การพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยว ในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่อง

Forecasting the Selling Prices of Non-Performing Asset Detached Houses in Bangkok by Machine Learning Techniques

พีรภัทร วัสแสง[1] และ กองกูณฑ์ โตชัยวัฒน์[2]

Peerapat Wassaeng[1] and Kongkoon Tochaiwat[2]

สาขาวิชานวัตกรรมการพัฒนาอสังหาริมทรัพย์ คณะสถาปัตยกรรมศาสตร์และการผังเมือง มหาวิทยาลัยธรรมศาสตร์

Department of Innovative Real Estate Development, Faculty of Architecture and Planning, Thammasat University, Thailand

Corresponding Author, E-mail: [1]peerapat.was@dome.tu.ac.th

Received July 30, 2022; **Revised** March 28, 2024; **Accepted** March 29, 2024

บทคัดย่อ

การนำปัญญาประดิษฐ์มาใช้ในอุตสาหกรรมอสังหาริมทรัพย์ นอกจากจะช่วยให้ผู้บริโภคมีข้อมูลในการตัดสินใจซื้ออสังหาริมทรัพย์ที่มีประสิทธิภาพมากขึ้นแล้ว ยังช่วยให้ผู้ประกอบการธุรกิจอสังหาริมทรัพย์สามารถนำข้อมูลมาใช้ประโยชน์ได้อย่างเต็มที่ รวมถึงสร้างโอกาสและเพิ่มมูลค่าในการแข่งขัน บทความนี้มีวัตถุประสงค์เพื่อทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร โดยใช้วิธีการวิจัยเป็นการวิจัยเชิงปริมาณและใช้แนวคิดการเรียนรู้ของเครื่องเป็นกรอบการวิจัย ทำการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร จำนวน 446 ตัวอย่าง ใช้วิธีการคัดเลือกแบบสะดวกโดยเลือกทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่มีข้อมูลครบถ้วนตามตัวแปรที่กำหนด จากนั้นทำการบันทึกข้อมูลลงในแบบตรวจสอบรายการ วิเคราะห์ข้อมูลโดยใช้สถิติเชิงพรรณนาและเทคนิคการเรียนรู้ของเครื่อง ได้แก่ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคต้นไม้กาเดียนบูทสเต็ท เทคนิคโครงข่ายประสาทเทียม และเทคนิคการทำเหมืองข้อมูลแบบรวมกลุ่ม

ผลการวิจัยพบว่า สมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote ให้ค่าความคลาดเคลื่อนเฉลี่ยกำลังสองในการพยากรณ์ต่ำที่สุดเมื่อเทียบกับแบบจำลองที่เป็นส่วนประกอบ โดยมีค่า RMSE, ค่า R^2 และค่า Beta ของแบบจำลองเท่ากับ 3,746,335.580 บาท 0.5377 และ 0.4919 ตามลำดับ จะเห็นได้ว่า

การสร้างสมการพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่องที่หลากหลาย ช่วยเพิ่มประสิทธิภาพโดยรวมของสมการพยากรณ์ เนื่องจากการนำตัวจำแนกข้อมูลที่หลากหลายมารวมกันด้วยวิธีการรวมกลุ่ม สามารถช่วยลดปัญหาการเกิดไบแอสของข้อมูลได้และเสริมประสิทธิภาพในการจำแนกข้อมูล ทำให้โมเดลที่ได้มีประสิทธิภาพมากขึ้น ทั้งนี้ผู้ประกอบการธุรกิจอสังหาริมทรัพย์สามารถนำเทคนิคการเรียนรู้ของเครื่องไปใช้วิเคราะห์ข้อมูลเพื่อแสวงหาผลประโยชน์จากข้อมูลประกอบการตัดสินใจ โดยเทคนิคการเรียนรู้ของเครื่องสามารถสร้างสมการพยากรณ์ได้โดยใช้ข้อมูลที่สามารถหาได้จากเครือข่ายอินเทอร์เน็ตแม้ว่าจะมีข้อมูลจำนวนไม่มากนัก

คำสำคัญ: การพยากรณ์ราคาขาย; ทรัพย์สินรอการขาย; เทคนิคการเรียนรู้ของเครื่อง; บ้านเดี่ยว; ปัจจัยทางกายภาพ

Abstract

Artificial intelligence is being used in the real estate industry to provide consumers with information that will make effective decisions to buy real estate. Additionally, it facilitates the optimal utilization of data by real estate business operators, which includes generating possibilities and raising competitive worth. This Article aimed to test the effectiveness of machine learning techniques in modeling for forecasting the selling prices of non-performing asset (NPA) detached houses in Bangkok. The research method is quantitative and uses machine learning concepts as the research framework. Data were gathered from 446 samples of non-performing asset (NPA) detached houses in Bangkok. They were selected by convenience sampling that have complete information according to the specified variables. Then, data were recorded in the checklist and analyzed by Descriptive statistics and machine learning techniques, including Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN), and Ensemble Vote.

The research results were found that the obtained model derived from Ensemble Vote technique has the least Root Mean Square Error (RMSE) compared with Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN), and Ensemble Vote techniques. The RMSE, R^2 and Beta of the best model are 3,746,335.580 Bath, 0.5377 and 0.4919, respectively. Evidently, constructing a modeling for forecasting the selling prices of non-performing asset (NPA) detached houses in Bangkok using a variety of machine learning techniques will increase the overall efficiency of the

forecast equation due to combining multiple classifiers with a clustering method can help reduce data bias. They can help each other to enhance the efficiency of data classification make a model more efficient. Real estate business entrepreneurs can apply machine learning techniques to analyze data to exploit data for decision–making. The results showed the potential of using machine learning techniques in predicting the prices of non–performing asset (NPA) detached houses by the data available from websites although there is not much information.

