

การพยากรณ์ผู้ที่มีแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยในงานมหกรรมบ้านและคอนโด
โดยใช้เทคนิคการเรียนรู้ของเครื่อง

นางลักษณ์ ตาดทอง

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตวิทยาลัย สาขาวิชาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2563

HOUSING PURCHASE TREND PREDICTION OF THE HOUSE AND CONDO FAIR VISITORS
USING MACHINE LEARNING TECHNIQUES

Nongluck Tadthong

TNI

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology

Graduate School

Thai-Nichi Institute of Technology

Academic Year 2020

หัวข้อวิทยานิพนธ์

การพยากรณ์ผู้ที่มีแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยใน
งานมหรหรรณบ้านและคอนโดโดยใช้เทคนิคการเรียนรู้ของ
เครื่อง

โดย

นางลักษณ์ ตาดทอง

สาขาวิชา

เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษา

ดร. สรมย์พร เจริญพิทย์

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. พิชิต สุขเจริญพงษ์)

วันที่ เดือน พ.ศ.

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ดร. ธรรมภูณ สุระศักดิ์)

.....กรรมการ

(ดร. ณัฐกิตติ จิตรเอื้อตระกูล)

.....กรรมการ

(ดร. ชัชไชย วรณบุญ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ดร. สรมย์พร เจริญพิทย์)

นางลักษณ์ ตาตทอง : การพยากรณ์ผู้ที่มีแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยในงาน
มหกรรมบ้านและคอนโดโดยใช้เทคนิคการเรียนรู้ของเครื่อง. อาจารย์ที่ปรึกษา :
ดร. สรมย์พร เจริญพิทย์, 72 หน้า.

ที่อยู่อาศัยนับเป็นหนึ่งในปัจจัย 4 ที่เป็นปัจจัยพื้นฐานในการดำรงชีวิต และเป็นสิ่งสำคัญใน
การสร้างคุณภาพชีวิตและสร้างคุณภาพของหน่วยย่อยที่สุดของสังคม คือ “ครอบครัว” ทั้งหน่วยงาน
ภาครัฐ และภาคเอกชนสนับสนุนให้ประชาชนคนไทยมีที่อยู่อาศัยเป็นของตนเอง โดยภาครัฐสนับสนุน
ด้วยการออกมาตรการเกี่ยวกับที่อยู่อาศัย และภาคเอกชนมีการพัฒนาที่อยู่อาศัยผ่านโครงการต่าง ๆ
ออกมาสู่ท้องตลาด และในแต่ละปีทางสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยจะมีการจัดงาน
มหกรรมบ้านและคอนโดขึ้น เพื่อให้ผู้ประกอบการนำสินค้ามาแสดงในงาน โดยถูกจัดขึ้น ณ บริเวณหอ
แสดงสินค้าในกรุงเทพมหานคร ทางสมาคมได้เก็บข้อมูลผู้ที่มาลงทะเบียนเข้างาน เพื่อนำมาวิเคราะห์
และนำข้อมูลที่ได้ไปทำการตลาดเพื่อเพิ่มโอกาสในการขาย แต่ไม่เกิดประสิทธิภาพเท่าที่ควร อีกทั้งยัง
ใช้งบประมาณจำนวนมากในการทำการตลาด

ผู้วิจัยจึงพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงาน
มหกรรมบ้านและคอนโดโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อเข้ามาช่วยเพิ่มโอกาสในการขายอยู่
อาศัยให้มีประสิทธิภาพสูงสุด และเพิ่มความสำเร็จในการขายได้ โดยใช้ข้อมูลย้อนหลัง 5 ปีในการทำ
การทดลองสร้างแบบจำลอง โดยใช้อัลกอริทึม ได้แก่ Decision Tree, Naïve Bayes, K-Nearest
Neighbors, Random Forest และเทคนิค Ensemble ในการสร้างแบบจำลองการพยากรณ์

ผลวิจัยพบว่า การประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้าน
และคอนโด โดยใช้ เทคนิค Ensemble มีประสิทธิภาพมากที่สุด ให้ค่าความถูกต้องร้อยละ 79.57
ค่าความแม่นยำร้อยละ 82.18 ค่าความระลึกร้อยละ 59.32 และค่าถ่วงดุลร้อยละ 68.90 และ
สามารถประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างเหมาะสม โดยเมื่อนำไปทดลองใช้กับ
ข้อมูลผู้ที่ตัดสินใจซื้อที่อยู่อาศัยจากงานมหกรรมบ้านและคอนโด พบว่า สามารถประเมินแนวโน้มการ
ตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้อง มีค่าความถูกต้องร้อยละ 78.92 ค่าความแม่นยำร้อยละ 86.42
ค่าความระลึกร้อยละ 56.82 และค่าถ่วงดุลร้อยละ 68.56 นอกจากนี้ทำให้ทราบว่า ปัจจัยที่ส่งผลต่อ
แนวโน้มการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด 3 อันดับแรก ได้แก่ สถานภาพสมรส รองลงเป็นเรื่องของ
รายได้ครอบครัว ความสามารถในการผ่อนชำระ ตามลำดับ

บัณฑิตวิทยาลัย

ลายมือชื่อนักศึกษา

สาขาวิชา เทคโนโลยีสารสนเทศ

ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2563

NONGLUCK TADTHONG: HOUSING PURCHASE TREND PREDICTION OF THE HOUSE AND CONDO FAIR VISITORS USING MACHINE LEARNING TECHNIQUES.
ADVISOR: DR. SAROMPORN CHAROENPIT, 72 PP.

House is one element of 4 basic needs of life. It is also important to improve people's quality of life and create quality for the smallest unit of society, so called family. Both governmental and private agencies have encouraged Thai people to have their own homes. The government has encouraged by issuing various housing measures and policies, while private sector has developed and launched numbers of housing projects to the market. In every year, 3 major housing developer societies organize "House and Condo Fairs", which create opportunities for developers to show their products. Furthermore, they also collected the visitors' and buyers' information for conducting market analyses, which can improve sale opportunities, but they were still inefficiency and high expenses.

The researcher had developed models to predict decision trends of House Purchase of House and Condo Visitors by using Machine Learning Techniques, which could assist the developers to increase their sale opportunity efficiently. The research initially tried with 5 years data sets, and several algorithms, including the 'Decision Tree', the 'Naïve Bays', the 'K-Nearest Neighbors', the 'Radom Forest' and 'Ensemble'.

The research revealed that the highest efficiency for prediction model for decision trends of House Purchase of House and Condo Visitors was using Ensemble technique. It achieved 79.57 per cent of accuracy rate, 82.18 per cent of precision rate 59.32 per cent of recall rate and 68.90 per cent of f1-score rate. Moreover, this model can be appropriately applied to predict decision trends of House Purchase of House and Condo Visitors, which had 78.95 per cent of accuracy rate, 86.42 per cent of precision rate, 56.82 per cent of recall rate and 68.56 per cent of f1-score rate. In addition, there were three significant factors which strongly affecting the decision trend of house purchase, including marital status, family income, and repayment ability, respectively.

Graduate School

Student's Signature.....

Field of Study Information Technology

Advisor's Signature.....

Academic Year 2020

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความกรุณา และการช่วยเหลือสนับสนุนอย่างยิ่งจาก ดร. สรสมัยพร เจริญพิทย์ ที่ให้คำปรึกษาแนะนำที่เป็นประโยชน์ในการทำวิจัย และให้คำแนะนำในการปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเมตตา ด้วยความเอาใจใส่อย่างยิ่ง ตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้ และขอขอบคุณสมาคมธุรกิจบ้านจัดสรร สมาคมอาคารชุดไทย สมาคมอสังหาริมทรัพย์ไทยที่ให้ความอนุเคราะห์ข้อมูลเพื่อนำมาใช้ในการวิจัยครั้งนี้

อีกทั้งผู้วิจัยตระหนักถึงความตั้งใจจริงและความทุ่มเทของคณะกรรมการทุกท่าน ได้แก่ ดร.ณัฐกิตติ จิตรเอื้อตระกูล ดร.ธรรตัญญู สุระศักดิ์ และ ดร.ชัชไชย วรรณบุรณ์ ที่กรุณาให้ข้อเสนอแนะ ที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์ฉบับนี้ พร้อมทั้งแนะนำแนวทางในการปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ ด้วยความดูแลเอาใจใส่เป็นอย่างดี จนสามารถทำให้วิทยานิพนธ์ฉบับนี้เสร็จสิ้นสมบูรณ์

นอกจากนี้ทางผู้วิจัยขอขอบคุณแรงสนับสนุนจากครอบครัว ส่งเสริมให้ความรัก ความห่วงใย เป็นกำลังใจให้ตลอดมา อาจารย์ทุกท่านที่ได้ให้ความรู้และให้คำปรึกษาตลอดระยะเวลาในการศึกษา เพื่อนร่วมชั้นเรียน และบุคคลรอบตัวที่ประกอบประกอบ สนับสนุน ให้กำลังใจ จนสามารถจัดทำวิทยานิพนธ์ฉบับนี้จนเสร็จสิ้น

ผู้วิจัยหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะก่อให้เกิดประโยชน์แก่ผู้ที่สนใจ และสามารถนำไปต่อยอดเพื่อประยุกต์ใช้กับองค์ความรู้ที่เกี่ยวข้อง อีกทั้งสามารถนำไปใช้ในการศึกษาหรือเป็นแนวทางในการนำไปเพิ่มโอกาสในการทำการตลาดของธุรกิจให้ประสบความสำเร็จมากยิ่งขึ้น

นางลักษณ์ ตาดทอง

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตงานวิจัย.....	4
1.4 นิยามศัพท์เฉพาะ.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	4
2 แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 แนวคิดและทฤษฎีเกี่ยวกับกระบวนการตัดสินใจซื้อ (The stages of the buying decision process).....	5
2.1.2 เทคนิคการพยากรณ์ด้วยการเรียนรู้ของเครื่อง (Machine Learning).....	8
2.1.3 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP.....	16
2.1.4 เทคนิคการแบ่งข้อมูลในการสร้างแบบจำลองด้วยวิธีการไขว้ข้อมูล (Cross-validation).....	17
2.1.5 การประเมินประสิทธิภาพของตัวแบบจำลอง (Evaluation Model).....	18

สารบัญ (ต่อ)

บทที่		หน้า
	2.1.6 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient).....	20
2	2.2 งานวิจัยที่เกี่ยวข้อง.....	21
	2.2.1 งานวิจัยเชิงเศรษฐศาสตร์.....	21
	2.2.2 งานวิจัยเชิงเทคนิคในการพัฒนาแบบจำลอง.....	26
3	วิธีดำเนินการวิจัย.....	30
	3.1 ปัญหาของงานวิจัย หรือวัตถุประสงค์ของการวิจัย.....	31
	3.2 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง.....	32
	3.3 กระบวนการสร้างแบบจำลอง.....	33
	3.4 ศึกษาปัจจัย (Attributes).....	43
	3.5 การสรุปผลการวิจัย.....	43
4	ผลการวิจัย.....	44
	4.1 ผลการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจ ซื้อที่อยู่อาศัยของผู้เช่าชมงานมหรรรรมบ้านและคอนโดโดยใช้การ เรียนรู้ของเครื่อง.....	44
	4.1.1 ส่วนที่ 1 เป็นการศึกษาเปรียบเทียบแบบจำลอง โดยการสร้าง แบบจำลอง 5 ครั้ง การประเมินประสิทธิภาพของแบบจำลอง ในส่วนนี้จะวัดผลแบบจำลองโดยดูจากค่าความถูกต้อง.....	44
	4.1.2 ส่วนที่ 2 เป็นการศึกษาเปรียบเทียบแบบจำลอง จากแบบจำลอง ที่ดีที่สุดที่ได้จากการทดลองในส่วนที่ 1 การประเมิน ประสิทธิภาพของแบบจำลองในส่วนนี้ วัดผลตัวแบบจำลองด้วย ตารางสำคัญการวัดความสามารถของการเรียนรู้ของเครื่อง.....	47
	4.1.3 ส่วนที่ 3 เป็นการศึกษาเปรียบเทียบแบบจำลอง จากแบบจำลอง ที่ดีที่สุดที่ได้จากการทดลองในส่วนที่ 1 มาเปรียบเทียบกับ การสร้างแบบจำลองด้วยเทคนิค Ensemble ในการประเมิน ประสิทธิภาพของแบบจำลอง.....	48

สารบัญ (ต่อ)

บทที่		หน้า
4	4.2 ผลการศึกษาการพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจ ซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับ ผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้าน อสังหาริมทรัพย์ในประเทศไทย.....	49
	4.3 การศึกษาปัจจัยที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการ ตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด...	50
5	บทสรุป อภิปราย และข้อเสนอแนะ.....	52
	5.1 สรุปผลการวิจัย.....	52
	5.2 อภิปรายผล.....	54
	5.3 ปัญหาและอุปสรรคในการทำวิจัย.....	56
	5.4 ข้อเสนอแนะงานวิจัย.....	56
	บรรณานุกรม.....	57
	ภาคผนวก.....	61
	ภาคผนวก ก. ผลการเปรียบเทียบค่าความถูกต้องของทดลองสร้าง แบบจำลอง.....	62
	ภาคผนวก ข. ผลการเปรียบเทียบเวลาที่ใช้ในการดำเนินการสร้างแบบ แบบจำลอง.....	66
	ภาคผนวก ค. ผลการสร้างแบบจำลองด้วยเทคนิค Ensemble.....	68
	ประวัติย่อผู้วิจัย.....	72

สารบัญตาราง

ตาราง		หน้า
1.1	แสดงสัดส่วนมูลค่าการก่อสร้างของภาครัฐและภาคเอกชน ปี 2536 – 2562P1..	3
2.1	ปัจจัยที่มีผลต่อความต้องการที่อยู่อาศัย.....	24
3.1	ระยะเวลาการดำเนินการวิจัย.....	31
3.2	แสดงโครงสร้างข้อมูลที่จัดเก็บรวบรวมในงานมหกรรมบ้านและคอนโด.....	34
3.3	แสดงข้อมูลปี 2557 – 2561 เป็นข้อมูลที่ทำให้ความสะดวกและเป็นข้อมูลที่นำมา สร้างแบบจำลอง.....	38
3.4	การเตรียมข้อมูลเพื่อการพัฒนาแบบจำลอง.....	39
4.1	แสดงการเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัย ของผู้เข้าชมงานมหกรรมบ้านและคอนโด.....	45
4.2	แสดงการเปรียบเทียบเวลาที่ใช้ในการดำเนินการสร้างแบบจำลองการ ประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้าน และคอนโด.....	46
4.3	ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลอง 4 เทคนิค.....	47
4.4	ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลองเทคนิค Decision Tree และเทคนิค Ensemble.....	49
4.5	ผลการประเมินประสิทธิภาพของตัวแบบจำลอง Ensemble เมื่อนำไปทดลองใช้ งาน (Confusion Matrix)	49
4.6	ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลองเทคนิค Ensemble.....	50
4.7	ลำดับความสำคัญของปัจจัย (Attributes) ที่มีต่อการประเมินแนวโน้มการ ตัดสินใจซื้อที่อยู่อาศัย.....	51

สารบัญรูป

รูป		หน้า
1.1	แสดงสัดส่วนการโอนกรรมสิทธิ์ที่อยู่อาศัยระหว่าง กรุงเทพฯ ปริมณฑล และ ภูมิภาค ปี 2558 – 2562.....	2
2.1	กระบวนการการตัดสินใจซื้อ 5 ขั้นตอนของผู้บริโภค คอทเลอร์.....	6
2.2	ตัวอย่างเทคนิคต้นไม้ตัดสินใจ (Decision Tree)	10
2.3	เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรส (Random Forest).....	11
2.4	ตัวอย่างการหาเพื่อนบ้านที่ใกล้ที่สุด.....	12
2.5	แผนภาพแสดงเทคนิค Ensemble แบบ Bagging.....	13
2.6	แผนภาพแสดงเทคนิค Ensemble แบบ Boosting.....	14
2.7	แผนภาพแสดงเทคนิค Ensemble แบบ Voting.....	15
2.8	แสดงขั้นตอนของแบบจำลอง CRISP-DM	16
2.9	การแบ่งชุดข้อมูลเพื่อใช้ในการสร้างแบบจำลองด้วยวิธีการไขว้ข้อมูล 10 ชุด เท่ากัน.....	17
2.10	ตารางสำคัญในการวัดความสามารถของการเรียนรู้ของเครื่อง (Confusion Matrix).....	18
2.11	จำนวนที่อยู่อาศัยและอัตราการจ้างงานของเมือง Washoe County ปี ค.ศ. 2001-2015	23
2.12	ปัจจัย (Attribute) ที่สำคัญ.....	26
2.13	ผลการศึกษาเปรียบเทียบเครื่องมือพยากรณ์	27
2.14	ปัจจัย (Attribute) ที่ใช้และเข้ารหัสแฮช.....	28
2.15	ผลการศึกษาเปรียบเทียบเครื่องมือพยากรณ์.....	28
3.1	ขั้นตอนการดำเนินงาน.....	30
3.2	ตัวอย่างข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหรรมบ้านและคอนโด ปี 2557 – 2562 ที่บันทึกลง Excel.....	33
3.3	ขั้นตอนการสร้างแบบจำลอง.....	34
3.4	แสดงข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหรรมบ้านและคอนโด ปี 2557 – 2562 ในโปรแกรม RapidMiner Studio Education.....	37
3.5	ขั้นตอนการเตรียมข้อมูล.....	37

สารบัญรูป (ต่อ)

รูป		หน้า
3.6	แสดงจำนวนการแบ่งชุดข้อมูล.....	39
3.7	กระบวนการสร้างแบบจำลอง.....	40
3.8	การสร้างแบบจำลองด้วย 4 เทคนิค.....	40
3.9	กระบวนการสร้างแบบจำลองด้วยเทคนิค Ensemble	41
3.10	กระบวนการสร้างแบบจำลองด้วยเทคนิค Ensemble ด้วยเทคนิค Voting	41
3.11	กระบวนการทดลองนำแบบจำลองที่มีประสิทธิภาพมากที่สุดมาทดสอบใช้งาน..	42
3.12	กระบวนการศึกษาปัจจัยที่มีนัยสำคัญด้วยเทคนิคความสัมพันธ์ (Correlation)..	43
4.1	กราฟแท่งแสดงการเปรียบเทียบค่าความแม่นยำของแบบจำลอง.....	46

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

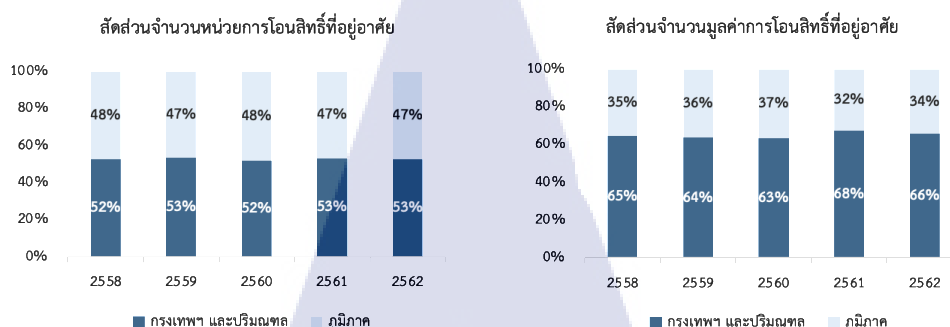
บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ที่อยู่อาศัยนับเป็นหนึ่งในปัจจัย 4 ที่เป็นปัจจัยพื้นฐานในการดำรงชีวิต และเป็นสิ่งสำคัญในการสร้างคุณภาพชีวิตและสร้างคุณภาพของหน่วยย่อยที่สุดของสังคม คือ “ครอบครัว” ในปัจจุบันการผลิตที่อยู่อาศัยเพื่อรองรับความต้องการที่อยู่อาศัยของประชาชนได้มีการดำเนินการทั้งหน่วยงานภาครัฐบาล และภาคเอกชน โดยภาครัฐบาลดำเนินการด้วยการเคหะแห่งชาติ ผ่านโครงการบ้านลำนหลังธนาคารอาคารสงเคราะห์ และบ้านประชารัฐกรมธนารักษ์ ร่วมมือกับการเคหะแห่งชาติ และธนาคารอาคารสงเคราะห์จัดทำโครงการที่อยู่อาศัย สำหรับผู้มีรายได้น้อย โดยใช้พื้นที่ที่ราชพัสดุที่มีอยู่ทั่วประเทศ จัดให้การเคหะแห่งชาติเช่าในอัตราต่ำสุด เพื่อให้การเคหะชื่อนำไปก่อสร้างที่อยู่อาศัยสำหรับผู้มีรายได้น้อย ซึ่งส่วนใหญ่เป็นการพัฒนาที่อยู่อาศัยสำหรับผู้มีรายได้น้อย หรือสำหรับเป็นบ้านสวัสดิการของข้าราชการในหน่วยงานต่าง ๆ และสำหรับภาคเอกชนโดยบริษัทผู้ประกอบการอสังหาริมทรัพย์ ทั้งที่เป็นขนาดใหญ่ที่เป็นบริษัทในตลาดหลักทรัพย์ บริษัทขนาดกลางและขนาดเล็ก ซึ่งได้มีการพัฒนาที่อยู่อาศัยสำหรับกลุ่มผู้มีรายได้น้อย รายได้ปานกลางจนถึงกลุ่มผู้รายได้สูง ผ่านโครงการต่าง ๆ ออกมาสู่ท้องตลาดโดยไม่สามารถคาดการณ์ถึงความต้องการของที่อยู่อาศัยที่แท้จริง

จากการติดตามด้านการลงทุนที่อยู่อาศัยในภาคเอกชนของศูนย์ข้อมูลอสังหาริมทรัพย์ ธนาคารอาคารสงเคราะห์ พบว่า ธุรกิจการพัฒนาที่อยู่อาศัยในประเทศไทยได้รับความสนใจจากผู้ประกอบการภาคเอกชนจำนวนมาก โดยพบว่า ในแต่ละปีมีการเปิดตัวโครงการใหม่ในกรุงเทพมหานครและปริมณฑลรวมถึงจังหวัดหลักๆ ในจังหวัดภูมิภาคเฉลี่ยปีละ 131,400 หน่วยและมีมูลค่าเฉลี่ยสูงถึงปีละ 497,300 ล้านบาท ทั้งนี้ในแต่ละปีจะมีการซื้อขายและโอนกรรมสิทธิ์ที่อยู่อาศัยเฉลี่ยสูงถึงปีละ 335,900 หน่วย และมีมูลค่าเฉลี่ยสูงถึงปีละ 737,100 ล้านบาท จากข้อมูลการโอนกรรมสิทธิ์ที่อยู่อาศัยในปี 2562 พบว่า ตลาดที่อยู่อาศัยในพื้นที่กรุงเทพมหานคร และปริมณฑล มีจำนวนหน่วยโอนกรรมสิทธิ์คิดเป็นสัดส่วนมากถึงร้อยละ 53 ของจำนวนหน่วยการโอนกรรมสิทธิ์ทั่วประเทศ ในขณะที่มูลค่าการโอนกรรมสิทธิ์มีสัดส่วนมากถึงร้อยละ 66 ของมูลค่าการโอนกรรมสิทธิ์ทั่วประเทศ เช่นกัน เห็นได้ว่าตลาดที่อยู่อาศัยในพื้นที่กรุงเทพมหานคร และปริมณฑลเป็นตลาดหลักของประเทศไทย [1]



รูปที่ 1.1 แสดงสัดส่วนการโอนกรรมสิทธิ์ที่อยู่อาศัยระหว่าง กรุงเทพฯ ปริมณฑล และภูมิภาค ปี 2558 – 2562 [1]

ในแต่ละปีทางสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยที่ประกอบไปด้วยสมาชิกผู้ประกอบการของภาคเอกชนได้ร่วมมือกันจัดงานมหกรรมบ้านและคอนโดขึ้นเพื่อเป็นช่องทางการตลาดในการเพิ่มโอกาสขายได้ให้กับสมาชิกในสมาคม โดยมีการเริ่มจัดงานมาตั้งแต่ปี 2555 – 2562 ถูกจัดขึ้นในบริเวณหอแสดงสินค้าในกรุงเทพมหานคร และการจัดงานมหกรรมในแต่ละครั้งมีการเก็บข้อมูลผู้เข้าชมงานมหกรรมเพื่อนำมาเป็นฐานข้อมูลในการวางแผนการตลาด โดยผู้ประกอบการทำการตลาดโดยส่งแผ่นพับโฆษณา ข้อมูลรายละเอียดที่อยู่อาศัย รวมถึงข้อมูลส่งเสริมการตลาดต่าง ๆ ไปให้ผู้ที่มาลงทะเบียนเข้าชมงานมหกรรมตามรายชื่อที่เก็บในงาน ส่งไปทางอีเมล ทางไปรษณีย์ รวมถึงการติดตามเสนอขายที่อยู่อาศัยทางโทรศัพท์ แต่ผลตอบรับการขายไม่มีประสิทธิภาพเท่าที่ควร เพราะเป็นการทำการตลาดแบบหว่านแหต้องส่งข้อมูลหาลูกค้าทุกคนโดยทางผู้ประกอบการเอง ก็ไม่สามารถประเมินหรือตัดสินใจได้ว่า ลูกค้าคนไหนมีโอกาสหรือแนวโน้มที่จะซื้อที่อยู่อาศัยจริง จึงทำให้เสียเวลา เสียค่าใช้จ่ายจำนวนมาก โดยค่าใช้จ่ายในการทำการตลาดทางบริษัทจะมิงงบประมาณไว้ประมาณร้อยละ 3 ของมูลค่าแต่ละโครงการ [2] และเสียโอกาสทางธุรกิจอีกด้วย

นอกจากนี้ ธุรกิจอสังหาริมทรัพย์ยังมีความสำคัญอย่างมากต่อเศรษฐกิจของประเทศไทย โดยมีมูลค่าเฉลี่ยคิดเป็นร้อยละ 3 ของผลิตภัณฑ์มวลรวมของประเทศไทย (Gross Domestic Product : GDP) [3] เป็นหนึ่งในธุรกิจที่สำคัญและขับเคลื่อนเศรษฐกิจประเทศไทย โดยเฉพาะอสังหาริมทรัพย์ที่เป็นประเภทที่อยู่อาศัย ทำให้เกิดการจ้างงานเพิ่มรายได้ และเกิดความเชื่อมโยงในอุตสาหกรรมอื่น ได้แก่ ธุรกิจการเงินการธนาคาร ธุรกิจการก่อสร้าง ธุรกิจวัสดุก่อสร้าง ธุรกิจเกี่ยวกับตกแต่งภายใน เป็นต้น

ตารางที่ 1.1 แสดงสัดส่วนมูลค่าการก่อสร้างของภาครัฐและภาคเอกชน ปี 2536 – 2562P1 [3]

ปี	Gross Domestic Product (GDP)	มูลค่าการก่อสร้างทั้งภาครัฐและภาคเอกชน (ล้านบาท)	% สัดส่วน	ปี	Gross Domestic Product (GDP)	มูลค่าการก่อสร้างทั้งภาครัฐและภาคเอกชน (ล้านบาท)	% สัดส่วน
2536	3,263,426	236,097	7%	2550	9,076,301	269,027	3%
2537	3,689,090	272,623	7%	2551	9,706,929	285,426	3%
2538	4,217,614	296,012	7%	2552	9,658,667	229,891	2%
2539	4,638,604	301,151	6%	2553	10,808,145	267,856	2%
2540	4,710,310	158,727	3%	2554	11,306,906	295,180	3%
2541	4,701,553	77,969	2%	2555	12,357,342	325,919	3%
2542	4,789,826	60,590	1%	2556	12,915,158	329,262	3%
2543	5,069,820	73,398	1%	2557	13,230,301	329,715	2%
2544	5,345,001	89,883	2%	2558	13,743,463	314,573	2%
2545	5,769,578	112,143	2%	2559	14,792,700	311,248	2%
2546	6,317,303	137,748	2%	2560 P	15,573,705	312,773	2%
2547	6,954,282	182,250	3%	2561 P	16,318,033	328,233	2%
2548	7,614,411	224,631	3%	2562 P1	16,258,986	344,457	2%
2549	8,400,644	253,288	3%	ค่าเฉลี่ย			3%

