การเปรียบเทียบประสิทธิภาพการทำนายจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทยโดยใช้วิธีการวิเคราะห์ทางสถิติและการเรียนรู้ของเครื่อง

ณัฐธิดา ศรีศิลปอุดม[1], ศุภกานต์ วงศาธรรมกุล[1], นลิณี จตุรพรชัยรักษา[1*] และ สุปราณี ลิสวัสดิ์[1]

[1]สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
*nalinee.chat@dome.tu.ac.th

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบหาวิธีที่ดีที่สุดสำหรับการพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางและกรุงเทพมหานคร ด้วยวิธีการวิเคราะห์การถดถอย (Regression Analysis) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคป่าสุ่ม (Random Forest) โดยใช้ข้อมูลจากสำนักงานสถิติแห่งชาติ ซึ่งเป็นข้อมูลย้อนหลัง 14 ปี ตั้งแต่ปี พ.ศ. 2550-2563 มีจำนวนปัจจัยที่เกี่ยวข้องทั้งหมด 7 ปัจจัย และข้อมูลจำนวน 2,058 ข้อมูล โดยจะทำการแบ่งข้อมูลในการวิเคราะห์ออกเป็นสองชุด คือ ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร และข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร โดยการวิเคราะห์ข้อมูลใช้โปรแกรมสำเร็จรูปทางสถิติ วิเคราะห์ปัจจัยที่มีความเกี่ยวข้อง ด้วยวิธีการลดตัวแปร (Backward Elimination) และทดสอบประสิทธิภาพด้วยโปรแกรม WEKA โดยอัลกอริทึมที่ใช้ในเทคนิคต้นไม้ตัดสินใจ ได้แก่ อัลกอริทึม REPTree ใช้เกณฑ์ในการวัดประสิทธิภาพ 2 เกณฑ์ ประกอบด้วย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) และค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) ผลการวิจัยพบว่าปัจจัยที่ส่งผลต่อข้อมูลในภาคกลาง มีจำนวน 3 ปัจจัย ได้แก่ อัตราเงินเฟ้อ รายได้เฉลี่ยต่อครัวเรือน จำนวนประชากร และปัจจัยที่ส่งผลต่อข้อมูลในกรุงเทพมหานคร มีจำนวน 2 ปัจจัย ได้แก่ อัตราเงินเฟ้อ จำนวนคนว่างงาน ในส่วนของผลการทดสอบประสิทธิภาพวิธีที่ดีที่สุดของชุดข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานครและข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร ได้แก่ วิธีการวิเคราะห์การถดถอย มีค่า RMSE เท่ากับ 0.3193 และ 1032.4407 คดี และค่า MAE เท่ากับ 0.2405 และ 858.1238 คดี ตามลำดับ งานวิจัยนี้สามารถนำไปใช้ในการพยากรณ์จำนวนคดีอาญาเพื่อหาแนวทางในการรับมือและแก้ไขสถานการณ์ที่จะเกิดขึ้นในอนาคตต่อไป

คำสำคัญ: คดีอาญา ต้นไม้ตัดสินใจ เทคนิคป่าสุ่ม การวิเคราะห์การถดถอย ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์

Efficiency Comparison of Property Crimes in Thailand by using Statistical Analysis and Machine Learning

Nuttida Srisilapa-udom[1], Supakarn Wongsathammakul[1], Nalinee chaturaphonchairaksa[1*] and Supranee Lisawadi[1]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University
*nalinee.chat@dome.tu.ac.th

Abstract

The purpose of this research is to compare the most suitable methods for forecasting the number of criminal cases against property in the central region and Bangkok by using three techniques, regression analysis method, decision tree method, and random forest method. The data was collected by the National Statistical Office which is a 14-year historical data from 2007-2020 with seven relevant factors and 2,058 data. The data for analysis was divided into two sets, the number of criminal cases in the central region excluding Bangkok and the number of criminal cases in Bangkok. Statistical package was used to analyze related factors by using backward elimination method and testing the efficiency with the WEKA program. Algorithm, which was used in the decision tree is the REPTree algorithm including the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE) for comparing the prediction efficiency. The results showed that there were 3 factors such as inflation rate, average monthly household income (AMHI), and population which affected data in the central region and 2 factors such as inflation rate and unemployment rate which affected data in Bangkok. As for performance testing, the best method is regression analysis with RMSE value equaled to 0.3193 and 1032.4407 cases, and MAE value was 0.2405 and 858.1238 cases, respectively. This research can be used in forecasting the number of criminal cases in order to find procedures to deal with and solve situations that would occur in the future.

