



# โครงการส่งเสริมการวิเคราะห์ และบริหารข้อมูลขนาดใหญ่ภาครัฐ

ต้นแบบระบบแนะนำที่อยู่อาศัยของการเคหะแห่งชาติ

Prototype of a House Recommender  
System for National Housing Agency

จัดทำโดย

สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi)

ในสังกัดสำนักงานส่งเสริมเศรษฐกิจดิจิทัล (depa)

ตุลาคม 2564

## บทสรุปผู้บริหาร

การเคหะแห่งชาติมีความมุ่งหมายในการให้บริการประชาชน โดยเฉพาะผู้มีรายได้น้อย ในการแนะนำและจัดสร้างที่อยู่อาศัย เพื่อให้สร้างหลักประกันการมีความเป็นอยู่ที่ดีสำหรับคนทุกระดับและทุกวัย ตลอดจนลดปัญหาเกี่ยวกับความเหลื่อมล้ำในสังคม และปัญหาการขาดแคลนที่อยู่อาศัย ได้เล็งเห็นถึงความสำคัญในการนำข้อมูลผู้จองในอดีตมาใช้ประโยชน์ในการพัฒนาและปรับปรุงการให้บริการให้ตรงตามความต้องการของประชาชนมากขึ้น ทางการเคหะแห่งชาติจึงร่วมมือกับ GBDi จัดทำระบบแนะนำที่อยู่อาศัยในลักษณะ Web Application เพื่ออำนวยความสะดวกให้แก่ผู้ที่สนใจโครงการที่อยู่อาศัยของการเคหะแห่งชาติ

ในการพัฒนาด้านแบบของระบบแนะนำที่อยู่อาศัย ทางทีมงาน GBDi นำข้อมูลสองส่วนที่ได้รับจากการเคหะแห่งชาติมาใช้ ซึ่งมีดังนี้ 1.) ข้อมูลทางประชากรของผู้จองโครงการการเคหะแห่งชาติ อาทิเช่น เพศ อายุ สถานภาพสมรส อาชีพ รายได้ เป็นต้น ตั้งแต่ พ.ศ.2550 ถึง พ.ศ.2562 2.) ข้อมูลโครงการการเคหะแห่งชาติ ณ ปัจจุบันที่มีหน่วยว่างทั้งหมด 543 โครงการซึ่งอัปเดตล่าสุดในเดือนกรกฎาคม พ.ศ.2564

หลังจากทำความสะอาดข้อมูลข้างต้นเรียบร้อยแล้ว ทางทีมงานได้นำข้อมูลทางประชากรในอดีตมาพัฒนาแบบจำลองในการทำนายประเภทบ้าน หลังจากนั้นจึงพัฒนา Web Application เพื่อรับข้อมูลของผู้ที่สนใจในโครงการของการเคหะ

แห่งชาติ ก่อนจะคัดกรองโครงการที่อยู่ใกล้ที่อยู่ปัจจุบันหรือที่ทำงานของผู้ที่กรอกข้อมูลมาแสดงบนหน้าจอซึ่งสามารถกรองโครงการที่ทางระบบคาดว่าผู้กรอกข้อมูลจะผ่อนไหวได้เพิ่มเติมอีกด้วย

ทางทีมงาน GBDi มุ่งหวังว่าระบบแนะนำที่อยู่อาศัยต้นแบบนี้จะเป็นประโยชน์ในการพัฒนาการบริการของการเคหะแห่งชาติและส่งเสริมการใช้ข้อมูลในการตัดสินใจในองค์กรต่อไป

# Executive Summary

National Housing Agency (NHA) determines to provide housing services, such as suggestions on finding appropriate houses and projects to construct affordable housing, especially for low-income people. Its missions are to guarantee that citizens have a good standard of life, reduce social inequality, and tackle the housing shortage. Currently, NHA aims to exploit historical data to have a better understanding of customers and improve the services. Therefore, NHA has collaborated with GBDi (Government Big Data institute) to design a prototype of a house recommender system, which is run on a web application, for prospective customers of NHA projects.

To develop the web application, we received two datasets from NHA: 1.) Dataset of NHA customers between 2007 and 2019 in terms of demography, such as sex, age, marital status, occupation, and income; 2.) Dataset of 543 NHA projects which still had available units (updated on July 2021).

After cleansing the data, we used the customer dataset to develop a machine learning model to predict types of housing. We designed a web application to receive necessary data from a prospective customer. Then, a list of projects, that are close to the customer's working place or current living place, is returned to the web application. Additionally, the customer can further filter the results that they tend to successfully pay for by instalments.

We hope that the prototype would be beneficial for improving NHA services and enhance NHA to be a data-driven organisation.

# สารบัญ

บทสรุปผู้บริหาร .....	1
Executive Summary .....	2
สารบัญ.....	3
สารบัญภาพ .....	5
สารบัญตาราง.....	6
บทที่ 1 บทนำ.....	7
1.1 หลักการและเหตุผล .....	7
1.2 วัตถุประสงค์.....	8
1.3 ผลผลิตที่คาดหวัง .....	8
1.4 ขอบเขตการดำเนินงานและผู้เกี่ยวข้อง .....	8
บทที่ 2 ระเบียบวิธีดำเนินการ.....	9
2.1 การเตรียมข้อมูล.....	9
2.1.1 การเตรียมข้อมูลลูกค้าจากการเคหะแห่งชาติ .....	9
2.1.2 การเตรียมข้อมูลโครงการ .....	12
2.2 การเตรียมฐานข้อมูล .....	14
2.3 แบบจำลองคณิตศาสตร์ทำนายประเภทที่อยู่อาศัย .....	17
2.3.1 การแบ่งข้อมูลเพื่อใช้ฝึกฝนและทดสอบแบบจำลองทางคณิตศาสตร์ (Train-Test Split).....	17
2.3.2 การเลือกตัวแปรที่นำไปใช้ในแบบจำลองทางคณิตศาสตร์ (Feature Selection).....	17
2.4 การเลือกแบบจำลองทางคณิตศาสตร์ .....	18
2.4.1 วิธีการวัดประสิทธิภาพของแบบจำลองทางคณิตศาสตร์ .....	18
2.4.2 แบบจำลองทางคณิตศาสตร์ประเภท Random Forest.....	19
2.5 ระบบแนะนำที่อยู่อาศัย .....	19
บทที่ 3 ผลการดำเนินการ.....	21
3.1 ประสิทธิภาพของแบบจำลองทางคณิตศาสตร์ .....	21
3.2 Web Application .....	21
บทที่ 4 บทสรุป .....	25
4.1 สรุปผลการดำเนินการ.....	25
4.2 การดำเนินการในอนาคตและการสนับสนุนเพิ่มเติม .....	25
ภาคผนวก.....	26

ก. ลักษณะเครื่อง Server ที่แนะนำสำหรับการติดตั้ง Web Application และ API ของระบบแนะนำที่อยู่อาศัย.....	26
บรรณานุกรม.....	27

## สารบัญภาพ

รูปที่ 1 Dashboard สำหรับการศึกษาค้นคว้าเบื้องต้นของผู้จองโครงการการเคหะแห่งชาติในอดีต.....	9
รูปที่ 2 ขั้นตอนการแนะนำที่อยู่อาศัย (รูป icon จาก Freepik (2020)).....	19
รูปที่ 3 ตัวอย่างหน้าจอที่ใช้ในการกรอกข้อมูลของลูกค้าในการแนะนำที่อยู่อาศัย .....	22
รูปที่ 4 ตัวอย่างหน้าจอในการตรวจสอบข้อมูลก่อนการยืนยันข้อมูลเพื่อให้ระบบแสดงผลลัพธ์ต่อไป .....	22
รูปที่ 5 ตัวอย่างหน้าจอแสดงผลโครงการของการเคหะที่คาดว่าลูกค้าจะสนใจ .....	24

# สารบัญตาราง

ตารางที่ 1	รายชื่อตัวแปรในการพัฒนาแบบจำลองทำนายประเภทบ้าน .....	12
ตารางที่ 2	รายชื่อกลุ่มโครงการในการรวบรวมข้อมูลราคาและขนาดพื้นที่ .....	14
ตารางที่ 3	ฐานข้อมูลทั้งหมดของระบบแนะนำที่อยู่อาศัย .....	17
ตารางที่ 4	ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลของแบบจำลองในการทำนายประเภทบ้าน .....	18
ตารางที่ 5	Confusion Matrix.....	19
ตารางที่ 6	ค่าเปอร์เซ็นต์ความถูกต้องของแต่ละแบบจำลองทางคณิตศาสตร์ .....	21
ตารางที่ 7	ลักษณะเครื่อง Server ที่แนะนำสำหรับการติดตั้ง Web Application และ API .....	26

