การพยากรณ์ราคาคอนโดมิเนียมในกรุงเทพมหานครโดยเทคนิค Web Mining

PREDICTING CONDOMINIUM PRICE IN BANGKOK USING WEB MINING TECHNIQUES

จิรพล สังข์โพธิ์ *ศราวุธ แรมจันทร์**

***Jirapon Sunkpho**, Sarawut Ramjan**

วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์

College of Innovation, Thammasat University.

**Corresponding author, e-mail: sarawut@citu.tu.ac.th*

***Received:** 2 July 2019; **Revised:** 13 January 2020; **Accepted:** 24 February 2020*

บทคัดย่อ

ในกรุงเทพมหานคร คอนโดมิเนียมกลายเป็นที่อยู่อาศัยหลังแรกของคนส่วนใหญ่โดยเฉพาะอย่างยิ่งในกลุ่มคนเจเนเรชัน Y โดยราคาเป็นหนึ่งในปัจจัยสำคัญเมื่อลูกค้าตัดสินใจซื้อคอนโดมิเนียม บทความวิจัยฉบับนี้มุ่งสร้างโมเดลในการพยากรณ์ราคาขายต่อตารางเมตรของคอนโดมิเนียมในกรุงเทพมหานครเพื่อให้ผู้ตัดสินใจซื้อคอนโดมิเนียมนำไปใช้ในการประมาณการมูลค่าของคอนโดมิเนียมในอนาคตประกอบการตัดสินใจซื้อ การวิจัยได้รวบรวมข้อมูลจากเว็บไซต์ hipflat.co.th ผ่านเทคนิค Web Crawling และเทคนิค Mining โดยได้รับข้อมูล คือ ข้อมูลโครงการคอนโดมิเนียม จำนวน 1,465 แห่ง ซึ่งถูกสกัดให้อยู่ในรูปแบบระเบียนที่มีจำนวน 15 แอตทริบิวต์ โมเดลถูกพัฒนาขึ้นโดยใช้รูปแบบ Deep Learning ซึ่งมีความแม่นยำในการพยากรณ์สูงสุดหลังจากการทดสอบด้วยโมเดลในลักษณะ Supervised Learning จำนวน 5 รูปแบบ จากการวิจัยพบว่า ราคาของคอนโดมิเนียมในกรุงเทพมหานครสอดคล้องกับผลลัพธ์ที่ได้รับจากการพยากรณ์ด้วยโมเดลโดยข้อมูลในการพยากรณ์ผันแปรตามจำนวนห้องของคอนโดมิเนียมในแต่ละโครงการ ระยะทางจากสถานีรถไฟฟ้าและจำนวนปีที่คอนโดมิเนียมถูกสร้างขึ้นมา โมเดลสามารถพยากรณ์ราคาคอนโดมีเนียมโดยมีความคลาดเคลื่อนของการพยากรณ์อยู่ที่ร้อยละ 17.53

คำสำคัญ: เว็บครอว์ลิงค์ เหมืองข้อมูล การเรียนรู้ของเครื่อง

Abstract

In Bangkok, condominium is becoming a first home for many people especially Generation Y who prefer convenience of lifestyle. Price of the condominium is one of the most important factors for buyers' decision-making. This study aims to formulate a model that predict a selling price per square meter of a condominium in Bangkok. This is to support buyer's decision-making whether to buy or not by using the price prediction from the model. The training dataset was gathered using web crawling techniques from hipflat.co.th. There were 1,465 condominium projects in Bangkok were extracted with each record contains

15 attributes of information. This model was formulated by using a deep learning method based on the best performed after testing the model with testing dataset. Results show that the price of the condominium in Bangkok highly affected by the number of units in the condominium project, its distance from the BTS metro line, and the year of its construction, with the relative error of 17.53%.

Keywords: Web Crawling, Mining, Machine Learning

บทนำ

จากประชากรมากกว่า 10 ล้านคน กรุงเทพมหานครกลายเป็นหนึ่งในเมืองขนาดใหญ่แห่งหนึ่งของโลก ประกอบกับการดำเนินธุรกิจจำนวนมากในกรุงเทพมหานครยิ่งทำให้เกิดการย้ายถิ่นฐานของประชากรต่างจังหวัดเข้ามาทำงานในกรุงเทพมหานครส่งผลให้ความต้องการที่อยู่อาศัยประเภทคอนโดมิเนียมเพิ่มขึ้นตามไปด้วย เมื่อเทียบกับที่อยู่อาศัยในลักษณะพื้นราบซึ่งมีอยู่อย่างจำกัดและมีราคาสูงในเขตพื้นที่ทางธุรกิจ [1] โดยเฉพาะอย่างยิ่งในกลุ่มคนเจเนเรชัน Y ซึ่งมีความต้องการสูงสุดในการซื้อคอนโดมิเนียมในเขตชั้นในฝั่งตะวันตกและตอนบนของกรุงเทพมหานคร โดยกลุ่มคนเจเนเรชัน Y ยังมีอัตราความต้องการในการซื้อคอนโดมิเนียมสูงที่สุดเมื่อเทียบกับกลุ่มคนเจเนเรชันอื่น ๆ เพราะอยู่ในช่วงวัยซึ่งเริ่มต้นทำงานโดยเฉพาะในย่านธุรกิจ [2] ในปี ค.ศ. 2018 มูลค่าตลาดการซื้อขายคอนโดมิเนียมอยู่ที่ประมาณ 500 ล้านบาท [3] จนถึงก่อนกลางปี ค.ศ. 2017 มีคอนโดมีเนียมถึง 538,920 ห้อง ที่พร้อมจำหน่ายในกรุงเทพมหานครและมีส่วนหนึ่งที่ได้รับการจำหน่ายไปแล้วจำนวน 412,274 ห้อง คิดเป็นประมาณร้อยละ 76.5 ของอัตราการจำหน่าย [4] ราคาเฉลี่ยของคอนโดมิเนียมที่สร้างใหม่ในปี ค.ศ. 2017 อยู่ที่ 153,220 บาท ต่อตารางเมตร ซึ่งเพิ่มขึ้นจากปีก่อนหน้านั้นคิดเป็นร้อยละ 5.9 [4] ในย่านธุรกิจราคาค่อนโดมิเนียมมีราคาสูงถึง 241,585 บาท ต่อตารางเมตร ในขณะที่ย่านอื่น ๆ มีราคาอยู่ที่ 131,906 บาท [4] โดยราคาดังกล่าวยังผันแปรไปตามระยะทางจากสถานีรถไฟลอยฟ้าและสถานีรถไฟฟ้าใต้ดินอีกด้วย