จากเหตุผลดังกล่าวทำให้ผู้วิจัยมีความสนใจและมีแนวคิดที่จะนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาเพื่อพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด เพื่อเป็นเครื่องมือสำหรับให้ผู้ประกอบการอสังหาริมทรัพย์ วางแผนการตลาดในการขายที่อยู่อาศัย โดยมีเป้าหมายเพื่อเพิ่มโอกาสในการขายที่อยู่อาศัยให้มีประสิทธิภาพสูงสุด และเพิ่มความสำเร็จในการขายได้ ทำให้เศรษฐกิจของประเทศเติบโตยิ่งขึ้นไป

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

1.2.2 เพื่อพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย

1.2.3 เพื่อศึกษาปัจจัย (Attributes) ที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

1.3 ขอบเขตการวิจัย

1.3.1 ขอบเขตด้านชุดข้อมูล การศึกษาแบบจำลองการประเมินแนวโน้มการซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรณพบ้านและคอนโด ใช้ข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหรณพบ้านและคอนโด โดยสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยเป็นผู้ดำเนินการจัดงานข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหรณพบ้านและคอนโด ตั้งแต่ปี 2557 ถึง ปี 2562 มีจำนวน 118,197 รายการ โดยมีปัจจัย (Attributes) จำนวน 9 ปัจจัย ได้แก่ เพศ อายุ อาชีพ สถานภาพสมรส รายได้ ครอบครัว ระยะเวลาที่ตัดสินใจซื้อ ราคาที่อยู่อาศัยที่ซื้อ ความสามารถในการผ่อนชำระ และวัตถุประสงค์ที่ซื้อที่อยู่อาศัย

1.3.2 ขอบเขตเทคนิคและวิธีการ การศึกษาและพัฒนาแบบจำลองโมเดลการเรียนรู้ของเครื่อง (Machine Learning) เพื่อการประเมินแนวโน้มการซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรณพบ้านและคอนโด

1.4 นิยามศัพท์เฉพาะ

1.4.1 สมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย หมายถึง สมาคมอสังหาริมทรัพย์ไทย สมาคมธุรกิจบ้านจัดสรร และสมาคมอาคารชุดไทย

1.4.2 ผู้เข้าชมงานมหรณพบ้านและคอนโด หมายถึง เข้าชมงานมหรณพบ้านและคอนโด ครั้งที่ 31 – 41 (ปี 2557 – 2562)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้แบบจำลองการประเมินแนวโน้มการซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรณพบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย เพื่อเพิ่มโอกาสในการขายที่อยู่อาศัยให้มีประสิทธิภาพสูงสุด และเพิ่มความสำเร็จในการขายได้

1.5.2 ได้ศึกษาและวัดผลเปรียบเทียบแบบจำลองการประเมินแนวโน้มการซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรณพบ้านและคอนโด

1.5.3 ได้ทราบถึงปัจจัย (Attributes) มีนัยสำคัญสำหรับการประเมินแนวโน้มการซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรณพบ้านและคอนโด

บทที่ 2

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยนี้เป็นการศึกษาเกี่ยวกับการสร้างเครื่องมือในการพยากรณ์ผู้ที่มีแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยในงานมหกรรมบ้านและคอนโดซึ่งผู้วิจัยได้ทำการศึกษาภายในกรอบแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อให้เป็นแนวทางในการศึกษางานวิจัย โดยทำการศึกษาหัวข้อต่าง ๆ ดังนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 แนวคิดและทฤษฎีเกี่ยวกับกระบวนการตัดสินใจซื้อ (The stages of the buying decision process)

2.1.2 เทคนิคการพยากรณ์ด้วยการเรียนรู้ของเครื่อง (Machine Learning)

2.1.3 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP

2.1.4 เทคนิคการแบ่งข้อมูลในการสร้างแบบจำลองด้วยวิธีการไขว้ข้อมูล (Cross-validation)

2.1.5 การประเมินประสิทธิภาพของตัวแบบจำลอง (Evaluation Model)

2.1.6 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยเชิงเศรษฐศาสตร์

2.2.2 งานวิจัยเชิงเทคนิคในการพัฒนาแบบจำลอง

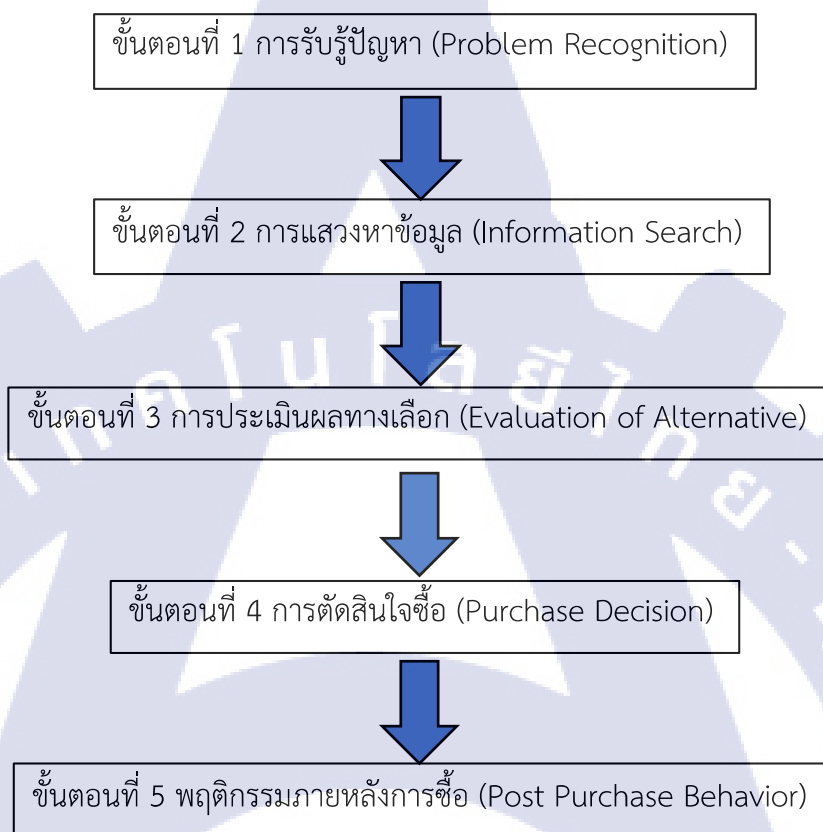
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 แนวคิดและทฤษฎีเกี่ยวกับกระบวนการตัดสินใจซื้อ (The stages of the buying decision process)

ศิริวรรณ เสรีรัตน์ และคณะ [4] กล่าวว่า กระบวนการตัดสินใจของผู้บริโภคจะช่วยให้การตลาดสามารถกำหนดสิ่งกระตุ้นที่เหมาะสม ให้ผู้บริโภคเกิดการตัดสินใจซื้อสินค้าและบริการอย่างมีประสิทธิภาพ เช่น การส่งเสริมการขายด้วยการลดราคาอาจจะสามารถกระตุ้นผู้บริโภคบางกลุ่มให้เกิดการซื้อสินค้าและบริการนั้น ๆ เพิ่มขึ้นได้

อดุลย์ จาตุรงค์กุล [5] กระบวนการตัดสินใจของผู้ซื้อ (Buyer's Decision Process) ตามทฤษฎีของคอตเลอร์ สามารถแบ่งได้เป็น ลำดับขั้นตอนที่สำคัญ ๆ ได้ 5 ขั้นตอน คือ การตระหนัก

ถึงความต้องการ การเสาะแสวงหาข่าวสาร การประเมินทางเลือก การตัดสินใจซื้อและพฤติกรรมหลังการซื้อ



รูปที่ 2.1 กระบวนการการตัดสินใจซื้อ 5 ขั้นตอนของผู้บริโภค คอทเลอร์ [4]

(1) การรับรู้ปัญหา (Problem Recognition)

ในขั้นตอนแรกผู้บริโภคจะตระหนักรู้ถึงปัญหาหรือความต้องการในสินค้าหรือการบริการ ซึ่งเกิดจาก

1. สิ่งกระตุ้นภายใน (Internal Stimuli) เช่น ความรู้สึกหิว กระหายน้ำ เป็นต้น
2. สิ่งกระตุ้นภายนอก (External Stimuli) อาจเกิดจากการกระตุ้นของส่วน

ประสมทางการตลาด (4 P's) เช่น เห็นโฆษณาที่อยู่อาศัยในโทรทัศน์ กิจกรรมส่งเสริมการตลาดจึงเกิดความรู้สึกอยากซื้อ อยากได้ การกระตุ้นมาตรการส่งเสริมของภาครัฐ เศรษฐกิจที่ดี รวมถึงถ้าเป็นการกู้ ถ้ามีข้อเสนอของเรื่องอัตราดอกเบี้ยถูกก็จะกระตุ้นการซื้อได้ การแต่งงานนำไปสู่การเลือกซื้อของที่จำเป็นต่อครอบครัว เป็นต้น

(2) การแสวงหาข้อมูล (Information Search)

เมื่อผู้บริโภคทราบถึงความต้องการในสินค้าหรือบริการแล้ว ลำดับขั้นตอนต่อไป ผู้บริโภคก็จะแสวงหาข้อมูล เพื่อใช้ประกอบการตัดสินใจ โดยแหล่งข้อมูลของผู้บริโภค แบ่งเป็น

1. แหล่งบุคคล (Personal Sources) เช่น การสอบถามจากเพื่อน ครอบครัว คนรู้จักที่มีประสบการณ์ในการใช้สินค้านั้น ๆ
2. แหล่งทางการค้า (Commercial Sources) เช่น การหาข้อมูลจากโฆษณาตามสื่อต่าง ๆ พนักงานขาย ร้านค้า เป็นต้น
3. แหล่งสาธารณชน (Public Sources) เช่น การสอบถามจากรายละเอียดของสินค้าหรือบริการจากสื่อมวลชน หรือองค์กรคุ้มครองผู้บริโภค หรือจากการรวมตัวการจัดงานแสดงสินค้าของผู้ประกอบการ เช่น การจัดงานมหกรรมบ้านและคอนโดของสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย
4. แหล่งประสบการณ์ (Experiential Sources) เกิดจากการประสบการณ์ส่วนตัวของ ผู้บริโภคที่เคยทดลองใช้ผลิตภัณฑ์นั้น ๆ มาก่อน

(3) การประเมินทางเลือก (Evaluation of Alternative)

เมื่อได้ข้อมูลจากขั้นตอนที่ 2 คือ การแสวงหาข้อมูลแล้ว ในขั้นตอนต่อไปผู้บริโภคก็จะทำการประเมินทางเลือก โดยในการประเมินทางเลือกนั้น ผู้บริโภคต้องกำหนดเกณฑ์หรือคุณสมบัติที่จะใช้ในการประเมิน เช่น ยี่ห้อ ราคา รูปแบบ บริการหลังการขาย ราคาขายต่อหน่วย เป็นต้น

(4) การตัดสินใจซื้อ (Purchase Decision)

หลังจากที่ได้ทำการประเมินทางเลือกแล้ว ผู้บริโภคก็จะเข้าสู่ขั้นของการตัดสินใจซื้อซึ่งต้องมีการตัดสินใจในด้านต่าง ๆ ดังนี้

1. ตรายี่ห้อที่ซื้อ (Brand Decision)
2. ร้านค้าที่ซื้อ (Vendor Decision)
3. ปริมาณที่ซื้อ (Quantity Decision)
4. ระยะเวลาที่ซื้อ (Timing Decision)
5. วิธีการในการชำระเงิน (Payment-method Decision)
6. รายได้ (Income)
7. วัตถุประสงค์ในการซื้อ (Purchase Purpose)

(5) พฤติกรรมภายหลังการซื้อ (Post purchase Behavior)

หลังจากที่ลูกค้าได้ตัดสินใจซื้อสินค้าหรือบริการไปแล้วนั้น นักการตลาดจะต้องทำการตรวจสอบความพึงพอใจภายหลังการซื้อ ซึ่งความพึงพอใจนั้นเกิดขึ้นจากการที่ลูกค้าได้เปรียบเทียบสิ่งที่เกิดขึ้นจริงกับสิ่งที่คาดหวัง ถ้าสินค้าหรือบริการที่ได้รับจริง ตรงกับที่คาดหวังหรือสูงกว่าที่ได้คาดหวังเอาไว้ ลูกค้าก็จะเกิดความพึงพอใจในสินค้าหรือ บริการนั้น

จากการศึกษาแนวคิดเกี่ยวกับกระบวนการตัดสินใจซื้อ ทำให้ทราบถึงพฤติกรรมของผู้ซื้อ หรือผู้บริโภคในแต่ละขั้นตอนก่อนที่จะตัดสินใจซื้อผลิตภัณฑ์ต่าง ๆ ในขั้นตอนที่ 1 ผู้ซื้อมีความตระหนักรู้จะพิจารณาถึงภาวะเศรษฐกิจ อัตราดอกเบี้ยต่าง ๆ รวมถึงพิจารณาว่ารัฐบาลมีมาตรการอะไรออกมากระตุ้นบ้างก่อนที่จะตัดสินใจซื้อ ดังนั้นผู้วิจัยจึงไม่นำปัจจัยทางเศรษฐกิจเข้ามาวิเคราะห์ ด้วยเพราะปัจจัยเหล่านี้มีผลต่อการตัดสินใจซื้ออยู่แล้ว และจะเห็นว่าในขั้นตอนที่ 4 ผู้ซื้อจะพิจารณาในด้านต่าง ๆ ก่อนที่จะตัดสินใจซื้อโดยจะพิจารณาว่าตัวเองมีรายได้เพียงพอหรือไม่ มีความสามารถในการผ่อนชำระสินค้าน้อยแค่ไหน รวมถึงวัตถุประสงค์ในการซื้อเพื่ออะไร และระยะเวลาที่จะซื้อเมื่อไหร่ ดังนั้นทำให้ผู้วิจัยได้ทราบถึงปัจจัยที่มีผลต่อการตัดสินใจซื้อจากขั้นตอนนี้

2.1.2 เทคนิคการพยากรณ์ด้วยการเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องจักร หรือ Machine Learning (ML) คือ เป็นกระบวนการที่ทำให้คอมพิวเตอร์มีความสามารถในการเรียนรู้ด้วยตนเอง หรือศาสตร์ที่ว่าด้วย การศึกษาและสร้าง อัลกอริทึม (Algorithm) ที่สามารถเรียนรู้ และทำนายข้อมูลได้ [6] โดยเรียนรู้จากโมเดล (Model) ของข้อมูลที่น่าเข้าไปเป็นตัวอย่าง ทำให้เครื่องจักรสามารถเรียนรู้ มีนิยามของการเรียนรู้ของเครื่องคือการเรียนรู้จากประสบการณ์ เพื่อทำงานได้อย่างมีประสิทธิภาพ และสามารถทำนายข้อมูลได้

2.1.2.1 การแบ่งประเภทการเรียนรู้ของเครื่องคอมพิวเตอร์

ปัจจุบันการเรียนรู้ของเครื่องคอมพิวเตอร์ถูกแบ่ง ออกเป็น 3 ประเภท ตามลักษณะของข้อมูลนำเข้า หรือข้อมูลฝึก (Training Data) โดย Shalev-Shwartz, S., & Ben-David, S. (2009) [7] ได้อธิบายไว้ดังนี้

(1) Supervised Learning คือ การเรียนรู้จากข้อมูลแบบมีโครงสร้าง หรือการเรียนรู้แบบมีผู้สอน เป็นกลุ่มของอัลกอริทึม (Algorithm) ที่เน้นสอนคอมพิวเตอร์จากข้อมูลตัวอย่างในอดีตที่เฉลยผลลัพธ์ที่เป็นลาเบล (Label) เป็นการเรียนรู้โดยอาศัยข้อมูลที่ป้อนเข้าไปเพื่อให้คอมพิวเตอร์นำไปใช้เปรียบเทียบกับข้อมูลที่ป้อนเข้ามาใหม่ จากนั้นเครื่องคอมพิวเตอร์จะทำนายผลหรือจัดหมวดหมู่ที่มีความเหมือนกันมากที่สุดให้อยู่ด้วยกัน ผลลัพธ์ที่ได้เป็นการจัดหมวดหมู่ (Classification) เช่น ต้องการให้คอมพิวเตอร์แยกภาพของชนิดที่อยู่อาศัยระหว่างแนวราบ กับอาคารชุด ก็ต้องมีข้อมูลภาพตัวอย่างของที่อยู่อาศัย 2 นี้มาป้อนให้คอมพิวเตอร์เรียนรู้ว่าถ้ารูปภาพมีลักษณะแบบนี้คือรูปแนวราบ อาคารชุด หรือต้องการให้คอมพิวเตอร์ประมวลผลว่าคนที่มาขอสินเชื่อควรได้รับการอนุมัติหรือไม่ ต้องทำการหากลุ่มตัวอย่างเพื่อมาสอนคอมพิวเตอร์ว่าคนที่มีลักษณะแบบนี้ มีประวัติการเงินแบบนี้

(2) Unsupervised Learning คือ การเรียนรู้จากข้อมูลแบบไม่มีโครงสร้าง หรือการเรียนรู้แบบไม่มีผู้สอน หรือไม่มีข้อมูลเฉลยผลลัพธ์ที่เป็นลาเบล (Label) ให้

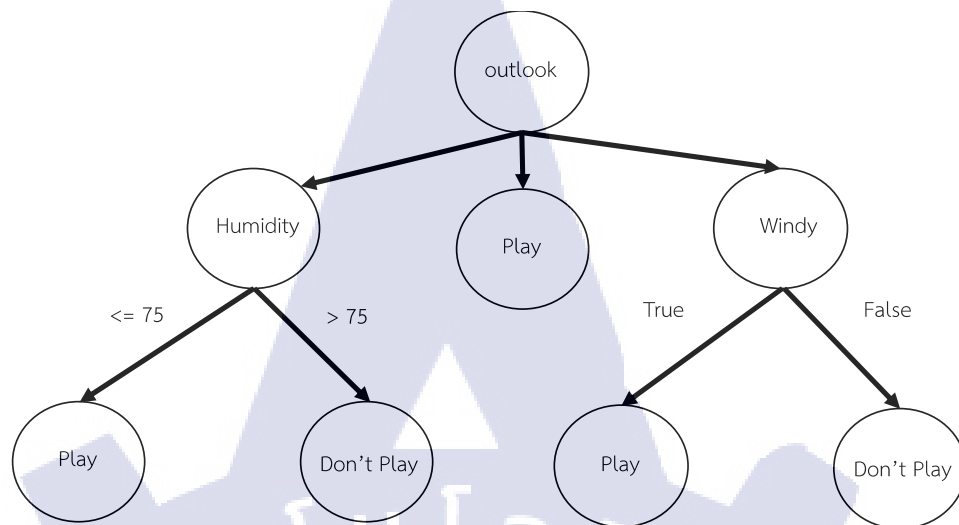
เครื่องจักรได้เรียนรู้ ซึ่งเครื่องคอมพิวเตอร์ต้องหาโครงสร้างของข้อมูลนั่นเอง เช่นต้องการแบ่งกลุ่มข้อมูล (Clustering) แต่ไม่สามารถระบุผลลัพธ์ที่จะได้ เครื่องคอมพิวเตอร์จะทำการแบ่งกลุ่มข้อมูลตามเงื่อนไขที่เราระบุเท่านั้น เช่น ต้องการแบ่งกลุ่มลูกค้าจากระยะเวลาการใช้งานเครือข่ายโทรศัพท์ และค่าใช้จ่ายรายเดือนของการโทรศัพท์ เครื่องคอมพิวเตอร์จะทำการแบ่งกลุ่มเองจากข้อมูลดังกล่าว

(3) reinforcement learning คือ การเรียนรู้ในรูปแบบที่ใกล้เคียงกับการเรียนรู้ของมนุษย์มากที่สุด เป็นการเรียนรู้แบบไม่มีผู้สอนแต่จะมีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนแปลงตลอดเวลา ทำให้เกิดการกระตุ้นในการสร้างแบบจำลอง เพื่อตอบสนองว่าจะต้องทำอะไรต่อไป การกระทำแต่ละครั้งจะเปลี่ยนไปตามสภาพแวดล้อมในขณะนั้น

2.1.2.2 ประเภทของอัลกอริทึม (Algorithm) ที่เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning)

จากการศึกษาการแบ่งกลุ่มการเรียนรู้ของเครื่องคอมพิวเตอร์ ได้ข้อสรุปในการทำงานวิจัย การศึกษาการสร้างเครื่องมือในการพยากรณ์ผู้ที่มีแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยในงานมหกรรมบ้านและคอนโด เหมาะกับประเภทการเรียนรู้แบบมีผู้สอน (Supervised Learning) เนื่องจากมีข้อมูลที่บอกลักษณะ (Attribute) ของคนที่ตัดสินใจซื้อที่อยู่อาศัย และข้อมูลลักษณะ (Attribute) ของคนที่ไม่ตัดสินใจซื้อที่อยู่อาศัย ดังนั้นจึงทำการศึกษาประเภทของโมเดลที่จะนำมาใช้ในการวิจัย ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเบย์เซียนแบบง่าย (Naïve Bayes) เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เทคนิคแรนดอมฟอร์เรส (Random Forest) โดยมีรายละเอียดดังนี้

(1) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) [8] เป็นการคัดเลือกปัจจัย (Attribute) ที่มีความสัมพันธ์กับคลาสมากที่สุด ขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาปัจจัย (Attribute) ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ ของปัจจัย (Attribute) นี้จะใช้ตัววัดที่เรียกว่า Information Gain (IG) การคำนวณค่า Information Gain จะใช้ค่า Entropy ซึ่งเป็นการวัดความแตกต่างหรือการกระจายกระจายของข้อมูลถ้าข้อมูลมีความแตกต่างกันมากค่า Entropy ก็จะมีค่าสูง ในทางตรงข้ามถ้าข้อมูลมีความคล้ายกันมากค่า Entropy ก็จะมีค่าที่ต่ำ



รูปที่ 2.2 ตัวอย่างเทคนิคต้นไม้ตัดสินใจ (Decision Tree)

(2) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอว์เรส (Random Forest) เป็นการจำแนกประเภทแบบไม่ตัดแต่งกิ่ง (unpruned) หรือต้นไม้ถดถอย (Regression Trees) โดยจะค้นหาโหนด (Node) แต่ละโหนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้โดยใช้การสุ่มเลือกคุณสมบัติจาก N คุณสมบัติ แล้วนำมาสร้างโมเดลป่าหลาย ๆ ต้น แล้วทำการให้คะแนน (Vote) จนได้ผลลัพธ์ที่ดีที่สุด หากต้นไม้ต้นใดได้คะแนนสูงสุด จะถูกนำมาสร้างโมเดลต่อไป

โมเดลป่าแบบสุ่ม (Random Forest) จะช่วยเพิ่มความสามารถในการทำนายให้กับตัวแปรต้นผ่านการสร้างกลุ่มของโมเดลพื้นฐาน (Base prediction) ด้วยวิธีต้นไม้ตัดสินใจ (Decision tree) ในขั้นตอนแรก ก่อนที่จะสร้างโมเดลสุดท้าย (Final prediction) จากการรวบรวมแต่ละโมเดลพื้นฐานเข้าด้วยกัน โดยใช้สูตรในการรวบรวมดังนี้

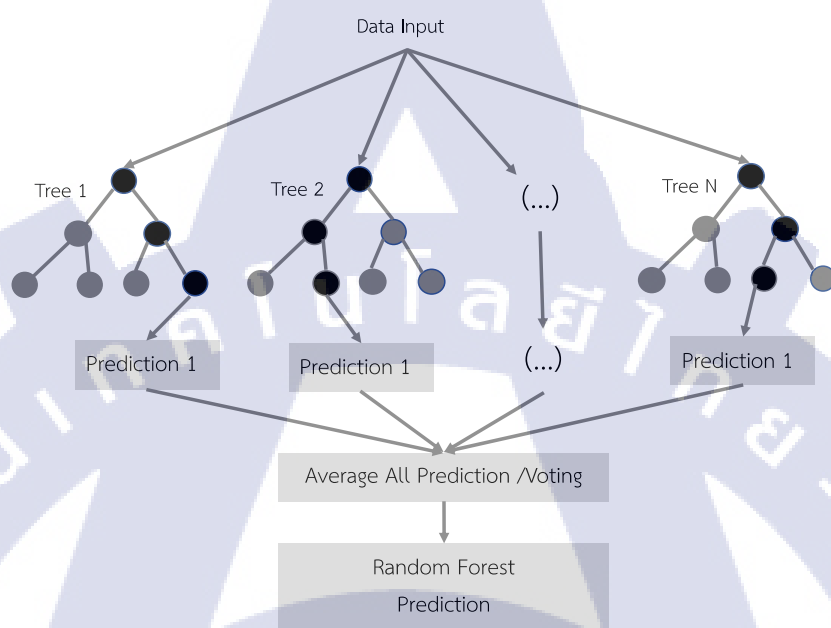
$$E(x, M) = \frac{1}{S} \sum_{s=1}^S \beta_s M_s$$

โดยที่ M คือ $(M_1, M_2, \dots, M_S)$, $M_s(x)$ เป็นแบบจำลองทำนายของแต่ละโมเดลพื้นฐาน

S คือ จำนวนชุดของโมเดลพื้นฐาน และ

β_s คือ สัดส่วนน้ำหนักของแต่ละโมเดลพื้นฐานในโมเดลสุดท้าย

โดยเทคนิคนี้สามารถใช้ได้ดีกับข้อมูลที่มีขนาดใหญ่ และมีตัวแปรต้นจำนวนมาก เช่นเดียวกับข้อมูลในงานวิจัยนี้ นอกจากนี้ยังมีประสิทธิภาพในการจัดการกับข้อมูลที่ขาดหาย และข้อมูลที่มีความสมดุลได้ด้วย [9]



รูปที่ 2.3 เทคนิคการจำแนกข้อมูลด้วยวิธีเร้นดอมฟอว์เรส (Random Forest)

(3) เทคนิคเบเซียนแบบง่าย (Naïve Bayes) เป็นเทคนิคการใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องมากที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ เป็นการจำแนกข้อมูลโดยใช้ความน่าจะเป็นและคำนวณการแจกแจงความน่าจะเป็นตามสมมติฐานที่ตั้งให้กับข้อมูล จากการคำนวณตัวอย่างใหม่ที่ได้จะถูกนำมาปรับเปลี่ยนการแจกแจง ซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็นของข้อมูล ข้อมูลใหม่ที่เกิดขึ้นและตัวแบบที่ตั้งไว้ให้กับข้อมูล จะถูกปรับเปลี่ยนไปตามข้อมูลใหม่ โดยผนวกกับข้อมูลที่มีอยู่เดิม

การเรียนรู้แบบเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน ใช้การคำนวณการหาความน่าจะเป็น ซึ่งถูกใช้ในการทำนายผลเป็นวิธีในการแก้ปัญหาแบบการจำแนกที่สามารถคาดการณ์ผลลัพธ์ได้ โดยจะวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมาก และปัจจัย (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน มีสมการดังนี้ [10]