# บทที่ 1

## บทนำ

### 1.1 หลักการและเหตุผล

การเคหะแห่งชาติคือหน่วยงานหลักในประเทศไทยที่ทำหน้าที่ในการพัฒนาและจัดสร้างที่อยู่อาศัยให้กับประชาชนในทุกระดับรายได้ โดยเฉพาะผู้ที่มีรายได้น้อย ตลอดจนผู้ด้อยโอกาสทั้งหลายในสังคม เพื่อให้สร้างหลักประกันการมีความเป็นอยู่ที่ดีสำหรับคนทุกระดับและทุกวัย ตลอดจนลดปัญหาเกี่ยวกับความเหลื่อมล้ำในสังคม และปัญหาการขาดแคลนที่อยู่อาศัย ซึ่งสืบเนื่องมาจากปัญหาการเพิ่มขึ้นของประชากรอย่างต่อเนื่อง โดยทางการเคหะแห่งชาตินั้นได้ดำเนินการพัฒนาที่อยู่อาศัยอย่างต่อเนื่อง เพื่อให้ประชาชนจะได้มีที่อยู่อาศัยและมีคุณภาพชีวิตที่ดีทั้งด้านสังคม เศรษฐกิจ และสิ่งแวดล้อม โดยได้มีการจัดตั้งระบบต่าง ๆ เช่น ระบบสาธารณูปโภค สาธารณูปการ สิ่งอำนวยความสะดวกต่าง ๆ แก่ผู้อยู่อาศัย เป็นต้น นอกจากนี้ยังช่วยเหลือทางการเงินแก่ประชาชนที่ต้องการเช่า เช่าซื้อหรือซื้อ ตลอดจนการประกอบธุรกิจต่าง ๆ ที่เกี่ยวข้องกับการก่อสร้าง หรือการจัดการที่ดิน เพื่อให้ประชาชนได้มีที่อยู่อาศัยเป็นของตนเอง

สืบเนื่องจากหน้าที่ดังกล่าวนี้นำมาสู่แผนงานและการดำเนินการต่าง ๆ ทางการเคหะแห่งชาติ โดยการเคหะแห่งชาตินั้นเล็งเห็นถึงความสำคัญในการทำความเข้าใจถึงความต้องการของประชาชนในด้านที่อยู่อาศัยอย่างแท้จริง เพื่อนำมาพัฒนาปรับปรุงการให้บริการของการเคหะแห่งชาติที่ต้องการให้ประชาชนมีที่อยู่อาศัยที่ตรงตามความต้องการ หนทางหนึ่งที่จะสามารถช่วยให้การเคหะแห่งชาติบรรลุวัตถุประสงค์นี้คือการนำข้อมูลของลูกค้าและโครงการที่ลูกค้าเลือกในอดีตมาศึกษาและวิเคราะห์ให้เกิดประโยชน์ ซึ่งจากผลการวิเคราะห์จะช่วยให้ทางการเคหะแห่งชาติเข้าใจถึงลักษณะหรือรูปแบบของผู้จองที่อยู่อาศัย แต่ละลักษณะของผู้จองที่อยู่อาศัยจะได้รับการแนะนำที่อยู่อาศัยที่อาจจะแตกต่างกันไป โดยทั้งนี้ขึ้นอยู่กับลักษณะของผู้จองที่อยู่อาศัย ซึ่งเครื่องมือนี้จะเป็นตัวช่วยในการแนะนำแบบที่อยู่อาศัยให้ตรงกับความต้องการของประชาชนที่มาใช้บริการได้อย่างมีประสิทธิภาพ

จากความสำคัญในการศึกษาลักษณะหรือรูปแบบของผู้จองที่อยู่อาศัย เพื่อนำมาพัฒนาปรับปรุงการให้บริการของการเคหะแห่งชาติที่ต้องการให้ประชาชนมีที่อยู่อาศัยที่ตรงตามความต้องการนั้น การเคหะแห่งชาติจึงได้ร่วมมือกับสถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (GBDi) จัดทำระบบแนะนำแบบที่อยู่อาศัยขึ้น ซึ่งเป็นเครื่องมือที่สำคัญสำหรับการแนะนำที่อยู่อาศัยให้กับประชาชน โดยมีเป้าหมายเพื่อที่จะได้เข้าใจลักษณะของลูกค้าและความต้องการในด้านที่อยู่อาศัยของลูกค้าการเคหะแห่งชาติ ตลอดจนสามารถนำไปใช้ในการพัฒนาการให้บริการแก่ประชาชนโดยรวม เพื่อที่จะตอบสนองต่อเป้าหมายหลักขององค์กรเรื่องการช่วยเหลือผู้ที่มีรายได้น้อยเป็นหลัก อีกทั้งยังช่วยเสริมสร้างให้เกิดการใช้ประโยชน์จากข้อมูลและผลักดันส่งเสริมให้การเคหะแห่งชาติก้าวเข้าสู่การเป็นองค์กรที่ขับเคลื่อนด้วยข้อมูล (Data-Driven Organization) อันเป็นภารกิจหนึ่งที่สำคัญของ GBDi อีกด้วย



## 1.2 วัตถุประสงค์

เพื่อพัฒนาเครื่องมือที่ใช้การการแนะนำที่อยู่อาศัยในรูปแบบของ Web Application โดยสามารถรับข้อมูลของผู้ที่สนใจโครงการของการเคหะแห่งชาติและนำเสนอโครงการต่าง ๆ ให้ตรงกับความต้องการของบุคคลนั้น ซึ่งเป็นการพัฒนาบริการของการเคหะแห่งชาติให้ดียิ่งขึ้น

## 1.3 ผลผลิตที่คาดหวัง

1. ระบบแนะนำที่อยู่อาศัยซึ่งเรียนรู้มาจากข้อมูลในหัวข้อต่าง ๆ เช่น ข้อมูลการจอง และชุดข้อมูลรายละเอียดโครงการของการเคหะแห่งชาติ เป็นต้น
2. แบบจำลองทางคณิตศาสตร์สำหรับการแนะนำแบบที่อยู่อาศัยตามลักษณะของลูกค้าของทางการเคหะแห่งชาติ
3. Web Application สำหรับลูกค้าของทางการเคหะแห่งชาติ เพื่อใช้ในการแสดงผลรูปแบบที่อยู่อาศัยที่ได้รับการแนะนำจากระบบแนะนำที่อยู่อาศัย

## 1.4 ขอบเขตการดำเนินงานและผู้เกี่ยวข้อง

ข้อมูลในการพัฒนาระบบแนะนำที่อยู่อาศัยจากการเคหะแห่งชาติ มาจาก 2 ส่วน ดังนี้

1. ข้อมูลในการพัฒนาแบบจำลองนั้นนำจากฐานข้อมูลผู้จองโครงการของการเคหะแห่งชาติ ตั้งแต่ พ.ศ.2550 ถึง พ.ศ.2562 จำนวนทั้งสิ้น 1,271,461 รายการ
2. ข้อมูลโครงการของการเคหะแห่งชาติที่มีหน่วยว่างทั้งหมด 543 โครงการ โดยข้อมูลจำนวนหน่วยว่างอัปเดตล่าสุดในเดือนกรกฎาคม พ.ศ.2564 และข้อมูลราคาของโครงการอัปเดตล่าสุดในช่วง พ.ศ.2562 ถึง พ.ศ.2563