ผู้คนจำนวนมากซื้อคอนโดมิเนียมเพื่อการอยู่อาศัยแต่ก็มีผู้คนอีกส่วนหนึ่งที่ซื้อเพื่อการลงทุนโดยคาดหวังมูลค่าที่เพิ่มขึ้นทุกปีและมีคนอีกจำนวนหนึ่งที่นิยมซื้อคอนโดมิเนียมที่ผ่านการซื้อขายมาแล้วแทนที่คอนโดมิเนียมสร้างใหม่ซึ่งมีราคาสูง การตัดสินใจซื้อคอนโดมิเนียมในราคาที่เหมาะสมเป็นเรื่องยากโดยผู้ซื้อจะพิจารณาจากทำเลที่ตั้ง สภาพอาคาร และความรับผิดชอบจากเจ้าของโครงการ มีงานวิจัยจำนวนมากซึ่งวิเคราะห์เกี่ยวกับปัจจัยที่ส่งผลต่อราคาคอนโดมิเนียมในประเทศโปแลนด์ มีงานวิจัยของ Chrostek และ Kopczewska เกี่ยวกับปัจจัยที่ส่งผลต่อราคาคอนโดมิเนียมโดยพบว่า ขนาดและอายุของคอนโดมิเนียมส่งผลอย่างมีนัยสำคัญเชิงบวกต่อราคาคอนโดมิเนียม [5] ในประเทศสิงค์โปร์ ทำเลที่ตั้ง ลักษณะของอาคารและสิทธิ์การถือครอง ส่งผลต่อราคาคอนโดมิเนียม [6] สำหรับคอนโดมิเนียมในกรุงเทพมหานคร จากงานวิจัยของโชติวุฒิ เหล่าไพโรจน์ พบว่ามี 3 ปัจจัย คือ ทำเลที่ตั้ง โครงสร้างอาคารและค่าบำรุงรักษาต่อเดือน ที่ส่งผลอย่างมากต่อราคาคอนโดมิเนียมในกรุงเทพมหานคร [7] สอดคล้องกับมนตรี ล้อเลิศสกุล และอานนท์ ศักดิ์วรวิชญ์ ซึ่งศึกษาโมเดลในการพยากรณ์ราคาขายเฉลี่ยต่อตารางเมตรของอาคารชุดในกรุงเทพมหานคร โดยใช้ช่วงระยะทางจากโครงการถึงสถานีรถไฟลอยฟ้า จำนวนช่องทางจราจรถนนผ่านหน้าโครงการและจำนวนชั้นของอาคารชุด มาใช้เป็นปัจจัยในการพยากรณ์ราคาคอนโดมิเนียม [8] อีกทั้งยังพบรายงานวิจัยเกี่ยวกับการกำหนดราคาสำหรับอาคารชุดระดับหรูหราในเขตกรุงเทพมหานคร โดยนุสรพร นัสบุสย์ ซึ่งระยะทางจากคอนโดมิเนียมไปสู่สถานีรถไฟลอยฟ้า ส่งผลต่อการกำหนดราคาจำหน่ายคอนโดมิเนียม [9] เช่นเดียวกับศศิธร กลันทกสุวรรณ ซึ่งศึกษาเกี่ยวกับปัจจัยที่มีอิทธิพลต่อการกำหนดราคาคอนโดมิเนียมระดับหรูและระดับหรูพิเศษ

ในเขตกรุงเทพมหานคร พบว่า ปัจจัยที่มีนัยสำคัญต่อการกำหนดราคาประกอบไปด้วย พื้นที่ทำเล ความผ่อนคลาย และความบันเทิง ระยะทางจากโรงพยาบาล ระยะทางจากสถานศึกษาและระยะทางจากสวนสาธารณะ [10]

วัตถุประสงค์ของการวิจัย

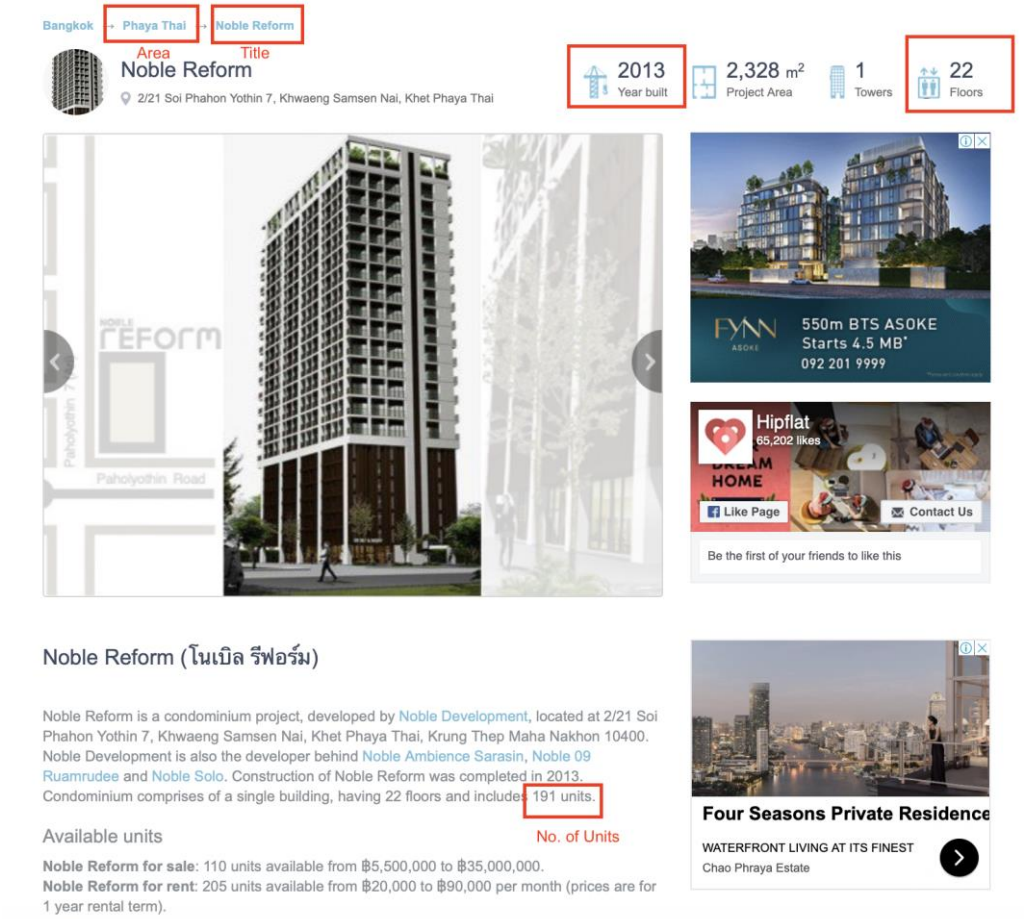
การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลในการพยากรณ์ราคาคอนโดมิเนียมโดยใช้เทคนิค Machine Learning สำหรับใช้ในการประมาณการมูลค่าราคาคอนโดมิเนียมในอนาคตประกอบการตัดสินใจซื้อคอนโดมิเนียมของผู้ซื้อ โดยงานวิจัยนี้ประกอบด้วยขั้นตอนต่าง ๆ เริ่มตั้งแต่การรวบรวมข้อมูลด้วยเทคนิค Web Crawling การสำรวจข้อมูล การเตรียมข้อมูล และการสร้างแบบจำลอง และการตรวจสอบความถูกต้อง