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

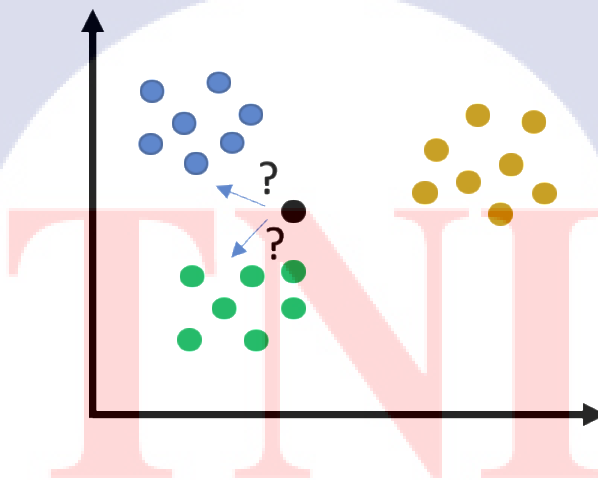
$P(c|x)$ คือ Posterior Probability

$P(x|c)$ คือ Likelihood

$P(c)$ คือ Class Prior Probability

$P(x)$ คือ Predictor Prior Probability

(4) เทคนิคการค้นหาลูกบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) เป็นวิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน ในขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุด ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่างๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด เป็นการหาระยะห่างระหว่างแต่ละตัวแปร (Attribute) ในข้อมูล จากนั้นก็คำนวณค่าออกมา มาเป็นฐานสำหรับการจัดคลาสในเงื่อนไขใหม่ ๆ โดยมีขั้นตอนวิธีการเพื่อนบ้านใกล้ที่สุดมีขั้นตอนโดยสรุปดังนี้ [11]

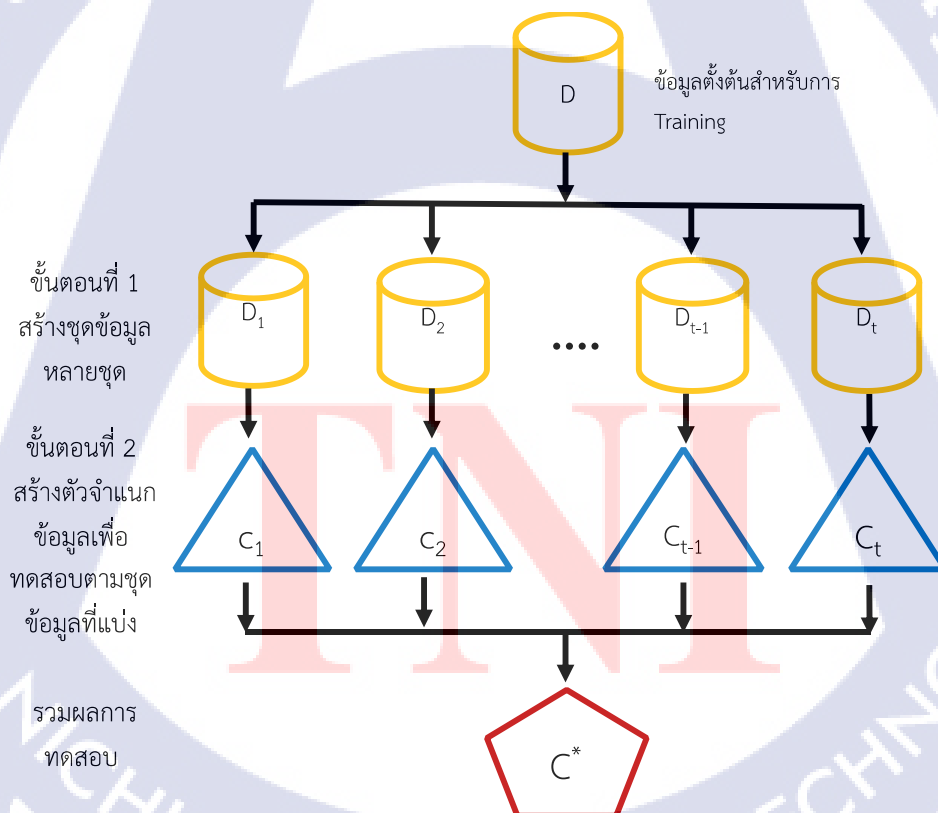


รูปที่ 2.4 ตัวอย่างการหาเพื่อนบ้านที่ใกล้ที่สุด

1. กำหนดขนาดของ K (ควรกำหนดให้เป็นเลขคี่)
2. คำนวณระยะห่าง (Distance) ของข้อมูลที่ต้องการพิจารณาในกลุ่ม

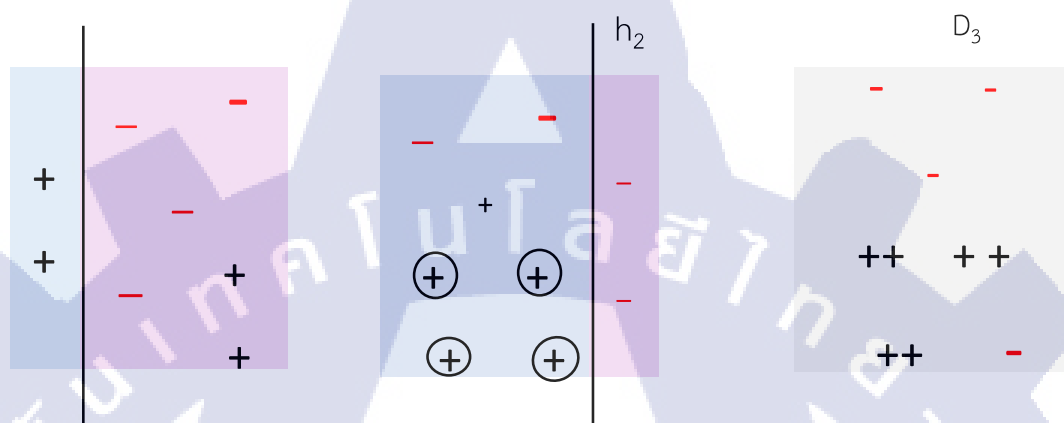
ข้อมูลตัวอย่าง

3. จัดเรียงลำดับของระยะห่าง และเลือกพิจารณาชุดข้อมูลที่ใกล้จุดที่ต้องการพิจารณาตามจำนวน K ที่กำหนดไว้
 4. พิจารณาข้อมูลจำนวน k ชุด และสังเกตว่ากลุ่ม (class) ไหนที่ใกล้จุดที่พิจารณาเป็นจำนวนมากที่สุด
 5. กำหนด class ให้กับจุดที่พิจารณา (class) ที่ใกล้จุดพิจารณามากที่สุด
- (5) เทคนิค Ensemble เป็นการรวมเทคนิคการทำแบบจำลองจำแนกข้อมูล (classification) หลายๆ แบบจำลองมารวมกัน โดยมีวัตถุประสงค์เพื่อเป็นการเพิ่มประสิทธิภาพการทำนายของแบบจำลอง [12] และมีเทคนิคดังนี้
1. Bagging เป็นการสร้างแบบจำลองหลายชุดแต่เป็นประเภทเดียวกัน โดยสร้างจากข้อมูลหลายๆ ชุด จากข้อมูลตั้งต้นชุดเดียวกัน เพื่อนำมาทำการทดสอบกับ subset ของข้อมูล จากนั้นนำผลการทำนายของโมเดลต่าง ๆ มารวมกัน โดยการหาค่าเฉลี่ยของผลการทำนาย ตัวอย่างเช่น Random Forest, Decision Trees, และ Extra Trees



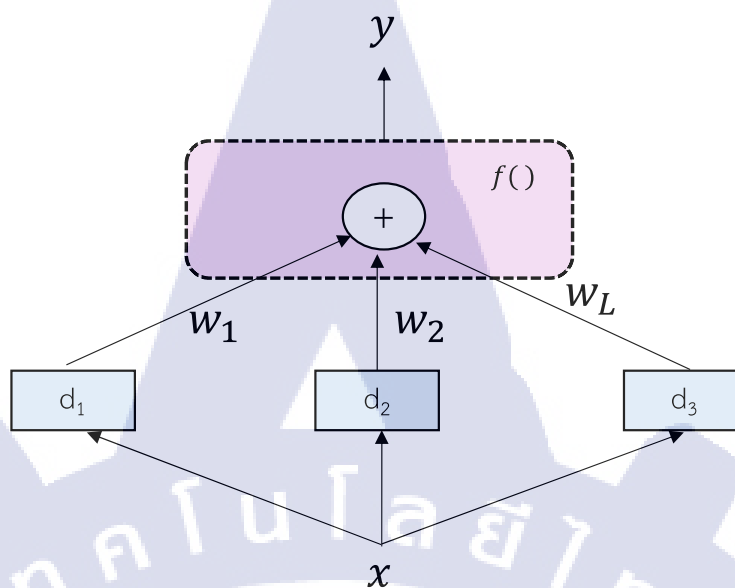
รูปที่ 2.5 แผนภาพแสดงเทคนิค Ensemble แบบ Bagging

2. Boosting เป็นการสร้างแบบจำลองหลายชุดแต่ประเภทเดียวกัน โดยทำการทดสอบแบบ iteration วนลูป ปรับค่าน้ำหนัก เพื่อให้ผลการทำนายของแบบจำลองดีขึ้น ปรับปรุงเรื่อย ๆ มีความยืดหยุ่นใช้ได้กับทุก Learning algorithm สามารถปรับลด Bias error ของแบบจำลองได้ดี ตัวอย่างเช่น AdaBoost และ Stochastic Gradient Boosting



รูปที่ 2.6 แผนภาพแสดงเทคนิค Ensemble แบบ Boosting

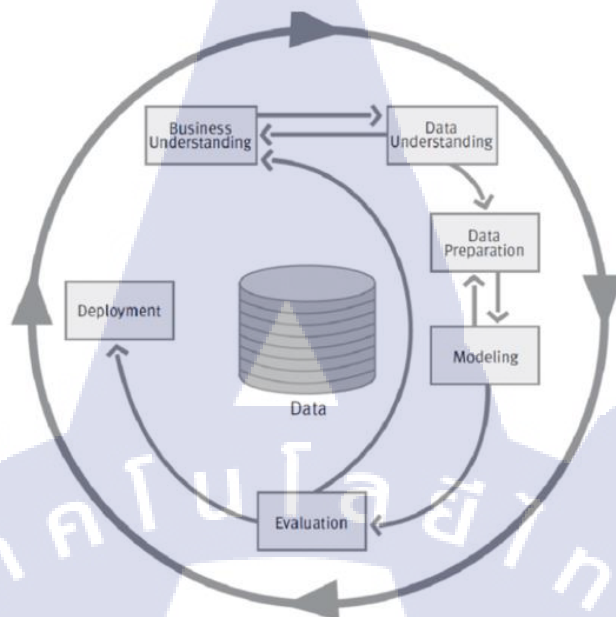
3. Voting เป็นการสร้างแบบจำลองหลากหลาย เช่น Decision Tree, SVM, K-Nearest Neighbors จากนั้นทดสอบบนข้อมูลแบบเดียวกัน เพื่อดูผลการ prediction สุดท้ายของแต่ละแบบจำลอง ใช้การโหวตผลที่เหมือนกัน คล้ายกัน เป็นคำตอบสุดท้าย โดยจะใช้แบบจำลองจำนวนกี่ ในการสร้างแบบจำลองด้วยเทคนิคนี้



รูปที่ 2.7 แผนภาพแสดงเทคนิค Ensemble แบบ Voting

2.1.3 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

ผู้วิจัยได้ศึกษากระบวนการวิเคราะห์ข้อมูลด้วยคริสป์-ดีเอ็ม (Cross Industry Standard Process for Data Mining) ซึ่งเป็นขั้นตอนมาตรฐานที่ใช้ในการวิเคราะห์ข้อมูลสำหรับการทำเหมืองข้อมูลเพื่อนำมาปรับใช้เป็นกรอบแนวคิดในการทำวิจัย โดยสรุปขั้นตอนสำคัญทั้งหมดที่ดำเนินการแบ่งกระบวนการทำงานออกเป็น 6 ขั้นตอนมาตรฐานดังแสดงดังรูปที่ 2.8 [13]



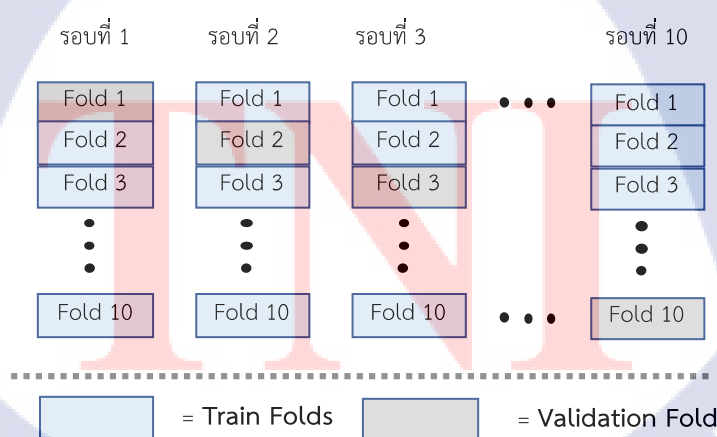
รูปที่ 2.8 แสดงขั้นตอนของแบบจำลอง CRISP-DM [13]

1. การทำความเข้าใจธุรกิจ (Business Understanding) เป็นการศึกษาถึงวัตถุประสงค์ในการทำงาน วิเคราะห์ถึงโอกาสเชิงธุรกิจ โดย ระบุผลลัพธ์ที่ต้องการได้จากการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูล (Data Mining) และวางแผนงาน
2. การทำความเข้าใจข้อมูล (Data Understanding) เป็นการพิจารณาว่าข้อมูลมีความถูกต้องน่าเชื่อถือ มีปริมาณที่เหมาะสมและมีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์ และการสร้างแบบจำลอง
3. การเตรียมข้อมูล (Data Preparation) หรือการทำความสะอาดข้อมูล (Data Cleaning) ในขั้นตอนนี้จะทำการคัดเลือกข้อมูล (Data Selection) และการแปลงข้อมูล (Data Transformation) เพื่อเตรียมไปใช้ในการสร้างแบบจำลอง
4. การสร้างแบบจำลอง (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลต่าง ๆ เช่น เทคนิคกฎความสัมพันธ์ เทคนิคการจัดกลุ่ม (Clustering) และเทคนิคการจำแนกกลุ่ม (Classification)
5. การประเมินประสิทธิภาพ (Evaluation) เป็นขั้นตอนความน่าเชื่อถือของตัวแบบจำลองว่าสามารถใช้งานได้มีประสิทธิภาพบรรลุตามเป้าหมายที่ตั้งไว้

6. การนำไปใช้งาน (Deployment) เป็นการนำข้อมูลที่วิเคราะห์ได้ไปใช้งาน เช่น การนำองค์ความรู้ที่ได้ไปกำหนดกลยุทธ์ต่าง ๆ ในการทำธุรกิจ หรือนำผลการพยากรณ์จากแบบจำลองที่สร้างขึ้นไปในการวางแผนธุรกิจ

2.1.4 เทคนิคการแบ่งข้อมูลในการสร้างแบบจำลองด้วยวิธีการไขว้ข้อมูล (Cross-validation)

เป็นวิธีการ Cross-validation หรือที่นิยมเรียกว่า “k-fold cross validation” จะเริ่มจากการสุ่มแบ่ง ข้อมูลออกเป็นจำนวน k ส่วน ซึ่งจะมีข้อมูลไม่ซ้ำกัน โดยชุดข้อมูลย่อย $D_1, D_2, \dots, D_k$ จะมีจำนวนรายการข้อมูลเท่ากัน จากนั้นจะทำการเรียนรู้ และทดสอบตัวจำแนกข้อมูลทั้งสิ้น k ครั้ง โดยในการทำงานในรอบที่ i จะกำหนดให้ชุดข้อมูลย่อย D_i ให้เป็นชุดข้อมูลทดสอบ (Test Data) และชุดข้อมูลอื่น ๆ เป็นชุดข้อมูลสอน (Training Data) ตัวอย่างเช่น ในการทำงานรอบแรก ชุดข้อมูลย่อย $D_2, \dots, D_k$ จะเป็นชุดข้อมูลสอน และชุดข้อมูลย่อย D_1 จะเป็นชุดข้อมูลที่ใช้ในการทดสอบ แต่สำหรับการทำงานในรอบที่สอง ชุดข้อมูลย่อย $D_1, D_3, \dots, D_k$ จะเป็นชุดข้อมูลสอน และชุดข้อมูลย่อย D_2 จะเป็นชุดข้อมูลทดสอบ ตามลำดับ โดยส่วนใหญ่ของการทำมักนิยมกำหนดค่า $k = 10$ ที่ซึ่งจะเรียกว่า 10-fold cross-validation [14] ถ้า $k=10$ ต้องสอนแบบจำลองทั้งหมด 10 รอบด้วย train folds และทดสอบแบบจำลองทั้งหมด 10 รอบด้วย validation fold วิธีการนี้เป็นการกระจายข้อมูลสอน และข้อมูลทดสอบให้มีความเสมอภาคกันจะช่วยลดไม่ให้เกิดการโน้มเอียง (Bias) ของผลการทดลอง ดังรูปที่ 2.9



รูปที่ 2.9 การแบ่งชุดข้อมูลเพื่อใช้ในการสร้างแบบจำลองด้วยวิธีการไขว้ข้อมูล 10 ชุดเท่ากัน

2.1.5 การประเมินประสิทธิภาพของตัวแบบจำลอง (Evaluation Model)

การประเมินประสิทธิภาพของตัวแบบจำลองจากเทคนิคการสร้างแบบจำลองด้วยวิธีการนำมาสร้างตารางสำคัญในการวัดความสามารถของการเรียนรู้ของเครื่อง หรือเรียกว่า Confusion Matrix [15]

		Actual Values	
		Yes	No
Predicted Values	Yes	TP	FP
	No	FN	TN

รูปที่ 2.10 ตารางสำคัญในการวัดความสามารถของการเรียนรู้ของเครื่อง (Confusion Matrix)

โดยที่ True = ทายถูก, False = ทายผิด, Positive = ทายว่าใช่, Negative = ทายว่าไม่ใช่

1. True Positive (TP) คือ ทายว่าใช่ แล้วใช่จริง ๆ (Hit)
2. True Negative (TN) คือ ทายว่าไม่ใช่ แล้วไม่ใช่จริง ๆ (Correct Rejection)
3. False Positive (FP) คือ ทายว่าใช่ แต่จริง ๆ แล้วไม่ใช่ (False Alarm, Type I error)
4. False Negative (FN) คือ ทายว่าไม่ใช่ แต่จริง ๆ แล้วมันใช่ (Miss, Type II error)
5. Condition Positive (P) คือ จำนวนของที่ใช่ทั้งหมด ที่อยู่ในข้อมูล = TP + FN
6. Condition Negative (N) คือ จำนวนของที่ไม่ใช่ทั้งหมด ที่อยู่ในข้อมูล = FP + TN

ในการทำวิจัยครั้งนี้เป็นการสร้างแบบจำลองเพื่อประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยโดยมีจำนวน 2 คลาส คือ ซื้อ และ ไม่ซื้อ ดังนั้นจึงกำหนดค่าในตาราง Confusion Matrix ดังนี้

TP คือ การจำแนกหมวดหมู่ได้อย่างถูกต้อง โดยที่แบบจำลองทำนายว่า “ซื้อ” และมีคำตอบเป็น “ซื้อ”

TN คือ การจำแนกหมวดหมู่ได้อย่างถูกต้อง โดยที่แบบจำลองทำนายว่า “ไม่ซื้อ” และมีคำตอบเป็น “ไม่ซื้อ”

FP คือ การจำแนกหมวดหมู่ผิดพลาด โดยที่แบบจำลองทำนายว่า “ซื้อ” แต่มีคำตอบเป็น “ไม่ซื้อ”

FN คือ การจำแนกหมวดหมู่ผิดพลาด โดยที่แบบจำลองทำนายว่า “ไม่ซื้อ” แต่มีคำตอบเป็น “ซื้อ”

ในการประเมินแบบจำลองนิยมทำการวัดค่าซึ่งประกอบด้วย 4 ด้านได้แก่

1. ด้านค่าความถูกต้อง (Accuracy) หมายถึง ค่าความถูกต้องของตัวแบบแบบจำลองโดยพิจารณาทุกคลาส มีสมการดังนี้

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

เป็นค่าที่ใช้พิจารณา ในกรณีอยากทราบว่าแบบจำลองที่ทำนายได้ตรงกับสิ่งที่เกิดขึ้นจริงมากน้อยแค่ไหน หรือเป็นค่าผลรวมของตัวเลขบนเส้นทแยงมุมในตาราง ต่อจำนวนเหตุการณ์ที่เกิดขึ้นทั้งหมด

ในงานวิจัยหมายถึง แบบจำลองสามารถทำนายว่าลูกค้าจะซื้อที่อยู่อาศัย และลูกค้าก็ซื้อที่อยู่อาศัยจริง หรือ ทำนายว่าลูกค้าจะไม่ซื้อที่อยู่อาศัย และลูกค้าก็ไม่ซื้อที่อยู่อาศัยจริง ได้ตรงกันมากที่สุด

2. ด้านค่าความแม่นยำ (Precision) หมายถึง ค่าความแม่นยำของตัวแบบจำลองโดยพิจารณาแยกทีละคลาส มีสมการดังนี้

$$\text{Precision} = \frac{TP}{TP + FP}$$

เป็นค่าที่ใช้พิจารณา ในกรณีอยากทราบว่าแบบจำลองที่ทำนายมีความแม่นยำขนาดไหน เป็นการเปรียบเทียบระหว่าง การทำนายที่ถูกต้องว่า จริง และก็เกิดขึ้นจริง (TP) กับ การทำนายว่าจริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง (FP)

ในงานวิจัยนี้ หมายถึง แบบจำลองสามารถทำนายถูกต้องว่าลูกค้าซื้อที่อยู่อาศัยจริง จากการทำนายว่าจะซื้อทั้งหมด

3. ด้านค่าความระลึก (Recall) หมายถึง ค่าความถูกต้องของตัวแบบจำลอง โดยจะพิจารณาทีละคลาส มีสมการดังนี้

$$\text{Recall} = \frac{TP}{TP + FN}$$

เป็นค่าที่ใช้พิจารณา ในกรณีอยากทราบว่าแบบจำลองที่สร้างขึ้น มีความถูกต้องของการทำนายว่าจะเป็น “จริง” เทียบกับ จำนวนครั้งของเหตุการณ์ทั้งทำนาย และ เกิดขึ้น ว่า “เป็นจริง”

ในงานวิจัยนี้ หมายถึง แบบจำลองสามารถทำนายถูกต้องว่าลูกค้าซื้อที่อยู่อาศัยจริง จากการซื้อที่อยู่อาศัยจริงของลูกค้าทั้งหมด

4. ด้านค่าความถ่วงดุล (F1 Score) หมายถึง การวัดค่า Precision และ Recall พร้อมกันของแบบจำลอง โดยพิจารณาแยกทีละคลาส มีสมการดังนี้

$$F1 - Score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

เป็นค่าที่ใช้พิจารณา ค่าเฉลี่ยแบบ Harmonic Mean ระหว่าง Precision และ Recall จุดประสงค์ของการสร้าง F1 ขึ้นมาเพื่อเป็น Single Metric ที่วัดความสามารถของโมเดล

ในงานวิจัยนี้ หมายถึง แบบจำลองสามารถทำนายถูกต้องว่าลูกค้าจะซื้อที่อยู่อาศัยจริง และ ในขณะเดียวกันก็ต้องซื้อที่อยู่อาศัยจริงด้วย

2.1.6 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

โดยทั่วไปนิยมใช้สัญลักษณ์ (r) แทนสัมประสิทธิ์สหสัมพันธ์ของกลุ่มการ บอกระดับหรือขนาดของความสัมพันธ์จะใช้ตัวเลขของค่าสัมประสิทธิ์สหสัมพันธ์หากค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ -1 หรือ +1 แสดงถึงการมีความสัมพันธ์กันในระดับสูง แต่หากมีค่าเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันในระดับน้อย หรือไม่มีเลย สำหรับการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์โดยทั่วไปอาจใช้เกณฑ์ดังนี้ [16]

ค่า (r) ระดับของความสัมพันธ์

0.90 - 1.00	หมายถึง มีความสัมพันธ์กันสูงมาก
0.70 - 0.90	หมายถึง มีความสัมพันธ์กันในระดับสูง
0.50 - 0.70	หมายถึง มีความสัมพันธ์กันในระดับปานกลาง
0.30 - 0.50	หมายถึง มีความสัมพันธ์กันในระดับต่ำ
0.00 - 0.30	หมายถึง มีความสัมพันธ์กันในระดับต่ำมาก

เครื่องหมาย (+, -) หน้าตัวเลขสัมประสิทธิ์สหสัมพันธ์ หมายถึงการบอกถึงทิศทางของความสัมพันธ์โดยหาก

r มีเครื่องหมาย + (บวก) หมายถึง การมีความสัมพันธ์กันไปในทิศทางเดียวกัน อธิบายได้ว่าตัวแปรหนึ่งมีค่าสูง อีกตัวหนึ่งจะมีค่าสูงตามไปด้วย

r มีเครื่องหมาย - (ลบ) การมีความสัมพันธ์กันไปในทิศทางตรงกันข้าม อธิบายได้ว่าตัวแปรหนึ่งมีค่าสูง อีกตัวหนึ่งจะมีค่าต่ำ

ยกเว้นค่าสัมประสิทธิ์สหสัมพันธ์บางชนิดที่มีลักษณะ $0 \leq r \leq 1$ ซึ่งจะบอกได้เพียงขนาดหรือระดับของความสัมพันธ์เท่านั้น ไม่สามารถบอกทิศทางของความสัมพันธ์ได้

2. งานวิจัยที่เกี่ยวข้อง

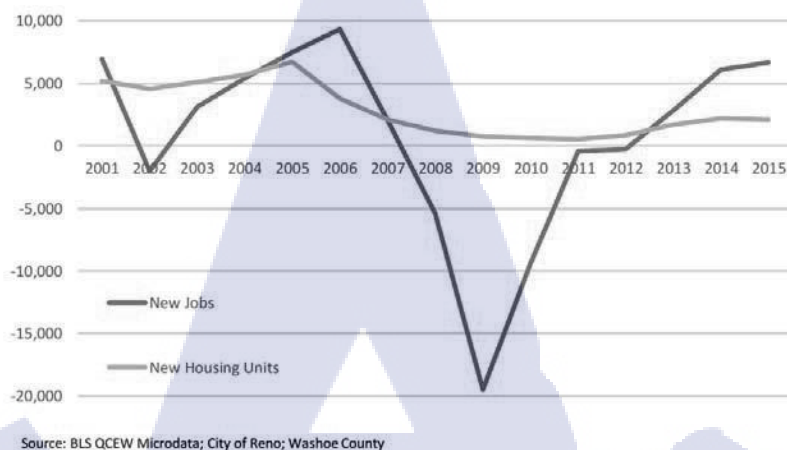
2.2.1 งานวิจัยเชิงเศรษฐศาสตร์

ที่ผ่านมาผู้วิจัยได้ศึกษาถึงปัจจัยที่ส่งผลต่อความต้องการซื้อที่อยู่อาศัย ดังเช่น งานวิจัยของอนุรัตน์ ช่างอง และจิรนนท์ อนุสรณ์พานิช [17] ได้ทำการศึกษถึงปัจจัยที่ส่งผลต่อการตัดสินใจเลือกซื้ออสังหาริมทรัพย์ประเภททรัพย์สินรอกการขายของธนาคารออมสิน พบว่า ลักษณะทางประชากรศาสตร์ที่มีผลต่อการตัดสินใจเลือกซื้อทรัพย์สินรอกการขายนั้นได้แก่ เพศ อาชีพ รายได้ และจำนวนสมาชิกในครอบครัว ส่วนปัจจัยทางเศรษฐกิจได้แก่ อัตราดอกเบี้ย และอัตราค่าธรรมเนียมการโอนกรรมสิทธิ์ ซึ่งได้สอดคล้องกับงานวิจัยของ ศักดิ์สิทธิ์ มิตรเจริญถาวร [18] ศึกษาพฤติกรรมผู้ซื้อทรัพย์สินรอกการขาย กรณีศึกษาผู้เข้าชมทรัพย์สินรอกการขายของสถาบันการเงินในงานบ้านและคอนโดปี 2557 และงานวิจัยของ อลงกรณ์ จาฟู [19] ทำการศึกษาเปรียบเทียบปัจจัยในการเลือกอาคารชุดพักอาศัยเป็นบ้านหลังที่สองในเขตชั้นกลางของกรุงเทพมหานคร พบว่า อาชีพ รายได้ และราคาที่อยู่อาศัยที่เหมาะสมกับรายได้มีผลต่อการตัดสินใจซื้อที่อยู่อาศัย งานวิจัยทั้ง 2 ฉบับให้ผลการศึกษาที่ตรงกันว่ากลุ่มเป้าหมายที่มีความต้องการซื้อที่อยู่อาศัยเป็นกลุ่มที่มีอาชีพเป็นพนักงานบริษัทมากที่สุด รองลงมาประกอบอาชีพอิสระ นอกจากนี้มีงานวิจัยของภูริณัฐ ธนวิบูลย์ชัย [20] ศึกษาปัจจัยที่ส่งผลต่อการตัดสินใจเลือกซื้อที่อยู่อาศัยในเขตกรุงเทพมหานครและปริมณฑลของพนักงานบริษัทเอกชน