Keyword: Criminal Case, Decision Tree, Random Forest, Regression Analysis, Mean Absolute Error

บทนำ

อาชญากรรมเป็นปัญหาที่สำคัญปัญหาหนึ่งที่เกี่ยวข้องกับการอยู่ร่วมกันของคนในสังคม ซึ่งเกิดขึ้นอย่างต่อเนื่องตั้งแต่ในอดีตจนถึงปัจจุบัน จะเห็นได้อย่างชัดเจนว่ามีข่าวอาชญากรรมเกี่ยวกับทรัพย์สินเกิดขึ้นอย่างมาก โดยสำนักงานตำรวจแห่งชาติได้สรุป 5 อันดับสถิติคดีในโลกออนไลน์ที่มีประชาชนโดนหลอกมากที่สุด ซึ่งนี้เป็นเพียงอาชญากรรมออนไลน์เท่านั้น ยังไม่รวมสถิติอาชญากรรมในรูปแบบปกติ ซึ่งพบว่ามีคดีความผิดเกี่ยวกับทรัพย์ รับแจ้งถึง 45,033 ราย และล่าสุดในปี 2565 ยังมีข่าวเกี่ยวกับ Forex3D ซึ่งเป็นข่าวที่เกี่ยวกับทรัพย์สินที่ยิ่งใหญ่และเป็นคดีกันมาอย่างยาวนาน ซึ่งทั้งหมดที่กล่าวมา

ข้างต้นนี้แสดงให้เห็นถึงข่าวอาชญากรรมทางทรัพย์สินที่ต่อให้นานเท่าไหร่ก็คงไม่หมดไป แล้วเพราะสาเหตุใดคดีอาชญากรรมต่างๆ ถึงเกิดขึ้นมาได้

ด้วยเหตุนี้ผู้วิจัยจึงได้เล็งเห็นถึงความสำคัญของคดีอาชญากรรมหรือคดีอาญาทางทรัพย์สินซึ่งมีมากและคาดว่าจะเพิ่มขึ้นอีกในอนาคต จึงได้ศึกษาว่าปัจจัยใดบ้างที่มีผลที่ทำให้เกิดจำนวนคดีอาญาที่เกี่ยวกับทรัพย์สิน โดยหากเราทราบถึงปัจจัยที่ส่งผลต่อการเกิดของจำนวนอาญาที่เกี่ยวกับทรัพย์สินแล้วนั้นจะทำให้เราสามารถแก้ปัญหาได้อย่างตรงจุดและสามารถลดอาชญากรรมที่กำลังจะเกิดขึ้นในอนาคตได้ไม่มากก็น้อย โดยผู้วิจัยได้ศึกษาปัจจัยทั้งสิ้น 7 ปัจจัย ได้แก่ จำนวนคนว่างงาน อัตราเงินเฟ้อ จำนวนคนจน การออมสุทธิต่อคน หนี้สินเฉลี่ยต่อครัวเรือน รายได้โดยเฉลี่ยต่อครัวเรือน และจำนวนประชากร โดยวิธีทางสถิติที่ใช้ในการวิเคราะห์ได้แก่ การวิเคราะห์การถดถอย (Regression Analysis) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคป่าสุ่ม (Random Forest)

วัตถุประสงค์ของการศึกษา

1. เพื่อเปรียบเทียบวิธีที่ดีที่สุดสำหรับการพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทย

2. ต้องการทราบว่าปัจจัยใดบ้างที่ส่งผลให้เกิดคดีอาญาในประเทศไทย เพื่อหาแนวทางในการรับมือและแก้ไขสถานการณ์ที่จะเกิดขึ้น

งานวิจัยที่เกี่ยวข้อง

Witsarut Kaewmaha & Varis Punyachatporn [1] ได้ทำงานวิจัยเกี่ยวกับการคาดการณ์ผลตอบแทนในอนาคตของตราสารทุนหุ้นสามัญ โดยการใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง โดยใช้ข้อมูลราคารายวันของตราสารทุนในตลาดหลักทรัพย์แห่งประเทศไทย ข้อมูลในงบการเงิน ข้อมูลอัตราส่วนทางการเงิน ข้อมูลปัจจัยทางเทคนิคข้อมูลเศรษฐศาสตร์มหภาค อัตราแลกเปลี่ยน ดัชนีหลักทรัพย์ และดัชนีทองคำ ซึ่งเป็นข้อมูลช่วง 2552 - 2564 ในส่วนอัลกอริทึมของ Machine Learning นั้น ผู้วิจัยได้ใช้เทคนิค Artificial neural network (ANN), Random Forest (RF) และ Long Short-Term Memory (LSTM) ได้ผลการศึกษาแสดงค่าพยากรณ์ผลตอบแทนด้วยแบบจำลองคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองว่าแบบจำลองที่มีความแม่นยำที่สุด คือแบบจำลอง Random Forest

Weerapun Panich [2] ได้ทำการศึกษาการวิจัยโดยการนำเสนอโมเดลการทำนายผลการเรียนวิชา Web Database ที่มีประสิทธิภาพดีที่สุด ผู้วิจัยดำเนินการสังเคราะห์ Attribute จากเอกสารงานวิจัยที่เกี่ยวข้องและคัดกรองด้วยการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) แบบลำดับขั้น (Stepwise Selection) ได้ Attribute จำนวน 6 Attribute ทำการทดสอบและเปรียบเทียบประสิทธิภาพโมเดลระหว่างวิธี Decision Tree, Naive Bayes, Neural Network และSupport Vector Machine โดยใช้โปรแกรม RapidMiner Studio แบ่งเป็นกลุ่มเรียนรู้ (Train Set) 80% และกลุ่มทดสอบ (Test set) 20% ผลการวิจัยพบว่าประสิทธิภาพทำนายด้วยวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด 88.89% รองลงมาวิธี Support Vector Machine 85.19% ลำดับที่สามวิธี Decision Tree 79.63% ลำดับสุดท้ายวิธี Naive Bayes 74.07

