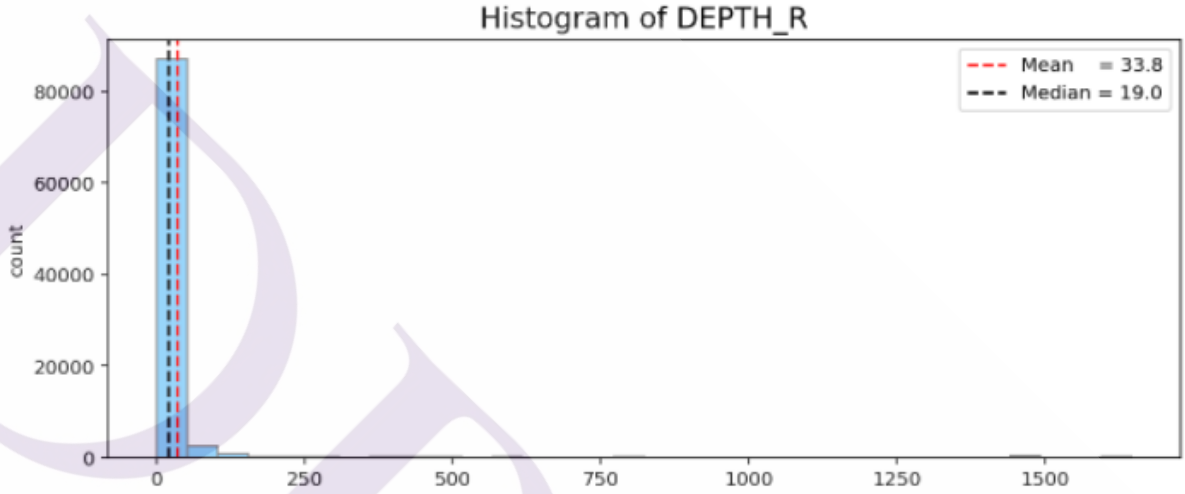
- "WIDTH" หรือ ความกว้างของแปลงที่ดิน
  - มีพิสัยตั้งแต่ 0.09 ถึง 273.90 เมตร
  - มีค่าเฉลี่ย (simple average) อยู่ที่ 10.2 เมตร
  - มีค่ามัธยฐาน (median) อยู่ที่ 7.6 เมตร
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 37,544 ค่า (40.86%)



ฮิสโทแกรมของตัวแปร "WIDTH"

- "DEPTH_R" หรือ "อัตราส่วนพื้นที่อาคารรวมต่อพื้นที่ดิน"

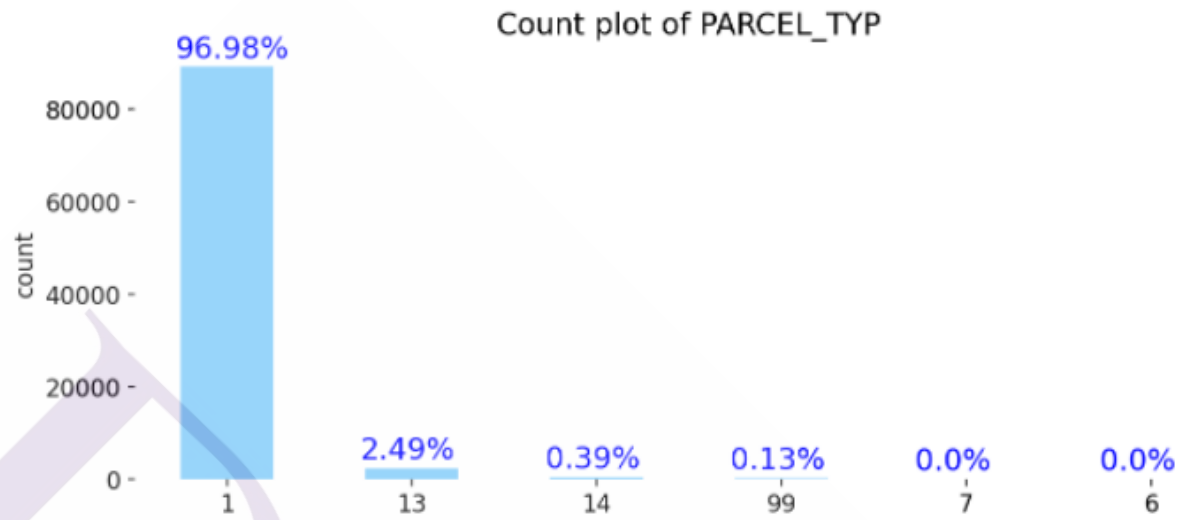
- มีพิสัยตั้งแต่ 0.0 ถึง 1,647.0 เมตร
- มีค่าเฉลี่ย (simple average) อยู่ที่ 33.8 เมตร
- มีค่ามัธยฐาน (median) อยู่ที่ 19.0 เมตร
- ไม่พบค่าสูญหาย (missing values)



ฮิสโทแกรมของตัวแปร "DEPTH_R"

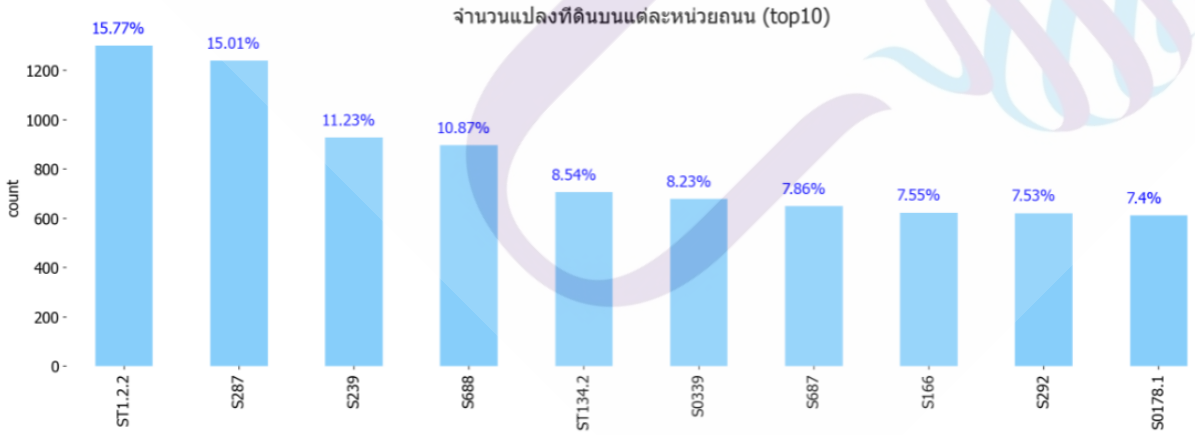
2.2 ผลการสำรวจข้อมูลโดยวิธี Univariate Analysis สำหรับข้อมูลเชิง "ประเภท" สำหรับแปลงที่ดินตั้งต้น (91,882 แปลง) มีรายละเอียดดังต่อไปนี้

