

การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล

นายสุรเดช ลีรุจิกุล

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาระบบสารสนเทศเพื่อการจัดการ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2568
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล

นายสุรเดช ลีรุจิกุล

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาระบบสารสนเทศเพื่อการจัดการ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2568
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



ใบรับรองการค้นคว้าอิสระ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล

โดย นายสุรเดช ลีรุจิกุล

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบสารสนเทศเพื่อการจัดการ

คณบดีคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

.....

(ผู้ช่วยศาสตราจารย์ ดร.สุนันฑา สดสี)

คณะกรรมการสอบการค้นคว้าอิสระ

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.เดช ธรรมศิริ)

..... อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.มณเฑียร รัตนศิริวงศ์วุฒิ)

ชื่อ : นายสุรเดช ลีรุจิกุล
ชื่อการค้นคว้าอิสระ : การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล
สาขาวิชา : ระบบสารสนเทศเพื่อการจัดการ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก : ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล
ปีการศึกษา : 2568

บทคัดย่อ

งานวิจัยนี้พัฒนาระบบสนับสนุนการตัดสินใจอัจฉริยะในการเลือกซื้อหรือเช่าที่อยู่อาศัยในเขตกรุงเทพมหานครและปริมณฑล โดยเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง 3 รูปแบบ ได้แก่ แคทบูสต์, มัลตินโนเมียล ไนอีฟ เบย์ และโครงข่ายประสาทเทียมแบบคอนโวลูชัน ร่วมกับการใช้เทคนิคการตัดคุณลักษณะแบบเรียกซ้ำเพื่อระบุปัจจัยพยากรณ์หลัก ซึ่งพบว่าช่วงอายุ รายได้เฉลี่ยต่อเดือน และจำนวนวันที่เดินทางไปที่ทำงานในแต่ละสัปดาห์ เป็นปัจจัยที่มีอิทธิพลสูงสุดในกระบวนการประเมินผล ได้แบ่งข้อมูลออกเป็นส่วนฝึกสอนร้อยละ 80 และส่วนทดสอบร้อยละ 20 ร่วมกับการทดสอบความถูกต้องแบบไขว้ 10 ส่วน เพื่อความแม่นยำทางสถิติ ซึ่งพบว่าแบบจำลองที่เลือกใช้คือแคทบูสต์ให้ค่าความถูกต้องสูงสุดที่ร้อยละ 82.50 นอกจากนี้ ระบบยังได้บูรณาการเทคนิคการอธิบายเชิงเหตุผลแบบโต้แย้งที่หลากหลายเพื่อสร้างคำแนะนำเชิงรุกที่นำไปปฏิบัติได้จริง และผลการประเมินความพึงพอใจต่อระบบอยู่ในระดับดี ผลการศึกษาชี้ให้เห็นว่าการใช้เทคนิคการทำเหมืองข้อมูลช่วยให้การวางแผนทางการเงินมีความแม่นยำและส่งเสริมการตัดสินใจอย่างเป็นเหตุเป็นผลในสภาวะเศรษฐกิจปัจจุบัน

(การค้นคว้าอิสระมีจำนวนทั้งสิ้น 93 หน้า)

คำสำคัญ : ระบบสนับสนุนการตัดสินใจ การเรียนรู้ของเครื่อง การตัดสินใจซื้อหรือเช่าที่อยู่อาศัย เหมืองข้อมูล อัลกอริทึม

_____ อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก

Name : Mr. Suradech Leerujikul
Independent Study Title : Development of a Decision Support System for Housing Purchase or Rental Selection Using Data Mining Techniques
Major Field : Management Information Systems
King Mongkut's University of Technology North Bangkok
Independent Study Advisor : Assistant Professor Dr. Watchareewan Jitsakul
Academic Year : 2025

ABSTRACT

This research develops an intelligent decision support system for residential selection in the Bangkok Metropolitan Region. Three algorithms—CatBoost, Multinomial Naive Bayes, and CNN—were evaluated alongside Recursive Feature Elimination (RFE). Findings reveal that age range, average monthly income, and weekly office commute days are the most significant predictors. The model was evaluated using an 80/20 train-test split and 10-fold cross-validation, which found that the selected model, CatBoost, achieved the highest accuracy of 82.50%. Furthermore, the system integrates Diverse Counterfactual Explanations (DiCE) to provide actionable insights for homeownership, and the system's satisfaction evaluation was at a good level. The results demonstrate that this framework effectively enhances financial planning and rational decision-making for residential choices in the current economic climate.

(Total 93 Pages)

Keywords: Decision Support System (DSS), Machine Learning, Housing Decision, Data Mining, Algorithm Comparison

_____ Advisor

กิตติกรรมประกาศ

โครงงานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีด้วยความกรุณาและความช่วยเหลืออย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล อาจารย์ที่ปรึกษาโครงงาน ที่ได้กรุณาสละเวลาให้คำปรึกษา แนะนำแนวทางการวิเคราะห์ข้อมูล การพัฒนาระบบสนับสนุนการตัดสินใจด้วยเทคนิคการเรียนรู้ของเครื่อง ตลอดจนตรวจทานและแก้ไขข้อบกพร่องต่าง ๆ ของงานวิจัยให้มีความสมบูรณ์ยิ่งขึ้น ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.เดช ธรรมศิริ ประธานกรรมการสอบ และ ผู้ช่วยศาสตราจารย์ ดร.มณเฑียร รัตนศิริวงศ์วุฒิ กรรมการสอบ ที่ได้กรุณาสละเวลาให้คำแนะนำและข้อเสนอแนะอันเป็นประโยชน์ยิ่งในการปรับปรุงงานวิจัยให้มีความถูกต้องและครบถ้วนตามหลักวิชาการ

ขอขอบพระคุณคณะกรรมการสอบโครงงานทุกท่านที่ได้ให้ข้อเสนอแนะอันเป็นประโยชน์ในการปรับปรุงงานวิจัยให้มีความถูกต้องตามหลักวิชาการ ขอบคุณเพื่อน ๆ ทุกคนที่คอยช่วยเหลือและเป็นกำลังใจในการแก้ไขปัญหาทางเทคนิคต่าง ๆ ตลอดการดำเนินงาน

นอกจากนี้ ผู้วิจัยขอขอบคุณกลุ่มตัวอย่างทุกท่านที่ได้สละเวลาในการตอบแบบสอบถามข้อมูล และผู้ร่วมทดสอบระบบสนับสนุนการตัดสินใจจำนวน 30 ท่าน ที่ได้ให้ข้อมูลอันเป็นประโยชน์ในการประเมินประสิทธิภาพและความพึงพอใจของระบบ ซึ่งเป็นส่วนสำคัญที่ทำให้งานวิจัยฉบับนี้สามารถสรุปผลได้อย่างเป็นรูปธรรม

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณบิดา มารดา และครอบครัว ที่ให้การสนับสนุนทั้งด้านกำลังใจและทุนทรัพย์ในการศึกษาเล่าเรียนเสมอมา ความดีและประโยชน์อันพึงมีจากงานวิจัยฉบับนี้ ผู้วิจัยขอมอบให้แก่ผู้มีพระคุณทุกท่านที่ได้กล่าวมาข้างต้น

สุรเดช ลีรุจิกุล

สารบัญ

หน้า

บทคัดย่อภาษาไทย ข

บทคัดย่อภาษาอังกฤษ ค

กิตติกรรมประกาศ ง

สารบัญ จ

สารบัญตาราง ช

สารบัญรูปภาพ ซ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 1

1.3 ขอบเขตของการวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 2

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 4

2.1 ระบบสนับสนุนการตัดสินใจ (Decision Support System: DSS) 4

2.2 อัลกอริทึมในการเรียนรู้ของเครื่อง (Machine Learning Algorithms) 5

2.3 งานวิจัยที่เกี่ยวข้อง (Related Research) 5

บทที่ 3 วิธีดำเนินงานวิจัย 24

3.1 ขั้นตอนการสร้างกรอบแนวความคิดของงานวิจัย 24

3.2 การศึกษา รวบรวมและจัดเตรียมข้อมูล 26

3.3 ขั้นตอนการเปรียบเทียบและคัดเลือกอัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุด 27

3.4 ขั้นตอนการออกแบบ และพัฒนาระบบ 29

3.5 ขั้นตอนการประเมินการใช้งานระบบ 31

บทที่ 4 ผลการดำเนินงานวิจัย 34

4.1 ผลการศึกษา และการรวบรวมข้อมูล 34

4.2 ผลการเตรียมชุดข้อมูล 37

4.3 ผลการสร้างและประเมินประสิทธิภาพแบบจำลองการพยากรณ์ 37

4.4 ผลการสร้างและประเมินประสิทธิภาพแบบจำลองการพยากรณ์ 63

สารบัญ (ต่อ)

หน้า

4.5 ผลการประเมินความพึงพอใจต่อการใช้งานระบบสนับสนุนการตัดสินใจ 65
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 67
5.1 สรุปผลการดำเนินงาน 67
5.2 อภิปรายผล 68
5.3 ข้อเสนอแนะ 69
บรรณานุกรม 70
ภาคผนวก ก 75
ภาคผนวก ข 78
ภาคผนวก ค 83
ภาคผนวก ง 89
ประวัติผู้เขียน 93

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

สารบัญตาราง

ตารางที่ หน้า

2-1 สรุปงานวิจัยที่เกี่ยวข้อง 14

3-1 เกณฑ์การประเมินความพึงพอใจ 31

3-2 เกณฑ์ในการแปลผลระดับประเมินความพึงพอใจ 32

4-1 ผลการประเมินค่าดัชนีความสอดคล้อง (IOC) ของแบบสอบถามโดยผู้เชี่ยวชาญ 35

4-2 ผลการรันแบบจำลอง Catboost เพื่อหา Best Parameter 39

4-3 ผลการรันแบบจำลอง CNN เพื่อหา Best Parameter 43

4-4 ผลการรันแบบจำลอง Multinomial Naive Bayes เพื่อหา Best Parameter 46

4-5 ผลการจัดลำดับความสำคัญโดย REF 47

4-6 ผลการรันแบบจำลอง Catboost ด้วย 10 คุณลักษณะ 49

4-7 ผลการรันแบบจำลอง CNN ด้วย 10 คุณลักษณะ 53

4-8 ผลการรันแบบจำลอง Multinomial Naive Bayes ด้วย 10 คุณลักษณะ 56

4-9 ผลการรันแบบจำลอง Catboost ด้วย 7 คุณลักษณะ 57

4-10 ผลการรันแบบจำลอง CNN ด้วย 7 คุณลักษณะ 61

4-11 ผลการรันแบบจำลอง Multinomial Naive Bayes ด้วย 7 คุณลักษณะ 63

4-12 ผลการเก็บข้อมูลความพึงพอใจระบบ 66

สารบัญรูปภาพ

ภาพที่ หน้า

3-1 กรอบแนวคิด 24
3-2 ขั้นตอนการดำเนินงานวิจัย 26
3-3 ออกแบบหน้าจอรับข้อมูลของระบบ 30
3-4 ออกแบบหน้าจอผลลัพธ์ของระบบ 30
4-1 หน้าจอรับข้อมูลของระบบ 64
4-2 หน้าจอผลลัพธ์ของระบบ 65
ก-1 คิวอาร์โค๊ดสำหรับจัดเก็บข้อมูลที่พักอาศัย 76
ก-2 คิวอาร์โค๊ดสำหรับจัดเก็บข้อมูลความพึงพอใจ 77

พระจอมเกล้าพระนครเหนือ มหาวิทยาลัยเทคโนโลยี

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันประเทศไทยกำลังก้าวเข้าสู่สังคมผู้สูงอายุระดับสุดยอด โดยมีอัตราการเกิดของประชากรรุ่นใหม่ในกลุ่มเจเนอเรชันซีและเจเนอเรชันอัลฟาลดลงอย่างต่อเนื่อง พร้อมกันนี้ยังเผชิญกับภาวะเงินเฟ้อและสภาพเศรษฐกิจที่เปราะบาง ส่งผลให้กลุ่มวัยทำงานต้องแบกรับภาระค่าใช้จ่ายและค่าเลี้ยงดูผู้สูงอายุสูงขึ้น ขณะที่ราคาที่อยู่อาศัยมีแนวโน้มเพิ่มขึ้นร้อยละ 5–10 ต่อปี ประชาชนจำนวนมากจึงประสบปัญหาด้านความสามารถในการซื้อที่พักอาศัยและขาดความมั่นคงทางการเงิน

สำหรับการศึกษาก่อนหน้านี้ที่เกี่ยวข้องกับการตัดสินใจเลือกที่อยู่อาศัย พบว่ามีการพิจารณาความสามารถในการจ่ายและทางเลือกที่อยู่อาศัยในเขตกรุงเทพมหานครและปริมณฑลมาบ้างแล้ว เช่น งานวิจัยของ กิตติพร พูลสวัสดิ์ และคณะ [1] ที่ระบุว่าอายุและรายได้ต่อเดือนเป็นปัจจัยสำคัญที่มีอิทธิพลต่อการตัดสินใจซื้อ แต่งานส่วนใหญ่ยังคงพึ่งพาวิธีการทางสถิติแบบดั้งเดิมหรือเน้นตัวแปรทางเศรษฐกิจสังคมที่จำกัด ซึ่งอาจไม่เพียงพอต่อการพยากรณ์พฤติกรรมในสภาวะเศรษฐกิจที่มีความผันผวนสูง งานวิจัยนี้จึงมุ่งพัฒนาระบบสนับสนุนการตัดสินใจ (DSS) โดยประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลร่วมกับอัลกอริทึม CatBoost, Multinomial Naive Bayes และ CNN พร้อมเทคนิค RFE และ DiCE เพื่อสร้างทางเลือกจำลองที่ช่วยให้ผู้ใช้เห็นแนวทางที่นำไปปฏิบัติได้จริง ผ่านการพัฒนาด้วยภาษา Python และ React เพื่อประสิทธิภาพสูงสุด

จากปัญหาและงานวิจัยต่างๆ ข้างต้น แม้จะมีการพัฒนาระบบสนับสนุนการตัดสินใจในรูปแบบต่างๆ แต่ความซับซ้อนของปัจจัยในปัจจุบันทำให้ผู้บริโภคต้องการข้อมูลที่มากกว่าเพียงแค่ผลลัพธ์การทำนาย เพราะเหตุนี้ ผู้วิจัยจึงได้พัฒนางานวิจัยนี้ขึ้นเพื่อ สร้างระบบที่ไม่เป็นเพียงแค่ระบบสนับสนุนการตัดสินใจ (DSS) ทั่วไป แต่ยังมีความสามารถในการช่วยให้คำแนะนำและเสนอแนวทางปฏิบัติที่ชัดเจนแก่ผู้ใช้ โดยการนำเทคนิคการเรียนรู้ของเครื่องมาช่วยวิเคราะห์และแสดงผลในลักษณะที่เข้าใจง่าย เพื่อให้มั่นใจว่าระบบจะสามารถสนับสนุนการตัดสินใจเลือกที่พักอาศัยให้กับประชาชนได้อย่างมีประสิทธิภาพ และเกิดประโยชน์สูงสุดในการสร้างความมั่นคงด้านที่อยู่อาศัยในสถานการณ์จริง

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อเปรียบเทียบอัลกอริทึมสำหรับการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัย

1.2.2 เพื่อประเมินประสิทธิภาพของระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัย

1.2.3 เพื่อพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัย

1.2.4 เพื่อประเมินความพึงพอใจของระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัย

1.3 ขอบเขตของการวิจัย

1.3.1 ขอบเขตด้านข้อมูล

การเก็บรวบรวมข้อมูลในงานวิจัยนี้มุ่งเน้นที่การรวบรวมปัจจัยที่มีผลต่อการตัดสินใจเลือกซื้อหรือเช่าที่พักอาศัย โดยทำการสำรวจกลุ่มตัวอย่างจำนวน 1,000 คน ซึ่งเป็นประชากรที่อาศัยอยู่ในพื้นที่กรุงเทพมหานครและปริมณฑล ข้อมูลที่เก็บรวบรวมครอบคลุมตัวแปรด้านเศรษฐกิจ สังคม พฤติกรรมการเงิน ความต้องการด้านที่อยู่อาศัย และความมั่นคงทางการเงิน เพื่อใช้เป็นฐานข้อมูลสำหรับการวิเคราะห์และพัฒนาโมเดลระบบสนับสนุนการตัดสินใจ โดยมีการออกแบบแบบสอบถามและเครื่องมือเก็บข้อมูลให้มีความเหมาะสมและสอดคล้องกับวัตถุประสงค์การวิจัย รวมถึงมีการตรวจสอบความเที่ยงและความน่าเชื่อถือของข้อมูลก่อนนำไปใช้ในการประมวลผลและวิเคราะห์ต่อไป

1.3.2 ขอบเขตด้านอัลกอริทึม

1.1..1. อัลกอริทึมในการตัดสินใจ ได้แก่ อัลกอริทึม Catboost อัลกอริทึม Multinomial Naive Bayes และอัลกอริทึม CNN

1.1..2. อัลกอริทึมในการวิเคราะห์พฤติกรรม ได้แก่ RFE

1.1..3. อัลกอริทึมในการแนะนำพฤติกรรม ได้แก่ DiCE

1.3.3 ขอบเขตด้านการเปรียบเทียบประสิทธิภาพของอัลกอริทึม โดยพิจารณาค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าการถ่วงดุล (F - Measure) ค่าการวิเคราะห์เส้นโค้งคุณลักษณะสมบัติการทำงาน (ROC)

1.3.4 ขอบเขตด้านการพัฒนาระบบ ได้แก่

1.3.4.1 พัฒนาส่วนติดต่อผู้ใช้งาน (Frontend) โดยใช้เฟรมเวิร์ก React

1.3.4.2 พัฒนาส่วนประมวลผลของระบบ (Backend) ด้วยภาษา Python

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้ระบบต้นแบบของระบบสนับสนุนการตัดสินใจที่มีประสิทธิภาพ เพื่อนำไปประยุกต์ใช้ในการเลือกที่พักอาศัยให้เหมาะสมคนไทย

1.4.2 ได้ทราบถึงประสิทธิภาพของอัลกอริทึมต่างๆ ที่นำมาเปรียบเทียบกัน ซึ่งจะเป็นประโยชน์ต่อเหล่านักพัฒนาในการเลือกใช้โมเดลที่แม่นยำที่สุด ในอนาคต

1.4.3 ช่วยลดความเสี่ยงทางการเงิน ของผู้บริโภคจากการตัดสินใจผิดพลาดในอสังหาริมทรัพย์ ซึ่งเป็นสินทรัพย์ที่มีมูลค่าสูงและมีภาระผูกพันระยะยาว

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ทำการศึกษาค้นคว้า บทความวิจัยที่มีความเกี่ยวข้องกับระบบสนับสนุนการตัดสินใจด้วย เทคนิคต่าง ๆ รวมถึงแนวทางขั้นตอนวิธีการในการเปรียบเทียบ เพื่อที่จะนำมาใช้ในการวิเคราะห์เกี่ยวกับระบบสนับสนุนการตัดสินใจ ดังงานวิจัยที่เกี่ยวข้องต่อไปนี้

2.1 ระบบสนับสนุนการตัดสินใจ (Decision Support System: DSS)

2.2 อัลกอริทึมในการเรียนรู้ของเครื่อง (Machine Learning Algorithms)

2.3 งานวิจัยที่เกี่ยวข้อง (Related Research)

2.1 ระบบสนับสนุนการตัดสินใจ (Decision Support System: DSS)

ระบบสนับสนุนการตัดสินใจ หมายถึง ระบบสารสนเทศที่ผสมผสานระหว่างข้อมูลและโมเดลการวิเคราะห์ เพื่อสนับสนุนการตัดสินใจในปัญหาที่มีความซับซ้อนและมีปัจจัยที่เปลี่ยนแปลงอยู่ตลอดเวลา [2] โดยเฉพาะในสภาวะเศรษฐกิจที่เปราะบางและราคาที่พักอาศัยที่มีความผันผวนสูงในปัจจุบัน ระบบ DSS จึงเป็นเครื่องมือสำคัญที่ช่วยลดความเสี่ยงทางการเงินและช่วยในการพยากรณ์พฤติกรรมที่มีความไม่แน่นอนสูงได้แม่นยำกว่าวิธีการทางสถิติแบบดั้งเดิม [3]

2.1.1 องค์ประกอบของระบบสนับสนุนการตัดสินใจด้วยการเรียนรู้ของเครื่อง ในงานวิจัยนี้ระบบสนับสนุนการตัดสินใจถูกพัฒนาโดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) และการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ โดยมีองค์ประกอบสำคัญดังนี้

2.1.1.1 ส่วนการจัดการข้อมูล ครอบคลุมการรวบรวมปัจจัยด้านเศรษฐกิจ สังคม และพฤติกรรมการเงินจากกลุ่มตัวอย่าง 1,000 คน เพื่อใช้เป็นฐานข้อมูลหลัก

2.1.1.2 ส่วนการประมวลผลโมเดล การเปรียบเทียบอัลกอริทึม CatBoost, Multinomial Naive Bayes และ CNN เพื่อหาโมเดลที่ให้ค่า Accuracy และ F-Measure สูงที่สุด [4]

2.1.1.3 ส่วนการวิเคราะห์ปัจจัยอิทธิพล การใช้เทคนิค Recursive Feature Elimination (RFE) เพื่อคัดเลือกตัวแปรที่มีนัยสำคัญต่อการตัดสินใจเลือกที่พักอาศัย [5]

2.1.2 การอธิบายเหตุผลและการจำลองทางเลือกด้วยเทคนิค DiCE เพื่อให้ระบบมีความโปร่งใสและนำไปปฏิบัติได้จริง งานวิจัยนี้ได้บูรณาการเทคนิค Diverse Counterfactual Explanations

(DiCE) ซึ่งช่วยให้ผู้ใช้เห็นแนวทางจำลอง (Counterfactuals) เช่น หากต้องการขยับจากสถานะเช่าไปเป็นการซื้อ ผู้ใช้ควรปรับเปลี่ยนปัจจัยใดบ้างเพื่อให้ได้ผลลัพธ์ที่ต้องการ [6] โดยส่วนติดต่อผู้ใช้งานถูกพัฒนาด้วย React เพื่อความทันสมัยและตอบสนองต่อผู้ใช้ได้อย่างมีประสิทธิภาพ

2.2 อัลกอริทึมในการเรียนรู้ของเครื่อง (Machine Learning Algorithms)

2.2.1 K-Nearest Neighbors: K-NN เป็นอัลกอริทึมสำหรับจัดหมวดหมู่ข้อมูล (Classification) ที่ใช้หลักการเปรียบเทียบความคล้ายคลึงระหว่างข้อมูลที่สนใจกับข้อมูลข้างเคียง โดยพิจารณาจากระยะห่างที่ใกล้ที่สุด (K) หากข้อมูลที่พิจารณาอยู่ใกล้กับกลุ่มข้อมูลใดมากที่สุด ระบบจะจัดหมวดหมู่ข้อมูลนั้นตามกลุ่มดังกล่าว อัลกอริทึมนี้มีลักษณะเด่นคือการไม่สร้างแบบจำลองจากข้อมูลชุดเรียนรู้ล่วงหน้า แต่จะใช้ข้อมูลชุดเรียนรู้เป็นตัวแบบจำลองในการตัดสินใจโดยตรง ซึ่งเหมาะสำหรับการจัดการข้อมูลเชิงตัวเลขที่สามารถวัดระยะห่างระหว่างคุณลักษณะได้ [7]

2.2.2 Random Forest เป็นเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ที่สร้างแบบจำลองโดยการรวมต้นไม้ตัดสินใจ (Decision Tree) จำนวนมากเข้าด้วยกัน (ตั้งแต่หลักสิบจนถึงหลักพันโมเดล) โดยแต่ละต้นไม้จะถูกฝึกสอนด้วยข้อมูลส่วนย่อย (Subset) ที่แตกต่างกัน เมื่อเข้าสู่ขั้นตอนการพยากรณ์ แต่ละต้นไม้จะทำการพยากรณ์ผลลัพธ์แยกจากกัน และใช้หลักการลงคะแนนเสียงส่วนมาก (Voting) สำหรับการจำแนกประเภท หรือการหาค่าเฉลี่ย (Mean) สำหรับการพยากรณ์ค่าต่อเนื่อง เพื่อให้ได้ผลลัพธ์ที่มีความแม่นยำและลดการเกิดปัญหาการเรียนรู้เกิน (Overfitting) [8]

2.2.3 Naive Bayes เป็นแบบจำลองการจำแนกประเภทข้อมูลโดยอาศัยหลักการทางทฤษฎีความน่าจะเป็นของเบย์ (Bayes' Theorem) ซึ่งทำหน้าที่คำนวณหาความน่าจะเป็นของเหตุการณ์ที่จะเกิดขึ้นตามเงื่อนไขของข้อมูลที่ได้รับ เช่น การวินิจฉัยความน่าจะเป็นของโรคจากอาการที่ปรากฏของผู้ป่วย เพื่อสรุปผลลัพธ์ที่มีความน่าจะเป็นสูงสุด โดยอัลกอริทึมนี้มีประสิทธิภาพสูงในกรณีที่ข้อมูลมีคุณลักษณะจำนวนมากและมีความเป็นอิสระต่อกัน [9]

2.3 งานวิจัยที่เกี่ยวข้อง (Related Research)

ณิชา นภาพร และจงกะสิกิจ [10] ได้ทำการรวบรวมข้อมูลจากสองแหล่งหลัก ได้แก่ การสัมภาษณ์บุคคลที่เกี่ยวข้องกับการคัดกรองผู้ป่วย และการศึกษาเอกสารรวมถึงฐานข้อมูลที่เกี่ยวข้องกับการคัดกรองความเสี่ยงของโรคเรื้อรัง เช่น โรคเบาหวานและความดันโลหิตสูง โดยใช้เทคนิคต้นไม้ตัดสินใจและ K-Nearest Neighbors (KNN) ในการวิเคราะห์ ข้อมูลทดสอบที่ใช้มีจำนวน 2,000 รายการ โดยมีการจัดกลุ่มผู้ป่วยตามหลักเกณฑ์ของปิงปองจราจรชีวิต 7 สี และใช้เทคนิค 10-Fold Cross Validation เพื่อประเมินประสิทธิภาพของโมเดล พบว่าเทคนิคต้นไม้ตัดสินใจมีความถูกต้องสูงกว่าที่ร้อยละ 96.45 การจัดกลุ่มตามเกณฑ์ประกอบด้วย กลุ่มปกติ (สีขาว), กลุ่มเสี่ยง (สีเขียวอ่อน),

กลุ่มป่วยระดับที่ 0 (สีเขียวเข้ม), กลุ่มป่วยระดับที่ 1 (สีเหลือง) กลุ่มป่วยระดับที่ 2 (สีส้ม) กลุ่มป่วยระดับที่ 3 (สีแดง) และกลุ่มป่วยรุนแรง (สีดำ).

รัชพล และจรัญ [11] ได้ดำเนินการศึกษาเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา โดยทำการเปรียบเทียบ 3 อัลกอริธึม ได้แก่ อัลกอริธึมต้นไม้ตัดสินใจ (DT : J48), อัลกอริธึม Naive Bayes (NB) และอัลกอริธึมกฎการอุปนัย (Rule Induction) ใช้ข้อมูลนักเรียนจากสถานศึกษารัฐบาลแห่งหนึ่งในประเทศไทยซึ่งครอบคลุม 9 สาขาวิชาในช่วงปี พ.ศ. 2550 - 2559 รวมจำนวน 5,100 ระเบียน การคัดเลือกคุณลักษณะใช้วิธี Forward Selection และการคัดเลือกคุณลักษณะทั้งหมด การทดสอบประสิทธิภาพการทำนายใช้เทคนิค 10-Fold Cross Validation และวัดค่าความถูกต้อง ผลการวิจัยพบว่าอัลกอริธึม DT : J48 ร่วมกับ Forward Selection มีค่าความถูกต้องสูงสุดที่ร้อยละ 83.08 ขณะที่การคัดเลือกคุณลักษณะทั้งหมดให้ค่าความถูกต้องร้อยละ 81.71 การทดสอบด้วยวิธี T-Test แสดงให้เห็นว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ระหว่างทั้งสองวิธี.