วิธีดำเนินการวิจัย

การรวบรวมข้อมูลโดยเทคนิค Web Crawling

เทคนิค Web Crawling เป็นเทคนิคในการสกัดข้อมูลจากเว็บไซต์โดยมุ่งให้ความสำคัญกับการจัดลำดับผลลัพธ์ที่ได้รับจากการสกัดข้อมูล เพื่อประโยชน์ในการจัดเตรียมข้อมูลที่เหมาะสมต่อการเรียนรู้ของโมเดล [11] มีรายงานวิจัยจำนวนหนึ่งได้ใช้เทคนิค Web Crawling ในการสกัดข้อมูลเพื่อการวิเคราะห์ไม่ว่าจะเป็นการใช้ข้อมูลนักท่องเที่ยวไทยเป็นกลุ่มตัวอย่างในการทดลองจากเว็บไซต์ ODP.org [12] หรืองานวิจัยของ Soonthornpipt, 2015 ซึ่งนำเทคนิค Crawling มาใช้สกัดข้อมูลการจำหน่ายบุหรี่ในประเทศไทยผ่านสังคมเครือข่าย อาทิ Facebook และ Twitter โดยมุ่งลดความซ้ำซ้อนของเว็บไซต์ที่มีข้อมูลคล้ายคลึงกัน [13]

ในการการศึกษาในครั้งนี้ ข้อมูลคอนโดมิเนียมได้ถูกรวบรวมจากเว็บไซต์ hipflat.co.th ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมข้อมูลที่เป็นปัจจัยในการประกอบการตัดสินใจซื้ออสังหาริมทรัพย์ที่เป็นประโยชน์ต่อการตัดสินใจซื้อของผู้ซื้ออย่างครบถ้วนมากที่สุดเมื่อเปรียบเทียบกับเว็บไซต์เกี่ยวกับการซื้ออสังหาริมทรัพย์อื่น ๆ ในประเทศไทย โดยใช้เทคนิค Web Crawling โดยเว็บไซต์ดังกล่าวมีข้อมูลเกี่ยวกับโครงการคอนโดมิเนียมแต่ละแห่งไม่ว่าจะเป็นข้อมูลที่เกี่ยวข้องกับทำเลที่ตั้ง ลักษณะของโครงการ และราคา ดังแสดงในภาพที่ 1 และการรวบรวมข้อมูลด้วยวิธี Web Crawling ได้ดำเนินการโดยการเขียนสคริปต์ XPath Query บนเครื่องมือ RapidMiner ดังแสดงในภาพที่ 2



ภาพที่ 1 ข้อมูลคอนโดมิเนียมจากเว็บไซต์ Hipflat.com

attribute name	query expression
Title	substring-before(//*[@class='property-description']/h:html/h:body/h:header/h:h1/text(),' (')
Area	//*[@property='position' and @content='2']/../h:li[1]/h:a/h:span/text()
Year Built	//*[@class='project-header-year']/h:div[1]/h:span/text()
No. of Floors	//*[@class='project-header-floor']/h:div[1]/h:span/text()
No. of Units	substring-after(//*[@class='property-description']/h:html/h:body/h:p, ' includes')
Nearest BTS	//*[@class='marker marker-bts']/../../*[@class='media-body']/h:h4/text()
Distance to Nearest BTS	substring-before(//*[@class='marker marker-bts']/../../*[@class='media-body']/h:small/text(),'/')
Nearest MRT	//*[@class='marker marker-mrt']/../../*[@class='media-body']/h:h4/text()
Distance to Nearest MRT	substring-before(//*[@class='marker marker-mrt']/../../*[@class='media-body']/h:small/text(),'/')
Price per sqm	//*[@class='market-data']/*[@class='indicator']/*[@class='indicator__value']/*[@class='indicator__amount']/*[@class='money']/*[@class='number']/text()
Price change from Previous Quarter	//*[@class='market-data']/*[@class='indicator'][2]/*[@class='indicator__value']/*[@class='indicator__amount']/text()
Price change from Last Year	//*[@class='market-data']/*[@class='indicator'][3]/*[@class='indicator__value']/*[@class='indicator__amount']/text()
Rental change from Last year	//*[@class='market-data']/*[@class='indicator'][5]/*[@class='indicator__value']/*[@class='indicator__amount']/text()
Yield	//*[@class='market-data']/*[@class='indicator'][4]/*[@class='indicator__value']/*[@class='indicator__amount']/text()
Developer	//*[@class='property-description']/h:html/h:body/h:p/h:a/text()

ภาพที่ 2 XPath Query สำหรับการสกัดข้อมูลคอนโดมิเนียม

ผลที่ได้จากการทำ Web Crawling จะได้ข้อมูลของคอนโดมิเนียม จำนวน 1,465 แห่ง ดังแสดงในภาพที่ 3 โดยจะเห็นได้ว่าข้อมูลมีความไม่สมบูรณ์โดยเฉพาะข้อมูลที่ขาดหาย (Missing Attribute) โดยข้อมูลที่ขาดหายไปมีสาเหตุมาจากเว็บไซต์ hipflat.co.th ไม่มีข้อมูลดังกล่าว ยกตัวอย่างเช่น ในบางคอนโดมิเนียมที่ไม่ได้อยู่ในบริเวณที่มีรถไฟฟ้า BTS หรือรถไฟฟ้าใต้ดินผ่านก็จะไม่มีข้อมูลดังกล่าว อย่างไรก็ดีข้อมูลที่ขาดหายนี้จะถูกแก้ไขในขั้นตอนของการเตรียมข้อมูลที่จะอธิบายในส่วนถัดไป

Row No.	Title	No. of Floors	Price per s...	Price chang...	Developer	Price chang...	Rental cha...	Nearest MRT
1	Premier Condominium	12	111391	-11.680	?	5.300	43.090	Queen Sirikit...
2	Tai Ping Towers	33	68921	-4.990	Thailand Ca...	33.760	-1.720	?
3	Siam Penthouse 1	10	79762	-6.630	?	3.330	-9.170	Sukhumvit
4	Pradipat Condominium	5	131662	0	?	7.650	0	Bang Sue
5	Grandville House Co...	13	126648	2.870	?	22.070	-23.480	Queen Sirikit...
6	Grand Ville House 1	13	113125	2.670	?	3.100	-3.270	Queen Sirikit...
7	The Heritage	14	86007	21.390	?	25.520	0.340	Sukhumvit
8	Bangrak Condominium	11	54781	0	?	0	0	?
9	Nana Condo	8	62188	0	?	0	256.230	?
10	Saphan Kwai Condo...	23	145161	0	Pruksa Real ...	0	0	Kamphaeng...
11	The Garden Place	6	60748	8.890	?	64.970	32.020	?
12	Hua Mark Condomini...	8	40137	0	?	0	6.610	Huai Khwang
13	Pinklao Condominium	13	57122	0	Sansiri	0	-23.080	Hua Lamph...
14	Indra Condominium	5	61437	0	?	0	0	?
15	Ruenrudee Condomi...	17	68096	2.550	?	-18.280	-4.020	?