- "PARCEL_TYP" หรือ "ชนิด/ลักษณะของแปลงที่ดิน"
  - มีทั้งหมด 6 ประเภท ได้แก่ 1 (แปลงปกติ), 6, 7, 13, 14, และ 99
  - ส่วนใหญ่เป็นประเภท 1 คือ แปลงปกติ (96.98%)
  - ไม่พบค่าสูญหาย (missing values)



Count plot ของตัวแปร "PARCEL_TYP"

- "STREET_COD" หรือ "รหัสถนน/หน่วยถนน"
  - เป็นรหัสที่กำหนดโดยกรมธนารักษ์
  - มีทั้งหมด 1,335 รหัสในเขตจตุจักร
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 803 ค่า (0.87%)



แสดงจำนวนแปลงที่ดินบนแต่ละหน่วยถนน – top 10

- “STREET_NAM” หรือ “ชื่อถนน” – เป็นชื่อที่ใช้สื่อความหมายกับ STREET_COD จึงมีลักษณะเช่นเดียวกับ STREET_COD
- “STREET_SMG” หรือ “รหัสถนนย่อย”
  - เป็นรหัสที่กำหนดโดยกรมธนารักษ์ มีทั้งหมด 4 หลัก ได้แก่

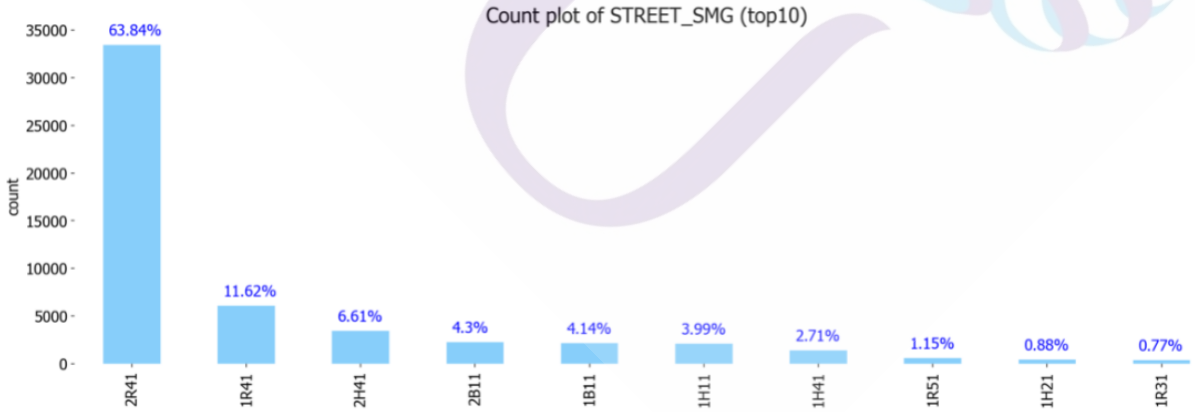
    รหัสตัวแรก คือ โซนพื้นที่ดำเนินการ เช่น 1, 2, 3 เป็นต้น

    รหัสตัวที่สอง คือ การกำหนดรหัสการใช้ประโยชน์ที่ดิน เช่น R หมายถึง กลุ่มที่อยู่อาศัย, B หมายถึง กลุ่มพาณิชยกรรม เป็นต้น

    รหัสตัวที่สาม คือ การกำหนดรหัสที่ตั้งของแปลงที่ดิน เช่น 1 คือ ถนนหลัก, 2 คือ ถนนรอง, 3 คือ ซอยเชื่อม, 4 คือ ซอย เป็นต้น

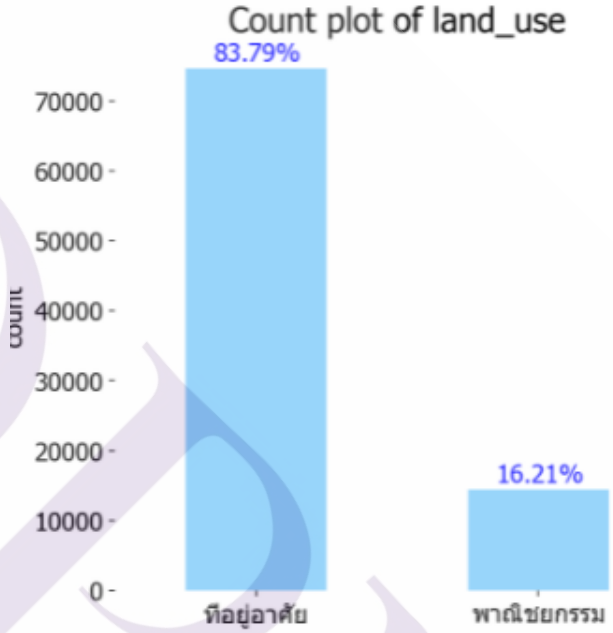
    รหัสตัวที่สี่ คือ การกำหนดรหัสผิวจราจร เช่น 1 คือ คอนกรีต/คอนกรีตหน้าแอสฟัลต์, 2 คือ ลาดยาง, 3 คือ หินคลุก/ลูกรัง เป็นต้น

  - มีทั้งหมด 29 รหัสในเขตจตุจักร
  - รหัสที่มีจำนวนมากที่สุด คือ 2R41 (พื้นที่ดำเนินการโซน 2, กลุ่มที่อยู่อาศัย, ซอย, ถนนคอนกรีต)
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 37,089 ค่า (40.37%)