Keywords: Forecasting the Selling Prices; Non–Performing Asset; Machine Learning Techniques; Detached House; Physical Factors

บทนำ

เหมืองข้อมูล (Data Mining) เป็นศาสตร์หนึ่งที่ได้รับความนิยมอย่างแพร่หลาย เนื่องจากเป็นกระบวนการค้นหารูปแบบและความสัมพันธ์จากข้อมูลจำนวนมากที่ซ่อนอยู่ในชุดข้อมูลโดยอาศัยหลักสถิติ การรู้จำ รวมถึงการเรียนรู้ของเครื่อง (Machine Learning) (Yang, 2019) ทั้งนี้ เหมืองข้อมูลกับการเรียนรู้ของเครื่องมักจะใช้วิธีการเหมือนกันและมีส่วนเกี่ยวข้องกัน แต่มีความแตกต่างกัน คือ การทำเหมืองข้อมูลเน้นการค้นหาคุณสมบัติที่ไม่รู้จากข้อมูลที่ได้มา กล่าวได้ว่าเป็นขั้นตอนการวิเคราะห์เพื่อค้นหาความรู้ในฐานข้อมูล ในขณะที่การเรียนรู้ของเครื่องเน้นการพยากรณ์ข้อมูลจากคุณสมบัติที่รู้แล้วที่ได้เรียนรู้มาจากข้อมูล ซึ่งอัลกอริทึมในการทำเหมืองข้อมูลสามารถแบ่งได้เป็น 2 ประเภท คือ การวิเคราะห์เหมืองข้อมูลเพื่อการทำนาย (Predictive) และการวิเคราะห์เหมืองข้อมูลเพื่อการอธิบาย (Descriptive) (Berry, Mohamed & Yap, 2020)

ปัจจุบันเทคนิคการเรียนรู้ของเครื่อง (Machine Learning Techniques) ซึ่งเป็นองค์ประกอบหนึ่งของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) ได้ถูกนำมาใช้ในธุรกิจที่หลากหลาย โดยธุรกิจอสังหาริมทรัพย์เป็นอีกหนึ่งธุรกิจที่ได้นำการเรียนรู้ของเครื่องมาใช้ในการประมวลผลข้อมูลที่มีจำนวนมากและมีโครงสร้างซับซ้อนได้ภายในเวลาอันรวดเร็ว ช่วยเพิ่มประสิทธิภาพการตัดสินใจในกระบวนการต่าง ๆ ของการดำเนินธุรกิจอสังหาริมทรัพย์ ทั้งการศึกษาความเป็นไปได้ของโครงการ การตัดสินใจซื้อที่ดิน การกำหนดรูปแบบโครงการ การออกแบบ การจัดซื้อจัดจ้าง การบริหารงานก่อสร้าง การบริหารหลังการขาย และการบริหารทรัพย์สิน เนื่องจากธุรกิจอสังหาริมทรัพย์เป็นธุรกิจที่มีมูลค่าสูงถึงประมาณร้อยละ 8 ของผลิตภัณฑ์มวลรวมภายในประเทศ อีกทั้งการตัดสินใจต่าง ๆ ต้องอาศัยข้อมูลสนับสนุนค่อนข้างมาก รวมถึงต้องการบูรณาการความรู้หลายศาสตร์ เช่น การเงิน การตลาด การขาย กฎหมาย

สถาปัตยกรรม วิศวกรรม การจัดซื้อจัดจ้าง การก่อสร้าง การบริหารทรัพย์สิน เข้าด้วยกัน (Tochaiwat, 2023) ทั้งนี้ ในต่างประเทศได้มีการศึกษาเกี่ยวกับการนำปัญญาประดิษฐ์มาใช้ในอุตสาหกรรมอสังหาริมทรัพย์ด้านต่าง ๆ เช่น การค้นหาอสังหาริมทรัพย์ (Home Search Experience) การประเมินราคา (Property Value Prediction) การศึกษาพฤติกรรมของผู้บริโภค (Lead Generation) การเพิ่มโอกาสในการกู้ (Efficient Mortgage Lending) (Kuflinski, 2020) อาทิ บริษัท Zillow จำกัด ซึ่งเป็นตัวอย่างหนึ่งของบริษัทที่มีการนำปัญญาประดิษฐ์มาใช้ในอุตสาหกรรมอสังหาริมทรัพย์ โดยได้พัฒนาโมเดลชื่อ Zestimate เพื่อใช้ประเมินราคาอสังหาริมทรัพย์ที่ประกาศขายอยู่ในเว็ปไซต์ของตนเอง ซึ่งช่วยให้ผู้บริโภคสามารถมีข้อมูลในการตัดสินใจซื้ออสังหาริมทรัพย์ได้อย่างมีประสิทธิภาพมากขึ้น

อย่างไรก็ตาม แม้ว่าการเรียนรู้ของเครื่องจะสามารถตอบโจทย์ของธุรกิจอสังหาริมทรัพย์ได้หลายประการ แต่การเรียนรู้ของเครื่องสามารถนำไปใช้งานได้หลากหลายรูปแบบ อีกทั้งประสิทธิภาพของการเรียนรู้ของเครื่องยังขึ้นอยู่กับปัจจัยหลายประการ เช่น จำนวนข้อมูล รูปแบบการเรียนรู้ เทคนิคการเรียนรู้ของเครื่องที่ใช้ในการสร้างแบบจำลอง เป็นต้น ทั้งนี้ ที่ผ่านมามีผู้สนใจศึกษาการนำเทคนิคการเรียนรู้มาประยุกต์ใช้ในธุรกิจอสังหาริมทรัพย์หลายงาน เช่น การพัฒนาแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์ความนิยมอาคารชุดของผู้บริโภคออนไลน์ด้วยภาพถ่าย (Sawangum, 2022) การศึกษารูปแบบการลงทุนพัฒนาอสังหาริมทรัพย์เพื่อการพักอาศัย กรณีศึกษา จังหวัดนครศรีธรรมราช (Vongsung, 2022) การพยากรณ์ราคาคอนโดมิเนียมในกรุงเทพมหานครโดยเทคนิค Web Mining (Sunkpho & Ramjan, 2020) การคาดการณ์ผลตอบแทนในอนาคตของพอร์ตโฟลิโอของตราสารทุนหมวดพัฒนาอสังหาริมทรัพย์โดยการใช้โครงข่ายประสาทเทียม (Jiamwiriyakul, 2019) เป็นต้น จะเห็นได้ว่า งานวิจัยที่พบส่วนใหญ่มุ่งเน้นไปที่การศึกษาการลงทุนในหลักทรัพย์หรือการพัฒนาโครงการอสังหาริมทรัพย์มือหนึ่ง ยังไม่ค่อยพบการนำเทคนิคการเรียนรู้มาประยุกต์ใช้กับอสังหาริมทรัพย์มือสองหรือทรัพย์สินรอการขาย

