DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

การปรับปรุงประสิทธิภาพของการประเมินราคาบ้านพักอาศัย โดยใช้ข้อมูลเชิงพื้นที่และรูปภาพ

วรภร เล็กอาราม

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิศวกรรมศาตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2565

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ENHANCING HOUSE PRICE ESTIMATION USING SPATIAL INFORMATION AND IMAGES

WORAPORN LAKARAM

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
Academic Year 2022

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ใบรับรองสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่

หัวข้อสารนิพนธ์ การปรับปรุงประสิทธิภาพของการประเมินราคาบ้านพักอาศัยโดยใช้ข้อมูลเชิงพื้นที่และรูปภาพ

เสนอโดย วรภร เล็กอาราม

สาขาวิชา วิศวกรรมข้อมูลขนาดใหญ่

อาจารย์ที่ปรึกษาสารนิพนธ์ ดร.ธนภัทร ฆังคะจิตร

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

---------- ประธานกรรมการ
(ดร.สรรพฤทธิ์ มฤคทัต)

---------- กรรมการที่ปรึกษาสารนิพนธ์
(ดร.ธนภัทร ฆังคะจิตร)

---------- กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ รับรองแล้ว

---------- คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
(ดร.ชัยพร เขมะภาคะพันธ์)

วันที่ 31 เดือน กรกฎาคม พ.ศ. 2566

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

หัวข้อสารนิพนธ์	การปรับปรุงประสิทธิภาพของการประเมินราคาบ้านพักอาศัย โดยใช้ข้อมูลเชิงพื้นที่และรูปภาพ
ชื่อผู้เขียน	วรภร เล็กอาราม
อาจารย์ที่ปรึกษา	ดร.ธนภัทร ฆังคะจิตร
หลักสูตร	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2565

บทคัดย่อ

การประเมินราคาบ้านพักอาศัยมีส่วนสำคัญในการพัฒนาอสังหาริมทรัพย์ โดยเป็นเครื่องมือที่ช่วยในการตัดสินใจในการซื้อขายทั้งผู้พัฒนาอสังหาริมทรัพย์และผู้ซื้อ โดยทั่วไปการประเมินราคามีค่าใช้จ่ายที่สูงและมีขั้นตอนการดำเนินการที่ใช้เวลา เนื่องจากผู้ประเมินต้องลงพื้นที่เพื่อเก็บข้อมูล ดังนั้นหลายงานวิจัยมีการนำเสนอการประยุกต์ใช้เทคนิคจักรกลเรียนรู้ในการประเมินราคาบ้านพักอาศัยที่ให้ความแม่นยำค่อนข้างสูง เป็นที่รู้กันว่าข้อมูลที่เกี่ยวข้องกับบ้าน และข้อมูลเชิงพื้นที่ซึ่งบ่งบอกถึงทำเลที่ตั้งของบ้าน เป็นปัจจัยสำคัญในการประเมินราคาบ้าน เช่นเดียวกันกับสภาพบ้าน ดังนั้นงานวิจัยนี้จึงนำเสนอการเพิ่มประสิทธิภาพการประเมินราคาบ้านพักอาศัยโดยใช้ข้อมูล 1) สภาพของบ้าน 2) ข้อมูลเชิงพื้นที่ และ3) ข้อมูลที่เกี่ยวข้องกับบ้านร่วมกันในการประเมินราคา โดยงานวิจัยนี้นำเสนอวิธีการสกัดตัวแทนสภาพบ้านจากรูปภาพบ้านด้านหน้าโดยใช้เทคนิคการเรียนรู้เชิงลึก เริ่มจากรูปภาพบ้านด้านหน้าจะถูกคัดเลือกมาใช้งานด้วย AutoEncoder หลังจากนั้นจะนำรูปภาพด้านหน้าไปสร้างแบบจำลองโดยใช้อัลกอริทึม CNN-based model (ResNet50) เพื่อจำแนกประเภทราคาบ้านเป็นราคาต่ำหรือราคาสูง แล้วจึงนำแบบจำลองที่ได้ตัดส่วนปลายทิ้งไปเพื่อใช้เป็นตัวแทนของปัจจัยรูปภาพบ้านด้านหน้า แล้วจึงนำมาใช้เป็นปัจจัยร่วมกับข้อมูลบ้านและข้อมูลเชิงพื้นที่ในการทำนายราคาประเมินบ้านโดยใช้แบบจำลอง XGBoost Regression

โดยข้อมูลที่นำมาใช้เป็นข้อมูลบ้านพักอาศัยในเขตปทุมธานีในช่วงเดือนพฤษภาคมถึงเดือนมิถุนายน ปี พ.ศ. 2566 ที่ประกาศขายใน www.baania.com ผลการศึกษาพบว่าแบบจำลองที่นำเสนอให้ประสิทธิภาพในการประเมินที่แม่นยำมากกว่าแบบจำลองที่ใช้เพียงข้อมูลบ้านเพียงอย่างเดียว อย่างไรก็ตามปัจจัยสภาพบ้านไม่ได้เพิ่มประสิทธิภาพในการอย่างมีนัยสำคัญ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ดังนั้นรูปภาพบ้านด้านหน้าอาจจะยังไม่เพียงพอที่จะเป็นตัวแทนแสดงถึงสภาพของบ้าน ดังนั้น ควรมีการเพิ่มรูปภาพห้องต่างๆในบ้านซึ่งจะช่วยบอกถึงสภาพของบ้านได้อย่างละเอียดและทำให้การประเมินราคามีประสิทธิภาพยิ่งขึ้น

คำสำคัญ: การประเมินราคาบ้าน, รูปภาพบ้านด้านหน้า, สภาพของบ้าน, การเรียนรู้เชิงลึก, เทคนิคจักรกลเรียนรู้, ข้อมูลเชิงพื้นที่

อาจารย์ที่ปรึกษา

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Thematic Paper Title	ENHANCING HOUSE PRICE ESTIMATION USING SPATIAL INFORMATION AND IMAGES
Author	WORAPORN LAKARAM
Thematic Paper Advisor	Dr.Thanapat Kangkachit
Program	Big Data Engineering
Academic Year	2023

ABSTRACT

House Price Estimation plays a crucial role in real estate development as a decision-making tool for both developers and customers. Generally, it is an expensive and time-consuming task since the appraisers need the collected data from the field. Therefore, several machine learning techniques are utilized and produced relatively high predictive performance. Indeed, house information and spatial information are the essential features to estimate house price as well as the house's condition. In this work, we propose an approach to enhance house price estimation by incorporating the house condition, spatial information, and house information. To extract the house's condition, we take the house front images into account. Initially, only the house-front pictures are selected using AutoEncoder. Then, the CNN-based model (ResNet50) is trained on those house front images to classify them as high or low price. To represent house condition features, the last layer of the trained CNN model is removed. Therefore, the final input features are generated from those features combined with the house information and spatial information.

Finally, the regression model (XGBoostRegression) is built to predict house prices. The experimental data are the houses for sale in Pathum Thani province during May and June 2023 collected from www.baania.com. The experimental results showed that the predictive performance of our model outperforms the based-line model using only house information. However, the house condition features seem not to increase predictive performance immensely. Hence the house front pictures are insufficient to represent the house's condition.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Thus, the images of different rooms in the house are further determined to obtain a better predictive performance of house price estimation.

Keywords: House Price Estimation, Spatial Information, Computer vision

Advisor

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จลุล่วงอย่างสมบูรณ์ โดยผู้วิจัยขอกราบขอบพระคุณในความกรุณาของอาจารย์ที่ปรึกษาสารนิพนธ์ ดร.ธนภัทร ฆังคะจิตร เป็นอย่างสูงที่เสียสละเวลาอันมีค่าเพื่อให้คำปรึกษาและคำแนะนำในการทำวิจัยตลอดระยะเวลาที่ผ่านมา อาจารย์ได้ให้ข้อเสนอแนะ ความคิดเห็นและทรัพยากรที่มีประโยชน์ ต่องานวิจัยชิ้นนี้ รวมถึงเอาใจใส่ผู้วิจัยเป็นอย่างดี

ขอขอบพระคุณ ดร.สรรพฤทธิ์ มฤคทัต และ ผศ.ดร.ดวงใจ จิตตคงชื่น ที่สละเวลาเพื่อมาเป็นกรรมการในการสอบสารนิพนธ์ และได้ให้คำแนะนำแนวทางเป็นประโยชน์ต่องานวิจัย

หากมีสิ่งได้ที่ผู้วิจัยได้ทำผิดพลาดหรือบกพร่องประการใด ผู้วิจัยต้องกราบขออภัยเป็นอย่างสูงมา ณ โอกาสนี้ ผู้วิจัยหวังเป็นอย่างยิ่งว่าสารนิพนธ์ฉบับนี้ จะเป็นพื้นฐานในการต่อยอดองค์ความรู้ของผู้ที่สนใจศึกษาในงานด้านนี้ต่อไป

วรภร เล็กอาราม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... ฉ

กิตติกรรมประกาศ..... ซ

สารบัญ..... ฌ

สารบัญตาราง..... ฎ

สารบัญภาพ..... ฏ

บทที่

1. บทนำ..... 1
   - 1.1 ที่มาและความสำคัญของปัญหา..... 1
   - 1.2 วัตถุประสงค์การศึกษาหรือวิจัย..... 1
   - 1.3 ขอบเขตของงานวิจัย..... 1
   - 1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 2
2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง..... 3
   - 2.1 แนวคิด..... 3
   - 2.2 ทฤษฎี..... 4
   - 2.3 งานวิจัยที่เกี่ยวข้อง..... 6
3. ระเบียบวิธีวิจัย..... 8
   - 3.1 การเตรียมข้อมูล..... 9
   - 3.2 การสร้างแบบจำลองเพื่อขัดแยกรูปบ้าน..... 11
   - 3.3 การสร้างแบบจำลองเพื่อสกัดตัวแปรจากรูปบ้าน..... 13
   - 3.4 การสร้างแบบจำลองประเมินราคาบ้าน..... 14
   - 3.5 เครื่องมือที่ใช้ในการศึกษา..... 15
4. ผลการวิจัย..... 16
   - 4.1 ผลการวัดประสิทธิภาพของแบบจำลองในการทำนายช่วงราคาบ้านจากรูปภาพ..... 14
   - 4.2 ผลการวัดประสิทธิภาพของแบบจำลองของการประเมินราคา..... 20
5. สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 22
   - 5.1 สรุปผลการศึกษา..... 22
   - 5.2 ข้อเสนอแนะ..... 22