ที่พบว่า ปัจจัยที่ส่งผลต่อการตัดสินใจในการซื้อสังหาริมทรัพย์คือ ราคามีความเหมาะสมกับทำเลที่ตั้ง มีพื้นที่ใช้สอยตรงตามความต้องการ ราคาเหมาะสมกับขนาดที่ดินและพื้นที่ใช้สอย มีบริการให้คำปรึกษาเรื่องอัตราดอกเบี้ย เจ้าของโครงการมีความน่าเชื่อถือ มีของแถมต่าง ๆ ซึ่งสอดคล้องกับผลการวิจัยของ สรลรัชว์ สุรธรรมทวี และกฤษฎา พัชรวานิซ [21] ได้ทำการศึกษาปัจจัยที่มีอิทธิพลต่อการตัดสินใจซื้อคอนโดมิเนียมของผู้บริโภค : กรณีศึกษากรุงเทพมหานคร โดยใช้กลุ่มตัวอย่างที่พักอยู่ในเขตกรุงเทพมหานคร พบว่า ผู้บริโภคให้ความสำคัญกับปัจจัยด้านราคามากที่สุด ความสัมพันธ์ระหว่างปัจจัยทางการตลาดที่ส่งผลต่อการตัดสินใจซื้อพบว่า การซื้อจะเพิ่มขึ้นหากเป็นบริษัทที่มีชื่อเสียง ส่วนปัจจัยบุคคลและปัจจัยทางการเงินที่ส่งผลพบว่า การซื้อจะเพิ่มขึ้นหากผู้ซื้อไม่มีรายได้เพิ่มขึ้น และสามารถผ่อนชำระกับสถาบันการเงินได้

ธวัชชัย อมรพิสิทธิกุล และ สุกัญญา ภูสุวรรณรัตน์ [22] ได้ศึกษาถึงปัจจัยทางเศรษฐกิจที่มีผลกระทบต่อจำนวนห้องชุดที่จดทะเบียนในกรุงเทพมหานคร โดยได้นำเอาปัจจัยเศรษฐกิจ ได้แก่ ผลิตภัณฑ์มวลรวมในประเทศภาคก่อสร้าง สินเชื่อสังหาริมทรัพย์ผู้ประกอบการค่าง ดัชนีราคาวัสดุก่อสร้าง รายได้ต่อประชากร และอัตราดอกเบี้ยเงินกู้ยืมแบบมีระยะเวลา มาใช้ทำการศึกษาเพื่อให้ทราบว่าปัจจัยทางเศรษฐกิจด้านใดที่ส่งผลกระทบต่อจำนวนห้องชุดที่จดทะเบียนในกรุงเทพมหานคร พบว่า ปัจจัยทางเศรษฐกิจ คือ ผลิตภัณฑ์มวลรวมในประเทศภาคก่อสร้าง และอัตราดอกเบี้ยเงินกู้ยืมแบบมีระยะเวลา มีผลกระทบต่อจำนวนห้องชุดที่จดทะเบียนในกรุงเทพมหานครในทิศทางเดียวกัน ส่วนปัจจัยทางเศรษฐกิจดัชนีราคาวัสดุก่อสร้าง และรายได้ต่อประชากรมีผลกระทบต่อจำนวนห้องชุดที่จดทะเบียนในกรุงเทพมหานคร ในทิศทางตรงกันข้าม

นอกจากนี้ยังพบว่าความผันผวนของการจ้างงานกับจำนวนที่อยู่อาศัยได้มีผลกระทบต่อราคาที่อยู่อาศัยและความสามารถในการมีที่อยู่อาศัยด้วยเช่นกัน กรณีของเมือง Renu พบว่า การเติบโตของการจ้างงานมีความผันผวนมากกว่าจำนวนที่อยู่อาศัยที่สร้างขึ้น อย่างไรก็ตามจำนวน ที่อยู่อาศัย (Housing Units) ที่สร้างขึ้น ก็มีทิศทางไปทางเดียวกับอัตราการจ้างงาน จากภาวะเศรษฐกิจซบเซาได้ทำให้สูญเสียการจ้าง 25,000 อัตรา ทำให้จำนวนที่อยู่อาศัยลดลงอย่างมากด้วยเช่นกัน นอกจากนี้พบว่า ตลาดที่อยู่อาศัยสามารถฟื้นตัวได้ช้า ถึงแม้จะมีการจ้างมากขึ้นอย่างเห็นได้ชัดเจน ในช่วงปี พ.ศ. 2555 – 2558 การชะลอตัวของตลาดที่อยู่อาศัยทำให้นักลงทุนชะลอตัวในการลงทุนโครงการที่อยู่อาศัยด้วยเช่นกัน ด้วยเหตุนี้จึงทำให้ความต้องการที่อยู่อาศัย (Housing Demand) กับจำนวนที่อยู่อาศัย (Housing Units) มีความแตกต่างกัน จึงทำให้ที่อยู่อาศัยมีราคาสูงขึ้น อ้างจากสำนักงานศูนย์วิจัยและให้คำปรึกษา มหาวิทยาลัยธรรมศาสตร์ [23] ดังรูปที่ 2.11



รูปที่ 2.11 จำนวนที่อยู่อาศัยและอัตราการจ้างงานของเมือง Washoe County ปี ค.ศ.2001-2015 [23]

จากทฤษฎี และงานวิจัยดังกล่าวสามารถสรุปมาเป็นปัจจัยที่ส่งผลต่อความต้องการซื้อที่อยู่อาศัยได้เป็น 2 ส่วน คือปัจจัยทางด้านประชากรศาสตร์ และปัจจัยทางด้านเศรษฐกิจ ได้แก่ อัตราดอกเบี้ย ค่าธรรมเนียมการโอน ผลิตภัณฑ์มวลรวมของประเทศ ซึ่งได้สอดคล้องกับปัจจัยที่ทางผู้วิจัยจะนำมาเป็นปัจจัยในการบ่งบอกถึงแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยจากข้อมูลประชากรศาสตร์ ได้แก่ เพศ อายุ สถานภาพสมรส รายได้ครอบครัว ระยะเวลาที่ตัดสินใจซื้อ ราคาที่อยู่อาศัยที่ซื้อ ความสามารถในการผ่อนชำระ วัตถุประสงค์ที่ซื้อที่อยู่อาศัย ส่วนปัจจัยทางด้านเศรษฐกิจที่ผู้วิจัยจะไม่นำมาพิจารณาด้วยเนื่องจากข้อมูลที่น่ามาทำการวิจัยเป็นกลุ่มคนที่ตระหนักรู้ถึงความต้องการที่อยู่อาศัย และพิจารณาภาวะเศรษฐกิจก่อนเป็นลำดับแรกก่อนจะซื้อที่อยู่อาศัยอยู่แล้ว

ในการพัฒนาสร้างแบบจำลองผู้วิจัยต้องการศึกษาถึงปัจจัย (Attributes) ที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยด้วยจึงรวบรวมและสรุปปัจจัยที่มีผลต่อความต้องการซื้อที่อยู่อาศัยจากงานวิจัยเชิงเศรษฐศาสตร์ได้ดังนี้

ตารางที่ 2.1 ปัจจัยที่มีผลต่อความต้องการที่อยู่อาศัย

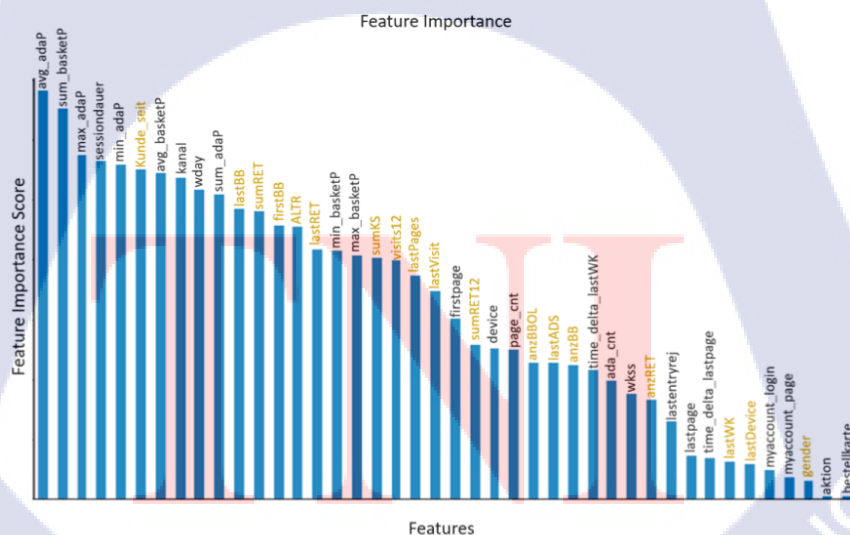
ลำดับ	ชื่อหัวข้อ	ปี	ปัจจัย
1	ปัจจัยที่ส่งผลต่อการตัดสินใจเลือกซื้ออสังหาริมทรัพย์ประเภททรัพย์สินรอการขายของธนาคารออมสิน ผู้เขียน: อนุฉัตร ชำของ และ จิรนนท์ อนุสรณ์พานิช [17]	2560	ปัจจัยทางประชากรศาสตร์ 1. อาชีพ 2. รายได้ 3. จำนวนสมาชิกในครอบครัว ปัจจัยทางเศรษฐกิจ 1. อัตราดอกเบี้ย 2. อัตราค่าธรรมเนียมการโอนกรรมสิทธิ์
2	พฤติกรรมผู้ซื้อทรัพย์สินรอการขาย กรณีศึกษาผู้เข้าชมทรัพย์สินรอการขายของสถาบันการเงินในงานบ้านและคอนโดปี 2557 ผู้เขียน: ศักดิ์สิทธิ์ มิตรเจริญถาวร [18]	2557	1. อาชีพ 2. รายได้ 3. ราคาที่อยู่อาศัยที่เหมาะสมกับรายได้
3	ศึกษาเปรียบเทียบปัจจัยในการเลือกอาคารชุดพักอาศัยเป็นบ้านหลังที่สองในเขตชั้นกลางของกรุงเทพมหานคร ผู้เขียน: อลงกรณ์ จาฟู [19]	2554	1. อาชีพ 2. รายได้ 3. ราคาที่อยู่อาศัยที่เหมาะสมกับรายได้
4	ปัจจัยที่มีผลต่อการตัดสินใจเลือกซื้อที่อยู่อาศัยในเขตกรุงเทพมหานครและปริมณฑล ผู้เขียน: ภูริณัฐ ธนวิบูลย์ชัย [20]	2557	1. ราคามีความเหมาะสมกับทำเลที่ตั้ง 2. พื้นที่ใช้สอยตรงตามความต้องการ 3. ราคาเหมาะสมกับขนาดที่ดิน 4. เจ้าของโครงการมีความน่าเชื่อถือ 5. มีบริการให้คำปรึกษาเรื่องอัตราดอกเบี้ย 4. มีของแถม

ตารางที่ 2.1 ปัจจัยที่มีผลต่อความต้องการที่อยู่อาศัย (ต่อ)

ลำดับ	ชื่อหัวข้อ	ปี	ปัจจัย
5	ปัจจัยที่มีอิทธิพลต่อการตัดสินใจซื้อคอนโดมิเนียมของผู้บริโภค กรณีศึกษา กรุงเทพมหานคร ผู้เขียน: สรลรัชว์ สุธรรมทวี และกฤษฎา พัชรวานิช [21]	2554	1. ราคาที่อยู่อาศัย 2. บริษัทที่มีชื่อเสียง 3. สามารถผ่อนชำระกับสถาบันการเงิน
6	ปัจจัยทางเศรษฐกิจที่มีผลกระทบต่อจำนวนห้องชุดที่จดทะเบียนใน กรุงเทพมหานคร ผู้เขียน: ธวัชชัย อมรพิสิทธิกุล และสุกัญญา ภูสุวรรณรัตน์ [22]	2557	1. ผลิตภัณฑ์มวลรวมในประเทศภาคก่อสร้าง 2. สินเชื่ออสังหาริมทรัพย์ผู้ประกอบการค้ำ 3. ดัชนีราคาวัสดุก่อสร้าง 4. รายได้ต่อประชากร 5. อัตราดอกเบี้ยเงินกู้ยืม
7	การคาดประมาณความต้องการที่อยู่อาศัยของประเทศไทยปี 2560 – 2580 ผู้เขียน: สำนักงานศูนย์วิจัยและให้คำปรึกษา มหาวิทยาลัยธรรมศาสตร์ [23]	2560	1. การจ้างงาน 2. จำนวนประชากร 3. อายุของประชากร 4. อัตราการย้ายถิ่นฐาน 6. ความสามารถในการมีที่อยู่อาศัย

2.2.2 งานวิจัยเชิงเทคนิคในการพัฒนาแบบจำลอง

ที่ผ่านมาผู้วิจัยได้นำเทคนิคการพยากรณ์ด้วยการเรียนรู้ของเครื่องไปประยุกต์ใช้ในการพยากรณ์เพื่อการวางแผนธุรกิจดังเช่นงานวิจัยของ Hannah Sophia Seippel [24] ได้ทำการศึกษาพฤติกรรมของลูกค้าออนไลน์ในการซื้อสินค้าในเว็บไซต์ร้านค้าปลีกเสื้อผ้ารายใหญ่ของเยอรมันเพื่อเพิ่มโอกาสในการแข่งขันการตลาดในโลกของอีคอมเมิร์ซ วัตถุประสงค์เพื่อการจัดประเภทของลูกค้าเป็นผู้ซื้อ และไม่ซื้อ หากแบ่งกลุ่มลูกค้าเป็นประเภทผู้ซื้อก็จะให้บริการที่มีคุณภาพสูงขึ้นแก่ลูกค้าที่มีแนวโน้มจะซื้อมากขึ้นเพื่อจูงใจให้ตัดสินใจซื้อเร็วขึ้น ข้อมูลที่ใช้คือข้อมูลคลิกสตรีมจากเว็บไซต์ของร้านค้า โดยการพัฒนาซอฟต์แวร์เพื่อนำมาบันทึกข้อมูลและจัดการเก็บการโต้ตอบของผู้เยี่ยมชมกับร้านค้าบนเว็บโดยละเอียด และใช้ปัจจัย (Attribute) 43 ตัวที่มีความสำคัญต่อการซื้อ และไม่ซื้อสินค้า ดังรูปที่ 2.7 และใช้อัลกอริทึม Boosted RF FNN SVM LR และ RNN ในการทำเครื่องมือพยากรณ์ จากรูปที่ 2.8 ผลการศึกษาเปรียบเทียบอัลกอริทึมที่ใช้ในการสร้างเครื่องมือพยากรณ์การซื้อ อัลกอริทึม RF มีประสิทธิภาพมากที่สุดในการทำนายผล มีค่าความแม่นยำ (Accuracy) สูงที่สุดเท่ากับ 0.76 ค่าความถ่วงดุล (F1 – Score) เท่ากับ 0.83 นอกจากนี้ยังเป็นอัลกอริทึมเดียวที่สามารถทำงานได้ดีกว่าแบบจำลองพื้นฐานซึ่งเป็นโครงสร้างที่ได้รับการปรับปรุงเกี่ยวกับเมตริกทั้งหมด ส่วน SVM มีประสิทธิภาพที่แย่ที่สุดทำให้ได้ความแม่นยำต่ำสุด



รูปที่ 2.12 ปัจจัย (Attribute) ที่สำคัญ [24]

All Data		Threshold: 0.5				
Algorithm	Class	Precision	Recall	F1-score	Accuracy	Roc Auc
Boosted tree	no buy	0.79	0.86	0.82		
	buy	0.62	0.5	0.56		
	total	0.73	0.75	0.74	0.75	0.79
RF	no buy	0.8	0.86	0.83		
	buy	0.64	0.64	0.59		
	total	0.75	0.75	0.75	0.76	0.82
FNN	no buy	0.78	0.84	0.81		
	buy	0.6	0.5	0.55		
	total	0.72	0.73	0.72	0.73	0.73
SVM	no buy	0.7	0.96	0.81		
	buy	0.51	0.1	0.17		
	total	0.64	0.68	0.6	0.68	0.67
LR	no buy	0.86	0.71	0.78		
	buy	0.56	0.76	0.64		
	total	0.76	0.73	0.74	0.73	0.8
RNN	no buy	0.78	0.83	0.81		
	buy	0.56	0.49	0.52		
	total	0.72	0.72	0.72	0.72	0.74

รูปที่ 2.13 ผลการศึกษาเปรียบเทียบเครื่องมือพยากรณ์ [24]

นอกจากนี้ยังมีงานวิจัยของ Saavi Stubseid and Ognjen Arandjelovic' [25] ได้ทำการศึกษาพฤติกรรมของลูกค้าที่ซื้อของออนไลน์ในประเทศสหราชอาณาจักร วัตถุประสงค์ของงานวิจัยเพื่อต้องการพยากรณ์ว่าลูกค้าคนไหนมีแนวโน้มที่จะตัดสินใจซื้อหรือไม่ซื้อสินค้าออนไลน์เพื่อสร้างโอกาสทางธุรกิจในการเพิ่มยอดขาย โดยพิจารณาจาก 72 ปัจจัย (Attribute) ที่เข้ารหัสข้อมูลไว้และใช้ข้อมูลการเข้ามาซื้อออนไลน์ในแต่ละเดือนรวมทั้งหมด 642,709 รายการดังแสดงในรูปที่ 2.9 และใช้อัลกอริทึม Naïve Bayes และ Random Forest ในการใช้สร้างเครื่องมือพยากรณ์เปรียบเทียบ ดังแสดงรูปที่ 2.10 จากผลการศึกษาพบว่า อัลกอริทึมของ Random Forest มีประสิทธิภาพในการพยากรณ์มากกว่า Naïve Bayes โดยมีค่าความถูกต้องเท่ากับ 0.72 และมีความคล่องตัว เท่ากับ 0.72 ซึ่งหมายความว่า การสร้างเครื่องมือเพื่อพยากรณ์ลูกค้ามีแนวโน้มที่จะตัดสินใจซื้อสินค้าจากอัลกอริทึมของ Random Forest ให้ผลที่แม่นยำมีประสิทธิภาพมากที่สุด

Transaction #	Feature 1 (C_1)	Feature 2 (C_2)	...	Feature 72 (C_{72})	Consumer decision (C)
1	BC5F4DF1E7	1582934400	...	-0.1216277	No purchase (0)
2	0AB04FC49F	1585612800	...	-0.5361754	Purchase (1)
...					
642,709	055D5DBE79	1596153600	...	+0.56486328	Purchase (1)

รูปที่ 2.14 ปัจจัย (Attribute) ที่ใช้และเข้ารหัสแฮช [25]

Measure	Naïve Bayes	Random forest
Accuracy	0.66	0.72
AUC	0.71	0.79
F1-score	0.66	0.72

รูปที่ 2.15 ผลการศึกษาเปรียบเทียบเครื่องมือพยากรณ์ [25]

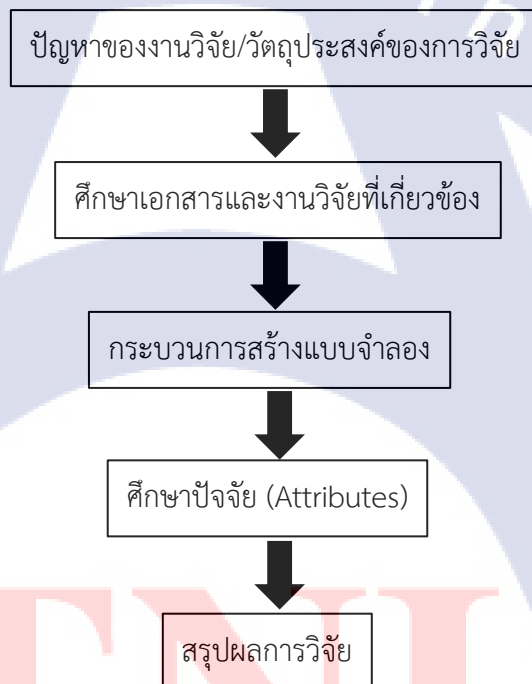
ในอุตสาหกรรมสินเชื่อก็นำเทคนิคการเรียนรู้ของเครื่องเข้ามาประยุกต์ใช้ในการวางแผนธุรกิจเช่นกันดังเช่นงานวิจัยของ ไพจิตร สุขสมบูรณ์ และณิชา นภาพร จงกะสิกิจ [26] ได้ทำการศึกษาเรื่องผลของตัวแบบที่ส่งผลต่อความสามารถในการชำระสินเชื่อ ผู้วิจัยได้นำเทคนิคทางด้านเหมืองข้อมูล มาประยุกต์ใช้ระบุตัวแปรที่เกี่ยวข้องกับความสามารถในการชำระสินเชื่อของสมาชิกสหกรณ์เครดิตยูเนียน มีวัตถุประสงค์เพื่อทำการค้นหาว่าควรจะต้องพิจารณาอะไรบ้างในการอนุมัติสินเชื่อ ซึ่งจะส่งผลให้มีอัตราการเกิดหนี้สูญลดลง โดยใช้ข้อมูลจากสหกรณ์เครดิตยูเนียนในเขตอำเภอเมือง จังหวัดลำปาง จำนวน 5 แห่ง พบว่าสหกรณ์แต่ละแห่งจะมีหลักเกณฑ์ในการให้อนุมัติสินเชื่อที่คล้ายคลึงกัน โดยสามารถกำหนดเป็นปัจจัยที่เกี่ยวข้องในการพิจารณาสินเชื่อจำนวน 19 ปัจจัย และนำมาวิเคราะห์โดยวิธีการ Evolutionary Selection พบว่ามีปัจจัยที่มีผลกระทบโดยตรงกับการชำระสินเชื่อ ของสมาชิกเพียง 12 ปัจจัย จากนั้นจึงนำปัจจัยที่ได้มาทำการสร้างตัวแบบ ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเบย์เซียนแบบง่าย (Naive Bayes) และเทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) และทดสอบด้วยวิธีการทดสอบแบบไขว้ (Cross Validation) แล้วทำการวัดค่าความถูกต้อง (Correct) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F1 -Score) โดยในการทดสอบเพื่อหาค่าตัวแบบที่เหมาะสมจะนำไปใช้ได้ใช้ข้อมูลตัวอย่างจำนวน 200 ตัวอย่าง จากผลการศึกษาพบว่าตัวแบบที่สร้างขึ้นด้วยวิธีการต้นไม้

ตัดสินใจ (Decision Tree) ให้ผลลัพธ์ในการทดสอบที่ดีที่สุด โดยมีค่าความถูกต้องที่ร้อยละ 95.96 มีค่าความแม่นยำที่ร้อยละ 95.71 มีค่าความระลึก ที่ร้อยละ 91.67 และมีค่าความถ่วงดุลที่ร้อยละ 91.94 ทั้งนี้เมื่อนำไปใช้ในการทดสอบกับข้อมูลสมาชิกที่ขอสินเชื่อและติดตามข้อมูลการชำระสินเชื่อพบว่า มีสมาชิกเพียง 2 ราย จาก 5 ราย ที่ได้รับการประเมินว่ามีโอกาสเกิดหนี้สูญเท่านั้นที่จะมียอดค้างชำระสินเชื่อ

จากการศึกษางานวิจัยที่เกี่ยวข้องจะเห็นว่าการนำเทคนิคการเรียนรู้ของเครื่องไปประยุกต์ใช้ในการพยากรณ์แนวโน้มการตัดสินใจของผู้บริโภคในการซื้อสินค้าออนไลน์ และนำไปประยุกต์ใช้กับการพยากรณ์สินเชื่อว่าจะเกิดแนวโน้มหนี้ดี หรือหนี้เสีย เพื่อเพิ่มโอกาสในธุรกิจนั้น ๆ ผู้วิจัยจึงได้แนวคิดและนำมาประยุกต์ใช้ในอุตสาหกรรมอสังหาริมทรัพย์ โดยจะทำการศึกษาแนวโน้มในการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรรรรมบ้านและคอนโด เพื่อจะได้เพิ่มโอกาสในการทำการตลาดให้มีประสิทธิภาพมากขึ้น โดยจากการศึกษางานวิจัยที่เกี่ยวข้องจะเห็นว่าอัลกอริทึม Random Forest เป็นอัลกอริทึมที่มีประสิทธิภาพดีที่สุดในการพยากรณ์ซึ่งเป็นประเภทของอัลกอริทึม (Algorithm) ที่เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) ผู้วิจัยจึงมีความสนใจที่จะศึกษาอัลกอริทึม Random Forest รวมถึงอัลกอริทึมต่าง ๆ ที่อยู่ในกลุ่มนี้ด้วย เพื่อต้องการหาแบบจำลองที่มีประสิทธิภาพและเหมาะสมกับชุดข้อมูลที่ผู้วิจัยนำมาศึกษาในครั้งนี้

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยนี้มีเป้าหมายเพื่อพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย พร้อมทั้งเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด และศึกษาถึงปัจจัย (Attributes) ที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด ซึ่งมีขั้นตอนการดำเนินงานดังรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนการดำเนินงาน

ตารางที่ 3.1 ระยะเวลาการดำเนินการวิจัย

การดำเนินงาน	ระยะเวลา									
	ปี 2563							ปี 2564		
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.
ขั้นตอนที่ 1 ปัญหาของ งานวิจัย/ วัตถุประสงค์ ของการวิจัย										
ขั้นตอนที่ 2 ศึกษาเอกสาร และงานวิจัยที่ เกี่ยวข้อง										
ขั้นตอนที่ 3 กระบวนการ สร้าง แบบจำลอง										
ขั้นตอนที่ 4 ศึกษาปัจจัย (Attributes)										
ขั้นตอนที่ 5 สรุป ผลการวิจัย										

3.1 ปัญหาของงานวิจัย หรือวัตถุประสงค์ของการวิจัย

จากที่หน่วยงานศูนย์ข้อมูลสังหาริมทรัพย์ได้ทำงานร่วมกับสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยในการเก็บรวบรวมข้อมูลผู้เช่าชมงานมหกรรมบ้านและคอนโด และวิเคราะห์ข้อมูลลูกค้าให้กับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมฯ โดยใช้การวิเคราะห์สถิติแบบพรรณนา เป็นระยะเวลาตั้งแต่ปี 2558 จนถึงปี 2562 ผู้ประกอบการได้นำข้อมูลดังกล่าวไปทำการตลาดแต่ไม่เกิดประสิทธิภาพเท่าที่ควร และยังต้องใช้งบประมาณมากในการทำการตลาด