# บทที่ 2

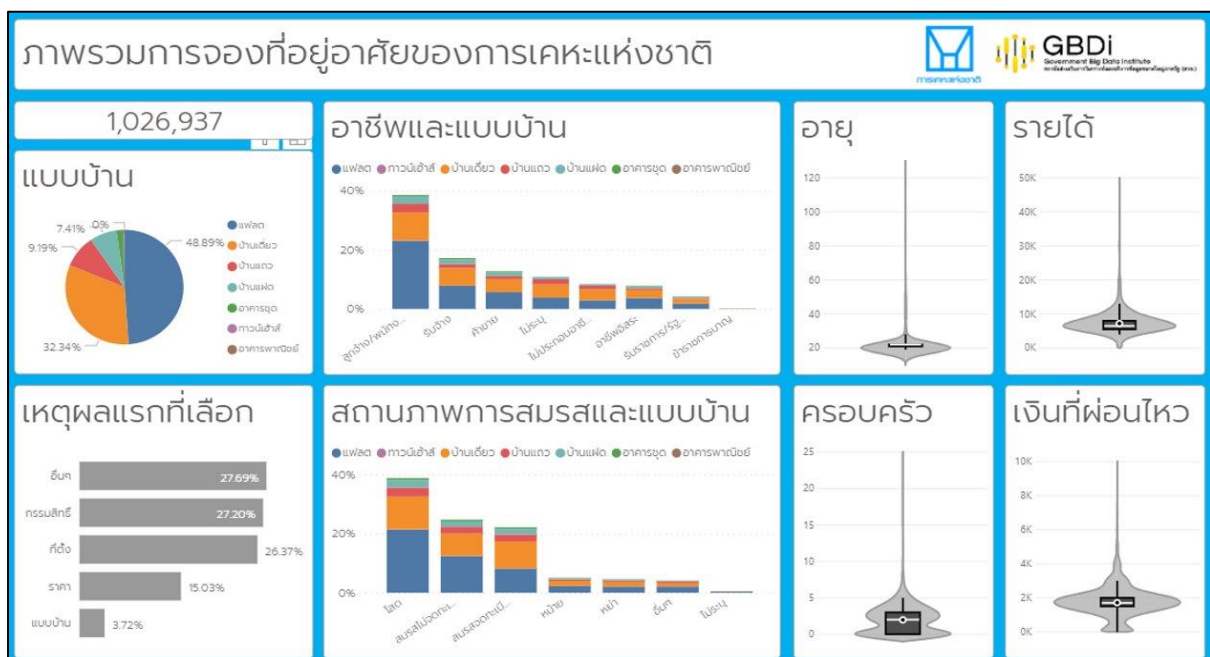
## ระเบียบวิธีดำเนินการ

### 2.1 การเตรียมข้อมูล

#### 2.1.1 การเตรียมข้อมูลลูกค้าจากการเคหะแห่งชาติ

ข้อมูลที่นำมาใช้ในการดำเนินการวิเคราะห์ลักษณะของลูกค้าของการเคหะแห่งชาติ เพื่อใช้ในการแนะนำแบบที่อยู่อาศัยให้ลูกค้า นั่นคือ ข้อมูลผู้จอง (F5) และข้อมูลประเภทที่อยู่อาศัย (cl\_model) โดยก่อนที่จะดำเนินการทำความสะอาดและเตรียมข้อมูล เราได้ทำการศึกษาข้อมูลเบื้องต้น (Exploratory Data Analysis) ในมุมมองต่าง ๆ ของตัวแปรบางประเภท และได้มีการจัดทำหน้าจอสถูมข้อมูล (Dashboard) นำร่องการวิเคราะห์ข้อมูลเชิงพรรณนาในหัวข้อต่าง ๆ ดังนี้ โดยมีการศึกษาข้อมูลประเภทที่อยู่อาศัย อาชีพ จำนวนสมาชิกในครอบครัว จำนวนเงินที่คาดว่าจะผ่อนได้จริง และอายุ รายได้ เหตุผลที่เลือกจองบ้าน ตลอดจนสถานภาพการสมรส

ข้อมูลที่แสดงใน Dashboard ในรูปที่ 1 เป็นข้อมูลที่ได้รับจากการเคหะแห่งชาติเพื่อใช้ในการศึกษาข้อมูลลูกค้าเบื้องต้น ซึ่งแสดงให้เห็นว่า ประเภทที่อยู่อาศัยจะมีข้อมูลจะมีลักษณะข้อมูลไม่สมดุล (Imbalanced Data) นั่นคือข้อมูลที่มีจำนวนข้อมูลในกลุ่มหนึ่งมากกว่าจำนวนข้อมูลของอีกกลุ่มหนึ่งเป็นจำนวนมาก



รูปที่ 1 Dashboard สำหรับการศึกษข้อมูลเบื้องต้นของผู้จองโครงการการเคหะแห่งชาติในอดีต

ซึ่งจากกราฟต่าง ๆ ใน Dashboard นั้นจะสังเกตเห็นได้เลยว่าลูกค้าจองประเภทที่อยู่อาศัยที่เป็นแฟลตกับบ้านเดี่ยวจำนวนมากกว่าประเภทที่อยู่อาศัยแบบอื่น ๆ รวมถึงลูกค้าส่วนใหญ่เป็นคนที่ประกอบอาชีพลูกจ้าง/พนักงานบริษัท

จากนั้นเมื่อได้ข้อมูลผู้จองเพิ่มเติมจากทางการเคหะแห่งชาติมาแล้ว จึงจะเริ่มทำความสะอาดข้อมูลที่ได้รับมาทั้งหมด ซึ่งข้อมูลเหล่านี้มีรายละเอียดในการจัดเก็บรูปแบบข้อมูลและการเตรียมข้อมูล ดังต่อไปนี้

1. ใช้ข้อมูลผู้จอง (F5) ที่เป็นข้อมูลเชิงประชากร (Demographic data) ซึ่งเป็นข้อมูลประวัติล่าสุดของผู้เช่าซื้อแต่ละสัญญา นำไปรวม (Merge) กับข้อมูลประเภทบ้าน (cl\_model) ผ่านรหัสโครงการ (PROJ\_ID) เพื่อให้ได้ประเภทบ้าน (MODEL\_NAME) จากนั้นทำการจัดกลุ่มประเภทบ้าน (MODEL\_NAME) ให้เหลือ 6 ประเภทได้แก่ บ้านเดี่ยว บ้านแฝด บ้านแถว ทาวน์เฮ้าส์ อาคารพาณิชย์ แฟลตหรืออาคารชุด
2. ทำความสะอาดคอลัมน์จำนวนสมาชิกในครอบครัว (FAMILY\_TOTAL) และจำนวนบุตร (SON\_TOTAL) โดยจัดการเปลี่ยนรูปแบบข้อมูลให้อยู่ในเชิงตัวเลขเหมือนกัน จากนั้นเลือกข้อมูล (Filter) จำนวนสมาชิกในครอบครัว (FAMILY\_TOTAL) และจำนวนบุตร (SON\_TOTAL) ที่น้อยกว่าหรือเท่ากับ 15 คนเท่านั้น โดยไม่นำข้อมูลที่คาดว่าจะรอกผิดมาใช้
3. ทำความสะอาดคอลัมน์ประเภทอาชีพ (OCCUPATION) และชื่ออาชีพ (OCCUPATION\_TH) โดยทำการจัดกลุ่มส่วนที่ไม่มีข้อมูลหรือไม่ระบุเป็นอีกประเภท
4. ทำความสะอาดคอลัมน์ประเภทสถานภาพการสมรส (MARITAL\_STATUS) และชื่อสถานภาพการสมรส (MARITAL\_STATUS\_TH) โดยทำการจัดกลุ่มส่วนที่ไม่มีข้อมูลหรือไม่ระบุเป็นอีกประเภท
5. ทำความสะอาดคอลัมน์รายได้ของตนเอง (OWN\_INCOME) รายได้คู่สมรส (SPOUSE\_INCOME) และ ค่าผ่อนชำระที่มากที่สุด (MAX\_RATE) เติมข้อมูลให้ถูกต้อง เช่น กรณีที่เป็นคนโสด แต่มีรายได้ของคู่สมรส เนื่องจากข้อมูลรายได้ เป็นต้น รวมถึงมีการโฟกัสไปที่ผู้มีรายได้น้อย โดยจะเลือกข้อมูลผู้ที่มีรายได้น้อยกว่าหรือเท่ากับ 500,000 มาใช้
6. ตรวจสอบข้อมูลรายได้ โดยการสร้างคอลัมน์ใหม่ที่ชื่อ รายได้โดยรวม (TOTAL\_INCOME) ซึ่งคำนวณมาจากคอลัมน์รายได้ของตนเอง (OWN\_INCOME) รวมกับรายได้คู่สมรส (SPOUSE\_INCOME) หักลบกับค่าเช่าบ้านหรือค่าใช้จ่ายในบ้าน (RENT\_RATE) จากนั้นเลือกข้อมูล (Filter) ที่รายได้โดยรวมที่ค่ามากกว่าศูนย์ โดยไม่นำข้อมูลที่คาดว่าจะรอกผิดมาใช้
7. ทำความสะอาดคอลัมน์อายุ (AGE) โดยเลือกข้อมูล (Filter) อายุที่บรรลุนิติภาวะเท่านั้น
8. ในคอลัมน์ลักษณะผู้จอง (PERSONAL\_TYPE) เลือกข้อมูล (Filter) เฉพาะบุคคลทั่วไปและชาวต่างชาติเท่านั้น
9. ทำความสะอาดคอลัมน์เพศ (SEX) โดยทำการจัดกลุ่มส่วนที่ไม่มีข้อมูลหรือไม่ระบุเป็นอีกประเภท

