

การเปรียบเทียบแบบจำลองจำแนกประเภทเพื่อการพัฒนาการท่องเที่ยว
อำเภอเขาค้อ จังหวัดเพชรบูรณ์ด้วยเทคนิคเหมืองข้อมูล
MODEL COMPARISONS OF CASSIFICATION FOR TOURISM
DEVELOPMENT OF KHAO KHO DISTRICT, PHETCHABUN PROVINCE
USING DATA MINING TECHNIQUES

เจษฎาพร ปาคำวัง^{1*} กาญจน์ คุ่มทรัพย์¹ และวรชัย ศรีเมือง²
Jetsadaporn Pakamwang^{1*}, Kan Khoomsab¹ and Worachai Srimuang²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบแบบจำลองจำแนกประเภทเพื่อใช้พัฒนาการท่องเที่ยว อำเภอเขาค้อ จังหวัดเพชรบูรณ์ด้วยเทคนิคเหมืองข้อมูล กลุ่มตัวอย่างคือนักท่องเที่ยวที่เดินทางมาท่องเที่ยวอำเภอเขาค้อ จังหวัดเพชรบูรณ์ จำนวน 1,474 คน โดยการสุ่มตัวอย่างแบบบังเอิญ ดำเนินการเก็บข้อมูลระหว่างเดือนตุลาคม พ.ศ.2561 ถึงเดือนมีนาคม พ.ศ.2562 เครื่องมือวิจัยเป็นแบบสอบถาม จำนวน 7 แอททริบิวต์ การสร้างแบบจำลองการวิจัยนี้ใช้เทคนิคการจำแนกประเภทข้อมูล จำนวน 7 วิธี ได้แก่ 1) Bayes network classifier 2) Naive bayes 3) Multi-layer perceptron 4) Logistic regression 5) Sequential minimal optimization 6) Decision tree และ 7) Random forest ด้วยโปรแกรม RapidMiner Studio เพื่อทดสอบความแม่นยำของโมเดลจากความคิดเห็นของนักท่องเที่ยว ต่อการพัฒนาแหล่งท่องเที่ยวอำเภอเขาค้อ จังหวัดเพชรบูรณ์ การประเมินประสิทธิภาพของโมเดลด้วยวิธีการตรวจสอบไขว้ (k-Fold Cross Validaton) โดยกำหนด k=10 ผลเปรียบเทียบประสิทธิภาพแบบจำลองพบว่าแบบจำลองที่มีค่า Correctly มากที่สุด คือแบบจำลอง Decision tree มีค่าเท่ากับ 94.57 การเปรียบเทียบประสิทธิภาพของแบบจำลองด้วย Confusion matrix พบว่าแบบจำลองที่มีค่าประสิทธิภาพดีที่สุดได้แก่ Decision tree ผลจากการวิจัยจะถูกนำไปใช้ในการพัฒนาแบบจำลองทางสารสนเทศเพื่อทำนายการพัฒนาการท่องเที่ยวของอำเภอเขาค้อ จังหวัดเพชรบูรณ์ต่อไป

คำสำคัญ: แบบจำลอง เหมืองข้อมูล การท่องเที่ยว

¹คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

¹Faculty of Science and Technology, Phetchabun Rajabhat University, Muang District, Phetchabun Province 67000

²คณะเทคโนโลยีการเกษตรและอุตสาหกรรม มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

²Faculty of Agricultural and Industrial Technology, Phetchabun Rajabhat University, Muang District, Phetchabun Province 67000

*corresponding author e-mail: jetsadaporn.pa@pcru.ac.th

Received: 1 May 2020; Revised: 21 May 2020; Accepted: 1 June 2020

Abstract

The objective of this research was to compare the classification models for tourism development in Khao Kho District, Phetchabun Province using data mining technique. The sample of this study consisted of 1,474 tourists who travel to Khao Kho District, Phetchabun Province. The sample was selected based on an accidental sampling. Data were collected between October 2018 and March 2019. A research instrument was a questionnaire composed of 7 attributes. The modeling of this research was based on seven data classification techniques, namely 1) Bayes network classifier, 2) Naive bayes, 3) Multi-layer perceptron, 4) Logistic regression, 5) Sequential minimal optimization, 6) Decision tree and 7) Random forest using RapidMiner Studio program to test the accuracy of the model from the opinions of tourists towards the development of tourist attractions in Khao Kho District, Phetchabun Province. The model performance was evaluated with k-Fold Cross Validation whereas $k = 10$. The results of comparing the efficiency of the model showed that the decision tree gives the highest accuracy with a value of 94.57. To compare the efficiency of the models, confusion matrix was selected from the model with the best efficiency, which was the decision tree. The experimental results could be applied to develop a predictive information modeling of tourism development in Khao Kho District, Phetchabun Province.