วิโรจน์ [12] ได้ศึกษาการวิเคราะห์คุณลักษณะทางเสียงสำคัญที่ส่งผลต่อคะแนนการทำงานสมองด้านแอตเทนชันในแบบประเมินไทยโมคาโดยใช้ข้อมูลเสียงจากแบบทดสอบทวนตัวเลขตามลำดับจากหน้าไปหลัง (DFS) และแบบทดสอบทวนตัวเลขตามลำดับจากหลังไปหน้า (DBS) พัฒนาโมเดลด้วยวิธี Recursive Feature Elimination (RFE) ร่วมกับตัวประเมินหลายชนิด เช่น Decision Tree, Gradient Boosting และ Random Forest เพื่อคัดเลือกคุณลักษณะทางเสียงที่มีผลมากที่สุด ผลการทดลองพบว่าชุดคุณลักษณะสำคัญ 3 ชุด ได้แก่ ชุด DFS จำนวน 10 ตัว ประกอบด้วยระดับเสียง ค่าเฉลี่ยฮาร์มอนิก ช่วงเวลาขณะเกิดเสียง และระดับการศึกษา ชุด DBS จำนวน 10 ตัว ได้แก่ ระดับความดัง ฮาร์มอนิกช่วงต่าง ๆ ช่วงเวลาขณะเกิดเสียง และระดับการศึกษา และชุดรวม DFS และ DBS จำนวน 3 ตัว ซึ่งเกี่ยวข้องกับผลรวมช่วงเวลาขณะเกิดเสียงที่ตรวจสอบจากความเข้มและคอนทัวร์ของระดับเสียง เมื่อเทียบกับการใช้คุณลักษณะทางเสียงทั้งหมด ชุดคุณลักษณะที่คัดเลือกให้ผลความแม่นยำสูงกว่า โดยชุดรวม DFS และ DBS ให้ความแม่นยำสูงสุดถึงร้อยละ 72 ซึ่งสะท้อนความสำคัญของช่วงเวลาขณะเกิดเสียงต่อคะแนนแอตเทนชัน นอกจากนี้พบว่าคุณลักษณะเสียงสำคัญเหล่านี้สามารถแยกความแตกต่างระหว่างกลุ่มปกติและกลุ่มผู้มีภาวะความบกพร่องเล็กน้อยของสมองได้ชัดเจน งานวิจัยชี้แนะว่าการผสมผสานคุณลักษณะจากทั้งสองแบบทดสอบช่วยเพิ่มประสิทธิภาพในการจำแนกข้อมูล และเน้นว่าคุณลักษณะที่เกี่ยวกับช่วงเวลาของเสียงมีบทบาทสำคัญในการประเมินแอตเทนชันในแบบประเมินไทยโมคา

ภูมิพัฒน์ และรัจนา [13] ได้ทำการศึกษาการสร้างแบบจำลองการทำนายสำหรับอากาศยานอัตโนมัติที่ไม่พึ่งพามนุษย์ โดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวของกำลังทางอากาศจาก

ประเทศรอบบ้าน การศึกษาเปรียบเทียบประสิทธิภาพการจำแนกด้วย 4 อัลกอริธึม ได้แก่ Naive Bayes (NB), การเรียนรู้เชิงลึกแบบ Multi-Layer Feed-Forward Artificial Neural Network (Deep Learning) อัลกอริธึมต้นไม้ตัดสินใจ (DT) และ Random Forest (RF) โดยใช้ข้อมูลจากกรมข่าวทหารอากาศจำนวนประมาณ 37,000 ระเบียน และจัดการกับปัญหาความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด การทดสอบประสิทธิภาพการทำนายใช้เทคนิค 10-Fold Cross Validation และวัดประสิทธิภาพด้วยค่าความถูกต้อง ความแม่นยำ และค่าระลึก (Recall) ผลการวิจัยพบว่าอัลกอริธึม RF มีประสิทธิภาพสูงที่สุดด้วยค่าความถูกต้องร้อยละ 90.2 ความแม่นยำร้อยละ 90.47 และค่าระลึกร้อยละ 90.24 รองลงมา คือ อัลกอริธึม DT โดยมีค่าความถูกต้องร้อยละ 87.94 ความแม่นยำร้อยละ 88.28 และค่าระลึกร้อยละ 88 ขณะที่อัลกอริธึม NB และ Deep Learning มีประสิทธิภาพต่ำกว่าตามลำดับ.

ศิริลักษณ์ และคณะ [14] ได้ศึกษาการพยากรณ์รูปแบบการกระทำความผิดซ้ำของเด็กและเยาวชน โดยใช้ข้อมูลจากการสัมภาษณ์เด็กและเยาวชนที่กระทำผิดจากกรมพินิจและคุ้มครองเด็กและเยาวชน รวมถึงข้อมูลการกระทำผิดจากการบันทึกคดีของสถานีตำรวจ การศึกษานี้มุ่งเน้นการวัดประสิทธิภาพการจำแนกประเภทด้วยอัลกอริธึม XGBoost โดยแบ่งข้อมูลออกเป็นชุดเรียนรู้และชุดทดสอบ และประเมินประสิทธิภาพด้วยวิธีการ 5-Fold Cross Validation ผลการวิจัยพบว่าอัลกอริธึม XGBoost ให้ค่าความถูกต้องและค่าความแม่นยำสูงถึงร้อยละ 97.80 และค่าระลึก (Recall) สูงถึงร้อยละ 100 ปัจจัยที่สำคัญในการกระทำความผิดของเด็กและเยาวชนประกอบด้วย 6 คุณลักษณะ ได้แก่ การดำรงชีพด้วยการประกอบมิจฉาชีพ, พฤติกรรมงัดแงะเข้าบ้าน อาคาร หรือรถของผู้อื่น, การรับฝากหรือจำหน่ายยาเสพติด, ประวัติการถูกดำเนินคดี, ประวัติการฝึกอบรม/จำคุก/คุมประพฤติ/การฟื้นฟูสมรรถภาพผู้ติดยาเสพติด, และอายุขณะกระทำผิด.

ธนภัทร และคณะ [15] ได้ศึกษาการพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวานโดยใช้เทคนิคการเรียนรู้ของเครื่องเพื่อปรับปรุงการจำแนกประเภทข้อมูล โดยเปรียบเทียบประสิทธิภาพของอัลกอริธึม Decision Tree (DT) Random Forest (RF) และ Gradient Boosting Machine (GBM) การศึกษานี้ใช้ข้อมูลจากโรงพยาบาลในจังหวัดเชียงใหม่ระหว่างปี พ.ศ. 2562-2564 จำนวน 10,000 รายการ ซึ่งประกอบด้วยข้อมูลปัจจัยเสี่ยงเช่น อายุ ดัชนีมวลกาย (BMI) ระดับน้ำตาลในเลือด และประวัติครอบครัว การทดสอบประสิทธิภาพการทำนายใช้วิธีการ 10-Fold Cross Validation โดยวัดค่าความถูกต้อง ค่าความแม่นยำ และค่าระลึก ผลการวิจัยพบว่าอัลกอริธึม RF มีค่าความถูกต้องสูงสุดที่ร้อยละ 92.5 ค่าความแม่นยำร้อยละ 91.7 และค่าระลึกร้อยละ 92.8 รองลงมาคือ GBM โดยมีค่าความถูกต้องร้อยละ 89.4 ค่าความแม่นยำร้อยละ 88.6 และค่าระลึกร้อยละ 89.0 ขณะที่อัลกอริธึม DT ให้ผลลัพธ์ต่ำสุดในทั้งสามมาตรวัด

Zhuoqing, W., et al. [16] ได้ศึกษาการพยากรณ์ผลลัพธ์โรค MAFLD โดยพัฒนาระบบ

พยากรณ์อัจฉริยะที่ผสานข้อมูลเชิงสัมพันธ์ตามช่วงเวลา (TARs) กับความรู้ทางการแพทย์จริงจากข้อมูลผู้ป่วยในโรงพยาบาลตติยภูมิช่วงปี 2017–2020 ระบบนี้เน้นการเพิ่มความแม่นยำและความสามารถในการอธิบายผลของแบบจำลองด้วยการใช้เทคนิค Gradient Boosting Machine (GBM) พร้อมวิเคราะห์ประสิทธิภาพด้วยการแบ่งข้อมูลเป็นชุดฝึกสอนและทดสอบในอัตราส่วน 7:3 ผลการวิจัยพบว่า การนำ TARs จำนวน 20–30 กฎเข้าร่วมในแบบจำลองช่วยเพิ่มประสิทธิภาพของการพยากรณ์ได้อย่างชัดเจน โดย GBM มีความทนทานสูงต่อความไม่สมดุลของข้อมูลผู้ป่วย MAFLD ที่เป็นบวกและลบ นอกจากนี้ ยังวิเคราะห์การอธิบายผลทั้งในระดับภาพรวมและรายบุคคลด้วย SHAP พบว่าค่าตัวแปรสำคัญ ได้แก่ ไตรกลีเซอไรด์ (TG), ดัชนีมวลกาย (BMI) และตัวชี้วัดการทำงานของไต รวมทั้งกฎ TARs ที่เกี่ยวข้องกับการทำงานของไตและตัวชี้วัดทางร่างกายอื่น ๆ ได้รับการยืนยันว่ามีผลต่อความเสี่ยงของ MAFLD ระบบต้นแบบถูกเปิดใช้งานออนไลน์และได้รับการประเมินโดยแพทย์กว่า 30 คน พบว่าผู้ใช้งานมีความพึงพอใจในประสิทธิภาพและความน่าเชื่อถือของระบบ โดยเฉพาะอย่างยิ่งระบบช่วยเพิ่มความเข้าใจและความไว้วางใจในผลการพยากรณ์ของโมเดล งานวิจัยชี้แนวทางต่อไปในการพัฒนาระบบ ได้แก่ การทดสอบกับข้อมูลในระยะยาวจากหลายศูนย์ การใช้งานจริงในสถานพยาบาล และการปรับปรุงโมเดลด้วยการเรียนรู้ต่อเนื่อง (Continual Learning) เพื่อเพิ่มประสิทธิภาพและการใช้งานในโลกความเป็นจริง

Xuelong, C., Yiping, C. and Guojie, Y. [17] ได้ศึกษาการตรวจจับพฤติกรรมความรุนแรงทางไซเบอร์ (Cyberviolence Behavior: CVB) โดยนำแนวคิดทฤษฎีพฤติกรรมตามแผน (Theory of Planned Behavior: TPB) มาประยุกต์ใช้ร่วมกับเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ในการออกแบบกรอบการวิเคราะห์และตรวจจับพฤติกรรม CVB ทั้งแบบชัดแจ้ง (Explicit) และแฝงเร้น (Implicit) จากความคิดเห็นบนสื่อสังคมออนไลน์ (Social Media Platforms: SMPs) การศึกษาได้ออกแบบชุดคุณลักษณะเชิงปริมาณโดยอิงจากองค์ประกอบหลักของ TPB ได้แก่ ทัศนคติ (Attitude) บรรทัดฐานทางสังคม (Subjective Norm) และการรับรู้การควบคุมพฤติกรรม (Perceived Behavioral Control: PBC) ร่วมกับปัจจัยบริบททางสังคมที่มีผลทางอ้อม โมเดลตรวจจับที่ใช้มีชื่อว่า M²EDCI ซึ่งรวมคุณลักษณะจากสามส่วนหลัก ได้แก่ 1) ข้อมูลข้อความผ่าน Word2Vec และ BiLSTM เพื่อดึงบริบทระดับคำและประโยค 2) บริบทของบทสนทนาผ่าน Graph Attention Network (GAT) แบบหลายชั้นเพื่อวิเคราะห์ความสัมพันธ์ระหว่างความคิดเห็นในเครือข่ายการตอบกลับ และ 3) ข้อมูลบริบทที่ไม่ใช่เชิงภาษา เช่น เพศ การศึกษา ซึ่งมีความสัมพันธ์กับปัจจัยตามแนวคิด TPB โมเดลสุดท้ายรวมคุณลักษณะทั้งหมดเข้าด้วยกันและจำแนกประเภทของความคิดเห็นออกเป็น 3 ประเภท ได้แก่ ความคิดเห็นปกติ ความรุนแรงทางไซเบอร์แบบแฝง และแบบชัดแจ้ง ผลลัพธ์ของการวิจัยชี้ให้เห็นว่าโมเดลสามารถจับพฤติกรรม CVB ได้อย่างมีประสิทธิภาพ และชี้ให้เห็นถึงความสำคัญของการบูรณาการองค์ความรู้จากทฤษฎีทางสังคมร่วมกับเทคโนโลยีปัญญาประดิษฐ์ใน

การวิเคราะห์พฤติกรรมออนไลน์.

Xue, S. and et al. [18] ได้ทำการศึกษาผลของการให้ข้อมูล Bayesian Updating ต่อพฤติกรรมการตัดสินใจแบบซ้ำในการซื้อสินค้าออนไลน์ โดยมุ่งเน้นไปที่การลดอคติจากสถานะเดิม (status quo bias) ซึ่งเกิดจากการเลือกซ้ำแบบอัตโนมัติผ่านฟีเจอร์ “Buy Again” หรือ “Order Again” บนแพลตฟอร์มออนไลน์ การวิจัยใช้รูปแบบการทดลองโดยอิงจากปริศนา Monty Hall และวิเคราะห์ข้อมูลด้วยแบบจำลอง HDDM เพื่อวัดกลไกทางการรับรู้ในการทดลองแบ่งเป็น 2 การศึกษา 1. ผู้เข้าร่วมสามารถเลือกตัวเลือกเริ่มต้นเอง พบว่า การให้ข้อมูล Bayesian Updating ทำให้ผู้เข้าร่วมเปลี่ยนใจจากตัวเลือกเดิมมากขึ้น มี Reaction Time นานขึ้น และมีค่า Non-decision time, Starting Point, Drift Rate และ Threshold สูงกว่าอย่างมีนัยสำคัญทางสถิติ 2. ระบบเป็นผู้กำหนดตัวเลือกเริ่มต้นแทนผู้เข้าร่วม ผลการทดลองพบว่า Bayesian Updating ทำให้มีการเปลี่ยนใจมากขึ้นและมีการเลือกที่ถูกต้องมากขึ้น โดยมีค่า Drift Rate และ Starting Point สูงกว่ากรณีไม่มีข้อมูล Bayesian อย่างมีนัยสำคัญ แต่ไม่มีผลต่อ Non-Decision time และ Threshold จากผลการศึกษาทั้งสองส่วนสรุปได้ว่า ข้อมูล Bayesian Updating มีบทบาทสำคัญในการยับยั้งพฤติกรรมการตัดสินใจแบบอัตโนมัติ โดยช่วยเพิ่มความหลากหลายในการเลือกสินค้าและลดการพึ่งพาตัวเลือกเดิม

Shuixia, C. and et al. [19] ได้ศึกษาการพยากรณ์พฤติกรรมการยกเลิกการจองห้องพักของลูกค้าในธุรกิจโรงแรม โดยใช้เทคนิค Machine Learning หลายรูปแบบ ได้แก่ Lasso Regression, Artificial Neural Network (ANN), และ XGBoost พร้อมทั้งผสานตัวแปรจาก Bayesian Network (BN) และสร้างตัวแปรปฏิสัมพันธ์ (Interactive Features) เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ การศึกษานี้ประเมินผลด้วยการทดสอบซ้ำแบบ Bootstrap จำนวน 30 ชุด ผลการวิจัยพบว่าโมเดล ANN ที่ผนวก BN และ Interactive Features (ANN-original-Bayesian-interaction) ให้ประสิทธิภาพดีที่สุด โดยได้ค่าความถูกต้อง (Accuracy) เท่ากับ 83.09%, ค่าความระลึก (Recall) 58.85%, ค่าความแม่นยำ (Precision) 73.81% และค่า F1-score 65.49% ซึ่งสูงกว่า Lasso และ XGBoost ทุกรูปแบบ ตัวแปรที่มีอิทธิพลมากที่สุดในโมเดล ได้แก่ Lead_point, Lead_likehood, Deposit_likehood และ Special_likehood ซึ่งเป็นตัวแปรจาก BN ที่สะท้อนความน่าจะเป็นตามประวัติการจองของลูกค้า รวมถึงตัวแปรปฏิสัมพันธ์ เช่น Hotel_country2 และ Lead_time_babies1 ซึ่งแสดงให้เห็นถึงความสำคัญของการสร้างตัวแปรใหม่เพื่อเสริมข้อมูลเดิม ผลการศึกษาเน้นย้ำว่าการผสานข้อมูลจาก BN และตัวแปรปฏิสัมพันธ์สามารถยกระดับความแม่นยำของโมเดลได้อย่างมีนัยสำคัญ ทั้งในโมเดลเชิงเส้นและไม่เชิงเส้น

Hemang, S. and et al. [20] ได้ศึกษาการพยากรณ์ราคาบิตคอยน์โดยใช้สัญญาณข้อมูลหลากหลายประเภทร่วมกับเทคนิคการเรียนรู้เชิงลึกแบบ LSTM เพื่อปรับปรุงการทำนายและ

การจัดการความเสี่ยง โดยเน้นการคัดเลือกสัญญาณที่เกี่ยวข้องกับความเคลื่อนไหวของตลาดสกุลเงินดิจิทัล การศึกษานี้ใช้ชุดข้อมูลรวม 55 สัญญาณที่สะท้อนทั้งข้อมูลเชิงอารมณ์ ข่าวสาร และข้อมูลราคาตลาด การประเมินผลใช้เกณฑ์มาตรฐานทางการเงิน ได้แก่ อัตราส่วน Sharpe, การวัดความเบ้ (Skewness), ความแหลม (Kurtosis), การลดลงสูงสุด (Maximum Drawdown) และผลตอบแทนสะสม (Cumulative Returns) ผลการวิเคราะห์พบว่าสัญญาณผสม (Mixed Signals) ที่ผ่านการคัดเลือกโดย LSTM มีประสิทธิภาพสูงกว่าสัญญาณกลุ่มต่าง ๆ และเกณฑ์มาตรฐานที่ใช้เปรียบเทียบ โดยสามารถสร้างผลตอบแทนที่มีความเสี่ยงเหมาะสมและลดการขาดทุนได้ดีกว่า นอกจากนี้ โมเดล LSTM ที่ใช้ยังช่วยแยกสัญญาณข้อมูลที่มีประโยชน์จากสัญญาณรบกวนได้อย่างมีประสิทธิภาพ ทำให้การพยากรณ์มีความน่าเชื่อถือและเหมาะสมกับลักษณะข้อมูลในตลาดคริปโตเคอเรนซี ทั้งนี้ ผลการวิจัยชี้ให้เห็นถึงศักยภาพในการนำไปใช้จริงสำหรับนักลงทุนและผู้ปฏิบัติงานในวงการ รวมถึงโอกาสขยายผลไปสู่การใช้สัญญาณเชิงข้อความและโมเดลการเรียนรู้อื่น ๆ ในอนาคต

Li-Chen, C. and et al. [21] ได้ศึกษาโมเดลการตรวจจับผู้รีวิวปลอมโดยใช้เครือข่ายบริบททางสังคมร่วมกับเทคนิคกราฟนิวรัลเน็ตเวิร์ก (GCN และ GAT) ภายใต้โครงสร้างสองขั้นตอน พร้อมใช้ Focal Loss เพื่อจัดการปัญหาความไม่สมดุลของข้อมูล โดยสร้างกราฟผู้ใช้–ผู้ใช้ และผู้ใช้–โพสต์ เพื่อเรียนรู้พฤติกรรมและเนื้อหาที่เกี่ยวข้อง ผลการทดลองกับชุดข้อมูลจริงที่มีอัตราผู้รีวิวปลอมต่ำมาก (312:1) แสดงว่าโมเดลที่รวมข้อมูลผู้ใช้และโพสต์ด้วยกลไก GAT (User_PostGAT) มีประสิทธิภาพสูงสุด โดยให้ F2-Score, AUC และ AUPRC ดีกว่าโมเดลอื่นอย่างชัดเจน ทั้งนี้ โมเดลสามารถตรวจจับผู้รีวิวปลอมได้อย่างครอบคลุม (Recall สูงถึง 88.1%) แม้ในสภาวะข้อมูลจำกัด และสามารถประยุกต์ใช้ในระบบตรวจสอบของแพลตฟอร์มออนไลน์ได้จริง พร้อมเปิดโอกาสต่อยอดการวิจัยในบริบทอื่น ๆ ที่มีโครงสร้างสังคมคล้ายคลึงกัน

Jiwei, L. and et al. [22] ได้ศึกษาวิธีการตรวจจับรีวิวปลอมโดยใช้วิธี Probabilistic Modeling ในชื่อ FRDGMPM ซึ่งเน้นวิเคราะห์ความแตกต่างของการแจกแจงคุณลักษณะระหว่างรีวิวปลอมและรีวิวจริง โดยใช้คุณลักษณะเชิงภาษา พฤติกรรม และความสัมพันธ์ระหว่างผู้รีวิว สร้างการแจกแจงผสมสำหรับแยกประเภทรีวิว จากนั้นใช้ MLP ในการจำแนกรีวิวที่สร้างข้อมูลสังเคราะห์จากการแจกแจงนี้ ผลการทดลองกับชุดข้อมูลจริง 3 ชุด ได้แก่ YelpZip, YelpNYC และ YelpChi พบว่า FRDGMPM ให้ผลการตรวจจับที่แม่นยำกว่าระบบเดิมอย่างมีนัยสำคัญ โดยมีความแม่นยำ (Accuracy) สูงถึง 86.79%, 89.73% และ 86.62% ตามลำดับ และมี F-Measure สูงถึง 92.92%, 94.59% และ 92.83% ตามลำดับ โดยเฉพาะในชุดข้อมูลขนาดใหญ่ YelpZip และ YelpNYC ซึ่งมีจำนวนรีวิวมากกว่า YelpChi ถึงเกือบ 100 เท่า นอกจากนี้ โมเดลยังสามารถตรวจจับรีวิวปลอมได้ดีแม้ใช้ข้อมูลรีวิวต้นฉบับเพียงเล็กน้อยเพียง 100 ตัวอย่าง ส่งผลให้โมเดลสามารถทำงานแบบกึ่งเรียนรู้ (semi-supervised) ได้ และพบว่าคุณลักษณะพฤติกรรมและความสัมพันธ์ระหว่างผู้รีวิวเป็นปัจจัย

สำคัญที่ช่วยเพิ่มประสิทธิภาพของโมเดล ทั้งนี้ FRDGMPM มีความยืดหยุ่นในการเพิ่มเติมคุณลักษณะใหม่ ๆ และสามารถนำไปใช้ในบริบทอื่น ๆ เช่น การตรวจจับบอทปลอมในแพลตฟอร์มอีคอมเมิร์ซได้จริง

Stephanie, D., Kristof, C. and Arno, C. [23] ได้ศึกษาการพยากรณ์เหตุการณ์ชีวิตของลูกค้าโดยใช้ข้อมูลลูกค้าทางการเงินจากผู้ให้บริการทางการเงินในยุโรป โดยเปรียบเทียบวิธีการแปลงข้อมูลย้อนหลังในรูปแบบต่างๆ เพื่อพยากรณ์เหตุการณ์สำคัญต่างๆ เช่น การซื้อบ้านครั้งแรก หรือการเปลี่ยนแปลงสถานะความสัมพันธ์ ผลการศึกษาพบว่าวิธีการรวมคุณลักษณะย้อนหลังแบบ Aggregation ช่วยเพิ่มประสิทธิภาพการพยากรณ์มากกว่าการใช้ข้อมูลแบบ Snapshot และอัลกอริธึม XGBoost ให้ผลลัพธ์ที่ดีที่สุด นอกจากนี้ โมเดล LSTM ที่ใช้ข้อมูลย้อนหลังยังให้ประสิทธิภาพดีกว่าโมเดลแบบดั้งเดิม ผลการวิเคราะห์ยังแสดงให้เห็นว่าคุณลักษณะเชิงพฤติกรรม (Dynamic Features) มีความสำคัญต่อการพยากรณ์เหตุการณ์ชีวิตมากกว่าคุณลักษณะคงที่ (Static Features) และช่วงเวลาที่อยู่ใกล้เหตุการณ์ยิ่งมีผลต่อการพยากรณ์มากขึ้น ผลการศึกษายังชี้ให้เห็นว่าเหตุการณ์ชีวิตบางประเภทส่งผลต่อความภักดีของลูกค้า ขณะที่เหตุการณ์บางประเภทอาจเพิ่มความเสี่ยงต่อการเลิกใช้บริการ โดยข้อมูลนี้สามารถนำไปใช้วางแผนการตลาดที่เหมาะสม เช่น การกำหนดเวลาติดต่อหรือเสนอผลิตภัณฑ์เฉพาะกลุ่ม เพื่อเพิ่มประสิทธิภาพการบริหารความสัมพันธ์กับลูกค้า และแนะนำแนวทางการวิจัยในอนาคต เช่น การทดสอบ A/B เพื่อวัดผลการตลาดแบบเฉพาะเจาะจง และการพัฒนาโมเดลที่ซับซ้อนขึ้นเพื่อเพิ่มประสิทธิภาพการพยากรณ์.