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ (ต่อ)

หน้า

บรรณานุกรม..... 23
ประวัติผู้เขียน..... 25

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญตาราง

ตารางที่ **หน้า**

3.1 ตารางแสดงตัวอย่างข้อมูลที่ได้จากเว็บไซต์..... 9
4.1 แสดงประสิทธิภาพทำนายช่วงราคาบ้านจากรูปภาพ..... 19
4.2 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองของการประเมินราคา..... 20

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ

ภาพที่ **หน้า**

2.1 ตัวอย่างแสดงปัจจัยของทำเลมีผลต่อราคาที่ดิน..... 3

3.1 ภาพสรุปแนวคิดวิธีวิจัย..... 8

3.2 ตำแหน่งบ้าน..... 10

3.3 ตำแหน่ง POIs..... 11

3.4 โครงสร้างแบบจำลอง AutoEncoder..... 12

3.5 ตัวอย่างภาพที่ทำการแยกออก..... 12

3.6 ตัวอย่างภาพด้านหน้าบ้านที่ทำไปใช้ในการทำนายราคาบ้าน..... 13

3.7 โครงสร้างส่วนปลายของแบบจำลอง ResNet50 ที่ทำการปรับปรุง..... 14

4.1 กราฟแสดงผลการทดลองแบ่ง 3 ช่วงโดยการแบ่งที่ 3 และ 7 ล้านบาท..... 16

4.2 กราฟแสดงผลการทดลองแบ่ง 3 ช่วงโดยการแบ่งที่ 3 และ 7 ล้านบาท และทำการ under-sampling..... 17

4.3 ฮิสโตแกรมของราคาบ้าน..... 17

4.4 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 4 ล้านบาท..... 17

4.5 กราฟการกระจายตัวของราคาในส่วนของรูปที่มีราคาสูงกว่า 4 ล้านแต่แบบจำลองทำนายต่ำกว่า 4 ล้านบาท (ซ้าย) และ confusion matrix ผลการทำนายด้วยแบบจำลองที่แบ่งที่ราคา 4 ล้านบาท (ขวา)..... 18

4.6 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 5 ล้านบาท..... 18

4.7 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 5 ล้านบาท และใช้ loss เป็น BinaryFocalCrossentropy..... 19

4.8 แสดงตัวแปรที่สำคัญ (Feature Importance)..... 18

4.9 กราฟแสดงราคาจริง และราคาที่ทำนายโดยแบบจำลอง..... 18

4.10 แสดง % error ของราคาที่ทำนายได้..... 19

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การทำนายราคาบ้านมีส่วนช่วยในการตัดสินใจซื้อขายทั้งกับผู้พัฒนาอสังหาริมทรัพย์และผู้ซื้อ ผู้พัฒนาอสังหาริมทรัพย์และผู้ซื้อสามารถเปรียบเทียบได้ว่าเป็นราคาที่เหมาะสมหรือไม่ในการซื้อขาย อีกทั้งราคาอสังหาริมทรัพย์ยังแสดงถึงสภาพเศรษฐกิจของประเทศได้อย่างหนึ่ง

โดยทั่วไปการประเมินราคาบ้านต้องใช้ผู้เชี่ยวชาญในการลงพื้นที่ เพื่อไปเก็บข้อมูลและประเมินราคา มีขั้นตอนและค่าใช้จ่ายในการดำเนินงาน เวลาที่ใช้ดำเนินการค่อนข้างนาน ทั้งนี้เพื่อลดเวลา และค่าใช้จ่ายในการดำเนินงานโดยผู้เชี่ยวชาญ เพิ่มความรวดเร็วในการตัดสินใจในการซื้อขายอสังหาริมทรัพย์ ทำให้การพัฒนาแบบจำลองเพื่อทำนายราคาบ้านเป็นที่น่าสนใจ

งานวิจัยนี้มีวัตถุประสงค์เพื่อปรับปรุงประสิทธิภาพของแบบจำลองที่ใช้ในการประเมินราคาบ้านมือสอง โดยใช้ข้อมูลเชิงพื้นที่ และข้อมูลรูปภาพบ้าน มาเพื่อเพิ่มประสิทธิภาพในการประเมินราคา

ผลการทดลองบนข้อมูลบ้านมือสองในจังหวัดปทุมธานี พบว่า การใช้ข้อมูลเชิงพื้นที่ และรูปภาพด้านหน้าบ้าน สามารถเพิ่มประสิทธิภาพการประเมินราคาบ้านได้

1.2 วัตถุประสงค์การศึกษาหรือวิจัย

1. เพื่อเพิ่มประสิทธิภาพในการประเมินราคาบ้านมือสอง โดยใช้ปัจจัย
   1) ข้อมูลบ้าน
   2) ข้อมูลเชิงพื้นที่
   3) รูปภาพบ้าน

1.3 ขอบเขตของงานวิจัย

ใช้ปัจจัยที่มีผลต่อราคาบ้าน ได้แก่ ข้อมูลบ้าน ข้อมูลเชิงพื้นที่ และรูปภาพบ้าน โดยข้อมูลที่นำมาใช้มีดังต่อไปนี้

1.3.1 ข้อมูลบ้านมือสองที่มีการลงประกาศขาย ในเดือนพฤษภาคมถึงมิถุนายน ปี พ.ศ. 2566 บนเว็บไซต์ประกาศขายบ้าน https://www.baania.com จากการทำ web scrapping ข้อมูลในจังหวัดปทุมธานี ซึ่งเป็นจังหวัดหนึ่งในเขตปริมณฑล ประเทศไทย ในช่วงราคาไม่เกิน 10 ล้านบาท ข้อมูลที่ใช้ ได้แก่ จำนวนห้องนอน จำนวนห้องน้ำ พื้นที่ รูปภาพด้านข้างหน้าตัวบ้าน ตำแหน่งละติจูด ลองจิจูด บน google map

1.3.2 ข้อมูล POIs ในจังหวัดปทุมธานี ได้แก่ โรงเรียน, สถานีบริการพลังงาน, ร้านสะดวกซื้อ, ไฮเปอร์มาร์เก็ต, ธนาคาร, บริการสาธารณสุข, มหาวิทยาลัย, ร้านกาแฟ, ร้านขายยา, ศูนย์การค้า, สถานีรถไฟฟ้า BTS

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เพื่อลดขั้นตอน และเวลาในการประเมินราคาบ้านโดยผู้เชี่ยวชาญ

1.4.2 เพื่อเป็นตัวช่วยในการตัดสินใจ ในการซื้อขายบ้านได้อย่างรวดเร็วกว่ายิ่งขึ้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 2
แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิด

ปัจจัยที่มีผลต่อราคาบ้าน นอกเหนือจากข้อมูลบ้าน ได้แก่ จำนวนห้องนอน จำนวนห้องน้ำ พื้นที่ดิน ปัจจัยเชิงพื้นที่ เช่น ทำเลที่ตั้ง และสิ่งอำนวยความสะดวก ในบริเวณพื้นที่นั้น มีผลต่อราคาบ้าน ที่ดินที่ใกล้ ห้างสรรพสินค้า รถไฟฟ้า จะมีราคาสูงกว่าทำเลที่ไกลออกไป และนอกจากทำเลที่ตั้ง สภาพบ้านและลักษณะบ้านก็เป็นปัจจัยที่ส่งผลต่อราคา ข้อมูลเชิงพื้นที่ และรูปภาพบ้าน ถูกพิจารณาว่าเป็นปัจจัยที่ส่งผลต่อราคาบ้าน การใช้ข้อมูลเชิงพื้นที่ และรูปภาพจะสามารถเพิ่มประสิทธิภาพการประเมินราคาบ้านได้

ภาพที่ 2.1 ตัวอย่างภาพแสดงปัจจัยของทำเลส่งผลต่อราคาที่ดิน

ที่มา: https://www.bangkokassets.com/articledetail/623/อัพเดต_10_ทำเล_ราคาที่ดินแพงที่สุดในกรุงเทพฯ-ปริมณฑล_2565

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2.2 ทฤษฎี

เพื่อศึกษาการปรับปรุงประสิทธิภาพการประเมินราคาบ้านพักอาศัย โดยใช้ข้อมูลเชิงพื้นที่ และรูปภาพ มีทฤษฎีที่เกี่ยวข้องดังต่อไปนี้

2.2.1 Spatial Information ข้อมูลเชิงพื้นที่

Spatial information (ข้อมูลเชิงพื้นที่) เป็นข้อมูลที่เกี่ยวข้องกับตำแหน่งทางภูมิศาสตร์ของวัตถุหรือเหตุการณ์ในโลก ซึ่งสามารถอธิบายลักษณะเชิงพื้นที่ของวัตถุได้ เช่น พิกัดทางภูมิศาสตร์ (geographic coordinates) เช่น ละติจูดและลองจิจูด รูปร่างและขอบเขตทางภูมิศาสตร์ และความสัมพันธ์กับวัตถุหรือเหตุการณ์อื่นๆ ในพื้นที่เดียวกันหรือบริเวณใกล้เคียง

Geopandas เป็นไลบรารี (library) ที่ใช้สำหรับการจัดการและการวิเคราะห์ข้อมูลเชิงพื้นที่ในภาษา Python โดย Geopandas เป็นการรวมความสามารถของไลบรารี Pandas (ที่ใช้สำหรับการจัดการข้อมูลตาราง) และไลบรารี Geospatial (ที่ใช้สำหรับการจัดการข้อมูลเชิงพื้นที่) เพื่อให้สามารถจัดการข้อมูลเชิงพื้นที่ในรูปแบบตารางได้

Geopandas สามารถทำงานกับข้อมูลเชิงพื้นที่ที่มีรูปแบบของเส้นทาง (lines) หรือรูปร่าง (shapes) ได้ เช่น การอ่านและเขียนไฟล์ที่มีข้อมูลเชิงพื้นที่ เช่น ไฟล์รูปร่าง (shapefile) และไฟล์ GeoJSON นอกจากนี้ยังมีฟังก์ชันที่สามารถดำเนินการเกี่ยวกับข้อมูลเชิงพื้นที่ได้ เช่น การคำนวณระยะทาง การคำนวณพื้นที่ การเชื่อมต่อข้อมูลตาราง และการทำแผนที่