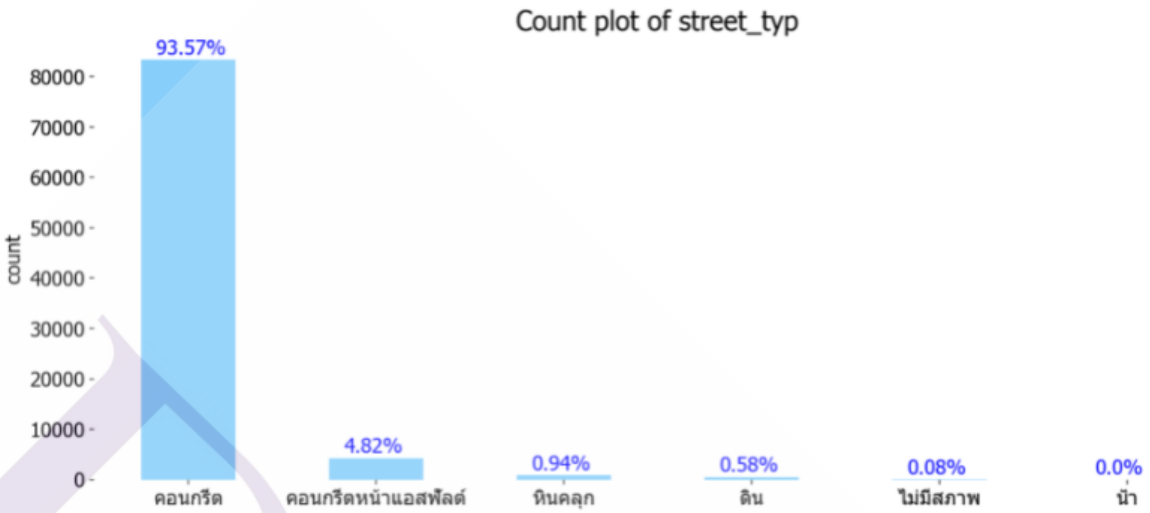
Count plot ของตัวแปร “STREET_SMG” – top 10

- “LAND_USE” หรือ “ลักษณะการใช้ประโยชน์ที่ดิน”
  - ในเขตจตุจักรมี 2 ประเภทหลัก คือ ที่อยู่อาศัย และ พาณิชยกรรม
  - ส่วนใหญ่เป็นประเภทที่อยู่อาศัย (83.79%)
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 2,794 ค่า (3.04%)



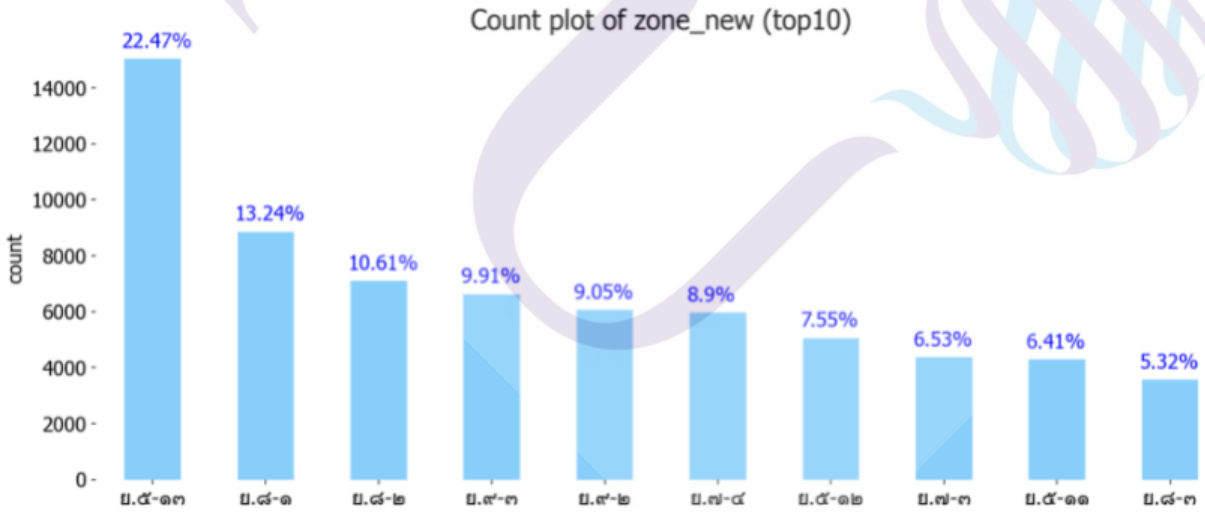
Count plot ของตัวแปร “LAND_USE”

- “STREET_TYP” หรือ “ชนิดผิวจราจร”
  - ในเขตจตุจักรมี 6 ประเภท คือ คอนกรีต, คอนกรีตหน้าแอสฟัลต์, หินคลุก, ดิน, ไม่มีสภาพ, และ น้ำ
  - ส่วนใหญ่เป็นประเภทคอนกรีต (93.57%)
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 2,794 ค่า (3.04%)



Count plot ของตัวแปร "STREET_TYP"

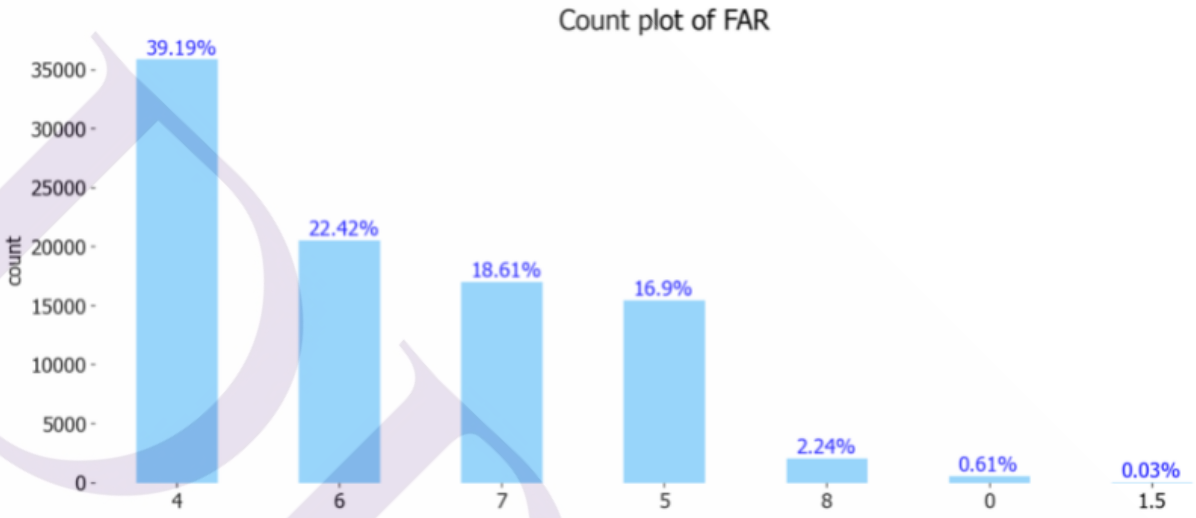
- "ZONE_NEW" หรือ "โซนผังเมือง"
  - ในเขตจตุจักรมีทั้งหมด 36 โซน เป็นโซนที่แบ่งตามกฎหมายผังเมืองซึ่งเป็นแบบแผนเพื่อการพัฒนาเมืองให้เป็นไปตามทิศทางที่เหมาะสม
  - ส่วนใหญ่เป็นโซนประเภท "ย" ซึ่งหมายถึง โซนประเภทที่อยู่อาศัย
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 351 ค่า (0.38%)