Siwakorn Banluesapy & Waraporn Jirapanthong [3] ได้ทำการศึกษาข้อมูลผู้ป่วยจำนวน 1,608,923 ราย จากกรมควบคุมโรค โดยรวบรวมข้อมูลตั้งแต่วันที่ 1 มกราคม 2563 - 1 ตุลาคม 2564 แล้วทำการพัฒนาระบบคัดแยกผู้ป่วยโควิด-19 ในรูปแบบเว็บแอปพลิเคชัน โดยใช้เทคนิคการเรียนรู้ของเครื่องในการจำแนกข้อมูล จากนั้นทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทำนายการคัดแยกผู้ป่วยโควิด-19 โดยใช้การจำแนกข้อมูลทั้งหมด 3 แบบได้แก่ Random Forest, Neural network และ Naive Bayes ซึ่งจากผลการวิจัยพบว่า Random Forest มีค่าความถูกต้องเท่ากับ 93.33%,

Neural network มีค่าความถูกต้องเท่ากับ 92.7% และ Naive Bayes มีค่าความถูกต้องเท่ากับ 92.1% โดยผลการเปรียบเทียบในครั้งนี้ Random Forest มีค่าสูงสุด จึงไปใช้ในการพัฒนาโปรแกรมระบบในการคัดแยกผู้ป่วยโควิด-19 ได้

วิธีดำเนินการวิจัย

1. การศึกษาข้อมูล

ในการศึกษาข้อมูลเกี่ยวกับการพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary Data) ที่นำมาจากสำนักงานสถิติแห่งชาติย้อนหลัง 14 ปี โดยข้อมูลมีลักษณะรายปี ตั้งแต่ปี พ.ศ. 2550-2563 โดยที่ ประชากร คือ จำนวนการกระทำความผิดทางคดีอาญาที่เกี่ยวกับทรัพย์ในภาคกลางและ กลุ่มตัวอย่าง คือ จำนวนการกระทำความผิดทางคดีอาญาที่เกี่ยวกับทรัพย์ในภาคกลาง ที่ได้รับแจ้งความและเก็บสถิติโดยสำนักงานสถิติแห่งชาติ

2. การศึกษาตัวแปรที่ใช้ในงานวิจัย

2.1 ตัวแปรอิสระ ได้แก่

1. จำนวนคนว่างงาน
2. อัตราเงินเฟ้อ
3. จำนวนคนจน
4. การออมสุทธิต่อคน
5. หนี้สินเฉลี่ยต่อครัวเรือน
6. รายได้เฉลี่ยต่อครัวเรือน
7. จำนวนประชากร

2.2 ตัวแปรตาม ได้แก่ จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์

ตารางที่ 1 ตัวอย่างข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์และปัจจัยที่ส่งผล

จังหวัด	ปี	จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์	จำนวนคนว่างงาน	อัตราเงินเฟ้อ	จำนวนคนจน	การออมสุทธิต่อคน	หนี้สินเฉลี่ยต่อครัวเรือน	รายได้เฉลี่ยต่อครัวเรือน	จำนวนประชากร
กรุงเทพ	2550	18,451	46,813	81.70	269.356	8661.06	155,396.00	39,020.00	5,716,248
	2551	19,151	61,228	86.20	183.478	7681.53	200,726.02	44,325.71	5,710,883
	2552	16,957	55,476	85.50	190.480	8835.05	207,665.00	42,379.83	5,702,595
	2553	14,192	37,805	88.29	186.659	8925.31	200,726.02	44,325.71	5,701,394
	2554	8,464	37,922	91.65	647.923	9385.44	218,741.40	48,951.00	5,674,843
	2555	10,370	28,041	94.41	161.461	8737.99	200,726.02	44,325.71	5,673,560
	2556	10,488	28,962	96.47	90.400	8347.38	275,576.80	49,190.80	5,686,252
	2557	9,954	44,923	98.30	140.600	9617.05	200,726.02	44,325.71	5,692,284

3. การวิเคราะห์หาปัจจัยที่ส่งผลต่อตัวแปรตาม

ใช้วิธีการคัดเลือกตัวแปรโดยวิธีลดตัวแปร เพื่อวิเคราะห์หาปัจจัยที่ส่งผลต่อตัวแปรตามด้วยโปรแกรมสำเร็จรูปทางสถิติ เพื่อหาว่ามีตัวแปรใดบ้างที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลางของประเทศไทย [2]

4. การวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ทางสถิติและการเรียนรู้ของเครื่อง

4.1 เทคนิคการวิเคราะห์การถดถอยพหุคูณ

ในการวิเคราะห์ข้อมูลหาตัวแบบการวิเคราะห์การถดถอยพหุคูณ เป็นการวิเคราะห์ความสัมพันธ์เชิงเส้นตรงของตัวแปรตาม (Y) ตัวแปรอิสระ (X) ซึ่งมีรูปแบบทั่วไปดังนี้

$$Y_{CASES} = \beta_0 + \beta_1 X_{UE} + \beta_2 X_{CPI} + \beta_3 X_{POOR} + \beta_4 X_{SAVE} + \beta_5 X_{LIAB} + \beta_6 X_{INC} + \beta_7 X_{POP} + \varepsilon$$