ด้วยความสามารถของ Geopandas ที่รวมการจัดการข้อมูลตารางและข้อมูลเชิงพื้นที่ไว้ด้วยกัน นักวิเคราะห์ข้อมูลสามารถใช้งานเพื่อวิเคราะห์และแสดงข้อมูลเชิงพื้นที่ในรูปแบบตารางและแผนที่ได้อย่างมีประสิทธิภาพ

Point of Interests (POIs) (จุดสนใจ) เป็นตำแหน่งทางภูมิศาสตร์ที่น่าสนใจหรือมีความสำคัญในเชิงท่องเที่ยว ธุรกิจ หรือการใช้ชีวิตประจำวัน ตัวอย่างของ POIs ได้แก่ ร้านอาหาร โรงแรม สถานีบริการน้ำมัน โรงเรียน สวนสาธารณะ สถานที่ท่องเที่ยว ห้างสรรพสินค้า ศาลากลางเมือง เป็นต้น ข้อมูล POIs ช่วยให้ผู้ใช้งานสามารถค้นหาและนำทางไปยังสถานที่ที่สนใจได้อย่างสะดวกและรวดเร็ว

2.2.2 AutoEncoder

AutoEncoder [1] เป็นโครงข่ายประสาทเทียมที่ออกแบบขึ้นเพื่อเรียนรู้การเข้ารหัส (encoding) และการถอดรหัส (decoding) ข้อมูล เพื่อสร้างภาพรวม (representation) ที่มีความสามารถในการรักษาข้อมูลต้นฉบับโดยใช้มาตรการการบีบอัด (compression) และการนำความหมาย (meaning) ออกมาจากข้อมูลโดยมีลักษณะโครงสร้างหลักประกอบไปด้วยส่วนของ Encoder, Latent layer และส่วนของ Decoder

Latent layer จะเป็นส่วนแก่นของข้อมูลที่ถูกย่อมาจาก Encoder ซึ่งประกอบไปด้วย node เพียงชั้นเดียว และส่วนของ Decoder คือส่วนขยายใจความของข้อมูล ถ้าในส่วนของ latent สามารถสกัดแก่นของข้อมูลได้ ข้อมูลหลังจากขยายจะไม่แตกต่างจากข้อมูลขาเข้ามาก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

AutoEncoder มักถูกนำมาใช้กับการตรวจจับข้อมูลแปลกปลอม เนื่องด้วยถ้าเป็นข้อมูลแปลกปลอมหลังจาก Decoder จะได้ข้อมูลแตกต่างไปจากเดิมมาก

2.2.3 ResNet50

ResNet50 [2] ปีนแบบจำลองแบบจำลองสำหรับการเรียนรู้ของโครงข่ายประสาทเชิงลึก (deep neural network) ที่ถูกพัฒนาขึ้นโดยทีมวิจัยจาก Microsoft Research ในปี 2015 โดยมีชื่อเต็มว่า "Residual Network 50 layers" (ResNet50) แบบจำลองนี้เป็นส่วนหนึ่งของครอบครัวของแบบจำลอง ResNet ที่ได้รับความนิยมและนำมาใช้ในงานประมวลผลภาพหลากหลายเช่นการจดจำใบหน้า (facial recognition) การตรวจจับวัตถุ (object detection) และการจำแนกภาพ (image classification) ซึ่งมีการนำไปประยุกต์ใช้ในหลายๆ งานที่เกี่ยวข้องกับการประมวลผลภาพ

โครงสร้างของ ResNet50 มีจำนวนเลเยอร์ทั้งหมด 50 เลเยอร์ ซึ่งประกอบไปด้วยบล็อกหลักสี่ส่วนหลักคือ convolutional layer, residual block, identity block และ fully connected layer. ในส่วนของ convolutional layer เป็นการทำคอนโวลูชันเพื่อสกัดลักษณะของภาพ ส่วน residual block ใช้เพื่อลดปัญหาการซึมผ่านและการแพร่กระจายของข้อมูลในเครือข่าย เมื่อเรามีแบบจำลองที่ลึกขึ้น อาจเกิดปัญหาในการฝึกสอนและการเรียนรู้ได้ยากขึ้น แต่ ResNet ใช้การผ่านข้ามที่ช่วยในการส่งผ่านข้อมูลผ่านเครือข่ายได้อย่างมีประสิทธิภาพ เมื่อผ่านเลเยอร์นั้นแล้ว ข้อมูลจะถูกรวมกันด้วยการบวกเพื่อให้ข้อมูลที่เข้ามาบวกกับข้อมูลก่อนหน้านั้น โดยไม่สูญเสียข้อมูลหรือลดความแม่นยำ

ResNet50 มีความสามารถในการจดจำและรู้จำวัตถุภายในภาพได้อย่างแม่นยำ โดยมีการฝึกสอนด้วยชุดข้อมูลที่มีมากถึงหลายล้านภาพ เช่น ImageNet ที่มีการจำแนกภาพออกเป็นกลุ่มหมวดหมู่ต่างๆ โดย ResNet50 มีความสามารถในการจำแนกภาพออกเป็นหมวดหมู่ได้ถึง 1,000 หมวดหมู่

โดยทั่วไปแล้ว ResNet50 เป็นแบบจำลองที่มีประสิทธิภาพและนิยมใช้ในงานประมวลผลภาพที่ต้องการความละเอียดสูงและความแม่นยำในการจำแนกวัตถุภาพ

2.2.4 XGboost

XGBoost (Extreme Gradient Boosting) [3] เป็นตัวจัดการเรียนรู้แบบมีผู้สอน (supervised learning) ที่เป็นแบบอัลกอริทึม (algorithm) สำหรับการจัดการกับปัญหาการจำแนก (classification) และการทำนาย (regression) ในข้อมูลที่มีลักษณะตาราง (tabular data) โดย XGBoost ใช้เทคนิคต่อเติมต้นแบบ (boosting) เพื่อปรับปรุงแบบจำลองให้มีความแม่นยำสูงขึ้น

โดยอัลกอริทึม XGBoost จะใช้การทำงานของต้นแบบเชิงพื้นที่ (tree-based model) เช่น ต้นไม้เชิงทำนาย (decision trees) เพื่อสร้างแบบจำลองที่มีความแม่นยำและความซับซ้อนต่ำพร้อมกับการป้องกันการเกิดการเรียนรู้ที่เกินไป (overfitting) โดยใช้เทคนิคการลดค่าเชิงบวก (regularization) เพื่อควบคุมความซับซ้อนของแบบจำลอง

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

XGBoost มีคุณสมบัติที่น่าสนใจเช่น การจัดการกับความหนาแน่นของข้อมูลได้ดี การทำงานแบบการกระจาย (distributed computing) เพื่อประมวลผลข้อมูลได้รวดเร็ว และการตรวจสอบคุณลักษณะที่สำคัญในการทำนาย (feature importance) เพื่อวิเคราะห์และตรวจสอบคุณลักษณะที่มีผลกระทบสูงสุดในการตัดสินใจของแบบจำลอง

2.3 งานวิจัยที่เกี่ยวข้อง

G. Srirutchataboon et al. “Stacking Ensemble Learning for Housing Price Prediction: a Case Study in Thailand” (2021) [4] นำเสนอการใช้ CNN ทำการสกัด features จากรูปภาพบ้าน โดยทำการเทรนแบบจำลอง เพื่อทำการจำแนกหรือแบ่งกลุ่มข้อมูลช่วงราคาบ้านจากรูปภาพ โดยแบ่งออกเป็น 3 ช่วง (1) ราคาต่ำกว่า 3 ล้านบาท (2) ราคาในช่วง 3-7 ล้านบาท (3) ราคาตั้งแต่ 7 ล้านบาทขึ้นไป โดยใช้ VGG19 ที่มีการตัดส่วนสุดท้ายออกและเพิ่มชั้นโครงข่ายประสาทเทียม 5 จุด และชั้นสำหรับแบ่งกลุ่ม 3 กลุ่ม และนำตัวแปรที่ได้จากรูปภาพ ซึ่งได้จากชั้นโครงข่ายประสาทเทียม 5 จุด ไปใช้ร่วมกับข้อมูลบ้านเพื่อสร้างแบบจำลองทำนายราคาด้วย XGBoost ซึ่งผลการทดสอบให้ MAPE 17.83% บนข้อมูลบ้านในเขตจังหวัดกรุงเทพมหานครและปริมณฑล

P. -Y. Wang et al. “Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism” (2021) [5] นำเสนอการใช้ข้อมูลสิ่งอำนวยความสะดวกส่วนกลางสาธารณะ (Public Facilities Data) ภาพถ่ายดาวเทียม (Google Satellite Image) และข้อมูลบ้าน (House Transaction Data) ในการใช่ร่วมกันทำนายราคาบ้าน ซึ่งสามารถปรับปรุงประสิทธิภาพการทำนายราคาบ้านได้

ในส่วนของข้อมูลสิ่งอำนวยความสะดวกสาธารณะที่ถูกนำมาใช้ ได้แก่ ร้านค้า (Stores) สวนสาธาณะ (Park) รถไฟฟ้าขนส่งมวชน (MRT) โรงพยาบาล (Hospital) โดยมีการคำนวณหาระยะทางต่างที่กำหนด ระหว่างบ้านและสิ่งอำนวยความสะดวกเหล่านั้น

A. Nouriani and L. Lemke “Vision-based Housing Price Estimation Using Interior, Exterior & Satellite Images” (2022) [6] นำเสนอการแบ่ง 5 ระดับความหรูหรา ของภาพห้องต่างๆภายในบ้าน หน้าบ้านและภาพถ่ายดาวเทียม ด้วย VGG16 ที่มีการปรับปรุงส่วนปลาย และนำระดับความหรูหรามาใช้เป็น feature ที่ใช่ร่วมทำนายกับข้อมูลบ้าน ด้วยแบบจำลอง XGBoost ซึ่งผลลัพธ์ที่ได้มีค่า MSE น้อยที่สุด

แนวคิดในงานวิจัยนี้ ไม่ได้ถูกนำมาประยุกต์ใช้ในสารนิพนธ์นี้ เนื่องจากข้อมูลที่ได้จากการทำ web scrapping ไม่มีข้อมูลรูปภาพห้องต่างๆ ภายในบ้าน และในงานวิจัยนี้มีการให้อาสาสมัครทำการประเมินระดับความหรูหราของบ้าน เพื่อใช้เป็นข้อมูลในการตรวจสอบ (Validation) ความแม่นยำของแบบจำลอง