และผู้วิจัยได้มีโอกาสทำงานโดยตรงกับทางสมาคมฯ จึงคิดหาแนวทางการวิเคราะห์ข้อมูลใหม่เพื่อให้ผู้ประกอบการนำไปใช้ให้เกิดประโยชน์สูงสุด และเพื่อเพิ่มโอกาสในการขายที่อยู่อาศัยให้มีประสิทธิภาพสูงสุดอีกด้วย ด้วยเหตุผลนี้ผู้วิจัยจึงมีความสนใจที่จะวิเคราะห์ข้อมูลลูกค้าโดยนำเทคนิคการเรียนรู้ของเครื่อง มาประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้ที่เข้ามาชมงานมหกรรมบ้านและคอนโด โดยประเมินว่าลูกค้าคนไหนมีแนวโน้มที่จะซื้อที่อยู่อาศัย หรือมีแนวโน้มที่จะไม่ซื้อที่อยู่อาศัย เพื่อให้ผู้ประกอบการอสังหาริมทรัพย์ นำไปวางแผนการตลาดในการขายที่อยู่อาศัย ให้ตรงกลุ่มเป้าหมายสำหรับลูกค้าที่แนวโน้มจะซื้อที่อยู่อาศัย โดยจัดโปรโมชั่น หรือให้ของแถมต่าง ๆ เพื่อจูงใจให้ลูกค้าตัดสินใจซื้อได้อย่างรวดเร็วขึ้น

3.2 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์ในการวิจัยเพื่อเปรียบเทียบแบบจำลอง เพื่อสร้างแบบจำลอง และเพื่อศึกษาปัจจัยที่มีนัยสำคัญต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสรุปได้ว่า ปัจจัยที่ส่งผลกระทบต่อความต้องการซื้อที่อยู่อาศัยแบ่งเป็น 2 ส่วน คือปัจจัยทางด้านประชากรศาสตร์ และปัจจัยทางด้านเศรษฐกิจ ได้แก่ อัตราดอกเบี้ย ค่าธรรมเนียมการโอน ผลผลิตมวลรวมของประเทศ ดังนั้นผู้วิจัยจึงนำปัจจัยทางด้านประชากรศาสตร์มาศึกษาปัจจัยที่มีนัยสำคัญต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัย ได้แก่ เพศ อายุ อาชีพ สถานภาพสมรส รายได้ครอบครัว ระยะเวลาที่ตัดสินใจซื้อ ราคาที่อยู่อาศัยที่ซื้อ ความสามารถในการผ่อนชำระ และวัตถุประสงค์ที่ซื้อที่อยู่อาศัย ส่วนปัจจัยทางด้านเศรษฐกิจที่ผู้วิจัยจะไม่นำมาพิจารณาด้วยเนื่องจากข้อมูลที่นำมาทำการวิจัยเป็นกลุ่มคนที่ตระหนักรู้ถึงความต้องการที่อยู่อาศัย และพิจารณาภาวะเศรษฐกิจก่อนเป็นลำดับแรกก่อนที่จะตัดสินใจซื้อที่อยู่อาศัยอยู่แล้ว จากการศึกษาเทคนิคการเรียนรู้ของเครื่องคอมพิวเตอร์ ในงานวิจัยนี้เหมาะสำหรับประเภทการเรียนรู้แบบมีผู้สอน (Supervised Learning) เนื่องจากมีข้อมูลที่บ่งบอกลักษณะของคนที่ตัดสินใจซื้อที่อยู่อาศัย และข้อมูลลักษณะของคนที่ไม่ตัดสินใจซื้อที่อยู่อาศัยว่ามีลักษณะอย่างไรบ้าง ในการแบ่งข้อมูลเพื่อใช้ในการสร้างแบบจำลองงานวิจัยนี้ใช้วิธีการไขว้ข้อมูลออกเป็น 10 ชุดข้อมูลเท่ากัน (10-Fold Cross-Validation) ทำการแบ่งเป็นข้อมูลฝึกฝน (Training Data) ใช้ข้อมูลจำนวน 9 ชุด และข้อมูลทดสอบ (Test Data) ใช้ข้อมูลจำนวน 1 ชุด โดยเลือกใช้การทดลองด้วย 4 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเบย์เซียนแบบง่าย (Naïve Bayes) เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) และเทคนิคแรนดอมฟอเรส (Random Forest) เพื่อเป็นการเพิ่มประสิทธิภาพของแบบจำลองให้ดีขึ้น ดังนั้นงานวิจัยนี้จึงนำเทคนิคการสร้างแบบจำลองด้วยเทคนิค Ensemble และใช้วิธีการสร้างแบบจำลองด้วยเทคนิค Voting เนื่องจากใช้แบบจำลองที่

มีประสิทธิภาพมากที่สุดมาช่วยกันในการพยากรณ์ เพื่อให้แบบจำลองพยากรณ์ได้มีความถูกต้องมากขึ้น ส่วนการประเมินประสิทธิภาพของแบบจำลองนั้น ในงานวิจัยนี้จะใช้ตารางสำคัญในการวัดความสามารถของการเรียนรู้ของเครื่อง (Confusion Matrix) ที่นิยมใช้วัดประสิทธิภาพของแบบจำลอง ซึ่งประกอบด้วย 4 ด้านได้แก่ ด้านค่าความถูกต้อง (Accuracy) ด้านค่าความแม่นยำ (Precision) ด้านค่าความระลึก (Recall) และด้านค่าถ่วงดุล (F1-Score)

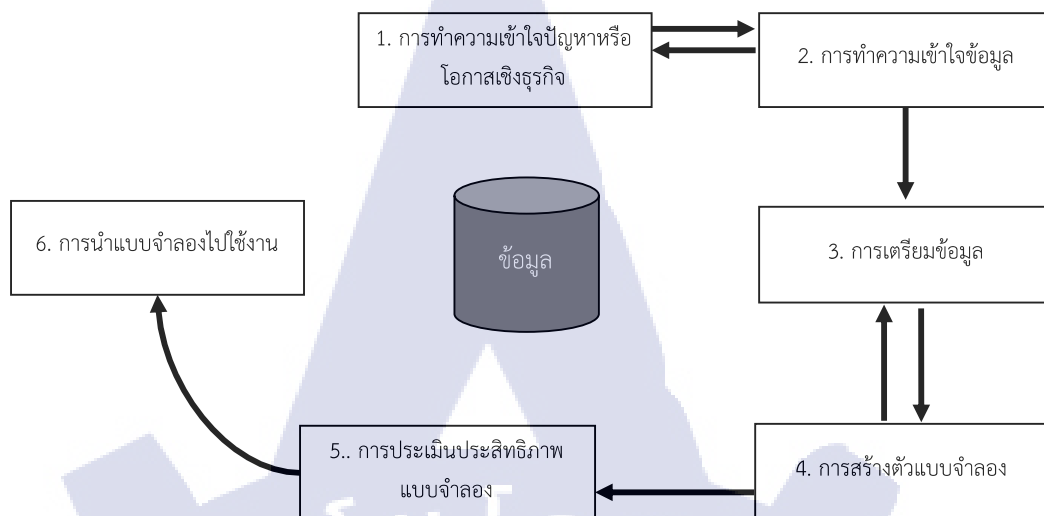
3.3 กระบวนการสร้างแบบจำลอง

ผู้วิจัยได้เก็บรวบรวมข้อมูลจากงานมหกรรมบ้านและคอนโดที่จัดขึ้นโดยสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย ในบริเวณหอแสดงสินค้าศูนย์ประชุมสิริกิติ์กรุงเทพมหานคร เป็นข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหกรรมบ้านและคอนโดโดยใช้ข้อมูลตั้งแต่ปี 2557 ถึงปี 2562 มีจำนวน 118,197 รายการ และบันทึกข้อมูลอยู่ในรูปแบบไฟล์นามสกุล .xlsx ดังรูปที่ 3.2

GuestID	ExpoNo	FirstName	LastName	Gender	Married	Age	Carreer	FamilyIncome	FExpectBuy	FReason	FBudget	FCanpay	Label
9000096	2558			2	1	3	4	5	5	1	6	3	Y
9000101	2558			1	1	2	7	1	7	2	2	1	N
9000111	2558			1	2	5	1	4	1	4	5	4	Y
9000115	2558			1	2	2	3	2	2	2	1	1	Y
9000120	2558			2	1	2	1	2	6	6	3	2	Y
9000116	2558			1	2	2	2	6	1	1	8	5	N
9000121	2558			1	1	3	1	6	2	5	3	3	Y
9000127	2558			1	2	3	1	2	4	6	3	2	Y
9000132	2558			1	1	4	1	2	5	2	4	2	Y
9000138	2558			1	1	3	1	3	5	5	3	2	Y
9000139	2558			1	2	5	7	5	1	3	4	5	Y
9000234	2558			1	1	5	3	2	1	3	1	1	Y
9000235	2558			2	2	5	2	3	4	3	2	1	N
9000236	2558			1	1	5	3	6	1	3	2	2	Y
9000237	2558			2	1	3	1	5	4	5	3	5	Y
9000238	2558			2	1	3	5	2	3	2	2	1	N
9000240	2558			2	1	3	1	2	3	3	2	1	Y
9000241	2558			2	1	3	1	1	4	2	2	1	Y
9000242	2558			1	2	4	3	4	3	3	4	2	Y
9000243	2558			2	1	2	1	6	5	5	1	1	Y
9000244	2558			1	1	3	1	4	1	1	2	2	Y
9000245	2558			1	1	3	1	1	5	2	2	1	Y

รูปที่ 3.2 ตัวอย่างข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหกรรมบ้านและคอนโด ปี 2557 – 2562
ที่บันทึกลง Excel

จากการศึกษาแบบจำลอง CRISP-DM ข้างต้นผู้วิจัยได้นำมาปรับใช้ตามความเหมาะสมในขั้นตอนการสร้างเครื่องมือพยากรณ์ดังรูปที่ 3.3



รูปที่ 3.3 ขั้นตอนการสร้างแบบจำลอง

3.3.1 การทำความเข้าใจปัญหาหรือโอกาสเชิงธุรกิจ

ผู้วิจัยได้ทำการศึกษาถึงโอกาสเชิงธุรกิจโดยนำข้อมูลผู้เข้าชมงานมหกรรมบ้านและคอนโด มาสร้างแบบจำลองการพยากรณ์แนวโน้มการตัดสินใจซื้อ เพิ่มโอกาสในการขายที่อยู่อาศัยของผู้ประกอบการให้มีโอกาสมากขึ้น เพื่อนำไปวางแผนทำการตลาดได้ตรงกลุ่มเป้าหมาย

3.3.2 การทำความเข้าใจข้อมูล

ในขั้นตอนนี้เป็นการรวบรวมข้อมูลที่จะใช้ในการสร้างแบบจำลอง และดูความเหมาะสม ความสมบูรณ์ของข้อมูล โดยใช้ข้อมูลผู้ที่มาลงทะเบียนเข้าชมงานมหกรรมบ้านและคอนโด ตั้งแต่ปี 2557 ถึงปี 2562 มีจำนวนทั้งหมด 10 ปัจจัย โดยปัจจัยที่ 1 – 9 เป็นปัจจัยที่ใช้สร้างแบบจำลอง ส่วนปัจจัยตัวที่ 10 เป็นปัจจัยที่แสดงผล (Label) โดยมีโครงสร้างข้อมูลดังตารางที่ 3.2

ตารางที่ 3.2 แสดงโครงสร้างข้อมูลที่จัดเก็บรวบรวมในงานมหกรรมบ้านและคอนโด

ชื่อปัจจัย	คำอธิบาย	จำนวน class	รายละเอียด class
1. Gender	ข้อมูลเพศ	2	1 = ชาย 2 = หญิง
2. Married	ข้อมูลสถานะภาพสมรส	2	1 = โสด 2 = สมรส

ตารางที่ 3.2 แสดงโครงสร้างข้อมูลที่จัดเก็บรวบรวมในงานมหกรรมบ้านและคอนโด (ต่อ)

ชื่อปัจจัย	คำอธิบาย	จำนวน class	รายละเอียด class
3. Age	ข้อมูลอายุ	6	1 = ต่ำกว่า 21 ปี 2 = 21-30 ปี 3 = 31-40 ปี 4 = 41-50 ปี 5 = 51-60 ปี 6 = มากกว่า 60 ปี
4. Career	ข้อมูลอาชีพ	6	1 = พนักงานบริษัทเอกชน 2 = รับราชการ 3 = พนักงานรัฐวิสาหกิจ 4 = เจ้าของกิจการ 5 = ค้าขาย 6 = วิชาชีพอิสระ เช่น แพทย์ วิศวกร สถาปนิก
5. Family Income	ข้อมูลรายได้ ครอบครัว ต่อเดือน	7	1 = ไม่เกิน 15,000 บาท 2 = 15,001 - 30,000 บาท 3 = 30,001 - 50,000 บาท 4 = 50,001 - 70,000 บาท 5 = 70,001 - 100,000 บาท 6 = 100,001 - 150,000 บาท 7 = เกิน 150,000 บาท
6. FExpectBuy	ระยะเวลาที่ คิดว่าจะซื้อ	7	1 = ซื้อภายในงาน 2 = ภายใน 1 เดือน 3 = 1 - 3 เดือน 4 = 3 - 6 เดือน 5 = 6 - 12 เดือน 6 = 1 - 2 ปี 7 = เกิน 2 ปี

ตารางที่ 3.2 แสดงโครงสร้างข้อมูลที่จัดเก็บรวบรวมในงานมหกรรมบ้านและคอนโด (ต่อ)

ชื่อ Attribute	คำอธิบาย	จำนวน class	รายละเอียด class
7. FReason	เหตุผลที่ ต้องการที่อยู่ ใหม่	6	1 = ต้องการแยกครอบครัว/แต่งงาน 2 = ต้องการมีที่อยู่อาศัยของตนเอง 3 = ต้องการความสะดวกในการเดินทาง 4 = ต้องการซื้อเพื่อเป็นทรัพย์สิน 5 = ต้องการซื้อเพื่อลงทุน/ให้เช่า 6 = ต้องการสภาพแวดล้อมที่ดีขึ้น
8. FBudget	งบประมาณ	8	1 = ไม่เกิน 1 ล้านบาท 2 = 1 - 1.99 ล้านบาท 3 = 2 - 2.99 ล้านบาท 4 = 3 - 3.99 ล้านบาท 5 = 4 - 5.99 ล้านบาท 6 = 6 - 9.99 ล้านบาท 7 = 10 - 14.99 ล้านบาท 8 = เกิน 15 ล้านบาท
9. FCanpay	ความสามารถ ในการผ่อน ชำระต่อเดือน	5	1 = ไม่เกิน 10,000 บาท 2 = 10,001 - 20,000 บาท 3 = 20,001 - 30,000 บาท 4 = เกิน 30,000 บาท 5 = ไม่ผ่อน - ชื้อเงินสด
10. Label	ผลการ ตัดสินใจซื้อ	2	Y = ซื้อ N = ไม่ซื้อ

Open in Turbo Prep Auto Model

Filter (118,197 / 118,197 examples):

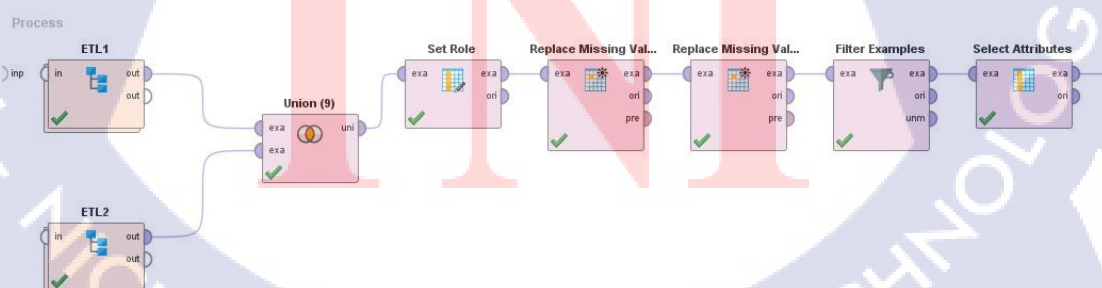
Row No.	Gender	Married	Age	Career	FamilyInco...	FExpectBuy	FReason	FBudget	FCanpay	Label
1	2	1	3	4	5	5	1	6	3	Y
2	1	1	2	7	1	7	2	2	1	N
3	1	2	5	1	4	1	4	5	4	Y
4	1	2	2	3	2	2	2	1	1	Y
5	2	1	2	1	2	6	6	3	2	Y
6	1	2	2	2	6	1	1	8	5	N
7	1	1	3	1	6	2	5	3	3	Y
8	1	2	3	1	2	4	6	3	2	Y
9	1	1	4	1	2	5	2	4	2	Y
10	1	1	3	1	3	5	5	3	2	Y
11	1	2	5	7	5	1	3	4	5	Y
12	1	1	5	3	2	1	3	1	1	Y
13	2	2	5	2	3	4	3	2	1	N
14	1	1	5	3	6	1	3	2	2	Y
15	2	1	3	1	5	4	5	3	5	Y
16	2	1	3	5	2	3	2	2	1	N
17	2	1	3	1	2	3	3	2	1	Y
18	2	1	3	1	1	4	2	2	1	Y
19	1	2	4	3	4	3	3	4	2	Y
20	2	1	2	1	6	5	5	1	1	Y

ExampleSet (118,197 examples, 0 special attributes, 14 regular attributes)

รูปที่ 3.4 แสดงข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหกรรมบ้านและคอนโด ปี 2557 – 2562 ในโปรแกรม RapidMiner Studio Education

3.3.3 การเตรียมข้อมูล

นำข้อมูลผู้ที่มาลงทะเบียนเข้ามาชมงานมหกรรมบ้านและคอนโด ตั้งแต่ปี 2557 ถึง ปี 2561 มีจำนวน 113,737 รายการมาทำความสะอาดข้อมูล ดูความสมบูรณ์ ความครบถ้วนของข้อมูล เพื่อเตรียมไปใช้ในการสร้างแบบจำลอง โดยมีขั้นตอนการเตรียมข้อมูลดังรูปที่ 3.5



รูปที่ 3.5 ขั้นตอนการเตรียมข้อมูล

ในการเตรียมข้อมูลมีขั้นตอนการทำงานดังนี้

1. ทำการรวมข้อมูลตั้งแต่ปี 2557 ถึง 2561 โดยการนำ Extract-Transform-Load (ETL) เนื่องจากข้อมูลถูกจัดเก็บไว้หลายไฟล์ โดยในขั้นตอนนี้ ต้องทำการปรับแต่งข้อมูลในแต่ละปี ให้มีรูปแบบการจัดเก็บที่เหมือนกันก่อน แล้วจึงนำมารวมไว้ในไฟล์เดียวกัน

2. แทนที่ข้อมูลที่จัดเก็บไม่ครบ (Missing Data) โดยกำหนดให้มีการแทนค่าข้อมูลที่ไม่ว่างมีอยู่ 2 กรณี

2.1 ข้อมูลใน Attribute นั้นถูกจัดเก็บเป็นตัวเลขที่เป็นแบบนามบัญญัติ (Nominal Scale) ให้แทนค่าด้วย 0

2.2 ข้อมูลใน Attribute นั้นถูกจัดเก็บเป็นตัวหนังสือ (Text) ให้แทนค่าด้วย N/A

3. ทำความสะอาดข้อมูล (Cleansing Data) ในขั้นตอนนี้จะเลือกข้อมูลในจำนวนแถวที่มีความครบถ้วนสมบูรณ์ โดยตัดแถวที่มีข้อมูลเป็น 0 หรือ N/A ทิ้งไป

4. ได้ข้อมูลที่มีความครบถ้วนสมบูรณ์จำนวน 113,554 รายการ จากจำนวน 113,737 รายการ คิดเป็นร้อยละ 99.8 ของข้อมูลทั้งหมดที่สามารถนำไปสร้างแบบจำลองได้ ดังตารางที่ 3.3

ตารางที่ 3.3 แสดงข้อมูลปี 2557 – 2561 เป็นข้อมูลที่ทำความสะอาดและเป็นข้อมูลที่นำมาสร้างแบบจำลอง

ปี	จำนวนข้อมูลก่อนทำความสะอาด (หน่วย : รายการ)	จำนวนข้อมูลที่ทำความสะอาดแล้วในแต่ละปี (หน่วย : รายการ)	สัดส่วน % ข้อมูลที่นำไปสร้างแบบจำลอง
2557	21,205	21,165	99.8%
2558	25,180	25,149	99.9%
2559	26,833	26,757	99.7%
2560	19,586	19,564	99.9%
2561	20,933	20,919	99.9%
รวม	113,737	113,554	99.8%

3.3.4 การสร้างตัวแบบจำลอง

ในการสร้างแบบจำลอง ดำเนินการโดยใช้ซอฟต์แวร์ RapidMiner Studio Education 9.8.001 โดยใช้ข้อมูลจากการเตรียมข้อมูลในขั้นตอน 3.3.3 ที่มีจำนวน 118,197 รายการ แบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 ข้อมูลปี 2557 – 2561 มีจำนวน 113,737 รายการ เป็นข้อมูลที่จะ

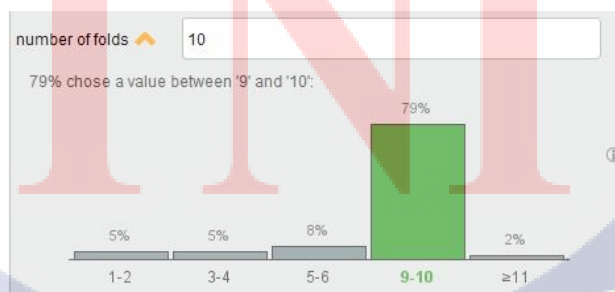
นำมาใช้ในการสร้างแบบจำลอง ส่วนที่ 2 ข้อมูลปี 2562 มีจำนวน 4,460 รายการ เป็นข้อมูลที่จะใช้ในการทดลองนำแบบจำลองที่มีประสิทธิภาพมากที่สุดมาทดสอบใช้งาน (Deployment)

เมื่อนำข้อมูลในส่วนที่ 1 มาทำความสะอาดแล้วได้ข้อมูลที่มีความครบถ้วนสมบูรณ์จำนวน 113,554 รายการ ที่สามารถนำมาสร้างแบบจำลองการพยากรณ์แนวโน้มการตัดสินใจซื้อที่อยู่อาศัย ดังตารางที่ 3.4

ตารางที่ 3.4 การเตรียมข้อมูลเพื่อการพัฒนาแบบจำลอง

ครั้งที่สร้างจำลอง	ข้อมูลที่ใช้สร้างแบบจำลอง (Data Set)	จำนวนข้อมูล (หน่วย : รายการ)
แบบจำลองครั้งที่ 1	ข้อมูลระหว่าง 2557- 2561	113,554
แบบจำลองครั้งที่ 2	ข้อมูลระหว่าง 2558- 2561	92,389
แบบจำลองครั้งที่ 3	ข้อมูลระหว่าง 2559- 2561	67,240
แบบจำลองครั้งที่ 4	ข้อมูลระหว่าง 2560- 2561	40,483
แบบจำลองครั้งที่ 5	ข้อมูล 2561	20,919

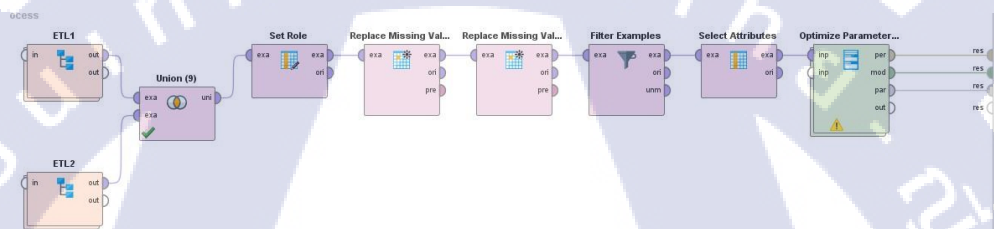
นำข้อมูลที่ทำความสะอาด และมีความครบถ้วนสมบูรณ์แล้ว มาแบ่งชุดข้อมูลสำหรับการสร้างแบบจำลอง ด้วยวิธีการ Cross-validation แบ่งชุดข้อมูลออกเป็น 10 ชุด เนื่องจากเป็นจำนวน k ที่ดีที่สุด จะเห็นว่าเมื่อแบ่งจำนวนชุดข้อมูลออกเป็น 9 – 10 ชุด จะทำให้แบบจำลองมีความถูกต้องมากถึงร้อยละ 79 เนื่องจากเป็นวิธีการที่ลดการโน้มเอียงของผลการทดสอบของแบบจำลองให้มีความใกล้เคียงกันในชุดข้อมูลที่ใช้ในการสอน และข้อมูลที่ใช้ในการทดสอบให้มีความเสมอภาคกัน ดังรูปที่ 3.6



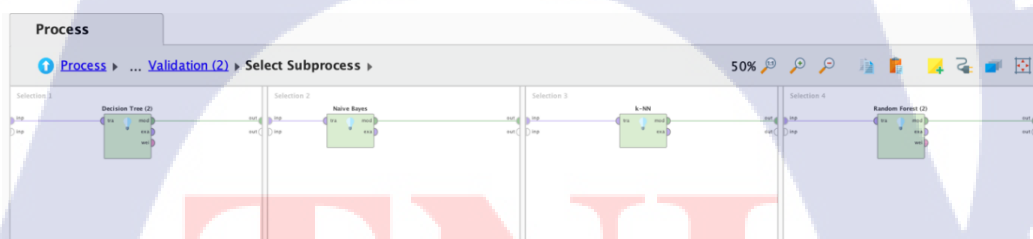
รูปที่ 3.6 แสดงจำนวนการแบ่งชุดข้อมูล

การสร้างแบบจำลองจะทำการทดลองโดยแบ่งการสร้างแบบจำลองออกเป็น 5 ครั้ง ในแต่ละครั้งจะลดข้อมูลในการสร้างแบบจำลองตัดออกไปครั้งละ 1 ปี เป็นเทคนิคการพยากรณ์ข้อมูลที่ให้ความสำคัญกับข้อมูลที่ใกล้กับข้อมูลในปัจจุบันมากที่สุด หรือเรียกว่าเทคนิคการพยากรณ์ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก [26] ที่ให้ความสำคัญกับข้อมูลที่ใกล้กับปัจจุบันโดยให้น้ำหนักกับข้อมูลในปัจจุบันจากมากไปหาน้อยที่เป็นข้อมูลในอดีต จะเห็นว่าการสร้างแบบจำลองจะเริ่มจากการใช้ข้อมูลทั้งหมด 5 ปี ที่มีความครบถ้วนสมบูรณ์แล้วมาสร้างแบบจำลอง ดังตารางที่ 3.4

ทำการสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคเบย์เซียนแบบง่าย (Naïve Bayes) เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) และเทคนิคแรนดอมฟอเรส (Random Forest) ตามกระบวนการสร้างแบบจำลองในรูปที่ 3.7 และ รูปที่ 3.8



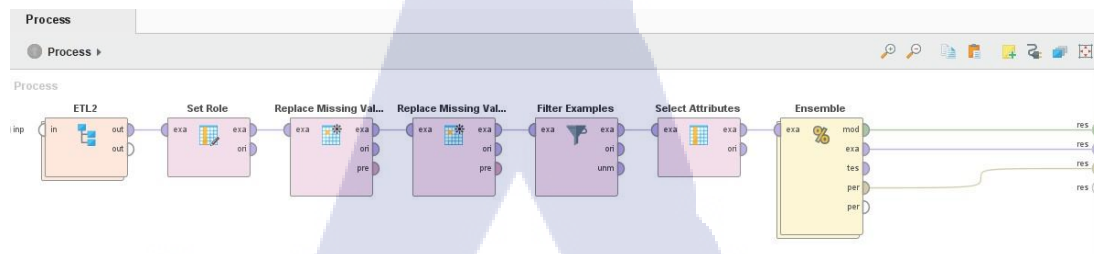
รูปที่ 3.7 กระบวนการสร้างแบบจำลอง



รูปที่ 3.8 การสร้างแบบจำลองด้วย 4 เทคนิค

เมื่อทำการทดลองสร้างแบบจำลองทั้ง 5 ครั้งเสร็จสิ้นแล้ว จะเลือกใช้แบบจำลองจากครั้งที่มีประสิทธิภาพมากที่สุด มาสร้างแบบจำลองที่จะนำไปใช้งาน โดยนำเทคนิค Ensemble มาใช้ในการสร้างแบบจำลอง จะเป็นการนำเทคนิคที่มีประสิทธิภาพสูงสุดใน 3 อันดับแรก เลือกจาก 4 เทคนิคที่ทดลอง มาช่วยกันทำนายผลด้วยเทคนิค Voting เพื่อเพิ่มประสิทธิภาพให้กับแบบจำลอง เหตุผลที่ใช้เพียง 3 เทคนิคแรกที่มีประสิทธิภาพสูงสุดเพราะการ Voting ต้องเป็นจำนวนคี่ถึงจะสามารถตัดสินผล