ทั้งนี้ จากรายงานของศูนย์ข้อมูลอสังหาริมทรัพย์ ธนาคารอาคารสงเคราะห์ (Real Estate Information Center, 2021) ซึ่งได้ทำการจัดเก็บข้อมูลที่อยู่อาศัยมือสองทั่วประเทศในไตรมาส 4 ปี 2563 จากการประกาศขายผ่านเว็บไซต์ของบริษัทภาคเอกชน รวมถึงข้อมูลที่อยู่อาศัยมือสองของสถาบันการเงินของรัฐ บริษัทบริหารสินทรัพย์ภาครัฐ และกรมบังคับคดี ที่ประกาศขายผ่านเว็บไซต์ตลาดนัดบ้านมือสอง พบว่า ที่อยู่อาศัยมือสองมีจำนวนหน่วยเสนอขายเฉลี่ยเดือนละ 113,211 หน่วย และมีมูลค่าเฉลี่ยเดือนละ 842,253 ล้านบาท โดยบ้านเดี่ยวหรือบ้านแฝดมีสัดส่วนการประกาศขายมากที่สุด รองลงมาเป็นอาคารชุดพักอาศัย ทาวน์เฮ้าส์ และอาคารพาณิชย์ ตามลำดับ และเมื่อพิจารณาด้านมูลค่าที่อยู่อาศัยมือสอง พบว่า บ้านเดี่ยวและบ้านแฝดมีสัดส่วนมากที่สุด รองลงมาคือ อาคารชุดพักอาศัย ทาวน์เฮ้าส์ และอาคารพาณิชย์ ตามลำดับ

ด้วยเหตุนี้ ผู้วิจัยจึงสนใจศึกษาการพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อช่วยให้นักลงทุนหรือผู้ที่สนใจซื้อทรัพย์สินรอการขายประเภทบ้านเดี่ยวทราบปัจจัยที่มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครเพื่อใช้เป็นข้อมูลประกอบการตัดสินใจลงทุน รวมถึงช่วยให้ธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์สามารถนำไปใช้เป็นแนวทางในการกำหนดราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครได้อย่างเหมาะสมกับลักษณะทางกายภาพของทรัพย์สินรอการขาย โดยในการวิจัยครั้งนี้มุ่งทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร ซึ่งเป็นเครื่องมือที่ใช้ในการประเมินสถานการณ์ของธุรกิจในปัจจุบันได้อย่างแม่นยำและให้ข้อมูลเชิงลึก (Insight) สำหรับใช้เป็นแนวทางการสร้างแบบจำลองพยากรณ์ที่สามารถนำมาศึกษาเปรียบเทียบกับงานวิจัยในอดีตที่ทำการแบบจำลองพยากรณ์ด้วยวิธีการอื่น

วัตถุประสงค์การวิจัย

เพื่อทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร

การทบทวนวรรณกรรม

ความก้าวหน้าของเทคโนโลยี ทำให้ผู้ประกอบการธุรกิจอสังหาริมทรัพย์จัดเก็บข้อมูลและสารสนเทศในรูปแบบอิเล็กทรอนิกส์มากขึ้น แต่การนำข้อมูลมาใช้ประโยชน์ยังไม่สามารถทำได้อย่างเต็มที่ เนื่องจากข้อมูลมีจำนวนมาก จึงมีความจำเป็นที่จะต้องสกัดหาองค์ความรู้ที่เป็นประโยชน์ต่อองค์กรเพื่อสร้างโอกาสและเพิ่มมูลค่าในการแข่งขัน ดังเช่นในต่างประเทศที่มีการศึกษาเกี่ยวกับการนำปัญญาประดิษฐ์มาใช้ในอุตสาหกรรมอสังหาริมทรัพย์ด้านต่าง ๆ เช่น การค้นหาอสังหาริมทรัพย์ (Home Search Experience) การประเมินราคา (Property Value Prediction) การศึกษาพฤติกรรมของผู้บริโภค (Lead Generation) การเพิ่มโอกาสในการกู้ (Efficient Mortgage Lending) (Kuflinski, 2020) อาทิ บริษัท Zillow จำกัด ซึ่งเป็นตัวอย่างหนึ่งของบริษัทที่มีการนำปัญญาประดิษฐ์มาใช้ในอุตสาหกรรมอสังหาริมทรัพย์ โดยได้พัฒนาโมเดลชื่อ Zestimate เพื่อใช้ประเมินราคาอสังหาริมทรัพย์ที่ประกาศขายอยู่ในเว็ปไซต์ของตนเอง ช่วยให้ผู้บริโภคสามารถมีข้อมูลในการตัดสินใจซื้ออสังหาริมทรัพย์ได้อย่างมีประสิทธิภาพมากขึ้น

เหมืองข้อมูล (Data Mining) เป็นเทคนิคหนึ่งในการวิเคราะห์ข้อมูลเพื่อค้นหารูปแบบ (Pattern) ความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติและการเรียนรู้ของเครื่องคอมพิวเตอร์

แล้วนำผลวิเคราะห์ที่ได้ไปใช้ประโยชน์เพื่อการตัดสินใจ กระบวนการในการทำเหมืองข้อมูลใช้โมเดลในการเรียนรู้ต้องนำอัลกอริทึมของการเรียนรู้ของเครื่อง (Machine Learning) มาใช้ โดยชนิดของอัลกอริทึมในการทำเหมืองข้อมูล (Data Mining Algorithms) (Yang, 2019) ทั้งนี้ การเรียนรู้ของเครื่อง (Machine Learning) เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: Ai) ซึ่งเป็นกระบวนการพัฒนาอัลกอริทึมที่สามารถเรียนรู้และปรับตัวตามข้อมูลที่ได้รับ โดยอาศัยแบบจำลอง (Model) ที่สร้างจากชุดข้อมูลที่มีความสัมพันธ์กัน และสามารถนำไปทำนายหรือใช้ในการตัดสินใจภายหลัง ประเภทการเรียนรู้ของเครื่องได้ผ่านการสังเคราะห์ โดยแบ่งรูปแบบการเรียนรู้ออกเป็น 3 รูปแบบ (Alpaydin, 2014; Armstrong, 2015; Cates et al., 2017; Yang, 2019) ได้แก่ (1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) ผลลัพธ์ที่ได้ คือ การจัดหมวดหมู่ (Classification) และการวิเคราะห์การถดถอย (Regression) (2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ผลลัพธ์ที่ได้ คือ การลดมิติข้อมูล (Dimensionality Reduction) การจัดกลุ่ม (Clustering) และการเรียนรู้แบบซับซ้อน (Manifold Learning) และ (3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นการเรียนรู้แบบไม่มีผู้สอน แต่จะมีปฏิสัมพันธ์กับสิ่งแวดล้อม ทำให้เกิดการกระตุ้นในการสร้างแบบจำลองตามสภาพแวดล้อมในขณะนั้น