10. ทำความสะอาดคอลัมน์สถานภาพการอยู่อาศัย (LIVE\_STATUS) โดยจัดกลุ่มข้อมูลที่เป็นข้อความอิสระ (Free text) ตามประเภทสถานภาพการอยู่อาศัยและจัดกลุ่มส่วนที่ไม่มีข้อมูลหรือไม่ระบุเป็นอีกประเภท
  11. ทำความสะอาดคอลัมน์เหตุที่เลือกซื้อที่อยู่อาศัยนั้น ๆ (REASON1, REASON2, REASON3, REASON4, REASON5, REASON6, REASON6\_DESCR) กรณีที่ไม่มีข้อมูล ได้มีการเติมข้อมูลให้ถูกต้อง
  12. ทำความสะอาดข้อมูลเหตุผลประการแรกที่เลือกที่อยู่อาศัย (EXPECTED) โดยทำการจัดกลุ่มส่วนที่ไม่มีข้อมูลหรือไม่ระบุเป็นอีกประเภท
  13. จากนั้นจัดการข้อมูลขั้นตอนสุดท้าย โดยการเลือกข้อมูล (Filter) ประเภทสถานภาพการสมรส (MARITAL\_STATUS) และสถานภาพการอยู่อาศัย (LIVE\_STATUS) กรณีอื่น ๆ ยกเว้นกรณีเดียวคือที่ไม่ระบุ
  14. เมื่อทำความสะอาดเสร็จเรียบร้อยแล้ว ใช้เทคนิค One-hot Encoding จัดการคอลัมน์ต่าง ๆ ได้แก่ อาชีพ (OCCUPATION), สถานภาพการสมรส (MARITAL\_STATUS) และ สถานภาพการอยู่อาศัย (LIVE\_STATUS) เป็นต้น
- ซึ่งผลลัพธ์ที่ได้จากการเตรียมข้อมูลนั้น มีข้อมูลที่สมบูรณ์จำนวนประมาณ 1,271,461 รายการ และมีจำนวนตัวแปรทั้งหมด 31 ตัวแปรดังที่แสดงในตารางที่ 1 สำหรับนำไปให้แบบจำลองเรียนรู้ เพื่อใช้ในการแนะนำประเภทที่อยู่อาศัยที่ตรงกับความต้องการของลูกค้า

ตัวแปร	ความหมาย
1. AGE	อายุ (ปี)
2. FAMILY_TOTAL	จำนวนสมาชิกในครอบครัว
3. RENT_RATE	ค่าเช่าบ้านหรือค่าใช้จ่ายในบ้าน
4. MAX_RATE	ค่าผ่อนชำระที่มากที่สุด
5. OWN_INCOME	รายได้ของตนเอง
6. SPOUSE_INCOME	รายได้คู่สมรส
7. REASON1	จองเพราะอยู่ใกล้แหล่งชุมชนและบริการ
8. REASON2	จองเพราะการคมนาคมสะดวก
9. REASON3	จองเพราะอยู่ใกล้ที่ทำงานหรือโรงเรียนบุตร
10. REASON4	จองเพราะอัตราค่าผ่อนถูก/ราคาถูก
11. REASON5	จองเพราะเป็นโครงการบ้านพร้อมที่ดิน
12. REASON6	จองเพราะสาเหตุอื่น ๆ

ตัวแปร	ความหมาย
13. EXPECTED__1	เหตุผลประการแรกที่เลือกที่อยู่อาศัยของท่าน คือ แบบบ้าน
14. EXPECTED__2	เหตุผลประการแรกที่เลือกที่อยู่อาศัยของท่าน คือ ที่ตั้ง
15. EXPECTED__3	เหตุผลประการแรกที่เลือกที่อยู่อาศัยของท่าน คือ ราคา
16. EXPECTED__4	เหตุผลประการแรกที่เลือกที่อยู่อาศัยของท่าน คือ กรรมสิทธิ์
17. OCCUPATION__ค้าขาย	อาชีพเป็นค้าขาย
18. OCCUPATION__รับจ้าง	อาชีพเป็นรับจ้าง
19. OCCUPATION__รับราชการ/รัฐวิสาหกิจ	อาชีพเป็นรับราชการ/รัฐวิสาหกิจ
20. OCCUPATION__ลูกจ้าง/พนักงานบริษัท	อาชีพเป็นลูกจ้าง/พนักงานบริษัท
21. OCCUPATION__อาชีพอิสระ	อาชีพเป็นอาชีพอิสระ
22. OCCUPATION__ไม่ประกอบอาชีพ	อาชีพเป็นไม่ประกอบอาชีพ
23. MARITAL_STATUS__สมรสไม่จดทะเบียน	สถานภาพการสมรสเป็นสมรสไม่จดทะเบียน
24. MARITAL_STATUS__หม้าย	สถานภาพการสมรสเป็นหม้าย
25. MARITAL_STATUS__หย่า	สถานภาพการสมรสเป็นหย่า
26. MARITAL_STATUS__อื่น ๆ	สถานภาพการสมรสเป็นอื่น ๆ
27. MARITAL_STATUS__โสด	สถานภาพการสมรสเป็นโสด
28. LIVE_STATUS__ อยู่บ้านผู้อื่นบนที่ดินตัวเอง	สถานภาพการอยู่อาศัยอยู่บ้านผู้อื่นบนที่ดินตัวเอง
29. LIVE_STATUS__อาศัยผู้อื่น	สถานภาพการอยู่อาศัยกับผู้อื่น
30. LIVE_STATUS__อื่น ๆ	สถานภาพการอยู่อาศัยอื่น ๆ
31. LIVE_STATUS__เช่าบ้าน	สถานภาพการอยู่อาศัยเช่าบ้าน

ตารางที่ 1 รายชื่อตัวแปรในการพัฒนาแบบจำลองทำนายประเภทบ้าน

### 2.1.2 การเตรียมข้อมูลโครงการ

ข้อมูลที่สำคัญของโครงการการเคหะแห่งชาติ ได้แก่ จำนวนหน่วยที่ว่าง ตำแหน่งโครงการ ราคาและขนาดพื้นที่ของโครงการ เป็นต้น ซึ่งข้อมูลแต่ละส่วนนั้นมีวิธีการทำความเข้าใจแตกต่างกัน แต่ส่วนนี้มีรายละเอียดดังต่อไปนี้

1. การเตรียมข้อมูลรายชื่อโครงการ ณ ปัจจุบัน หน่วยว่างของประเภทที่อยู่อาศัยแต่ละประเภท และตำแหน่งของโครงการ

ข้อมูลส่วนนี้ประกอบด้วยรหัสโครงการ ชื่อโครงการ จำนวนหน่วยว่างของแต่ละประเภทที่อยู่อาศัยในโครงการ และที่อยู่ระดับบ้านเลขที่ ถนน ตำบล อำเภอและจังหวัด แต่ข้อมูลที่อยู่โครงการนั้นขาดความสมบูรณ์และส่วนใหญ่จะมีข้อมูลเพียงระดับตำบล อำเภอ และจังหวัดเท่านั้น ทางทีมงานจึงตัดสินใจเติมพิกัดทางภูมิศาสตร์ระดับ Latitude และ Longitude โดยประมาณจากข้อมูลอ้างอิง SERVICE.meewebfree (2020) ซึ่งนอกเหนือจากมีข้อมูลจังหวัด ตำบล อำเภอ พร้อมทั้ง Latitude และ Longitude แล้ว ยังมีข้อมูลรหัสจังหวัด รหัสอำเภอและรหัสตำบลอีกด้วย

อย่างไรก็ตาม ข้อมูลอ้างอิง SERVICE.meewebfree (2020) ยังขาดข้อมูลรหัสไปรษณีย์ และ Latitude และ Longitude ในบางตำแหน่ง ดังนั้นทีมงานจึงเริ่มต้นด้วยการเติมข้อมูลรหัสไปรษณีย์เองดังนี้

- a. ในกรณีรหัสอำเภอเดียวกันมีรหัสไปรษณีย์เดียว ทีมงานจะเติมรหัสไปรษณีย์ตามรหัสอำเภอเดียวกัน
- b. ในกรณีที่รหัสอำเภอเดียวกันแต่มีรหัสไปรษณีย์หลายรหัส ทีมงานจะตรวจสอบตำแหน่งนั้นเพิ่มเติมจาก Thailandpostcode (ม.ป.ป.) เพื่อใช้ในการเติมรหัสไปรษณีย์ให้แม่นยำมากยิ่งขึ้น

หลังจากทุกตำแหน่งมีรหัสไปรษณีย์ครบแล้ว จึงเติมตำแหน่ง Latitude และ Longitude ของตำแหน่งที่ขาดไปให้ตรงกับข้อมูลของตำแหน่งของรหัสไปรษณีย์เดียวกันด้วยวิธี ffill

2. การเตรียมข้อมูลราคาและขนาดพื้นที่ของโครงการ (ในหน่วยตารางวาหรือตารางเมตร)

**ตารางที่ 2** แสดงถึงกลุ่มโครงการที่ใช้ในการหาข้อมูลเพิ่มเติมในส่วนของราคาและขนาดพื้นที่ รวมถึงปีที่มีการอัปเดตข้อมูลเหล่านั้นล่าสุด