ภาพที่ 3 ข้อมูลที่ได้หลังจากทำ Web Crawling

ในการพยากรณ์ราคาคอนโดมิเนียมซึ่งอาศัยปัจจัยในการสนับสนุนการตัดสินใจซื้อจากเว็บไซต์ Hipflat.com รายงานวิจัยฉบับนี้ได้กำหนด 15 แอตทริบิวต์ เพื่อใช้ในการพยากรณ์โดยสามารถถูกแบ่งออกได้เป็น 3 กลุ่ม ดังนี้

กลุ่มที่ 1 ปัจจัยด้านทำเลที่ตั้ง ซึ่งประกอบไปด้วยพื้นที่ที่คอนโดมิเนียมนั้นตั้งอยู่ ระยะทางจากสถานีรถไฟลอยฟ้า และระยะทางจากสถานีรถไฟฟ้าใต้ดิน โดยแต่ละปัจจัยมีรายละเอียดดังต่อไปนี้

1. พื้นที่ตั้งของโครงการ (Area) โดยเขตที่มีคอนโดมิเนียมมากที่สุด คือ เขตวัฒนา จำนวน 234 แห่ง

2. สถานีรถไฟฟ้า BTS ที่ใกล้ที่สุด (Nearest BTS) โดยสถานีรถไฟฟ้า BTS พร้อมพงศ์ มีคอนโดมิเนียมที่อยู่ใกล้สถานีดังกล่าวมากที่สุด จำนวน 89 แห่ง

3. สถานีรถไฟฟ้าใต้ดินที่ใกล้ที่สุด (Nearest MRT) โดยสถานีรถไฟฟ้า MRT สุขุมวิท มีคอนโดมิเนียมที่อยู่ใกล้สถานีดังกล่าวมากที่สุด จำนวน 95 แห่ง

4. ระยะทางจากสถานีรถไฟฟ้า BTS (Distance to Nearest BTS) ระยะทางโดยเฉลี่ยจากคอนโดมิเนียมไปสู่สถานีรถไฟฟ้า BTS ประมาณ 2 กิโลเมตร

5. ระยะทางจากสถานีรถไฟฟ้าใต้ดิน (Distance to Nearest MRT) โดยมีระยะทางโดยเฉลี่ย 2.9 กิโลเมตร จากคอนโดมิเนียมไปสู่สถานีรถไฟฟ้าใต้ดิน

กลุ่มที่ 2 ปัจจัยด้านลักษณะของอาคาร ประกอบไปด้วย 5 แอตทริบิวต์ ได้แก่

1. ชื่อของโครงการคอนโดมิเนียม (Title)

2. จำนวนชั้นของคอนโดมิเนียม (No. of Floors) โดยจำนวนชั้นโดยเฉลี่ยอยู่ที่ 18 ชั้น จำนวนต่ำสุดอยู่ที่ 3 ชั้น และสูงสุดอยู่ที่ 77 ชั้น คอนโดมิเนียมส่วนใหญ่สร้างที่ 8 ชั้น เป็นไปตามกฎหมายในการก่อสร้างคอนโดมิเนียม โดยคอนโดมิเนียมที่มีจำนวนชั้นไม่มากจะมีกฎหมายในการก่อสร้างเกี่ยวกับการใช้ที่ดินน้อยกว่าการสร้างอาคารสูง

3. จำนวนยูนิต (No. of Units) โดยจำนวนยูนิตโดยเฉลี่ยของแต่ละโครงการ คือ 388 ห้อง ซึ่งจำนวนต่ำสุดอยู่ที่ 1 ยูนิต และจำนวนสูงสุดอยู่ที่ 3,500 ยูนิต

4. ปีที่สร้าง (Year Built) ระบุปี ค.ศ. ที่คอนโดมิเนียมถูกสร้าง โดยคอนโดมิเนียมที่เก่าแก่ที่สุดถูกสร้างในปี ค.ศ. 1981 และคอนโดมิเนียมที่ใหม่ที่สุดถูกสร้างในปี ค.ศ. 2020

5. ผู้พัฒนาโครงการ (Developer) เป็นการระบุชื่อบริษัทพัฒนาอสังหาริมทรัพย์ อย่างไรก็ดีมีข้อมูลที่ขาดหายไปประมาณ 650 โครงการ ที่ไม่ได้ถูกระบุชื่อผู้พัฒนาโครงการจากเว็บไซต์ https://www.hipflat.co.th

กลุ่มที่ 3 ปัจจัยความเคลื่อนไหวของราคา

1. ราคาต่อตารางเมตร (Price Per sqm) ระบุราคาโดยเฉลี่ยต่อตารางเมตรของคอนโมมิเนียมแต่ละแห่ง โดยแอตทริบิวต์ดังกล่าวเป็นตัวแปรเป้าหมายที่สามารถใช้เพื่อการพยากรณ์ราคาคอนโดมิเนียมได้ ราคาโดยเฉลี่ยต่อตารางเมตรสำหรับคอนโดมิเนียมทั้ง 1,465 แห่ง คือ 105,000 บาท

2. อัตราราคาที่เปลี่ยนแปลงไปจากสิ้นปีที่ผ่านมา (Price Change from Last Year) ระบุร้อยละของการเปลี่ยนแปลงราคาคอนโดมิเนียมโดยเฉลี่ยต่อตารางเมตรนับจากสิ้นปีที่ผ่านมา โดยการเปลี่ยนแปลงโดยเฉลี่ยอยู่ที่ร้อยละ 9.3

3. อัตราราคาที่เปลี่ยนแปลงไปจากสิ้นไตรมาสที่ผ่านมา (Price Change from Previous Quarter) ระบุร้อยละของการเปลี่ยนแปลงราคาคอนโดมิเนียมโดยเฉลี่ยต่อตารางเมตรนับจากสิ้นไตรมาสที่ผ่านมา โดยมีการเปลี่ยนแปลงโดยเฉลี่ยอยู่ที่ร้อยละ 2.6

4. อัตราค่าเช่าคอนโดมิเนียมที่เปลี่ยนแปลงไปจากสิ้นปีที่ผ่านมา (Rental Change from Last Year) ระบุอัตราการเปลี่ยนแปลงจากสิ้นปีที่ผ่านมาอยู่ที่ร้อยละ 38

5. ผลตอบแทนการลงทุน (Yield) ระบุผลตอบแทนต่อปีหากลงทุนในคอนโดมิเนียมดังกล่าว โดยอัตราผลตอบแทนเฉลี่ยของคอนโดมิเนียมในชุดข้อมูลนี้อยู่ที่ร้อยละ 7

การสำรวจข้อมูล (Data Exploration)

รายงานวิจัยฉบับนี้ทำการศึกษาประสิทธิภาพในการสนับสนุนการพยากรณ์ราคาคอนโดมิเนียมของแต่ละแอตทริบิวต์ผ่านขั้นตอนการสำรวจข้อมูล และพบว่าแต่ละแอตทริบิวต์มีระดับนัยสำคัญในการพยากรณ์ผลลัพธ์ที่แตกต่างกัน ยกตัวอย่างจากการสำรวจข้อมูล เช่น ราคาขายคอนโดมิเนียมที่สูงที่สุดอยู่ในเขตชิดลม ซึ่งอยู่ใกล้สถานีรถไฟลอยฟ้าชิดลมและรถไฟฟ้าใต้ดินสีลม ดังนั้นระยะทางจากคอมโดมิเนียมไปสู่สถานีรถไฟลอยฟ้าและสถานีรถไฟฟ้าใต้ดินน่าจะส่งผลต่อการตั้งราคาจำหน่ายคอนโดมิเนียม