โดยที่ Y_{CASES} หมายถึง จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์

X_i หมายถึง ตัวแปรอิสระแทนปัจจัยต่างๆ ซึ่งปัจจัยที่นำมาศึกษามีดังนี้

X_{UE} หมายถึง จำนวนคนว่างงาน

X_{CPI} หมายถึง อัตราเงินเฟ้อ

X_{POOR} หมายถึง จำนวนคนจน

X_{SAVE} หมายถึง การออมสุทธิต่อคน

X_{LIAB} หมายถึง หนี้สินเฉลี่ยต่อครัวเรือน

X_{INC} หมายถึง รายได้เฉลี่ยต่อครัวเรือน

X_{POP} หมายถึง จำนวนประชากร

โดยผู้วิจัยได้นำผลลัพธ์ที่ได้จากวิธีการคัดเลือกตัวแปรโดยวิธีลดตัวแปร ในข้อที่ 3 มาสร้างตัวแบบการวิเคราะห์การถดถอยพหุคูณ เทคนิคต้นไม้ตัดสิน และเทคนิคป่าสุ่ม

4.2 เทคนิคต้นไม้ตัดสิน

คือแบบจำลองทางคณิตศาสตร์เพื่อพยากรณ์ ซึ่งจะนำข้อมูลมาสร้างแบบจำลองในรูปแบบของโครงสร้างต้นไม้ โดยในงานวิจัยนี้ ผู้วิจัยได้ใช้อัลกอริทึม REPTree และนำชุดข้อมูลมาสร้างโมเดลด้วยโปรแกรม WEKA ด้วยเทคนิค 10-folds cross validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน ครั้งที่ 1 ใช้ส่วนที่ 1 เป็น Test set ส่วนที่เหลือทั้งหมดอีก 9 ส่วนเป็น Training set ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ โดย Evaluation accuracy จะเป็นค่าเฉลี่ยของทั้ง 10 ครั้ง จากนั้นจะรันโมเดลอีกครั้ง โดยใช้ Full training set ทั้งหมดในการรันโมเดล [3]

4.3 เทคนิคป่าสุ่ม

เป็นเทคนิคที่ทำการสุ่มเลือกคุณสมบัติออกมาจากชุดข้อมูลหลายๆชุดจากนั้นนำเอาชุดของคุณสมบัติเหล่านั้นมาสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจหลายๆต้น [4] โดยผู้วิจัยนำชุดข้อมูลมาสร้างโมเดลด้วยโปรแกรม WEKA ด้วยเทคนิค 10-folds cross validation เช่นเดียวกับเทคนิคต้นไม้ตัดสินใจ [4]

5. การเปรียบเทียบประสิทธิภาพ

ทำการเปรียบเทียบประสิทธิภาพการพยากรณ์ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทย โดยใช้เกณฑ์การเปรียบเทียบประสิทธิภาพ 2 เกณฑ์ ประกอบด้วย

1. ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE)

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

2. ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) [3]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

ผลการดำเนินการวิจัย

การวิเคราะห์ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทย ทั้ง 2 ชุด โดยมีขั้นตอนในการวิเคราะห์ข้อมูล ดังนี้

1. การวิเคราะห์หาปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทย

1.1 ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร

จากการวิเคราะห์หาปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร ด้วยวิธีการคัดเลือกตัวแปรแบบลดตัวแปรแล้วนั้น จึงได้ตัวแปรอัตราเงินเฟ้อ การออมสุทธิต่อคน รายได้เฉลี่ยต่อครัวเรือน และจำนวนประชากร แต่พบว่าค่าระดับนัยสำคัญทางสถิติ (Significance value: Sig) ของตัวแปรการออมสุทธิต่อคน เท่ากับ 0.084 กล่าวคือ ตัวแปรการออมสุทธิต่อคนไม่มีผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ผู้วิจัยจึงได้ตัดตัวแปรนี้ทิ้งไป ฉะนั้นแล้วปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร จึงเหลือเพียงตัวแปรอัตราเงินเฟ้อ รายได้เฉลี่ยต่อครัวเรือน และจำนวนประชากร ที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร

ตารางที่ 2 ปัจจัยที่ส่งผลต่อจำนวนคดีอาญาในภาคกลาง

ตัวแปร	B	Std. Error	Beta	T	Sig.	Tolerance	VIF
(Constant)	4.151	0.234		17.728	0.000		
CPI	-0.054	0.007	-0.484	-8.000	0.000	0.486	2.056
SAVE	0.00003287	0.000	0.101	1.738	0.084	0.527	1.897
INC	0.00003794	0.000	0.301	6.606	0.000	0.858	1.165
POP	0.000001794	0.000	0.612	13.999	0.000	0.930	1.076

1.2 ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร

จากการวิเคราะห์หาปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร ด้วยวิธีการคัดเลือกตัวแปรแบบลดตัวแปร พบว่าตัวแปรจำนวนคนว่างงานและอัตราเงินเฟ้อ ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร

ตารางที่ 3 ปัจจัยที่ส่งผลต่อจำนวนคดีอาญาในกรุงเทพมหานคร

ตัวแปร	B	Std. Error	Beta	T	Sig.	Tolerance	VIF
(Constant)	95844.075	5443.505		17.607	0.000		
UE	0.082	0.024	0.189	3.385	0.012	0.986	1.014
CPI	-620.093	36.595	-0.949	-16.945	0.000	0.986	1.014

2. การสร้างแบบจำลอง

2.1 ข้อมูลจำนวนคดีอาญาในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร

2.1.1 เทคนิคการวิเคราะห์การถดถอยพหุคูณ

นำข้อมูลที่ได้ทำการคัดเลือกมาสร้างแบบจำลองการวิเคราะห์การถดถอยพหุคูณในโปรแกรมสำเร็จรูปทางสถิติ ได้สมการพยากรณ์ ดังนี้

$$Ln_PREDICTCASES = -0.046x_{CPI} + 0.00003612x_{INC} + 0.000001783x_{POP} + 3.975$$

โดยที่ $Ln_PREDICTCASES$ หมายถึง จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์

x_{CPI} หมายถึง อัตราเงินเฟ้อ

x_{INC} หมายถึง รายได้เฉลี่ยต่อครัวเรือน

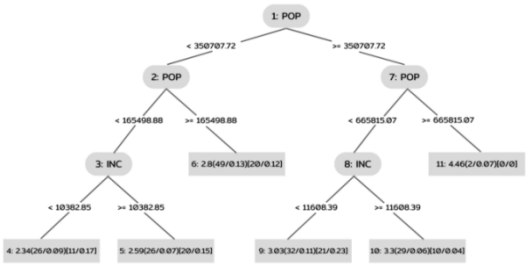
x_{POP} หมายถึง จำนวนประชากร

2.1.2 เทคนิคต้นไม้ตัดสินใจ

ผู้วิจัยได้นำข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร มาวิเคราะห์ในโปรแกรม WEKA เทคนิคต้นไม้ตัดสินใจ โดยใช้อัลกอริทึม REPTree ได้ผลการวิเคราะห์ ดังนี้

ตารางที่ 4 ผลการวิเคราะห์ข้อมูลในโปรแกรม WEKA โดยเทคนิคต้นไม้ตัดสินใจ

ผลการวิเคราะห์เทคนิคต้นไม้ตัดสินใจ	
Correlation coefficient (r)	0.5961
Mean absolute error (MAE)	0.2722
Root mean squared error (RMSE)	0.3856
Relative absolute error (RAE)	72.8499%
Root relative squared error (RRSE)	80.5171%
Total Number of Instances	246



ภาพที่ 1 ส่วนประกอบต้นไม้ตัดสินใจ ของข้อมูลภาคกลางที่แสดงปัจจัยที่เกี่ยวข้อง

จากภาพที่ 1 จะเห็นว่าปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานครมากที่สุด คือ จำนวนประชากร ซึ่งมีค่าเฉลี่ยเท่ากับ 350707.72 คน โดยในกรณีที่ 1 จำนวนประชากรน้อยกว่าค่าเฉลี่ย ให้พิจารณาทางด้านซ้ายมือของต้นไม้ตัดสินใจ โดยถ้าจำนวนประชากรมีค่ามากกว่าหรือเท่ากับค่าเฉลี่ย ซึ่งมีค่าเท่ากับ 165498.88 คน แสดงว่าค่าพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง มีค่าเท่ากับ 2.8 คดี แต่ถ้าจำนวนประชากรน้อยกว่าค่าเฉลี่ย ให้พิจารณาปัจจัยรายได้เฉลี่ยต่อครัวเรือน ซึ่งถ้ารายได้เฉลี่ยต่อครัวเรือนมีค่าน้อยกว่าค่าเฉลี่ย คือ

10382.85 บาท จะได้ค่าพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง มีค่าเท่ากับ 2.34 คดี และในกรณีที่ 2 จำนวนประชากร มีค่ามากกว่าหรือเท่ากับค่าเฉลี่ย 350707.72คน ให้พิจารณาทางด้านขวามือของต้นไม้ตัดสินใจ โดยใช้หลักเกณฑ์ในการพิจารณาเดียวกันกับกรณีที่ 1

เทคนิคป่าสุ่ม โดยในครั้งนี้ผู้วิจัยได้กำหนดจำนวนต้นไม้ตัดสินใจ ทั้งหมดจำนวน 100 ต้น

2.1.3 เทคนิคป่าสุ่ม

ผู้วิจัยได้นำข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร มาวิเคราะห์ในโปรแกรม WEKA

ตารางที่ 5 ผลการวิเคราะห์ข้อมูลในโปรแกรม WEKA โดยเทคนิคป่าสุ่ม

ผลการวิเคราะห์เทคนิคป่าสุ่ม	
Correlation coefficient (r)	0.7223
Mean absolute error (MAE)	0.2453
Root mean squared error (RMSE)	0.3294
Relative absolute error (RAE)	65.6622%
Root relative squared error (RRSE)	68.7854%
Total Number of Instances	246