จากการที่เทคนิคการเรียนรู้ของเครื่องมีหลากหลายรูปแบบ การวิจัยนี้จึงมีวัตถุประสงค์เพื่อทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร เพื่อช่วยให้นักลงทุนหรือผู้ที่สนใจซื้อทรัพย์สินรอการขายประเภทบ้านเดี่ยวทราบปัจจัยที่มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร รวมถึงใช้เป็นข้อมูลประกอบการตัดสินใจลงทุนธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ ตลอดจนใช้เป็นแนวทางในการกำหนดราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครได้อย่างเหมาะสมกับลักษณะทางกายภาพของทรัพย์สินรอการขาย โดยทำการทบทวนวรรณกรรมที่เกี่ยวข้องกับเทคนิคการเรียนรู้ของเครื่องที่ใช้ในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร ดังนี้

1. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นอัลกอริทึมในการคัดแยกกลุ่มเพื่อจัดประเภทหรือจำแนกประเภทข้อมูลที่มีการนำมาใช้ในด้านการประมวลผลภาพ เป็นวิธีการจำแนกกลุ่มข้อมูลที่อาศัยระนาบการตัดสินใจที่เรียกว่าระนาบเกินหรือไฮเปอร์เพลน (Hyperplane) มาใช้ในการจำแนกกลุ่มข้อมูลโดยใช้สมการเส้นตรงในการแบ่งข้อมูลออกเป็น 2 กลุ่มแยกออกจากกัน (Saigal, 2020) ซัพพอร์ตเวกเตอร์แมชชีนมีรูปแบบในการเรียนรู้เป็นกระบวนการเลือกแบบจำลองที่เหมาะสมที่สุดจะทำให้ได้ค่าที่เหมาะสมที่สุดเป็นคำตอบ จึงเป็นที่นิยมและถูกนำไปใช้ในงานด้านการรู้จำรูปแบบ ทั้งนี้การเลือกใช้ซัพพอร์ตเวกเตอร์แมชชีนแบบแบ่งกลุ่มซัพพอร์ตเวกเตอร์แมชชีนสำหรับการแบ่งกลุ่มข้อมูลนั้น

จะใช้ระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane)ในการแบ่งกลุ่ม ซึ่งการสร้างระนาบเกินที่ใช้ในการแบ่งกลุ่มข้อมูลสามารถสร้างได้หลายแบบ แต่จะมีระนาบเกินที่เหมาะสมที่สุดเพียงระนาบเดียวเท่านั้นที่สามารถรักษาระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่มที่ใกล้กันมากที่สุดได้

2. เทคนิคต้นไม้กาเดียนบูทสเต็ท (Gradient Boosting Tree: GBT)

Gradient Boosting Tree (GBT) เป็นการสร้างโมเดลการตัดสินใจแบบลำดับขั้น (Sequence) โดย Input ในแต่ละ Tree จะเป็น Output จาก Tree ก่อนหน้า โดยมีแนวคิด คือ GBT จะทำการสร้างแต่ละ Tree เพื่อลดค่า Error ที่เกิดจาก Tree ก่อนหน้า ก่อนนำผลลัพธ์ที่ได้มารวมกัน ซึ่งมีข้อดี คือ Bias และ Variance ลดลง เนื่องจาก Error ถูกแก้ไข ข้อเสีย คือ มีค่าตัวแปรหลายตัวที่ต้องปรับเพื่อให้ได้ค่า Performance ที่ดี และเลี่ยงการเกิด Over–Fitting Gradient Boosting Tree (GBT) เป็นโมเดลที่นำเอา Decision Tree มาทำการ Train ต่อกันหลาย ๆ รอบ โดยที่ Decision Tree จะเรียนรู้จากความผิดพลาดของต้นไม้ก่อนหน้า โดยให้น้ำหนักที่แตกต่างกันออกไป ถ้า Classifier ตัวใดที่มีการทำนายผิดบ่อย ๆ จะให้น้ำหนักมากกว่าเพื่อให้มีการสุ่มขึ้นมาใหม่อีกครั้ง และทำการทำนายค่าไปเรื่อย ๆ ซึ่งจะทำให้มีความแม่นยำขึ้นเรื่อย ๆ ในแต่ละรอบ และโมเดลจะทำการหยุดก็ต่อเมื่อมีการ Train ที่ลึกเกินไป และโมเดลไม่สามารถหารูปแบบของความผิดพลาดได้จากต้นไม้ก่อนหน้าได้หรือได้ผลรวมของ Classifier ที่ดีที่สุด

3. โครงข่ายประสาทเทียม (Artificial Neural Network)

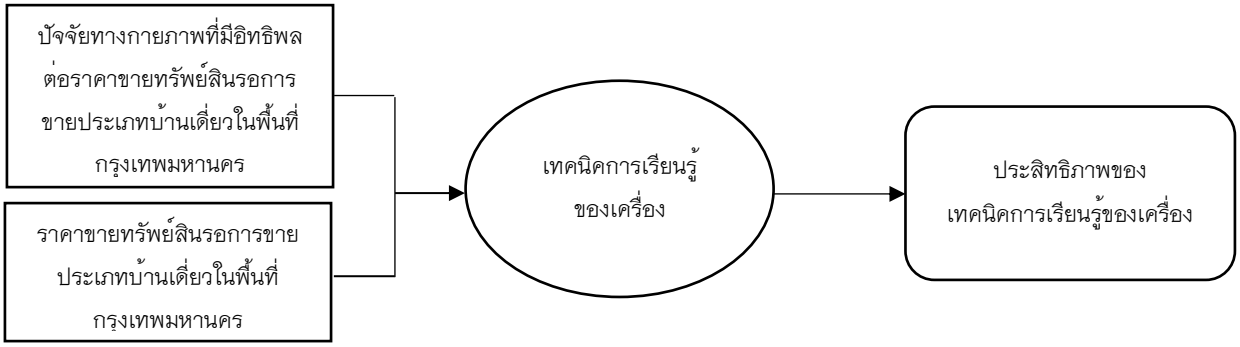
โครงข่ายประสาทเทียม (Artificial Neural Network) เป็นวิธีการที่มีพื้นฐานมาจากการเลียนแบบการทำงานของสมอ มนุษย์ซึ่งประกอบด้วยนิวรอน (Neurons) ที่ใช้ในการเชื่อมโยงเพื่อแก้ปัญหาต่าง ๆ (Dastres & Soori, 2021) โดยโครงข่ายประสาทเทียมจะนำข้อมูลเข้าและค่าน้ำหนักแต่ละปลายประสาทมาคูณกัน และนำค่าทั้งหมดมารวมกัน หากค่าที่รวมกันนี้เกินกว่าค่าแบ่งที่กำหนดไว้ (Threshold: T) จะทำการส่งต่อข้อมูลไปยังฟังก์ชันกระตุ้น โดยฟังก์ชันกระตุ้นหรือฟังก์ชันการส่งต่อ (Transfer Function) เป็นฟังก์ชันในการคำนวณ เพื่อทำนายค่าของข้อมูลออก รูปแบบที่นิยมมากที่สุดและมีประโยชน์ คือ รูปแบบที่ไม่เป็นฟังก์ชันสมการเส้นตรง (Non–Linear Function) เนื่องจากปัญหาในโลกความเป็นจริงมีลักษณะเป็นแบบสมการเส้นตรงน้อยมาก (Minsky, Papert, & Bottou, 2017) ฟังก์ชันกระตุ้นทำหน้าที่ในการตัดสินใจว่านิวรอนควรจะถูกกระตุ้นหรือไม่ โดยดูค่าผลรวมของข้อมูลเข้าและค่าน้ำหนัก ฟังก์ชันกระตุ้นจะถูกนำไปใช้ทั้งโหนดซ่อน (Hidden Node) และโหนดข้อมูลออก (Output Node) ซึ่งทั้ง 2 โหนดอาจจะใช้ฟังก์ชันกระตุ้นที่เหมือนหรือต่างกันก็ได้ แต่ส่วนใหญ่จะใช้ฟังก์ชันแบบไม่เป็นเชิงเส้น เนื่องจากในโหนดซ่อนจะมีการคำนวณแบบการรวมเชิงเส้น (Linear Combination) ถ้าฟังก์ชันกระตุ้นของโหนดซ่อนจะมีการคำนวณแบบเชิงเส้นอีก จะเป็นการทำงานซ้ำซ้อนกับการคำนวณแบบการรวมเชิงเส้นในชั้นข้อมูลออก และจะทำให้ผลลัพธ์เทียบเท่ากับสมการถดถอยโลจิสติกส์