นอกเหนือไปจากทำเลที่ตั้งของคอนโดมิเนียมแล้ว ยังพบว่าคอนโดมิเนียมสร้างใหม่มีราคาสูงกว่าคอนโดมิเนียมที่สร้างมานาน เช่นเดียวกันกับที่คอนโดมิเนียมที่มีจำนวนชั้นมากกว่า (High Rise) มักจะมีราคาสูงกว่าคอนโดมิเนียมที่มีจำนวนชั้นน้อย (Low Rise) นอกจากนี้ คอนโดมิเนียมที่มีจำนวนยูนิตที่มากกว่า มักจะมีราคาต่อตารางเมตรที่ต่ำกว่าคอนโดมิเนียมที่มีจำนวนยูนิตน้อยกว่า และราคาจำหน่ายโดยเฉลี่ยของคอนโดมิเนียมที่สร้างโดยบริษัทที่อยู่ในตลาดหลักทรัพย์มีราคาเฉลี่ยต่อตารางเมตรสูงกว่าคอนโดมิเนียมที่สร้างโดยบริษัทที่อยู่นอกตลาดหลักทรัพย์

ส่วนสุดท้ายของการสำรวจข้อมูลเพื่อศึกษาประสิทธิภาพในการสนับสนุนการพยากรณ์ของแต่ละแอตทริบิวต์ คือ การวิเคราะห์สหสัมพันธ์ (Corelation Analysis) ของแอตทริบิวต์ของข้อมูลทั้งหมด จากภาพที่ 4 พบว่า ระยะทางจากสถานีรถไฟลอยฟ้า (Distance to Nearest BTS) และสถานีรถไฟฟ้าใต้ดิน (Distance to Nearest MRT) รวมถึงอัตราผลตอบแทน (Yield) และอัตราการเปลี่ยนแปลงของค่าเช่าจากสิ้นปี (Rental Change from Last Year) มีนัยสำคัญต่อการพยากรณ์ข้อมูลสูง โดยราคาคอนโดมิเนียมต่อตารางเมตรมีนัยสำคัญเชิงบวกต่อจำนวนชั้นของคอนโดมิเนียมและปีที่ทำการก่อสร้าง อย่างไรก็ตามกลับมีนัยสำคัญเชิงลบกับระยะทางจากสถานีรถไฟลอยฟ้าและสถานีรถไฟฟ้าใต้ดิน ข้อมูลเหล่านี้สามารถถูกนำมาใช้เพื่อการปรับปรุงความแม่นยำของแบบจำลองในการวิเคราะห์ข้อมูลได้

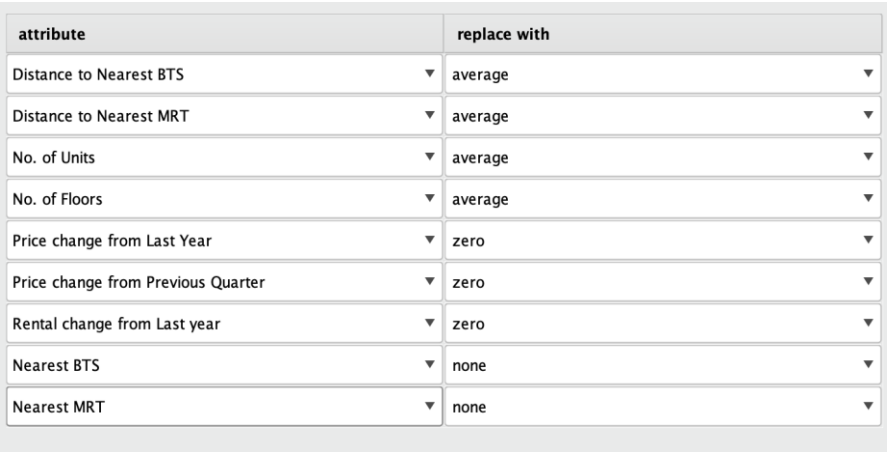
Attributes	No. of ...	Price p...	Price c...	Price c...	Rental ...	Distan...	Year B...	Distan...	Yield	No. of ...
No. of Floors	1	0.460	-0.008	-0.009	-0.043	-0.261	0.153	-0.154	-0.050	0.217
Price per sqm	0.460	1	0.095	0.103	-0.040	-0.387	0.401	-0.404	-0.056	-0.152
Price change from Previous Quarter	-0.008	0.095	1	0.244	-0.003	-0.000	0.029	0.030	-0.009	0.004
Price change from Last Year	-0.009	0.103	0.244	1	-0.004	-0.012	0.024	-0.019	-0.010	0.023
Rental change from Last year	-0.043	-0.040	-0.003	-0.004	1	-0.017	0.034	0.049	0.997	-0.011
Distance to Nearest MRT	-0.261	-0.387	-0.000	-0.012	-0.017	1	-0.049	0.925	-0.006	0.297
Year Built	0.153	0.401	0.029	0.024	0.034	-0.049	1	-0.046	0.033	0.177
Distance to Nearest BTS	-0.154	-0.404	0.030	-0.019	0.049	0.925	-0.046	1	0.055	0.346
Yield	-0.050	-0.056	-0.009	-0.010	0.997	-0.006	0.033	0.055	1	-0.008
No. of Units	0.217	-0.152	0.004	0.023	-0.011	0.297	0.177	0.346	-0.008	1

ภาพที่ 4 สัมประสิทธิ์สหสัมพันธ์ของแต่ละแอตทริบิวต์

การจัดเตรียมข้อมูล (Data Preparation)

จากภาพที่ 3 จะเห็นได้ว่าข้อมูลซึ่งถูกรวบรวมจาก https://www.hipflat.com ด้วยเทคนิค Crawling ดังที่ได้อธิบายไปแล้วนั้นพบข้อมูลที่ไม่สมบูรณ์อยู่จำนวนหนึ่ง (Missing Value) เช่น ข้อมูลระยะทางจากสถานีรถไฟฟ้า BTS และสถานีรถไฟใต้ดินไปยังคอนโดมิเนียมตลอดจนจำนวนห้องและจำนวนชั้นของคอนโดมิเนียม ดังนั้นผู้วิจัยจึงได้ทำการปรับปรุงข้อมูลของแอตทริบิวต์ จำนวน 9 แอตทริบิวต์ ซึ่งเป็นแอตทริบิวต์ที่พบข้อมูลซึ่งขาดความ

สมบูรณ์ ดังแสดงในภาพที่ 5 โดยระยะทางจากสถานีรถไฟลอยฟ้าและสถานีรถไฟใต้ดิน รวมไปถึงจำนวนยูนิตและจำนวนชั้นของโครงการคอนมิเนียมถูกแทนด้วยค่าเฉลี่ยของข้อมูล ในขณะที่ข้อมูลการเปลี่ยนแปลงของราคาจากไตรมาสที่ผ่านมาและการเปลี่ยนแปลงของค่าเช่าจากปีที่ผ่านมาถูกแทนด้วย 0 ซึ่งเป็นข้อมูลที่ขาดความสมบูรณ์เช่นเดียวกัน อย่างไรก็ดีในข้อมูลอื่น ๆ ประกอบไปด้วย รายชื่อสถานีรถไฟฟ้าใต้ดินและรายชื่อสถานีรถไฟลอยฟ้าที่ใกล้ที่สุดยังคงให้ข้อมูลมีลักษณะที่ไม่สมบูรณ์ไว้เช่นเดิม โดยรายงานวิจัยฉบับนี้ใช้ซอฟต์แวร์ RapidMiner ในการพัฒนาและทดสอบโมเดล โดยมีฟังก์ชันที่ช่วยในการเปลี่ยนค่าข้อมูลที่ไม่สมบูรณ์ได้อย่างสะดวก และเมื่อดำเนินการแล้วข้อมูลที่ได้ถูกนำมาสร้างแบบจำลองในส่วนถัดไป



attribute	replace with
Distance to Nearest BTS	average
Distance to Nearest MRT	average
No. of Units	average
No. of Floors	average
Price change from Last Year	zero
Price change from Previous Quarter	zero
Rental change from Last year	zero
Nearest BTS	none
Nearest MRT	none