กลุ่มโครงการ	ปีของข้อมูล (อย่างไม่เป็นทางการ)
โครงการเคหะชุมชน	2564
โครงการพัฒนาที่อยู่อาศัย ชุดที่ 1 แผนงานปี 2557 - 2559 : โครงการรายได้น้อยถึงปานกลาง และ โครงการหารายได้	2562
โครงการพัฒนาที่อยู่อาศัย ชุดที่ 2 ระยะที่ 1 : โครงการเคหะชุมชนและบริการชุมชน (เชิงสังคมและเชิงพาณิชย์)	2562
โครงการพัฒนาที่อยู่อาศัยตามแนวเส้นทางรถไฟฟ้ามหานคร และปริมณฑล	2562

กลุ่มโครงการ	ปีของข้อมูล (อย่างไม่เป็นทางการ)
โครงการบ้านเคหะกตัญญู คลองหลวง 1	2562
โครงการบ้านเอื้ออาทร	2563
โครงการบ้านสวัสดิการข้าราชการ (เช่าซื้อ)	2562

## ตารางที่ 2 รายชื่อกลุ่มโครงการในการรวบรวมข้อมูลราคาและขนาดพื้นที่

ทีมงานได้ดำเนินการทำความสะอาดข้อมูลด้วยการเลือกเก็บเพียงข้อมูลรหัสโครงการ ชื่อโครงการ รูปแบบที่อยู่อาศัย รูปแบบที่อยู่อาศัยอย่างละเอียด ราคา และขนาดพื้นที่ ในกรณีโครงการที่มีรูปแบบที่อยู่อาศัยประเภทแฟลตและมีราคาต่างกันตามชั้นของอาคาร ทางทีมงานจะคำนวณราคาเฉลี่ยเพื่อนำมาเป็นตัวแทนของราคาสำหรับโครงการนั้น หลังจากได้ข้อมูลเหล่านี้ครบถ้วนแล้วจึงเชื่อมกับข้อมูลโครงการที่มีหน่วยว่างด้วยรหัสโครงการ

อย่างไรก็ตาม ข้อมูลในส่วนของราคาและขนาดพื้นที่ไม่ได้ครอบคลุมทุกโครงการที่มีหน่วยว่าง ณ ปัจจุบัน ทางทีมงานจึงได้ตัดโครงการเหล่านี้ออกจากการนำมาแสดงผลที่หน้าจอของ Web Application

## 2.2 การเตรียมฐานข้อมูล

ทางทีมงานได้มีการเตรียมฐานข้อมูล เพื่อใช้ในการจัดเก็บข้อมูลโครงการที่อยู่อาศัยต่าง ๆ ของทางการเคหะแห่งชาติ และเพื่อรองรับการใช้งาน Web Application สำหรับลูกค้าของทางการเคหะแห่งชาติ ที่ใช้ในการแสดงผลรูปแบบที่อยู่อาศัยที่ได้รับการแนะนำจากระบบแนะนำที่อยู่อาศัย ซึ่งจะประกอบไปด้วยตารางต่าง ๆ ดังในตารางที่ 3

ชื่อตาราง	ชื่อ Feature	ความหมายของ Feature ต่าง ๆ
nha_house เพื่อใช้ในการจัดเก็บข้อมูล โครงการที่อยู่อาศัยต่าง ๆ ของการเคหะแห่งชาติ	price	ราคา
	lat	ละติจูด
	long	ลองจิจูด
	space	ขนาดพื้นที่
	zipcode	รหัสไปรษณีย์
	model_plan	ประเภทแบบบ้าน
	model_name	ชื่อโครงการบ้านแบบละเอียด
	model_plan_id	รหัสประเภทแบบบ้าน

ชื่อตาราง	ชื่อ Feature	ความหมายของ Feature ต่าง ๆ
	model_name_id	รหัสโครงการบ้านแบบละเอียด
	nha_id	ลำดับโครงการและบ้าน
	model_plan_name	ประเภทบ้านแบบละเอียด
	source	ชื่อไฟล์อ้างอิง
	note	หมายเหตุ
	available_unit	จำนวนหน่วยว่าง
provinces เพื่อใช้ในการจัดเก็บข้อมูล เกี่ยวกับพิกัดทางภูมิศาสตร์ ของจังหวัดต่าง ๆ ใน ประเทศไทย	province_id	รหัสจังหวัด
	name_th	ชื่อจังหวัดภาษาไทย
	name_en	ชื่อจังหวัดภาษาอังกฤษ
	distinct_id	รหัสอำเภอ/เขต
	name_th_x	ชื่ออำเภอ/เขตภาษาไทย
	name_en_x	ชื่ออำเภอ/เขตภาษาอังกฤษ
	sub_dist_id	รหัสตำบล/แขวง
	name_th_y	ชื่ออำเภอ/เขตภาษาไทย
	name_en_y	ชื่ออำเภอ/เขตภาษาอังกฤษ
	lat	ละติจูด
	long	ลองจิจูด
	zipcode	รหัสไปรษณีย์
	note	หมายเหตุ
input_records เพื่อใช้ในการจัดเก็บข้อมูลที่ ทางลูกค้ากรอกจาก Web Application	family_total	จำนวนสมาชิกในครอบครัว
	rent_rate	ค่าเช่าบ้านหรือค่าใช้จ่ายในบ้านที่ท่านอยู่แต่ ละเดือน
	max_rate	ค่าผ่อนที่สามารถให้ชำระได้มากที่สุด
	own_income	รายได้ของตนเอง
	spouse_income	รายได้ของคู่สมรส
	occupation_1	อาชีพเป็นค้าขาย
	occupation_2	อาชีพเป็นรับจ้าง
	occupation_3	อาชีพเป็นรับราชการ/รัฐวิสาหกิจ
	occupation_4	อาชีพเป็นลูกจ้าง/พนักงานบริษัท
occupation_5	อาชีพเป็นอาชีพอิสระ	



ชื่อตาราง	ชื่อ Feature	ความหมายของ Feature ต่าง ๆ
	occupation_6	อาชีพเป็นหรือไม่ประกอบอาชีพ
	reason_1	เหตุผลที่จองเพราะอยู่ใกล้แหล่งชุมชนและบริการ
	reason_2	เหตุผลที่จองเพราะการคมนาคมสะดวก
	reason_3	เหตุผลที่จองเพราะอยู่ใกล้ที่ทำงานหรือโรงเรียนบุตร
	reason_4	เหตุผลที่จองเพราะอัตราค่าผ่อนถูก/ราคาถูก
	reason_5	เหตุผลที่จองเพราะเป็นโครงการบ้านพร้อมที่ดิน
	reason_6	เหตุผลที่จองอื่น ๆ
	y_pred_1	ผลการแนะนำบ้านแบบแรก
	y_pred_2	ผลการแนะนำบ้านแบบที่สอง
	age	อายุ
	gbdi_id	ลำดับข้อมูลที่กรอกเข้ามาผ่านหน้าเว็บ
	house_zipcode	รหัสไปรษณีย์ของบ้าน
	work_zipcode	รหัสไปรษณีย์ของที่ทำงาน
	result_show เพื่อใช้ในการจัดเก็บข้อมูลที่ใช้ ใช้ แสดงผลบน Web Application	price
model_plan_id		รหัสประเภทแบบบ้าน
available_unit		จำนวนหน่วยว่าง
work_distance		ระยะทางจากที่ทำงานปัจจุบันถึงโครงการบ้าน จัดสรรนั้น ๆ ที่แนะนำ
house_distance		ระยะทางจากที่บ้านที่อาศัยปัจจุบันถึงโครงการ บ้านจัดสรรนั้น ๆ ที่แนะนำ
affordable		Flag ที่ระบุว่าลูกค้าสามารถจ่ายค่าบ้านได้จริง หรือไม่
lat		ละติจูด
long		ลองจิจูด
space		ขนาดพื้นที่
zipcode		รหัสไปรษณีย์
model_name_id		รหัสโครงการบ้านแบบละเอียด

ชื่อตาราง	ชื่อ Feature	ความหมายของ Feature ต่าง ๆ
	house_lat	ละติจูดของบ้าน
	house_long	ลองติจูดของที่บ้าน
	work_lat	ละติจูดของที่ทำงาน
	work_long	ลองติจูดของที่ทำงาน
	house_zipcode	รหัสไปรษณีย์ของบ้าน
	work_zipcode	รหัสไปรษณีย์ของที่ทำงาน
	gbdi_id	ลำดับข้อมูลที่กรอกเข้ามาผ่านหน้าเว็บ
	run_no	ลำดับข้อมูลในตาราง