การทำนายที่ชนะ หรือแพ้ได้ หากเป็นจำนวนคู่ ถ้าผลการทำนายคลาสออกมาเท่ากันก็จะไม่สามารถตัดสินผลการทำนายที่ชนะ หรือแพ้ได้ ดังรูปที่ 3.9 และรูปที่ 3.10



รูปที่ 3.9 กระบวนการสร้างแบบจำลองด้วยเทคนิค Ensemble



รูปที่ 3.10 กระบวนการสร้างแบบจำลองด้วยเทคนิค Ensemble ด้วยเทคนิค Voting

3.3.5 การประเมินประสิทธิภาพของแบบจำลอง

นำแบบจำลองที่สร้างขึ้น มาเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยจะแบ่งออกเป็น 3 ส่วน

ส่วนที่ 1 เป็นการสร้างแบบจำลอง 5 ครั้ง การประเมินประสิทธิภาพของแบบจำลองในส่วนนี้จะวัดผลแบบจำลองโดยดูจากค่าความถูกต้อง (Accuracy) หากการทดลองครั้งไหนมีค่าความถูกต้องสูงสุดในครั้งนั้น ๆ จะถูกประเมินว่าเป็นแบบจำลองที่ดีที่สุด ในส่วนนี้ใช้เพียงค่าความถูกต้องในการวัดผลแบบจำลอง เนื่องจากเป็นการสร้างแบบจำลองจากข้อมูลที่เป็นข้อมูลในอดีต เป็นข้อมูลที่เกิดขึ้นมาแล้ว จึงสามารถเปรียบเทียบค่าความถูกต้องระหว่างผลที่ทำนาย และผลที่เกิดขึ้นจริงได้ หากพิจารณาจากตาราง Confusion Matrix [15] จะพบว่า ค่าความถูกต้องของแบบจำลองได้มาจากการที่แบบจำลองสามารถทำนายได้ตรงกับสิ่งที่เกิดขึ้นจริง เมื่อเทียบกับเหตุการณ์ที่เกิดขึ้นจริงทั้งหมด

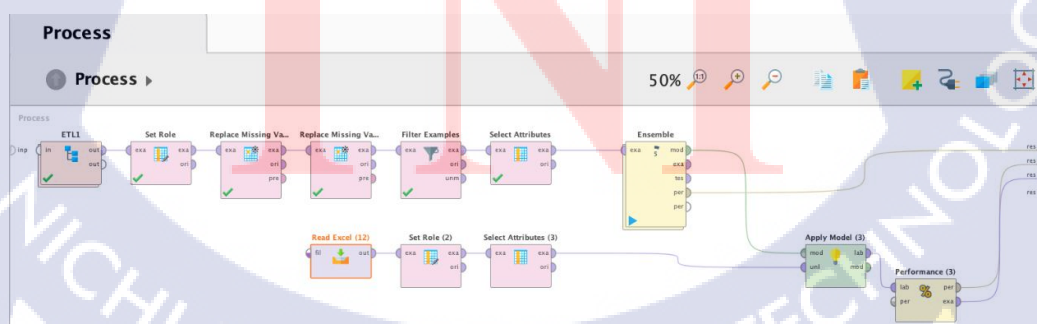
ดังนั้นจึงเลือกใช้เฉพาะค่าความถูกต้องมาประเมินการเลือกใช้แบบจำลอง ซึ่งเป็นค่าที่เหมาะสมกับการวัดในสิ่งที่เกิดขึ้นแล้ว

ส่วนที่ 2 เป็นการนำแบบจำลองที่ดีที่สุดที่ได้จากการทดลองในส่วนที่ 1 และสร้างแบบจำลองด้วยเทคนิค Ensemble การประเมินประสิทธิภาพของแบบจำลองในส่วนนี้ วัดผลตัวแบบจำลองด้วยตารางสำคัญการวัดความสามารถของการเรียนรู้ของเครื่อง หรือที่เรียกว่า Confusion Matrix โดยพิจารณาจาก 4 ด้าน ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F1-Score)

ส่วนที่ 3 เป็นการทดลองนำแบบจำลองไปใช้งาน การประเมินประสิทธิภาพของแบบจำลองในส่วนนี้จะวัดผลแบบจำลองโดยดูจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F1-Score)

3.3.6 การนำแบบจำลองไปใช้งาน

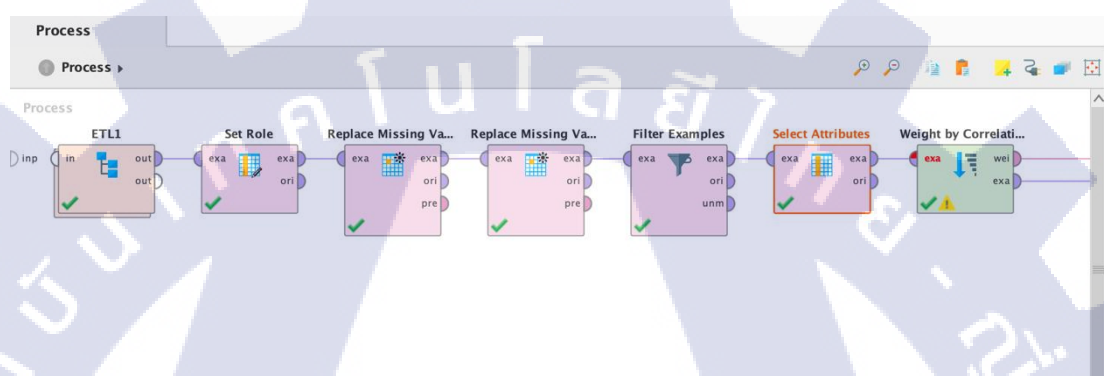
ในขั้นตอนนี้เป็นการนำแบบจำลองที่สร้างขึ้นมาในข้อ 3.2.5 ส่วนที่ 2 ไปทดลองใช้งาน เนื่องจากมีเหตุการณ์การแพร่ระบาดของไวรัส Covid-19 ในช่วงปี 2563 ทำให้สมาคมหลักด้านสาธารณสุขในประเทศไทยที่เป็นผู้จัดงานยกเลิกการจัดงานมหกรรมบ้านและคอนโดออกไป จนกว่าสถานการณ์การแพร่ระบาดของไวรัส Covid-19 จะดีขึ้น ประกอบกับรัฐบาลออกมาตรการเพื่อลดการแพร่ระบาดของไวรัส Covid-19 ไม่ให้มีวงกว้าง จึงมีการล็อกดาวน์ประเทศ และให้งดเว้นการจัดงานกิจกรรมต่าง ๆ ตามมาตรการ Social Distancing ผู้วิจัยจึงไม่สามารถนำแบบจำลองไปใช้งานกับข้อมูลที่จะเก็บในปี 2563 ได้ จากเหตุการณ์ดังกล่าว ผู้วิจัยจึงแบ่งข้อมูลในปี 2562 ที่มีจำนวนข้อมูล 4,460 รายการ มาเป็นข้อมูลที่ใช้ในการทดลองนำแบบจำลองที่มีประสิทธิภาพมากที่สุดมาทดสอบใช้งาน (Deployment) ตามขั้นตอนกระบวนการ ดังรูปที่ 3.11



รูปที่ 3.11 กระบวนการทดลองนำแบบจำลองที่มีประสิทธิภาพมากที่สุดมาทดสอบใช้งาน

3.4 ศึกษาปัจจัย (Attributes)

ในขั้นตอนนี้เป็นการศึกษาถึงปัจจัยที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรรมบ้านและคอนโด โดยใช้เทคนิคความสัมพันธ์ (Correlation) เป็นการพิจารณาจากค่าน้ำหนักของปัจจัยแต่ละตัวที่มีนัยสำคัญต่อการทำนายผลแบบจำลอง ถ้าปัจจัยไหนมีค่าน้ำหนักมาก หมายความว่าปัจจัยนั้นจะมีความสำคัญมากในการสร้างแบบจำลองตามลำดับของค่าน้ำหนัก ดังรูปที่ 3.12 เป็นกระบวนการศึกษาปัจจัยที่มีนัยสำคัญด้วยเทคนิคความสัมพันธ์ (Correlation)



รูปที่ 3.12 กระบวนการศึกษาปัจจัยที่มีนัยสำคัญด้วยเทคนิคความสัมพันธ์ (Correlation)

3.5 การสรุปผลการวิจัย

ผู้วิจัยเปรียบเทียบแบบจำลองที่สร้างขึ้นโดยใช้เทคนิคการเรียนรู้ของเครื่อง และสรุปผลการศึกษาแบบจำลองไหนมีประสิทธิภาพมากที่สุดโดยใช้เกณฑ์การตัดสินใจจากค่าต่าง ๆ ในข้อขั้นตอนที่ 3.2.5 และสรุปถึงปัจจัย (Attributes) ที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรรมบ้านและคอนโดตามวัตถุประสงค์ที่ได้กล่าวไว้ในบทที่ 1

บทที่ 4

ผลการวิจัย

การศึกษาวิจัยการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดโดยใช้เทคนิคการเรียนรู้ของเครื่อง ได้ทำการศึกษา ไว้ดังนี้

1. ผลการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดโดยใช้การเรียนรู้ของเครื่อง
2. ผลการศึกษาการพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย
3. การศึกษาปัจจัยที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

4.1 ผลการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดโดยใช้การเรียนรู้ของเครื่อง

จากการศึกษาการพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดด้วยเทคนิคการเรียนรู้ของเครื่องจำนวน 4 เทคนิค ได้แก่ เทคนิค Decision Tree, Naïve Bayes, K-Nearest Neighbors, Random Forest และเทคนิค Ensemble ได้ทำการทดลองเปรียบเทียบออกเป็น 3 ส่วน ดังนี้

4.1.1 ส่วนที่ 1 เป็นการศึกษาเปรียบเทียบแบบจำลอง โดยการสร้างแบบจำลอง 5 ครั้ง การประเมินประสิทธิภาพของแบบจำลองในส่วนนี้จะวัดผลแบบจำลองโดยดูจากค่าความถูกต้อง (Accuracy)

จากการทดลองสร้างแบบจำลองทั้งหมด 5 ครั้งเพื่อค้นหาว่า แบบจำลองในครั้งไหนมีประสิทธิภาพมากที่สุดและประเมินผลด้วยค่าความถูกต้อง (Accuracy) เมื่อพิจารณา จากตารางที่ 4.1 และ ตารางที่ 4.2 สามารถสรุปการทดลองในแต่ละครั้งได้ดังนี้

การทดลองสร้างแบบจำลองครั้งที่ 1 เป็นการใช้อัตราส่วน 2557 – 2561 จำนวน 113,554 รายการ พบว่า เทคนิค Naïve Bayes มีประสิทธิภาพในการพยากรณ์ได้ดีที่สุด มีค่าความถูกต้องร้อยละ 58.78 และใช้เวลาในการดำเนินการสร้างแบบจำลอง 1 ชั่วโมง 7 นาที 23 วินาที

การทดลองสร้างแบบจำลองครั้งที่ 2 เป็นการใช้ข้อมูลปี 2558 – 2561 จำนวน 92,389 รายการ พบว่า เทคนิค Naïve Bayes มีประสิทธิภาพในการทำนายได้ดีที่สุด มีค่าความถูกต้อง ร้อยละ 59.10 และใช้เวลาในการดำเนินการสร้างแบบจำลอง 40 นาที 18 วินาที

การทดลองสร้างแบบจำลองครั้งที่ 3 เป็นการใช้ข้อมูลปี 2559 – 2561 จำนวน 67,240 รายการ พบว่า เทคนิค Naïve Bayes มีประสิทธิภาพในการทำนายได้ดีที่สุด มีค่าความถูกต้อง ร้อยละ 62.39 และใช้เวลาในการดำเนินการสร้างแบบจำลอง 21 นาที 54 วินาที

การทดลองสร้างแบบจำลองครั้งที่ 4 เป็นการใช้ข้อมูลปี 2560 – 2561 จำนวน 40,483 รายการ พบว่า เทคนิค Decision Tree มีประสิทธิภาพในการทำนายได้ดีที่สุด มีค่าความถูกต้อง ร้อยละ 69.42 และใช้เวลาในการดำเนินการสร้างแบบจำลอง 8 นาที 10 วินาที

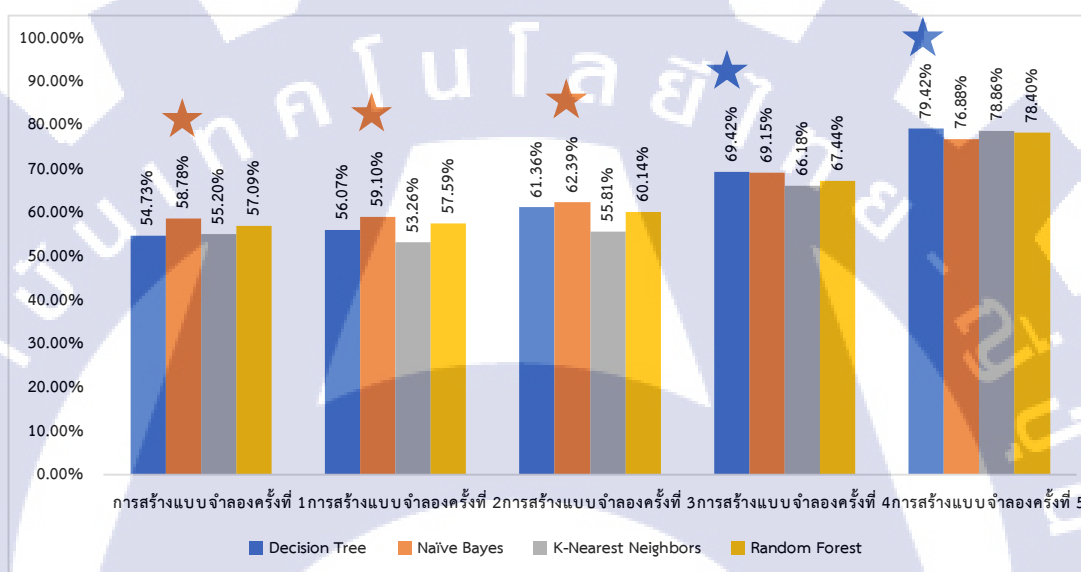
การทดลองสร้างแบบจำลองครั้งที่ 5 เป็นการใช้ข้อมูลปี 2561 จำนวน 20,919 รายการ พบว่า เทคนิค Decision Tree มีประสิทธิภาพในการทำนายได้ดีที่สุด มีค่าความถูกต้อง ร้อยละ 79.42 และใช้เวลาในการดำเนินการสร้างแบบจำลอง 2 นาที 13 วินาที

ตารางที่ 4.1 แสดงการเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เช่าชมงานมหรสพบ้านและคอนโด

	การสร้าง แบบจำลอง ครั้งที่ 1	การสร้าง แบบจำลอง ครั้งที่ 2	การสร้าง แบบจำลอง ครั้งที่ 3	การสร้าง แบบจำลอง ครั้งที่ 4	การสร้าง แบบจำลอง ครั้งที่ 5
แบบจำลอง	ข้อมูล 2557 – 2561 (113,554)	ข้อมูล 2558 – 2561 (92,389)	ข้อมูล 2559 – 2561 (67,240)	ข้อมูล 2560 – 2561 (40,483)	ข้อมูล 2561 (20,919)
Decision Tree	54.73%	56.07%	61.36%	69.42%	79.42%
Naïve Bayes	58.78%	59.10%	62.39%	69.15%	76.88%
K-Nearest Neighbors	55.20%	53.26%	55.81%	66.18%	78.86%
Random Forest	57.09%	57.59%	60.14%	67.44%	78.40%

ตารางที่ 4.2 แสดงการเปรียบเทียบเวลาที่ใช้ในการดำเนินการสร้างแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

	การสร้าง แบบจำลอง ครั้งที่ 1	การสร้าง แบบจำลอง ครั้งที่ 2	การสร้าง แบบจำลอง ครั้งที่ 3	การสร้าง แบบจำลอง ครั้งที่ 4	การสร้าง แบบจำลอง ครั้งที่ 5
Execution Time	01 : 07 : 23	40 : 18	21 : 54	8 : 10	2 : 13



รูปที่ 4.1 กราฟแท่งแสดงการเปรียบเทียบค่าความแม่นยำของแบบจำลอง

จากการทดลองสร้างแบบจำลองทั้ง 5 ครั้งจะเห็นว่าครั้งที่ 1 ถึง ครั้งที่ 3 พบว่า เทคนิค Naïve Bayes มีประสิทธิภาพในการทำนายได้ดีที่สุด เนื่องจากเทคนิคนี้มีคุณสมบัติเหมาะสมกับชุดข้อมูลที่มีปริมาณมากและใช้ความน่าจะเป็นในการทำนายความถูกต้องโดยใช้การเรียนรู้จากข้อมูลในอดีตที่มีปริมาณมากมาจำแนกข้อมูลใหม่ [10] ซึ่งจะเห็นว่าในการทดลองในครั้งที่ 1 ถึงครั้งที่ 3 ใช้ข้อมูลจำนวนมากในการสร้างแบบจำลอง แต่ในการทดลองสร้างแบบจำลองในครั้งที่ 4 และ 5 กลับพบว่า เทคนิค Decision Tree [8] มีประสิทธิภาพในการทำนายได้ดีที่สุด เนื่องจากเทคนิคนี้มีคุณสมบัติเหมาะสมกับชุดข้อมูลที่มีลักษณะคำตอบที่หลากหลาย หรือข้อมูลที่มีการกระจายของคำตอบมาก ดังเช่นในงานวิจัยนี้มีลักษณะของคนที่ซื้อ และไม่ซื้อที่อยู่อาศัยที่มีความหลากหลายมาก ดังนั้นเทคนิค Decision Tree จึงเหมาะสมกับการสร้างแบบจำลองในครั้งที่ 4 และ 5

เมื่อพิจารณาผลการทดลองทั้ง 5 ครั้ง จากรูปที่ 4.1 พบว่า การสร้างแบบจำลองครั้งที่ 5 ด้วยการใช้ข้อมูลปี 2561 มาสร้างแบบจำลองทั้ง 4 เทคนิค ได้ค่าความถูกต้องสูงเกินร้อยละ 70 และสังเกตเห็นว่าเมื่อลดจำนวนปีลงตัดปีที่เก่าออกครั้งละ 1 ปี โดยไม่นำปีเก่ามาสร้างแบบจำลอง ค่าความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญ จากการทดลองนี้สรุปได้ว่า ข้อมูลที่มีปริมาณมาก หรือเป็นข้อมูลที่เก่าเกินไป อาจจะไม่สะท้อนถึงข้อมูลในปัจจุบัน หรือทำนายข้อมูลได้ดีเท่ากับ ข้อมูลที่ใกล้เคียงกับเหตุการณ์ปัจจุบัน

4.1.2 ส่วนที่ 2 เป็นการศึกษาเปรียบเทียบแบบจำลอง จากแบบจำลองที่ดีที่สุดที่ได้จากการทดลองในส่วนที่ 1 การประเมินประสิทธิภาพของแบบจำลองในส่วนนี้ วัดผลตัวแบบจำลองด้วยตารางสำคัญการวัดความสามารถของการเรียนรู้ของเครื่อง หรือที่เรียกว่า Confusion Matrix โดยพิจารณาจาก 4 ด้าน ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F1-Score)

จากการศึกษาเปรียบเทียบแบบจำลอง จากการการสร้างแบบจำลองจำนวน 5 ครั้งในส่วนที่ 1 พบว่า การทดลองสร้างแบบจำลองในครั้งที่ 5 เป็นแบบจำลองที่มีประสิทธิภาพสูงที่สุดในส่วนนี้ ทำการการประเมินประสิทธิภาพของแบบจำลอง ด้วย Confusion Matrix โดยวัดค่าจาก 4 ด้าน ได้แก่ วัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F1-Score)

ตารางที่ 4.3 ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลอง 4 เทคนิค

ตัวแบบจำลอง	ค่าประสิทธิภาพ			
	ร้อยละ ค่าความถูกต้อง (Accuracy)	ร้อยละ ค่าความแม่นยำ (Precision)	ร้อยละ ค่าความระลึก (Recall)	ร้อยละ ค่าถ่วงดุล (F1-Score)
1. Decision Tree	79.42	85.20	55.70	67.36
2. K-Nearest Neighbors	78.86	81.35	57.84	67.61
3. Random Forest	78.25	72.33	69.62	70.95
4. Naïve Bayes	76.88	70.83	66.99	68.86

จากตารางที่ 4.3 จะเห็นได้ว่า ตัวแบบจำลองที่สร้างขึ้นด้วยเทคนิคต้นไม้ตัดสินใจ ให้ประสิทธิภาพดีที่สุดโดยมีค่าความถูกต้องร้อยละ 79.42 ค่าความแม่นยำร้อยละ 85.20 ค่าความระลึกร้อยละ 55.70 และค่าถ่วงดุลร้อยละ 67.36

และเพื่อเป็นการเพิ่มประสิทธิภาพของแบบจำลองให้สูงขึ้น ในงานวิจัยนี้จึงนำเทคนิค Ensemble มาใช้ในการสร้างแบบจำลอง โดยจะเป็นการนำเทคนิคที่มีประสิทธิภาพสูงสุดใน 3 อันดับแรก จากตารางที่ 4.3 ได้แก่ เทคนิค Decision Tree, K-Nearest Neighbors และ Random Forest มาช่วยกันทำนายผลด้วยเทคนิค Voting

4.1.3 ส่วนที่ 3 เป็นการศึกษาเปรียบเทียบแบบจำลอง จากแบบจำลองที่ดีที่สุดที่ได้จากการทดลองในส่วนที่ 1 มาเปรียบเทียบกับสร้างแบบจำลองด้วยเทคนิค Ensemble ในการประเมินประสิทธิภาพของแบบจำลองในส่วนนี้ วัดผลตัวแบบจำลองด้วยตารางสำคัญการวัดความสามารถของการเรียนรู้ของเครื่อง หรือที่เรียกว่า Confusion Matrix โดยพิจารณาจาก 4 ด้าน ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F1-Score)

จากการศึกษาเปรียบเทียบแบบจำลองส่วนที่ 2 ตัวแบบจำลองที่สร้างขึ้นด้วยเทคนิคต้นไม้ตัดสินใจให้ประสิทธิภาพดีที่สุด เมื่อนำมาเปรียบเทียบกับสร้างแบบจำลองด้วยเทคนิค Ensemble จากตารางที่ 4.4 พบว่า เทคนิค Ensemble เป็นแบบจำลองที่มีประสิทธิภาพดีกว่าเทคนิค Decision Tree โดยมีค่าความถูกต้องร้อยละ 79.57 ค่าความแม่นยำร้อยละ 82.18 ค่าความระลึกร้อยละ 59.32 และค่าถ่วงดุล 68.90 สูงกว่าเทคนิค Decision Tree ยกเว้นค่าความแม่นยำจะน้อยกว่าเทคนิค Decision Tree ซึ่งมาจากการทำนายที่ผิดพลาดของแบบจำลองซึ่งเป็นความผิดพลาดแบบ False Positive หรือความผิด Type I Error ที่มีค่าสูงขึ้นจึงทำให้ค่าความแม่นยำลดลง แต่ในงานวิจัยนี้ถือว่าสามารถยอมรับการทำนายที่ผิดพลาดนี้ได้ เนื่องจากการทำนายว่าซื้อที่อยู่อาศัยจริงแต่ในความเป็นจริงไม่ได้ซื้อที่อยู่อาศัย จึงไม่ทำให้เกิดความเสียหายต่อธุรกิจการซื้อขายที่อยู่อาศัยนี้ ในทางกลับกันหากค่าความระลึกของแบบจำลองเทคนิค Ensemble มีค่าลดลง ซึ่งเป็นความผิดพลาดแบบ False Negative หรือความผิดพลาด Type II Error ที่มีค่าสูงขึ้น จะทำให้ธุรกิจนี้สูญเสียโอกาสในการขายที่อยู่อาศัย เนื่องจากแบบจำลองทำนายว่าไม่ซื้อที่อยู่อาศัยแต่ในความเป็นจริงคือลูกค้าซื้อที่อยู่อาศัยจึงทำให้สูญเสียโอกาสในการขายไป หากพิจารณาเปรียบเทียบค่าความระลึกของทั้งสองแบบจำลอง พบว่า แบบจำลอง Ensemble มีค่าความระลึกที่สูงขึ้น จึงสรุปได้ว่าเทคนิค Ensemble เป็นเทคนิคการสร้างแบบจำลองที่เพิ่มประสิทธิภาพให้กับแบบจำลองได้ดี ดังนั้นในงานวิจัยนี้จึงนำ เทคนิค Ensemble นี้ไป พัฒนาแบบจำลองประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหรหรมนบ้านและคอนโดได้อย่างเหมาะสม

ตารางที่ 4.4 ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลองเทคนิค Decision Tree และเทคนิค Ensemble

ตัวแบบจำลอง	ค่าประสิทธิภาพ			
	ร้อยละ ค่าความถูกต้อง (Accuracy)	ร้อยละ ค่าความแม่นยำ (Precision)	ร้อยละ ค่าความระลึก (Recall)	ร้อยละ ค่าถ่วงดุล (F1-Score)
1. Ensemble	79.57	82.18	59.32	68.90
2. Decision Tree	79.42	85.20	55.70	67.36

4.2 ผลการศึกษาการพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย

จากการศึกษาเปรียบเทียบแบบจำลอง พบว่า แบบจำลองสร้างด้วยเทคนิค Ensemble มีประสิทธิภาพดีที่สุด จึงนำไปทดลองใช้งาน เนื่องจากมีเหตุการณ์การแพร่ระบาดของไวรัส Covid-19 ในช่วงปี 2563 ทำให้สมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยที่เป็นผู้จัดงานยกเลิกการจัดงานมหกรรมบ้านและคอนโดออกไปจนกว่าสถานการณ์การแพร่ระบาดของไวรัส Covid-19 จะดีขึ้น ประกอบกับรัฐบาลประกาศให้มีการล็อกดาวน์ประเทศ และให้งดเว้นการจัดงานกิจกรรมต่าง ๆ ตามมาตรการ Social Distancing ผู้วิจัยจึงไม่สามารถนำแบบจำลองไปใช้งานกับข้อมูลที่จะเก็บในปี 2563 ได้ จากเหตุการณ์ดังกล่าว จึงใช้ข้อมูลในปี 2562 ที่มีจำนวนข้อมูล 4,460 รายการ มาเป็นข้อมูลที่ใช้ในการทดสอบการใช้งานของแบบจำลอง (Deployment) ซึ่งสามารถสรุปผลการทดสอบได้ดังนี้

ตารางที่ 4.5 ผลการประเมินประสิทธิภาพของตัวแบบจำลอง Ensemble เมื่อนำไปทดลองใช้งาน (Confusion Matrix)

การทำนายของตัวแบบจำลอง	ผลการซื้อที่อยู่อาศัยจริง	
	ซื้อ	ไม่ซื้อ
ซื้อ	1,025	161
ไม่ซื้อ	779	2,495

จากตารางที่ 4.5 จะเห็นได้ว่า เมื่อนำตัวแบบที่พัฒนาขึ้นมาไปใช้ในการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด มีการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้องจำนวน 1,025 รายการ และมีการประเมินแนวโน้มการไม่ตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้องจำนวน 2,495 รายการ มีค่าความถูกต้องร้อยละ 78.92 ค่าความแม่นยำร้อยละ 86.42 ค่าความระลึกร้อยละ 56.82 และค่าถ่วงดุลร้อยละ 68.56