ภาพที่ 5 การแทนที่ข้อมูลที่ขาดความสมบูรณ์

ผลการวิจัย

การสร้างแบบจำลอง (Data Modeling)

ข้อมูลที่ถูกจัดเตรียมจากกระบวนการ Data Preparation ถูกนำไปประมวลผลด้วยอัลกอริทึมทางวิทยาศาสตร์ข้อมูลถึง 5 รูปแบบ ดังแสดงในภาพที่ 6 โดยทำการนำข้อมูลเข้าสู่ซอฟต์แวร์ RapidMiner และทำการพัฒนาโมเดลที่ใช้ในการพยากรณ์ราคาคอนโดมิเนียม จำนวน 5 รูปแบบ ประกอบด้วย Generalized Linear Model, Deep Learning, Decision Tree, Random Forest และ Gradient Boosted Trees ตามลำดับ เพื่อทำการทดสอบความแม่นยำของโมเดลแต่ละรูปแบบที่ใช้ในการพยากรณ์ราคาคอนโดมิเนียม

เพื่อที่จะให้โมเดลเรียนรู้ลักษณะของข้อมูล (Train) และสร้างความสามารถในการพยากรณ์ราคาคอนโดมิเนียมต่อตารางเมตร รายงานวิจัยฉบับนี้ใช้เทคนิค Split Test ที่มีการสุ่มข้อมูลเพียงครั้งเดียวซึ่งเหมาะสมกับชุดข้อมูลขนาดใหญ่ โดยชุดข้อมูลได้ถูกแบ่งออกเป็นชุดข้อมูลสำหรับการสอนให้โมเดลเกิดการเรียนรู้ (Training Set) และชุดข้อมูลสำหรับการทดสอบประสิทธิภาพของโมเดล (Testing Set) ในสัดส่วน 80 : 20 โดยข้อมูลทั้ง 2 ส่วน มีลักษณะคล้ายคลึงกันอันเป็นการส่งเสริมประสิทธิภาพที่ดีของการพยากรณ์ราคาคอนโดมิเนียม ประสิทธิภาพของโมเดลทุกรูปแบบได้ถูกอธิบายไว้ในภาพที่ 6 โดยแสดงให้เห็นว่า Deep Learning เป็นโมเดลที่มีประสิทธิภาพสูงสุด ซึ่งมีค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Least Root Mean Square Error: RMSE) อยู่ที่ 37,652

Model	Root Mean Squared Error	Runtime
Generalized Linear Model	63232.472	513 ms
Deep Learning	37652.143	3 s
Decision Tree	54649.960	2 s
Random Forest	44202.914	24 s
Gradient Boosted Trees	41850.678	25 s

ภาพที่ 6 ประสิทธิภาพของโมเดลแต่ละแบบ

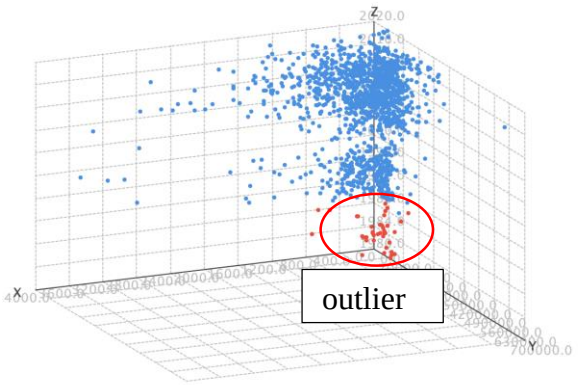
หลังจากการทดสอบประสิทธิภาพของโมเดลในการพยากรณ์ราคาคอนโดมิเนียมทั้ง 5 รูปแบบ รายงานวิจัยฉบับนี้ได้พบว่า Deep Learning เป็นโมเดลที่มีความแม่นยำในการพยากรณ์สูงสุด และเพื่อที่จะทำให้โมเดลเกิดประสิทธิภาพในการประมวลผลมากขึ้นก่อนที่ผู้ซื้อคอนโดมิเนียมจะนำโมเดลดังกล่าวไปใช้ในการพยากรณ์ราคาคอนโดมิเนียมในอนาคตเพื่อสนับสนุนการตัดสินใจซื้อ ขั้นตอนในการปรับแต่งให้โมเดลมีประสิทธิภาพในการพยากรณ์ที่สูงขึ้นจึงมีความสำคัญต่อการพัฒนาโมเดล ดังนั้นรายงานวิจัยฉบับนี้จึงได้ปรับแต่งโมเดลเพิ่มเติม จำนวน 3 กระบวนการ ดังนี้

1. แอตทริบิวต์ระยะทางจากคอนโดมิเนียมไปยังสถานีรถไฟฟ้า MRT และสถานีรถไฟฟ้า BTS ได้ถูกแบ่งออกเป็นเป็น 4 คลาส (Binning) โดยใช้ Discretize Process ในซอฟต์แวร์ RapidMiner ดังแสดงในภาพที่ 7 หลังจากนั้น โมเดล Deep Learning ได้ถูกสอนให้เกิดการเรียนรู้และทดสอบประสิทธิภาพอีกครั้ง อย่างไรก็ดีค่า RMSE กลับเพิ่มขึ้นเป็น 39,824 ซึ่งมีสาเหตุจากการที่ราคาค่าเฉลี่ยต่อตารางเมตรของคอนโดมิเนียมมีความอ่อนไหวต่อการเปลี่ยนแปลงของระยะทางจากสถานีรถไฟลอยฟ้าและสถานีรถไฟใต้ดิน

class names	upper limit
<500m	500.0
500-1000m	1000.0
1km-2km	2000.0
>2km	50000.0

ภาพที่ 7 การแบ่ง 2 แอททริบิวต์ออกเป็น 4 คลาส (Binning)

2. แบบจำลองค่าเฉลี่ยราคาต่อตารางเมตรยังได้ถูกปรับแต่งอีกครั้งด้วยการคัดกรอง (Filtering) ข้อมูลที่มีค่าผิดปกติ (Outlier) โดยใช้ Distance Measure ในซอฟต์แวร์ RapidMiner ข้อมูลที่มีค่าผิดปกติและได้ถูกคัดกรองออกไปซึ่งส่วนใหญ่เป็นคอนโดมิเนียมที่ถูกสร้างในปี ค.ศ. 1980 หลังการคัดกรองคอนโดมิเนียมจำนวน 30 โครงการออกไป แบบจำลอง Deep Learning มีประสิทธิภาพสูงขึ้นอย่างมีนัยสำคัญ โดย RMSE ลดลงสู่ 29,271 ด้วย Relative Error ที่ร้อยละ 17.53 ดังแสดงในภาพที่ 8