### ตารางที่ 3 ฐานข้อมูลทั้งหมดของระบบแนะนำที่อยู่อาศัย

## 2.3 แบบจำลองคณิตศาสตร์ทำนายประเภทที่อยู่อาศัย

หลังจากที่ได้ผ่านการศึกษาค้นคว้าข้อมูลเบื้องต้นและจัดการข้อมูลเบื้องต้น เพื่อเปลี่ยนประเภทให้เหมาะสม และตัดหรือเติมข้อมูลที่ขาด พบว่ามีจำนวนข้อมูลประมาณ 1,271,461 รายการที่มีความสมบูรณ์ โดยก่อนที่จะนำข้อมูลเข้าไปใช้ให้แบบจำลองทางคณิตศาสตร์เรียนรู้นั้น จำเป็นต้องดำเนินการอีกสองขั้นตอนคือ การแบ่งข้อมูลเพื่อใช้ฝึกฝนแบบจำลองทางคณิตศาสตร์และใช้ทดสอบแบบจำลองทางคณิตศาสตร์ และการเลือกตัวแปรที่จะนำไปใช้ในแบบจำลองทางคณิตศาสตร์

### 2.3.1 การแบ่งข้อมูลเพื่อใช้ฝึกฝนและทดสอบแบบจำลองทางคณิตศาสตร์ (Train-Test Split)

ข้อมูลทั้งหมดจะถูกแบ่งออกเป็น 2 ส่วน โดยใช้ส่วนหนึ่งในการฝึกฝนและอีกส่วนใช้ในการทดสอบ ส่วนแรก Training Set เป็นปริมาณ 70% ของข้อมูลทั้งหมด เพื่อใช้สำหรับให้แบบจำลองทางคณิตศาสตร์ฝึกฝน เรียนรู้ และอีกส่วน Test Set เป็นปริมาณ 30% ของข้อมูลทั้งหมด จะถูกแยกออกไปเพื่อทดสอบประสิทธิภาพแบบจำลองทางคณิตศาสตร์ ซึ่งแบบจำลองทางคณิตศาสตร์จะไม่เคยเห็นข้อมูลชุดนี้มาก่อน

### 2.3.2 การเลือกตัวแปรที่นำไปใช้ในแบบจำลองทางคณิตศาสตร์ (Feature Selection)

เนื่องจากการใช้ตัวแปรที่ไม่เกี่ยวข้องกับสิ่งที่ต้องการจะทำนายจะไม่ช่วยให้แบบจำลองทางคณิตศาสตร์ทำนายได้ดีขึ้น และอาจส่งผลเสียกับแบบจำลองทางคณิตศาสตร์ในด้านอื่น ๆ ได้ ทางทีมงานจึงได้เลือกตัวแปรที่มีผลในการแนะนำแบบที่อยู่อาศัยของลูกค้า โดยการหาความสำคัญของตัวแปร (Feature Importance) ด้วยวิธีที่เหมาะสมกับแบบจำลองทางคณิตศาสตร์แต่ละประเภท ตัวแปรที่ถูกเลือกนำไปใช้ในแบบจำลองทางคณิตศาสตร์แสดงในตารางที่ 4 ได้แก่ อายุ จำนวนสมาชิกในครอบครัว ค่าเช่าบ้านหรือค่าใช้จ่ายในบ้านค่าผ่อนชำระที่มากที่สุด รายได้ตนเอง รายได้คู่สมรส เหตุผลในการจอง และ อาชีพ ตัวแปรอื่น ๆ ที่อยู่ในชุดข้อมูลนอกเหนือจากนี้จะไม่ถูกนำมาใช้

ตัวแปร	ความหมาย
1. AGE	อายุ (ปี)
2. FAMILY_TOTAL	จำนวนสมาชิกในครอบครัว
3. RENT_RATE	ค่าเช่าบ้านหรือค่าใช้จ่ายในบ้าน
4. MAX_RATE	ค่าผ่อนชำระที่มากที่สุด
5. OWN_INCOME	รายได้ของตนเอง
6. SPOUSE_INCOME	รายได้คู่สมรส
7. REASON1	จงเพราะอยู่ใกล้แหล่งชุมชนและบริการ
8. REASON2	จงเพราะการคมนาคมสะดวก
9. REASON3	จงเพราะอยู่ใกล้ที่ทำงานหรือโรงเรียนบุตร
10. REASON4	จงเพราะอัตราค่าผ่อนถูก/ราคาถูก
11. REASON5	จงเพราะเป็นโครงการบ้านพร้อมที่ดิน
12. REASON6	จงเพราะสาเหตุอื่น ๆ
13. OCCUPATION__ ค้าขาย	อาชีพเป็นค้าขาย
14. OCCUPATION__ รับจ้าง	อาชีพเป็นรับจ้าง
15. OCCUPATION__ รับราชการ/รัฐวิสาหกิจ	อาชีพเป็นรับราชการ/รัฐวิสาหกิจ
16. OCCUPATION__ ลูกจ้าง/พนักงานบริษัท	อาชีพเป็นลูกจ้าง/พนักงานบริษัท
17. OCCUPATION__ อาชีพอิสระ	อาชีพเป็นอาชีพอิสระ
18. OCCUPATION__ ไม่ประกอบอาชีพ	อาชีพเป็นไม่ประกอบอาชีพ

ตารางที่ 4 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลของแบบจำลองในการทำนายประเภทบ้าน

## 2.4 การเลือกแบบจำลองทางคณิตศาสตร์

ทางทีมงานจัดทำได้ศึกษาและทดลองใช้แบบจำลองทางคณิตศาสตร์ 3 ประเภท ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Network), Random Forest และ XGBOOST หลังจากได้ทดลองและเปรียบเทียบผลลัพธ์ จึงได้ตัดสินใจนำแบบจำลองทางคณิตศาสตร์ประเภท Random Forest ไปพัฒนาต่อเพื่อให้เหมาะกับการนำไปใช้งานจริง โดยมีรายละเอียดเกี่ยวกับวิธีการวัดประสิทธิภาพ และเหตุผลในการเลือกแบบจำลองทางคณิตศาสตร์ ดังนี้

### 2.4.1 วิธีการวัดประสิทธิภาพของแบบจำลองทางคณิตศาสตร์

วิธีการวัดประสิทธิภาพของแบบจำลองทางคณิตศาสตร์ในโครงการนี้นั้น ขึ้นอยู่กับความถูกต้องของแบบจำลองทางคณิตศาสตร์ จึงเลือกใช้ Recall โดยคำนวณจากความน่าจะเป็นที่ข้อมูลที่ทำนายตรงกับข้อมูลจริงจากค่าเหตุการณ์จริงทั้งหมดซึ่งมีสูตรการคำนวณดังนี้

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

โดยการคำนวณของ True Positive และ False Negative เป็นไปตาม Confusion Matrix ดังในตารางที่ 5

Predicted / Actual	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

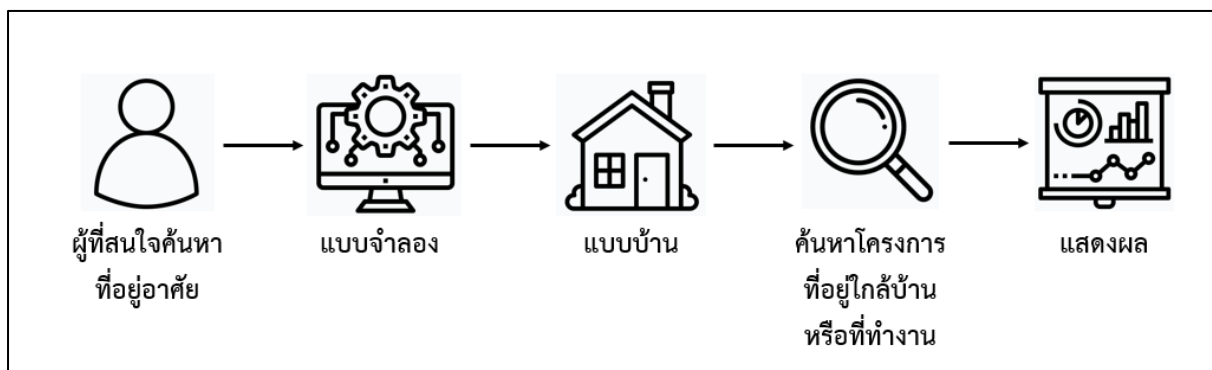
ตารางที่ 5 Confusion Matrix

#### 2.4.2 แบบจำลองทางคณิตศาสตร์ประเภท Random Forest

แบบจำลองทางคณิตศาสตร์ประเภท Random Forest เป็นแบบจำลองที่เป็นที่นิยมและได้รับความสนใจอย่างมากในปัจจุบัน โดยแบบจำลองประเภทนี้นั้นสร้างจากต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ ต้น โดยที่แต่ละต้นจะได้เรียนรู้ข้อมูลในแต่ละชุดข้อมูลไม่เหมือนกัน ทำให้ต้นไม้แต่ละต้นนั้นมีความหลากหลายและความอิสระจากกัน โดยการทำนายผลของต้นไม้แต่ละต้นนั้น จะทำนายแยกกัน จากนั้นผลการทำนายในแต่ละต้นจะถูกสรุปออกมา Random Forest นั้นเป็นแบบจำลองที่ค่อนข้างแม่นยำ เหมาะกับข้อมูลที่มีโครงสร้าง (Structured Data) และไม่มีโครงสร้าง (Unstructured Data)