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

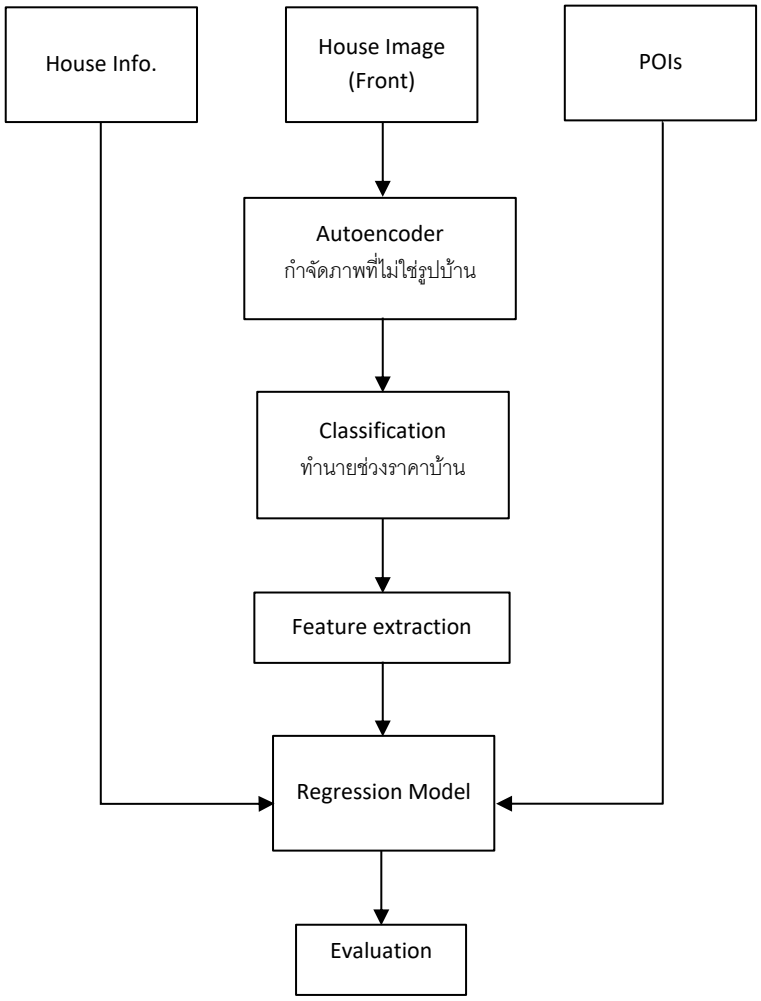
J. K. Chow et al. “Anomaly detection of defects on concrete structures with the convolutional autoencoder” (2020) [7] นำเสนอการใช้ Convolution AutoEncoder ในการตรวจจับการแตกร้าวของโครงสร้างคอนกรีต โดยใช้การเทรนแบบจำลองด้วยภาพคอนกรีตที่มีสภาพปกติและไม่แตก เมื่อนำแบบจำลองดังกล่าวไปทดสอบกับภาพคอนกรีตที่มีรอยแตกร้าวหรือไม่ปกติจะมีค่า ภาพที่ได้จะมีค่า reconstruction error ที่สูงทำให้สามารถตรวจจับภาพแตกร้าวได้ เป็นการเรียนรู้โดยที่ไม่ต้องมีการ label ภาพคอนกรีตที่แตกให้กับตัวข้อมูล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 3
ระเบียบวิธีวิจัย

งานวิจัยนี้นำเสนอการปรับปรุงประสิทธิภาพในการประเมินราคาบ้านพักอาศัยโดยใช้ข้อมูลเชิงพื้นที่ และรูปภาพ โดยทำการศึกษาปัจจัย 3 ส่วน ได้แก่ (1) ตัวแปรข้อมูลบ้าน ได้แก่ จำนวนห้องนอน จำนวนห้องน้ำ พื้นที่ดิน (2) ตัวแปรเชิงพื้นที่ ได่แก่ ละติจูด ลองจิจูด ของทรัพย์สิน จำนวน POIs ในระยะ 1 กิโลเมตร 3 กิโลเมตร 5 กิโลเมตร และระยะทางใกล้สุดจากบ้านไปยัง POIs (3) รูปภาพด้านหน้าบ้าน

ภาพสรุปแนวคิดวิธีวิจัย เป็นดังภาพต่อไปนี้



ภาพที่ 3.1 ภาพสรุปแนวคิดวิธีวิจัย

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.1 การเตรียมข้อมูล

ขั้นตอนในการเตรียมข้อมูลบ้าน และ ข้อมูลเชิงพื้นที่ มีดังต่อไปนี้

3.1.1 รวบรวมข้อมูลโดยการทำ web scrapping บนเว็บไซต์ประกาศขายบ้านมือสอง https://www.baania.com ข้อมูลที่ทำการรวบรวม เป็นข้อมูลบ้านในจังหวัดปทุมธานี ที่ประกาศขาย ในเดือนพฤษภาคมถึงมิถุนายน ปี พ.ศ. 2566 ซึ่งข้อมูลที่ได้มีดังต่อไปนี้

1) ชื่อคำประกาศขาย (name)

2) ราคา (price)

3) พื้นที่ (Space)

4) พื้นที่ใช้สอย (Area_usable)

5) จำนวนห้องนอน (Bed)

6) จำนวนห้องน้ำ (Bath)

7) จำนวนชั้น (Stair)

8) จำนวนที่จอดรถ (Parking)

9) จำนวนห้องครัว (Kitchen)

10) จำนวนห้องนั่งเล่น (Living)

11) จำนวนห้องแม่บ้าน (Maid)

12) url ของรูปภาพด้านข้างหน้าตัวบ้าน

13) url ของตำแหน่งละติจูต ลองจิจูด บน google map

ตารางที่ 3.1 ตารางแสดงตัวอย่างข้อมูลที่ได้จากเว็บไซต์

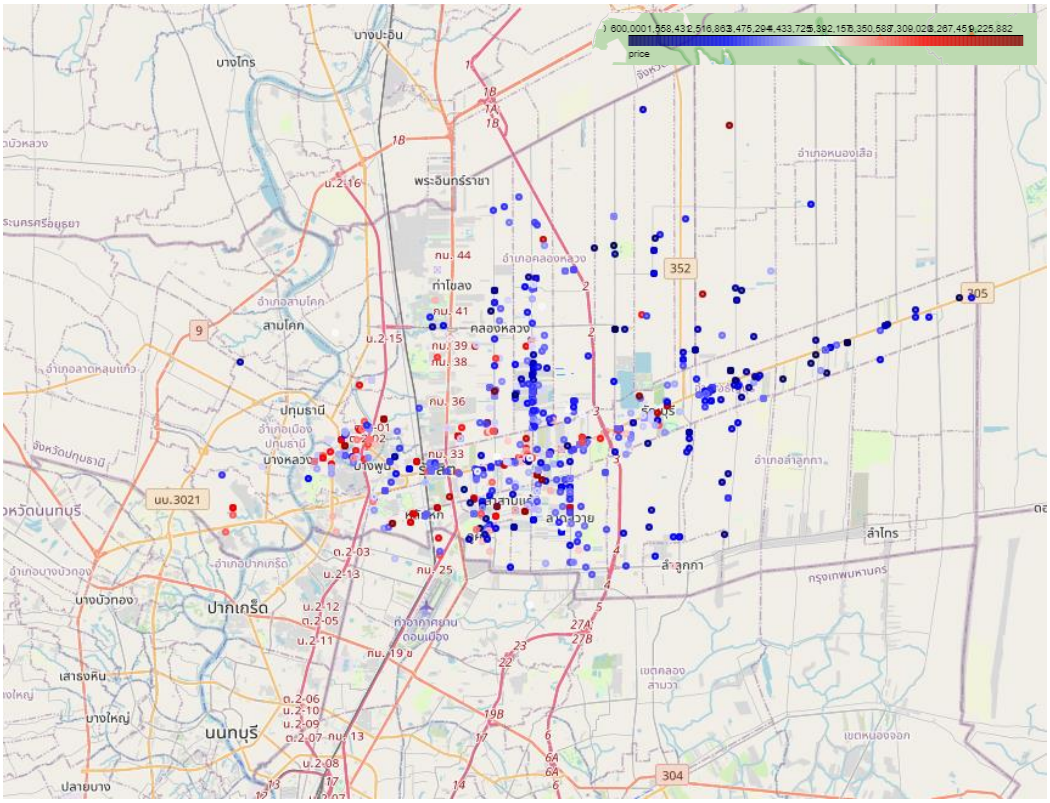
	name	price	latlong	pic	stair	bed	bath	space	area_usable	parking	kitchen	living	maid
0	ขาย บ้านเดี่ยว ราคาคุ้มสุด เวนิว ติวานนท์-รังสิต 175 ตรม. 50 ตร.วา SC ASSET Brand	6,790,000 THB	"https://www.google.com/maps/search/?api=1&query=13.963,100.495"	https://cdn.baania.com/b20/listings/6373b862b53b839c18e3fd1c/main/original/main.jpeg	2 ชั้น	3 ห้องนอน	3 ห้องน้ำ	50 ตร.ว.	175 ตร.ม.	NaN	NaN	NaN	NaN
1	ขาย บ้านสวย ราคาสุดคุ้ม !! บุราสิริ-รังสิต หลังมุม 61.8 ตร.ว 4 นอน 4 น้ำ แถมฟรี เฟอร์นิเจอร์ Build in และเครื่องใช้ไฟฟ้าทั้งหลัง ใกล้ทางด่วน และ รถไฟฟ้ารังสิต	7,300,000 THB	href="https://www.google.com/maps/search/?api=1&query=13.993385991223946,100.56884765625001	https://s3.ap-southeast-1.amazonaws.com/cdn.baania.com/Baania2/listing/2022/April/18/main/1650276644851-1650276627684-1150D51A-D3E1-4BF1-8E12-5CFCC0F16345.jpeg	2 ชั้น	4 ห้องนอน	4 ห้องน้ำ	61.8 ตร.ว.	163 ตร.ม.	2 ที่จอดรถ	1 ห้องครัว	1 ห้องนั่งเล่น	NaN