Shuixia, C. and et al. [24] ได้ศึกษาการพยากรณ์ผลการยกเลิกการจองโรงแรม (Booking Cancellation) โดยใช้ข้อมูล PNR จากสองชุดข้อมูลหลักในงานวิจัย โดยมีการเตรียมข้อมูลล่วงหน้าด้วยการจัดการค่าว่างและการตรวจสอบค่าผิดปกติ เพื่อลดผลกระทบจากข้อมูลที่ไม่สมบูรณ์และข้อมูลผิดปกติ ในการศึกษาชุดที่หนึ่งมีจำนวน 85,274 ตัวอย่าง และ 27 ตัวแปรอิสระ ส่วนชุดที่สองมี 4,235 ตัวอย่าง และ 15 ตัวแปรอิสระ เพื่อสร้างแบบจำลองการพยากรณ์ที่มีประสิทธิภาพและความน่าเชื่อถือสูง ในการเปรียบเทียบผลการทำงานของแบบจำลองต่าง ๆ เช่น XGBoost (XGB), Bayesian Network (BN), Lasso regression และ Artificial Neural Network (ANN) พบว่าแบบจำลองที่ผสมผสานเทคนิค Bayesian และการวิเคราะห์ปฏิสัมพันธ์ของตัวแปรมีประสิทธิภาพเหนือกว่าแบบจำลองพื้นฐานอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) โดยผลการทดสอบ paired t-test บ่งชี้ว่าแบบจำลองผสม (เช่น Lasso-original-Bayesian-interaction และ ANN-original-Bayesian-interaction) มีค่า Accuracy, Recall, Precision และ F1-score สูงกว่ารูปแบบอื่นอย่างชัดเจน ทั้งนี้ ผลการทดสอบความทนทานของแบบจำลองยังชี้ให้เห็นว่าแบบจำลองผสมสามารถรักษาความแม่นยำได้ดีแม้เมื่อลดการเลือกคุณสมบัติและเปลี่ยนวิธีการฝึกอบรม ผลการวิจัยนี้ชี้ให้เห็นถึงความสำคัญของการรวมความรู้เชิงสถิติและความสัมพันธ์ของตัวแปรในการพัฒนาระบบพยากรณ์ที่มี

ประสิทธิภาพสำหรับการจัดการการยกเลิกการจองโรงแรม โดยแนวทางในอนาคต คือ การทดสอบแบบจำลองกับชุดข้อมูลจากแหล่งที่หลากหลายและสถานการณ์จริง รวมถึงการปรับปรุงโมเดลให้เหมาะสมกับการใช้งานเชิงปฏิบัติและเพิ่มความสามารถในการเรียนรู้แบบต่อเนื่อง (Continual Learning) เพื่อการประยุกต์ใช้ในระบบบริการลูกค้าอย่างยั่งยืน

Haoqiang, J., Paulo, G., Debra, V. [25] ได้ศึกษาการแก้ปัญหาการมอบหมายพยาบาลในหอผู้ป่วยใน (NPA) โดยเน้นความสำคัญของความต่อเนื่องในการดูแล (Continuity of Care) เสนออัลกอริธึมฮิวริสติกที่คำนึงถึงทั้งความสมดุลภาระงาน (AAI) และความสัมพันธ์ระหว่างพยาบาลกับผู้ป่วยเดิม สร้างแบบจำลอง Discrete-Event Simulation เพื่อประเมินผลลัพธ์ในช่วงเวลา 56 เวร เมื่อกำหนด Threshold ความต่อเนื่องที่ 0.7 สามารถให้การดูแลต่อเนื่องแก่ผู้ป่วยอาการหนักได้ถึง 63% ในขณะที่ค่า AAI เพิ่มจาก 0.5 เป็น 1 สะท้อนภาระงานที่เพิ่มขึ้นแต่ได้ความต่อเนื่องที่มากขึ้น โมเดลสามารถปรับใช้กับข้อมูลจริงของโรงพยาบาลที่มีคะแนน Acuity จากระบบต่าง ๆ เช่น QuadMed มีความยืดหยุ่นต่อสถานการณ์ เช่น การรับผู้ป่วยใหม่ หรือการเปลี่ยนเวรของพยาบาล สามารถนำไปใช้ในระบบช่วยตัดสินใจเพื่อจัดตารางมอบหมายพยาบาลรายวัน รองรับแนวทางที่กำหนดความต่อเนื่องเฉพาะกลุ่ม เช่น ผู้ป่วยอาการหนัก แนะนำให้ศึกษาเพิ่มเติมถึงผลกระทบทางอารมณ์และความพึงพอใจของพยาบาลในงานต่อเนื่อง

สุภาวดี [26] ได้ศึกษาปัจจัยทำนายภาวะสับสนเฉียบพลันในผู้สูงอายุที่เข้ารับบริการในแผนกฉุกเฉินโรงพยาบาลตติยภูมิในเขตกรุงเทพมหานคร โดยใช้แบบประเมิน 4AT, IQCODE, NRS และ PHQ-9 วิเคราะห์ด้วยถดถอยโลจิสติก พบว่าปัจจัยที่มีนัยสำคัญ ได้แก่ ภาวะการรู้คิดบกพร่อง (OR=17.24), ภาวะเลือดออกในสมอง (OR=7.69), ภาวะซึมเศร้า (OR=6.31), ประวัติโรคหลอดเลือดสมอง (OR=3.54), ระดับความปวด (NRS) (OR=1.19) และระยะเวลาอยู่ในแผนกฉุกเฉิน (OR=1.04) โมเดลมีค่า Nagelkerke R^2 = 0.716 และความแม่นยำ 90.9% สะท้อนศักยภาพในการใช้ทำนายผู้สูงอายุที่มีความเสี่ยง delirium ได้อย่างมีประสิทธิภาพ แนะนำให้ประยุกต์แบบประเมินร่วมกับข้อมูลเวชระเบียนเพื่อคัดกรองผู้ป่วยกลุ่มเสี่ยง ตั้งแต่แรกรับบริการในห้องฉุกเฉิน และพัฒนาระบบช่วยตัดสินใจทางคลินิกเพื่อสนับสนุนการจัดการดูแลแบบเฉพาะกลุ่ม เช่น ผู้ป่วยที่มีภาวะสมองเสื่อมหรือภาวะซึมเศร้า ควรศึกษาเพิ่มเติมถึงผลของการแทรกแซงเชิงระบบต่อการลดอุบัติการณ์ของ Delirium

กฤต [27] ได้ศึกษาการระบุสภาวะความเครียดของพืชในระยะเริ่มต้นโดยใช้การเรียนรู้เชิงลึกที่พัฒนาบน FPGA (Early Stress Detection) เพื่อเพิ่มผลผลิตในอนาคต โดยใช้แบบจำลอง CNN และแบบจำลอง CNN+LSTM กับภาพต้นข้าวฟ่างกว่า 40,000 ภาพจากหลายมุมมอง ภายใต้เงื่อนไขการขาดสารอาหาร 3 กลุ่ม ทดลองใช้แบบจำลอง CNN บนบอร์ด PYNQ-Z1 (FPGA) พบว่า ความแม่นยำสูงสุดเพียง 61% และเกิดปัญหา Uunderfit เนื่องจากข้อจำกัดด้านทรัพยากรของ FPGA แม้

ประมวลผลเร็วกว่า CPU ถึง 38 เท่า แต่ไม่เหมาะกับการใช้งานจริง ขณะที่แบบจำลอง VGG16+LSTM สามารถระบุสภาวะความเครียดได้แม่นยำสูงถึง 85% ภายใน 3 วันหลังพืชขาดสารอาหาร และสูงถึง 91% ในช่วงวันที่ 13 แบบจำลองนี้สามารถเรียนรู้ลักษณะเชิงเวลาได้ดีกว่า CNN แบบเดี่ยว แม้ใช้เวลาประมวลผลนานกว่าถึง 20.72 เท่า แต่เหมาะกับการใช้งานที่ไม่เน้น Real-time แสดงให้เห็นว่าโมเดลเชิงลำดับอย่าง LSTM เหมาะสำหรับการตรวจจับความเปลี่ยนแปลงทางกายภาพของพืชในระยะเริ่มต้น แนะนำให้ศึกษาการปรับขนาดโมเดลเพื่อนำไปใช้ในระบบฝังตัวจริง และพัฒนา dataset สำหรับพืชชนิดอื่นหรือสภาวะเครียดรูปแบบต่าง ๆ เพิ่มเติม

ปวริศ [28] ได้ศึกษาการพัฒนาต้นแบบการคาดการณ์ระยะเวลาเดินทางบนถนนในกรุงเทพฯ ด้วยการเรียนรู้ของเครื่อง โดยใช้ข้อมูล Mobile Probe จากมูลนิธิ iTIC ซึ่งครอบคลุม 4,634 คัน และแปลงเป็นข้อมูล Origin-Destination พร้อมตัวแปร 29 รายการ เช่น พิกัด เวลา เขต และสภาพอากาศ หลังจากทำความสะอาดข้อมูลและใช้ K-Fold Cross Validation (K=5) พร้อม RandomizedSearchCV 100 ครั้ง พบว่า Random Forest มีประสิทธิภาพสูงสุด (MAE = 134.72, MAPE = 43.06%) แต่ใช้ทรัพยากรมาก ขณะที่ XGBoost และ LightGBM ให้ผลใกล้เคียงแต่ใช้เวลาน้อยลง สำหรับ CatBoost แม้ประสิทธิภาพโดยรวมยังน้อยกว่า (MAE = 148.41, MAPE = 52.55%) แต่เด่นที่ใช้เวลาฝึกสั้นกว่าโมเดลอื่นอย่างมาก จึงมีศักยภาพในการประยุกต์ใช้ในงานที่ต้องการการฝึกโมเดลอย่างรวดเร็ว เสนอให้ศึกษาต่อในด้านการปรับพารามิเตอร์เพิ่มเติม และอัปเดตข้อมูลหลังยุค COVID-19 เพื่อเพิ่มความแม่นยำของต้นแบบ

ณัฐชา [29] ได้ศึกษาการจำแนกรหัสโรคแบบหลายฉลาก (Multi-label ICD Classification) เพื่อสนับสนุนการเบิกจ่ายค่ารักษาพยาบาลอย่างถูกต้องและครบถ้วน โดยพัฒนาแบบจำลองผสมผสานระหว่างโมเดลเรียนรู้เชิงลึก (Bi-LSTM, CNN, Transformer Encoder) สำหรับรหัสโรคทั่วไป และโมเดลทางสถิติ (Multinomial Naïve Bayes) สำหรับรหัสโรคที่พบได้ยาก จากข้อมูลเวชระเบียน 2 ประเภท ได้แก่ บันทึกการวินิจฉัยโรคและบันทึกการให้ยา/หัตถการ ผลการทดลองแบ่งเป็น 3 ส่วน พบว่า โมเดล Bi-LSTM มีความแม่นยำสูงสุดในการทำนายรหัส ICD-10 (Jaccard index = 0.792) และการใช้วิธีการผสมแบบ MaxEnsemble ให้ผลดีที่สุดเมื่อรวมทั้งสองกลุ่มรหัส (Jaccard index = 0.138) อย่างไรก็ตาม แบบจำลองยังมีข้อจำกัดในการทำนายรหัสโรคที่พบน้อยซึ่งมีตัวอย่างไม่เพียงพอ ส่งผลให้ค่า Precision ต่ำ แนะนำให้ศึกษาต่อยอดโดยเพิ่มเทคนิคการถ่วงน้ำหนักกลุ่มรหัสหายาก หรือใช้โมเดลเฉพาะทางสำหรับกลุ่มรหัสที่มีข้อมูลจำกัด เพื่อยกระดับความครอบคลุมและความถูกต้องของระบบจำแนกรหัสในระบบเวชระเบียนอิเล็กทรอนิกส์ต่อไป

ตารางที่ 2-1 สรุปงานวิจัยที่เกี่ยวข้อง

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอรึทึม	ผลการวิจัย
ระบบสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรัง กรณีของโรคเบาหวานและความดันโลหิตสูง [10]	2559	- Decision Tree - K-Nearest Neighbor (K-NN)	จากการวิจัยพบว่าาเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องที่ดีกว่าคือร้อยละ 96.45 และได้ทำการจัดกลุ่มตามเกณฑ์ปิงปองจราจรชีวิต 7 สี
การเปรียบเทียบประสิทธิภาพ อัลกอริทึมและการคัดเลือก คุณลักษณะที่เหมาะสม เพื่อการทำนายผลสัมฤทธิ์ ทางการเรียนของนักศึกษา ระดับอาชีวศึกษา [11]	2561	- Decision Tree - Naïve Bayes - Rule Induction -Forward Selection	ผลการทำนายผลสัมฤทธิ์ทางการเรียน ของนักศึกษาระดับอาชีวศึกษา ด้วยอัลกอริทึม Decision Tree อัลกอริทึม J48 graft ร่วมกับ การคัดเลือกคุณลักษณะ แบบ Forward Selection มีความเหมาะสมในการนำไปทำนาย ค่าความถูกต้องร้อยละ 83.08

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
ได้ศึกษาการวิเคราะห์คุณลักษณะทางเสียงสำคัญที่ส่งผลต่อคะแนนการทำงานสมองด้านแอตเทนชันในแบบประเมินไทยโมคา [12]	2565	- Decision Tree - Gradient Boosting - Random Forest - Linear Regression - Perceptron	ผลการวิจัยพบว่าการใช้ Recursive Feature Elimination (RFE) ร่วมกับหลายตัวประเมิน สามารถคัดเลือกคุณลักษณะเสียงสำคัญที่ทำนายคะแนนแอตเทนชันได้แม่นยำ โดยมีความแม่นยำสูงสุดถึง 72.0%
แบบจำลองการทำนายอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ [13]	2562	- Decision Tree - Random Forest - Naïve Bayes - Deep Learning	ผลการสร้างแบบจำลองการทำนาย แบบอากาศยานอัตโนมัติที่ไม่พึ่งพา มนุษย์ พบว่า อัลกอริทึม Random Forest ให้ประสิทธิภาพสูงที่สุด ค่าความถูกต้องร้อยละ 90.2 ค่าความแม่นยำร้อยละ 90.47 และค่าระลึกร้อยละ 90.24
การพยากรณ์รูปแบบ ของการกระทำความผิดซ้ำ ของเด็กและเยาวชน [14]	2562	eXtreme Gradient Boosting	ผลการพยากรณ์โดยใช้ปัจจัย ที่คัดเลือกพบว่าอัลกอริทึม eXtreme Gradient Boosting ค่าความถูกต้องร้อยละ 97.80 ค่าความแม่นยำร้อยละ 97.80 และค่าระลึกร้อยละ 100

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
การพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวาน [15]	2566	- Decision Tree - Random Forest - Gradient Boosting Machine	ผลการสร้างแบบจำลองความเสี่ยงในการเกิดโรคเบาหวาน พบว่า Random Forest ให้ประสิทธิภาพสูงที่สุด ค่าความถูกต้องร้อยละ 92.5 ค่าความแม่นยำร้อยละ 91.7 ค่าระลึกร้อยละ 92.8
ระบบพยากรณ์โรคตับไขมันชนิดเมตาบอลิก (MAFLD) โดยใช้กฎสัมพันธ์ตามช่วงเวลา [16]	2025	- Gradient Boosting Machine - CatBoost - Random Forest	ผลการศึกษาพบว่า การนำกฎสัมพันธ์ตามช่วงเวลา (TARs) จำนวน 20-30 กฎ เข้ามาใช้ในแบบจำลองช่วยเพิ่มประสิทธิภาพการทำนายโรค MAFLD ได้อย่างมีนัยสำคัญ โดยโมเดล GBM มีความทนทานสูงสุดต่อความไม่สมดุลของข้อมูล ให้ค่า F1-Score สูงสุดและมีความสามารถในการอธิบายผลลัพธ์ทั้งในระดับภาพรวมและรายบุคคล

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
การตรวจจับพฤติกรรมความรุนแรงทางไซเบอร์ [17]	2025	- Word2Vec - BiLSTM - GAT - Multilayer Attention Network - M²EDCI- อัลกอริทึมเพื่อนบ้านใกล้ที่สุด - อัลกอริทึมนาอีฟเบย์	งานวิจัยนี้พัฒนาแบบจำลอง M²EDCI เพื่อตรวจจับพฤติกรรมความรุนแรงทางไซเบอร์จากความคิดเห็นออนไลน์ โดยอิงทฤษฎี TPB และใช้เทคนิค Deep Learning ผสานข้อมูลภาษา โครงสร้างสนทนา และบริบททางสังคม เพื่อจำแนกความคิดเห็นได้อย่างมีประสิทธิภาพ
การศึกษาผลของการให้ข้อมูล Bayesian Updating ต่อพฤติกรรมและกลไกทางการรับรู้ในการตัดสินใจแบบซ้ำภายใต้สถานการณ์เลือกเองและระบบเลือกให้ [18]	2025	- Bayesian Updating - Monty Hall Paradigm - Hierarchical Drift Diffusion Model (HDDM) - Within-subject Experimental Design	การให้ข้อมูล Bayesian Updating เพิ่มอัตราการเปลี่ยนใจ (change choice) จาก 61.55% เป็น 69.71% และเพิ่มจำนวนการเลือกที่ถูกต้องจาก 12.17 เป็น 21.66 ครั้ง โดยมีผลต่อพารามิเตอร์ drift rate (0.241 ⟶ 0.468)

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
ศึกษาการพยากรณ์พฤติกรรมการยกเลิกการจองห้องพักของลูกค้าในธุรกิจโรงแรม โดยใช้เทคนิค Machine Learning หลายรูปแบบ [19]	2023	- Lasso Regression - Artificial Neural Network (ANN) - Bayesian Network (BN) - XGBoost (XGB)	งานวิจัยนี้เปรียบเทียบประสิทธิภาพของอัลกอริทึม Lasso Regression, ANN และ XGB ในการพยากรณ์ โดยทดสอบการผนวกผลลัพธ์จาก Bayesian Network และการสร้างคุณลักษณะเชิงปฏิสัมพันธ์ ผลพบว่า ANN ที่รวม BN และ Interactive Features ให้ผลดีที่สุด (Accuracy = 0.8309, Recall = 0.5885, Precision = 0.7381, F1-score = 0.6549) ขณะที่ XGB มี Accuracy = 0.8047 แต่ยังด้อยกว่ารูปแบบ ANN ที่พัฒนาเพิ่มเติม
การพยากรณ์ราคาบิตคอยน์โดยใช้สัญญาณข้อมูลหลากหลายประเภทร่วมกับเทคนิคการเรียนรู้เชิงลึกแบบ LSTM [20]	2024	- Long Short-Term Memory (LSTM) - ARMA (AutoRegressive Moving Average)	การใช้สัญญาณข้อมูลผสมในโมเดล LSTM ช่วยเพิ่ม Sharpe Ratio และผลตอบแทนสะสม พร้อมลดความเสี่ยงจากการขาดทุนสูงสุด โดยปรับปรุงพารามิเตอร์ Drift Rate และ Starting Point ทำให้โมเดลทำนายราคาบิตคอยน์ได้แม่นยำและน่าเชื่อถือขึ้น

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
การศึกษาโมเดลการตรวจจับผู้รีวิวปลอมโดยใช้เครือข่ายบริบททางสังคมร่วมกับเทคนิคกราฟนิวรัลเน็ตเวิร์ก (GCN และ GAT) [21]	2024	- Graph Convolutional Network (GCN) - Graph Attention Network (GAT) - Two-Stage Architecture - Focal Loss	การใช้สัญญาณข้อมูลผสมในโมเดล LSTM ช่วยเพิ่ม Sharpe ratio เป็น 1.048 และผลตอบแทนสะสมเป็น 83.8% พร้อมลดความเสี่ยงจากการขาดทุนสูงสุด (maximum drawdown) เหลือ 20.0% โดยปรับพารามิเตอร์ drift rate และ starting point ทำให้โมเดลทำนายราคาบิตคอยน์ได้แม่นยำและน่าเชื่อถือมากขึ้น
การศึกษาวิธีการตรวจจับรีวิวปลอมโดยใช้วิธี probabilistic modeling [22]	2023	- FRDGMPM - MLP (Multi-Layer Perceptron)	งานวิจัยนี้ใช้ฟีเจอร์ผสมเชิงภาษา พฤติกรรม และความสัมพันธ์ในโมเดล FRDGMPM ร่วมกับ MLP เพื่อเพิ่มความแม่นยำในการตรวจจับรีวิวปลอม โดยได้ Accuracy สูงสุด 89.73% และ F-Measure 94.59% ในชุดข้อมูล YelpNYC อีกทั้งการใช้ฟีเจอร์ความสัมพันธ์ช่วยเพิ่ม Accuracy ใน YelpZip จาก 60.15% เป็น 86.79% แม้ใช้ข้อมูลฝึกเพียง 100 ตัวอย่าง ก็ยังคงรักษาประสิทธิภาพได้ดี เหมาะสำหรับใช้งานจริงในระบบอีคอมเมิร์ซ

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
ศึกษาการพยากรณ์เหตุการณ์ชีวิตของลูกค้าโดยใช้ข้อมูลลูกค้าทางการเงินจากผู้ให้บริการทางการเงินในยุโรป [23]	2024	- XGBoost - LSTM - SaML Classifiers	การใช้ฟีเจอร์สถิตและไดนามิกในโมเดล LSTM ร่วมกับการฟีเจอไรเซชันแบบ aggregation ช่วยเพิ่มความแม่นยำพยากรณ์เหตุการณ์ชีวิตลูกค้าได้สูงสุด 78.5% โดยฟีเจอร์ไดนามิกมีอิทธิพลมากกว่าถึง 78% ทำให้สามารถกำหนดช่วงเวลาที่เหมาะสมสำหรับทำการตลาดเฉพาะบุคคล เพิ่มอัตรารักษาลูกค้าได้มากกว่าร้อยละ 10 และพัฒนากลยุทธ์ CRM ได้มีประสิทธิภาพยิ่งขึ้น
ศึกษาการพยากรณ์ผลการยกเลิกการจองโรงแรม (Booking Cancellation) [24]	2023	- XGBoost (XGB) - Bayesian Network (BN) - Lasso Regression - Artificial Neural Network (ANN)	งานวิจัยนี้ศึกษาการพยากรณ์การยกเลิกการจองโรงแรมจากข้อมูล PNR จำนวน 85,274 และ 4,235 ตัวอย่าง โดยพัฒนาโมเดลด้วย XGBoost, Bayesian Network, Lasso Regression, ANN และการผสมผสานระหว่าง Bayesian กับ Lasso และ ANN พร้อมวิเคราะห์ปฏิสัมพันธ์ของตัวแปร ผลการทดสอบด้วย paired t-test พบว่าโมเดลที่รวม Bayesian และวิเคราะห์ปฏิสัมพันธ์ให้ผลลัพธ์ดีกว่าอย่างมีนัยสำคัญ ($p < 0.05$ ถึง $p < 0.01$) และมีความคงทนภายใต้สถานการณ์ต่าง ๆ

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
ศึกษาการแก้ปัญหาการมอบหมายพยาบาลในหอผู้ป่วยใน (NPA) [25]	2023	- Heuristic Algorithm - Mixed-Integer Programming (MIP) - Discrete-Event Simulation	ศึกษาการพยากรณ์การยกเลิกจองโรงแรมจากข้อมูล PNR โดยพัฒนาโมเดลด้วย XGBoost, Bayesian Network, Lasso และ ANN พร้อมผสานการวิเคราะห์ปฏิสัมพันธ์ตัวแปร ผลพบว่าโมเดลที่รวม Bayesian ให้ประสิทธิภาพสูง (Accuracy > 85%, F1-score ≈ 81%) และยังคงความแม่นยำได้แม้ลดตัวแปรและเปลี่ยนวิธีฝึกอบรม
ศึกษาปัจจัยทำนายภาวะสับสนเฉียบพลันในผู้สูงอายุที่เข้ารับบริการในแผนกฉุกเฉิน [26]	2566	- Binary Logistic Regression - Decision Tree	ศึกษาการพยากรณ์ภาวะสับสนเฉียบพลันในผู้สูงอายุที่แผนกฉุกเฉินด้วย Logistic Regression พบว่าโมเดลสามารถทำนายได้อย่างมีนัยสำคัญ (Nagelkerke R^2 = .370, $p < .05$) โดยมีปัจจัยสำคัญ ได้แก่ ภาวะรู้คิดบกพร่อง, โรคหลอดเลือดสมอง, ระยะเวลาเข้ารับบริการ, เลือดออกในสมอง และภาวะซึมเศร้า.

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
การระบุสภาวะความเครียดของพืชในระยะเริ่มต้นโดยใช้การเรียนรู้เชิงลึกที่พัฒนาบน FPGA [27]	2566	- CNN - VGG16+LSTM - GoogLeNet	ศึกษาการระบุสภาวะความเครียดของต้นข้าวฟ่างจากภาพถ่ายกว่า 40,000 ภาพภายใต้การขาดสารอาหาร 3 กลุ่ม โดยเปรียบเทียบแบบจำลอง CNN บนบอร์ด PYNQ-Z1 กับ VGG16+LSTM พบว่า CNN แม้ประมวลผลเร็วกว่า CPU 38 เท่า แต่ความแม่นยำต่ำ (≤ 61%) ขณะที่ VGG16+LSTM ระบุสภาวะเครียดได้แม่นยำถึง 85% ภายใน 3 วัน และ 91% ในวันที่ 13 เหมาะสำหรับระบบที่ไม่เน้น real-time และมีศักยภาพในเกษตรอัจฉริยะ
การคาดการณ์เวลาเดินทางในกรุงเทพฯ ด้วยการเรียนรู้ของเครื่อง[28]	2566	- Random Forest - XGBoost - LightGBM - CatBoost	ศึกษาการสร้างต้นแบบคาดการณ์เวลาเดินทางในกรุงเทพฯ โดยใช้ข้อมูล Mobile Probe และตัวแปรเชิงพิกัด-เวลา-สภาพอากาศ 29 ตัว พบว่า Random Forest ให้ผลดีที่สุด ขณะที่ CatBoost แม่นยำน้อยกว่าแต่ฝึกเร็วกว่า เหมาะกับการพัฒนาเร็ว ข้อจำกัดคือข้อมูลปี 2020 ได้รับผลกระทบ COVID-19 และต้องปรับแต่งโมเดลเพิ่มเพื่อความแม่นยำสูงขึ้น

ตาราง 2-1 (ต่อ)

ชื่องานวิจัย/ทฤษฎี	ปี	เทคนิค/อัลกอริทึม	ผลการวิจัย
การจำแนกรหัสโรคแบบหลายฉลากด้วยโมเดลผสมผสานเพื่อช่วยการเบิกจ่ายค่ารักษาพยาบาลของโรงพยาบาล [29]	2566	-Multinomial Naïve Bayes -Bidirectional LSTM -Convolutional Neural Network (CNN) -Transformer Encoder -Ensemble Models (Average & Maximizing)	ศึกษาการจำแนกรหัส ICD แบบหลายฉลากจากเวชระเบียน โดยใช้ Naïve Bayes สำหรับรหัสหายากและ Deep Learning สำหรับรหัสทั่วไป พบว่า Bi-LSTM มีความแม่นยำสูงสุด (Jaccard = 0.792/0.9364) และการรวมโมเดลแบบ MaxEnsemble ช่วยเพิ่มผลลัพธ์ (Jaccard = 0.138) ข้อจำกัดคือข้อมูลรหัสหายากมีน้อย แนะนำเพิ่มการถ่วงน้ำหนักและพัฒนาโมเดลเฉพาะทางเพื่อความแม่นยำสูงขึ้น

จากการทบทวนทฤษฎี แนวคิด และงานวิจัยที่เกี่ยวข้องข้างต้น แสดงให้เห็นว่าการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลและอัลกอริทึมการเรียนรู้ของเครื่อง เช่น CatBoost, Random Forest และ Naive Bayes มีประสิทธิภาพสูงในการจำแนกประเภทข้อมูลและพยากรณ์พฤติกรรมในหลากหลายโดเมน ทั้งยังมีการใช้เทคนิค RFE เพื่อคัดเลือกคุณลักษณะสำคัญและการใช้เทคนิค DiCE เพื่อเพิ่มความสามารถในการอธิบายผลด้วยทางเลือกจำลอง ผู้วิจัยจึงมีแนวคิดในการพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยสำหรับคนไทยในพื้นที่กรุงเทพมหานครและปริมณฑลขึ้น เพื่อช่วยลดความเสี่ยงทางการเงินในสภาวะเศรษฐกิจเปราะบาง โดยจะนำอัลกอริทึมที่ได้ศึกษามาเปรียบเทียบประสิทธิภาพเพื่อให้ได้โมเดลที่มีความแม่นยำและเหมาะสมที่สุดในการนำไปพัฒนาเป็นระบบต้นแบบที่ตอบสนองต่อความต้องการของผู้ใช้งานได้อย่างมีประสิทธิภาพต่อไป

บทที่ 3
วิธีดำเนินงานวิจัย

ผู้วิจัยได้ทำการศึกษาค้นคว้า ทฤษฎี แนวคิด และงานวิจัยที่เกี่ยวข้องต่าง ๆ เพื่อนำมาประยุกต์ใช้ในงานวิจัย โดยมีรายละเอียด และขั้นตอนดังต่อไปนี้

3.1 ขั้นตอนการสร้างกรอบแนวความคิดของงานวิจัย

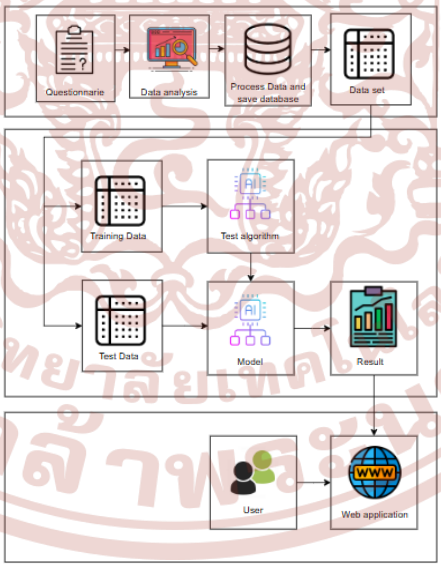
3.2 การศึกษา รวบรวมและจัดเตรียมข้อมูล

3.3 ขั้นตอนการเปรียบเทียบและคัดเลือกอัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุด

3.4 ขั้นตอนการออกแบบ และพัฒนาระบบ

3.5 ขั้นตอนการประเมินการใช้งานระบบ

3.1 ขั้นตอนการสร้างกรอบแนวความคิดของงานวิจัย



ภาพที่ 3-1 กรอบแนวคิด

ผู้วิจัยได้กำหนดแนวทางการดำเนินงานวิจัยและจัดทำกรอบแนวคิดที่ครอบคลุมขั้นตอนการทำงานอย่างเป็นระบบ โดยสามารถแบ่งส่วนการดำเนินงานออกเป็น 3 ส่วนหลัก ดังนี้

3.1.1 ส่วนการรวบรวมและเตรียมข้อมูล เริ่มต้นจากการเก็บรวบรวมข้อมูลผ่านเครื่องมือแบบสอบถามที่ผ่านการประเมินค่าดัชนีความสอดคล้อง (IOC) จากผู้เชี่ยวชาญ เพื่อให้ได้ชุดข้อมูลที่มีความน่าเชื่อถือจากประชากรในเขตกรุงเทพมหานครและปริมณฑล จากนั้นเข้าสู่กระบวนการเตรียม

ข้อมูล (Data Preparation) โดยการปรับสมดุลชุดข้อมูล (Data Resampling) เพื่อแก้ไขปัญหาความไม่สมดุลของข้อมูล (Class Imbalance) และประยุกต์ใช้เทคนิคการตัดคุณลักษณะแบบเรียกซ้ำ (Recursive Feature Elimination: RFE) เพื่อคัดเลือกปัจจัยที่มีอิทธิพลสำคัญ ซึ่งช่วยลดความซับซ้อนและเพิ่มประสิทธิภาพให้กับแบบจำลอง

3.1.2 ส่วนการประมวลผลและคัดเลือกแบบจำลอง ในส่วนนี้เป็นการนำชุดข้อมูลที่ผ่านการเตรียมความพร้อมมาแบ่งออกเป็นข้อมูลส่วนการเรียนรู้ (Training Data) และข้อมูลส่วนการทดสอบ (Test Data) เพื่อดำเนินการเปรียบเทียบประสิทธิภาพของอัลกอริทึมจำนวน 3 รูปแบบ ได้แก่ CatBoost, Convolutional Neural Network (CNN) และ Multinomial Naive Bayes โดยมีจุดมุ่งหมายเพื่อประเมินและคัดเลือกแบบจำลองที่ให้ค่าความแม่นยำสูงสุด เพื่อนำไปใช้เป็นกลไกหลักในการประมวลผลของระบบสนับสนุนการตัดสินใจ

3.1.3 ส่วนการพัฒนาระบบและการแสดงผล ผู้วิจัยได้นำแบบจำลองที่มีประสิทธิภาพดีที่สุดมาบูรณาการเข้ากับระบบสนับสนุนการตัดสินใจที่พัฒนาขึ้นในรูปแบบเว็บแอปพลิเคชัน (Web Application) โดยใช้ภาษา Python ในการประมวลผลส่วนหลัง (Backend) ร่วมกับ React ในการพัฒนาส่วนติดต่อผู้ใช้งาน (Frontend) นอกจากนี้ยังได้ประยุกต์ใช้เทคนิคการอธิบายเชิงเหตุผลแบบโต้แย้งที่หลากหลาย (Diverse Counterfactual Explanations: DiCE) เพื่อสร้างผลลัพธ์การทำนายพร้อมคำแนะนำเชิงรุก ซึ่งช่วยให้ผู้ใช้เห็นแนวทางในการปรับเปลี่ยนเงื่อนไขส่วนบุคคล เพื่อประกอบการตัดสินใจเลือกซื้อหรือเช่าที่พักอาศัยได้อย่างแม่นยำและเหมาะสมกับสภาวะเศรษฐกิจในปัจจุบัน

โดยมีขั้นตอนการดำเนินงาน แสดงดังภาพที่ 3-2



ภาพที่ 3-2 ขั้นตอนการดำเนินงานวิจัย

3.2 การศึกษา รวบรวมและจัดเตรียมข้อมูล

ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลและจัดเตรียมชุดข้อมูล (Dataset) เพื่อนำมาใช้ในการวิเคราะห์และพัฒนาแบบจำลอง โดยมีรายละเอียดขั้นตอนดังนี้

3.2.1 การรวบรวมข้อมูล (Data Collection) ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูลปฐมภูมิ (Primary Data) โดยใช้แบบสอบถามออนไลน์เป็นเครื่องมือในการเก็บข้อมูลจากกลุ่มตัวอย่างที่เป็นประชาชนในเขตกรุงเทพมหานครและปริมณฑล จำนวนทั้งสิ้น 1,107 ราย โดยเครื่องมือวิจัยได้ผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) ด้วยค่าดัชนีความสอดคล้อง (IOC) จากผู้เชี่ยวชาญ ชุดข้อมูลที่รวบรวมได้ประกอบด้วยปัจจัยสำคัญ ได้แก่ ข้อมูลทางประชากรศาสตร์ ประกอบด้วย เพศ อายุ ระดับการศึกษา ปัจจัยทางเศรษฐกิจประกอบด้วย รายได้เฉลี่ยต่อเดือน ภาระหนี้สิน) และพฤติกรรมการตัดสินใจเลือกที่อยู่อาศัย

3.2.2 การทำความสะอาดข้อมูล (Data Cleansing) ข้อมูลที่รวบรวมได้จะถูกนำมาเข้าสู่กระบวนการตรวจสอบคุณภาพเพื่อกำจัดข้อมูลที่มีความซ้ำซ้อน ข้อมูลที่ไม่สมบูรณ์ (Missing Values) หรือข้อมูลที่ผิดปกติ (Outliers) เพื่อให้แน่ใจว่าชุดข้อมูลมีความสะอาดและพร้อมสำหรับการประมวลผลด้วยอัลกอริทึมการเรียนรู้ของเครื่อง

3.2.3 การปรับสมดุลชุดข้อมูล (Data Balancing) เนื่องจากชุดข้อมูลเบื้องต้นมีลักษณะความไม่สมดุลของระดับชั้นข้อมูล (Class Imbalance) ระหว่างกลุ่มผู้ตัดสินใจซื้อและกลุ่มผู้ตัดสินใจเช่า ซึ่งอาจส่งผลให้แบบจำลองเกิดความลำเอียง (Bias) ผู้วิจัยจึงได้ประยุกต์ใช้เทคนิค Synthetic Minority Over-sampling Technique (SMOTE) เพื่อสร้างข้อมูลสังเคราะห์สำหรับกลุ่มที่มีจำนวนน้อย (Minority Class) ให้มีความสมดุลกับกลุ่มที่มีจำนวนมาก (Majority Class) กระบวนการนี้ส่งผลให้ได้ชุดข้อมูลรวมสำหรับการทดลองเพิ่มขึ้นเป็นจำนวน 1,456 ตัวอย่าง ซึ่งช่วยเพิ่มประสิทธิภาพให้แบบจำลองสามารถเรียนรู้ลักษณะเด่นของทุกกลุ่มข้อมูลได้อย่างเท่าเทียมกัน

3.2.4 ผู้วิจัยดำเนินการแปลงข้อมูลคุณลักษณะที่อยู่ในรูปแบบกลุ่มหรือข้อความ (Categorical Data) ให้เป็นรหัสตัวเลข (Label Encoding) เพื่อให้สอดคล้องกับข้อกำหนดของอัลกอริทึมที่จะนำมาทดสอบ พร้อมทั้งจัดทำโครงสร้างข้อมูลให้อยู่ในรูปแบบตาราง (Tabular Data) ที่พร้อมสำหรับการนำไปเข้าสู่กระบวนการคัดเลือกคุณลักษณะสำคัญ (Feature Selection) ในส่วนของกระบวนการเตรียมความพร้อมก่อนการฝึกฝนแบบจำลอง ผู้วิจัยได้ดำเนินการแบ่งชุดข้อมูล (Data Splitting) ออกเป็น 2 ส่วน คือ ชุดข้อมูลสำหรับฝึกฝน (Training Set) ในสัดส่วนร้อยละ 80 และชุดข้อมูลสำหรับทดสอบ (Test Set) ในสัดส่วนร้อยละ 20 เพื่อใช้ประเมินความแม่นยำของแบบจำลองกับข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data) ควบคู่ไปกับการประยุกต์ใช้เทคนิคการประเมินผลแบบ 10-Fold Cross-Validation เพื่อยืนยันความทนทานและความน่าเชื่อถือทางสถิติของผลการวิจัย ซึ่งสอดคล้องกับแนวทางการทดสอบในงานวิจัยที่เกี่ยวข้องที่เน้นความแม่นยำและการลดความคลาดเคลื่อนในการพยากรณ์

3.3 ขั้นตอนการเปรียบเทียบและคัดเลือกอัลกอริทึมที่มีประสิทธิภาพที่ดีที่สุด

ในขั้นตอนนี้ ผู้วิจัยมุ่งเน้นการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง เพื่อคัดเลือกอัลกอริทึมที่เหมาะสมที่สุดสำหรับระบบสนับสนุนการตัดสินใจ โดยมีรายละเอียดขั้นตอนดังนี้

3.3.1 การคัดเลือกคุณลักษณะสำคัญ (Feature Selection) ผู้วิจัยประยุกต์ใช้เทคนิค RFE ร่วมกับอัลกอริทึม CatBoost เพื่อวิเคราะห์และจัดลำดับความสำคัญของคุณลักษณะโดยกระบวนการ RFE จะทำการตัดตัวแปรที่มีอิทธิพลน้อยที่สุดออกทีละลำดับ เพื่อค้นหาจำนวนคุณลักษณะที่น้อยที่สุดแต่ยังคงให้ประสิทธิภาพในการทำนายสูงสุด ซึ่งช่วยลดความซับซ้อนของแบบจำลอง และลดระยะเวลาในการประมวลผล

3.3.2 การกำหนดอัลกอริทึมสำหรับการทดลอง ผู้วิจัยได้เลือกอัลกอริทึมที่มีสถาปัตยกรรมแตกต่างกัน 3 รูปแบบ เพื่อนำมาเปรียบเทียบประสิทธิภาพ ดังนี้ CatBoost เป็นอัลกอริทึมในกลุ่ม Gradient Boosting Decision Tree ที่มีประสิทธิภาพสูงในการจัดการข้อมูลประเภทกลุ่ม โดยไม่ต้องทำการทำ One-hot Encoding ในทุกกรณี CNN แบบจำลองโครงข่ายประสาทเทียมเชิงลึก โดย

ผู้วิจัยได้ปรับเปลี่ยนโครงสร้างให้เหมาะสมกับการประมวลผลข้อมูลในรูปแบบตาราง (Tabular Data) Multinomial Naive Bayes: อัลกอริทึมพื้นฐานทางสถิติที่ใช้หลักความน่าจะเป็นของเบย์ (Bayes' Theorem) เพื่อใช้เป็นฐานในการเปรียบเทียบประสิทธิภาพ (Baseline Model)

3.3.3 การฝึกฝนและการทดสอบแบบจำลอง (Model Training and Testing) ผู้วิจัยดำเนินการแบ่งชุดข้อมูล (Data Splitting) ออกเป็น 2 ส่วน คือ ชุดข้อมูลสำหรับฝึกฝน (Training Set) และชุดข้อมูลสำหรับทดสอบ (Test Set) ในสัดส่วนที่เหมาะสม พร้อมทั้งทำการปรับจูนพารามิเตอร์ (Hyperparameter Tuning) ของแต่ละอัลกอริทึมเพื่อให้ได้ประสิทธิภาพสูงสุดในแต่ละบริบทของคุณลักษณะที่ถูกเลือกโดย RFE

3.3.4 เกณฑ์การประเมินประสิทธิภาพ (Performance Evaluation Metrics) เพื่อให้ได้แบบจำลองที่มีความน่าเชื่อถือ ผู้วิจัยได้ใช้เกณฑ์การวัดผลระดับสากล 5 ด้าน ประกอบด้วย

1.1.4.1. ค่าความถูกต้อง (Accuracy): อัตราส่วนของการทำนายที่ถูกต้องทั้งหมด

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (3\text{-}1)$$

1.1.4.2. ค่าความแม่นยำ (Precision): ความถูกต้องของการทำนายในแต่ละคลาส

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3\text{-}2)$$

1.1.4.3. ค่าความระลึก (Recall): ความสามารถในการตรวจจับข้อมูลในคลาสนั้น ๆ

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3\text{-}3)$$

1.1.4.4. ค่าการถ่วงดุล (F-Measure): ค่าเฉลี่ยฮาร์มอนิกระหว่าง Precision และ Recall

$$\text{F} - \text{Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+ Recall}} \quad (3\text{-}4)$$

1.1.4.5. ค่าพื้นที่ใต้กราฟ ROC (AUC): ตัวบ่งชี้ประสิทธิภาพการจำแนกประเภทในภาพรวม

1.1.4.5.1. TPR (True Positive Rate): ซึ่งก็คือค่า Recall

1.1.4.5.2. FPR (False Positive Rate): อัตราการทายผิดว่า "ใช่" ทั้งที่จริงๆ คือ "ไม่ใช่”

$$\mathrm{FRP} = 2 \times \frac{\mathrm{FP}}{\mathrm{FP+TN}} \tag{3-5}$$

1.1.4.5.3. AUC

$$\mathrm{AUC} = \int_0^1 TRP(FPR)\, dFPR \tag{3-6}$$

3.4 ขั้นตอนการออกแบบ และพัฒนาระบบ

ในลำดับสุดท้ายของการดำเนินงานวิจัย ผู้วิจัยได้อัลกอริทึม CatBoost มีประสิทธิภาพสูงสุดมาเป็นกลไกหลักในการพัฒนาระบบสนับสนุนการตัดสินใจต้นแบบ เพื่อให้สามารถใช้งานได้จริงในบริบทของการวิเคราะห์ข้อมูลที่พักอาศัย โดยมีขั้นตอนดังนี้

3.4.1 การออกแบบสถาปัตยกรรมระบบ (System Architecture) ผู้วิจัยกำหนดโครงสร้างการทำงานของระบบในลักษณะการเชื่อมต่อระหว่างส่วนหน้าบ้านและหลังบ้าน เพื่อให้รองรับการประมวลผลโมเดลการเรียนรู้ของเครื่องได้อย่างมีประสิทธิภาพ ประกอบด้วย

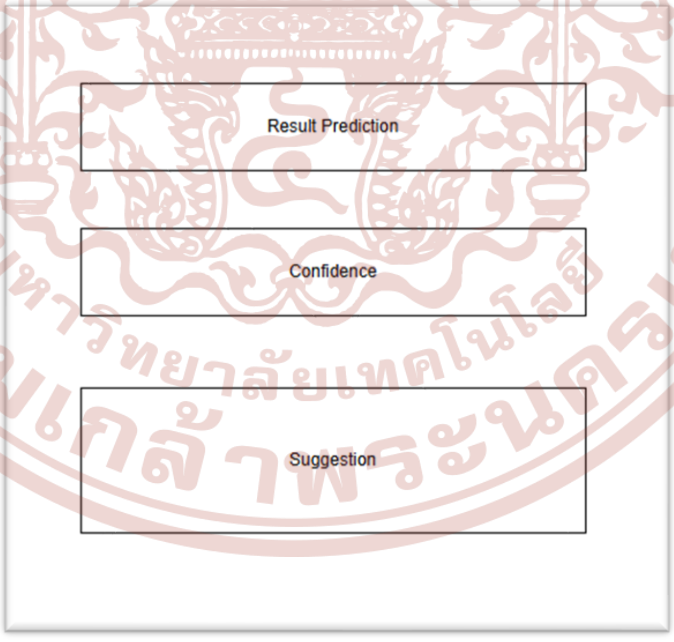
3.4.1.1 Frontend (ส่วนติดต่อผู้ใช้งาน): พัฒนาด้วยเทคโนโลยี React เพื่อสร้างส่วนแสดงผลที่เป็นมิตรต่อผู้ใช้ (User-friendly Interface) รองรับการรับค่าปัจจัยทางประชากรศาสตร์และเศรษฐกิจจากผู้ใช้งานประกอบด้วย

1.1.1.1.1. หน้าจอรับข้อมูลของระบบดังภาพที่ 3-2



ภาพที่ 3-3 ออกแบบหน้าจอรับข้อมูลของระบบ

1.1.1.1.2. หน้าจอผลลัพธ์ของระบบดังภาพที่ 3-3



ภาพที่ 3-4 ออกแบบหน้าจอผลลัพธ์ของระบบ

3.4.1.2 Backend (ส่วนประมวลผล): พัฒนาด้วยภาษา Python ซึ่งทำหน้าที่รับชุดข้อมูลจาก Frontend มาประมวลผลผ่านโมเดล CatBoost ที่ผ่านการฝึกฝน (Trained Model) แล้วและส่งผลลัพธ์การทำนายกลับไปแสดงผล

3.4.1.3 Model Integration: นำแบบจำลอง CatBoost ที่มีค่า Accuracy, Precision และ Recall สูงที่สุดตามสมมติฐาน มาทำการ Deploy เพื่อใช้สำหรับการพยากรณ์ข้อมูลแบบ Real-time

3.4.2 การประยุกต์ใช้เทคนิค Diverse Counterfactual Explanations (DiCE) เพื่อเพิ่มศักยภาพของระบบสนับสนุนการตัดสินใจ ผู้วิจัยได้บูรณาการเทคนิคการอธิบายผลด้วย DiCE เพื่อแก้ปัญหา "Black Box" ของปัญญาประดิษฐ์ โดยระบบจะไม่เพียงแค่บอกผลลัพธ์ว่า "ซื้อ" หรือ "เช่า" เท่านั้น แต่จะสามารถสร้างสถานการณ์สมมติ (Counterfactual) เพื่อให้คำแนะนำแก่ผู้ใช้งานได้ว่าหากต้องการเปลี่ยนผลลัพธ์จากการเช่าเป็นการซื้อ จะต้องปรับปรุงปัจจัยพื้นฐาน เช่น รายได้ต่อเดือน หรือการจัดการหนี้สินในระดับใด จึงจะเข้าเกณฑ์ความปลอดภัยทางการเงิน

3.5 ขั้นตอนการประเมินการใช้งานระบบ

.1.1. ประเมินความพึงพอใจสำหรับผู้ใช้งานระบบสนับสนุนการตัดสินในในการซื้อหรือเช่าที่พักจากกลุ่มเป้าหมาย 30 ราย โดยผู้วิจัยจะสร้างแบบสอบถามตามวิธี Likert ซึ่งมีลักษณะเป็นแบบมาตราส่วนประมาณค่า (Rating Scale) มีค่าตัวเลือก 5 ระดับ ให้ 5 หมายถึงมีความพึงพอใจมากที่สุด และ 1 หมายถึงมีความพึงพอใจน้อยที่สุด โดยแบ่งการประเมินเป็น 3 ด้าน ได้แก่

.1.1.1. ด้านฟังก์ชันการทำงานของระบบ (Functional Requirements) เป็นการประเมิน

.1.1.2. ด้านความง่ายต่อการใช้งาน (Usability) ประเมินความสะดวกในการใช้งานระบบผ่านเว็บแอปพลิเคชัน ความชัดเจนของการกรอกข้อมูล ความเข้าใจง่ายของคำแนะนำ และความพึงพอใจของผู้ใช้ต่อประสบการณ์ใช้งาน

.1.1.3. ด้านประสิทธิภาพ (Performance) ประเมินความรวดเร็วในการประมวลผลข้อมูลและแสดงผลคำแนะนำ รวมถึงความเสถียรของระบบในขณะใช้งาน

ตารางที่ 3-1 เกณฑ์การประเมินความพึงพอใจ

ระดับเกณฑ์การประเมินความพึงพอใจ		ความหมาย
เชิงคุณภาพ	เชิงปริมาณ	
ดีมาก	5	มีความพึงพอใจต่อระบบในระดับดีมาก
ดี	4	มีความพึงพอใจต่อระบบในระดับมาก

ปานกลาง	3	มีความพึงพอใจต่อระบบในระดับปานกลาง
น้อย	2	มีความพึงพอใจต่อระบบในระดับน้อย
น้อยที่สุด	1	มีความพึงพอใจต่อระบบในระดับน้อยที่สุด

3.5.2 วิเคราะห์ และสรุปผลการประเมินความพึงพอใจสำหรับผู้ใช้งานระบบสนับสนุนการตัดสินในในการซื้อหรือเช่าที่พัก โดยการประเมินความพึงพอใจมีเกณฑ์การแปลผล และพิจารณาจากการประเมินผลข้อมูลด้านความพึงพอใจในการใช้แอปพลิเคชัน ใช้วิธีการคำนวณผลโดยการหาค่าเฉลี่ย และการหาค่าเบี่ยงเบนมาตรฐาน (S.D.) โดยเลือกใช้วิธีการหาค่าเฉลี่ยเพื่อให้ทราบผลเฉลี่ยของความพึงพอใจในแต่ละด้าน และนำผลของค่าเฉลี่ยที่ได้มาตีความเทียบกับเกณฑ์ที่กำหนด โดยมีสูตรการคำนวณหาค่าเฉลี่ยดังนี้

สูตรการคำนวณหาค่าเฉลี่ย แสดงดังสมการที่ 3-1

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n} \quad (3\text{-}1)$$

เมื่อ $\bar{x}$ คือ ค่าเฉลี่ย

x_i คือ ค่าของข้อมูลตัวที่ i

n คือ จำนวนข้อมูลทั้งหมดที่มี

ตารางที่ 3-2 เกณฑ์ในการแปลผลระดับประเมินความพึงพอใจ

ระดับเกณฑ์การแปลความหมาย	ความหมาย
4.51 - 5.00	มีความพึงพอใจต่อระบบในระดับดีมาก
3.51 - 4.50	มีความพึงพอใจต่อระบบในระดับดี
2.51 - 3.50	มีความพึงพอใจต่อระบบในระดับปานกลาง
1.51 - 2.50	มีความพึงพอใจต่อระบบในระดับน้อย
1.00 - 1.50	มีความพึงพอใจต่อระบบในระดับน้อยมาก

จากนั้นใช้วิธีการหาค่าเบี่ยงเบนมาตรฐาน (S.D.) เพื่อวัดการกระจายทางสถิติโดยเปรียบเทียบระดับความพึงพอใจในแต่ละด้านว่ามีการกระจายตัวมากน้อยเพียงใด หากค่าเบี่ยงเบนมาตรฐานที่ได้มีค่ามาก แสดงว่า ผลรวมของแบบสอบถามความพึงพอใจในแต่ละด้านมีระดับความพึงพอใจที่แตกต่างกันสูง หากค่าเบี่ยงเบนมาตรฐานที่ได้มีค่าน้อย หรือเท่ากับศูนย์ แสดงว่า ผลรวมของแบบสอบถามความพึงพอใจ ในแต่ละด้านมีระดับความพึงพอใจที่แตกต่างกันน้อย หรือไม่แตกต่างกันโดยมีสูตรการคำนวณหาค่าเบี่ยงเบนมาตรฐาน ดังนี้

สูตรการคำนวณหาค่าเบี่ยงเบนมาตรฐาน แสดงดังสมการที่ 3-2

$$S.D. = \sqrt{\frac{(x - \bar{x})^2}{n - 1}} \qquad (3\text{-}2)$$

เมื่อ S.D. คือ ค่าเบี่ยงเบนมาตรฐาน

x_i คือ ค่าของข้อมูลตัวที่ i

$\bar{x}$ คือ ค่าเฉลี่ย

n คือ จำนวนข้อมูลทั้งหมดที่มี

พระจอมเกล้าพระนครเหนือ มหาวิทยาลัยเทคโนโลยี

บทที่4
ผลการดำเนินงานวิจัย

จากการสร้างแบบจำลองการพยากรณ์เพื่อจำแนกพฤติกรรมการตัดสินใจเลือกซื้อหรือเช่าที่พักอาศัย ประกอบไปด้วยการทดลองเปรียบเทียบอัลกอริทึมและจำนวนคุณลักษณะที่แตกต่างกัน เพื่อให้ได้แบบจำลองที่เหมาะสมที่สุดตามวิธีการดำเนินงานในบทที่ 3 ผู้วิจัยขอเสนอผลการดำเนินงานวิจัยดังนี้

4.1 ผลการศึกษา และรวบรวมข้อมูล

4.2 ผลการเตรียมชุดข้อมูล

4.3 ผลการสร้างและประเมินประสิทธิภาพแบบจำลองการพยากรณ์

4.4 ผลการออกแบบ และพัฒนาระบบ

4.5 ผลการประเมินความพึงพอใจต่อการใช้งานระบบสนับสนุนการตัดสินใจ

4.1 ผลการศึกษา และการรวบรวมข้อมูล

4.1.1 ผู้วิจัยได้นำแบบสอบถามที่สร้างขึ้นเสนอต่อผู้เชี่ยวชาญจำนวน 3 ท่าน เพื่อตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) โดยพิจารณาความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์และนิยามศัพท์เชิงปฏิบัติการ ซึ่งกำหนดเกณฑ์การให้คะแนนเป็น +1 สำหรับข้อที่แน่ใจว่าสอดคล้อง, 0 สำหรับข้อที่ไม่แน่ใจ และ -1 สำหรับข้อที่แน่ใจว่าไม่สอดคล้อง จากนั้นนำคะแนนที่ได้มาคำนวณหาค่า IOC รายข้อ โดยยึดเกณฑ์การคัดเลือกข้อคำถามที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไป ผลการพิจารณาของผู้เชี่ยวชาญ แสดงดังตารางที่ 4-1

ตารางที่ 4-1 ผลการประเมินค่าดัชนีความสอดคล้อง (IOC) ของแบบสอบถามโดยผู้เชี่ยวชาญ

ข้อคำถาม	ผลการประเมิณ			
	คนที่1	คนที่2	คนที่3	IOC
ช่วงอายุของผู้ทำแบบสอบถาม	1	1	1	1
เพศของผู้ทำแบบสอบถาม	1	1	1	1
พื้นที่อยู่อาศัย (กรุงเทพชั้นใน, กรุงเทพชั้นนอก, ปริมณฑล)	1	1	1	1
ภาระหนี้สินเทียบกับสินทรัพย์	1	1	1	1
มีเงินสำรองฉุกเฉิน 6 เดือนหรือไม่	1	1	1	1
รายได้ต่อเฉลี่ยต่อเดือน	1	1	1	1
ในมุมมองของท่านมองว่าที่พักอาศัยเป็นสินทรัพย์มากเพียงใด	1	1	1	1
หากท่านป่วยและต้องใช้ค่ารักษาพยาบาล 2 ล้านบาท ท่านมีความพร้อมรับมือเพียงใด	1	1	1	1
ท่านวางแผนการเกษียณไว้แล้ว	1	0	1	0.67
สถานะในที่ทำงาน เช่น พนักงานเอกชน ราชการ ฟรีแลนซ์	0	1	1	0.67
ในระยะเวลา 5 ปีที่ผ่านมามีการย้ายหรือเปลี่ยนสถานที่ทำงานบ่อยแค่ไหน	1	1	1	1
ต้องเดินทางไปที่ทำงาน สัปดาห์ละกี่วัน	1	1	1	1
การเดินทางไปสถานที่ทำงานใช้ค่าเดินทางเดือนละเท่าไร	1	1	1	1
ท่านต้องการอาศัยร่วมกัน หรือใกล้เคียงครอบครัว	1	1	1	1
สถานะปัจจุบัน	1	1	1	1
สถานที่พักที่อาศัยอยู่เป็นของใคร	1	1	1	1
คุณมีความพึงพอใจในที่พักอาศัยมากเพียงใด	0	1	1	0.67
ท่านมีบุตรหรือต้องการมีบุตรหรือไม่	1	1	1	1
ท่านต้องการรายได้จากที่พักอาศัยของท่านหรือไม่	1	1	1	1
ท่านมีความกดดันจากคนรอบตัวในการซื้อที่พักอาศัยหรือไม่	1	1	1	1
พฤติกรรมความชื่นชอบส่วนบุคคล				0

ตารางที่ 4-1 (ต่อ)