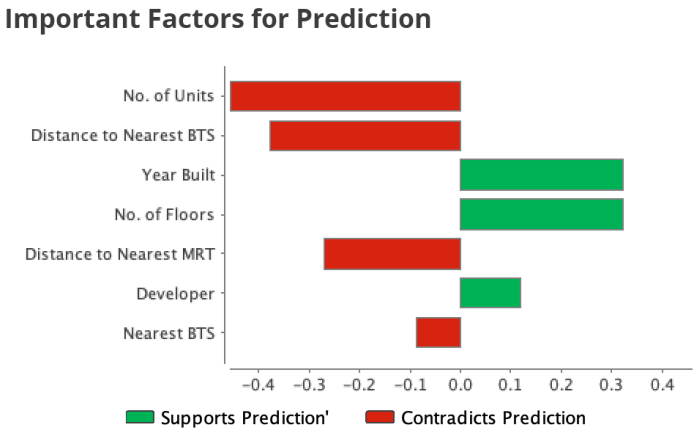


No. of Outlier	RMSE
20	33587
30	29271
40	29883

ภาพที่ 8 การตรวจสอบค่าผิดปกติและประสิทธิภาพของโมเดล

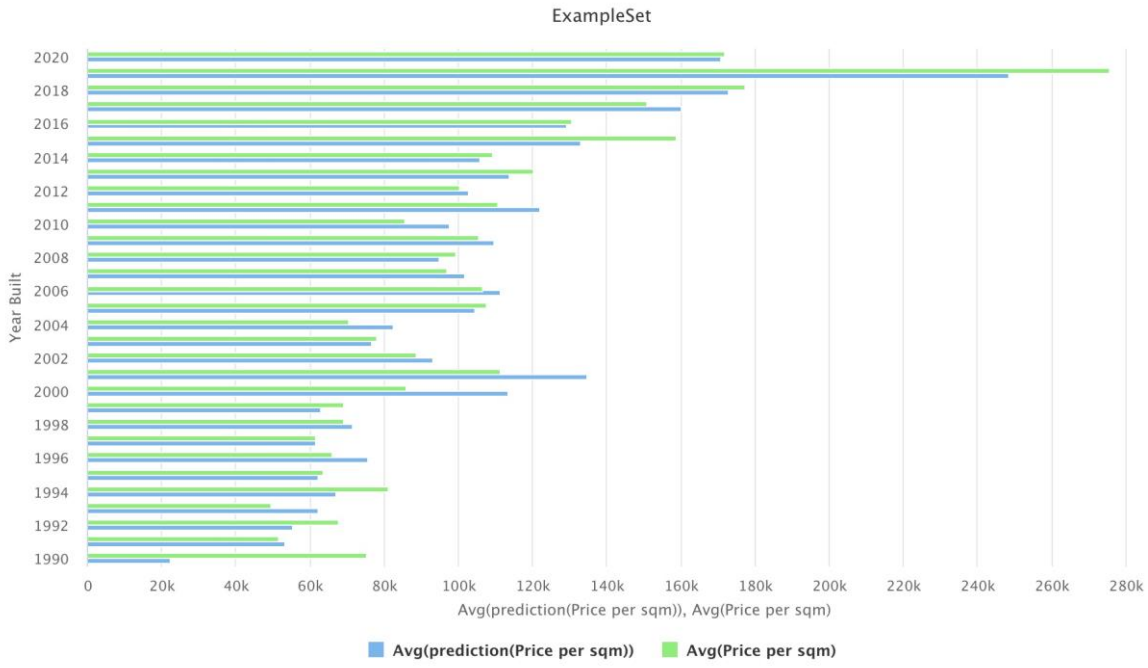
3. ในขั้นตอนสุดท้าย แอตทริบิวต์ในกลุ่มที่เป็นปัจจัยความเคลื่อนไหวของราคา ซึ่งประกอบไปด้วย ข้อมูลผลตอบแทน (Yield) อัตราการเปลี่ยนแปลงราคาจากสิ้นปี (Price Change from Last Year) อัตราการเปลี่ยนแปลงราคาจากไตรมาสสุดท้าย (Price Change from Last Quarter) และอัตราการเปลี่ยนแปลงค่าเช่าคอนโดมิเนียมจากสิ้นปี (Rental Change) ได้ถูกนำไปรวมไว้กับชุดข้อมูลที่ใช้ในการสอนให้โมเดลเกิดการเรียนรู้ โดยหลังจากดำเนินการดังกล่าว ประสิทธิภาพโมเดลมีการเพิ่มขึ้นเล็กน้อย โดยมีค่า RMSE อยู่ที่ 28,907 ที่ลดลงในขณะที่ Relative Error อยู่ที่ร้อยละ 17.66 อย่างไรก็ตามการใช้ปัจจัยความเคลื่อนไหวของราคามาใช้ในการพยากรณ์ผลลัพธ์นั้นอาจไม่สมเหตุสมผลนัก เนื่องจากข้อมูลดังกล่าวมักจะไม่ปรากฏหรือหาได้ยาก ดังนั้นจากผลลัพธ์ในการวิจัย ปัจจัยที่ส่งผลต่อประสิทธิภาพในการพยากรณ์ของโมเดลมีเพียงปัจจัยด้านทำเลที่ตั้ง ได้แก่ ระยะทางจากคอนโดมิเนียมไปยังสถานีรถไฟฟ้า BTS และ MRT และลักษณะของอาคาร ได้แก่ จำนวนชั้นและปีที่ก่อสร้าง

จากการเรียนรู้ของโมเดล Deep Learning ที่ได้จากการวิจัยนี้พบว่า ตัวแปรที่ส่งผลเชิงบวกต่อราคาเฉลี่ยต่อตารางเมตรของคอนโดมิเนียมในกรุงเทพมหานคร ได้แก่ ปีที่ก่อสร้าง (คอนโดมิเนียมที่ใหม่กว่าจะมีราคาต่อตารางเมตรที่สูงกว่า) จำนวนชั้น (คอนโดมิเนียมแบบ High Rise จะมีราคาต่อตารางเมตรที่สูงกว่า) และคอนโดมิเนียมที่สร้างโดยผู้ประกอบการในตลาดหลักทรัพย์จะมีราคาต่อตารางเมตรที่สูงกว่า ในขณะที่ปัจจัยที่ส่งผลกระทบเชิงลบ ได้แก่ จำนวนยูนิต (ยิ่งจำนวนยูนิตมาก ราคาต่อตารางเมตรยิ่งต่ำลง) และระยะทางจากคอนโดมิเนียมไปยังสถานีรถไฟฟ้า BTS และ MRT (ยิ่งห่างจากสถานี BTS หรือ MRT ราคาต่อตารางเมตรยิ่งต่ำลง) ดังแสดงในภาพที่ 9