2.2 ข้อมูลจำนวนคดีอาญาในกรุงเทพมหานคร

2.2.1 เทคนิคการวิเคราะห์การถดถอยพหุคูณ

นำข้อมูลที่ได้ทำการคัดเลือกมาสร้างแบบจำลองการวิเคราะห์การถดถอยพหุคูณในโปรแกรม สำเร็จรูปทางสถิติ ได้สมการพยากรณ์ ดังนี้

$$PREDICTCASES = 0.082x_{UE} - 620.093x_{CPI} + 95844.075$$

โดยที่ $PREDICTCASES$ หมายถึง จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์

x_{UE} หมายถึง จำนวนคนว่างงาน

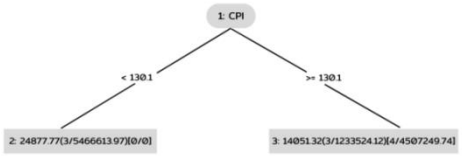
x_{CPI} หมายถึง อัตราเงินเฟ้อ

2.2.2 เทคนิคต้นไม้ตัดสินใจ

ผู้วิจัยได้นำข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร มาวิเคราะห์ในโปรแกรม WEKA เทคนิคต้นไม้ตัดสินใจ โดยใช้อัลกอริทึม REPTree ได้ผลการวิเคราะห์ ดังนี้

ตารางที่ 6 ผลการวิเคราะห์ข้อมูลในโปรแกรม WEKA โดยเทคนิคต้นไม้ตัดสินใจ

ผลการวิเคราะห์เทคนิคต้นไม้ตัดสินใจ	
Correlation coefficient (r)	0.6408
Mean absolute error (MAE)	2624.9698
Root mean squared error (RMSE)	04157.2168
Relative absolute error (RAE)	51.9555%
Root relative squared error (RRSE)	71.2355%
Total Number of Instances	10



ภาพที่ 2 ส่วนประกอบต้นไม้ตัดสินใจ ของข้อมูลกรุงเทพมหานครที่แสดงปัจจัยที่เกี่ยวข้อง

จากภาพที่ 2 จะเห็นว่าปัจจัยที่ส่งผลต่อจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานครมากที่สุด คือ อัตราเงินเฟ้อ ซึ่งมีค่าเฉลี่ยเท่ากับ 130.1 โดยกรณีที่อัตราเงินเฟ้อมีค่าน้อยกว่าค่าเฉลี่ย แสดงว่าค่าพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร มีค่าเท่ากับ 24877.77 คดี ในขณะเดียวกัน กรณีที่อัตราเงินเฟ้อมีค่ามากกว่าหรือเท่ากับค่าเฉลี่ย แสดงว่าค่าพยากรณ์จำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร มีค่าเท่ากับ 14051.32 คดี

2.2.3 เทคนิคป่าสุ่ม

ผู้วิจัยได้นำข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร มาวิเคราะห์ในโปรแกรม WEKA เทคนิคป่าสุ่ม โดยในครั้งนี้ผู้วิจัยได้กำหนดจำนวนต้นไม้ตัดสินใจ ทั้งหมดจำนวน 100 ต้น ได้ผลการวิเคราะห์ ดังนี้

ตารางที่ 7 ผลการวิเคราะห์ข้อมูลในโปรแกรม WEKA โดยเทคนิคป่าสุ่ม

ผลการวิเคราะห์เทคนิคป่าสุ่ม	
Correlation coefficient (r)	0.9357
Mean absolute error (MAE)	1701.0146
Root mean squared error (RMSE)	1945.7962
Relative absolute error (RAE)	33.6678%
Root relative squared error (RRSE)	33.342%
Total Number of Instances	10

3. การเปรียบเทียบประสิทธิภาพ

เปรียบเทียบการวิเคราะห์ข้อมูลทั้ง 3 วิธี คือ เทคนิคการวิเคราะห์การถดถอยพหุคูณ เทคนิคต้นไม้ตัดสินใจ และเทคนิคป่าสุ่ม วิธีการวิเคราะห์ข้อมูลแบบโดยใช้เกณฑ์การวัด 2 เกณฑ์ ประกอบด้วย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งวิธีที่เหมาะสมที่สุดจะพิจารณาจากวิธีที่ได้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยที่สุด

3.1 ค่าวัดประสิทธิภาพของข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง

ตารางที่ 8 ค่า RMSE และค่า MAE ของข้อมูลจำนวนคดีอาญาในภาคกลาง

วิธีการวิเคราะห์ข้อมูล	**ภาคกลาง**	
	RMSE	MAE
เทคนิคการวิเคราะห์การถดถอยพหุคูณ	**0.3193**	**0.2405**
เทคนิคต้นไม้ตัดสินใจ	0.3856	0.2722
เทคนิคป่าสุ่ม	0.3294	0.2453