ข้อคำถาม	ผลการประเมิณ			
	คนที่ 1	คนที่ 2	คนที่ 3	IOC
ท่านสบายใจเมื่อได้พักในที่พักที่เป็นของตนเองมากเพียงใด	1	1	1	1
ท่านมีความชื่นชอบที่จะตกแต่งที่พักอาศัยของท่านมากเพียงใด	1	1	1	1
ท่านชอบที่พักอาศัยแบบใด (บ้านเดี่ยว, คอนโด, ฯลฯ)	1	1	1	1
ท่านคำนึงถึงค่าใช้จ่ายในการซื้อ/เช่าที่อยู่อาศัย เช่น ดอกเบี้ยหรือค่าเช่า	1	1	1	1
กังวลในการผ่อนชำระที่อยู่อาศัยในอนาคตมากเพียงใด	1	1	1	1
ที่อยู่อาศัยของท่านเดินทางสะดวกเพียงใด	1	1	1	1
ท่านพึงพอใจสภาพแวดล้อมที่อยู่อาศัยของท่านมากเพียงใด	1	1	1	1
ท่านมีอำนาจในการตัดสินใจเลือกซื้อหรือเช่าที่อยู่อาศัยหลังแรกหรือไม่	1	0	1	0.67
ท่านซื้อที่พักอาศัยแล้วใช่หรือไม่	1	0	1	0.67

จากผลการประเมินพบว่า ผู้เชี่ยวชาญทั้ง 3 ท่าน มีความเห็นสอดคล้องกันว่าข้อคำถามทุกข้อมีความเที่ยงตรงเชิงเนื้อหาและสอดคล้องกับวัตถุประสงค์ของการวิจัย โดยมีค่าดัชนีความสอดคล้อง (IOC) เท่ากับ 1.00 ทุกข้อคำถาม ซึ่งสูงกว่าเกณฑ์ที่กำหนด (0.50) แสดงให้เห็นว่าแบบสอบถามมีความเหมาะสมและสามารถนำไปใช้ใน

4.1.2 ผลการเก็บรวบรวมข้อมูล ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลปฐมภูมิผ่านแบบสอบถามออนไลน์ โดยได้รับข้อมูลที่สมบูรณ์จากกลุ่มตัวอย่างจำนวน 1,107 ราย ซึ่งเครื่องมือที่ใช้ประกอบด้วยข้อคำถามทั้งหมด 27 ข้อคำถาม (ตัวแปรต้น) และ 1 ตัวแปรตาม โดยสามารถจำแนกรายละเอียดข้อมูลได้ดังนี้

1. เพศ
2. อายุ
3. ระดับการศึกษา
4. อาชีพปัจจุบัน
5. สถานภาพการทำงาน (ประจำ/ชั่วคราว)
6. รายได้เฉลี่ยต่อเดือน

7. สถานภาพการสมรส
8. จำนวนสมาชิกในครอบครัว
9. ที่อยู่อาศัยในปัจจุบัน (บ้านตนเอง/เช่า/อยู่กับผู้ปกครอง)
10. จำนวนเงินออมคงเหลือในปัจจุบัน
11. ภาระหนี้สินคงเดิมที่มีอยู่ (เช่น ผ่อนรถ, บัตรเครดิต
12. ความสามารถในการผ่อนชำระต่อเดือนที่คาดว่าไหว
13. จำนวนเงินดาวน์ที่เตรียมไว้
14. ความมั่นคงในรายอาชีพในระยะยาว
15. แหล่งที่มาของรายได้ (รายได้หลัก/รายได้เสริม)
16. ภาระค่าใช้จ่ายคงที่อื่นๆ ต่อเดือน
17. ทำเลที่ตั้งที่ต้องการ (ใกล้ที่ทำงาน/ชานเมือง)
18. ความต้องการสิ่งอำนวยความสะดวก (ใกล้รถไฟฟ้า/ห้างสรรพสินค้า)
19. ระยะเวลาที่คาดว่าจะพักอาศัย (ระยะสั้น/ระยะยาว)
20. วัตถุประสงค์หลัก (อยู่เอง/ลงทุน)
21. ค่าใช้จ่ายในการเดินทางเฉลี่ยต่อเดือนในปัจจุบัน
22. ประเภทที่พักอาศัยที่สนใจ (บ้านเดี่ยว/คอนโด/ทาวน์โฮม)
23. ปัจจัยด้านความปลอดภัยในพื้นที่
24. อิทธิพลจากบุคคลรอบข้าง (ครอบครัว/เพื่อน)
25. การเข้าถึงข้อมูลรีวิวและโซเชียลมีเดีย
26. ความพึงพอใจต่อที่พักปัจจุบัน
27. แนวโน้มภาวะเศรษฐกิจที่มีผลต่อการตัดสินใจ

ตัวแปรตาม (Target Variable) พฤติกรรมการตัดสินใจ (เลือกซื้อ หรือ เลือกเช่า)

4.2 ผลการเตรียมชุดข้อมูล

หลังจากรวบรวมข้อมูลได้ 1,107 ชุด ผู้วิจัยได้ดำเนินการดังนี้

4.2.1 Data Cleansing: ตรวจสอบความสมบูรณ์ของข้อมูลทั้ง 28 หัวข้อ

4.2.2 Label Encoding: แปลงคำตอบจากตัวอักษรเป็นตัวเลข เพื่อให้โมเดลสามารถคำนวณได้ ตัวอย่างเช่น เพศชายเป็น 1 และเพศหญิงเป็น 2

4.2.3 Data Balancing (SMOTE): ทำการปรับสมดุลข้อมูลจนได้ชุดข้อมูลรวม 1,456 ตัวอย่าง

4.3 ผลการสร้างและประเมินประสิทธิภาพแบบจำลองการพยากรณ์

ในขั้นตอนการสร้างแบบจำลอง ผู้วิจัยได้นำชุดข้อมูลที่ผ่านการเตรียมความพร้อมและปรับสมดุลแล้ว (Balanced Dataset) จำนวน 1,456 ตัวอย่าง มาดำเนินการทดลองเพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึม 3 รูปแบบ ประกอบด้วย CatBoost, CNN และ Multinomial Naive Bayes โดยมีกระบวนการดำเนินงานเริ่มจากการเตรียมและการแบ่งข้อมูล (Data Splitting) ซึ่งผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็น 2 ส่วน คือ ชุดข้อมูลสำหรับการฝึกฝน (Training Set) ในสัดส่วนร้อยละ 80 เพื่อให้แบบจำลองเรียนรู้รูปแบบของข้อมูล และชุดข้อมูลสำหรับการทดสอบ (Test Set) ในสัดส่วนร้อยละ 20 เพื่อใช้ประเมินประสิทธิภาพความแม่นยำกับข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data) นอกจากนี้ยังได้ประยุกต์ใช้เทคนิค K-Fold Cross-Validation (K=10) เพื่อลดความลำเอียงจากการแบ่งข้อมูลและทำให้การประเมินผลมีความน่าเชื่อถือสูงสุด โดยแต่ละโมเดลจะผ่านกระบวนการค้นหาค่าพารามิเตอร์ที่ดีที่สุด (Hyperparameter Tuning) เพื่อให้ได้ประสิทธิภาพสูงสุดในการพยากรณ์ก่อนนำผลลัพธ์มาเปรียบเทียบกันดังนี้

4.3.1 แบบจำลอง CatBoost ผู้วิจัยได้ดำเนินการรันแบบจำลอง CatBoost โดยใช้กระบวนการค้นหาค่าพารามิเตอร์ที่ดีที่สุด (Hyperparameter Tuning) เพื่อให้เหมาะสมกับชุดข้อมูลที่มีปัจจัยคุณลักษณะหลากหลาย และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-2 ผลการรันแบบจำลอง Catboost เพื่อหา Best Parameter

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	20	3	0.01	20	800	4	50	1	78	0.78	0.78	0.86	0.78
2	30	2	0.03	20	500	5	32	0.5	82	0.81	0.81	0.90	0.81
3	30	3	0.02	20	500	4	32	1	79	0.78	0.78	0.87	0.78
4	30	3	0.03	50	500	6	50	1.5	82	0.81	0.81	0.90	0.81
5	20	3	0.01	20	300	4	32	1	76	0.75	0.75	0.84	0.75
6	10	2	0.01	10	300	4	50	1.5	.76	0.75	0.75	0.85	0.75
7	20	3	0.03	30	800	5	32	1.5	82	0.82	0.82	0.90	0.82
8	20	1	0.02	10	300	5	32	1	79	0.78	0.78	0.87	0.78
9	20	1	0.01	5	300	6	32	1	76	0.76	0.76	0.85	0.76
10	20	3	0.01	20	800	4	32	0.5	78	0.77	0.78	0.86	0.77
11	30	3	0.03	20	800	6	50	1.5	81	0.81	0.81	0.89	0.81
12	20	3	0.03	10	800	5	75	0.5	82	0.82	0.82	0.89	0.82
13	20	3	0.02	20	800	4	32	1	82	0.81	0.81	0.89	0.81
14	30	2	0.01	10	500	5	75	1.5	78	0.77	0.78	0.86	0.77
15	30	3	0.03	20	500	6	32	1.5	82	0.82	0.82	0.90	0.82
16	30	2	0.03	3	300	6	32	1.5	81	0.80	0.80	0.88	0.80
17	30	2	0.02	40	500	5	32	1.5	80	0.80	0.80	0.87	0.80
18	10	3	0.01	3	800	6	50	1	81	0.80	0.80	0.89	0.80
19	30	1	0.02	50	800	6	32	0.5	81	0.81	0.81	0.90	0.81
20	20	2	0.02	3	500	5	32	0.5	82	0.81	0.81	0.89	0.81
21	30	3	0.03	50	500	4	32	1.5	81	0.81	0.81	0.89	0.81
22	10	3	0.01	3	500	4	75	1.5	78	0.77	0.77	0.86	0.77
23	20	1	0.02	40	300	4	50	1	76	0.76	0.76	0.85	0.76
24	10	2	0.01	50	300	5	50	0.5	76	0.76	0.76	0.84	0.76
25	20	1	0.03	3	500	4	32	1	82	0.82	0.82	0.89	0.82

ตารางที่ 4-2 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
26	30	1	0.01	40	800	6	75	0.5	77	0.77	0.77	0.86	0.77
27	10	2	0.03	50	300	6	32	1	79	0.78	0.78	0.87	0.78
28	30	2	0.03	5	300	6	32	0.5	80	0.79	0.79	0.88	0.79
29	20	1	0.03	3	300	6	75	0.5	80	0.80	0.80	0.88	0.80
30	20	1	0.02	50	800	6	50	0.5	82	0.81	0.81	0.90	0.81
31	10	3	0.03	3	800	4	50	1.5	82	0.82	0.82	0.89	0.82
32	10	3	0.03	40	300	6	32	1	80	0.79	0.79	0.88	0.79
33	10	2	0.03	50	800	5	50	1.5	82	0.81	0.81	0.89	0.81
34	30	3	0.03	30	500	5	32	1.5	82	0.82	0.82	0.90	0.82
35	30	3	0.03	40	300	4	75	1	77	0.76	0.76	0.86	0.76
36	20	2	0.02	3	500	5	75	1	82	0.82	0.82	0.89	0.82
37	30	1	0.03	20	500	5	50	0.5	82	0.82	0.82	0.90	0.82
38	20	1	0.03	20	800	5	50	1	82	0.82	0.82	0.90	0.82
39	20	2	0.03	40	500	4	50	0.5	81	0.81	0.81	0.89	0.81
40	30	1	0.03	20	800	5	50	0.5	82	0.82	0.82	0.90	0.82
41	30	2	0.02	20	300	5	75	1	78	0.77	0.77	0.86	0.77
42	30	3	0.01	50	500	6	75	1	77	0.76	0.76	0.85	0.76
43	10	1	0.01	3	500	4	32	0.5	77	0.77	0.77	0.86	0.77
44	30	1	0.02	50	800	6	75	1.5	82	0.81	0.81	0.89	0.81
45	20	3	0.01	20	800	5	50	0.5	80	0.79	0.79	0.87	0.79
46	30	3	0.01	50	300	6	50	0.5	76	0.75	0.75	0.84	0.75
47	30	2	0.02	30	300	6	75	0.5	77	0.76	0.76	0.85	0.76
48	10	1	0.02	3	500	4	50	0.5	81	0.81	0.81	0.89	0.81
49	10	2	0.02	50	500	5	50	1	81	0.80	0.80	0.88	0.80
50	20	3	0.03	30	300	6	32	0.5	79	0.79	0.79	0.87	0.79

ตารางที่ 4-2 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
51	30	1	0.03	5	800	6	75	1.5	82	0.82	0.82	0.89	0.82
52	20	3	0.02	3	800	6	32	0.5	83	0.82	0.82	0.90	0.82
53	10	1	0.01	10	300	6	32	1.5	78	0.77	0.77	0.85	0.77
54	20	1	0.03	40	300	4	75	0.5	78	0.78	0.78	0.86	0.78
55	30	1	0.01	30	800	4	75	0.5	78	0.77	0.77	0.86	0.77
56	10	1	0.02	40	500	4	75	1	79	0.79	0.79	0.88	0.79
57	20	1	0.03	50	300	5	75	1.5	79	0.79	0.79	0.87	0.79
58	20	1	0.02	30	500	5	32	1.5	80	0.80	0.80	0.88	0.80
59	10	2	0.01	40	500	5	75	1.5	78	0.77	0.77	0.86	0.77
60	30	1	0.03	20	500	4	32	1	81	0.81	0.81	0.89	0.81
61	20	3	0.02	5	800	4	32	1.5	82	0.81	0.81	0.90	0.81
62	30	3	0.02	20	500	4	75	1	79	0.79	0.79	0.87	0.79
63	20	3	0.03	40	500	4	32	1.5	81	0.80	0.80	0.89	0.80
64	30	1	0.03	5	300	4	32	1.5	78	0.78	0.78	0.87	0.78
65	10	1	0.02	30	800	6	50	1	81	0.81	0.81	0.90	0.81
66	10	1	0.01	5	300	4	32	1.5	76	0.75	0.75	0.85	0.75
67	10	1	0.01	10	500	5	32	1.5	79	0.78	0.78	0.87	0.78
68	20	3	0.01	40	300	5	50	1	76	0.75	0.75	0.84	0.75
69	30	1	0.03	40	500	5	50	0.5	82	0.81	0.81	0.90	0.81
70	10	2	0.03	5	500	4	50	1	82	0.81	0.81	0.90	0.81
71	10	3	0.03	20	800	5	75	1	82	0.81	0.81	0.89	0.81
72	20	3	0.03	50	800	5	32	1	82	0.81	0.81	0.90	0.81
73	30	2	0.01	3	800	4	32	1	79	0.78	0.78	0.87	0.78
74	20	3	0.03	20	500	4	50	1	81	0.81	0.81	0.89	0.81
75	20	2	0.01	5	300	6	75	1	77	0.77	0.77	0.85	0.77

ตารางที่ 4-2 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
76	30	1	0.01	50	800	4	32	1.5	77	0.76	0.76	0.85	0.76
77	20	1	0.03	3	500	5	32	0.5	82	0.82	0.82	0.90	0.82
78	30	1	0.01	30	500	4	50	1.5	76	0.75	0.76	0.85	0.75
79	10	1	0.03	20	500	6	50	1	83	0.82	0.82	0.90	0.82
80	10	3	0.02	50	500	5	75	1.5	81	0.80	0.80	0.88	0.80
81	30	1	0.03	30	500	6	50	0.5	82	0.81	0.81	0.90	0.81
82	20	2	0.03	30	800	6	75	1	82	0.81	0.81	0.90	0.81
83	20	1	0.03	10	800	6	32	0.5	83	0.82	0.82	0.90	0.82
84	20	1	0.01	50	800	4	75	0.5	78	0.77	0.77	0.86	0.77
85	20	2	0.03	50	300	5	50	1	78	0.78	0.78	0.87	0.78
86	20	2	0.03	5	500	6	50	0.5	83	0.82	0.82	0.90	0.82
87	20	2	0.03	10	800	5	32	0.5	82	0.82	0.82	0.89	0.82
88	20	3	0.02	30	500	6	32	1.5	81	0.80	0.80	0.88	0.80
89	30	2	0.01	5	300	5	32	1.5	76	0.76	0.76	0.85	0.76
90	20	3	0.03	50	500	6	50	1.5	82	0.81	0.81	0.90	0.81
91	10	1	0.02	10	800	4	75	0.5	81	0.81	0.81	0.90	0.81
92	10	3	0.03	5	300	6	50	1	82	0.81	0.81	0.89	0.81
93	30	3	0.03	5	300	6	75	1	79	0.79	0.79	0.88	0.79
94	10	2	0.01	40	800	5	50	1.5	79	0.78	0.78	0.87	0.78
95	20	2	0.03	50	800	5	75	1.5	82	0.81	0.81	0.90	0.81
96	10	3	0.02	10	300	4	50	1	78	0.77	0.77	0.86	0.77
97	30	2	0.01	3	300	5	32	1	77	0.76	0.76	0.85	0.76
98	30	3	0.03	40	300	4	32	0.5	77	0.76	0.76	0.86	0.76
99	20	2	0.01	40	500	5	75	0.5	77	0.76	0.76	0.86	0.76
100	10	2	0.03	40	800	6	75	1	82	0.81	0.81	0.90	0.81

จากการทดสอบหาค่าพารามิเตอร์ที่ดีที่สุดผ่านวิธี Grid Search พบว่าเมื่อกำหนดค่าอัตราการเรียนรู้ (Learning Rate) และความลึกของต้นไม้ (Depth) ที่เหมาะสม แบบจำลองให้ผลลัพธ์ที่ดีที่สุด

คือค่าความถูกต้อง (Accuracy) สูงถึงร้อยละ 99.52 โดยมีค่า F1-Score ที่ร้อยละ 99.52 และมีค่าความถูกต้องของข้อมูลทดลองอยู่ที่ร้อยละ 80

4.3.2 แบบจำลอง CNN ผู้วิจัยได้ทำการออกแบบโครงสร้างโครงข่ายประสาทเทียมและทำการหาค่าพารามิเตอร์ที่ดีที่สุด ทั้งในส่วนของจำนวนชั้น (Layers) และอัตราการเรียนรู้ เพื่อให้โมเดลสามารถสกัดลักษณะเด่นจากข้อมูลตาราง ได้อย่างมีประสิทธิภาพ และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-3 ผลการรันแบบจำลอง CNN เพื่อหา Best Parameter

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	96	1	FALSE	128	0.2	0.0047	-	-	71	0.75	0.63	0.68	0.79
2	96	1	TRUE	32	0.1	0.0013	32	1	67	0.70	0.62	0.65	0.75
3	64	6	FALSE	32	0.3	0.0040	96	6	73	0.72	0.75	0.74	0.81
4	128	3	TRUE	64	0	0.0004	96	1	70	0.72	0.67	0.69	0.79
5	64	1	FALSE	128	0.2	0.0004	128	3	72	0.75	0.67	0.70	0.79
6	96	3	TRUE	96	0	0.0001	32	3	67	0.69	0.63	0.65	0.74
7	64	1	TRUE	32	0.1	0.0060	128	6	68	0.71	0.65	0.67	0.77
8	64	1	FALSE	96	0.4	0.0001	128	6	70	0.70	0.69	0.69	0.77
9	32	3	FALSE	64	0	0.0019	96	1	73	0.75	0.71	0.72	0.81

ตารางที่ 4-3 (ต่อ)

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
10	96	3	TRUE	128	0	0.0045	64	3	73	0.75	0.70	0.72	0.80
11	96	3	FALSE	64	0.4	0.0004	128	1	72	0.76	0.66	0.70	0.80
12	128	6	TRUE	128	0.4	0.0032	128	6	73	0.73	0.73	0.72	0.80
13	96	1	TRUE	128	0.2	0.0023	32	1	68	0.71	0.63	0.66	0.76
14	96	1	TRUE	96	0.3	0.0002	96	3	69	0.71	0.63	0.67	0.76
15	32	3	TRUE	64	0.4	0.0001	64	1	67	0.69	0.62	0.65	0.73
16	64	1	TRUE	128	0.4	0.0006	128	1	67	0.71	0.62	0.65	0.75
17	96	6	FALSE	32	0.3	0.0011	128	3	73	0.75	0.69	0.72	0.81
18	32	6	FALSE	64	0.2	0.0001	32	3	68	0.69	0.63	0.66	0.76
19	96	6	FALSE	96	0.4	0.0013	32	1	72	0.74	0.71	0.71	0.81
20	64	6	FALSE	32	0.3	0.0092	64	6	73	0.73	0.72	0.73	0.82
21	32	1	FALSE	64	0.1	0.0060	96	6	67	0.72	0.62	0.64	0.78
22	128	6	TRUE	32	0.4	0.0003	32	6	68	0.71	0.62	0.66	0.77
23	32	1	FALSE	32	0.2	0.0003	96	3	70	0.72	0.65	0.68	0.76
24	128	6	TRUE	128	0.4	0.0001	32	6	67	0.70	0.61	0.65	0.73
25	96	3	FALSE	96	0.4	0.0003	32	3	72	0.72	0.71	0.72	0.80

ตารางที่ 4-3 (ต่อ)

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
26	64	6	FALSE	128	0.1	0.0039	64	3	73	0.76	0.71	0.72	0.83
27	32	3	TRUE	96	0.3	0.0015	96	3	75	0.73	0.79	0.76	0.81
28	96	1	FALSE	32	0.2	0.0002	96	1	69	0.70	0.68	0.69	0.75
29	64	3	FALSE	128	0.4	0.0005	32	3	73	0.75	0.71	0.72	0.81
30	32	6	FALSE	64	0.2	0.0001	96	3	69	0.72	0.63	0.67	0.77

จากการปรับจูนโครงสร้างเครือข่ายประสาทเทียม พบว่าเมื่อใช้พารามิเตอร์ในจุดที่โมเดลเรียนรู้ได้ดีที่สุด (Best Configuration) แบบจำลองให้ผลลัพธ์ที่ดีที่สุดคือค่าความถูกต้องอยู่ที่ร้อยละ 95.40 และมีค่าความถูกต้องของข้อมูลทดลองอยู่ที่ร้อยละ 76 แม้ว่าจะมีประสิทธิภาพสูงแต่ยังคงน้อยกว่าแบบจำลอง CatBoost เล็กน้อยในบริบทของข้อมูลชุดนี้

4.3.3 แบบจำลอง Multinomial Naive Bayes ผู้วิจัยได้รันโมเดล Multinomial Naive Bayes โดยมีการหาค่าพารามิเตอร์ที่ดีที่สุด (Smoothing Parameter) เพื่อปรับจูนความแม่นยำตามหลักความน่าจะเป็น และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-4 ผลการรันแบบจำลอง Multinomial Naive Bayes เพื่อหา Best Parameter

ลำดับที่	alpha	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	0.0001	69	0.69	0.69	0.69	0.76
2	0.001	69	0.69	0.69	0.69	0.76
3	0.01	69	0.69	0.69	0.69	0.76
4	0.1	69	0.69	0.69	0.69	0.76
5	0.5	69	0.69	0.69	0.69	0.76
6	1	69	0.69	0.69	0.69	0.76
7	5	69	0.69	0.69	0.69	0.76
8	10	69	0.69	0.69	0.69	0.76
9	50	69	0.70	0.68	0.68	0.76
10	100	69	0.70	0.67	0.68	0.76

จากการค้นหาค่า Alpha ที่ดีที่สุดเพื่อลดข้อผิดพลาดในการคำนวณความน่าจะเป็น พบว่าแบบจำลองให้ผลลัพธ์ที่ดีที่สุดคือค่าความถูกต้องอยู่ที่ร้อยละ 88.70 และมีค่าความถูกต้องของข้อมูลทดลองอยู่ที่ร้อยละ 66.50 ซึ่งถือเป็นระดับที่ยอมรับได้สำหรับการเป็นโมเดลฐาน (Baseline Model) แต่ยังไม่เพียงพอสำหรับการนำไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจที่มีความซับซ้อน

4.3.4 ผลการคัดเลือกคุณลักษณะสำคัญด้วยเทคนิค RFE มื่อได้ข้อสรุปว่าแบบจำลอง CatBoost เป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดในการจำแนกข้อมูล ผู้วิจัยจึงได้นำโมเดลดังกล่าวมาดำเนินการคัดเลือกคุณลักษณะสำคัญ (Feature Selection) เพื่อลดจำนวนตัวแปรที่ไม่มีนัยสำคัญหรือมีความซ้ำซ้อนจากทั้งหมด 28 ปัจจัย ให้เหลือเพียงปัจจัยหลักที่ส่งผลต่อการตัดสินใจจริง โดยใช้เทคนิค RFE ซึ่งมีผลการดำเนินงานดังนี้

ตารางที่ 4-5 ผลการจัดลำดับความสำคัญโดย REF

ลำดับความสำคัญ	ข้อคำถาม	คะแนนความสำคัญ	RFE Score
1	อายุ	6.874	24
1	รายได้เฉลี่ยต่อเดือน	6.77	24
1	ทำเลที่ตั้งที่ต้องการ	5.239	24
1	ค่าใช้จ่ายในการเดินทาง	3.225	24
1	เพศ	3.148	24
2	จำนวนเงินดาวน์ที่เตรียมไว้	6.434	23
3	ภาระหนี้สินคงเดิมที่มีอยู่	6.193	22
4	สถานภาพการทำงาน	3.893	21
5	ปัจจัยด้านความปลอดภัยในพื้นที่	2.827	20
6	ประเภทที่พักอาศัยที่สนใจ	4.639	19
7	อาชีพปัจจุบัน	3.553	18
8	ความสามารถในการผ่อนชำระ	2.819	17
9	วัตถุประสงค์หลัก (อยู่เอง/ลงทุน)	4.203	16
10	จำนวนเงินออมคงเหลือ	3.584	15
11	ระยะเวลาที่คาดว่าจะพักอาศัย	3.133	14
12	ความมั่นคงในอาชีพระยะยาว	2.863	13
13	จำนวนสมาชิกในครอบครัว	2.618	12
14	ความต้องการสิ่งอำนวยความสะดวก	2.84	11
15	แหล่งที่มาของรายได้	2.311	10
16	อิทธิพลจากบุคคลรอบข้าง	2.558	9

ตารางที่ 4-5 (ต่อ)

ลำดับความสำคัญ	ข้อคำถาม	คะแนนความสำคัญ	RFE Score
17	แนวโน้มภาวะเศรษฐกิจ	2.66	8
18	สถานภาพการสมรส	2.549	7
19	ที่อยู่อาศัยในปัจจุบัน	2.558	6
20	ระดับการศึกษา	2.714	5
21	ความพึงพอใจต่อที่พักปัจจุบัน	2.13	4
22	การเข้าถึงข้อมูลรีวิว/โซเชียล	2.594	3
23	ปัจจัยด้านความปลอดภัย (ส่วนเสริม)	2.703	2
24	ภาระค่าใช้จ่ายคงที่อื่นๆ	2.366	1

ในการคัดเลือกตัวแปรอิสระเพื่อให้ได้โมเดลที่มีความกระชับ และลดปัญหาการเกิด Overfitting ตามแนวทางของ Kuhn และ Johnson [30] ผู้วิจัยได้ประยุกต์ใช้เกณฑ์การคัดเลือกเชิงประจักษ์ (Heuristic Threshold) โดยพิจารณาจากค่าความสำคัญของตัวแปร (Feature Importance) จากอัลกอริทึม CatBoost ที่มีค่าตั้งแต่ 3.0 ขึ้นไป ซึ่งเป็นระดับที่สูงกว่าค่าเฉลี่ยความสำคัญโดยรวม (Mean Importance) ของตัวแปรทั้งหมด ร่วมกับการใช้เทคนิค RFE ตามหลักการของ Guyon และคณะ [31] โดยมุ่งเน้นการคัดเลือกเฉพาะกลุ่มตัวแปรที่มีลำดับความสำคัญสูงสุดในกลุ่ม Top 25% (RFE Score 18-24) ซึ่งเป็นกลุ่มตัวแปรที่ทรงประสิทธิภาพและมีอิทธิพลต่อความแม่นยำในการพยากรณ์สูงสุด เพื่อให้มั่นใจว่าตัวแปรที่นำไปพัฒนาโมเดลนั้นมีนัยสำคัญ