ตารางที่ 4.6 ค่าความสำคัญในการวัดความสามารถของตัวแบบจำลองเทคนิค Ensemble

ตัวแบบจำลอง	ค่าประสิทธิภาพ			
	ร้อยละ ค่าความถูกต้อง (Accuracy)	ร้อยละ ค่าความแม่นยำ (Precision)	ร้อยละ ค่าความระลึก (Recall)	ร้อยละ ค่าถ่วงดุล (F1-Score)
Ensemble	78.92	86.42	56.82	68.56

4.3 การศึกษาปัจจัยที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

จากการศึกษาปัจจัยที่มีนัยสำคัญต่อการการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยด้วยเทคนิคความสัมพันธ์ (Correlation) จากตารางที่ 4.7 พบว่า ปัจจัยที่มีความสำคัญต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด ได้แก่ สถานภาพสมรส ซึ่งมีค่าความสัมพันธ์สูงที่สุดมีค่าเท่ากับ 0.502 รองลงมาเป็นเรื่องของรายได้ครอบครัว มีค่าความสัมพันธ์เท่ากับ 0.428 ความสามารถในการผ่อนชำระ มีค่าความสัมพันธ์เท่ากับ 0.404 ราคาที่อยู่อาศัยที่ซื้อ มีค่าความสัมพันธ์เท่ากับ 0.282 อายุ มีค่าความสัมพันธ์เท่ากับ 0.209 อาชีพ มีค่าความสัมพันธ์เท่ากับ 0.196 วัตถุประสงค์ที่ซื้อที่อยู่อาศัย มีค่าความสัมพันธ์เท่ากับ 0.194 ระยะเวลาที่ตัดสินใจซื้อ มีค่าความสัมพันธ์เท่ากับ 0.188 ส่วนเรื่องเพศ มีนัยสำคัญต่อการตัดสินใจซื้อที่อยู่อาศัยน้อยที่สุด มีค่าความสัมพันธ์เท่ากับ 0.057

ตารางที่ 4.7 ลำดับความสำคัญของปัจจัย (Attributes) ที่มีต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัย

ลำดับที่	ปัจจัย (Attributes)	ค่าน้ำหนัก
1	สถานภาพสมรส	0.502
2	รายได้ครอบครัว	0.428
3	ความสามารถในการผ่อนชำระ	0.404
4	ราคาที่อยู่อาศัยที่ซื้อ	0.282
5	อายุ	0.209
6	อาชีพ	0.196
7	วัตถุประสงค์ที่ซื้อที่อยู่อาศัย	0.194
8	ระยะเวลาที่ตัดสินใจซื้อ	0.188
9	เพศ	0.057

บทที่ 5

บทสรุป อภิปราย และข้อเสนอแนะ

การศึกษาวิจัยการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดโดยใช้เทคนิคการเรียนรู้ของเครื่อง มีวัตถุประสงค์ในการทำวิจัยดังนี้

1. เพื่อเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด
2. เพื่อพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย
3. เพื่อศึกษาปัจจัย (Attributes) ที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด

5.1 สรุปผลการวิจัย

5.1.1 ผลการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดโดยใช้การเรียนรู้ของเครื่องด้วย จำนวน 4 เทคนิค ได้แก่ เทคนิค Decision Tree, Naïve Bayes, K-Nearest Neighbors, Random Forest และเทคนิค Ensemble สรุปผลการศึกษาได้ดังนี้

1. การศึกษาเปรียบเทียบแบบจำลอง โดยการสร้างแบบจำลอง 5 ครั้ง ด้วยจำนวน 4 เทคนิค นั้นพบว่า การสร้างแบบจำลองในครั้งที่ 5 ด้วยการใช้ข้อมูลปี 2561 ได้ค่าความถูกต้องสูงเกินร้อยละ 70 และสังเกตเห็นว่าเมื่อลดจำนวนปีลงตัดปีที่เก่าออกครั้งละ 1 ปี โดยไม่นำปีเก่ามาสร้างแบบจำลอง ค่าความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญ จากการทดลองนี้สรุปได้ว่า ข้อมูลที่มีปริมาณมากหรือเป็นข้อมูลที่เก่าเกินไป อาจจะไม่สะท้อนถึงข้อมูลในปัจจุบัน หรือทำนายข้อมูลได้ดีเท่ากับ ข้อมูลที่ใกล้เคียงกับเหตุการณ์ปัจจุบัน

2. การศึกษาเปรียบเทียบแบบจำลอง จากการการสร้างแบบจำลองจำนวน 5 ครั้ง พบว่า การทดลองสร้างแบบจำลองในครั้งที่ 5 เป็นแบบจำลองที่มีประสิทธิภาพสูงที่สุด เมื่อพิจารณาแบบจำลองที่สร้างขึ้นมาในครั้งที่ 5 สรุปได้ว่าตัวแบบจำลองที่สร้างขึ้นด้วยเทคนิค Decision Tree ให้ประสิทธิภาพดีที่สุดในครั้งนี้โดยมีค่าความถูกต้องร้อยละ 79.42 ค่าความแม่นยำร้อยละ 85.20 ค่าความระลึกร้อยละ 75.70 และค่าถ่วงดุลร้อยละ 67.36 เนื่องจากเทคนิคนี้มีคุณสมบัติเหมาะสมกับชุดข้อมูลที่มีลักษณะคำตอบที่หลากหลาย หรือข้อมูลที่มีการกระจายของคำตอบมาก ดังเช่นในงานวิจัยนี้มี

ลักษณะของคนที่ซื้อ และไม่ซื้อที่อยู่อาศัยที่มีความหลากหลายมาก ดังนั้น เทคนิค Decision Tree จึงเหมาะสมกับการสร้างแบบจำลอง

3. จากการศึกษาเปรียบเทียบแบบจำลองที่สร้างขึ้นด้วยเทคนิคต้นไม้ตัดสินใจ และแบบจำลองด้วยเทคนิค Ensemble พบว่า เทคนิค Ensemble ที่สร้างขึ้นจากเทคนิค Decision Tree, K-Nearest Neighbors และ Random Forest มาช่วยกันทำนายผลด้วยเทคนิค Voting เป็นแบบจำลองที่มีประสิทธิภาพดีกว่า เทคนิค Decision Tree โดยมีค่าความถูกต้องร้อยละ 79.57 ค่าความแม่นยำร้อยละ 82.18 ค่าความระลึกร้อยละ 59.32 และค่าถ่วงดุล 68.90 สรุปได้ว่า เทคนิค Ensemble เป็นเทคนิคการสร้างแบบจำลองที่เพิ่มประสิทธิภาพให้กับแบบจำลองได้ดี ดังนั้นในงานวิจัยนี้จึงนำ เทคนิค Ensemble นี้ไปพัฒนาสร้างแบบจำลองประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโดได้อย่างเหมาะสม

5.1.2 ผลการศึกษาการพัฒนาแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด สำหรับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทย พบว่า แบบจำลองสร้างด้วยเทคนิค Ensemble มีประสิทธิภาพ จึงนำไปทดลองใช้งาน เนื่องจากมีเหตุการณ์การแพร่ระบาดของไวรัส Covid-19 ในช่วงปี 2563 ทำให้สมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยที่เป็นผู้จัดงานยกเลิกการจัดงานมหกรรมบ้านและคอนโดออกไปจนกว่าสถานการณ์การแพร่ระบาดของไวรัส Covid-19 จะดีขึ้น ประกอบกับรัฐบาลประกาศให้มีการล็อกดาวน์ประเทศ และให้งดเว้นการจัดงานกิจกรรมต่าง ๆ ตามมาตรการ Social Distancing ผู้วิจัยจึงไม่สามารถนำแบบจำลองไปใช้งานกับข้อมูลที่จะเก็บในปี 2563 ได้ จากเหตุการณ์ดังกล่าว จึงใช้ข้อมูลในปี 2562 ที่มีจำนวนข้อมูล 4,460 รายการ มาเป็นข้อมูลที่ใช้ในการทดสอบการใช้งานของแบบจำลอง (Deployment) ซึ่งแบบจำลอง Ensemble สามารถประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้องจำนวน 1,025 รายการ และมีการประเมินแนวโน้มการไม่ตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้องจำนวน 2,495 รายการ มีค่าความถูกต้องร้อยละ 78.92 ค่าความแม่นยำร้อยละ 86.42 ค่าความระลึกร้อยละ 56.82 และค่าถ่วงดุลร้อยละ 68.56

5.1.3 การศึกษาปัจจัยที่มีนัยสำคัญสำหรับการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด พบว่า ปัจจัยที่มีความสำคัญต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด ได้แก่ สถานภาพสมรส รองลงมาเป็นเรื่องของรายได้ครอบครัว ความสามารถในการผ่อนชำระ ราคาที่อยู่อาศัยที่ซื้อ อายุ อาชีพ วัตถุประสงค์ที่ซื้อที่อยู่อาศัย ระยะเวลาที่ตัดสินใจซื้อ ส่วนเรื่องเพศ มีนัยสำคัญต่อการตัดสินใจซื้อที่อยู่อาศัยน้อยที่สุด

5.2 อภิปรายผล

จากการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด ผู้วิจัยได้สร้างแบบจำลองโดยใช้ข้อมูลปี 2557 – 2561 และมีทั้งหมด 9 ปัจจัย ในการสร้างแบบจำลองแบบมีผู้สอน (Supervised Learning) เนื่องจากมีข้อมูลที่บอกลักษณะของคนที่ตัดสินใจซื้อที่อยู่อาศัย และข้อมูลลักษณะของคนที่ไม่ตัดสินใจซื้อที่อยู่อาศัย ผู้วิจัยจึงได้สร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ เทคนิคเบเซียนแบบง่าย เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด และเทคนิคแรนดอมฟอเรส ซึ่งเป็นอัลกอริทึม (Algorithm) แบบมีผู้สอน [6] จากการทดลองสร้างแบบจำลองจากข้อมูลที่มีปริมาณมาก และเป็นข้อมูลในอดีต เปรียบเทียบกับการสร้างแบบจำลองจากข้อมูลที่มีปริมาณน้อย แต่เป็นข้อมูลที่ใกล้กับปัจจุบันนั้น เมื่อพิจารณาจากค่าความถูกต้อง พบว่า ข้อมูลที่มีปริมาณมาก และเป็นข้อมูลที่เก่าเกินไป เมื่อนำมาสร้างแบบจำลอง ทำให้ค่าความถูกต้องไม่สูงมาก มีค่าไม่ถึงร้อยละ 60 แต่เมื่อนำข้อมูลที่มีปริมาณน้อย และเป็นข้อมูลที่ใกล้กับปัจจุบัน เมื่อนำมาสร้างแบบจำลอง ทำให้ค่าความถูกต้องสูงเกินร้อยละ 70 อธิบายได้ว่า ข้อมูลที่เก่าเกินไปจะใช้อธิบายถึงข้อมูลในปัจจุบัน หรือสามารถทำนายข้อมูลในอนาคต ได้น้อยกว่าข้อมูลที่ใกล้กับปัจจุบัน ซึ่งสอดคล้องกับเทคนิคการพยากรณ์ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก [27] ที่ให้ความสำคัญกับข้อมูลที่ใกล้กับปัจจุบันโดยให้น้ำหนักกับข้อมูลในปัจจุบันจากมากไปหาน้อยที่เป็นข้อมูลในอดีต และเมื่อนำข้อมูลปี 2561 เพียง 1 ปี มาสร้างแบบจำลอง พบว่า ตัวแบบจำลองที่สร้างขึ้นด้วยเทคนิค Decision Tree ให้ประสิทธิภาพดีที่สุดโดยมีค่าความถูกต้องร้อยละ 79.42 ค่าความแม่นยำร้อยละ 85.20 ค่าความระลึกร้อยละ 55.70 และค่าถ่วงดุลร้อยละ 67.36 เช่นเดียวกับงานวิจัยผลของตัวแบบที่ส่งผลต่อความสามารถในการชำระสินเชื่อของสมาชิกสหกรณ์เครดิตยูเนียน [26] ซึ่งค้นพบว่าเทคนิค Decision Tree เป็นตัวแบบที่มีประสิทธิภาพในการทำนายมากที่สุด และสอดคล้องกับทฤษฎีหลักการทำงานของเทคนิค Decision Tree [8] ซึ่งจะมีประสิทธิภาพในการทำนายได้ดีที่สุด เนื่องจากเทคนิคนี้มีคุณสมบัติเหมาะกับชุดข้อมูลที่มีลักษณะคำตอบที่หลากหลายหรือข้อมูลที่มีการกระจายของคำตอบมาก ดังเช่นในงานวิจัยนี้มีลักษณะของคนที่ยื่นและไม่ยื่นที่อยู่อาศัยที่มีความหลากหลายมาก ดังนั้น เทคนิค Decision Tree จึงมีประสิทธิภาพสูงในการทำนายถึงแม้จะได้แบบจำลองที่มีประสิทธิภาพสูงแล้วก็ตามในการวิจัยครั้งนี้ได้เพิ่มการทดลองโดยนำเทคนิค Ensemble มาสร้างแบบจำลองเพื่อเพิ่มประสิทธิภาพการทำนายของแบบจำลอง โดยเป็นการรวมเทคนิคที่มีประสิทธิภาพสูงสุดใน 3 อันดับแรก ได้แก่ เทคนิค Decision Tree, K-Nearest Neighbors และ Random Forest มาช่วยกันทำนายผลด้วยเทคนิค Voting โดยมีค่าความถูกต้องร้อยละ 79.57 ค่าความแม่นยำร้อยละ 82.18 ค่าความระลึกร้อยละ 59.32 และค่าถ่วงดุล 68.90 เมื่อเปรียบเทียบแบบจำลองเทคนิค Decision Tree กับแบบจำลองที่สร้างด้วยเทคนิค Ensemble จะพบว่าค่าเกือบทุกค่าสูงกว่าเทคนิค Decision Tree ยกเว้นค่าความแม่นยำจะน้อยกว่าเทคนิค Decision Tree ซึ่งมา

จากการทำนายที่ผิดพลาดของแบบจำลองซึ่งเป็นความผิดพลาดแบบ False Positive หรือความผิดพลาด Type I Error [15] ที่มีค่าสูงขึ้นจึงทำให้ค่าความแม่นยำลดลงแต่ในงานวิจัยนี้ถือว่าสามารถยอมรับการทำนายที่ผิดพลาดนี้ได้ เนื่องจากการทำนายว่าซื้อที่อยู่อาศัยจริงแต่ในความเป็นจริงไม่ได้ซื้อที่อยู่อาศัย จึงไม่ทำให้เกิดความเสียหายต่อธุรกิจการซื้อขายที่อยู่อาศัยนี้ ในทางกลับกันหากค่าความระลึกรของแบบจำลองเทคนิค Ensemble มีค่าลดลง ซึ่งเป็นความผิดพลาดแบบ False Negative หรือความผิดพลาด Type II Error [15] ที่มีค่าสูงขึ้น จะทำให้ธุรกิจนี้สูญเสียโอกาสในการขายที่อยู่อาศัย เนื่องจาแบบจำลองทำนายว่าไม่ซื้อที่อยู่อาศัยแต่ในความเป็นจริงคือลูกค้าซื้อที่อยู่อาศัยจึงทำให้สูญเสียโอกาสในการขายไป หากพิจารณาเปรียบเทียบค่าความระลึกรของทั้งสองแบบจำลอง พบว่าแบบจำลอง Ensemble มีค่าความระลึกรที่สูงขึ้น จึงสรุปได้ว่า แบบจำลองที่สร้างด้วยเทคนิค Ensemble มีประสิทธิภาพในการทำนายได้ดีกว่า แบบจำลองที่สร้างด้วยเทคนิค Decision Tree ซึ่งสอดคล้องกับทฤษฎีของเทคนิค Ensemble [12] ที่มีคุณสมบัติหลักเป็นเทคนิคการสร้างแบบจำลองเพื่อเพิ่มประสิทธิภาพให้กับแบบจำลองให้ดีขึ้นกว่าเดิม นอกจากนี้ผู้วิจัยยัง พบว่า ปัจจัยที่มีความสำคัญต่อการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด ได้แก่ สถานภาพสมรส รองลงมาเป็นเรื่องของรายได้ครบครัว ความสามารถในการผ่อนชำระ ราคาที่อยู่อาศัยที่ซื้อ อายุ อาชีพ วัตถุประสงค์ที่ซื้อที่อยู่อาศัย ระยะเวลาที่ตัดสินใจซื้อ ส่วนเรื่องเพศ มีนัยสำคัญต่อการตัดสินใจซื้อที่อยู่อาศัยน้อยที่สุด ซึ่งสอดคล้องกับทฤษฎีกระบวนการตัดสินใจของผู้ซื้อ (Buyer's Decision Process) ตามทฤษฎีของคอตเลอร์ [5] จะมีสิ่งรบกวนนอกเข้ามาเพื่อให้ตัดสินใจซื้อ อย่างเช่น เรื่องของการแต่งงานจะทำให้ผู้บริโภคต้องการซื้อของที่จำเป็นต่อการใช้ชีวิตคู่ และจะตระหนักถึงรายได้รวมถึงความสามารถในการที่จะซื้อ ราคาที่จะซื้อ เป็นต้น และยังสอดคล้องกับงานวิจัยพฤติกรรมผู้ซื้อทรัพย์สินการขาย กรณีศึกษาผู้เข้าชมทรัพย์สินการขายของสถาบันการเงินในงานบ้านและคอนโดปี 2557 [18] และงานวิจัยศึกษาเปรียบเทียบปัจจัยในการเลือกอาคารชุดพักอาศัยเป็นบ้านหลังที่สองในเขตชั้นกลางของกรุงเทพมหานคร [19] ซึ่งค้นพบว่า รายได้ และราคาที่อยู่อาศัยที่เหมาะสมกับรายได้ส่งผลต่อการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด จากการทดลองนำแบบจำลองไปทดสอบใช้งานซึ่งผู้วิจัยได้ใช้ข้อมูลในปี 2562 จำนวน 4,460 รายการ เพื่อใช้ในการประเมินตัวแบบจำลอง Ensemble พบว่าแบบจำลองมีการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้อง มีค่าความถูกต้องร้อยละ 78.92 ทั้งนี้แสดงให้เห็นว่า ตัวแบบจำลองที่พัฒนาขึ้นมาสามารถเพิ่มโอกาสทางธุรกิจให้กับผู้ประกอบการอสังหาริมทรัพย์ที่เป็นสมาชิกในสมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยได้อย่างเหมาะสม เนื่องจากสามารถประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างถูกต้องถึงร้อยละ 66.8 จากการที่ซื้อจริงทั้งหมด หากนำแบบจำลองนี้ไปประเมินผู้เข้าชมงานมหกรรมบ้านและคอนโดจะสามารถเพิ่มโอกาสในการซื้อขายได้ถึงร้อยละ 66.8 จากการประเมินผู้เข้าชมงานทั้งหมด

กล่าวโดยสรุปการศึกษาเปรียบเทียบแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด พบว่า เทคนิค Ensemble มีประสิทธิภาพมากที่สุด และสามารถประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้อย่างเหมาะสม ทำให้ผู้ประกอบการสามารถใช้แบบจำลองนี้เพื่อเพิ่มโอกาสในการขายที่อยู่อาศัยให้มีประสิทธิภาพสูงสุด และเพิ่มความสำเร็จในการขายได้ นอกจากนี้ทำให้ทราบว่า ปัจจัยที่ส่งผลต่อแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยมากที่สุด 3 อันดับแรก ได้แก่ สถานภาพสมรส รองลงเป็นเรื่องของรายได้ครอบครัว ความสามารถในการผ่อนชำระ ตามลำดับ

5.3 ปัญหาและอุปสรรคในการทำวิจัย

5.3.1 ต้องศึกษาการเรียนรู้เครื่องมือที่เป็นโปรแกรมชุดคำสั่งสำเร็จรูปแบบบล็อก และในงานวิจัยนี้ใช้เครื่องมือโปรแกรม Rapid Miner Studio มาช่วยในการวิเคราะห์ จึงทำให้เกิดความล่าช้าในขั้นตอนการวิเคราะห์ข้อมูล เนื่องจากเป็นเครื่องมือที่ไม่เคยใช้มาก่อน

5.3.2 มีปัญหาอุปสรรคในขั้นตอนการนำแบบจำลองไปทดลองใช้งาน (Deployment) เนื่องจากในปี 2563 เกิดเหตุการณ์แพร่ระบาดของไวรัส Covid-19 ทำให้สมาคมหลักด้านอสังหาริมทรัพย์ในประเทศไทยที่เป็นผู้จัดงาน ต้องยกเลิกการจัดงานมหกรรมบ้านและคอนโดออกไป จนกว่าสถานการณ์การแพร่ระบาดของไวรัส Covid-19 จะมีสถานการณ์ดีขึ้น ผู้วิจัยจึงใช้ข้อมูลปี 2562 ที่เป็นปัจจุบันมากที่สุดในการทดลองนำแบบจำลองไปทดลองใช้งาน

5.4 ข้อเสนอแนะงานวิจัย

5.4.1 นำแบบจำลองการประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยของผู้เข้าชมงานมหกรรมบ้านและคอนโด ไปประยุกต์ใช้กับกลุ่มผู้ที่ต้องการที่จะซื้อที่อยู่อาศัยที่ไปเยี่ยมชมตามโครงการที่อยู่อาศัย โดยเก็บข้อมูลตามปัจจัยที่ใช้ในการสร้างแบบจำลองในงานวิจัยนี้ เพื่อให้เกิดประโยชน์ต่อธุรกิจอสังหาริมทรัพย์มากยิ่งขึ้น

5.4.2 จากการศึกษาปัจจัยที่ส่งผลต่อแนวโน้มการตัดสินใจซื้อที่อยู่อาศัย พบว่า 3 อันดับแรก ได้แก่ สถานภาพสมรส รายได้ครอบครัว และความสามารถในการผ่อนชำระ ส่งผลสูงสุดต่อการตัดสินใจซื้อที่อยู่อาศัย ดังนั้นสามารถนำแบบจำลองไปใช้งาน โดยสามารถเก็บข้อมูลเพียง 3 ปัจจัยนี้ก็จะสามารถประเมินแนวโน้มการตัดสินใจซื้อที่อยู่อาศัยได้ เพื่อลดเวลาและความยุ่งยากในการจัดเก็บข้อมูล

5.4.3 การศึกษาถึงเทคนิคการสร้างแบบจำลองในกลุ่มของการเรียนรู้แบบมีผู้สอน ต้องศึกษาเทคนิคแต่ละเทคนิคให้เข้าใจ เนื่องจากบางเทคนิคอาจจะไม่เหมาะกับชุดข้อมูลในการทำวิจัย และอาจจะส่งผลของการสร้างแบบจำลองไม่ถูกต้อง หรือมีค่าความถูกต้องที่ไม่สูงมาก

บรรณานุกรม

TNI

บรรณานุกรม

- [1] ฝ่ายสถิติและประมวลผลข้อมูล ศูนย์ข้อมูลอสังหาริมทรัพย์, “จำนวนหน่วยและมูลค่าการโอนกรรมสิทธิ์ ในกรุงเทพมหานครและปริมณฑล ปี 2558 – 2562,” [Online]. Available: <https://www.reic.or.th/Research/211>. [Accessed: 5 กันยายน 2563].
- [2] สมาคมอสังหาริมทรัพย์ไทย, “How to Real Estate Development & Research Center: วิธีการและขั้นตอนการลงทุนในอสังหาริมทรัพย์,” [Online]. Available: <http://thairealestate.org/content/lists/40/Journal>. [Accessed: 10 กันยายน 2563].
- [3] สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ, “ข้อมูลเศรษฐกิจและสังคม: รายงานภาวะเศรษฐกิจ,” [Online]. Available: https://www.nesdc.go.th/main.php?filename=QGDP_report#. [Accessed: 10 กันยายน 2563].
- [4] ศิริวรรณ เสรีรัตน์ และคณะ, *การบริหารการตลาดยุคใหม่*, กรุงเทพฯ: บริษัท อีระฟิล์ม และโซเท็กซ์ จำกัด, 2560.
- [5] อุดลย์ จาตุรงค์กุล, *พฤติกรรมผู้บริโภค*, กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์, 2543.
- [6] T. Mitchell and H. McGraw, *Machine Learning*, Pittsburgh: Carnegie Mellon University, 1997.
- [7] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning*, England: Cambridge University Press, 2009.
- [8] เอกสิทธิ์ พัทธวงศ์ศักดิ์, *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไม่นิ่ง เบื้องต้น*, กรุงเทพฯ: ดาต้าคิวบ์, 2557.
- [9] L. Breiman and R. Forests, “Random Forests,” *Kluwer Academic Publishers*, Vol. 45, no. 1, pp. 5-32, June 2001.
- [10] วรณสิริ ฐระชน และคณะ, “ระบบการจำแนกพันธุ์ ยางพาราโดยใช้ตัวจำแนกนาอ์ฟเบย์,” *การประชุมทางวิชาการระดับชาติ ด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 10, NCCIT 2014*, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ, 8-9 พฤษภาคม 2557, หน้า 20-25.
- [11] ธวัช รวมทรัพย์ และคณะ, “การจำแนกชนิดของพืชด้วยวิธีเพื่อนบ้านใกล้ที่สุด ร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอน วิธีเชิงพันธุกรรมโดยใช้คุณลักษณะรูปทรงและพื้นผิวของใบพืช,” *วิทยานิพนธ์ วท.บ. (เทคโนโลยี)*, มหาวิทยาลัยศิลปากร, กรุงเทพฯ, 2559.

- [12] เอกสิทธิ์ พัทธวงศ์ศักดิ์, “การสร้างโมเดล Ensemble แบบต่างๆ,” [Online]. Available: <http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model/>. [Accessed: 15 กันยายน 2563].
- [13] วัทธัญตา นีลาภาตระกูล และคณะ, “การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัท ประกันภัย,” วิทยานิพนธ์ วท.บ. (เทคโนโลยี), มหาวิทยาลัยรังสิต, กรุงเทพฯ, 2562.
- [14] G. Golub et al., “Generalized cross-validation as a method for choosing a good ridge parameter,” *Technometrics*, Vol. 21 no.2, pp. 215–223, April 1979.
- [15] J. Mohammed and J.R. Meira Wagner, *Data Mining and Analysis Fundamental Concepts and Algorithms*, England: Cambridge University Press, 2014.
- [16] Hinkle, D.E, William, W. and Stephen G. J., *Applied Statistics for the Behavior Sciences, 4th ed*, New York: Houghton Mifflin, 1998.
- [17] อนันต ช่างทอง และจิรนนท์ อนุสรณ์พานิช, “ปัจจัยที่ส่งผลต่อการตัดสินใจเลือกซื้ออสังหาริมทรัพย์ประเภททรัพย์สินรอการขายของธนาคารออมสิน,” ค้นคว้าอิสระ บธ.ม. (การจัดการ), มหาวิทยาลัยหอการค้าไทย, กรุงเทพฯ, 2560.
- [18] ศักดิ์สิทธิ์ มิตรเจริญถาวร, “พฤติกรรมผู้ซื้อทรัพย์สินรอการขาย กรณีศึกษาผู้เข้าชมทรัพย์สินรอการขายของสถาบันการเงินในงานบ้านและคอนโดปี 2557,” วิทยานิพนธ์ คพ.ม. (เคหการ), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2557.
- [19] อลงกรณ์ จาฟู, “ศึกษาเปรียบเทียบปัจจัยในการเลือกอาคารชุดพักอาศัยเป็นบ้านหลังที่สองในเขตชั้นกลางของกรุงเทพมหานคร,” วิทยานิพนธ์ คพ.ม. (เคหการ), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2554.
- [20] ภูริณัฐ ธนวิบูลย์ชัย, “ปัจจัยที่มีผลต่อการตัดสินใจเลือกซื้อที่อยู่อาศัยในเขตกรุงเทพมหานครและปริมณฑลของพนักงานบริษัทเอกชน,” การค้นคว้าอิสระ บธ.ม. (พาณิชยศาสตร์และการบัญชี), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2557.
- [21] สรลรัชว์ สุรธรรมทวี และกฤษฎา พัทธวานิช, “ปัจจัยที่มีอิทธิพลต่อการตัดสินใจซื้อคอนโดมิเนียมของผู้บริโภค กรณีศึกษากรุงเทพมหานคร,” วิทยานิพนธ์ บธ.ม. (การประกอบการ), มหาวิทยาลัยศิลปากร, กรุงเทพฯ, 2554.
- [22] ธวัชชัย อมรพิสิทธิกุล และ สุกัญญา ภูสุวรรณรัตน์, “ปัจจัยทางเศรษฐกิจที่มีผลกระทบต่อจำนวนห้องชุดที่จัดทะเบียนในกรุงเทพมหานคร,” ค้นคว้าอิสระ บธ.ม. (การเงิน), มหาวิทยาลัยหอการค้าไทย, กรุงเทพฯ, 2557.