Count plot ของตัวแปร "ZONE_NEW" – top 10

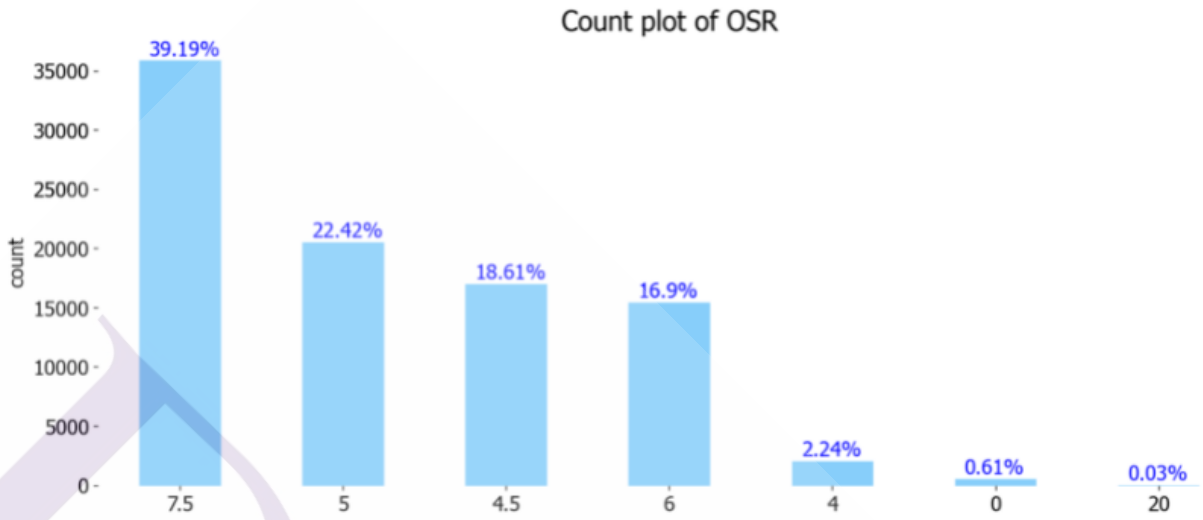
- "FAR" หรือ "อัตราส่วนพื้นที่อาคารรวมต่อพื้นที่ดิน"

- มี 7 อัตราส่วนในเขตจตุจักร ได้แก่ 0.0, 1.5, 4.0, 5.0, 6.0, 7.0, และ 8.0
- ส่วนใหญ่เป็นอัตราส่วน 4.0 หมายถึง สามารถสร้างอาคารหรือบ้านโดยมีพื้นที่รวมไม่เกิน 4 เท่าของพื้นที่แปลงที่ดินทั้งหมด
- จำนวนค่าสูญหาย (missing values) อยู่ที่ 314 ค่า (0.34%)



Count plot ของตัวแปร "FAR"

- "OSR" หรือ "อัตราส่วนพื้นที่ว่างปราศจากสิ่งปกคลุมต่อพื้นที่อาคารรวม"
  - มี 7 อัตราส่วนในเขตจตุจักร ได้แก่ 0.0, 4.0, 4.5, 5.0, 6.0, 7.5, และ 20.0
  - ส่วนใหญ่เป็นอัตราส่วน 7.5 หมายถึง ต้องจัดสรรพื้นที่ว่าง (ที่ไม่โดนปกคลุม) ให้มีอย่างน้อย 7.5 เท่าของพื้นที่อาคารรวม
  - จำนวนค่าสูญหาย (missing values) อยู่ที่ 314 ค่า (0.34%)



Count plot ของตัวแปร "OSR"