Keywords: Model, Data mining, Tourism

บทนำ

ปัจจุบันเทคโนโลยีสารสนเทศและการสื่อสาร ส่งผลกระทบต่ออุตสาหกรรมการท่องเที่ยวในหลายด้าน การจัดทำระบบสารสนเทศ ข้อมูลและสถิติด้านการท่องเที่ยวจึงเป็นสิ่งสำคัญในการนำข้อมูลมาใช้ในการตัดสินใจ โดยรัฐบาลหรือหน่วยงานของรัฐ สามารถนำข้อมูลเพื่อใช้ในการวางแผน ปรับปรุงแก้ไขการทำงาน นักท่องเที่ยวสามารถนำข้อมูลจากระบบสารสนเทศไปใช้ในการวางแผนการท่องเที่ยว (Caichompoo et al., 2017) และกลุ่มธุรกิจท่องเที่ยวมีนำไปใช้ในการปรับปรุงธุรกิจ วางแผนธุรกิจ การตลาด และประชาสัมพันธ์ของธุรกิจ ดังนั้น สิ่งที่ขาดไม่ได้คือข้อมูลและสถิติ ในอุตสาหกรรมการท่องเที่ยว ข้อมูลสถิติจะบอกถึงผลงานในอดีตที่ผ่านมาและใช้ในการวางแผนในอนาคต ซึ่งต้องได้มาด้วยกระบวนการทางวิจัย และใช้สถิติที่ถูกต้องเพื่อให้ข้อมูลมีคุณภาพน่าเชื่อถือ วิธีการหนึ่งในปัจจุบันที่ได้รับความนิยมในการนำข้อมูลมาใช้ประโยชน์เพื่อการพัฒนาดังกล่าว คือ เทคนิคการทำเหมืองข้อมูล (data mining) เพื่อการพยากรณ์และสืบค้นความรู้ที่เป็นประโยชน์และนำเสนอในฐานข้อมูลของนักท่องเที่ยว โดยจะนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้ หรือสิ่งสำคัญออกมา เพื่อใช้ในการวิเคราะห์