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.1.2 จากข้อมูลที่ได้จากเว็บไซต์จำนวน 1,551 ข้อมูล ทำการเลือกข้อมูลช่วงช่วงราคาไม่เกิน 10 ล้านบาท ทำการตัดข้อมูล missing และข้อมูลซ้ำอันเนื่องมาจากมีผู้ลงประกาศขายบ้านหลังเดียวกันซ้ำหลายเพจ เหลือข้อมูลที่นำมาพิจารณาต่อ 1,228 ข้อมูล

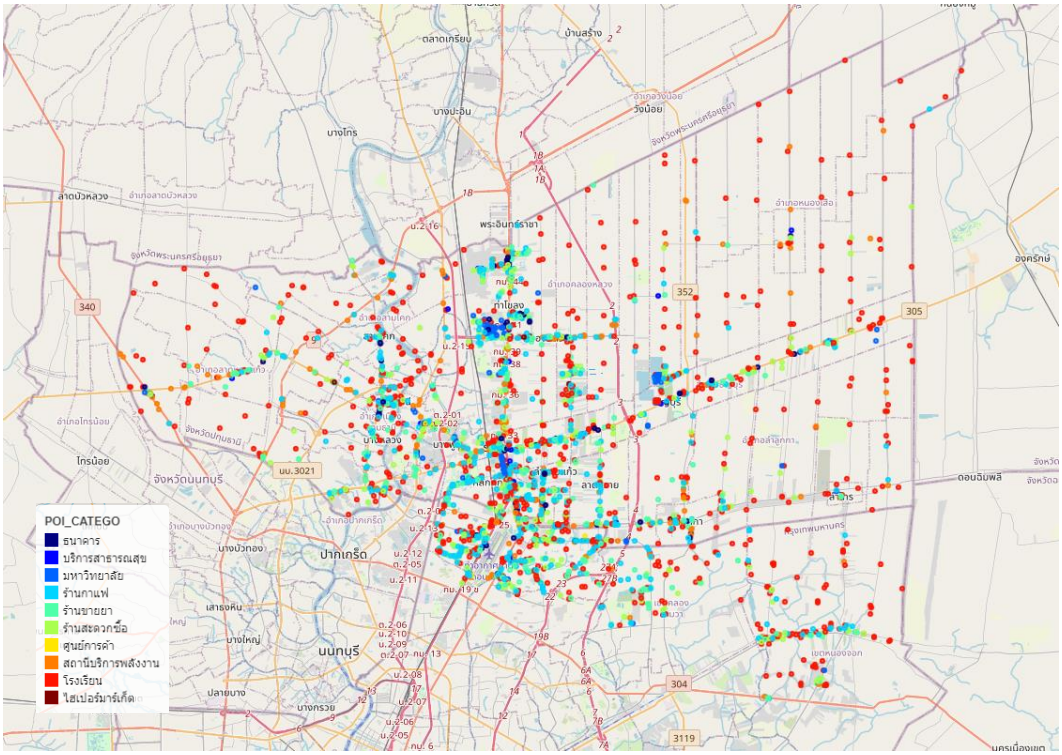
3.1.3 ข้อมูลที่ไม่ถูกนำมาพิจารณาใช้ในการทำนายราคาบ้าน ได้แก่ ชื่อคำประกาศขาย พื้นที่ใช้สอย จำนวนชั้น จำนวนที่จอดรถ จำนวนห้องครัว จำนวนห้องนั่งเล่น จำนวนห้องแม่บ้าน ส่วนข้อมูลที่ถูกพิจารณานำมาใช้ ได้แก่ จำนวนห้องนอน จำนวนห้องน้ำ พื้นที่ดิน ราคา url ของรูปภาพ แล url ของตำแหน่งละติจูด ลองจิจูด บน google map จากนั้นทำการแปลงข้อมูล จำนวนห้องนอน ห้องน้ำ พื้นที่ จากข้อมูลตัวหนังสือให้เป็นตัวเลขเพื่อใช้ในการคำนวน

3.1.4 ทำการแปลงข้อมูล ละติจูต ลองจิจูด เป็นจุด POINT ข้อมูลเชิงพื้นที่ด้วยไลบรารี geopandas



ภาพที่ 3.2 ตำแหน่งบ้าน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.3 ตำแหน่ง POIs

3.1.5 นำเข้าข้อมูล POIs (Point of Interests) จำนวน 11 ประเภท ในจังหวัดปทุมธานี ได้แก่ โรงเรียน, สถานีบริการพลังงาน, ร้านสะดวกซื้อ, ไฮเปอร์มาร์เก็ต, ธนาคาร, บริการสาธารณสุข, มหาวิทยาลัย, ร้านกาแฟ , ร้านขายยา, ศูนย์การค้า, สถานีรถไฟฟ้า BTS

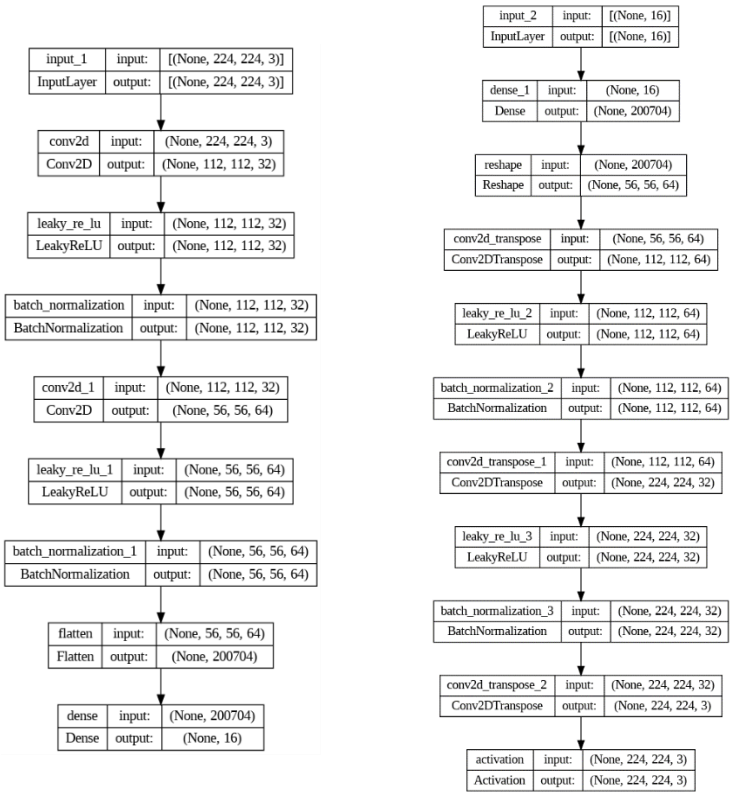
3.1.6 สร้างตัวแปรจาก POIs กับบ้าน ได้แก่ ระยะทางที่ใกล้ที่สุดของแต่ละ POIs จำนวน POI ในระยะ 1 กิโลเมตร จำนวน POI ในระยะ 3 กิโลเมตร และจำนวน POI ในระยะ 5 กิโลเมตร

3.2 การสร้างแบบจำลองเพื่อขัดแยกรูปบ้าน

เนื่องจากข้อมูลรูปภาพที่ได้จากเว็บไซต์ มีผู้ประกาศขายลงรูปที่ไม่ใช่รูปภาพด้านข้างหน้าบ้าน เช่น รูปภาพภายในบ้าน รูปป้าย เป็นต้น ในสารนิพนธ์นี้ต้องการใช้รูปภาพหน้าด้านหน้าบ้าน เป็นปัจจัยในการทำนายราคาบ้าน จึงต้องการที่คัดเลือกข้อมูลรูปภาพที่ไม่ใช่รูปด้านหน้าบ้านออกไป

3.2.1 ทำการสร้างแบบจำลอง AutoEncoder อ้างอิงจาก pyimagesearch [8] โดยมีโครงสร้างดังภาพที่ 3.4

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.4 โครงสร้างแบบจำลอง AutoEncoder

3.2.2 นำรูปภาพบ้านเป็นข้อมูลสำหรับการเทรนแบบจำลอง

3.2.3 หลังจากการเทรนแบบจำลอง นำแบบจำลองที่ได้ ไปใช้หา reconstruction error ของแต่ละภาพ

3.2.4 ภาพที่ไม่ใช่ภาพบ้านจะมีค่า reconstruction error ที่สูง ภาพที่มี reconstruction error เกินค่า threshold 75% จะถูกตัดออก ดังตัวอย่างภาพที่ 3.5 ทำให้เหลือข้อมูลที่ใช้ในการศึกษา 923 ข้อมูล ดังตัวอย่างภาพที่ 3.6

ภาพที่ 3.5 ตัวอย่างภาพที่ทำการแยกออก

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาพที่ 3.6 ตัวอย่างภาพด้านหน้าบ้านที่นำไปใช้ในการทำนายราคาบ้าน

3.3 การสร้างแบบจำลองเพื่อสกัดตัวแปรจากภาพ

เพื่อนำไปสู่การสกัดตัวแปรที่เป็นตัวแทนลักษณะสำคัญของรูปภาพด้านหน้าบ้าน ได้มีสร้างแบบจำลองจาก ResNet50 ที่มีการปรับปรุงส่วนปลายเพื่อใช้ทำนายช่วงของราคาบ้าน มีขั้นตอนดังต่อไปนี้