2.3 สรุปผลการสำรวจข้อมูลเชิงลึก

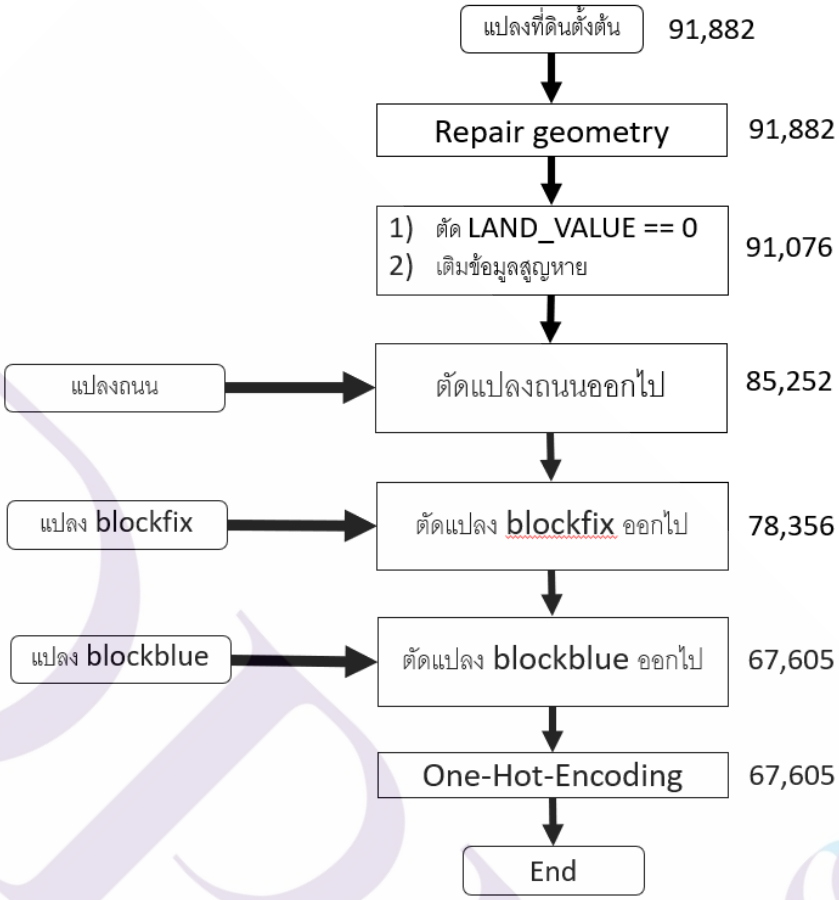
- แปลงที่ดินตั้งต้น (91,882 แปลง) จำเป็นต้องแก้ไขและสกัดส่วนที่ไม่เกี่ยวข้องออกไปก่อน ได้แก่
  - แก้ไขปัญหา geometry ที่มี topology เสียหาย
  - แปลงถนน ซึ่งจำเป็นต้องใช้ชั้นแปลงถนนเข้ามาช่วยสกัด
  - แปลงบล๊อคบลู ซึ่งจำเป็นต้องใช้ชั้นแปลงบล๊อคบลูเข้ามาช่วยสกัด
  - แปลงบล็อกฟิกซ์ ซึ่งจำเป็นต้องใช้ชั้นแปลงบล๊อคฟิกซ์เข้ามาช่วยสกัด
- ตัวแปรที่ใช้ชี้วัดผลการวิจัย ได้แก่ LAND_VALUE ซึ่งเป็นราคาประเมินจากกรมธนารักษ์ ซึ่งในที่นี้ประกาศใช้เมื่อปีพ.ศ. 2559
- ตัวแปรต้นที่มีแนวโน้มนำไปใช้ประโยชน์ต่อในงานวิจัย ได้แก่
  - STREET_COD – ตัวแปรเชิงประเภท
  - WIDTH – ตัวแปรเชิงตัวเลข
  - LAND_USE – ตัวแปรเชิงประเภท
  - STREET_TYP – ตัวแปรเชิงประเภท

- ZONE_NEW – ตัวแปรเชิงประเภท
- FAR – ตัวแปรเชิงประเภท
- OSR – ตัวแปรเชิงประเภท
- DEPTH_R – ตัวแปรเชิงตัวเลข
- STREET_WID – ตัวแปรเชิงตัวเลข

- ตัวแปรเหล่านี้จำเป็นต้องแก้ปัญหาเรื่องค่าสูญหาย (missing values) ได้แก่ STREET_COD, WIDTH, LAND_USE, STREET_TYP, ZONE_NEW, FAR, และ OSR
- ตัวแปรเชิงประเภทซึ่งจำเป็นต้องแปลงให้เป็นตัวเลขด้วยเทคนิคต่างๆ เช่น LabelEncoder หรือ OneHotEncoder ได้แก่ LAND_USE, STREET_TYP, ZONE_NEW, FAR, และ OSR (รวมถึง STREET_COD ถ้าจำเป็นในภายหลัง)

3. การทำความสะอาดข้อมูล

ขั้นตอนการทำความสะอาดข้อมูลเป็นตามกระบวนการด้านล่าง



กระบวนการทำความสะอาดข้อมูล

ขั้นตอนที่ 1: ทำการซ่อมแซมแก้ไข geometry ที่มีปัญหาด้วยเครื่องมือจาก ArcGIS Pro

ขั้นตอนที่ 2: ทำการตัดแปลงที่ดินที่ไม่มีราคาประเมินออกไป (LAND_VALUE = 0.0) ซึ่งส่วนใหญ่เป็นแปลงถนน หลังจากนั้น ให้ทำการเติมค่าสูญหายด้วยวิธีการหาฐานนิยมจากแปลงที่ดินพื้นบ้านที่ติดกันจำนวน 9 แปลงที่ดิน ซึ่งหลังจากการทำขั้นตอนนี้แล้วจะเหลือแปลงที่ดินอยู่ 91,076 แปลง

ขั้นตอนที่ 3: ทำการตัดแปลงถนนออกไป โดยนำชั้นแปลงถนนมาทาบลงบนแปลงที่ดินตั้งต้นแล้วตัดส่วนที่ทับกันออกไป หลังจากการทำขั้นตอนนี้แล้วจะเหลือแปลงที่ดินอยู่ 85,252 แปลง

ขั้นตอนที่ 4: ทำการตัดแปลงบล๊อคฟิกซ์ออกไป โดยนำชั้นแปลงบล๊อคฟิกซ์มาทาบลงบนแปลงที่ดินตั้งต้นแล้วตัดส่วนที่ทับกันออกไป หลังจากการทำขั้นตอนนี้แล้วจะเหลือแปลงที่ดินอยู่ 78,356 แปลง

ขั้นตอนที่ 5: ทำการตัดแปลงบล๊อคบลูออกไป โดยนำชั้นแปลงบล๊อคบลูมาทาบลงบนแปลงที่ดินตั้งต้นแล้วตัดส่วนที่ทับกันออกไป หลังจากการทำขั้นตอนนี้แล้วจะเหลือแปลงที่ดินอยู่ 67,605 แปลง

ขั้นตอนที่ 6: ทำการแปลงค่าข้อความในแต่ละตัวแปรเชิงประเภทให้กลายเป็นเชิงตัวเลขด้วยเทคนิค One-Hot-Encoding

- ตัวแปร STREET_WID หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: STREET_WID 0-2 เมตร
  - กลุ่ม 1: STREET_WID 2-4 เมตร
  - กลุ่ม 2: STREET_WID 4-6 เมตร
  - กลุ่ม 3: STREET_WID 6-10 เมตร
  - กลุ่ม 4: STREET_WID 10-20 เมตร
  - กลุ่ม 5: STREET_WID 20-30 เมตร
  - กลุ่ม 6: STREET_WID ตั้งแต่ 30 เมตร เป็นต้นไป
- ตัวแปร LAND_USE หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: ที่อยู่อาศัย
  - กลุ่ม 1: พาณิชยกรรม
- ตัวแปร ZONE_NEW หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: โซน 1
  - กลุ่ม 1: โซน 2
  - ……
  - กลุ่ม 35: โซน 36