หรือทำนายสิ่งต่าง ๆ ที่จะเกิดขึ้น ซึ่งการค้นหาคำรู้ ความจริงที่แฝงอยู่ในข้อมูล (knowledge discovery) เป็นกระบวนการขุดค้นสิ่งที่น่าสนใจในกองข้อมูลที่เราถืออยู่ (Han et al., 2012; Larose, 2015; Elatia et al., 2016) จากการทบทวนวรรณกรรมงานวิจัยในประเทศไทยพบว่าได้มีการนำเทคนิคการทำเหมืองข้อมูลมาใช้ในงานด้านการท่องเที่ยว โดย Kosinanon (2006) ได้ศึกษาระบบการพยากรณ์จำนวนนักท่องเที่ยว โดยใช้โครงข่ายประสาทเทียมกรณีศึกษา การท่องเที่ยวแห่งประเทศไทย เพื่อการพยากรณ์ไปช่วยในการวางแผนงานเรื่องการบริหารงานกับนักท่องเที่ยวโดยทราบจำนวนได้ล่วงหน้าจึงเกิดความสะดวกรวดเร็วทันเวลา ในขณะที่ Kittidachanupap (2015) ได้พัฒนาแบบจำลองการแนะนำสถานที่ท่องเที่ยวรายบุคคลกลุ่มของนักท่องเที่ยว ระหว่างวิธี J48 และ PCA-NN โดยพบว่า การทำนายกลุ่มนักท่องเที่ยวด้วยวิธี J48 มีประสิทธิภาพ ในการทำนายที่ดีกว่าวิธี PCA-NN Lake (2016) ได้พัฒนาแบบจำลองการพยากรณ์ปริมาณนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย เพื่อพัฒนาเศรษฐกิจด้านการท่องเที่ยวให้เติบโตขึ้น โดยใช้วิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้นแบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยข้อมูลที่น่ามาใช้ศึกษาเป็นข้อมูลนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2550-2558 จำนวน 108 เดือน ในด้านการหาความสัมพันธ์ Nuankaew et al. (2017) ใช้รูปแบบการท่องเที่ยวทางวัฒนธรรม ตามทฤษฎีการเรียนรู้ของกฎความสัมพันธ์ และต้นไม้เพื่อการตัดสินใจเพื่อเปรียบเทียบรูปแบบการท่องเที่ยวทางวัฒนธรรมในภาคเหนือและภาคตะวันออกเฉียงเหนือ เทคนิคและกระบวนการที่ใช้ในงานวิจัย คือ การวิเคราะห์สถิติพื้นฐาน ร่วมกับกระบวนการพัฒนาโมเดลเพื่อการเรียนรู้ของเครื่องคอมพิวเตอร์ โดยเก็บข้อมูลจากนักท่องเที่ยวที่เคยมีประสบการณ์การท่องเที่ยวทางวัฒนธรรม จำนวน 1,317 คน จาก 4 แห่ง 2 ภูมิภาคของประเทศไทย ได้แก่ 1) วัดร่องขุน จังหวัดเชียงราย 2) วัดศรีโคมคา จังหวัดพะเยา 3) วัดหนองแวง จังหวัดขอนแก่น และ 4) วัดพระธาตุนาดูน จังหวัดมหาสารคาม เครื่องมือที่ใช้ในการวิเคราะห์โมเดล คือ กระบวนการวัดประสิทธิภาพโมเดล ได้แก่ ความถูกต้อง ความแม่นยำ และความสามารถในการทำนาย Yotsawat et al. (2017) สร้างแบบจำลองกลุ่มนักท่องเที่ยวด้วยวิธีจัดกลุ่มข้อมูลร่วมกับวิธีการหาความสัมพันธ์ (association rule) โดยใช้กลุ่มตัวอย่างนักท่องเที่ยวจำนวน 704 ตัวอย่าง นำมาผ่านกระบวนการทำความสะอาดข้อมูลและจัดกลุ่มด้วยวิธีการจัดกลุ่มแบบสองขั้นตอน จากนั้นนำข้อมูลไปหาความสัมพันธ์ และ Rukpukdee & Morachart (2018) ศึกษาความต้องการและพฤติกรรมการตัดสินใจเลือกแหล่งท่องเที่ยวทางธรรมชาติของนักท่องเที่ยว และพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทางธรรมชาติ ในจังหวัดอุบลราชธานี ตัวอย่างที่ใช้ในการศึกษา คือ นักท่องเที่ยวที่เดินทางมาท่องเที่ยวหรือชาวบ้านที่อาศัย อยู่บริเวณแหล่งท่องเที่ยว จำนวน 1,840 คน โดยแบ่งแหล่งท่องเที่ยว ตามกิจกรรมการท่องเที่ยวออกเป็น 5 กิจกรรม ใช้เทคนิคการทำเหมืองข้อมูล คือ การเรียนรู้ต้นไม้ตัดสินใจมาใช้ในการวิจัย จากงานวิจัยดังกล่าวจะเห็นได้ว่าความสามารถของการใช้เทคนิคเหมืองข้อมูลสามารถนำมาใช้ในการบริหารจัดการสารสนเทศด้านการท่องเที่ยวได้

จากผลงานวิจัยที่ผ่านมาจะเห็นได้ว่าเทคนิคเหมืองข้อมูลมีหลายขั้นตอนวิธีที่หลากหลาย ได้แก่ การจัดกลุ่ม การจำแนกประเภท การทำนาย การหากฎความสัมพันธ์ ซึ่งแต่ละวิธีจะมีความเหมาะสมกับข้อมูลที่แตกต่างกัน จากการศึกษาข้อมูลสารสนเทศด้านการท่องเที่ยวอำเภอเขาค้อ จังหวัดเพชรบูรณ์ เป็นจุดท่องเที่ยวหลักที่นักท่องเที่ยวที่นิยมเดินทางมาเป็นจำนวนมาก และจากยุทธศาสตร์การพัฒนาจังหวัดเพชรบูรณ์ส่งเสริมการท่องเที่ยวระหว่างปี 2561-2564 กำหนดยุทธศาสตร์ 5 ด้าน โดยหน่วยงานราชการ งานวิจัยนี้จึงสนใจที่จะใช้ข้อมูลสารสนเทศจากนักท่องเที่ยวว่ามีความคิดเห็นตรงกับนโยบายของรัฐที่กำหนดขึ้นหรือไม่ โดยใช้แบบจำลองหาผลลัพธ์ความคิดเห็นของนักท่องเที่ยวจากการท่องเที่ยวของนักท่องเที่ยวโดยใช้เทคนิคเหมืองข้อมูล ประโยชน์ของงานวิจัยนี้จะสามารถนำไปใช้ในการทำนายแนวคิดการพัฒนาการท่องเที่ยวจากความคิดเห็นหรือข้อเสนอจากนักท่องเที่ยว และนำไปใช้สร้างนโยบายสาธารณะวางแผนการท่องเที่ยวและพัฒนาสารสนเทศด้านการท่องเที่ยว อำเภอเขาค้อ จังหวัดเพชรบูรณ์ต่อไป