4.3.5 ผลการทดสอบแบบจำลอง CatBoost ด้วย 10 คุณลักษณะผู้วิจัยนำคุณลักษณะสำคัญ 10 ลำดับแรกที่ได้จากการคัดเลือกด้วยเทคนิค RFE มาทำการฝึกฝนและทดสอบแบบจำลอง CatBoost โดยมีการหาค่าพารามิเตอร์ที่ดีที่สุด (Hyperparameter Tuning) และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-6 ผลการรันแบบจำลอง Catboost ด้วย 10 คุณลักษณะ

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	20	3	0.01	20	800	4	50	1	77	0.78	0.77	0.77	0.85
2	30	2	0.03	20	500	5	32	0.5	80	0.81	0.80	0.80	0.87
3	30	3	0.02	20	500	4	32	1	78	0.79	0.78	0.78	0.86
4	30	3	0.03	50	500	6	50	1.5	80	0.81	0.80	0.80	0.87
5	20	3	0.01	20	300	4	32	1	74	0.75	0.74	0.74	0.82
6	10	2	0.01	10	300	4	50	1.5	76	0.77	0.76	0.76	0.84
7	20	3	0.03	30	800	5	32	1.5	79	0.80	0.79	0.79	0.87
8	20	1	0.02	10	300	5	32	1	77	0.78	0.77	0.77	0.85
9	20	1	0.01	5	300	6	32	1	76	0.78	0.76	0.76	0.84
10	20	3	0.01	20	800	4	32	0.5	77	0.78	0.77	0.77	0.85
11	30	3	0.03	20	800	6	50	1.5	79	0.80	0.79	0.79	0.87
12	20	3	0.03	10	800	5	75	0.5	79	0.80	0.79	0.79	0.86
13	20	3	0.02	20	800	4	32	1	80	0.81	0.80	0.80	0.87
14	30	2	0.01	10	500	5	75	1.5	76	0.77	0.76	0.76	0.85
15	30	3	0.03	20	500	6	32	1.5	81	0.81	0.81	0.81	0.87
16	30	2	0.03	3	300	6	32	1.5	79	0.79	0.79	0.79	0.86
17	30	2	0.02	40	500	5	32	1.5	78	0.79	0.78	0.78	0.86
18	10	3	0.01	3	800	6	50	1	79	0.80	0.79	0.79	0.87
19	30	1	0.02	50	800	6	32	0.5	81	0.81	0.81	0.81	0.87
20	20	2	0.02	3	500	5	32	0.5	79	0.80	0.79	0.79	0.87
21	30	3	0.03	50	500	4	32	1.5	79	0.80	0.79	0.79	0.87
22	10	3	0.01	3	500	4	75	1.5	77	0.78	0.77	0.77	0.85
23	20	1	0.02	40	300	4	50	1	75	0.76	0.75	0.75	0.84
24	10	2	0.01	50	300	5	50	0.5	74	0.75	0.74	0.74	0.83
25	20	1	0.03	3	500	4	32	1	80	0.81	0.80	0.80	0.87

ตารางที่ 4-6 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
26	30	1	0.01	40	800	6	75	0.5	77	0.78	0.77	0.77	0.84
27	10	2	0.03	50	300	6	32	1	78	0.79	0.78	0.78	0.86
28	30	2	0.03	5	300	6	32	0.5	79	0.79	0.79	0.79	0.86
29	20	1	0.03	3	300	6	75	0.5	79	0.80	0.79	0.79	0.86
30	20	1	0.02	50	800	6	50	0.5	81	0.81	0.81	0.81	0.87
31	10	3	0.03	3	800	4	50	1.5	79	0.79	0.79	0.79	0.86
32	10	3	0.03	40	300	6	32	1	79	0.80	0.79	0.79	0.86
33	10	2	0.03	50	800	5	50	1.5	80	0.80	0.80	0.80	0.87
34	30	3	0.03	30	500	5	32	1.5	80	0.81	0.80	0.80	0.87
35	30	3	0.03	40	300	4	75	1	75	0.76	0.75	0.75	0.84
36	20	2	0.02	3	500	5	75	1	79	0.80	0.79	0.79	0.87
37	30	1	0.03	20	500	5	50	0.5	81	0.81	0.81	0.81	0.87
38	20	1	0.03	20	800	5	50	1	79	0.80	0.79	0.79	0.87
39	20	2	0.03	40	500	4	50	0.5	80	0.81	0.80	0.80	0.87
40	30	1	0.03	20	800	5	50	0.5	79	0.80	0.79	0.79	0.87
41	30	2	0.02	20	300	5	75	1	76	0.77	0.76	0.76	0.85
42	30	3	0.01	50	500	6	75	1	75	0.76	0.75	0.75	0.83
43	10	1	0.01	3	500	4	32	0.5	77	0.78	0.77	0.77	0.85
44	30	1	0.02	50	800	6	75	1.5	80	0.81	0.80	0.80	0.87
45	20	3	0.01	20	800	5	50	0.5	77	0.78	0.77	0.77	0.85
46	30	3	0.01	50	300	6	50	0.5	74	0.75	0.74	0.74	0.82
47	30	2	0.02	30	300	6	75	0.5	76	0.77	0.76	0.76	0.84
48	10	1	0.02	3	500	4	50	0.5	80	0.81	0.80	0.80	0.87
49	10	2	0.02	50	500	5	50	1	79	0.79	0.79	0.79	0.86
50	20	3	0.03	30	300	6	32	0.5	78	0.79	0.78	0.78	0.86

ตารางที่ 4-6 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
51	30	1	0.03	5	800	6	75	1.5	78	0.78	0.78	0.78	0.86
52	20	3	0.02	3	800	6	32	0.5	80	0.80	0.80	0.80	0.87
53	10	1	0.01	10	300	6	32	1.5	76	0.78	0.76	0.76	0.84
54	20	1	0.03	40	300	4	75	0.5	77	0.77	0.77	0.77	0.85
55	30	1	0.01	30	800	4	75	0.5	76	0.77	0.76	0.76	0.84
56	10	1	0.02	40	500	4	75	1	78	0.79	0.78	0.78	0.86
57	20	1	0.03	50	300	5	75	1.5	77	0.78	0.77	0.77	0.85
58	20	1	0.02	30	500	5	32	1.5	79	0.80	0.79	0.79	0.86
59	10	2	0.01	40	500	5	75	1.5	76	0.77	0.76	0.76	0.84
60	30	1	0.03	20	500	4	32	1	80	0.81	0.80	0.80	0.87
61	20	3	0.02	5	800	4	32	1.5	80	0.81	0.80	0.80	0.87
62	30	3	0.02	20	500	4	75	1	77	0.78	0.77	0.77	0.85
63	20	3	0.03	40	500	4	32	1.5	80	0.81	0.80	0.80	0.87
64	30	1	0.03	5	300	4	32	1.5	78	0.79	0.78	0.78	0.86
65	10	1	0.02	30	800	6	50	1	80	0.80	0.80	0.80	0.87
66	10	1	0.01	5	300	4	32	1.5	75	0.76	0.75	0.75	0.84
67	10	1	0.01	10	500	5	32	1.5	78	0.79	0.78	0.78	0.85
68	20	3	0.01	40	300	5	50	1	74	0.75	0.74	0.74	0.82
69	30	1	0.03	40	500	5	50	0.5	80	0.81	0.80	0.80	0.87
70	10	2	0.03	5	500	4	50	1	80	0.81	0.80	0.80	0.87
71	10	3	0.03	20	800	5	75	1	79	0.80	0.79	0.79	0.87
72	20	3	0.03	50	800	5	32	1	80	0.81	0.80	0.80	0.87
73	30	2	0.01	3	800	4	32	1	78	0.79	0.78	0.78	0.86
74	20	3	0.03	20	500	4	50	1	80	0.81	0.80	0.80	0.87
75	20	2	0.01	5	300	6	75	1	76	0.77	0.76	0.76	0.84

ตารางที่ 4-6 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
76	30	1	0.01	50	800	4	32	1.5	76	0.77	0.76	0.76	0.84
77	20	1	0.03	3	500	5	32	0.5	80	0.81	0.80	0.80	0.87
78	30	1	0.01	30	500	4	50	1.5	75	0.76	0.75	0.75	0.83
79	10	1	0.03	20	500	6	50	1	80	0.81	0.80	0.80	0.87
80	10	3	0.02	50	500	5	75	1.5	79	0.80	0.79	0.79	0.86
81	30	1	0.03	30	500	6	50	0.5	81	0.82	0.81	0.81	0.87
82	20	2	0.03	30	800	6	75	1	80	0.80	0.80	0.80	0.87
83	20	1	0.03	10	800	6	32	0.5	79	0.79	0.79	0.79	0.86
84	20	1	0.01	50	800	4	75	0.5	76	0.77	0.76	0.76	0.84
85	20	2	0.03	50	300	5	50	1	77	0.78	0.77	0.77	0.85
86	20	2	0.03	5	500	6	50	0.5	80	0.81	0.80	0.80	0.87
87	20	2	0.03	10	800	5	32	0.5	79	0.80	0.79	0.79	0.86
88	20	3	0.02	30	500	6	32	1.5	79	0.79	0.79	0.79	0.86
89	30	2	0.01	5	300	5	32	1.5	76	0.78	0.76	0.76	0.84
90	20	3	0.03	50	500	6	50	1.5	80	0.81	0.80	0.80	0.87
91	10	1	0.02	10	800	4	75	0.5	80	0.81	0.80	0.80	0.87
92	10	3	0.03	5	300	6	50	1	79	0.80	0.79	0.79	0.87
93	30	3	0.03	5	300	6	75	1	78	0.78	0.78	0.78	0.86
94	10	2	0.01	40	800	5	50	1.5	77	0.78	0.77	0.77	0.86
95	20	2	0.03	50	800	5	75	1.5	79	0.80	0.79	0.79	0.87
96	10	3	0.02	10	300	4	50	1	77	0.78	0.77	0.77	0.85
97	30	2	0.01	3	300	5	32	1	76	0.77	0.76	0.76	0.84
98	30	3	0.03	40	300	4	32	0.5	77	0.78	0.77	0.77	0.85
99	20	2	0.01	40	500	5	75	0.5	76	0.77	0.76	0.76	0.84
100	10	2	0.03	40	800	6	75	1	79	0.80	0.79	0.79	0.87

จากการทดลองลดจำนวนคุณลักษณะลงเหลือ 10 ปัจจัย พบว่าแบบจำลอง CatBoost ยังคงประสิทธิภาพการทำนายอยู่ในเกณฑ์ดีมาก โดยมีค่าความถูกต้อง (Accuracy) อยู่ที่ร้อยละ 90.29 และมีค่า Area Under Curve (ROC_AUC) สูงถึง 0.9641 และมีค่าความถูกต้องของข้อมูลทดลองอยู่ที่ร้อยละ 82.52 ซึ่งแสดงให้เห็นว่าการใช้ปัจจัยเพียง 10 ด้าน แต่แบบจำลองสามารถแยกแยะความแตกต่างระหว่างกลุ่มผู้ที่จะซื้อหรือเช่าที่พักอาศัยได้มีประสิทธิภาพเพิ่มขึ้นจากแบบจำลองชุดข้อมูลเต็ม (Full Features)

4.3.6 ผลการทดสอบแบบจำลอง CNN ด้วย 10 คุณลักษณะ ผู้วิจัยนำคุณลักษณะ 10 ลำดับแรกมาทดสอบกับโครงข่ายประสาทเทียม CNN โดยมีการปรับจูนโครงสร้าง Layer และพารามิเตอร์ให้เหมาะสมกับชุดข้อมูลใหม่ และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-7 ผลการรันแบบจำลอง CNN ด้วย 10 คุณลักษณะ

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	32	6	TRUE	96	0.2	0.0005	32	1	70	0.72	0.65	0.69	0.75
2	96	3	FALSE	64	0	0.0001	32	6	69	0.70	0.66	0.68	0.76
3	32	1	TRUE	32	0	0.0030	96	3	68	0.69	0.68	0.68	0.76
4	32	6	FALSE	96	0.1	0.0002	32	6	69	0.71	0.67	0.68	0.76
5	128	1	FALSE	64	0.2	0.0003	64	3	68	0.71	0.62	0.66	0.76
6	128	3	TRUE	96	0.2	0.0004	128	1	70	0.71	0.67	0.69	0.76
7	64	3	TRUE	64	0.1	0.0047	128	3	70	0.78	0.59	0.66	0.79
8	64	1	TRUE	96	0.2	0.0009	128	1	68	0.71	0.61	0.66	0.72

ตารางที่ 4-7 (ต่อ)

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
9	96	6	FALSE	64	0.1	0.0017	128	3	71	0.75	0.63	0.68	0.78
10	32	3	FALSE	128	0.2	0.0074	64	1	70	0.71	0.71	0.70	0.79
11	32	3	FALSE	96	0	0.0009	96	6	70	0.72	0.67	0.69	0.78
12	96	3	TRUE	96	0.1	0.0008	128	3	71	0.71	0.72	0.71	0.78
13	64	1	FALSE	128	0.4	0.0001	64	1	68	0.70	0.63	0.66	0.74
14	128	1	TRUE	32	0.3	0.0076	96	3	68	0.69	0.69	0.68	0.76
15	32	3	TRUE	96	0.3	0.0001	32	1	65	0.69	0.61	0.63	0.72
16	64	1	FALSE	96	0.3	0.0095	96	1	69	0.70	0.67	0.68	0.77
17	128	1	TRUE	128	0.2	0.0001	32	1	65	0.66	0.63	0.64	0.71
18	64	3	FALSE	96	0	0.0001	64	3	69	0.71	0.66	0.68	0.76
19	32	6	TRUE	64	0.3	0.0007	128	1	70	0.72	0.67	0.69	0.76
20	64	1	FALSE	128	0.4	0.0008	128	6	69	0.71	0.65	0.68	0.76
21	32	1	TRUE	128	0.2	0.0001	128	3	67	0.69	0.61	0.64	0.73
22	64	6	FALSE	128	0.3	0.0002	128	3	69	0.71	0.65	0.68	0.76
23	32	6	TRUE	32	0.3	0.0003	64	1	68	0.69	0.64	0.67	0.73

ตารางที่ 4-7 (ต่อ)

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
24	64	1	TRUE	64	0.1	0.0003	96	3	69	0.72	0.60	0.65	0.75
25	64	3	FALSE	64	0.3	0.0001	64	3	69	0.70	0.65	0.67	0.76
26	64	1	TRUE	96	0.4	0.0001	128	1	66	0.67	0.63	0.65	0.71
27	32	1	FALSE	32	0.2	0.0023	96	1	70	0.71	0.68	0.69	0.77
28	128	1	FALSE	96	0	0.0005	128	6	71	0.73	0.65	0.69	0.76

จากการทดสอบแบบจำลอง CNN เมื่อลดจำนวนคุณลักษณะเหลือ 10 ปัจจัย พบว่าประสิทธิภาพความแม่นยำลดลงมาอยู่ที่ร้อยละ 73.01 โดยค่า ROC_AUC อยู่ที่ 0.8004 และมีค่าความถูกต้องของข้อมูลทดลองอยู่ที่ร้อยละ 73.01 ซึ่งแสดงให้เห็นว่าโครงข่ายประสาทเทียมยังสามารถจำแนกข้อมูลได้ในระดับปานกลางค่อนข้างดี อย่างไรก็ตาม เมื่อเปรียบเทียบกับ CatBoost ในสัดส่วนปัจจัยที่เท่ากัน พบว่า CNN มีความอ่อนไหวต่อการลดจำนวนตัวแปรมากกว่า ส่งผลให้ความสามารถในการเรียนรู้รูปแบบข้อมูลลดลงอย่างชัดเจนจากตัวโมเดลเต็ม (Full Features)

4.3.7 ผลการทดสอบแบบจำลอง Multinomial Naive Bayes ด้วย 10 คุณลักษณะ ผู้วิจัยรันโมเดล Multinomial Naive Bayes บนชุดข้อมูล 10 ปัจจัย และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-8 ผลการรันแบบจำลอง Multinomial Naive Bayes ด้วย 10 คุณลักษณะ

ลำดับที่	alpha	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	0.0001	69	0.69	0.69	0.69	0.76
2	0.001	69	0.69	0.69	0.69	0.76
3	0.01	69	0.69	0.69	0.69	0.76
4	0.1	69	0.69	0.69	0.69	0.76
5	0.5	69	0.69	0.69	0.69	0.76
6	1	69	0.69	0.69	0.69	0.76
7	5	69	0.69	0.69	0.69	0.76
8	10	69	0.69	0.69	0.69	0.76
9	50	69	0.70	0.68	0.68	0.76
10	100	69	0.70	0.67	0.68	0.76

จากการทดลองลดจำนวนคุณลักษณะเหลือ 10 ปัจจัย พบว่าแบบจำลอง Multinomial Naive Bayes ให้ประสิทธิภาพ เท่าเดิมกับเมื่อใช้ชุดข้อมูลเต็ม (Full Features) โดยไม่มีการเปลี่ยนแปลงของค่าสถิติใด ๆ สิ่งนี้สะท้อนให้เห็นว่าปัจจัยอีก 18 ปัจจัยที่ถูกตัดออกไปนั้นไม่มีผลต่อกระบวนการคำนวณความน่าจะเป็นของโมเดลนี้เลย และยืนยันได้ว่าปัจจัย 10 ลำดับแรกที่ RFE คัดเลือกมาคือข้อมูลส่วนสำคัญ (Core Information) ที่โมเดลใช้ในการตัดสินใจทั้งหมด

4.3.8 ผลการทดสอบแบบจำลอง CatBoost ด้วย 7 คุณลักษณะ ผู้วิจัยลดจำนวนปัจจัยลงเหลือเพียง 7 ปัจจัยหลัก ซึ่งประกอบด้วยปัจจัยพื้นฐานทางเศรษฐกิจและประชากรศาสตร์ที่สำคัญที่สุด และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-9 ผลการรันแบบจำลอง Catboost ด้วย 7 คุณลักษณะ

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	20	3	0.01	20	800	4	50	1	78	0.79	0.78	0.78	0.84
2	30	2	0.03	20	500	5	32	0.5	80	0.81	0.80	0.79	0.85
3	30	3	0.02	20	500	4	32	1	77	0.78	0.77	0.77	0.85
4	30	3	0.03	50	500	6	50	1.5	78	0.79	0.78	0.78	0.86
5	20	3	0.01	20	300	4	32	1	75	0.76	0.75	0.75	0.82
6	10	2	0.01	10	300	4	50	1.5	77	0.78	0.77	0.77	0.83
7	20	3	0.03	30	800	5	32	1.5	79	0.80	0.79	0.79	0.85
8	20	1	0.02	10	300	5	32	1	78	0.79	0.78	0.78	0.84
9	20	1	0.01	5	300	6	32	1	77	0.78	0.77	0.77	0.83
10	20	3	0.01	20	800	4	32	0.5	77	0.78	0.77	0.77	0.84
11	30	3	0.03	20	800	6	50	1.5	78	0.79	0.78	0.78	0.84
12	20	3	0.03	10	800	5	75	0.5	79	0.79	0.79	0.79	0.84
13	20	3	0.02	20	800	4	32	1	80	0.81	0.80	0.80	0.86
14	30	2	0.01	10	500	5	75	1.5	77	0.78	0.77	0.77	0.83
15	30	3	0.03	20	500	6	32	1.5	79	0.80	0.79	0.79	0.85
16	30	2	0.03	3	300	6	32	1.5	79	0.80	0.79	0.79	0.85
17	30	2	0.02	40	500	5	32	1.5	77	0.78	0.77	0.77	0.84
18	10	3	0.01	3	800	6	50	1	79	0.80	0.79	0.79	0.85
19	30	1	0.02	50	800	6	32	0.5	79	0.80	0.79	0.79	0.85
20	20	2	0.02	3	500	5	32	0.5	79	0.80	0.79	0.79	0.85
21	30	3	0.03	50	500	4	32	1.5	80	0.81	0.80	0.80	0.86
22	10	3	0.01	3	500	4	75	1.5	79	0.80	0.79	0.79	0.84
23	20	1	0.02	40	300	4	50	1	76	0.77	0.76	0.76	0.83
24	10	2	0.01	50	300	5	50	0.5	76	0.77	0.76	0.76	0.82
25	20	1	0.03	3	500	4	32	1	80	0.81	0.80	0.80	0.86

ตารางที่ 4-9 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
26	30	1	0.01	40	800	6	75	0.5	77	0.78	0.77	0.77	0.83
27	10	2	0.03	50	300	6	32	1	78	0.79	0.78	0.78	0.84
28	30	2	0.03	5	300	6	32	0.5	78	0.79	0.78	0.78	0.84
29	20	1	0.03	3	300	6	75	0.5	79	0.80	0.79	0.79	0.85
30	20	1	0.02	50	800	6	50	0.5	79	0.80	0.79	0.79	0.85
31	10	3	0.03	3	800	4	50	1.5	79	0.79	0.79	0.79	0.84
32	10	3	0.03	40	300	6	32	1	78	0.79	0.78	0.78	0.85
33	10	2	0.03	50	800	5	50	1.5	79	0.80	0.79	0.79	0.85
34	30	3	0.03	30	500	5	32	1.5	79	0.80	0.79	0.79	0.85
35	30	3	0.03	40	300	4	75	1	76	0.77	0.76	0.76	0.83
36	20	2	0.02	3	500	5	75	1	78	0.79	0.78	0.78	0.85
37	30	1	0.03	20	500	5	50	0.5	79	0.80	0.79	0.79	0.85
38	20	1	0.03	20	800	5	50	1	79	0.80	0.79	0.79	0.85
39	20	2	0.03	40	500	4	50	0.5	79	0.80	0.79	0.79	0.86
40	30	1	0.03	20	800	5	50	0.5	79	0.80	0.79	0.79	0.85
41	30	2	0.02	20	300	5	75	1	76	0.77	0.76	0.76	0.83
42	30	3	0.01	50	500	6	75	1	75	0.76	0.75	0.75	0.82
43	10	1	0.01	3	500	4	32	0.5	78	0.79	0.78	0.78	0.84
44	30	1	0.02	50	800	6	75	1.5	79	0.80	0.79	0.79	0.85
45	20	3	0.01	20	800	5	50	0.5	78	0.79	0.78	0.78	0.84
46	30	3	0.01	50	300	6	50	0.5	74	0.75	0.74	0.74	0.81
47	30	2	0.02	30	300	6	75	0.5	76	0.77	0.76	0.76	0.83
48	10	1	0.02	3	500	4	50	0.5	79	0.80	0.79	0.79	0.86
49	10	2	0.02	50	500	5	50	1	78	0.79	0.78	0.78	0.85
50	20	3	0.03	30	300	6	32	0.5	78	0.79	0.78	0.78	0.84

ตารางที่ 4-9 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
51	30	1	0.03	5	800	6	75	1.5	77	0.78	0.77	0.77	0.83
52	20	3	0.02	3	800	6	32	0.5	78	0.78	0.78	0.78	0.84
53	10	1	0.01	10	300	6	32	1.5	77	0.78	0.77	0.77	0.83
54	20	1	0.03	40	300	4	75	0.5	77	0.78	0.77	0.77	0.84
55	30	1	0.01	30	800	4	75	0.5	76	0.77	0.76	0.76	0.83
56	10	1	0.02	40	500	4	75	1	78	0.79	0.78	0.78	0.85
57	20	1	0.03	50	300	5	75	1.5	77	0.78	0.77	0.77	0.84
58	20	1	0.02	30	500	5	32	1.5	78	0.79	0.78	0.78	0.85
59	10	2	0.01	40	500	5	75	1.5	76	0.77	0.76	0.76	0.83
60	30	1	0.03	20	500	4	32	1	80	0.81	0.80	0.80	0.86
61	20	3	0.02	5	800	4	32	1.5	80	0.81	0.80	0.80	0.86
62	30	3	0.02	20	500	4	75	1	78	0.79	0.78	0.78	0.84
63	20	3	0.03	40	500	4	32	1.5	79	0.80	0.79	0.79	0.86
64	30	1	0.03	5	300	4	32	1.5	78	0.79	0.78	0.78	0.85
65	10	1	0.02	30	800	6	50	1	79	0.80	0.79	0.79	0.85
66	10	1	0.01	5	300	4	32	1.5	77	0.78	0.77	0.77	0.83
67	10	1	0.01	10	500	5	32	1.5	78	0.79	0.78	0.78	0.84
68	20	3	0.01	40	300	5	50	1	75	0.76	0.75	0.75	0.82
69	30	1	0.03	40	500	5	50	0.5	79	0.80	0.79	0.79	0.86
70	10	2	0.03	5	500	4	50	1	79	0.80	0.79	0.79	0.86
71	10	3	0.03	20	800	5	75	1	79	0.79	0.79	0.79	0.85
72	20	3	0.03	50	800	5	32	1	80	0.81	0.80	0.80	0.85
73	30	2	0.01	3	800	4	32	1	79	0.80	0.79	0.79	0.85
74	20	3	0.03	20	500	4	50	1	80	0.81	0.80	0.80	0.86
75	20	2	0.01	5	300	6	75	1	77	0.78	0.77	0.77	0.83

ตารางที่ 4-9 (ต่อ)

ลำดับที่	random_strength	max_ctr_complexity	learning_rate	l2_leaf_reg	iterations	depth	border_count	bagging_temperature	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
76	30	1	0.01	50	800	4	32	1.5	76	0.77	0.76	0.76	0.83
77	20	1	0.03	3	500	5	32	0.5	79	0.80	0.79	0.79	0.85
78	30	1	0.01	30	500	4	50	1.5	76	0.77	0.76	0.76	0.82
79	10	1	0.03	20	500	6	50	1	79	0.79	0.79	0.79	0.85
80	10	3	0.02	50	500	5	75	1.5	78	0.79	0.78	0.78	0.85
81	30	1	0.03	30	500	6	50	0.5	79	0.80	0.79	0.79	0.85
82	20	2	0.03	30	800	6	75	1	79	0.79	0.79	0.79	0.85
83	20	1	0.03	10	800	6	32	0.5	78	0.78	0.78	0.78	0.83
84	20	1	0.01	50	800	4	75	0.5	76	0.77	0.76	0.76	0.83
85	20	2	0.03	50	300	5	50	1	78	0.79	0.78	0.78	0.84
86	20	2	0.03	5	500	6	50	0.5	79	0.79	0.79	0.78	0.84
87	20	2	0.03	10	800	5	32	0.5	79	0.79	0.79	0.79	0.84
88	20	3	0.02	30	500	6	32	1.5	78	0.79	0.78	0.78	0.84
89	30	2	0.01	5	300	5	32	1.5	77	0.78	0.77	0.77	0.83
90	20	3	0.03	50	500	6	50	1.5	79	0.80	0.79	0.78	0.85
91	10	1	0.02	10	800	4	75	0.5	79	0.81	0.79	0.79	0.86
92	10	3	0.03	5	300	6	50	1	78	0.79	0.78	0.78	0.85
93	30	3	0.03	5	300	6	75	1	78	0.79	0.78	0.77	0.85
94	10	2	0.01	40	800	5	50	1.5	78	0.79	0.78	0.78	0.84
95	20	2	0.03	50	800	5	75	1.5	79	0.80	0.79	0.79	0.85
96	10	3	0.02	10	300	4	50	1	78	0.79	0.78	0.78	0.84
97	30	2	0.01	3	300	5	32	1	78	0.79	0.78	0.78	0.83
98	30	3	0.03	40	300	4	32	0.5	77	0.78	0.77	0.77	0.83
99	20	2	0.01	40	500	5	75	0.5	76	0.77	0.76	0.76	0.83
100	10	2	0.03	40	800	6	75	1	79	0.80	0.79	0.79	0.85

จากการทดลองคัดเลือกปัจจัยที่สำคัญที่สุดเหลือเพียง 7 คุณลักษณะ พบว่าแบบจำลอง CatBoost ให้ค่าความถูกต้อง อยู่ที่ร้อยละ 87.02 และค่า F1-Score ที่ร้อยละ 86.97 แม้ว่าตัวเลข