- ตัวแปร STREET_TYP หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: ไม่มีสภาพ
  - กลุ่ม 1: หินคลุก
  - กลุ่ม 2: น้ำ
  - กลุ่ม 3: ดิน
  - กลุ่ม 4: คอนกรีตหน้าแอสฟัลต์
  - กลุ่ม 5: คอนกรีต
- ตัวแปร FAR หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: อัตราส่วน 0.0
  - กลุ่ม 1: อัตราส่วน 1.5
  - กลุ่ม 2: อัตราส่วน 4.0
  - กลุ่ม 3: อัตราส่วน 5.0
  - กลุ่ม 4: อัตราส่วน 6.0
  - กลุ่ม 5: อัตราส่วน 7.0
  - กลุ่ม 6: อัตราส่วน 8.0
- ตัวแปร OSR หลังจากทำ One-hot-encoding แล้วจะได้ตัวแปรใหม่ดังนี้
  - กลุ่ม 0: อัตราส่วน 0.0
  - กลุ่ม 1: อัตราส่วน 4.0
  - กลุ่ม 2: อัตราส่วน 4.5
  - กลุ่ม 3: อัตราส่วน 5.0
  - กลุ่ม 4: อัตราส่วน 6.0
  - กลุ่ม 5: อัตราส่วน 7.5
  - กลุ่ม 6: อัตราส่วน 20.0

แปลงที่ดินหลังจากทำความสะอาดข้อมูล (67,605 แปลง)

ภาคผนวก ค

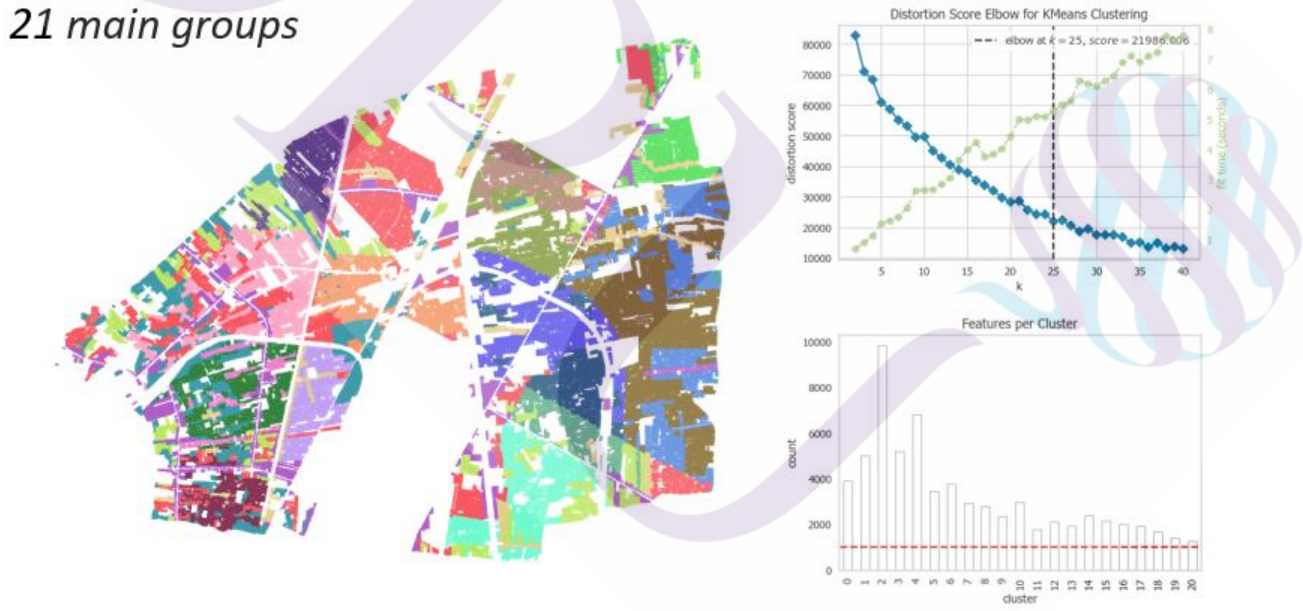
รายละเอียดการจัดกลุ่มแปลงที่ดิน

ในหัวข้อนี้ ต้องการที่จะจัดกลุ่มแปลงที่ดินที่มีคุณลักษณะคล้ายๆ กันไว้ด้วยกัน เพื่อที่จะนำไปใช้ในขั้นตอนประเมินราคาที่ดินต่อไป ด้วยสมมติฐานสำคัญที่ว่า แปลงที่ดินที่อยู่ใกล้กัน (spatial characteristics) และ/หรือ มีคุณลักษณะบางประการใกล้เคียงกัน ควรจะมีราคาประเมินเท่ากัน หรือ ไม่ควรแตกต่างกันมาก

การแบ่งกลุ่มได้เลือกใช้อัลกอริทึมหลายประเภทเพื่อเปรียบเทียบหาแบบจำลองที่เหมาะสมที่สุด โดยในทุกๆ อัลกอริทึมจะเลือกใช้แอตทริบิวต์ “LAND_USE” กับ “STREET_WID” ซึ่งเป็นผลลัพธ์ที่ได้มาจากขั้นตอนการหาค้นหาตัวแปรที่มีความสำคัญก่อนหน้า อย่างไรก็ดี ได้นำ “ZONE_NEW” เข้ามาร่วมเป็นตัวกรองด้วย เนื่องจากเป็นส่วนหนึ่งของกฎหมายที่บังคับใช้

<u>อัลกอริทึม K-Means</u>

เป็นการแบ่งกลุ่มออกเป็น K กลุ่มด้วยวิธีค่าเฉลี่ย โดยแปลงที่ดินที่มี “ค่าคุณลักษณะ” ใกล้เคียงกันจะอยู่ในกลุ่มเดียวกัน ในที่นี้ การวัดระยะห่างระหว่างคุณลักษณะของแปลงที่ดินใช้วิธี Euclidean และทำการ constrain ให้แต่ละกลุ่มมีจำนวนสมาชิกอย่างน้อย 1,000 แปลง