วิธีดำเนินการวิจัย

งานวิจัยนี้ใช้กลุ่มตัวอย่างคือนักท่องเที่ยวที่เดินทางมาท่องเที่ยวอำเภอเขาค้อ จังหวัดเพชรบูรณ์ จำนวน 1,474 คน โดยการสุ่มตัวอย่างแบบบังเอิญ (accidental sampling) เนื่องจากไม่ทราบจำนวนประชากรที่แท้จริง โดยดำเนินการเก็บข้อมูลระหว่างเดือนตุลาคม พ.ศ.2561 ถึงเดือนมีนาคม พ.ศ.2562 โดยใช้เครื่องมือวิจัยเป็นแบบสอบถาม โดยมีคุณลักษณะ (attribute) ดังนี้

Table 1 Attributes of tourist status

No.	Attributes	Type	Description
1	Education	Ordinal	Education
2	Career	Nominal	Career
3	Salary	Continuous	Income
4	Accommodation	Nominal	Accommodation
5	Duration stay	Continuous	Duration
6	Most season	Nominal	Travel season
7	Strategy (Class)	Nominal	Opinion

การคัดเลือกคุณลักษณะนำมาจากเอกสารแผนการพัฒนาจังหวัดเพชรบูรณ์ 4 ปี (พ.ศ.2561 - 2564) โดยทบทวนผลการวิเคราะห์ข้อมูลยุทธศาสตร์ด้านการท่องเที่ยว จากนั้นนำมาสร้างเป็นแบบสอบถามเพื่อใช้ในการเก็บข้อมูล ดังตารางที่ 1 และ 2 (Table 1 and 2) เมื่อเก็บข้อมูลจากกลุ่มตัวอย่างนักท่องเที่ยวแล้วนำข้อมูลมาดำเนินการตามวิธีการวิเคราะห์ตามเทคนิคเหมืองข้อมูลต่อไป

Table 2 Attributes of strategy classification

No.	Attributes	Type	Description
1	Strategy 1	Nominal	Strength or weakness of scenery and natural resource
2	Strategy 2	Nominal	Problem of traffic jam in high season.
3	Strategy 3	Nominal	The opportunity of sustainable tourism and development in South East Asia.
4	Strategy 4	Nominal	Accelerate development of travel following national strategies.
5	Strategy 5	Nominal	Good scenery, clean air, and creative tourism for sustainable tourism.

วิธีการวิเคราะห์ข้อมูลใช้กระบวนการมาตรฐานการทำเหมืองข้อมูล CRISP-DM (Larose & Larose, 2015) โดยมีขั้นตอนการเตรียมข้อมูลให้มีคุณภาพ ประกอบด้วย 6 ขั้นตอน ดังนี้

1. การทำความเข้าใจเกี่ยวกับธุรกิจ คำว่าธุรกิจในที่นี้หมายถึงเนื้อหาเรื่องราวข้อมูลที่สนใจ โดยในการวิจัยนี้เป็นการนำข้อมูลจากนโยบายทางรัฐบาลมาใช้เป็นแนวคำตอบเพื่อให้นักท่องเที่ยวได้ประเมินความคิดเห็นของตนเองในการต้องการพัฒนาและปรับปรุงการท่องเที่ยวในอำเภอเขาค้อ จังหวัดเพชรบูรณ์ โดยใช้แบบจำลองวิธีการทำเหมืองข้อมูลมาใช้เพื่อสร้างแบบจำลองสารสนเทศในการยืนยันข้อมูลของนักท่องเที่ยวกับทางรัฐบาลว่ามีความคิดเห็นสอดคล้องกัน จากนั้นจึงค้นหาแบบจำลองที่เหมาะสมที่สุดไปใช้ในการทำนายความต้องการพัฒนาแหล่งท่องเที่ยวในพื้นที่อื่น ๆ ในจังหวัดเพชรบูรณ์ได้