#### 2.5 ระบบแนะนำที่อยู่อาศัย

ขั้นตอนในการแนะนำที่อยู่อาศัยผู้ที่สนใจโครงการของการเคหะแห่งชาติบน Web Application นั้น จะมีขั้นตอนต่าง ๆ ดังรูปที่ 2



รูปที่ 2 ขั้นตอนการแนะนำที่อยู่อาศัย (รูป icon จาก Freepik (2020))

ขั้นตอนแรก เริ่มต้นจากผู้ที่สนใจโครงการของการเคหะแห่งชาติกรอกข้อมูลส่วนตัวที่มีความเกี่ยวข้องในการแนะนำที่อยู่อาศัย ได้แก่ ข้อมูลส่วนตัวต่าง ๆ ตาม**ตารางที่ 4** และข้อมูลรหัสไปรษณีย์ของที่อยู่ปัจจุบันและที่ทำงาน หลังจากนั้น ระบบจะนำข้อมูลของบุคคลนั้นเข้าสู่แบบจำลองทางคณิตศาสตร์เพื่อทำนายประเภทที่อยู่อาศัยที่บุคคลนั้นจะสนใจ 2 อันดับแรก

ในส่วนของการค้นหาโครงการของการเคหะแห่งชาติที่คาดว่าจะน่าสนใจสำหรับบุคคลนั้น ระบบจะนำข้อมูลรหัสไปรษณีย์ของที่อยู่ปัจจุบันและที่ทำงานของบุคคลนั้นที่ได้รับมาจาก Web Application มาเตรียมค่า Latitude และ Longitude โดยประมาณ โดยใช้ฐานข้อมูล province (ใน**ตารางที่ 3**) จากนั้นระบบจะคำนวณระยะทางระหว่างโครงการและที่อยู่ปัจจุบันและที่ทำงานโดยประมาณ ก่อนที่จะเลือกโครงการที่อยู่ใกล้ที่อยู่ปัจจุบันและที่ทำงาน 10 อันดับแรกมาแสดงผลบนหน้าจอ Web Application ในรูปแบบตารางรายชื่อโครงการและแผนที่ Google Maps นอกเหนือจากนี้ใน Web Application ผู้กรอกข้อมูลยังสามารถเลือกให้แสดงเฉพาะโครงการที่ระบบแนะนำว่าจะผ่อนชำระได้จริง (โดยใช้หลักการว่าลูกค้าสามารถผ่อนชำระค่าที่อยู่อาศัยหมดก่อนอายุ 60 ปี) ได้อีกด้วย

# บทที่ 3

## ผลการดำเนินการ

### 3.1 ประสิทธิภาพของแบบจำลองทางคณิตศาสตร์

ทางทีมงานได้มีการจัดทำและได้ดำเนินการทดสอบแบบจำลองทางคณิตศาสตร์ โดยมีการคัดเลือกแบบจำลองมา 3 ประเภท ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Network) Random Forest และ XGBOOST จากผลการทดสอบประสิทธิภาพแบบจำลองที่คัดเลือกมานั้นพบว่าแบบจำลองทางคณิตศาสตร์ประเภท Random Forest นั้นมีประสิทธิภาพมากที่สุด ซึ่งมีเปอร์เซ็นต์ความถูกต้องอยู่ที่ 84.32% เมื่อเปรียบเทียบกับแบบจำลองทางคณิตศาสตร์อีกสองประเภทดังแสดงในตารางที่ 6 แบบจำลองทางคณิตศาสตร์ที่มีประสิทธิภาพมากเป็นอันดับที่สองคือ Extreme Gradient Boosting ซึ่งมีเปอร์เซ็นต์ความถูกต้องอยู่ที่ 84.30% ซึ่งค่อนข้างใกล้เคียงกับแบบจำลองทางคณิตศาสตร์ประเภท Random Forest

ประเภทแบบจำลองทางคณิตศาสตร์	เปอร์เซ็นต์ความถูกต้อง
Baseline	57.80%
แบบจำลอง 1 (Artificial Neural Network)	81.77%
<b>แบบจำลอง 2 (Random Forest)</b>	<b>84.32%</b>
แบบจำลอง 3 (Extreme Gradient Boosting)	84.30%

ตารางที่ 6 ค่าเปอร์เซ็นต์ความถูกต้องของแต่ละแบบจำลองทางคณิตศาสตร์

### 3.2 Web Application

ต้นแบบ Web Application สำหรับการแนะนำที่อยู่อาศัยประกอบด้วยสองส่วนด้วยกัน ได้แก่

1. หน้าจอสำหรับการกรอกข้อมูลของลูกค้าที่ใช้ในการแนะนำที่อยู่

จากรูปที่ 3 แสดงตัวอย่างหน้าจอสำหรับการกรอกข้อมูล ซึ่งลูกค้าจำเป็นต้องกรอกข้อมูลให้ครบถ้วนทุกส่วน ยกเว้นในส่วนของ “เหตุผลในการเลือกซื้อบ้าน” ที่สามารถเว้นว่างได้ นอกจากนั้นต้องกรอกข้อมูลให้ถูกประเภท เช่น อายุต้องเป็นตัวเลข ถ้ากรอกข้อมูลไม่ครบถ้วนหรือข้อมูลผิดประเภท ระบบจะแจ้งเตือนให้ผู้กรอกข้อมูลกรอกใหม่อีกครั้ง

ถ้าข้อมูลที่กรอกครบถ้วนสมบูรณ์ หลังจากกดปุ่ม “ยืนยันข้อมูล” ดังแสดงในรูปที่ 3 จะมีหน้าจอแสดงข้อมูลที่ลูกค้ากรอกเพื่อให้ลูกค้าตรวจสอบดังในรูปที่ 4 ถ้าข้อมูลถูกต้อง ผู้ใช้กดปุ่ม “ยืนยันข้อมูล” อีกครั้งเพื่อให้ระบบแสดงผลลัพธ์ของโครงการของการเคหะแห่งชาติที่คาดว่าลูกค้าจะสนใจต่อไป

ระบบแนะนำที่อยู่อาศัย (House Recommendation System)

กรุณากรอกข้อมูลต่อไปนี้

อายุ	<input type="text" value="36"/>	สมาชิกในครอบครัว	<input type="text" value="3"/>
รายได้ต่อเดือน	<input type="text" value="18000"/>	รายได้คู่สมรส (ถ้ามี)	<input type="text" value="0"/>
ค่าผ่อนบ้าน ณ ปัจจุบัน	<input type="text" value="4500"/>	ค่าผ่อนบ้านที่จ่ายได้จริง	<input type="text" value="3000"/>
อาชีพ	<input type="text" value="กรุณาเลือกอาชีพ"/>	เหตุผลในการเลือกซื้อบ้าน	
รหัสไปรษณีย์ที่อยู่ปัจจุบัน	<input type="text" value="10900"/>	<input type="checkbox"/> ใกล้แหล่งชุมชน	<input type="checkbox"/> ใกล้กรมคมนาคมสะดวก
รหัสไปรษณีย์ที่ทำงาน	<input type="text" value="10900"/>	<input type="checkbox"/> ใกล้ที่ทำงานหรือโรงเรียนบุตร	<input type="checkbox"/> ราคาสมเหตุสมผล
		<input type="checkbox"/> เป็นโครงการบ้านพร้อมที่ดิน	<input type="checkbox"/> อื่น ๆ

รูปที่ 3 ตัวอย่างหน้าจอที่ใช้ในการกรอกข้อมูลของลูกค้าในการแนะนำที่อยู่อาศัย

กรุณาตรวจสอบข้อมูล

อายุ	<input type="text" value="36"/>
สมาชิกในครอบครัว	
รายได้ต่อเดือน	
รายได้คู่สมรส	
ค่าผ่อนบ้าน ณ ปัจจุบัน	
ค่าผ่อนบ้านที่จ่ายไหว	
อาชีพ	
รหัสไปรษณีย์ที่อยู่ปัจจุบัน	
รหัสไปรษณีย์ที่ทำงาน	
เหตุผลในการเลือกซื้อบ้าน	

รูปที่ 4 ตัวอย่างหน้าจอในการตรวจสอบข้อมูลก่อนการยืนยันข้อมูลเพื่อให้ระบบแสดงผลลัพธ์ต่อไป