3.3.1 ทำการแบ่งกลุ่มข้อมูลตามช่วงราคา ซึ่งในการทดลองมีการแบ่งกลุ่ม 3 รูปแบบ

รูปแบบที่ 1 แบ่ง 3 กลุ่ม

(1) ราคาต่ำกว่า 3 ล้านบาท

(2) ราคาในช่วง 3-7 ล้านบาท

(3) ราคา 7 ล้านบาทขึ้นไป

รูปแบบที่ 2 แบ่ง 2 กลุ่ม

(1) ราคาต่ำกว่า 4 ล้านบาท

(2) ราคา 4 ล้านบาทขึ้นไป

รูปแบบที่ 3 แบ่ง 2 กลุ่ม

(1) ราคาต่ำกว่า 5 ล้านบาท

(2) ราคา 5 ล้านบาทขึ้นไป

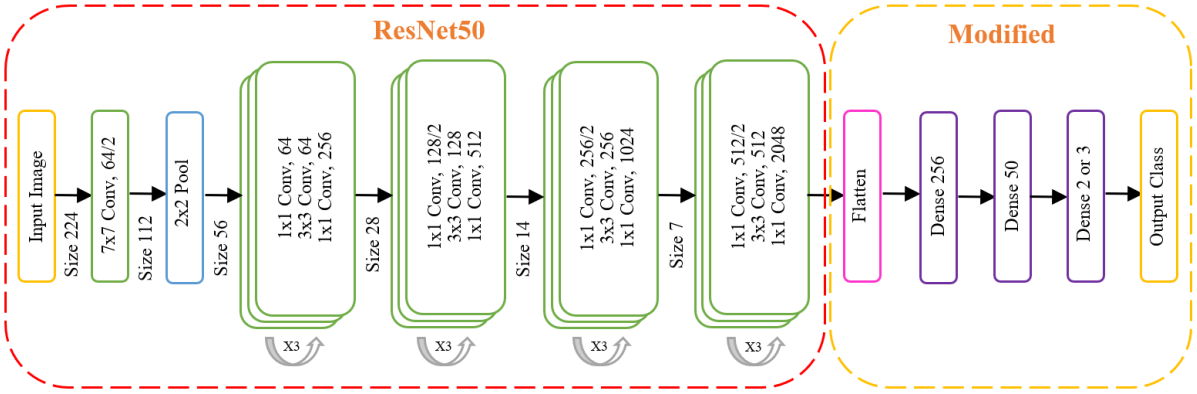
3.3.2 สร้างแบบจำลองเพื่อทำการจำแนกแบ่งกลุ่มของช่วงราคาบ้านจากรูปภาพ โดยใช้ แบบจำลอง ResNet50 Pre-trained model ใช้ weight เป็น ImageNet ที่มีการนำชั้นด้านบนออก เพิ่มชั้นโครงข่ายประสาทเทียม 256 จุด เพิ่มชั้นโครงข่ายประสาทเทียม 50 จุด และเพิ่มชั้นสุดท้ายเพื่อจำแนกแบ่งกลุ่ม 2 จุด หรือ 3 จุด ขึ้นอยู่กับจำนวนกลุ่มช่วงราคาที่แบ่ง ดังภาพที่ 3.7

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ค่าพารามิเตอร์ที่ตั้งค่ามีดังต่อไปนี้

optimizer = Adam (learning_rate = 0.0001)

loss = categorical_crossentropy หรือ BinaryFocalCrossentropy



ภาพที่ 3.7 โครงสร้างส่วนปลายของแบบจำลอง ResNet50 ที่ทำการปรับปรุง

3.3.3 ทำการเตรียมภาพเพื่อนำเข้าแบบจำลอง โดยแปลงเป็นขนาด 224x224 สำหรับขาเข้าของ ResNet50 และทำการทำ augmentation รูปภาพเพื่อลด overfitting

3.3.4 ให้แบบจำลองเรียนรู้ และทำการบันทึกค่าน้ำหนักที่ดีที่สุด

3.3.5 นำแบบจำลองที่ได้ตัดส่วนชั้นสุดท้ายที่มีการแยกกลุ่มออก และนำมาสร้าง 50 ตัวแปร จากรูปภาพทั้งหมดเพื่อใช้เป็นตัวแปรจากรูปภาพ สำหรับขั้นตอนถัดไป

3.4 การสร้างแบบจำลองประเมินราคาบ้าน

3.4.1 สร้างแบบจำลองสำหรับประเมินราคาบ้าน โดยใช้แบบจำลอง XGBoost และทำการตั้งค่า learning rate = 0.1

3.4.2 นำข้อมูลบ้าน 3 ตัวแปร ข้อมูลเชิงพื้นที่ 46 ตัวแปร และตัวแปรที่สกัดได้จากรูปภาพที่ได้จากการเรียนรู้ในขั้นตอน 3.3 ทั้ง 50 ตัวแปร ทำการเรียงต่อข้อมูลเพื่อเข้าสู่แบบจำลอง

3.4.2 นำข้อมูลมาทำการ normalization ทำการแบ่งข้อมูลออกเป็น train : test 75 : 25 ตามลำดับ และนำเข้าแบบจำลองเพื่อเรียนรู้ โดยมีการใช้ random_state = 100

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.5 เครื่องมือที่ใช้ในการศึกษา

3.4.1 ภาษา Python programming language

เป็นภาษาระดับสูงมุ่งที่มีไวยากรณ์ที่เป็นระเบียบ อ่านง่าย มีขนาดเล็ก และง่ายต่อการใช้งาน รองรับการเขียนโปรแกรมในรูปแบบทำงานแบบเชิงวตถุ มีบไลบรารีมาตรฐานจำนวนมาก เช่นโครงสร้างข้อมูลแบบซับซ้อน และไลบรารีสำหรับคณิตศาสตร์ ที่ช่วยการทำงานและพัฒนา เป็นภาษาโปรแกรมที่มีความหยืดหยุ่นสูง สามารถนำไปใช้ในงานหลากหลายอย่างทำให้เป็นที่นิยมในอุตสาหกรรม เช่น การพัฒนาเว็บแอพพลิเคชั่น การวิเคราะห์ข้อมูล

3.4.2 โปรแกรม Google Colab

เป็นเครื่องมือสำหรับเขียนและเรียกใช้งานโค้ด Python แบบออนไลน์ที่พัฒนาโดย Google ซึ่งให้บริการในรูปแบบของสมุดบันทึก (notebook) ที่ใช้งานผ่านเบราว์เซอร์ ซึ่งเป็นส่วนหนึ่งของ Google Drive ซึ่งเป็นบริการคลาวด์เก็บข้อมูลของ Google โดยสามารถ สร้าง Notebook เขียนโปรแกรมภาษา Python ได้ฟรีและมี GPU, TPU ให้ได้ใช้ฟรี

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

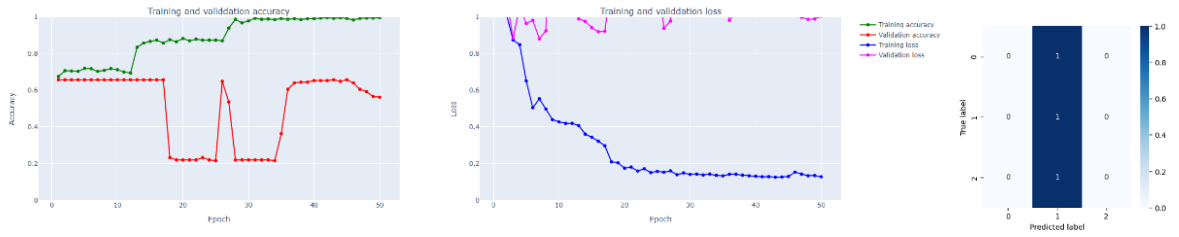
บทที่ 4
ผลการวิจัย

การศึกษาแนวทางในการปรับปรุงประสิทธิภาพของการประเมินราคาบ้านพักอาศัย โดยการใช้ข้อมูลเชิงพื้นที่ และรูปภาพ เพิ่มเติมนอกเหนือจากการประเมินราคาจากเพียงข้อมูลบ้าน มีรายละเอียดผลการทดสอบดังนี้

4.1 ผลการวัดประสิทธิภาพของแบบจำลองในการทำนายช่วงราคาบ้านจากรูปภาพ

เพื่อนำไปสู่การสกัดตัวแปรที่เป็นตัวแทนลักษณะสำคัญของรูปภาพ ได้มีสร้างแบบจำลองจาก ResNet50 ที่มีการปรับปรุงส่วนปลายเพื่อใช้ทำนายช่วงกลุ่มของราคาบ้านจากข้อมูลรูปภาพ มีการแบ่งกลุ่มของราคาบ้าน 3 รูปแบบและทำการทดสอบ ดังแสดงในตารางที่ 4.1

ในการทดลองแบ่งกลุ่มราคาออกเป็น 3 ช่วง นั่นคือในช่วงราคาต่ำกว่า 3 ล้านบาท (Class0) ราคาในช่วง 3 – 7 ล้านบาท (Class1) ราคา 7 ล้านบาทขึ้นไป (Class2) โดยอ้างอิงการแบ่งช่วงราคาตามงานวิจัย Gan S et al. 2021 [1] ซึ่งจำนวนของแต่ละ Class มีค่าดังต่อไปนี้ Class0 : Class1 : Class2 174 : 642 : 107 รูปภาพ ตามลำดับ ทำนายโดยการใช้ loss เป็น categorical_crossentropy ผลการทดลองที่ได้แบบจำลองทำนาย ราคาในช่วง 3 – 7 ล้านบาท (Class1) เพียงอย่างเดียว ดังภาพ ที่ 4.1 เนื่องจากจำนวนข้อมูลของ Class1 ที่มีจำนวนมาก

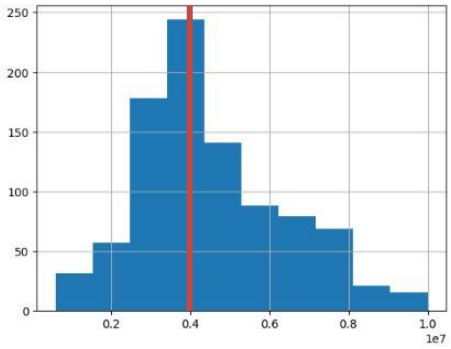


ภาพที่ 4.1 กราฟแสดงผลการทดลองแบ่ง 3 ช่วงโดยการแบ่งที่ 3 และ 7 ล้านบาท

เทคนิค under-sampling OneSideSelection ถูกพิจารณานำมาใช้ เนื่องจากข้อมูลมีจำนวนแตกต่างกันมาก จากการ under-sampling จำนวนของข้อมูลแต่ละกลุ่ม มีค่าดังต่อไปนี้ Class0 : Class1 : Class2 174 : 258 : 107 รูปภาพ ตามลำดับ ผลการทดลองที่ได้ แบบจำลองยังคงทำนาย ราคาในช่วง 3 – 7 ล้านบาท (Class1) เพียงอย่างเดียว ดังภาพ ที่ 4.2