ประสิทธิภาพจะลดลงจากชุดข้อมูลเต็มตามสัดส่วนของตัวแปรที่หายไป แต่เมื่อพิจารณาค่า ROC_AUC ที่ยังคงอยู่ในระดับสูงถึง 0.9369 แสดงให้เห็นว่าแบบจำลองยังคงมีความสามารถในการแยกแยะพฤติกรรมการตัดสินใจได้อย่างแม่นยำและมีความเสถียรสูง

ดังนั้น เนื่องจากผลลัพธ์ที่ได้ยังอยู่ในเกณฑ์ที่น่าพึงพอใจและมีความคุ้มค่าในแง่ของการลดภาระการกรอกข้อมูลของผู้ใช้งาน ผู้วิจัยจึงตัดสินใจเลือกใช้แบบจำลอง CatBoost ร่วมกับปัจจัยสำคัญทั้ง 7 ประการนี้ เป็นโครงสร้างหลักในการพัฒนาระบบสนับสนุนการตัดสินใจต่อไป

4.3.9 ผลการทดสอบแบบจำลอง CNN ด้วย 7 คุณลักษณะ ผู้วิจัยรันโมเดล CNN ด้วยชุดข้อมูล 7 ปัจจัยสำคัญ และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-10 ผลการรันแบบจำลอง CNN ด้วย 7 คุณลักษณะ

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	128	6	FALSE	64	0	0.0006	-	-	70	0.73	0.63	0.67	0.77
2	32	6	FALSE	32	0.2	0.0099	-	-	70	0.72	0.64	0.68	0.77
3	32	1	TRUE	128	0.2	0.0002	32	1	62	0.64	0.57	0.60	0.67
4	96	6	FALSE	96	0.4	0.0002	32	1	69	0.71	0.64	0.67	0.76
5	96	6	FALSE	64	0.3	0.0010	96	3	70	0.73	0.64	0.68	0.77
6	32	1	TRUE	64	0	0.0005	32	1	62	0.64	0.55	0.59	0.67
7	128	3	FALSE	96	0	0.0052	32	6	68	0.74	0.58	0.64	0.75

ตารางที่ 4-10 (ต่อ)

ลำดับที่	conv_1_filters	kernel_size	add_conv_2	dense_1_units	dropout_1	learning_rate	conv_2_filters	kernel2_kernel_size	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
8	32	3	FALSE	64	0.3	0.0002	64	1	67	0.70	0.60	0.65	0.72
9	128	1	FALSE	96	0	0.0089	64	6	67	0.70	0.60	0.64	0.74
10	96	1	FALSE	96	0.3	0.0001	128	1	68	0.70	0.62	0.66	0.72
11	128	1	FALSE	32	0.1	0.0007	64	3	67	0.70	0.62	0.65	0.73
12	64	6	FALSE	96	0.1	0.0018	128	6	69	0.73	0.60	0.66	0.77
13	32	3	FALSE	32	0.4	0.0023	96	6	69	0.69	0.67	0.68	0.74
14	64	3	FALSE	64	0.3	0.0002	128	3	67	0.70	0.60	0.65	0.72
15	64	3	FALSE	64	0.4	0.0004	64	1	68	0.70	0.64	0.67	0.73
16	96	1	FALSE	64	0	0.0067	128	6	67	0.68	0.64	0.65	0.74
17	32	1	FALSE	128	0.1	0.0005	32	6	67	0.69	0.63	0.66	0.73
18	32	6	FALSE	32	0.2	0.0023	128	3	69	0.71	0.64	0.67	0.76
19	64	3	TRUE	128	0.2	0.0070	96	1	67	0.71	0.61	0.64	0.73
20	32	6	FALSE	64	0.1	0.0084	64	3	70	0.74	0.62	0.67	0.78
21	128	3	TRUE	64	0	0.0014	64	1	65	0.67	0.65	0.65	0.73

จากการลดจำนวนปัจจัยลงเหลือเพียง 7 ปัจจัยหลัก พบว่าประสิทธิภาพของ CNN ยังคงใกล้เคียงกับกรณี 10 ปัจจัย โดยมีค่า Accuracy อยู่ที่ร้อยละ 72.37 และค่า ROC_AUC อยู่ที่ 0.8019 ซึ่งแสดงให้เห็นว่าโมเดลเริ่มถึงจุดอิ่มตัวในการเรียนรู้จากปัจจัยที่จำกัด แม้ผลลัพธ์จะอยู่ในเกณฑ์ที่

ยอมรับได้สำหรับการวิจัยเชิงวิชาการ แต่เมื่อเปรียบเทียบประสิทธิภาพกับ CatBoost ที่ใช้ปัจจัยเท่ากัน (87.02%) พบว่า CNN ยังมีประสิทธิภาพด้อยกว่าในบริบทของข้อมูลชุดนี้

4.3.10 ผลการทดสอบแบบจำลอง Multinomial Naive Bayes ด้วย 7 คุณลักษณะ ผู้วิจัยรันโมเดล Multinomial Naive Bayes ด้วยชุดข้อมูล 7 ปัจจัยสำคัญ และมีผลการรันเพื่อหา Best Parameter ดังนี้

ตารางที่ 4-11 ผลการรันแบบจำลอง Multinomial Naive Bayes ด้วย 7 คุณลักษณะ

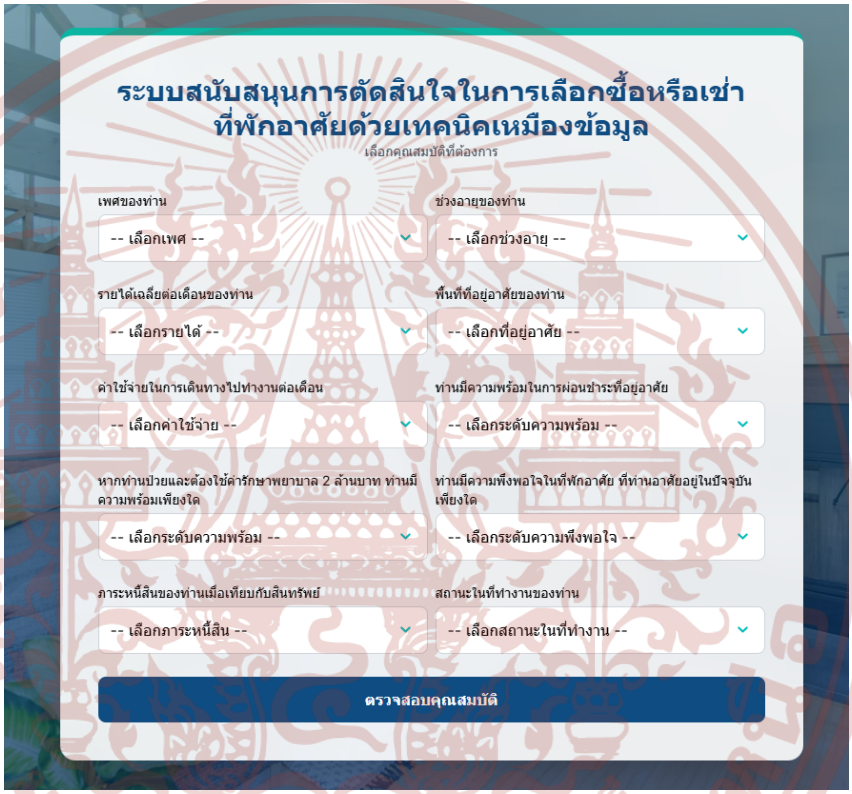
ลำดับที่	alpha	accuracy (%)	precision	recall	F-Measure	ROC (AUC)
1	0.0001	69	0.69	0.69	0.69	0.76
2	0.001	69	0.69	0.69	0.69	0.76
3	0.01	69	0.69	0.69	0.69	0.76
4	0.1	69	0.69	0.69	0.69	0.76
5	0.5	69	0.69	0.69	0.69	0.76
6	1	69	0.69	0.69	0.69	0.76
7	5	69	0.69	0.69	0.69	0.76
8	10	69	0.69	0.69	0.69	0.76
9	50	69	0.70	0.68	0.68	0.76
10	100	69	0.70	0.67	0.68	0.76

ในกรณีที่ลดปัจจัยเหลือเพียง 7 ปัจจัยหลัก ผลการทดสอบยังคงแสดงค่าประสิทธิภาพที่คงที่อยู่ที่ร้อยละ 68.95 ซึ่งพิสูจน์ให้เห็นว่าความสามารถในการทำนายของ Naive Bayes ในงานวิจัยนี้ถูกจำกัดอยู่ที่ระดับนี้ด้วยหลักการทางสถิติ แม้จะให้ผลลัพธ์ที่ด้อยกว่า CatBoost (87.02%) และ CNN (72.37%) อย่างชัดเจน แต่ความคงที่ของโมเดลก็ช่วยยืนยันความสำคัญของปัจจัย 7 ประการที่คัดเลือกมา ว่าเป็นชุดตัวแปรที่ทรงประสิทธิภาพและเพียงพอต่อการพยากรณ์พื้นฐาน

4.4 ผลการสร้างและประเมินประสิทธิภาพแบบจำลองการพยากรณ์

ในหัวข้อนี้ ผู้วิจัยได้นำแบบจำลอง CatBoost ร่วมกับ 7 คุณลักษณะสำคัญ ที่ให้ประสิทธิภาพดีที่สุดและมีความกระชับ มาพัฒนาเป็นเว็บแอปพลิเคชัน เพื่อให้บุคคลทั่วไปสามารถเข้ามาใช้งานเพื่อประเมินแนวโน้มการเลือกที่อยู่อาศัยของตนเองได้ โดยแบ่งการทำงานเป็น 2 ส่วนหลัก

4.4.1 ส่วนการรับข้อมูล (Input Interface) ผู้วิจัยออกแบบหน้าจอการรับข้อมูลโดยเน้นความเรียบง่ายและเป็นมิตรต่อผู้ใช้งาน (User-Friendly Design) โดยดึงปัจจัยสำคัญ 7 ประการจากผลการวิเคราะห์ RFE มาสร้างเป็นฟอร์มรับข้อมูล เพื่อให้ผู้ใช้กรอกข้อมูลน้อยที่สุดแต่ยังคงได้รับผลลัพธ์ที่แม่นยำแสดงดังภาพที่ 4-1



ภาพที่ 4-1 หน้าจอรับข้อมูลของระบบ

4.4.2 การแสดงผลการพยากรณ์และการอธิบายเหตุผลด้วย DiCE จุดเด่นของระบบที่ผู้วิจัยพัฒนาขึ้นคือการบูรณาการเทคนิค DiCE เพื่อสร้างความโปร่งใสให้กับแบบจำลอง (XAI) โดยเฉพาะในกรณีที่ผู้ใช้ยังไม่พร้อมสำหรับการซื้อ ระบบจะคำนวณและแสดง "คำแนะนำเพื่อเพิ่มโอกาสในการซื้อ" ซึ่งเป็นการจำลองสถานการณ์ว่าหากผู้ใช้ต้องการเปลี่ยนผลลัพธ์จากเช่าเป็นซื้อจะต้องปรับเปลี่ยนปัจจัยใดบ้างแสดงดังภาพที่ 4-2



ภาพที่ 4-2 หน้าจอผลลัพธ์ของระบบ

4.5 ผลการประเมินความพึงพอใจต่อการใช้งานระบบสนับสนุนการตัดสินใจ

เพื่อให้ทราบถึงประสิทธิภาพของระบบในมุมมองของผู้ใช้งาน ผู้วิจัยได้นำระบบที่พัฒนาขึ้นไปทำการทดสอบกับกลุ่มตัวอย่างจำนวน 30 คน และเก็บรวบรวมข้อมูลผ่านแบบสอบถามความพึงพอใจ โดยแบ่งเกณฑ์การประเมินออกเป็น 7 ข้อคำถามสำคัญ ตามมาตรวัดของ Likert Scale (5 ระดับ) ซึ่งผลการวิเคราะห์ข้อมูลสรุปได้ดังนี้

ตารางที่ 4-12 ผลการเก็บข้อมูลความพึงพอใจระบบ

ข้อคำถาม	คะแนนเฉลี่ย	SD
ท่านมีความเห็นว่าผลการทำนายที่ได้รับมีความสอดคล้องกับสถานะความเป็นจริงของท่านเพียงใด?	ดี	0.83
ท่านมีความเชื่อมั่นในผลการวิเคราะห์ที่ได้รับจากระบบปัญญาประดิษฐ์นี้มากน้อยเพียงใด?	ดี	0.74
คำแนะนำในการปรับเปลี่ยนข้อมูล (เช่น การเพิ่ม/ลดปัจจัยต่างๆ) มีความสมเหตุสมผลเพียงใด?	ดี	0.58
ท่านคิดว่าแนวทางที่ระบบแนะนำ (Counterfactuals) สามารถนำไปปรับใช้ได้จริงในสถานการณ์ปัจจุบันของท่าน?	ดี	0.88
การจัดวางรูปแบบข้อมูลและผลลัพธ์บนหน้าจอ เข้าใจได้ง่ายและไม่ซับซ้อน?	ดีมาก	0.68
ความรวดเร็วในการประมวลผลและการแสดงผลของระบบมีความเหมาะสม?	ดี	0.79
โดยรวมแล้ว ระบบนี้มีประโยชน์ต่อท่านในการใช้วางแผนหรือตัดสินใจในอนาคต?	ดี	0.92

พบว่าผู้ตอบแบบสอบถามมีความพึงพอใจต่อระบบในระดับ ดี ในทุกรายการ โดยครอบคลุมทั้งด้านความสอดคล้องของผลทำนาย ความเชื่อมั่นในระบบปัญญาประดิษฐ์ ความสมเหตุสมผลของคำแนะนำรวมถึงด้านการออกแบบส่วนติดต่อประสานงาน (UI) และความเร็วในการประมวลผล แสดงให้เห็นว่าระบบมีประสิทธิภาพและมีประโยชน์ต่อการตัดสินใจในอนาคตของผู้ใช้งานอย่างแท้จริง

บทที่5
สรุปผล อภิปรายผล และข้อเสนอแนะ

การศึกษาวิจัยเรื่อง "การพัฒนาระบบสนับสนุนการตัดสินใจเลือกที่พักอาศัยโดยใช้เทคนิคการเรียนรู้ของเครื่อง" สามารถสรุปผลการดำเนินงาน อภิปรายผล และให้ข้อเสนอแนะได้ดังนี้

5.1 สรุปผลการดำเนินงาน

5.2 อภิปรายผล

5.3 ข้อเสนอแนะ

5.1 สรุปผลการดำเนินงาน

การศึกษาวิจัยเรื่อง "การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล" สามารถสรุปผลการศึกษาตามวัตถุประสงค์ของการวิจัยได้ดังนี้

5.1.1 ข้อมูลทั่วไปของกลุ่มตัวอย่าง จากการเก็บรวบรวมข้อมูลผ่านช่องทางออนไลน์จากกลุ่มตัวอย่างทั้งหมด 1,107 ราย ผู้วิจัยพบว่าข้อมูลมีความหลากหลายครอบคลุมพฤติกรรมการเลือกที่พักอาศัยทั้งในเขตกรุงเทพมหานครและปริมณฑล โดยในกระบวนการเตรียมข้อมูล (Data Pre-processing) ผู้วิจัยได้ดำเนินการทำความสะอาดข้อมูล (Data Cleaning) และคัดเลือกตัวแปรที่มีความสำคัญ (Feature Selection) จนคงเหลือปัจจัยหลักที่ส่งผลต่อการตัดสินใจจำนวน 28 ข้อคำถาม (ตัวแปรต้น) และ 1 ตัวแปรตาม คือพฤติกรรมการเลือกซื้อหรือเช่าที่พักอาศัย

5.1.2 ผลการพัฒนาแบบจำลองสนับสนุนการตัดสินใจ ผู้วิจัยได้ทำการทดลองเปรียบเทียบอัลกอริทึมในกลุ่ม Supervised Learning หลายชนิด ผลการศึกษาพบว่า

5.1.1.1 แบบจำลอง CatBoost ให้ประสิทธิภาพสูงสุดในการแยกแยะพฤติกรรมระหว่างกลุ่มผู้เช่าและผู้ซื้อ โดยมีค่าความแม่นยำ (Accuracy) และค่า F1-Score อยู่ในระดับที่น่าพอใจ

5.1.1.2 การทดสอบด้วย Unseen Data: เมื่อนำแบบจำลองไปทดสอบกับชุดข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Evaluation) แบบจำลองยังคงรักษาประสิทธิภาพการทำนายได้ดีเยี่ยม แสดงถึงความสามารถในการนำไปประยุกต์ใช้จริง (Generalization Ability) โดยมีพารามิเตอร์ที่ผ่านการปรับจูน (Hyperparameter Tuning) ผ่านวิธี RandomizedSearchCV อย่างเหมาะสม

5.1.3 การพัฒนาและทดสอบระบบสนับสนุนการตัดสินใจ ผู้วิจัยได้นำแบบจำลองที่ได้ไปพัฒนาต่อยอดเป็นระบบสารสนเทศในรูปแบบเว็บแอปพลิเคชัน โดยมีจุดเด่นดังนี้

5.1.1.3 การวิเคราะห์ปัจจัยด้วย DiCE (Diverse Counterfactual Explanations): ระบบไม่ได้แจ้งเพียงผลทำนาย แต่สามารถให้คำแนะนำเชิงรุกแก่ผู้ใช้งานได้ว่า หากต้องการเปลี่ยนผลลัพธ์จากการเช่าเป็นการซื้อ (หรือในทางกลับกัน) ผู้ใช้งานควรปรับเปลี่ยนปัจจัยด้านใดบ้าง (เช่น รายได้ หรือเงินสำรองฉุกเฉิน)

5.1.1.4 การเชื่อมต่อผ่าน API: ระบบมีการเชื่อมต่อระหว่างส่วนหน้าจอและโมเดล CatBoost ผ่านภาษา Python API ทำให้การประมวลผลและการแนะนำผลลัพธ์ (Counterfactuals) เป็นไปอย่างรวดเร็ว

5.1.2 ผลประเมินความพึงพอใจและความสมเหตุสมผล จากการประเมินโดยกลุ่มผู้ใช้งาน พบว่าผู้ใช้งานมีความเห็นพ้องในระดับ ดี ต่อระบบในทุกประเด็นสำคัญ

5.1.2.1 ความสอดคล้อง (Relevance): ผลการทำนายสอดคล้องกับสถานะความเป็นจริงของผู้ใช้งาน

5.1.2.2 ความเชื่อมั่น (Trustworthiness): ผู้ใช้งานมีความเชื่อมั่นในผลวิเคราะห์ที่ได้รับจากปัญญาประดิษฐ์

5.1.2.3 การนำไปใช้ประโยชน์ (Practicality): คำแนะนำจากเทคนิค DiCE มีความสมเหตุสมผลและสามารถนำไปเป็นแนวทางในการวางแผนการเงินเพื่อการมีที่อยู่อาศัยในอนาคตได้จริง

5.2 อภิปรายผล

จากการดำเนินงานวิจัย มีประเด็นสำคัญที่สามารถนำมาอภิปรายผลได้ดังนี้

5.2.1 ความเหมาะสมของแบบจำลอง: การที่ CatBoost ให้ผลลัพธ์ดีที่สุด อธิบายได้ว่าเนื่องจากข้อมูลที่ใช้ในการตัดสินใจเลือกที่พักอาศัยส่วนใหญ่เป็นข้อมูลประเภทหมวดหมู่ (Categorical Data) ซึ่ง CatBoost มีจุดเด่นในการจัดการข้อมูลลักษณะนี้ได้ดีกว่า CNN ที่มักต้องการข้อมูลจำนวนมหาศาลและการปรับจูนโครงสร้างที่ซับซ้อน

5.2.2 อิทธิพลของปัจจัยเศรษฐศาสตร์: ผลการวิจัยยืนยันว่า "รายได้" และ "ภาระหนี้สิน" เป็นตัวแปรต้นที่สำคัญที่สุด ซึ่งสอดคล้องกับสภาพเศรษฐกิจปัจจุบันที่ความสามารถในการผ่อนชำระ (Affordability) เป็นปัจจัยชี้ขาดหลักในการเลือกซื้อที่อยู่อาศัย

5.2.3 การเพิ่มความเชื่อมั่นด้วย XAI: ประเด็นสำคัญที่ทำให้งานวิจัยนี้แตกต่างจากงานส่วนใหญ่ คือการเปลี่ยนจากระบบที่ยากจะเข้าใจ (Black-box) ให้กลายเป็นเครื่องมือที่โปร่งใสผ่านเทคนิค XAI (Explainable AI) โดยในบทที่ 1 ได้ระบุถึงความซับซ้อนของการตัดสินใจเลือกที่อยู่อาศัย ซึ่งงานในอดีต เช่น พูลสวัสดิ์ และคณะ [1] มักใช้สถิติดั้งเดิมและจำกัดอยู่เพียงตัวแปรด้านอายุหรือรายได้ ซึ่งอาจไม่เพียงพอต่อการพยากรณ์ในสภาวะเศรษฐกิจที่ผันผวน ผู้วิจัยจึงได้นำอัลกอริทึม CatBoost มา

ใช้ร่วมกับตัวแปรที่ครอบคลุมถึง 28 ข้อคำถาม ทำให้ระบบสามารถวิเคราะห์พฤติกรรมได้ละเอียดและแม่นยำกว่าแนวทางเดิม อีกทั้งการนำเทคนิค DiCE มาใช้ยังช่วยเปลี่ยนภาพลักษณ์ของ AI ให้มีความโปร่งใสและให้คำแนะนำที่จับต้องได้ผ่านการจำลองเงื่อนไขที่ต้องปรับเปลี่ยน (Counterfactuals) ซึ่งแตกต่างจากระบบทั่วไปที่แจ้งเพียงผลทำนายโดยไม่บอกเหตุผล ความโปร่งใสนี้ส่งผลโดยตรงต่อระดับความพึงพอใจของผู้ใช้งานในบทที่ 4 ที่อยู่ในระดับ มากที่สุด

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะในการนำไปใช้งาน

1.1.1.1. สำหรับภาคธุรกิจ: ผู้ประกอบการอสังหาริมทรัพย์สามารถใช้ระบบนี้ในการวิเคราะห์ความต้องการของลูกค้าเบื้องต้น เพื่อนำเสนอแพ็กเกจที่อยู่อาศัยหรือแผนการเงินที่ตรงใจลูกค้ามากขึ้น

1.1.1.2. สำหรับภาคประชาชน: สามารถใช้เป็นเครื่องมือประเมินความพร้อมและวางแผนเตรียมตัวก่อนตัดสินใจทำสัญญาซื้อขายจริง เพื่อลดความเสี่ยงทางการเงินในระยะยาว

5.3.1 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

1.1.1.1. การเพิ่มความหลากหลายของข้อมูล เนื่องจากหัวข้อการเลือกที่พักอาศัยเป็นเรื่องที่เกี่ยวข้องกับบุคคลทุกกลุ่มในสังคม ซึ่งมีพื้นฐานชีวิตที่แตกต่างกันมาก การวิจัยในอนาคตควรขยายการเก็บข้อมูลให้ครอบคลุมกลุ่มประชากรที่หลากหลายมากขึ้น

1.1.1.2. ข้อจำกัดด้านทรัพยากร เนื่องด้วยระยะเวลาและงบประมาณที่จำกัดในการวิจัยครั้งนี้ ทำให้การเก็บรวบรวมข้อมูลให้มีความสมดุล (Balance) ในทุกมิติเป็นเรื่องที่ทำได้ยาก ส่งผลให้ข้อมูลบางกลุ่มอาจมีปริมาณน้อยกว่ากลุ่มอื่น การศึกษาครั้งต่อไปจึงควรเพิ่มจำนวนกลุ่มตัวอย่างให้มากขึ้น และใช้ระยะเวลาที่นานขึ้นเพื่อให้ได้ฐานข้อมูลที่สมบูรณ์และลดความลำเอียง (Bias) ของแบบจำลอง

1.1.1.3. การเชื่อมโยงข้อมูลภายนอก ควรทดลองนำปัจจัยภายนอก เช่น อัตราดอกเบี้ยนโยบายในขณะนั้น หรือราคาประเมินที่ดินรายปี เข้ามาเป็นส่วนหนึ่งของปัจจัยนำเข้า เพื่อให้แบบจำลองมีความทันสมัยสอดคล้องกับสภาวะเศรษฐกิจที่เปลี่ยนแปลงตลอดเวลา

บรรณานุกรม

กิตติพร พูลสวัสดิ์ และคณะ. (2568). “การศึกษาการตัดสินใจเลือกซื้อที่อยู่อาศัย ในเขตกรุงเทพมหานคร และปริมณฑล”. ใน รายงานสืบเนื่องจากการประชุมวิชาการระดับชาติและนานาชาติ เบญจมิตรวิชาการ ครั้งที่ 15 (หน้า 137–147). มหาวิทยาลัยธนบุรี. [สืบค้นวันที่ 1 สิงหาคม 2568]. เข้าถึงได้จาก https://benjamit.thonburi-u.ac.th/ojs/bmv15/article/view/192

Turban, E., et al. (2011). Decision Support and Business Intelligence Systems. Pearson Education. [สืบค้นวันที่ 1 สิงหาคม 2568]. เข้าถึงได้จาก https://scribd.com/document/355912027/Decision-Support-and-Business-Intelligence-Systems-9th-Ed-Efraim-Turban

Arpaci, I., et al. (2023). Intelligence in Decision Support Systems: Trends and Applications. Springer. [สืบค้นวันที่ 1 สิงหาคม 2568]. เข้าถึงได้จาก https://link.springer.com/book/10.1007/978-3-031-14391-5

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for Big Data: An Interdisciplinary Review. [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00369-8

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for Big Data: An Interdisciplinary Review. Journal of Big Data, 7(1), 1-20

Mothilal, R. K., et al. (2020). “Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations”. ใน Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (หน้า 607–617)

Peterson, L. E. (2009). “K-nearest neighbor for statistical decision making and algorithm”. วารสารระบาดวิทยาเชิงคอมพิวเตอร์และชีวสถิติ (Computer Methods and Programs in Biomedicine), 96(3), 105–109.

Breiman, L. (2001). “Random Forests”. วารสารแมชชีนเลิร์นนิง (Machine Learning), 45(1), 5–32.

Rish, I. (2001). “An empirical study of the naive Bayes classifier”. ใน รายงานสืบเนื่องจากการประชุมเชิงปฏิบัติการว่าด้วยการเรียนรู้เชิงประจักษ์ (IJCAI 2001 Workshop on Empirical Methods in AI) (หน้า 41–46)

ณิชานภาพร จงกะสิกิจ. (2559). “ระบบสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรัง กรณีของโรคเบาหวานและความดันโลหิตสูง”. วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม, 3(1), 56–63.

รัชพล กลัดชื่น และจรัญ แสนราช. (2018) . “การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา.” วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี. ปีที่ 17 ฉบับที่ 1 : 1-10.

วิโรจน์ ตรีมงคลโชค. (2022). การวิเคราะห์คุณลักษณะของเสียงเพื่อทำนายคะแนนแอตเทนชันในแบบประเมินโมคา [วิทยานิพนธ์ปริญญามหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย]. คลังปัญญาจุฬาฯ (Chulalongkorn University Intellectual Repository). [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://cuir.car.chula.ac.th/handle/123456789/81533

ภูมิพัฒน์ ดวงกลาง และรัจนา เครือแก้ว. (2019). “แบบจำลองการทำนายแบบอากาศยานจากข้อมูเป้าหมายไม่ทราบฝ่ายอัตโนมัติ.” วารสารวิทยาศาสตร์และเทคโนโลยีนายเรืออากาศ. ปีที่ 15 (มกราคม - ธันวาคม) : 1-8.