4. เทคนิคการทำเหมืองข้อมูลแบบรวมกลุ่ม (Ensemble Vote)

เทคนิคการทำเหมืองข้อมูลแบบรวมกลุ่ม (Ensemble Vote) คือ เทคนิคของ Advanced Machine Learning Model ที่ช่วยทำให้ประสิทธิภาพการทำนายผล (Prediction Performance) โดยการใช้เทคนิคที่ใช้โมเดล Classification หลายโมเดล มาช่วยในการหาคำตอบ ซึ่งเป็นเทคนิคที่มีประสิทธิภาพสูง โดยมีอัลกอริทึมที่พัฒนาบนแนวคิด เทคนิค Ensemble ที่ถูกใช้กันมาก โดย Vote Ensemble เป็นการใช้ Training Data ชุดเดียวกัน แต่สร้างโมเดลด้วยเทคนิคหลากหลาย จึงเป็นการเลือกที่จะสร้างโมเดลด้วยเทคนิคการจำแนกข้อมูล (Classification) ที่แตกต่างกัน คือ หลังจากได้โมเดลมาชุดหนึ่งแล้วจะทำการนำไปทำนายข้อมูลและนำคำตอบมารวมกันเพื่อดูว่าคำตอบใดมีความเหมาะสมมากที่สุด โดยใช้วิธีการโหวต (Vote) คือ เลือกคำตอบที่ตอบตรงกันมากที่สุด

กรอบแนวคิดการวิจัย

งานวิจัยนี้ เป็นการวิจัยเชิงปริมาณ ผู้วิจัยกำหนดกรอบแนวคิดการวิจัย โดยมีรายละเอียดดังนี้



ภาพที่ 1 กรอบแนวคิดในการวิจัย

ระเบียบวิธีวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงปริมาณ พื้นที่วิจัย คือ กรุงเทพมหานคร ประชากร คือ ทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่ประกาศขายในเว็บไซต์ของธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ ซึ่งไม่ทราบจำนวนประชากรที่แน่นอน กลุ่มตัวอย่าง คือ ทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่ประกาศขายในเว็บไซต์ของธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ จำนวน 446 ตัวอย่าง โดย Hair et al. (2010) กล่าวว่า การวิเคราะห์สมการการถดถอยควรมีขนาดตัวอย่างอย่างน้อย 10 – 20 ตัวอย่างต่อจำนวนตัวแปรต้น 1 ตัวแปร แต่ต้องไม่ต่ำกว่า 100 ตัวอย่าง ทั้งนี้ ใช้วิธีการคัดเลือกแบบสะดวก โดยเลือกทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่มีข้อมูล

ครบถ้วนตามตัวแปรที่กำหนด เครื่องมือที่ใช้ในการวิจัย มี 1 ชนิด ได้แก่ แบบตรวจสอบรายการ รวบรวมข้อมูลเว็บไซต์ของธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ ระหว่างเดือนเมษายน ถึงเดือนพฤษภาคม พ.ศ. 2565 นำข้อมูลเชิงปริมาณมาวิเคราะห์ด้วยสถิติเชิงพรรณนา และใช้เทคนิคการเรียนรู้ของเครื่อง ได้แก่ เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN) และ Ensemble Vote ด้วยโปรแกรม Rapid Miner Studio

ทั้งนี้ จากการสำรวจข้อมูลทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่ประกาศขายในเว็บไซต์ของธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ พบว่า ปัจจัยทางกายภาพที่มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครที่มีข้อมูลทั้งจากเว็บไซต์ของธนาคารพาณิชย์และเว็บไซต์ของบริษัทบริหารสินทรัพย์ มีทั้งสิ้น 8 ตัวแปร ประกอบด้วย ที่ตั้งทรัพย์ เนื้อที่ดิน ขนาดพื้นที่ใช้สอย จำนวนห้องนอน จำนวนห้องน้ำ จำนวนที่จอดรถ แปลงหัวมุมถนน และสภาพทรัพย์สินรอการขาย ส่วนตัวแปรตามในการวิจัยครั้งนี้ ได้แก่ ราคาขายทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานคร

ผลการวิจัย

ผู้วิจัยทำการเก็บรวบรวมข้อมูลปัจจัยทางกายภาพที่มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานครและราคาเสนอขายทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานครที่ประกาศเสนอชายในเว็บไซต์ของธนาคารพาณิชย์และบริษัทบริหารสินทรัพย์ จากกลุ่มตัวอย่างทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานครประเภทบ้านเดี่ยว จำนวน 446 ตัวอย่าง พบว่ากลุ่มตัวอย่างทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครส่วนใหญ่ตั้งอยู่ในโซนพื้นที่กรุงเทพมหานครชั้นกลางและชั้นนอก ไม่ได้อยู่แปลงหัวมุมถนน และประกาศขายตามสภาพ โดยกลุ่มตัวอย่างทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร มีขนาดเนื้อที่ดินอยู่ระหว่าง 18.10 – 634.00 ตารางวา ขนาดพื้นที่ใช้สอยอยู่ระหว่างอยู่ระหว่าง 44.00 – 1,788 ตารางเมตร จำนวนห้องนอนอยู่ระหว่าง 1 – 10 ห้อง จำนวนห้องน้ำอยู่ระหว่าง 1 – 12 ห้องน้ำ จำนวนที่จอดรถอยู่ระหว่าง 0 – 8 คัน และมีราคาประกาศขายอยู่ระหว่าง600,000 – 119,700,000 บาท