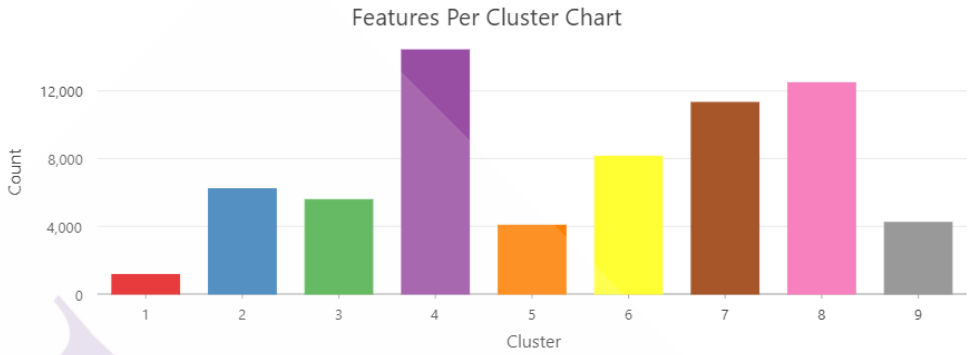


แสดงผลลัพธ์จากการจัดกลุ่มด้วย K-Means

ผลจากการรันอัลกอริทึมพบว่า K-Means ได้แบ่งกลุ่มออกมาเป็น 21 กลุ่ม โดยพบว่า แต่ละกลุ่มมีความกระจัดกระจายไปทั่วทุกบริเวณ ด้วยเหตุผลที่ว่า K-Means ไม่ได้สนใจว่าแปลงที่ดินจะตั้งอยู่ในบริเวณเดียวกันหรือไม่ (ไม่มีคุณสมบัติเชิงพื้นที่) ผลลัพธ์ที่ได้จึงไม่เหมาะสมที่จะนำไปใช้งานต่อ

อัลกอริทึม K-Medoids

เป็นอัลกอริทึมที่ต่อยอดมาจาก K-Means โดยแทนที่จะคำนวณตำแหน่ง centroid ของแต่ละกลุ่มด้วยค่าเฉลี่ย K-Medoids จะเลือกตำแหน่ง centroid ด้วยค่าจริงๆ ที่อยู่ในกลุ่มตัวอย่างแทน จึงมีความเหมาะสมกับข้อมูลประเภทที่เป็นเชิงกลุ่ม (categorical) ในการนำไปใช้งาน ได้ทำการ constrain ให้แต่ละกลุ่มมีจำนวนสมาชิกอย่างน้อย 1,000 แปลง เช่นเดียวกับวิธี K-Means

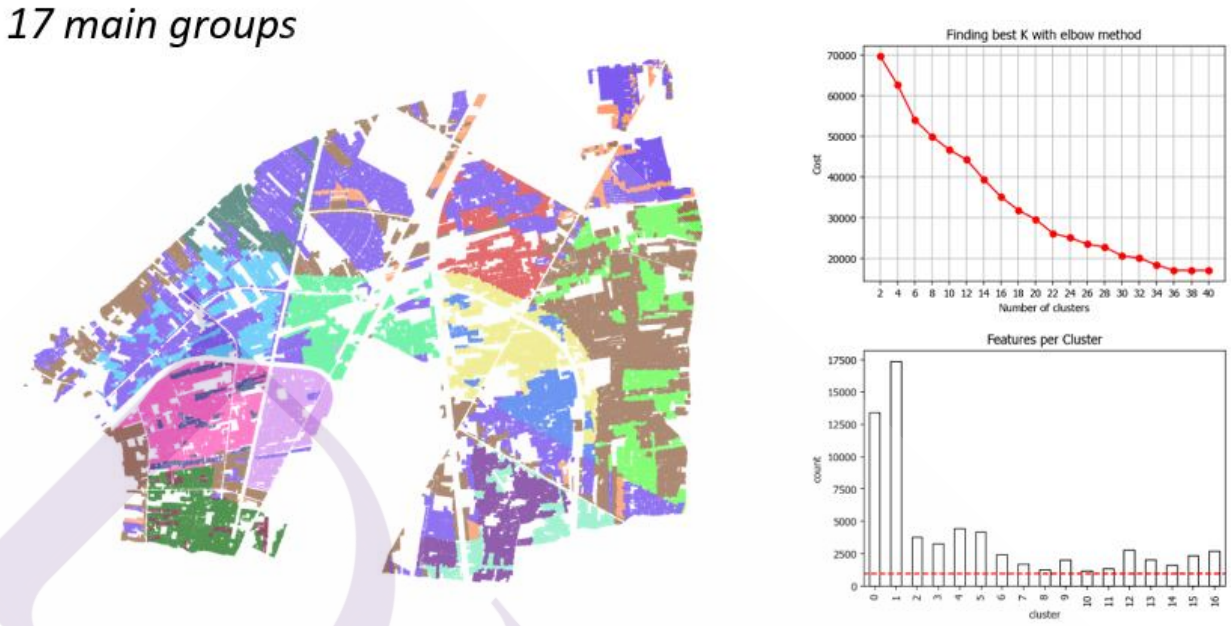


แสดงผลลัพธ์จากการจัดกลุ่มด้วย K-Medoids

ผลจากการรันอัลกอริทึม แบ่งออกได้ 9 กลุ่ม (น้อยกว่ากลุ่มที่ได้จาก K-Means) อย่างไรก็ดี แปลงที่ดินในกลุ่มเดียวกันก็ยังกระจัดกระจาย จึงยังไม่เหมาะสมในการนำไปใช้งานต่อ เนื่องจากยังขาดคุณสมบัติเชิงพื้นที่

อัลกอริทึม K-Prototypes

เป็นอัลกอริทึมที่ผสมผสานวิธี K-Means และวิธี K-Modes เข้าด้วยกัน เหมาะสำหรับชุดข้อมูลที่มีแอตทริบิวต์ทั้งแบบตัวเลข (numerical) และแบบเชิงกลุ่ม (categorical) รวมอยู่ด้วยกัน มีวิธีการทำงานคล้าย K-Means แตกต่างกันตรงวิธีการปรับตำแหน่งของ centroid ในแต่ละกลุ่ม ซึ่งจะถูกแยกคิด โดยถ้าเป็นแอตทริบิวต์เชิงตัวเลขยังคงใช้ค่าเฉลี่ยเหมือนเดิม แต่ถ้าเป็นแอตทริบิวต์เชิงกลุ่มจะเปลี่ยนไปใช้ค่าฐานนิยม (mode) แทน ซึ่งมีความเหมาะสมกับชุดข้อมูลแปลงที่ดินมาก จึงลองนำมาประยุกต์ใช้