ศิริลักษณ์ หล่อพันธ์มณีและคณะ. (2562). “การพยากรณ์รูปแบบของการกระทำความผิดซ้ำของเด็กและเยาวชน.” วารสารนวัตกรรมวิทยาศาสตร์เพื่อการพัฒนาอย่างยั่งยืน. ปีที่ 1 ฉบับที่ 1 (มกราคม - ธันวาคม) : 80-94.

ธนภัทร และคณะ. (2023). “การพยากรณ์ความเสี่ยงในการเกิดโรคเบาหวานโดยใช้เทคนิคการเรียนรู้ของเครื่อง: การเปรียบเทียบประสิทธิภาพของอัลกอริธึม”. วารสารการแพทย์และสุขภาพ, 17(3), 145-156.

Zhuoqing, W., and et al. (2025). “Prognosis of Metabolically Associated Fatty Liver Disease (MAFLD) by developing an intelligent forecasting system with Temporal Association Rules (TARs)”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 188, 114391.

Xuelong, C., Yiping, C., and Guojie, Y. (2025). “Predicting Cyberviolence Behavior (CVB) by incorporating the Theory of Planned Behavior into an intelligent forecasting system”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 188, 114382.

Xue, S., and et al. (2025). การศึกษาผลของการให้ข้อมูล Bayesian Updating ต่อพฤติกรรมการตัดสินใจแบบซ้ำในการซื้อสินค้าออนไลน์ โดยมุ่งเน้นไปที่การลดอคติจากสถานะเดิม https://www.sciencedirect.com/science/article/pii/S0167923624002045

Shuixia, C. (2023). “Mitigating the status quo bias in online repeat purchase: An empirical study of Bayesian information updating”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 188, 114383.

Hemang, S., and et al. (2024). “Hemang, S., และคณะ. (2024). Bitcoin price prediction using multi-source information and LSTM deep learning”. วารสารซอฟต์แวร์และวิศวกรรมข้อมูล (Journal of Software and Data Engineering), 3(2), 145–158.

Li-Chen, C., and et al. (2024). “Detecting fake reviewers by using social context networks with graph neural networks”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 177, 114041.

Jiwei Luo and et al. (2023). “A probabilistic modeling method for fake review detection”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 171, 113988.

Stephanie, D., Kristof, C., and Arno, C. (2024). “Predicting customer life events using financial data: Evidence from a European financial service provider”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 186, 114343.

Shuixia Chen and et al. (2023). “Explainable hotel booking cancellation prediction: A deep learning approach with attention mechanism”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 169, 113954.

Haoqiang, J., Paulo, G., and Debra, V. (2023). “The nurse practitioner assignment problem in inpatient wards”. วารสารระบบสนับสนุนการตัดสินใจ (Decision Support Systems), 167, 113915.

สุภาวดี แก้วสินธุ์. (2023). ปัจจัยทำนายภาวะสับสนเฉียบพลันในผู้สูงอายุที่เข้ารับบริการในแผนกฉุกเฉิน. วิทยานิพนธ์ปริญญาพยาบาลศาสตรมหาบัณฑิต สาขาวิชาการพยาบาลผู้ใหญ่และผู้สูงอายุ คณะพยาบาลศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://cuir.car.chula.ac.th/handle/123456789/83987

นายกฤต โรจน์รุ่งเรืองพรใ (2022). การระบุสภาวะความเครียดของพืชในระยะเริ่มต้นโดยใช้การเรียนรู้เชิงลึกที่พัฒนาบน FPGA. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://cuir.car.chula.ac.th/handle/123456789/77115

ปวริศ เวชวรรณกิจกุล. (2022). การคาดการณ์ทิศทางราคาทองคำจากความรู้สึกของข้อความบนทวิตเตอร์โดยใช้เทคนิคการเรียนรู้ด้วยเครื่อง. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://cuir.car.chula.ac.th/handle/123456789/81530

ณัฐชา ซาซุม. (2023). การเรียนรู้เชิงลึกสำหรับเข้ารหัสบัญชีการจำแนกโรคระหว่างประเทศ (ไอซีดี) จากบันทึกเวชระเบียนโดยใช้คลังข้อมูลภาษาไทยและภาษาอังกฤษ. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. [สืบค้นวันที่ 1 มีนาคม 2569]. เข้าถึงได้จาก https://cuir.car.chula.ac.th/handle/123456789/84315

Kuhn, M., and Johnson, K. “Applied Predictive Modeling”. สำนักพิมพ์ Springer, New York.

Guyon, I., and et al. “Gene selection for cancer classification using support vector machines”. วารสารการเรียนรู้ของเครื่อง (Machine Learning), 46, 389-422.

ภาคผนวก ก

แบบสอบถามสำหรับรวบรวมข้อมูล

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ในส่วนนี้ประกอบด้วยรายละเอียดของเครื่องมือที่ผู้วิจัยใช้ในการดำเนินงาน ทั้งในส่วนของการเก็บรวบรวมข้อมูลเพื่อพัฒนาแบบจำลอง และการประเมินความพึงพอใจของระบบ โดยผู้ที่สนใจสามารถเข้าถึงเครื่องมือดังกล่าวได้ผ่านการสแกนรหัสคิวอาร์ (QR Code) ที่ปรากฏด้านล่างนี้

1. แบบสอบถามเพื่อการเก็บรวบรวมข้อมูลพยากรณ์การเลือกที่พักอาศัย แบบสอบถามชุดนี้ใช้สำหรับเก็บข้อมูลปัจจัยด้านประชากรศาสตร์และเศรษฐกิจ จำนวน 28 ปัจจัย จากกลุ่มตัวอย่างจำนวน 1,107 ชุด เพื่อนำมาใช้ในการฝึกสอนและทดสอบแบบจำลองปัญญาประดิษฐ์

ภาพที่ ก-1 คิวอาร์โค๊ดสำหรับจัดเก็บข้อมูลที่พักอาศัย

https://shorturl.at/5VPCm

2. แบบประเมินความพึงพอใจต่อการใช้งานระบบสนับสนุนการตัดสินใจ แบบสอบถามชุดนี้ใช้สำหรับทดสอบประสิทธิภาพของเว็บแอปพลิเคชันในมุมมองของผู้ใช้งานจริง จำนวน 30 คน โดยเน้นการประเมินในด้านความแม่นยำ ความง่ายในการใช้งาน และประโยชน์ของคำแนะนำจากเทคนิค DiCE



ภาพที่ ก-2 คิวอาร์โค๊ดสำหรับจัดเก็บข้อมูลความพึงพอใจ

https://shorturl.at/Wf3ay

ภาคผนวก ข

ส่วนการพัฒนาระบบ

รายละเอียดรหัสต้นฉบับ (Source Code) และการฝึกสอนแบบจำลอง ในส่วนนี้แสดงรายละเอียดของกระบวนการฝึกสอนแบบจำลอง CatBoost ที่ผู้วิจัยได้พัฒนาขึ้น โดยมุ่งเน้นความโปร่งใสของกระบวนการจัดการข้อมูลและการปรับจูนไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning)

1. รหัสต้นฉบับสำหรับการพัฒนาแบบจำลอง CatBoost (7 ปัจจัยหลัก) ผู้วิจัยเลือกใช้เทคนิค Unseen Data Separation เพื่อแยกข้อมูลออกมาคลาสละ 100 ตัวอย่าง สำหรับการทดสอบประสิทธิภาพครั้งสุดท้าย เพื่อให้มั่นใจว่าแบบจำลองมีความแม่นยำต่อข้อมูลที่ไม่เคยเห็นมาก่อน (Generalization Ability)

```python
# การเตรียมข้อมูลและแยก Unseen Set เพื่อป้องกัน Data Leakage
df0 = df_latest_check[df_latest_check['class'] == 0].sample(100, random_state=42)
df1 = df_latest_check[df_latest_check['class'] == 1].sample(100, random_state=42)
df_unseen = pd.concat([df0, df1]).sample(frac=1, random_state=42)

# การกำหนด Parameter Grid สำหรับ RandomizedSearchCV
param_grid = {
    'iterations': [300, 500, 800],
    'learning_rate': [0.01, 0.02, 0.03],
    'depth': [4, 5, 6],
    'l2_leaf_reg': [3, 5, 10, 20, 30, 40, 50],
    'grow_policy': ['Lossguide'],
    'bagging_temperature': [0.5, 1.0, 1.5]
}

# การฝึกสอนด้วย 10-Fold Cross-Validation
kf = KFold(n_splits=10, shuffle=True, random_state=42)
random_search = RandomizedSearchCV(
    estimator=CatBoostClassifier(verbose=0),
    param_distributions=param_grid,
    n_iter=100,
    cv=kf,
    scoring='f1_weighted',
    n_jobs=-1
)
random_search.fit(X_processed, y_processed)
```

ภาพที่ ข-1 การสร้างแบบจำลอง CatBoost

2. กระบวนการประเมินผลแบบ Unseen Evaluation เพื่อให้ผลงานวิจัยมีความน่าเชื่อถือ ผู้วิจัยได้ทำการทดสอบด้วยชุดข้อมูล Unseen ทันทีหลังจากได้รับแบบจำลองที่ดีที่สุด (Best Estimator) เพื่อวัดค่า Accuracy, F1-Score และ ROC AUC ที่แท้จริง

```python
# การประเมินผลบนชุดข้อมูล Unseen เพื่อวัดประสิทธิภาพจริง
y_pred_unseen = best_model.predict(X_unseen)
print(classification_report(y_unseen, y_pred_unseen))

# บันทึกแบบจำลองเพื่อนำไปใช้งานบนระบบสนับสนุนการตัดสินใจ (React)
best_model.save_model("rfe_best_catboost_model.cbm")
```

ภาพที่ ข-2 การประเมิณผล CatBoost ด้วย Unseen

3. สรุปพารามิเตอร์ที่ดีที่สุด (Best Parameters) จากการทำ RandomizedSearchCV พบว่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับแบบจำลอง 7 ปัจจัยหลัก


```json
{
    "random_strength": 20,
    "max_ctr_complexity": 3,
    "learning_rate": 0.03,
    "l2_leaf_reg": 50,
    "iterations": 800,
    "grow_policy": "Lossguide",
    "depth": 5,
    "border_count": 32,
    "bagging_temperature": 1
}
```


ภาพที่ ข-3 Best Parameter ของแบบจำลองที่ใช้ในระบบ

4. การจัดการข้อมูลเพื่อติดต่อเข้าใช้งานแบบจำลองโมเดล CatBoost โดนเชื่อต่อ API

```
const handlePrediction = async () => {
    setIsLoading(true); // เริ่มแสดงผลหน้าจอรอโหลด
    try {
      // 1. จัดเตรียม Payload จาก State formData ให้ตรงกับความต้องการของ Model
      const payload = {
        age: Number(formData.ageRange),
        income: Number(formData.incomeRange),
        area: Number(formData.residentialArea),
        expenses_to_work: Number(formData.commuteCost),
        sex: Number(formData.gender),
        ready: Number(formData.readinessToPay),
        sick: Number(formData.healthcareReadiness),
        safty_in_the_house: Number(formData.residentialSatisfaction),
        property: Number(formData.debtToAsset),
        work: Number(formData.employmentStatus),
      };

      // 2. ส่งข้อมูลไปที่ Model เพื่อพยากรณ์ผล (ซื้อ หรือ เช่า)
      const data = await predict(payload);

      // 3. เงื่อนไขเชิงรุก: หาก Model ทำนายว่าเป็น "เช่า" (0) ให้ใช้เทคนิค DiCE ต่อทันที
      if(data.prediction === 0){
        const resp = await suggest(payload); // เรียก API ส่วน Explainable AI

        const counterfactuals = resp.counterfactuals;
        const result = {};

        // วนลูปเพื่อดึงค่าปัจจัยที่ Model แนะนำว่าควรปรับเปลี่ยน
        for (const item of counterfactuals) {
            for (const key in item) {
                if (!(key in result) && item[key] && typeof item[key].value !== "undefined") {
                    result[key] = item[key].value;
                }
            }
        }
        setSuggestions(result); // เก็บข้อมูลค่าแนะนำลงใน State
      }

      setPrediction(data); // เก็บผลการทำนาย (0 หรือ 1)
      setIsOpen(true);     // เปิด Modal แสดงผลลัพธ์ให้ผู้ใช้เห็น
    } catch (error) {
      console.error('Predict error:', error);
      setIsOpen(false);
    } finally {
      setIsLoading(false); // ปิดหน้าจอรอโหลด
    }
  };
```

ภาพที่ ข-4 การติดต่อ API เพื่อใช้งานแบบจำลอง

3. ส่วนของการใช้งานแบบจำลองด้วยภาษา Python

```python
@app.post("/dice")
def dice_counterfactual(input_data: Dict[str, Any]):
    permitted_range = {}
    features_to_vary = []

    for f in EXPECTED_FEATURES:
        if f == "sex":
            # ล็อกค่าคงที่: เพศเป็นปัจจัยที่ไม่สามารถเปลี่ยนแปลงได้ในทางสถิติ
            permitted_range[f] = [input_row[f], input_row[f]]
            continue

        if f == "income":
            # เงื่อนไขรายได้: ระบบจะแนะนำให้ "เพิ่มขึ้น" เท่านั้น (ค่าปัจจุบันถึง 5)
            permitted_range[f] = [input_row["income"], 5]
            features_to_vary.append(f)

        # ... การกำหนดช่วงตัวแปรอื่นๆ เช่น หนี้สิน (ต้องลดลง) หรือ ความพร้อม (ต้องเพิ่มขึ้น) ...

    # สร้างคำแนะนำจำลองสถานการณ์ (Counterfactuals) 1 รูปแบบที่เป็นไปได้
    cf = dice_exp.generate_counterfactuals(
        df_scaled, total_CFs=1, desired_class=1,
        features_to_vary=features_to_vary, permitted_range=permitted_range
    )
```

ภาพที่ ข-5 การแนะนำของระบบด้วย DiCE



ภาคผนวก ค

แบบประเมินค่าดัชนีความสอดคล้อง (IOC) ของผู้เชี่ยวชาญ

แบบประเมินค่าดัชนีความสอดคล้อง (IOC) ของผู้เชี่ยวชาญ
การศึกษาพฤติกรรมในการเลือกซื้อที่อยู่อาศัยหลังแรก หรือเช่าที่พักอาศัย
คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

คำชี้แจง

1. แบบสอบถามฉบับนี้มุ่งตรวจสอบเพื่อหาค่าความเที่ยงตรง (Validity) โดยการวิเคราะหด์ ัชนีความสอดคล้อง (Index of item objective congruence: IOC) ของแบบสอบถามและข้อเสนอแนะของผู้เชี่ยวชาญ เพื่อนำไปปรับปรุงแบบสอบถามให้สมบูรณ์ยิ่งขึ้น

2. แบบทดสอบฉบับนี้มีทั้งหมด 11ด้าน
ด้านที่ 1 ข้อมูลพื้นฐานของผู้ทำแบบสอบถามส่งผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 2 ความรู้ด้านการเงินมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 3 พฤติกรรมในการทำงานมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 4 สถานะภาพมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 5 สถานที่พักอาศัยมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 6 สถานที่พักอาศัยมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 7 พฤติกรรมความชื่นชอบส่วนบุคคลมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่
ด้านที่ 8 พฤติกรรมทางด้านการเงิน
ด้านที่ 9 ระยะทาง
ด้านที่ 10 สภาพแวดล้อม
ด้านที่ 12 อื่นๆ

3. ขอความกรุณาผู้ทรงคุณวุฒิหรือท่านผู้เชี่ยวชาญ พิจารณาแบบสอบถามว่ามีความสอดคล้องกับตัวแปร / นิยามศัพท์เชิงปฏิบัติการของการวิจัยเรื่องนี้หรือไม่ ด้วยการให้คะแนนแต่ละข้อคำถามในระบบ IOC โดยการทำเครื่องหมาย P ลงในช่องเกณฑ์การให้คะแนน

1) ให้+1 พฤติกรรมบ่งชี้สอดคล้องกับตัวแปร / นิยามศัพท์เชิงปฏิบัติการ
2) ให้0 พฤติกรรมบ่งชี้ไม่แน่ใจกับตัวแปร / นิยามศัพท์เชิงปฏิบัติการ
3) ให้-1 พฤติกรรมบ่งชี้ไม่สอดคล้องกับตัวแปร / นิยามศัพท์เชิงปฏิบัติการ

4. ผู้วิจัยขอความกรุณาท่านผู้ทรงวุฒิและผู้เชี่ยวชาญ ให้ข้อเสนอแนะหรือความคิดเห็นเพิ่มเติมในประเด็นที่ยังไม่สมบูรณ์โดยการเขียนข้อเสนอแนะไว้ท้ายแบบสอบถามในแต่ละข้อ

ผู้วิจัยขอขอบพระคุณในความกรุณาของท่านอย่างมาก

นายสรุ เดช ลีรุจิกุล

นักศึกษาปริญญาโท (การจัดการเทคโนโลยีสารสนเทศ)

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ร่างแบบสอบถามเพื่อการวิจัย (IOC)

เรื่อง “การศึกษาพฤติกรรมในการเลือกซื้อที่อยู่อาศัยหลังแรก หรือเช่าที่พักอาศัย”

วัตถุประสงค์/เนื้อหา	แบบทดสอบ/แบบสอบถาม	ระดับของการพิจารณา		
		+1	0	-1
1.ข้อมูลพื้นฐานของผู้ทำแบบสอบถามส่งผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 ช่วงอายุของผู้ทำแบบสอบถาม			
	2 เพศของผู้ทำแบบสอบถาม			
	3 พื้นที่อยู่อาศัย (กรุงเทพชั้นใน, กรุงเทพชั้นนอก,ปริมณฑล)			
2.ความรู้ด้านการเงินมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 ภาระหนี้สินเทียบกับสินทรัพย์			
	2 มีเงินสำรองฉุกเฉิน 6 เดือนหรือไม่			
	3 รายได้ต่อเฉลี่ยต่อเดือน			
	4 ในมุมมองของท่านมองว่าที่พักอาศัยเป็นสินทรัพย์มากเพียงใด			
	5 ถ้าหากท่านป่วยและต้องใช้ค่ารักษาพยาบาล 2 ล้านบาท ท่านมีความพร้อมรับมือ เหตุการณ์มีเพียงใด			
	6 ท่านวางแผนการเกษียณไว้แล้ว			
3.พฤติกรรมในการทำงานมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 สถานะในที่ทำงาน เช่น พนักงานเอกชน ราชการ ฟรีแลนซ์			
	2 ในระยะเวลา 5 ปีที่ผ่านมามีการย้ายหรือเปลี่ยนสถานที่ทำ			

วัตถุประสงค์/เนื้อหา	แบบทดสอบ/แบบสอบถาม	ระดับของการพิจารณา		
		+1	0	-1
	งานบ่อยแค่ไหน			
	3 ต้องเดินทางไปที่ทำงาน สัปดาห์ละกี่วัน			
	4 การเดินทางไปสถานที่ทำงานใช้ค่าเดินทางเดือนละ เท่าไหร่			
4.สถานะภาพมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 ท่านต้องการอาศัยร่วมกัน หรือใกล้เคียงครอบครัว			
	2 สถานะปัจจุบัน			
5.สถานที่พักอาศัยมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 สถานที่พักที่อาศัยอยู่เป็นของใคร			
	2 คุณมีความพึงพอใจในที่พักอาศัยของมากเพียงใด			
6.แผนการที่วางไว้ในอนาคตมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 ท่านมีบุตรหรือต้องการมี บุตรหรือไม่			
	2 ท่านต้องการรายได้จาก ที่พักอาศัยของท่านหรือไม่			
	3. ท่ามีความกดดันจากคนรอบตัวในการซื้อที่พักอาศัยหรือไม่			
7.พฤติกรรมความชื่นชอบส่วนบุคคลมีผลต่อการเลือกซื้อและเลือกเช่าที่พักอาศัยหรือไม่	1 ท่านสบายใจเมื่อได้พักใน ที่พักที่เป็นของตนเองมาก เพียงใด			
	2 ท่านมีความชื่นชอบที่จะ ตกแต่งที่พักอาศัยของท่าน มาก เพียงใด			

วัตถุประสงค์/เนื้อหา	แบบทดสอบ/แบบสอบถาม	ระดับของการพิจารณา		
		+1	0	-1
	3 ท่านชอบที่พักอาศัยแบบใด (บ้านเดี่ยว, คอนโด, ฯลฯ)			
8.พฤติกรรมทางด้านการเงิน	1 ท่านคำนึงถึงค่าใช้จ่าย ในการซื้อ/เช่าที่อยู่ อาศัยแรก เช่น ดอกเบี้ยหรือค่าเช่า			
	2 กังวลในการผ่อนชำระ ที่อยู่อาศัยในอนาคต			
9.ระยะทาง	1 ที่อยู่อาศัยของท่านเดินทางสะดวกเพียงใด			
10.สภาพแวดล้อม	1 ท่านพึงพอใจสภาพแวดล้อมที่อยู่อาศัยของท่านมากเพียงใด			
11. อื่นๆ	1 ท่านมีอำนาจในการตัดสินใจเลือกซื้อหรือเช่าที่อยู่อาสัยหลังแรกหรือไม่			
	2 ท่านซื้อที่พักอาศัยแล้วใช่หรือไม่			

ลงชื่อ .....

( )

ภาคผนวก ง

คู่มือการใช้งานระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัย

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ระบบถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันหน้าเดียว (Single Page Application) เพื่อความสะดวกรวดเร็วในการใช้งาน โดยระบบจะประมวลผลผ่านโมเดล CatBoost และใช้เทคนิค DiCE ในการวิเคราะห์เงื่อนไขที่ต้องปรับเปลี่ยน (Counterfactuals) เฉพาะในกรณีที่ผลการทำนายระบุว่าผู้ใช้งานควรเลือกการ "เช่า" ที่พักอาศัย

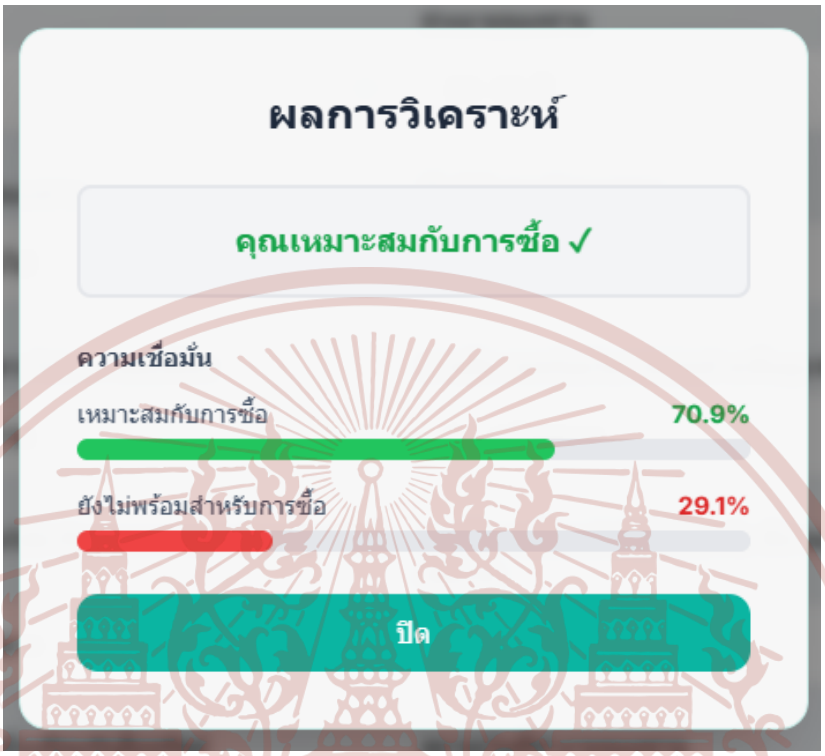
ขั้นตอนการใช้งาน

ขั้นตอนที่ 1: เมื่อเปิดระบบขึ้นมา ผู้ใช้งานจะพบกับส่วนของแบบฟอร์ม ให้ดำเนินการกรอกข้อมูลตามปัจจัยที่กำหนดให้ครบถ้วน (อาทิ ข้อมูลรายได้, อายุ, ภาระหนี้สิน) แล้วกดปุ่ม "วิเคราะห์ผล"



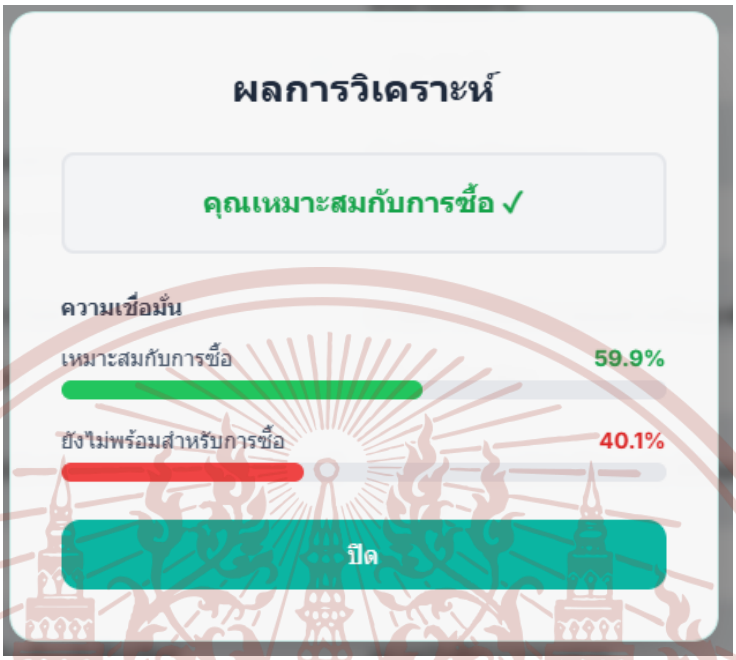
ภาพที่ ง-1 หน้าจอส่วนการรับข้อมูล (Input Form)

ขั้นตอนที่ 2: การแสดงผลการทำนายกรณีผลลัพธ์เป็น "ซื้อ" ในกรณีที่โมเดล CatBoost วิเคราะห์ว่าสถานะทางการเงินและปัจจัยต่างๆ เหมาะสมกับการซื้อที่พักอาศัย ระบบจะแสดงข้อความสรุปผลว่า "ควรเลือกซื้อที่พักอาศัย" โดยจะไม่แสดงตารางคำแนะนำเพิ่มเติม



ภาพที่ ง-2 การแสดงผลการทำนายกรณีความต้องการเหมาะสมกับการซื้อ

ขั้นตอนที่ 3: การแสดงผลการทำนายและคำแนะนำเชิงรุกกรณีผลลัพธ์เป็น "เช่า" ในกรณีที่ผลการทำนายระบุว่า "ควรเลือกเช่าที่พักอาศัย" ระบบจะแสดงส่วนของคำแนะนำเชิงรุก (Counterfactual Explanations) จากเทคนิค DiCE ขึ้นมาโดยอัตโนมัติที่ด้านล่างของผลทำนาย เพื่อแจ้งให้ผู้ใช้ทราบว่าต้องปรับเปลี่ยนปัจจัยใดบ้างเพื่อให้เปลี่ยนสถานะเป็น "ซื้อ" ได้ในอนาคต



ภาพที่ ง-3 การแสดงผลทำนายกรณีเช่าพร้อมคำแนะนำการปรับเปลี่ยนปัจจัย (DiCE)

ประวัติผู้เขียน

ชื่อ นายสุรเดช ลีรุจิกุล

ชื่อการค้นคว้าอิสระ การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกซื้อหรือเช่าที่พักอาศัยด้วยเทคนิคเหมืองข้อมูล

สาขาวิชา ระบบสารสนเทศเพื่อการจัดการ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ประวัติ ระดับปริญญาตรี เทคโนโลและการจัดการอุตสาหกรรม
สาขาเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

2562-2563 วิศวกร ซอฟต์แวร์
ทรู คอร์ปอเรชั่น

2563-2565 วิศวกร ซอฟต์แวร์
บริษัท ไบรต์ เอ็คซ์เพิร์ท จำกัด

2565-ปัจจุบัน วิศวกร ซอฟต์แวร์
ธนาคารทหารไทยธนชาต จำกัด (มหาชน)