2. การทำความเข้าใจเกี่ยวกับข้อมูล โดยข้อมูลที่ได้เป็นข้อมูลจากแบบสอบถามและข้อมูลทั้งหมดบันทึกในไฟล์เอกเซลและจัดเตรียมโครงสร้างข้อมูลเพื่อดำเนินการในขั้นตอนการเตรียมข้อมูลต่อไป

3. การเตรียมข้อมูล โดยการตรวจสอบข้อมูลที่ได้อาจมีความถูกต้อง จัดเก็บข้อมูลให้เป็นระเบียบ ทำความสะอาดข้อมูล (data cleaning) หรือตรวจสอบข้อมูลที่ผิดปกติ ข้อมูลที่ไม่สมบูรณ์หรือข้อมูลที่ขาดหายไปและดำเนินการแก้ไขให้เรียบร้อย ฟอแมตข้อมูลให้เหมาะสมและพร้อมใช้งาน

4. การสร้างแบบจำลองการวิจัยนี้ใช้เทคนิคการจำแนกประเภทข้อมูล (classification) ได้แก่ 1) Bayes network classifier 2) Naïve bayes 3) Multi-layer perceptron 4) logistic regression 5) Sequential minimal optimization 6) Decision tree และ 7) Random forest ด้วยโปรแกรม RapidMiner Studio Version Educational License Program Version 9.3 เพื่อทดสอบความแม่นยำของโมเดลจากความคิดเห็นของนักท่องเที่ยวต่อการพัฒนาแหล่งท่องเที่ยวอำเภอเขาค้อ จังหวัดเพชรบูรณ์

5. การประเมินผลใช้วิธีทดสอบประสิทธิภาพของโมเดลโดยวิธีการตรวจสอบไขว้ (k-Fold Cross Validation) โดยกำหนด $k = 10$ เพื่อให้ข้อมูลทุกตัวมีโอกาส เป็นชุดทดสอบและชุดสอน โดยแบ่งข้อมูลออกเป็น training data และ testing data โดยแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ใช้ 9 ส่วนเป็นชุดข้อมูลสอน และอีก 1 ส่วนเป็นชุดข้อมูลทดสอบ ซึ่งจะสลับกันจนครบทั้งหมด 10 รอบ จากนั้น

ทดสอบประสิทธิภาพด้วยค่า 1) Correctly 2) Incorrectly 3) TP Rate 4) FP Rate 5) Precision Recall และ 6) F-Measure เมื่อพิจารณาแบบจำลองที่มีประสิทธิภาพดีแล้ว นำผลการวิเคราะห์ Confusion matrix มาพิจารณาเพื่อตัดสินใจเลือกแบบจำลองที่เหมาะสมสำหรับนำไปใช้จริงอีกครั้งหนึ่ง ตามวิธีการของ Songpan (2017) โดย Confusion matrix มีตัวแปรดังนี้

True Positive (TP) หมายถึง ค่าคลาสเป้าหมาย ถูก และแบบจำลองทายถูก

False Negative (FN) หมายถึง ค่าคลาสเป้าหมาย ถูก และแบบจำลองทายผิด

True Negative (TN) หมายถึง ค่าคลาสเป้าหมาย ผิด และแบบจำลองทายผิด

True Positive (FP) หมายถึง ค่าคลาสเป้าหมาย ผิด และแบบจำลองทายถูก

6. การนำไปใช้งาน หลังจากได้แบบจำลองจากผลการวิจัยแล้วนำไปใช้โดยต้องมีการปรับแต่งที่เหมาะสมกับสภาวะจริง หรือนำไปใช้ร่วมในระบบสารสนเทศเพื่อการตัดสินใจด้านการพัฒนาแหล่งท่องเที่ยวอำเภอเขาต้อ จังหวัดเพชรบูรณ์