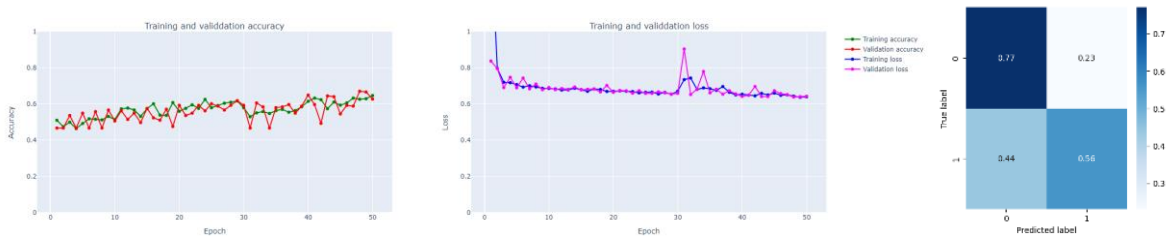
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ภาพที่ 4.2 กราฟแสดงผลการทดลองแบ่ง 3 ช่วงโดยการแบ่งที่ ที่ 3 และ 7 ล้านบาท และทำการ under-sampling ข้อมูล



ภาพที่ 4.3 ฮิสโตแกรม ของราคาบ้าน

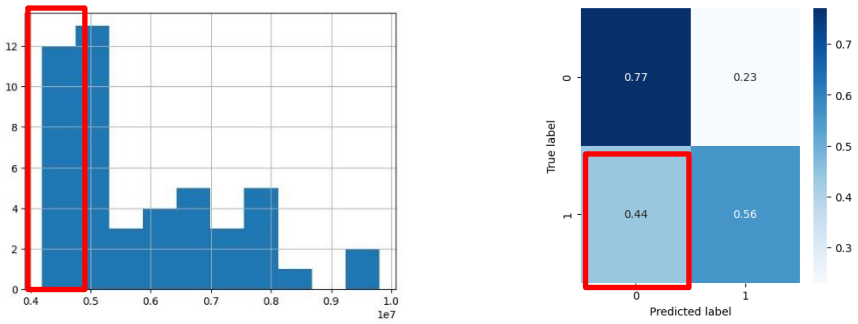
เนื่องจากจำนวนในแต่ละกลุ่มมีจำนวนที่แตกต่างกันมาก เมื่อตรวจสอบจากการกระจายตัวของข้อมูลในภาพที่ 4.3 ทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม ดังต่อไปนี้ ราคาต่ำกว่า 4 ล้านบาท (Class0) ราคา 4 ล้านบาทขึ้นไป (Class1) ข้อมูลของแต่ละกลุ่มมีจำนวนที่ใกล้เคียงกันดังต่อไปนี้ Class0 : Class1 442 : 481 รูปภาพ ตามลำดับ และผลที่ได้จากการทดลอง แบบจำลองสามารถแยกกลุ่มราคาจากรูป โดยมีประสิทธิภาพดังภาพ ที่ 4.4



ภาพที่ 4.4 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 4 ล้านบาท

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สำหรับรูปบ้านที่มีราคาสูงกว่า 4 ล้านบาท แต่แบบจำลองทำนายต่ำกว่า 4 ล้านบาท (44%) ภาพที่ 4.5 (ขวา) พบว่าอาจจะเกิดจาก ข้อมูลภาพบ้านราคา 4 ถึง 5 ล้านมีปริมาณใกล้เคียงกัน ดังรภาพที่ 4.5 (ซ้าย)



ภาพที่ 4.5 กราฟการกระจายตัวของราคาในส่วนของรูปที่มีราคาสูงกว่า 4 ล้านแต่แบบจำลองทำนายต่ำกว่า 4 ล้านบาท (ซ้าย) และ confusion matrix ผลการทำนายด้วยแบบจำลองที่แบ่งที่ราคา 4 ล้านบาท (ขวา)

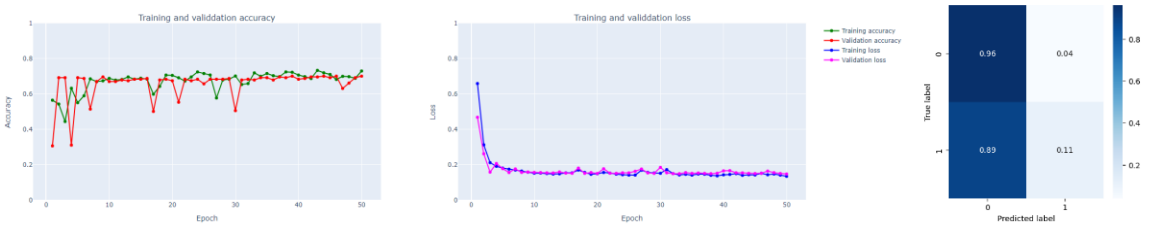
ดังนั้นจึงทดลองปรับเป็น 2 ช่วงคือ 1) ราคาต่ำกว่า 5 ล้านบาท และ 2) 5 ล้านบาทขึ้นไป โดยมีสมุติฐานว่าจะทำนายแม่นยำขึ้น

ทำการแบ่งกลุ่มข้อมูลออกเป็นราคาต่ำกว่า 5 ล้านบาท (Class0) ราคา 5 ล้านบาทขึ้นไป (Class1) มีจำนวนข้อมูล Class 0: Class 1 625: 298 รูปภาพ ตามลำดับ ผลการทดลองที่ได้ แบบจำลองทำนายรูปภาพส่วนใหญ่เป็น Class0 เนื่องจากข้อมูลที่ใช้เทรน Class0 มีจำนวนมาก

ภาพที่ 4.6 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 5 ล้านบาท

BinaryFocalCrossentropy เป็น loss ที่ถูกพิจารณาใช้เมื่อข้อมูลมีจำนวนที่แตกต่างกัน อย่างไรก็ตามผลการทดลองที่ได้ แบบจำลองทำนายรูปภาพส่วนใหญ่เป็น Class0

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.7 กราฟแสดงผลการทดลองแบ่ง 2 ช่วงโดยการแบ่งที่ ที่ 5 ล้านบาท และใช้ loss เป็น BinaryFocalCrossentropy

ท้ายที่สุด แบบจำลองที่เกิดการเรียนรู้โดยแบ่งข้อมูลเป็น 2 กลุ่ม โดยที่แบ่งเป็น 1) ต่ำกว่าราคา 4 ล้านบาท 2) ราคาตั้งแต่ 4 ล้านบาทขึ้นไป ให้ความแม่นยำมากที่สุดในการทำนายช่วงราคาจากรูปภาพ แบบจำลองนี้ถูกนำมาใช้ในการสกัดข้อมูลสำคัญของรูปภาพออกมาเป็นจำนวน 50 ตัวแปร ในการใช้เป็นตัวแปรตัวแทนของรูปภาพ เพื่อนำมาร่วมกับ ข้อมูลบ้าน และข้อมูลเชิงพื้นที่ ในการสร้าง Regression แบบจำลองต่อไป

ตารางที่ 4.1 แสดงประสิทธิภาพทำนายช่วงราคาบ้านจากรูปภาพ

	Number of Classes / สัดส่วนรูปภาพแต่ละ class	Description	loss	Validation accuracy	Validation loss
1	3 Class0 : Class1 : Class2 174 : 642 : 107	Class0: ต่ำกว่า 3 ล้านบาท Class1: ระหว่าง 3-7ล้านบาท Class2: 7 ล้านบาทขึ้นไป	categorical_crossentropy	0.6667	0.8400
2	3 และ under-sampling Class0 : Class1 : Class2 174 : 258 : 107	Class0: ต่ำกว่า 3 ล้านบาท Class1: ระหว่าง 3-7ล้านบาท Class2: 7 ล้านบาทขึ้นไป	categorical_crossentropy	0.6977	0.8561
3	2 Class0 : Class1 442 : 481	Class0: ต่ำกว่า 4 ล้านบาท Class1: 4 ล้านบาทขึ้นไป	categorical_crossentropy	0.6433	0.6277
4	2 Class 0 : Class 1 625: 298	Class0: ต่ำกว่า 5 ล้านบาท Class1: 5 ล้านบาทขึ้นไป	categorical_crossentropy	0.6926	0.6027
5	2 Class 0 : Class 1 625 : 298	Class0: ต่ำกว่า 5 ล้านบาท Class1: 5 ล้านบาทขึ้นไป	BinaryFocalCrossentropy	0.7013	0.1486

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

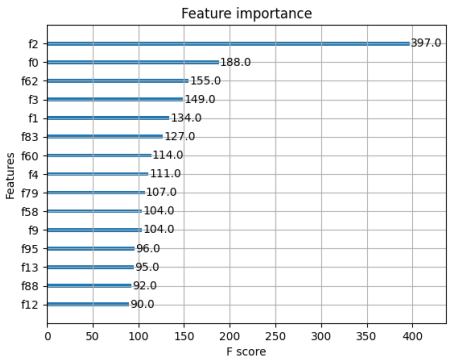
4.2 ผลการวัดประสิทธิภาพของแบบจำลองของการประเมินราคา

ผลการประเมินราคาบ้านโดยใช้แบบจำลอง XGBoost ในการทำนาย ซึ่งมีข้อมูลตัวแปรที่นำมาใช้ 3 กลุ่ม ได้แก่ (1) ข้อมูลบ้าน (2) ข้อมูลเชิงพื้นที่ (3) ข้อมูลตัวแปรที่สกัดได้จากรูปภาพ ซึ่งได้มาจากหัวข้อที่ 4.1 ให้ผลการทดสอบแสดงดัง ตารางที่ 4.1 และภาพที่ 4.8 โดยที่ การใช้ข้อมูลบ้านร่วมกับข้อมูลเชิงพื้นที่ทำให้แบบจำลองมีความแม่นยำมากขึ้นจากการประเมินราคาบ้านโดยใช้ข้อมูลบ้านอย่างเดียว และการใช้ข้อมูลรูปภาพด้านหน้าบ้านร่วมกับข้อมูลเชิงพื้นที่และข้อมูลบ้าน ให้แบบจำลองที่แม่นยำมากที่สุด และค่าความคลาดเคลื่อนในการทำนายราคา กระจายทั่วทั้งพื้นที่ไม่ได้เกาะกลุ่มในบริเวณใด ดังภาพที่ 4.10

ตารางที่ 4.2 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองของการประเมินราคา

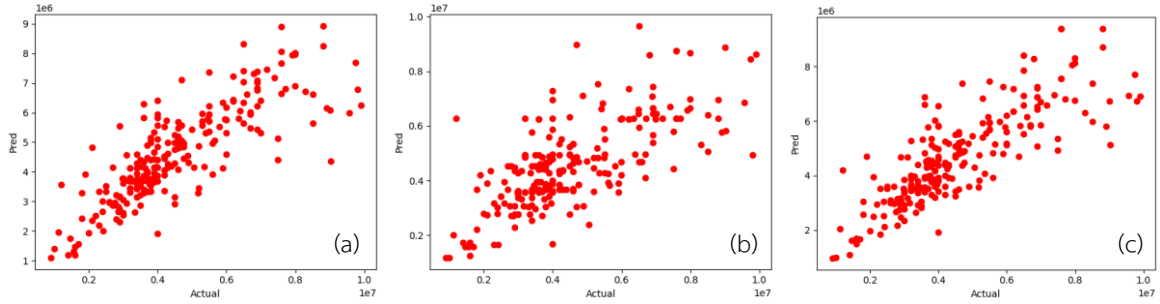
Features / หน่วยวัด	House Info.	House Info. + POI	House Info. + POI + Image
MAE	0.09848	0.07672	**0.07444**
MAPE	0.24538	0.18758	**0.18190**
RMSE	0.13445	0.10821	**0.10777**
R2	0.49239	0.67122	**0.67386**
STD	34.09 %	24.37 %	**21.81%**

สำหรับแบบจำลองที่ใช้ข้อมูลบ้าน ข้อมูลเชิงพื้นที่และรูปภาพ มีตัวแปรที่สำคัญ (Feature Importance) เรียงลำดับจากมากไปน้อย ดังต่อไปนี้ พื้นที่ (f2) จำนวนห้องนอน (f0) ตัวแปรรูปภาพ (f62) ละติจูด (f3) จำนวนห้องน้ำ (f1) ตัวแปรรูปภาพ (f83,f60) ลองกิจูด (f4) ตัวแปรรูปภาพ (f79,f58) ไฮเปอร์มาร์เก็ต (f9) ตัวแปรรูปภาพ (f95) โรงพยาบาล (f13) ตัวแปรรูปภาพ (f88) ศูนย์การค้า (f12)

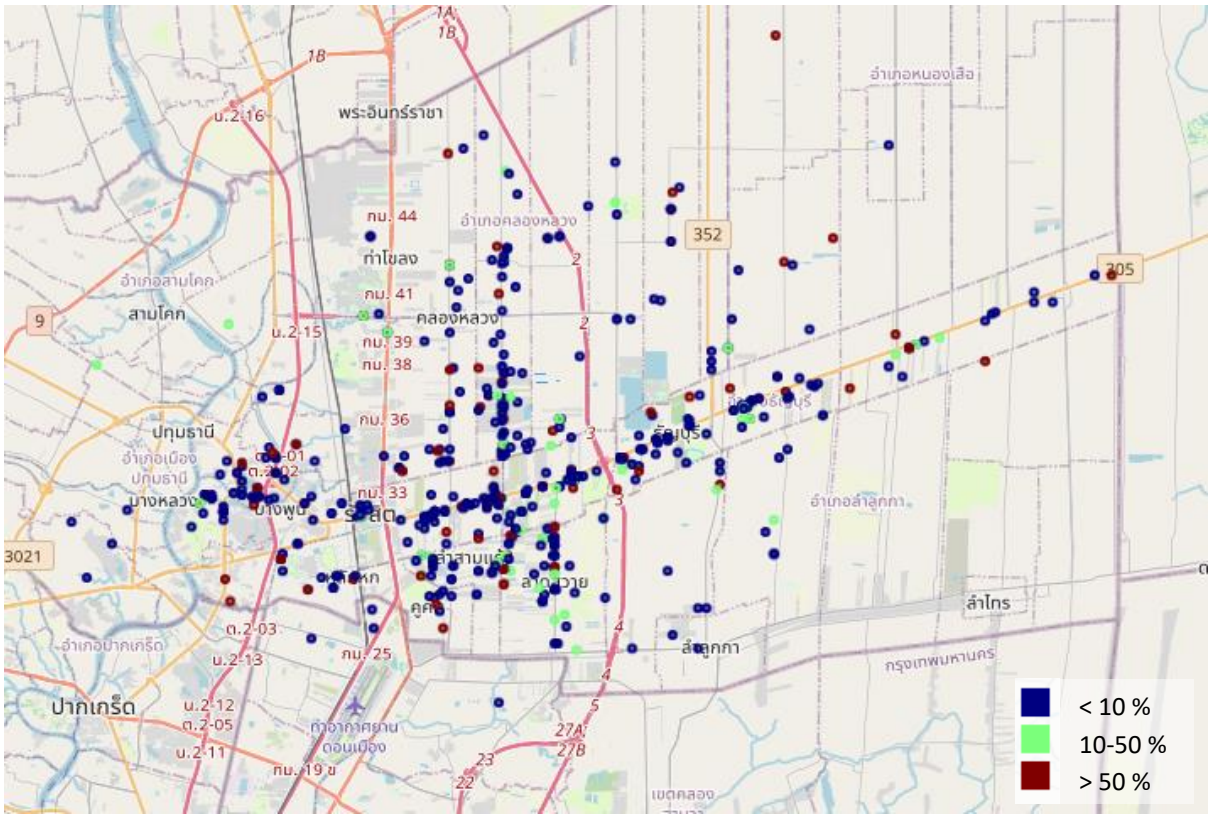


ภาพที่ 4.8 แสดงตัวแปรที่สำคัญ (Feature Importance)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 4.9 กราฟแสดงราคาจริง และราคาที่ทำนายโดยแบบจำลอง โดยใช้ปัจจัย (a) ใช้ข้อมูลบ้าน (b) ใช้ข้อมูลบ้านและข้อมูลเชิงพื้นที่ (c) ใช้ข้อมูลบ้าน ข้อมูลเชิงพื้นที่ และข้อมูลบ้าน



ภาพที่ 4.10 แสดง % error ของราคาที่ทำนายได้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาแนวทางในการปรับปรุงประสิทธิภาพของการประเมินราคาบ้านพักอาศัย โดยการใช้ข้อมูลเชิงพื้นที่ และรูปภาพ เพิ่มเติมนอกเหนือจากการประเมินราคาจากเพียงข้อมูลบ้าน สามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการศึกษา

ในสารนิพนธ์นี้เสนอแนวทางในการเพิ่มความแม่นยำในการประเมินราคาบ้านโดยใช้ปัจจัย 1) ข้อมูลบ้าน (House Info) 2) ข้อมูลเชิงพื้นที่ (Spatial Info) 3) ข้อมูลภาพบ้านด้านหน้า (Front) โดยทำการเสนอแนวทางในการสกัดปัจจัยจากข้อมูลรูปภาพบ้าน (Front) ดังขั้นตอนดังนี้ 1) กำจัดรูปภาพที่ไม่ใช้ภาพบ้าน (Front) โดยใช้ AutoEncoder 2) สร้าง Classifier สำหรับแบ่งข้อมูลออกเป็น 2 ช่วงราคาคือ ราคาต่ำกว่า 4 ล้านบาท และมากกว่าหรือเท่ากับ 4 ล้านบาท (ทั้งนี้ขึ้นกับการกระจายตัวของราคาบ้านที่นำมาสร้างแบบจำลอง) อย่างไรก็ตามการใช้ Features จากรูปภาพบ้าน (Front) มีการเพิ่มความแม่นยำในการทำนายราคาบ้านให้แบบจำลอง แต่ยังไม่ส่งผลอย่างมีนัยสำคัญ เนื่องจากรูปภาพบ้าน (Front) เพียงอย่างเดียวไม่อาจจะแยกความแตกต่างของราคาบ้านได้

5.2 ข้อเสนอแนะ และ future work

ในการศึกษานี้ทำการศึกษาเพียงจังหวัดเดียวปทุมธานี เพื่อให้ขอบเขตการศึกษากว้างขึ้น และเพิ่มโอกาสในการปรับปรุงประสิทธิภาพในการประเมินราคา จึงมีข้อเสนอแนะดังต่อไปนี้

5.2.1 สร้างแบบจำลองทำนายแยกแต่ละจังหวัด

5.2.2 เปรียบแบบจำลองแต่ละจังหวัด กับ แบบจำลองกลางที่รวมทุกจังหวัด

5.2.3 เพิ่มข้อมูลภาพถ่ายดาวเทียม และภาพการตกแต่งภายในตัวบ้าน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

[1] G. E. Hinton and R. S. Zemel. "Autoencoders, minimum description length and Helmholtz free energy." *Advances in neural information processing systems 6*. 1993.

[2] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[3] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

[4] G. Srirutchataboon, S. Prasertthum, E. Chuangsuwanich, P. N. Pratanwanich and C. Ratanamahatana, "Stacking Ensemble Learning for Housing Price Prediction: a Case Study in Thailand," *2021 13th International Conference on Knowledge and Smart Technology (KST)*, Bangsaen, Chonburi, Thailand, 2021, pp. 73-77, doi: 10.1109/KST51265.2021.9415771.

[5] P. -Y. Wang, C. -T. Chen, J. -W. Su, T. -Y. Wang and S. -H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism," in *IEEE Access*, vol. 9, pp. 55244-55259, 2021, doi: 10.1109/ACCESS.2021.3071306.

[6] A. Nouriani and L. Lemke, “Vision-based housing price estimation using interior, exterior & satellite images,” *Intelligent Systems with Applications*, vol. 14, p. 200081, May 2022, doi: https://doi.org/10.1016/j.iswa.2022.200081.

[7] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, “Anomaly detection of defects on concrete structures with the convolutional autoencoder,” *Advanced Engineering Informatics*, vol. 45, p. 101105, Aug. 2020, doi: https://doi.org/10.1016/j.aei.2020.101105.

[8] A. Rosebrock, “Anomaly detection with Keras, TensorFlow, and Deep Learning,” *PyImageSearch*, Mar. 02, 2020. https://pyimagesearch.com/2020/03/02/anomaly-detection-with-keras-tensorflow-and-deep-learning/

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ประวัติผู้เขียน

ชื่อ – นามสกุล วรภร เล็กอาราม

ประวัติการศึกษา

พ.ศ. 2555 - วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมเคมี มหาวิทยาลัยเกษตรศาสตร์

ประสบการณ์ทำงาน

พ.ศ. 2566 - Process Engineer Specialist
TTCL Public Company Limited

พ.ศ. 2556 - Technology Development Engineer
Mitsui Hygiene Material (Thailand) Company Limited