ผู้วิจัยจะทำการคัดเลือกและแปลงข้อมูลโดยใช้เทคนิคเหมืองข้อมูล (Data Mining) และการเรียนรู้ของเครื่อง (Machine Learning) ในการสร้างสมการพยากรณ์ (Predictive Modeling) ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร ด้วยการสร้างแบบจำลองโดยใช้เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN) และ Ensemble Vote ด้วยโปรแกรม Rapid Miner Studio สามารถแสดงผลการศึกษาจำแนกตามเทคนิคที่ใช้ในการสร้าง

สมการพยากรณ์ (Predictive Modeling) ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร โดยมีตัวแปรอิสระ จำนวน 9 attribute ซึ่งเป็นปัจจัยทางกายภาพที่มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานคร

ผู้วิจัยกำหนด Operator โดย Retrieve Data เพื่อนำชุดข้อมูลที่สร้างขึ้นสำหรับการสร้างสมการพยากรณ์ หลังจากนั้นทำการเลือก Validation เพื่อกำหนดสัดส่วนของชุดข้อมูลฝึกสอน (Train) และชุดข้อมูลทดสอบ (Test) ในส่วนของ Split Ratio โดยผู้วิจัยกำหนดค่าสำหรับชุดข้อมูลฝึกสอน (Train) คือ ร้อยละ 80 จึงทำให้มีชุดข้อมูลฝึกสอน (Train) จำนวน 357 ชุดข้อมูล และชุดข้อมูลทดสอบ (Test) จำนวน 89 ชุดข้อมูล (ภาพที่ 2) นอกจากนี้ ผู้วิจัยได้ทำการใช้ฟังก์ชันเคอร์เนล (Kernel Function) เพื่อแปลงข้อมูลไปยังมิติที่สูงขึ้นในปริภูมิคุณลักษณะ (Feature Space) สามารถจำแนกข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ จากนั้นทำการจัดเตรียมข้อมูลใน Spreadsheet และสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทที่อยู่อาศัยในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่อง ได้แก่ เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN) และ Ensemble Vote ด้วยโปรแกรม Rapid Miner Studio ตลอดจนทำการทดสอบประสิทธิภาพของแบบจำลองด้วยการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง

1. การสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิค Support Vector Machine (SVM)

ผู้วิจัยทำการเลือก Operator Support Vector Machine (SVM) จากนั้นเลือก Operator Apply Model และ Performance Regression ลากเส้นเชื่อมโยง Operator และทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Support Vector Machine (SVM) ผู้วิจัยทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Support Vector Machine (SVM) ซึ่งกำหนดค่าพารามิเตอร์โดยใช้ค่าตั้งต้น (Default Parameter) ทั้งนี้ เมื่อพิจารณา Performance ของการสร้างสมการพยากรณ์โดยใช้เทคนิค Support Vector Machine (SVM) สามารถแสดงค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ค่าความผิดพลาดสัมพันธ์ (Relative Error) และค่าสหสัมพันธ์ (Correlation) สามารถแสดงได้ดังตารางที่ 1

ตารางที่ 1 ประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Support Vector Machine

ประสิทธิภาพ	**ค่าที่ได้**
Root Mean Square Error	5,870,721.056
Relative Error	78.70%
Correlation	0.800

จากตารางที่ 1 แสดงประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Support Vector Machine (SVM) โดยพบว่าค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เท่ากับ 5,870,721.056 ค่าความผิดพลาดสัมพัทธ์ (Relative Error) เท่ากับร้อยละ 78.70 และค่าสหสัมพันธ์ (Correlation) เท่ากับ 0.800

2. การสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิค Gradient Boosted Trees (GBT)

ผู้วิจัยทำการเลือก Operator Gradient Boosted Trees (GBT) จากนั้นเลือก Operator Apply Model และ Performance Regression ลากเส้นเชื่อมโยง Operator และทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Gradient Boosted Trees (GBT) ผู้วิจัยทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Gradient Boosted Trees (GBT) ซึ่งกำหนดค่าพารามิเตอร์โดยใช้ค่าตั้งต้น (Default Parameter) ทั้งนี้ เมื่อพิจารณา Performance ของการสร้างสมการพยากรณ์โดยใช้เทคนิค Gradient Boosted Trees (GBT) สามารถแสดงค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ค่าความผิดพลาดสัมพันธ์ (Relative Error) และค่าสหสัมพันธ์ (Correlation) สามารถแสดงได้ดังตารางที่ 2

ตารางที่ 2 ประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Gradient Boosted Trees (GBT)

ประสิทธิภาพ	**ค่าที่ได้**
Root Mean Square Error	4,301,866.855
Relative Error	107.86%
Correlation	0.774

จากตารางที่ 2 แสดงประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Gradient Boosted Trees (GBT) โดยพบว่าค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เท่ากับ 4,301,866.855 ค่าความผิดพลาดสัมพัทธ์ (Relative Error) เท่ากับร้อยละ 107.86 และค่าสหสัมพันธ์ (Correlation) เท่ากับ 0.774

3. การสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิค Artificial Neural Network (ANN)

ผู้วิจัยทำการเลือก Operator Artificial Neural Network (ANN) จากนั้นเลือก Operator Apply Model และ Performance Regression ลากเส้นเชื่อมโยง Operator และทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Artificial Neural Network (ANN) ผู้วิจัยทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Artificial Neural Network (ANN) ซึ่งกำหนดค่าพารามิเตอร์โดยใช้ค่าตั้งต้น (Default Parameter) ทั้งนี้ เมื่อพิจารณา Performance ของการสร้างสมการพยากรณ์โดยใช้เทคนิค Artificial Neural Network (ANN) สามารถแสดงค่าความ

คลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ค่าความผิดพลาดสัมพันธ์ (Relative Error) และค่าสหสัมพันธ์ (Correlation) สามารถแสดงได้ดังตารางที่ 3

ตารางที่ 3 ประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Artificial Neural Network (ANN)