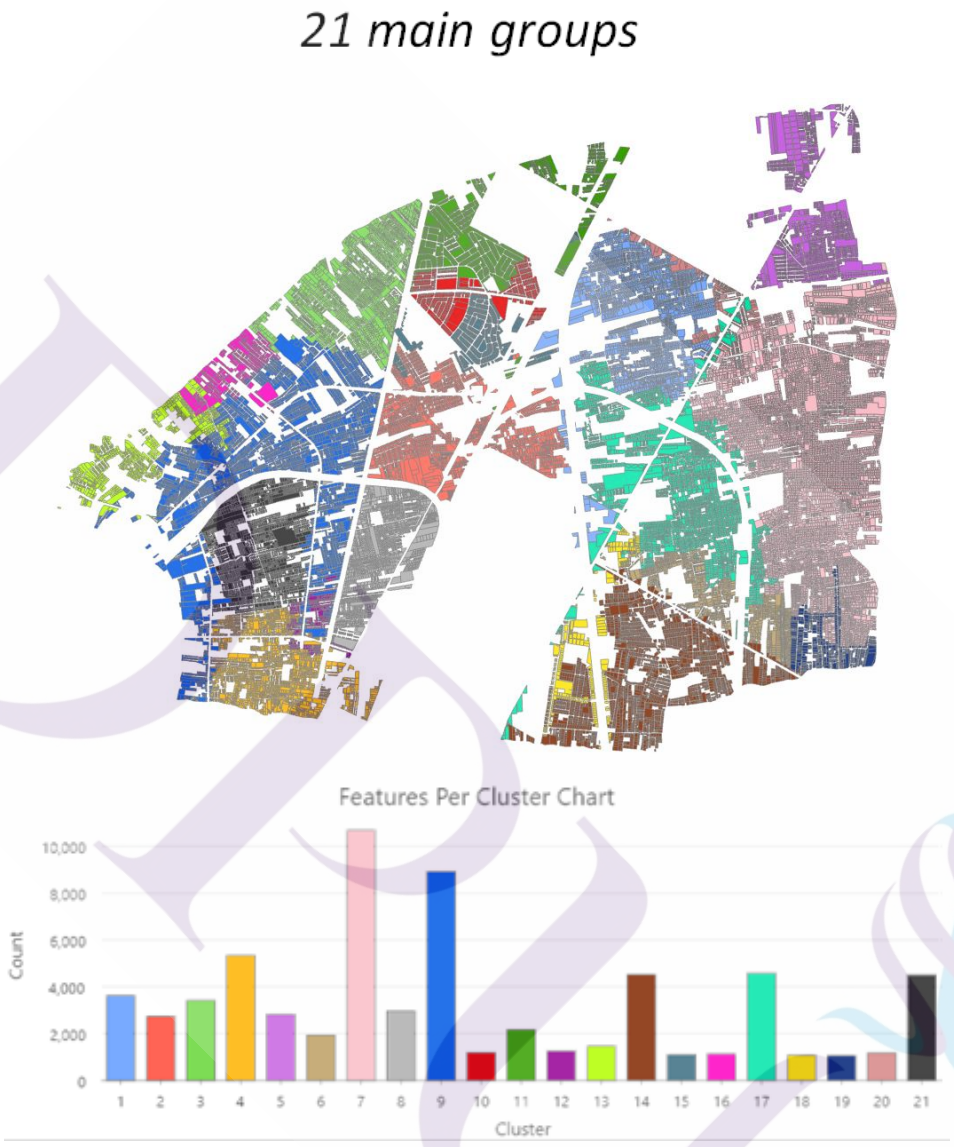


แสดงผลลัพธ์จากการจัดกลุ่มด้วย K-Prototypes

จะเห็นได้ว่าความกระจัดกระจายลดลงเมื่อเทียบกับวิธี K-Means และ K-Medoids แต่ก็ยังไม่หมดไป จึงจำเป็นต้องใช้วิธีสุดท้ายซึ่งเป็นวิธีการที่ยึดหลักการเชิงพื้นที่ (Spatially constrained) เพื่อแก้ปัญหาข้างต้นให้หมดไป

อัลกอริทึม Spatially Constrained Multivariate Clustering

เป็นวิธีการแบ่งกลุ่มที่คำนึงถึงความสัมพันธ์เชิงพื้นที่ระหว่างกัน โดยมีสมมติฐานว่า แปลงที่ดินที่อยู่ในบริเวณเดียวกันหรือใกล้เคียงกันเท่านั้น จึงจะถูกจัดให้อยู่ในกลุ่มเดียวกัน มิเช่นนั้นแล้ว จะถูกแบ่งออกเป็นอีกกลุ่มหนึ่ง การรันอัลกอริทึมได้อาศัยเครื่องมือที่ ArcGIS Pro จัดเตรียมไว้ให้แล้ว



แสดงผลลัพธ์จากการจัดกลุ่มด้วย Spatially Constrained Multivariate Clustering

แปลงที่ดินถูกแบ่งออกเป็น 21 กลุ่ม และผลลัพธ์ออกมาค่อนข้างดี เนื่องจากปัญหาความกระจัดกระจายได้ถูกแก้ไขออกไปจนเกือบหมด

สุดท้าย ต้องแก้ปัญหากรณี STREET_COD หรือหน่วยถนนเดียวกันแต่อยู่ในหลายกลุ่มคลัสเตอร์ จึงจำเป็นต้องเลือกกลุ่มคลัสเตอร์ใหม่ โดยอาศัยวิธีฐานนิยม หมายถึง กลุ่มใดมีจำนวนแปลงที่ดินที่นับจำนวน STREET_COD หรือหน่วยถนนเดียวกันได้มากกว่า ให้เปลี่ยนเป็นกลุ่มคลัสเตอร์นั้นทั้งหมด

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายเจตนา สมคำนึง
ประวัติการศึกษา	บัญชีบัณฑิต (บช.บ.)
	สาขาวิชาการบัญชี
	มหาวิทยาลัยธรรมศาสตร์
	ปีการศึกษา 2541
ประวัติการทำงาน	บริษัท ไทยเอ็มเอฟซี จำกัด
	บริษัท เอสซีจีเคมิคอลส์ จำกัด
	บริษัท ระยองโอเลฟินส์ จำกัด (สำนักงานสิงคโปร์)