- [23] สำนักงานศูนย์วิจัยและให้คำปรึกษา มหาวิทยาลัยธรรมศาสตร์, “การคาดประมาณความต้องการที่อยู่อาศัยของประเทศไทยปี 2560 – 2580,” กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์, 2560.
- [24] Hannah Sophia Seippel, “Customer purchase prediction through machine learning,” M.S. thesis (Electrical Engineering), University of Twente, Ensched, Netherlands, 2018.
- [25] S. Stubseid and O. Arandjelovic, “Machine Learning Based Prediction of Consumer Purchasing Decisions: The Evidence and Its Significance,” *Proceedings AI and Marketing Science Workshop, AAAI 2018*, New Orleans, United States, February 2, 2018, no page.
- [26] ไพจิตร สุขสมบูรณ์ และคณะ, “ผลของตัวแบบที่ส่งผลต่อความสามารถในการชำระสินเชื่อของสมาชิกสหกรณ์เครดิตยูเนี่ยน,” *วารสารวิชาการคณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏลำปาง*, ปีที่ 10, ฉบับที่ 1, หน้า 101-110, มกราคม – มิถุนายน, 2560.
- [27] สถาพร โอภาสานนท์, “การพยากรณ์อุปสงค์โดยใช้เทคนิค การพยากรณ์มากกว่าหนึ่งวิธี” *วารสารบริหารธุรกิจ มหาวิทยาลัยธรรมศาสตร์*, ปีที่ 34, ฉบับที่ 129, หน้า 1-4, มกราคม-มีนาคม, 2554.

ภาคผนวก

TNI

ภาคผนวก ก.

ผลการเปรียบเทียบค่าความถูกต้องของทดลองสร้างแบบจำลอง

TNI

คำนิยามของหมายเลขแบบจำลอง

- 1 หมายถึง แบบจำลอง Decision Tree
- 2 หมายถึง แบบจำลอง Naïve Bayes
- 3 หมายถึง แบบจำลอง K-Nearest Neighbors
- 4 หมายถึง แบบจำลอง Random Forest

ตารางที่ ก.1 ผลการเปรียบเทียบค่าความถูกต้องของทดลองสร้างแบบจำลอง

ครั้งที่ของ การสร้างแบบจำลอง และปีที่ใช้ในการสร้าง แบบจำลอง	ผลการดำเนินการการสร้างแบบจำลองจากโปรแกรม RapidMiner Studio Education 9.8.001																																			
ครั้งที่ 1 ข้อมูล 2557 – 2561 (113,554)	<table><tr><th>iteration</th><th>Select Subprocess.select_which</th><th>accuracy</th></tr><tr><td>1</td><td>1</td><td>0.5473</td></tr><tr><td>2</td><td>2</td><td>0.5876</td></tr><tr><td>4</td><td>4</td><td>0.5709</td></tr><tr><td>3</td><td>3</td><td>0.5520</td></tr></table> <table><tr><th colspan="4">accuracy: 58.76% +/- 0.49% (micro average: 58.76%)</th></tr><tr><th></th><th>true Y</th><th>true N</th><th>class precision</th></tr><tr><td>pred. Y</td><td>40060</td><td>24749</td><td>61.81%</td></tr><tr><td>pred. N</td><td>22085</td><td>26660</td><td>54.69%</td></tr><tr><td>class recall</td><td>64.46%</td><td>51.86%</td><td></td></tr></table>	iteration	Select Subprocess.select_which	accuracy	1	1	0.5473	2	2	0.5876	4	4	0.5709	3	3	0.5520	accuracy: 58.76% +/- 0.49% (micro average: 58.76%)					true Y	true N	class precision	pred. Y	40060	24749	61.81%	pred. N	22085	26660	54.69%	class recall	64.46%	51.86%	
iteration	Select Subprocess.select_which	accuracy																																		
1	1	0.5473																																		
2	2	0.5876																																		
4	4	0.5709																																		
3	3	0.5520																																		
accuracy: 58.76% +/- 0.49% (micro average: 58.76%)																																				
	true Y	true N	class precision																																	
pred. Y	40060	24749	61.81%																																	
pred. N	22085	26660	54.69%																																	
class recall	64.46%	51.86%																																		
ครั้งที่ 2 ข้อมูล 2558 – 2561 (92,389)	<table><tr><th>iteration</th><th>Select Subprocess.select_which</th><th>accuracy</th></tr><tr><td>2</td><td>2</td><td>0.5910</td></tr><tr><td>1</td><td>1</td><td>0.5607</td></tr><tr><td>4</td><td>4</td><td>0.5759</td></tr><tr><td>3</td><td>3</td><td>0.5326</td></tr></table> <table><tr><th colspan="4">accuracy: 59.10% +/- 0.53% (micro average: 59.10%)</th></tr><tr><th></th><th>true N</th><th>true Y</th><th>class precision</th></tr><tr><td>pred. N</td><td>27522</td><td>19971</td><td>57.95%</td></tr><tr><td>pred. Y</td><td>17814</td><td>27082</td><td>60.32%</td></tr><tr><td>class recall</td><td>60.71%</td><td>57.56%</td><td></td></tr></table>	iteration	Select Subprocess.select_which	accuracy	2	2	0.5910	1	1	0.5607	4	4	0.5759	3	3	0.5326	accuracy: 59.10% +/- 0.53% (micro average: 59.10%)					true N	true Y	class precision	pred. N	27522	19971	57.95%	pred. Y	17814	27082	60.32%	class recall	60.71%	57.56%	
iteration	Select Subprocess.select_which	accuracy																																		
2	2	0.5910																																		
1	1	0.5607																																		
4	4	0.5759																																		
3	3	0.5326																																		
accuracy: 59.10% +/- 0.53% (micro average: 59.10%)																																				
	true N	true Y	class precision																																	
pred. N	27522	19971	57.95%																																	
pred. Y	17814	27082	60.32%																																	
class recall	60.71%	57.56%																																		

ตารางที่ ก.1 ผลการเปรียบเทียบค่าความถูกต้องของทดลองสร้างแบบจำลอง (ต่อ)

ครั้งที่ของ การสร้างแบบจำลอง และปีที่ใช้ในการสร้าง แบบจำลอง	ผลการดำเนินการการสร้างแบบจำลองจากโปรแกรม RapidMiner Studio Education 9.8.001																																			
ครั้งที่ 3 ข้อมูล 2559 – 2561 (67,240)	<table><tr><th>iteration</th><th>Select Subprocess.select_which</th><th>accuracy</th></tr><tr><td>1</td><td>1</td><td>0.6136</td></tr><tr><td>2</td><td>2</td><td>0.6239</td></tr><tr><td>4</td><td>4</td><td>0.6014</td></tr><tr><td>3</td><td>3</td><td>0.5581</td></tr></table> <table><tr><th colspan="4">accuracy: 62.39% +/- 0.51% (micro average: 62.39%)</th></tr><tr><th></th><th>true N</th><th>true Y</th><th>class precision</th></tr><tr><td>pred. N</td><td>20738</td><td>13857</td><td>59.95%</td></tr><tr><td>pred. Y</td><td>11435</td><td>21210</td><td>64.97%</td></tr><tr><td>class recall</td><td>64.46%</td><td>60.48%</td><td></td></tr></table>	iteration	Select Subprocess.select_which	accuracy	1	1	0.6136	2	2	0.6239	4	4	0.6014	3	3	0.5581	accuracy: 62.39% +/- 0.51% (micro average: 62.39%)					true N	true Y	class precision	pred. N	20738	13857	59.95%	pred. Y	11435	21210	64.97%	class recall	64.46%	60.48%	
iteration	Select Subprocess.select_which	accuracy																																		
1	1	0.6136																																		
2	2	0.6239																																		
4	4	0.6014																																		
3	3	0.5581																																		
accuracy: 62.39% +/- 0.51% (micro average: 62.39%)																																				
	true N	true Y	class precision																																	
pred. N	20738	13857	59.95%																																	
pred. Y	11435	21210	64.97%																																	
class recall	64.46%	60.48%																																		
ครั้งที่ 4 ข้อมูล 2560 – 2561 (40,483)	<table><tr><th>iteration</th><th>Select Subprocess.select_which</th><th>accuracy</th></tr><tr><td>2</td><td>2</td><td>0.6915</td></tr><tr><td>1</td><td>1</td><td>0.6942</td></tr><tr><td>4</td><td>4</td><td>0.6744</td></tr><tr><td>3</td><td>3</td><td>0.6618</td></tr></table> <table><tr><th colspan="4">accuracy: 69.42% +/- 1.05% (micro average: 69.42%)</th></tr><tr><th></th><th>true Y</th><th>true N</th><th>class precision</th></tr><tr><td>pred. Y</td><td>12089</td><td>5597</td><td>68.35%</td></tr><tr><td>pred. N</td><td>6783</td><td>16014</td><td>70.25%</td></tr><tr><td>class recall</td><td>64.06%</td><td>74.10%</td><td></td></tr></table>	iteration	Select Subprocess.select_which	accuracy	2	2	0.6915	1	1	0.6942	4	4	0.6744	3	3	0.6618	accuracy: 69.42% +/- 1.05% (micro average: 69.42%)					true Y	true N	class precision	pred. Y	12089	5597	68.35%	pred. N	6783	16014	70.25%	class recall	64.06%	74.10%	
iteration	Select Subprocess.select_which	accuracy																																		
2	2	0.6915																																		
1	1	0.6942																																		
4	4	0.6744																																		
3	3	0.6618																																		
accuracy: 69.42% +/- 1.05% (micro average: 69.42%)																																				
	true Y	true N	class precision																																	
pred. Y	12089	5597	68.35%																																	
pred. N	6783	16014	70.25%																																	
class recall	64.06%	74.10%																																		

ตารางที่ ก.1 ผลการเปรียบเทียบค่าความถูกต้องของทดลองสร้างแบบจำลอง (ต่อ)

ครั้งที่ของ การสร้างแบบจำลอง และปีที่ใช้ในการสร้าง แบบจำลอง	ผลการดำเนินการการสร้างแบบจำลองจากโปรแกรม RapidMiner Studio Education 9.8.001																	
ครั้งที่ 5 ข้อมูล 2561 (20,919)	<table><tr><th>iteration</th><th>Select Subprocess.select_which</th><th>accuracy</th></tr><tr><td>2</td><td>2</td><td>0.7688</td></tr><tr><td>1</td><td>1</td><td>0.7942</td></tr><tr><td>4</td><td>4</td><td>0.7840</td></tr><tr><td>3</td><td>3</td><td>0.7886</td></tr></table>			iteration	Select Subprocess.select_which	accuracy	2	2	0.7688	1	1	0.7942	4	4	0.7840	3	3	0.7886
	iteration	Select Subprocess.select_which	accuracy															
	2	2	0.7688															
	1	1	0.7942															
	4	4	0.7840															
3	3	0.7886																
accuracy: 58.76% +/- 0.49% (micro average: 58.76%)																		
	true Y	true N	class precision															
pred. Y	40060	24749	61.81%															
pred. N	22085	26660	54.69%															
class recall	64.46%	51.86%																

ภาคผนวก ข.

ผลการเปรียบเทียบเวลาที่ใช้ในการดำเนินการสร้างแบบจำลอง

TNI

ตารางที่ ข.1 ผลการเปรียบเทียบเวลาที่ใช้ในการดำเนินการสร้างแบบจำลอง

ครั้งที่ของ การสร้างแบบจำลอง และปีที่ใช้ในการสร้าง แบบจำลอง	ผลการดำเนินการการสร้างแบบจำลองจากโปรแกรม RapidMiner Studio Education 9.8.001
ครั้งที่ 1 ข้อมูล 2557 – 2561 (113,554)	model by house Select Bast Model (3 results, Process results) Completed: Feb 13, 2021 5:40:05 AM (execution time: 01:07:23) Performance Vector (Performance (2)) Result not stored in repository. PerformanceVector: accuracy: 58.7562% +/- 0.4880% (micro average) ConfusionMatrix: True: Y N Y: 48068 24749 N: 22885 26668 precision: 54.6964% +/- 0.5976% (micro average) ConfusionMatrix: True: Y N Y: 48068 24749 N: 22885 26668 recall: 51.8586% +/- 0.6892% (micro average) ConfusionMatrix: True: Y N Y: 48068 24749 N: 22885 26668 Distribution Model (Naive Bayes) Result not stored in repository. Distribution model for label attribute Label Class Y (0.5473) 9 distributions Class N (0.4527) 9 distributions Parameter Set (Optimize Parameters (Grid)) Result not stored in repository. Parameter set: Performance: PerformanceVector [-----accuracy: 58.7562% +/- 0.4880% (micro average)-----precision: 54.6964% +/- 0.5976% (micro average)-----recall: 51.8586% +/- 0.6892% (micro average)-----confusion matrix: True: Y N Y: 48068 24749 N: 22885 26668]
ครั้งที่ 2 ข้อมูล 2558 – 2561 (92,389)	model by house Select Bast Model (3 results, Process results) Completed: Feb 13, 2021 8:43:51 AM (execution time: 40:18) Performance Vector (Performance (2)) Result not stored in repository. PerformanceVector: accuracy: 59.1823% +/- 0.5276% (micro average) ConfusionMatrix: True: N Y N: 27522 19971 Y: 17814 27882 precision: 68.3235% +/- 0.5624% (micro average) ConfusionMatrix: True: N Y N: 27522 19971 Y: 17814 27882 recall: 57.5564% +/- 0.6952% (micro average) ConfusionMatrix: True: N Y N: 27522 19971 Y: 17814 27882 Distribution Model (Naive Bayes) Result not stored in repository. Distribution model for label attribute Label Class N (0.4907) 9 distributions Class Y (0.5093) 9 distributions Parameter Set (Optimize Parameters (Grid)) Result not stored in repository. Parameter set: Performance: PerformanceVector [-----accuracy: 59.1823% +/- 0.5276% (micro average)-----precision: 68.3235% +/- 0.5624% (micro average)-----recall: 57.5564% +/- 0.6952% (micro average)-----confusion matrix: True: N Y N: 27522 19971 Y: 17814 27882]
ครั้งที่ 3 ข้อมูล 2559 – 2561 (67,240)	model by house Select Bast Model (3 results, Process results) Completed: Feb 13, 2021 9:34:52 AM (execution time: 21:54) Performance Vector (Performance (2)) Result not stored in repository. PerformanceVector: accuracy: 62.3855% +/- 0.5128% (micro average) ConfusionMatrix: True: N Y N: 28738 13857 Y: 11435 21210 precision: 64.9718% +/- 0.4739% (micro average) ConfusionMatrix: True: N Y N: 28738 13857 Y: 11435 21210 recall: 60.4842% +/- 1.1141% (micro average) ConfusionMatrix: True: N Y N: 28738 13857 Y: 11435 21210 Distribution Model (Naive Bayes) Result not stored in repository. Distribution model for label attribute Label Class N (0.4785) 9 distributions Class Y (0.5215) 9 distributions Parameter Set (Optimize Parameters (Grid)) Result not stored in repository. Parameter set: Performance: PerformanceVector [-----accuracy: 62.3855% +/- 0.5128% (micro average)-----precision: 64.9718% +/- 0.4739% (micro average)-----recall: 60.4842% +/- 1.1141% (micro average)-----confusion matrix: True: N Y N: 28738 13857 Y: 11435 21210]
ครั้งที่ 4 ข้อมูล 2560 – 2561 (40,483)	model by house Select Bast Model (3 results, Process results) Completed: Feb 13, 2021 10:14:56 AM (execution time: 8:10) Performance Vector (Performance (2)) Result not stored in repository. PerformanceVector: accuracy: 69.4193% +/- 1.0486% (micro average) ConfusionMatrix: True: Y N Y: 12889 5597 N: 6783 16814 precision: 70.2484% +/- 0.9613% (micro average) ConfusionMatrix: True: Y N Y: 12889 5597 N: 6783 16814 recall: 74.1812% +/- 1.1750% (micro average) ConfusionMatrix: True: Y N Y: 12889 5597 N: 6783 16814 Decision Tree (Decision Tree (2)) Result not stored in repository.  Parameter Set (Optimize Parameters (Grid)) Result not stored in repository. Parameter set: Performance: PerformanceVector [-----accuracy: 69.4193% +/- 1.0486% (micro average)-----precision: 70.2484% +/- 0.9613% (micro average)-----recall: 74.1812% +/- 1.1750% (micro average)-----confusion matrix: True: Y N Y: 12889 5597 N: 6783 16814]
ครั้งที่ 5 ข้อมูล 2561 (20,919)	model by house Select Bast Model (3 results, Process results) Completed: Feb 13, 2021 10:21:55 AM (execution time: 2:13) Performance Vector (Performance (2)) Result not stored in repository. PerformanceVector: accuracy: 79.4158% +/- 1.1641% (micro average) ConfusionMatrix: True: N Y N: 12168 3535 Y: 771 4445 precision: 85.2033% +/- 1.7616% (micro average) ConfusionMatrix: True: N Y N: 12168 3535 Y: 771 4445 recall: 55.7017% +/- 2.3589% (micro average) ConfusionMatrix: True: N Y N: 12168 3535 Y: 771 4445 Decision Tree (Decision Tree (2)) Result not stored in repository.  Parameter Set (Optimize Parameters (Grid)) Result not stored in repository. Parameter set: Performance: PerformanceVector [-----accuracy: 79.4158% +/- 1.1641% (micro average)-----precision: 85.2033% +/- 1.7616% (micro average)-----recall: 55.7017% +/- 2.3589% (micro average)-----confusion matrix: True: N Y N: 12168 3535 Y: 771 4445]

ภาคผนวก ค.

ผลการสร้างแบบจำลองด้วยเทคนิค Ensemble

TNI

ภาคผนวก ง.

ผลการประเมินประสิทธิภาพของตัวแบบจำลอง Ensemble เมื่อนำไปทดลองใช้งาน

TNI

ตารางที่ ง.1 ผลการประเมินประสิทธิภาพของตัวแบบจำลอง Ensemble เมื่อนำไปทดลองใช้งาน

แบบจำลอง Ensemble	ผลการดำเนินการการสร้างแบบจำลองจากโปรแกรม RapidMiner Studio Education 9.8.001																																																																																																																																																																																																																																						
Accuracy	accuracy: 78.9238% <table><tr><td></td><td>true N</td><td>true Y</td><td>class precision</td></tr><tr><td>pred. N</td><td>2495</td><td>779</td><td>76.2065%</td></tr><tr><td>pred. Y</td><td>161</td><td>1025</td><td>86.4250%</td></tr><tr><td>class recall</td><td>93.9383%</td><td>56.8182%</td><td></td></tr></table>		true N	true Y	class precision	pred. N	2495	779	76.2065%	pred. Y	161	1025	86.4250%	class recall	93.9383%	56.8182%																																																																																																																																																																																																																							
	true N	true Y	class precision																																																																																																																																																																																																																																				
pred. N	2495	779	76.2065%																																																																																																																																																																																																																																				
pred. Y	161	1025	86.4250%																																																																																																																																																																																																																																				
class recall	93.9383%	56.8182%																																																																																																																																																																																																																																					
เวลาที่ใช้ในการ ดำเนินการสร้าง แบบจำลอง	model by house Deploy (3 results, process results) Completed: Feb 13, 2021 12:01:57 PM (execution time: 22:47) Performance Vector (Performance) Result not stored in repository. PerformanceVector: accuracy: 79.5688% +/- 0.8533% (micro average) ConfusionMatrix: True: N Y N: 11911 3246 Y: 1028 4734 precision: 82.1833% +/- 1.8484% (micro average) ConfusionMatrix: True: N Y N: 11911 3246 Y: 1028 4734 recall: 59.3237% +/- 1.4129% (micro average) ConfusionMatrix: True: N Y N: 11911 3246 Performance Vector (Performance (3)) Result not stored in repository. PerformanceVector: accuracy: 78.9238% ConfusionMatrix: True: N Y N: 2495 779 Y: 161 1025 ExampleSet (Apply Model (3)) Result not stored in repository. Data Table Source: /Users/nungluckdthong/Desktop/My Final/data/บ้าน_ผลการประเมิน_ensemble/41-nsdata-2.xlsx Number of examples = 4460 13 attributes: <table><tr><th>Role</th><th>Name</th><th>Type</th><th>Range</th><th>Miss</th></tr><tr><td>-</td><td>Gender</td><td>polynomial</td><td>= {1, 2}</td><td>no miss value</td></tr><tr><td>-</td><td>Married</td><td>polynomial</td><td>= {1, 2}</td><td>no miss value</td></tr></table>	Role	Name	Type	Range	Miss	-	Gender	polynomial	= {1, 2}	no miss value	-	Married	polynomial	= {1, 2}	no miss value																																																																																																																																																																																																																							
Role	Name	Type	Range	Miss																																																																																																																																																																																																																																			
-	Gender	polynomial	= {1, 2}	no miss value																																																																																																																																																																																																																																			
-	Married	polynomial	= {1, 2}	no miss value																																																																																																																																																																																																																																			
ผลการทำนาย	Open in Turbo Prep Auto Model Filter (4,460 / 4,460 examples): all <table><tr><th>Row No.</th><th>Label</th><th>prediction</th><th>confidence</th><th>confidence</th><th>Gender</th><th>Married</th><th>Age</th><th>Carreer</th><th>FamilyInco...</th></tr><tr><td>1</td><td>N</td><td>N</td><td>0.6667</td><td>0.3333</td><td>2</td><td>1</td><td>3</td><td>1</td><td>6</td></tr><tr><td>2</td><td>Y</td><td>N</td><td>0.6667</td><td>0.3333</td><td>2</td><td>1</td><td>5</td><td>1</td><td>5</td></tr><tr><td>3</td><td>Y</td><td>Y</td><td>0.3333</td><td>0.6667</td><td>2</td><td>1</td><td>4</td><td>1</td><td>4</td></tr><tr><td>4</td><td>Y</td><td>Y</td><td>0.3333</td><td>0.6667</td><td>1</td><td>1</td><td>3</td><td>1</td><td>3</td></tr><tr><td>5</td><td>Y</td><td>N</td><td>0.6667</td><td>0.3333</td><td>2</td><td>1</td><td>2</td><td>3</td><td>7</td></tr><tr><td>6</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>2</td><td>5</td><td>3</td></tr><tr><td>7</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>3</td><td>6</td><td>2</td></tr><tr><td>8</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>4</td><td>5</td><td>3</td></tr><tr><td>9</td><td>N</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>3</td><td>1</td><td>2</td></tr><tr><td>10</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>2</td><td>1</td><td>6</td></tr><tr><td>11</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>2</td><td>1</td><td>3</td></tr><tr><td>12</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>5</td><td>5</td><td>7</td></tr><tr><td>13</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>3</td><td>4</td><td>3</td></tr><tr><td>14</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>2</td><td>4</td><td>3</td></tr><tr><td>15</td><td>N</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>4</td><td>7</td><td>1</td></tr><tr><td>16</td><td>N</td><td>N</td><td>0.6667</td><td>0.3333</td><td>1</td><td>1</td><td>3</td><td>4</td><td>2</td></tr><tr><td>17</td><td>N</td><td>N</td><td>1</td><td>0</td><td>1</td><td>2</td><td>4</td><td>1</td><td>4</td></tr><tr><td>18</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>2</td><td>1</td><td>1</td></tr><tr><td>19</td><td>N</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>2</td><td>1</td><td>3</td></tr><tr><td>20</td><td>Y</td><td>N</td><td>1</td><td>0</td><td>1</td><td>1</td><td>2</td><td>1</td><td>3</td></tr><tr><td>21</td><td>Y</td><td>Y</td><td>0</td><td>1</td><td>2</td><td>2</td><td>5</td><td>7</td><td>1</td></tr><tr><td>22</td><td>N</td><td>N</td><td>1</td><td>0</td><td>2</td><td>1</td><td>2</td><td>1</td><td>2</td></tr></table> ExampleSet (4,460 examples, 4 special attributes, 9 regular attributes)	Row No.	Label	prediction	confidence	confidence	Gender	Married	Age	Carreer	FamilyInco...	1	N	N	0.6667	0.3333	2	1	3	1	6	2	Y	N	0.6667	0.3333	2	1	5	1	5	3	Y	Y	0.3333	0.6667	2	1	4	1	4	4	Y	Y	0.3333	0.6667	1	1	3	1	3	5	Y	N	0.6667	0.3333	2	1	2	3	7	6	Y	N	1	0	1	1	2	5	3	7	Y	N	1	0	1	1	3	6	2	8	Y	N	1	0	1	1	4	5	3	9	N	N	1	0	1	1	3	1	2	10	Y	N	1	0	2	1	2	1	6	11	Y	N	1	0	2	1	2	1	3	12	Y	N	1	0	1	1	5	5	7	13	Y	N	1	0	2	1	3	4	3	14	Y	N	1	0	2	1	2	4	3	15	N	N	1	0	2	1	4	7	1	16	N	N	0.6667	0.3333	1	1	3	4	2	17	N	N	1	0	1	2	4	1	4	18	Y	N	1	0	2	1	2	1	1	19	N	N	1	0	1	1	2	1	3	20	Y	N	1	0	1	1	2	1	3	21	Y	Y	0	1	2	2	5	7	1	22	N	N	1	0	2	1	2	1	2
Row No.	Label	prediction	confidence	confidence	Gender	Married	Age	Carreer	FamilyInco...																																																																																																																																																																																																																														
1	N	N	0.6667	0.3333	2	1	3	1	6																																																																																																																																																																																																																														
2	Y	N	0.6667	0.3333	2	1	5	1	5																																																																																																																																																																																																																														
3	Y	Y	0.3333	0.6667	2	1	4	1	4																																																																																																																																																																																																																														
4	Y	Y	0.3333	0.6667	1	1	3	1	3																																																																																																																																																																																																																														
5	Y	N	0.6667	0.3333	2	1	2	3	7																																																																																																																																																																																																																														
6	Y	N	1	0	1	1	2	5	3																																																																																																																																																																																																																														
7	Y	N	1	0	1	1	3	6	2																																																																																																																																																																																																																														
8	Y	N	1	0	1	1	4	5	3																																																																																																																																																																																																																														
9	N	N	1	0	1	1	3	1	2																																																																																																																																																																																																																														
10	Y	N	1	0	2	1	2	1	6																																																																																																																																																																																																																														
11	Y	N	1	0	2	1	2	1	3																																																																																																																																																																																																																														
12	Y	N	1	0	1	1	5	5	7																																																																																																																																																																																																																														
13	Y	N	1	0	2	1	3	4	3																																																																																																																																																																																																																														
14	Y	N	1	0	2	1	2	4	3																																																																																																																																																																																																																														
15	N	N	1	0	2	1	4	7	1																																																																																																																																																																																																																														
16	N	N	0.6667	0.3333	1	1	3	4	2																																																																																																																																																																																																																														
17	N	N	1	0	1	2	4	1	4																																																																																																																																																																																																																														
18	Y	N	1	0	2	1	2	1	1																																																																																																																																																																																																																														
19	N	N	1	0	1	1	2	1	3																																																																																																																																																																																																																														
20	Y	N	1	0	1	1	2	1	3																																																																																																																																																																																																																														
21	Y	Y	0	1	2	2	5	7	1																																																																																																																																																																																																																														
22	N	N	1	0	2	1	2	1	2																																																																																																																																																																																																																														