ประสิทธิภาพ	**ค่าที่ได้**
Root Mean Square Error	5,339,240.604
Relative Error	116.51%
Correlation	0.824

จากตารางที่ 3 แสดงประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Gradient Boosted Trees (GBT) โดยพบว่าค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เท่ากับ 5,339,240.604ค่าความผิดพลาดสัมพัทธ์ (Relative Error) เท่ากับร้อยละ 116.51 และค่าสหสัมพันธ์ (Correlation) เท่ากับ 0.824

4. การสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิค Ensemble Vote

ผู้วิจัยทำการเลือก Operator Ensemble Vote จากนั้นเลือก Operator Apply Model และ Performance Regression ลากเส้นเชื่อมโยง Operator และทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Ensemble Vote ผู้วิจัยทำการสร้างสมการพยากรณ์โดยใช้เทคนิค Ensemble Vote ทั้งนี้ เมื่อพิจารณา Performance ของการสร้างสมการพยากรณ์โดยใช้เทคนิค Ensemble Vote สามารถแสดงค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ค่าความผิดพลาดสัมพันธ์ (Relative Error) และค่าสหสัมพันธ์ (Correlation) สามารถแสดงได้ดังตารางที่ 4

ตารางที่ 4 ประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Ensemble Vote

ประสิทธิภาพ	**ค่าที่ได้**
Root Mean Square Error	3,746,335.580
Relative Error	91.44%
Correlation	0.836

จากตารางที่ 4 แสดงประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Ensemble Vote โดยพบว่าค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) เท่ากับ 3,746,335.580 ค่าความผิดพลาดสัมพัทธ์ (Relative Error) เท่ากับร้อยละ 91.44 และค่าสหสัมพันธ์ (Correlation) เท่ากับ 0.836

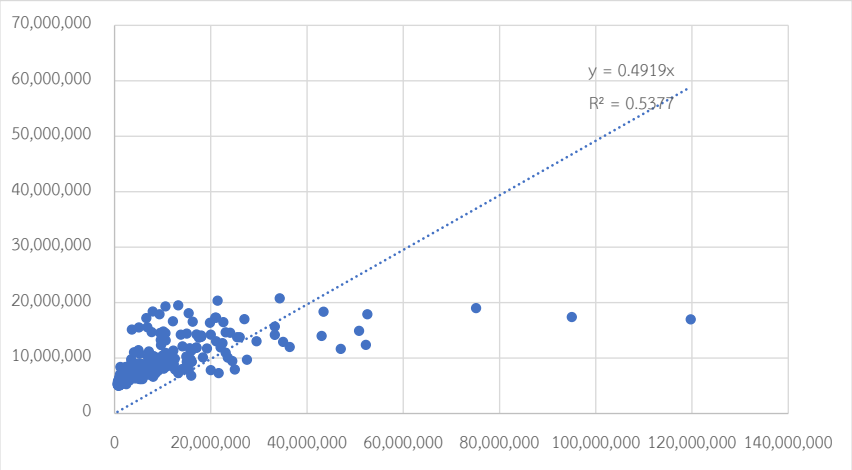
5. สรุปผลการเปรียบเทียบประสิทธิภาพของสมการพยากรณ์จำแนกตามเทคนิคการเรียนรู้ของเครื่องที่ใช้

ตารางที่ 5 ประสิทธิภาพของสมการพยากรณ์จำแนกโดยใช้เทคนิค Ensemble Vote

สมการพยากรณ์	ประสิทธิภาพของสมการพยากรณ์		
	Root Mean Square Error	Relative Error	Correlation
เทคนิค SVM	5,870,721.056	78.70%	0.800
เทคนิค GBT	4,301,866.855	107.86%	0.774
เทคนิค ANN	5,339,240.604	116.51%	0.824
เทคนิค Ensemble Vote	3,746,335.580	91.44%	0.836

จากตารางที่ 5 เมื่อพิจารณาจากค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) พบว่าสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote มีค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) น้อยที่สุดเมื่อเทียบกับสมการพยากรณ์ที่ได้จากการใช้เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT) และ Artificial Neural Network (ANN) จึงเป็นสมการพยากรณ์ที่มีประสิทธิภาพสูงกว่าสมการณ์พยากรณ์ที่ได้จากการใช้เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT) และ Artificial Neural Network (ANN) เนื่องจากค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) มีค่ายิ่งต่ำยิ่งดี เพราะแสดงให้เห็นถึงความสามารถในการทำนายราคาขายของทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครได้มากขึ้น

นอกจากนี้ ผู้วิจัยได้ทำการหาค่า R^2 และ Beta ของสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote ซึ่งสามารถแสดงได้ดังภาพที่ 2



ภาพที่ 2 ค่า R^2 และ Beta ของสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote

จากภาพที่ 2 ผู้วิจัยได้ทำการหาค่า R^2 และ Beta ของสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote พบว่าสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote มีค่า R^2 เท่ากับ 0.5377 แสดงว่าค่าความผันแปรของตัวแปรตอบสนองที่สามารถอธิบายได้มีอยู่ในสมการพยากรณ์นี้ เท่ากับร้อยละ 53.77 และค่า Beta เท่ากับ 0.4919 แสดงว่าตัวแปรอิสระที่อยู่ในสมการพยากรณ์ได้จากการใช้เทคนิค Ensemble Vote มีอิทธิพลต่อราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร เท่ากับร้อยละ 49.19

อภิปรายผลการวิจัย

ผลการทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร พบว่า สมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote เป็นสมการที่มีประสิทธิภาพมากที่สุดในการพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร เนื่องจากมีค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) น้อยที่สุด ทั้งนี้ เป็นเพราะมีการนำตัวจำแนกข้อมูลที่หลากหลายมารวมกันด้วยวิธีการรวมกลุ่มของตัวจำแนก ซึ่งช่วยลดปัญหาการเกิดไบแอส (Bias) ของข้อมูลโดยตัวจำแนกที่ดีแต่ละตัวจะสามารถช่วยกันเสริมประสิทธิภาพในการจำแนกข้อมูล ทำให้โมเดลที่ได้มีประสิทธิภาพมากยิ่งขึ้น ผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของปรเมษฐ์ ธันวานนท์ และคณะ (Thanwanont, 2017) ที่พบว่าการประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย 4 เทคนิค และเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Model) เพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์โดยใช้ข้อมูลของตลาดหลักทรัพย์แห่งประเทศไทยในช่วงเดือนมกราคม 2554 – ธันวาคม 2559 โดยเทคนิค Ensemble Model ด้วยวิธีการถ่วงน้ำหนักสามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ถึงร้อยละ 5 – 14