## 2. หน้าจอสำหรับการแนะนำโครงการที่อยู่อาศัย

ในหน้าจอนี้ จะแสดงผลสองส่วนดังรูปที่ 5

### a. รายการโครงการ 10 อันดับที่คาดว่าผู้กรอกข้อมูลน่าจะสนใจ

รายการนี้ประกอบไปด้วยรหัสโครงการ พื้นที่ใช้สอย ราคาโดยเฉลี่ย จำนวนหน่วยว่าง และ แบบบ้านที่มีรายละเอียด เช่น เป็นอาคารชุด 5 ชั้น เป็นต้น ผู้กรอกข้อมูลสามารถเลือกที่จะกดดูโครงการที่มีประเภทที่อยู่อาศัยทั้ง 2 แบบตามที่ระบบแนะนำ 10 รายการ (โดยระบบจะแสดงที่อยู่อาศัยแต่ละประเภทอย่างละ 5 รายการ) หรือจะเลือกดูเพียงแค่ประเภทที่อยู่อาศัยแบบใดแบบหนึ่งก็ได้

นอกเหนือจากนั้น ผู้กรอกข้อมูลสามารถเลือกที่จะดูเพียงแค่โครงการที่ระบบคาดการณ์ว่าจะผ่อนผ่อนหมดภายในอายุ 60 ปี ได้

### b. แผนที่แสดงตำแหน่งโครงการ

ผู้ที่กรอกข้อมูลสามารถกดเลือกดูแผนที่เพื่อดูระยะทางโดยประมาณ เปรียบเทียบระหว่างที่อยู่ปัจจุบันและที่ทำงานได้ ตัวอย่างที่แสดงในรูปที่ 5 เป็นแผนที่ที่เปรียบเทียบกับที่อยู่ปัจจุบันซึ่งแสดงโดยสัญลักษณ์รูปบ้าน ในขณะที่สัญลักษณ์หมุดสีแดงคือโครงการที่อยู่ในรายการ ผู้กรอกข้อมูลสามารถเอาเมาส์วางไว้ตรงหมุดสีแดงเพื่อดูชื่อโครงการและระยะทางที่ห่างจากที่อยู่ปัจจุบันโดยประมาณ นอกเหนือจากนี้ ผู้กรอกข้อมูลสามารถกดปุ่ม “ใกล้ที่ทำงาน” เพื่อเปลี่ยนไปเปรียบเทียบโครงการกับตำแหน่งที่ทำงานแทนได้อีกด้วย



ระบบแนะนำที่อยู่อาศัย (House Recommendation System)

**a.** แบบบ้านที่ระบบแนะนำ

ทั้งหมด    อาคารชุด    บ้านเดี่ยว

รายละเอียดโครงการที่อยู่อาศัย

ใช้เงื่อนไขการคำนวณราคาตามที่คาดว่าจะจ่ายได้จริง

รหัสโครงการ	พื้นที่ใช้สอย (ตร.ม./ตร.ว.)	ราคา (บาท)	จำนวนยูนิตว่าง (ยูนิต)	แบบบ้าน
			5	อาคารชุด 5 ชั้น
			73	อาคารชุด 5 ชั้น
			132	อาคารชุด 5 ชั้น
160059	35.0	678,000	73	อาคารชุด 5 ชั้น
160060	35.0	678,000	132	อาคารชุด 5 ชั้น
147203	210	580,000	1	บ้านเดี่ยว 2 ชั้น
150021	22.75	590,000	3	บ้านเดี่ยว 2 ชั้น
152025	22.75	590,000	2	บ้านเดี่ยว 2 ชั้น
146035	19.5	710,000	3	บ้านเดี่ยว 2 ชั้น
157006	210	598,000	1	บ้านเดี่ยว 2 ชั้น

**กดเพื่อแสดงเพียงแค่โครงการที่คาดว่าจะผ่อนไหว**

**b.** ตำแหน่งที่ต้องการเปรียบเทียบ

ใกล้ที่อยู่ปัจจุบัน    ใกล้ที่ทำงาน

Google    Keyboard shortcuts    Map data ©2021    Terms of Use

กลับหน้าหลัก

รูปที่ 5 ตัวอย่างหน้าจอแสดงผลโครงการของการเคหะที่คาดว่าจะถูกค่าจะสนใจ

# บทที่ 4

## บทสรุป

### 4.1 สรุปผลการดำเนินการ

เนื่องจากทางการเคหะแห่งชาตินั้นมีความมุ่งหมายในการให้บริการประชาชน โดยการแนะนำและจัดสร้างที่อยู่อาศัยสำหรับผู้มีรายได้น้อย รวมถึงได้เล็งเห็นถึงความสำคัญในการนำข้อมูลผู้จองในอดีตมาใช้ประโยชน์ในการพัฒนาและปรับปรุงการให้บริการให้ตรงตามความต้องการของประชาชนมากขึ้น ทางการเคหะแห่งชาติจึงร่วมมือกับ GBDi จัดทำระบบแนะนำที่อยู่อาศัยในลักษณะ Web Application เพื่ออำนวยความสะดวกให้แก่ผู้ที่สนใจโครงการที่อยู่อาศัยของการเคหะแห่งชาติ

ในการพัฒนาระบบแนะนำที่อยู่อาศัย ทางทีมงาน GBDi เริ่มจากการนำข้อมูลของผู้จองในอดีตมาศึกษา ทำความสะอาดและสืบหาลักษณะข้อมูลที่มีประโยชน์ พร้อมทั้งพัฒนาแบบจำลองทางคณิตศาสตร์ในการทำนายประเภทที่อยู่อาศัย หลังจากนั้นทางทีมงานได้พัฒนา Web Application ให้ผู้ที่สนใจโครงการบ้านของการเคหะแห่งชาติกรอกข้อมูลที่เกี่ยวข้อง เพื่อใช้แบบจำลองทำนายประเภทที่อยู่อาศัยและคัดกรองโครงการที่มีประเภทที่อยู่อาศัยเหล่านั้นที่คาดว่าผู้กรอกข้อมูลจะสนใจ โดยคำนึงถึงระยะทางโดยประมาณให้ใกล้ที่อยู่ปัจจุบันและที่ทำงานของผู้กรอกข้อมูล ก่อนที่จะแสดงผลลัพท์บนหน้าจอของ Web Application ซึ่งสามารถตรวจสอบตำแหน่งโดยประมาณของโครงการผ่านแผนที่และสามารถเลือกกรองโครงการเพิ่มเติมที่ทางระบบคาดว่าจะสามารถผ่อนชำระค่าบ้านหมดก่อนอายุ 60 ปีได้

### 4.2 การดำเนินการในอนาคตและการสนับสนุนเพิ่มเติม

1. ข้อมูลตำแหน่งของโครงการของการเคหะแห่งชาติควรมีความสมบูรณ์มากพอที่จะใช้ในการระบุ Latitude และ Longitude ได้
2. ข้อมูลหน่วยว่างของแต่ละโครงการควรมีข้อมูลที่ละเอียดมากขึ้น เช่น ระบุว่าหน่วยว่างนั้นเป็นหน่วยเช่าหรือหน่วยขาย หรือ ควรระบุหน่วยว่างให้ชัดเจนสำหรับในกรณีที่โครงการเดียวกัน มีบ้านประเภทเดียวกันแต่ขนาดต่างกัน
3. ควรมีการเชื่อมข้อมูลจำนวนหน่วยว่างเข้ากับฐานข้อมูลของการเคหะแห่งชาติเพื่อให้ได้ข้อมูลที่เป็นปัจจุบันมาก

## ภาคผนวก

### ก. ลักษณะเครื่อง Server ที่แนะนำสำหรับการติดตั้ง Web Application และ API ของระบบ แนะนำที่อยู่อาศัย

ในการติดตั้งเครื่อง Server สำหรับ Web Application และ API ของระบบแนะนำที่อยู่อาศัย ทาง  
ทีมงานแนะนำลักษณะเครื่องตามตารางที่ 7

คุณลักษณะ	ขนาด
จำนวน Core	2 core
Ram	4 GB
Capacity	50 GB

ตารางที่ 7 ลักษณะเครื่อง Server ที่แนะนำสำหรับการติดตั้ง Web Application และ API  
ของระบบแนะนำที่อยู่อาศัย

## บรรณานุกรม

Freepik. (October 2021). เข้าถึงได้จาก flaticon: <https://www.flaticon.com/>

SERVICE.meewebfree. (27 August 2020). *DATABASE CSV จังหวัด ตำบล อำเภอ รหัสไปรษณีย์ และ LAT LONG*. เข้าถึงได้จาก SERVICE DOT Meewebfree:

[http://www.service.meewebfree.com/internet/thailand\\_master](http://www.service.meewebfree.com/internet/thailand_master)

Thailandpostcode. (ม.ป.ป.). *ค้นหารหัสไปรษณีย์ทั่วประเทศ*. เข้าถึงได้จาก thailandpostcode:

<http://www.thailandpostcode.com/>