ภาพที่ 9 ปัจจัยที่ส่งผลต่อราคาคอนโดมิเนียมในกรุงเทพมหานคร

เพื่อให้เห็นถึงความสามารถในการทำนายของแบบจำลองที่พัฒนาขึ้น ภาพที่ 10 แสดงการเปรียบเทียบราคาต่อตารางเมตรของราคาขายจริงและราคาที่พยากรณ์ได้ด้วยโมเดล Deep Learning ที่ใช้ข้อมูลปัจจัยของโครงการและปัจจัยของอาคาร จากรูปจะเห็นได้ว่าราคาต่อตารางเมตรจากการพยากรณ์มีความใกล้เคียงกับราคาขายจริง ยกเว้นในกรณีของคอนโดมิเนียมที่สร้างในปี ค.ศ. 1990 ที่โมเดลทำนายราคาต่ำกว่าราคาจริงค่อนข้างมาก ทั้งนี้เนื่องจากโมเดลได้นำข้อมูลของคอนโดมิเนียมที่สร้างมานานแล้วออกจาก Training Dataset เนื่องจากเป็นข้อมูล Outlier ตามที่ได้กล่าวมาในหัวข้อที่แล้ว



ภาพที่ 10 เปรียบเทียบราคาที่พยากรณ์และราคาขายจริง

สรุปและอภิปรายผล

บทความวิจัยฉบับนี้ได้กล่าวถึงการพัฒนาโมเดลที่ใช้สำหรับพยากรณ์ราคาคอนโดมิเนียมต่อตารางเมตรในกรุงเทพมหานครเพื่อให้ผู้ซื้อได้นำโมเดลดังกล่าวไปประมาณมูลค่าของคอนโดมิเนียมในอนาคตก่อนที่จะทำการตัดสินใจซื้อ โดยรายงานวิจัยฉบับนี้ได้สกัดข้อมูลที่ใช้ในการพัฒนาและทดสอบโมเดลจากเว็บไซต์ hipflat.co.th โดยใช้เทคนิค Crawling และ Mining จากนั้นชุดข้อมูลได้ถูกจัดเตรียมและพัฒนาโมเดลในการพยากรณ์ผ่านการทดสอบด้วยโมเดลลักษณะ Supervised Learning จำนวน 5 รูปแบบ และได้พบว่าโมเดล Deep Learning มีประสิทธิภาพสูงสุดที่สุด จากนั้นรายงานวิจัยฉบับนี้จึงทำการปรับแต่โมเดล Deep Learning เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ผลของโมเดลให้สูงขึ้นหลังจากนั้นโมเดลจึงมีประสิทธิภาพในการพยากรณ์ราคาคอนโดมิเนียมที่สูงขึ้นโดยหลังจากการทดสอบการพยากรณ์ของโมเดลรายงานวิจัยฉบับนี้พบว่า ราคาจริงคอนโดมิเนียมในกรุงเทพมหานครมีความใกล้เคียงกับผลลัพธ์ที่ได้จากการพยากรณ์ผ่านโมเดลอย่างมีนัยสำคัญ

อีกทั้งรายงานวิจัยฉบับนี้ยังพบปัจจัยที่ใช้ในการสนับสนุนการตัดสินใจซื้อของผู้ซื้อและปัจจัยดังกล่าวยังสามารถนำมาใช้เป็นแอตทริบิวต์ในการพยากรณ์ราคาคอนโดยมิเนียมได้ โดยพยากรณ์จากจำนวนห้องในโครงการคอนโดมิเนียมและระยะทางจากสถานีรถไฟลอยฟ้าและปีที่ทำการก่อสร้าง โดยสอดคล้องกับรายงานวิจัยของ Diewert, W. E., & Shimizu, C. ในปี ค.ศ. 2016 ซึ่งกล่าวว่า จำนวนห้องในโครงการคอนโดมิเนียมและทำเลที่ตั้งส่งผลต่อราคาจำหน่ายคอนโดมิเนียมในเมืองโตเกียวประเทศญี่ปุ่น [14] อีกทั้งบทความวิจัยฉบับนี้ยังได้รับผลลัพธ์จากการวิจัยสอดคล้องกับรายงานวิจัยชื่อ Housing Price Prediction Using Neural Networks ซึ่งหนึ่งในปัจจัยที่ถูกนำมาใช้ในการพยากรณ์ราคาคอนโดมิเนียมในประเทศสิงค์โปร์ คือ จำนวนห้องในโครงการคอนโดมิเนียมและระยะทางจากสถานีรถไฟลอยฟ้า [15]

เอกสารอ้างอิง

[1] Choosun, T. (2014). *Factors Affecting Condominium Buying Decisions of Consumer.* Graduate School, Bangkok University.

[2] Ratpitaksunti, A. (2011). *Condominium Purchasing Demand of Generation Y*. Retrieved from https://positioningmag.com/54642

[3] Bangkok Post. (2018). *Condo market on a rollercoaster*. Retrieved from https://property.bangkokpost.com/news/1417998/condo-market-on-a-rollercoaster

[4] Knightfrank. (2018). *Bangkok Condominium Market Report Q3 2018*. Retrieved from https://content.knightfrank.com/research/1654/documents/en/bangkok-condo-market-thailand-q3-2018-q3-2018-5948.pdf

[5] Chrostek, K., & Kopczewska, K. (2013). Spatial Prediction Models for Real Estate Market Analysis. *Ekonomia Journal, 35*, 25-43.

[6] Lehner, M. (2017). *Modelling Housing Prices in Singapore Applying Spatial Hedonic Regression*. Master of Science Thesis, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

[7] Chotiwut Laophairoj. (2012). *Factors Affecting Condominium Price in Bangkok*. Thesis, M.Econ. (Economics). Bangkok: Faculty of Economics, Thammasat University.

[8] Lorlertsakool, M., & Sakworawit, A. (2017). Spatial Regression Model Predicting Condominium Price per Square Meter in Bangkok. *Academic Journal of Architecture*, *66*(1), 121-138.

[9] Nutsarapohn Nutsarabut. (2017). *Application of the Pricing Model for Appropriate Pricing in the Development of Luxury Class Condominium Project in Bangkok.* Thesis, M. Arch. (Architecture). Bangkok: Faculty of Architecture and Planning, Thammasat University.

[10] Sasithorn Kaluntakasuwan. (2017). *Factors Influencing Luxury and Super-Luxury Condominium Prices in Bangkok.* Independent Study, M.Econ. (Economics). Bangkok: Faculty of Economics, Thammasat University.

[11] Kollayuth Borpit. (2010). *An Algorithm of Product Information Extraction on Web Pages for Web Crawler in Product Search Engines.* Thesis, M.S. (Web Engineering). Bangkok: Faculty of Information Technology, Dhurakij Pundit University.

[12] T. Suebchua, B. Manaskasemsak, A. Rungsawang, & H. Yamana. (2018). *History-enhanced focused website segment crawler.* 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018. pp. 80-85.

[13] Soontornpipit, P. (2015). Monitoring and Indexing System for Illegal Tobacco Sales on Website in Thailand by Using Web Crawler Technique. *Applied Mechanics and Materials*, *781*, 129-132.

[14] Diewert, W. E., & Shimizu, C. (2016). Hedonic regression models for Tokyo condominium sales. *Regional Science & Urban Economics*, *60*, 300-315.

[15] W. T. Lim, L. Wang, Y. Wang, & Q. Chang. (2016). *Housing price prediction using neural networks*. 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, 2016. pp. 518-522.