วิธีที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลางที่ไม่รวมกรุงเทพมหานคร คือ วิธีการวิเคราะห์ข้อมูลโดยเทคนิคการวิเคราะห์การถดถอยพหุคูณ เนื่องจากได้ค่า RMSE เท่ากับ 0.3193 คดี ซึ่งต่ำกว่าวิธีการวิเคราะห์ข้อมูลโดยเทคนิคต้นไม้ตัดสินใจและเทคนิคป่าสุ่ม มีค่า RMSE เท่ากับ 0.3856 และ 0.3294 คดี ตามลำดับ สอดคล้องกับเกณฑ์การพิจารณาประสิทธิภาพค่า MAE ที่พบว่า วิธีการวิเคราะห์ข้อมูลโดยเทคนิคการวิเคราะห์การถดถอยพหุคูณ ได้ค่า MAE เท่ากับ 0.2405 คดี ซึ่งต่ำกว่า วิธีการวิเคราะห์ข้อมูลโดยเทคนิคต้นไม้ตัดสินใจและเทคนิคป่าสุ่ม ที่ได้ค่า MAE เท่ากับ 0.2722 และ 0.2453 คดี ตามลำดับ

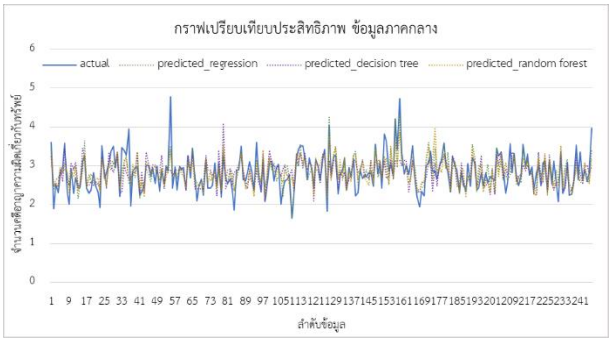
3.2 ค่าวัดประสิทธิภาพของข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร

ตารางที่ 9 แสดงค่า RMSE และค่า MAE ของข้อมูลจำนวนคดีอาญาในกรุงเทพมหานคร

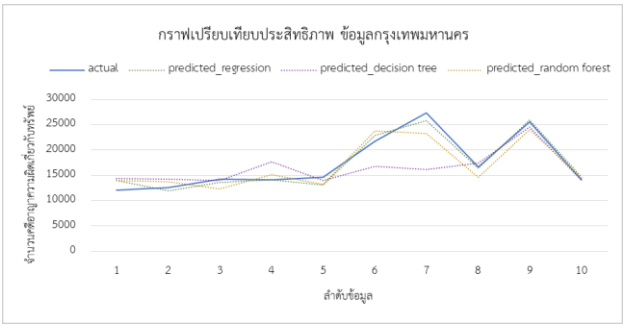
วิธีการวิเคราะห์ข้อมูล	**กรุงเทพมหานคร**	
	RMSE	MAE
เทคนิคการวิเคราะห์การถดถอยพหุคูณ	**1032.4407**	**858.1238**
เทคนิคต้นไม้ตัดสินใจ	4157.2168	2624.9698
เทคนิคป่าสุ่ม	1945.7962	1701.0146

วิธีที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร คือ วิธีการวิเคราะห์ข้อมูลโดยเทคนิคการวิเคราะห์การถดถอยพหุคูณ เนื่องจากได้ค่า RMSE เท่ากับ 1032.4407 คดี ซึ่งต่ำกว่าวิธีการวิเคราะห์ข้อมูลโดยเทคนิคต้นไม้ตัดสินใจและเทคนิคป่าสุ่ม มีค่า RMSE เท่ากับ 4157.2168 และ 1945.7962 คดี ตามลำดับ สอดคล้องกับเกณฑ์การพิจารณาประสิทธิภาพค่า MAE ที่พบว่าวิธีการวิเคราะห์ข้อมูลโดยเทคนิคการวิเคราะห์การถดถอยพหุคูณ ได้ค่า MAE เท่ากับ 858.1238 คดี ซึ่งต่ำกว่าวิธีการวิเคราะห์ข้อมูลโดยเทคนิคต้นไม้ตัดสินใจและเทคนิคป่าสุ่ม ที่ได้ค่า MAE เท่ากับ 2624.9698 และ 1701.0146 คดี ตามลำดับ

3.3 กราฟเปรียบเทียบประสิทธิภาพ



กราฟที่ 1 เปรียบเทียบประสิทธิภาพของข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง



กราฟที่ 2 เปรียบเทียบประสิทธิภาพของข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในกรุงเทพมหานคร

จากกราฟเปรียบเทียบประสิทธิภาพระหว่างค่าจริงและค่าพยากรณ์ของจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ทั้ง 3 วิธี ในกราฟที่ 1 และ 2 พบว่า ค่าพยากรณ์ของวิธีการวิเคราะห์การถดถอยนั้นใกล้เคียงกับค่าจริง มากกว่าวิธีต้นไม้ตัดสินใจและวิธีป่าสุ่ม

สรุปผลการวิจัย

1. ผลการทดสอบโดยใช้เกณฑ์การวัด

โดยในข้อมูลจำนวนคดีอาญาในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร และจำนวนคดีอาญาในกรุงเทพมหานคร จะทดสอบโดยใช้เกณฑ์ 2 เกณฑ์ ประกอบด้วย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย ได้ผลสรุปออกเป็น 2 ส่วน ดังนี้

1.1 ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ในภาคกลาง ที่ไม่รวมกรุงเทพมหานคร