องค์ความรู้ใหม่จากการวิจัย

ผลการทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร

การสร้างสมการพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่องที่หลากหลาย เป็นการเพิ่มประสิทธิภาพโดยรวมของสมการพยากรณ์ เนื่องจากการนำตัวจำแนกข้อมูลที่หลากหลายมารวมกันด้วยวิธีการรวมกลุ่มสามารถช่วยลดปัญหาการเกิดไบแอสของข้อมูลได้ โดยสามารถช่วยกันเสริมประสิทธิภาพในการจำแนกข้อมูล

ทำให้โมเดลที่ได้มีประสิทธิภาพมากขึ้น องค์กรต่าง ๆ สามารถนำเทคนิคการเรียนรู้ของเครื่องไปใช้วิเคราะห์ข้อมูลเพื่อแสวงหาผลประโยชน์จากข้อมูลประกอบการตัดสินใจ ทั้งนี้ เทคนิคการเรียนรู้ของเครื่องสามารถนำมาใช้ในการสร้างสมการพยากรณ์ได้โดยใช้ข้อมูลที่สามารถหาได้จากเครือข่ายอินเทอร์เน็ต แม้ว่าจะมีข้อมูลจำนวนไม่มากนัก

สรุป

เมื่อพิจารณาจากค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) พบว่าสมการพยากรณ์ที่ได้จากการใช้เทคนิค Ensemble Vote มีค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) น้อยที่สุดเมื่อเทียบกับสมการพยากรณ์ที่ได้จากการใช้เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT) และ Artificial Neural Network (ANN) จึงเป็นสมการพยากรณ์ที่มีประสิทธิภาพสูงกว่าสมการณ์พยากรณ์ที่ได้จากการใช้เทคนิค Support Vector Machine (SVM), Gradient Boosted Trees (GBT) และ Artificial Neural Network (ANN) เนื่องจากความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) มีค่ายิ่งต่ำยิ่งดี เพราะแสดงให้เห็นถึงความสามารถในการทำนายราคาขายของทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครได้มากขึ้น

ข้อเสนอแนะ

จากผลการวิจัย ผู้วิจัยมีข้อเสนอแนะ ดังนี้

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

ผู้ประกอบการธุรกิจอสังหาริมทรัพย์สามารถนำเทคนิคการการเรียนรู้ของเครื่องไปใช้ในการจัดการข้อมูลวิเคราะห์ข้อมูลที่มี เพื่อแสวงหาผลประโยชน์จากข้อมูลประกอบการตัดสินใจต่าง ๆ เนื่องจากเทคนิคการการเรียนรู้ของเครื่องสามารถนำมาใช้ในการสร้างสมการพยากรณ์ได้แม้ว่าจะมีข้อมูลไม่มากนัก

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

งานวิจัยนี้ได้ข้อค้นพบว่าเทคนิคการเรียนรู้ของเครื่องสามารถนำมาใช้ในการสร้างสมการพยากรณ์ได้โดยใช้ข้อมูลที่สามารถหาได้จากเครือข่ายอินเทอร์เน็ตแม้ว่าจะมีข้อมูลจำนวนไม่มากนัก สำหรับประเด็นในการวิจัยครั้งต่อไปควรทำวิจัยในประเด็นเกี่ยวกับทรัพย์สินรอการขายประเภทอื่น

การวิจัยครั้งนี้เลือกใช้เทคนิคในการพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครเพียง 4 เทคนิค คือ Support Vector Machine (SVM), Gradient Boosted Trees (GBT), Artificial Neural Network (ANN) และ Ensemble Vote ซึ่งในความเป็นจริงแล้ว ยังมีเทคนิคอีกหลายเทคนิค

ที่น่าสนใจ ผู้สนใจศึกษาจึงควรนำเทคนิคอื่นมาพิจารณาร่วมด้วยเพื่อทำให้การเปรียบเทียบความแม่นยำของสมการพยากรณ์มีประสิทธิภาพมากยิ่งขึ้น

References

Alpaydin, E. (2014). *Introduction to machine learning* (3rd ed). MIT Press.

Armstrong, H. (2015). *Machines Thai learn in the wild*. Nesta.

Berry, M. W., Mohamed, A., & Yap, B. W. (2020). *Supervised and unsupervised learning for data science.* Springer.

Cates, S., Lawrence, S., Penedo, C., & Samatova, V. (2017). A machine learning approach to research curation for investment process. *Journal of Investment Management, 15*(1), 39–49.

Dastres, R., & Soori, M. (2021). Artificial Neural Network systems. *International Journal of Imaging and Robotics (IJIR), 21*(2), 13–25.

Hair, J., et al. (2010). *Multivariate data analysis* (7th ed.). Pearson Education International.

Jiamwiriyakul, B. (2019). *Portfolio return prediction in property development sector using neural network* [Master' s Independent Study, Mahidol University].

Kuflinski, Y. (2020). *6 ways artificial intelligence is reshaping real estate.* Iflexion. https://www.iflexion.com/blog/artificial–intelligence–real–estate.

Minsky, M., Papert, S. A., & Bottou, L. (2017). *Perceptrons: An introduction to computational geometry*. MIT Press.

Real Estate Information Center. (2021). *The situation of the second–hand housing market across the country in the 4th quarter of 2020.* Real Estate Information Center. https://www.reic.or.th/Upload/PressRelease210304_32354_1614874695_ 87218.pdf.

Saigal, P. (2020). *Support Vector Machines: Evolution and applications.* Nova Science Publishers.

Sawangum, W. (2022). *The online consumer hits for condominium photo using machine learning technique* [Master' s Thematic Paper, Dhurakij Pundit University].

Sunkpho, J., & Ramjan, S. (2020). Predicting condominium price in Bangkok using web mining techniques. *Srinakharinwirot Research and Development Journal of Humanities and Social Sciences, 12*(24), 15–27.

Thanwanont, P., Yingseri, C., Pongpetch, W., & Kangkajit, T. (2017). Predict stock price trends in stock exchange of Thailand using ensemble model. *Journal of Information Science and Technology, 7*(1), 12–21.

Tochaiwat, K. (2023). *Machine learning transforms the real estate business.* Bangkokbiznews. https://www.bangkokbiznews.com/tech/innovation/1088762

Vongsung, S. (2022). *A study of investment model for residential real estate development: A case study of Nakhon Si Thammarat Province* [Master' s Independent Study, Prince of Songkla University].

Yang, X.S. (2019). *Introduction to algorithms for data mining and machine learning.* Candice Janco.