จากการศึกษาข้อมูลภาคกลางที่ไม่รวมกรุงเทพมหานคร โดยค่า RMSE ที่ได้จากการพยากรณ์โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ 0.3193 คดี ค่า RMSE จากการพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ 0.3856 คดี และค่า RMSE จากการพยากรณ์โดยใช้เทคนิคป่าสุ่ม 0.3294 คดี ในส่วนของค่า MAE ที่ได้จากการพยากรณ์โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ 0.2405 คดี ค่า MAE จากการพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ 0.2722 คดี และค่า MAE จากการพยากรณ์โดยใช้เทคนิคป่าสุ่ม 0.2453 คดี

1.2 ข้อมูลจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์กรุงเทพมหานคร

จากการศึกษาข้อกรุงเทพมหานครได้ผลดังนี้ โดยค่า RMSE ที่ได้จากการพยากรณ์โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ 1032.4407 คดี ค่า RMSE จากการพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ 4157.2168 คดี และค่า RMSE จากการพยากรณ์โดยใช้เทคนิคป่าสุ่ม 1945.7962 คดี ในส่วนของค่า MAE ที่ได้จากการพยากรณ์โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ 858.1238 คดี ค่า MAE จากการพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ 2624.9698 คดี และค่า MAE จากการพยากรณ์โดยใช้เทคนิคป่าสุ่ม 1701.0146 คดี

2. ปัจจัยที่ส่งผลให้เกิดคดีอาญาในประเทศไทย

ผลการวิจัยพบว่าปัจจัยที่ส่งผลต่อข้อมูลในภาคกลาง มีจำนวน 3 ปัจจัย ได้แก่ อัตราเงินเฟ้อ รายได้เฉลี่ยต่อครัวเรือน จำนวนประชากร และปัจจัยที่ส่งผลต่อข้อมูลในกรุงเทพมหานคร มีจำนวน 2 ปัจจัย ได้แก่ อัตราเงินเฟ้อ จำนวนคนว่างงาน

3. ผลการเปรียบเทียบการพยากรณ์ทั้ง 3 วิธี

ผลจากการศึกษาการสร้างสมการทำนายจำนวนคดีอาญาความผิดเกี่ยวกับทรัพย์ของภาคกลางในประเทศไทยและจังหวัดกรุงเทพมหานคร โดยใช้วิธีการวิเคราะห์ทางสถิติและการเรียนรู้ของเครื่อง พบว่าเมื่อเปรียบเทียบการพยากรณ์ทั้งสามวิธีดังกล่าว โดยพิจารณาจากค่าประมาณที่ดีที่สุดของพารามิเตอร์ซึ่งค่าประมาณที่ดีที่สุดหมายถึงค่าที่ทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยมีค่าต่ำที่สุด

พบว่าการพยากรณ์โดยเทคนิคการวิเคราะห์การถดถอย จะให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าการพยากรณ์โดยเทคนิคป่าสุ่มและการพยากรณ์โดยเทคนิคต้นไม้ตัดสินใจตามลำดับ ซึ่งจะสอดคล้องกับค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย

4. ผลสรุปการพยากรณ์ข้อมูลภาคกลาง และกรุงเทพมหานคร

จากผลสรุปจะเห็นได้ว่าวิธีที่เหมาะสมสำหรับข้อมูลของภาคกลาง และข้อมูลของกรุงเทพมหานครคือ วิธีเทคนิคการวิเคราะห์การถดถอย ดังนั้นเมื่อต้องการพยากรณ์ข้อมูลที่มีลักษณะคล้ายคลึงกับข้อมูลภาคกลางและข้อมูลของกรุงเทพมหานคร ที่ใช้ในการศึกษาครั้งนี้ควรเลือกใช้วิธีเทคนิคการวิเคราะห์การถดถอย เนื่องจากจะทำให้ได้สมการพยากรณ์ที่มีความคลาดเคลื่อนน้อยกว่า หรืออาจทำการศึกษาโดยวิธีอื่นๆ เพิ่มเติม เพื่อเปรียบเทียบกับการพยากรณ์ทั้งสามวิธีดังกล่าวว่าวิธีใดจะทำให้ได้สมการพยากรณ์ที่ดีที่สุด

เอกสารอ้างอิง (References)

[1] Witsarut Kaewmaha & Varis Punyachatporn. (2021). Predicting Stock Return Using Machine Learning. *Journal of innovation in Business, Management, and Social Sciences, 2*(3), 108-123. (In Thai)

[2] Weerapun Panich. (2021). *Apply of Machine Learning to Grade Prediction in Web Databases Course of Educational Technology Students at Faculty of Education Burapha University* [Master's thesis,Dhurakij Pundit University]. http://libdoc.dpu.ac.th/thesis/Weerapun.Pan.pdf. (In Thai)

[3] Siwakorn Banluesapy & Waraporn Jirapanthong. (2022). Towards Machine Learning Algorithm for Screening Prediction of COVID-19 Patients. *Journal of Information Science and Technology. 12*(1), 47-60. (In Thai)

[4] Wuttichai Chan-in. (2017). Regression Model Development for Preliminary Cost Estimation of High Density Polyethylene Natural Gas Pipeline Construction in Industrial Estates of Thailand. *Thai Science and Technology Journal, 25*(4), 656-669. (In Thai)