ผลการวิจัย

จากผลการวิเคราะห์ข้อมูลด้วยแบบจำลองทั้งหมด 7 แบบจำลองด้วยโปรแกรม RapidMiner Studio พบว่าค่า Correctly มีค่าระหว่าง 84.26-94.57 ค่า Incorrectly มีค่าระหว่าง 5.42-13.97 ค่า TP Rate มีค่าระหว่าง 0.84-0.94 ค่า FP Rate มีค่าระหว่าง 0.01-0.04 ค่า Precision มีค่าระหว่าง 0.84-0.94 ค่า Recall มีค่าระหว่าง 0.84-0.94 และค่า F-Measure มีค่าระหว่าง 0.84-0.94 จากตารางที่ 1 (Table 1) แสดงผลเปรียบเทียบประสิทธิภาพของแบบจำลองพบว่าแบบจำลองที่มีค่า Correctly มากที่สุดคือแบบจำลอง Decision tree มีค่าเท่ากับ 94.57 แบบจำลอง Incorrectly มีค่าน้อยที่สุดคือ Decision tree มีค่าเท่ากับ 5.42 แบบจำลองที่มีค่า TP Rate มากที่สุดคือ Decision tree มีค่าเท่ากับ 0.94 แบบจำลองที่มีค่า FP Rate มากที่สุดคือ Decision Tree มีค่าเท่ากับ 0.01 ค่า Precision Recall และ F-Measure ของแบบจำลอง Decision tree และ Random forest มีค่ามากที่สุด เท่ากับ 0.94

Table 3 Data Mining Techniques

	Data mining techniques						
	Bayes network classifier	Naive bayes	Multi-layer perceptron	Logistic regression	Sequential minimal optimization	Decision tree	Random forest
Correctly	86.43	86.56	93.82	84.26	86.02	94.57	93.96
Incorrectly	13.56	13.43	6.17	15.73	13.97	5.42	6.03
TP Rate	0.86	0.86	0.93	0.84	0.86	0.94	0.94
FP Rate	0.03	0.03	0.02	0.04	0.04	0.01	0.02
Precision	0.86	0.86	0.93	0.84	0.86	0.94	0.94
Recall	0.86	0.86	0.93	0.84	0.86	0.94	0.94
F-Measure	0.86	0.86	0.93	0.84	0.86	0.94	0.94

จากตารางที่ 4-6 (Tables 4-6) พิจารณาการเปรียบเทียบประสิทธิภาพของแบบจำลองได้เลือกการพิจารณา Confusion matrix จากแบบจำลองที่มีค่าประสิทธิภาพที่ดีที่สุดได้แก่ เทคนิค Decision tree Random forest และ Multi-layer perceptron เพื่อเปรียบเทียบความแม่นยำของผลการทำนายระหว่างผลลัพธ์จากแบบจำลองกับผลลัพธ์จริง พบว่า การทำนายผลที่ให้ความแม่นยำมากที่สุดคือ Confusion matrix ด้วย Decision tree รองลงมาคือ Random forest และ Multi-layer perceptron ตามลำดับ โดยพิจารณาจาก True Positive ซึ่งเป็นข้อมูลที่เป็นจริงและแบบจำลองทายถูก จากตารางที่ 2 (Table 2) พบว่าแบบจำลอง Decision tree มีผลการทำนาย Strategy ที่ 1 เท่ากับ 316 มากกว่าแบบจำลอง Random forest และ Multi-layer perceptron และมีแนวโน้มที่ทายถูกมากกว่าทุก Strategy จากผลการวิจัยนี้จึงสามารถสรุปได้ว่า การเลือกใช้แบบจำลองด้วย Decision tree จะมีประสิทธิภาพและความแม่นยำกว่าแบบจำลองอื่น ๆ ที่ได้ทดสอบเปรียบเทียบ

Table 4 Confusion matrix with decision tree

		Actual					Precision
		Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	
Predicted	Strategy 1	316	1	1	1	2	98%
	Strategy 2	0	183	0	0	0	100%
	Strategy 3	0	0	50	0	0	100%
	Strategy 4	2	0	1	98	1	96%
	Strategy 5	0	0	0	0	34	100%

จากตารางที่ 4 (Table 4) แสดงผลการคำนวณ Confusion matrix ด้วยวิธี Decision tree ผลจากการทำนายในแต่ละคลาสพบว่าความแม่นยำในแต่ละคลาส คือ 98% (Strategy 1) 100% (Strategy 2) 100% (Strategy 3) 96% (Strategy 4) และ 100% (Strategy 5) โดยมีค่าเฉลี่ยความแม่นยำเท่ากับ 98%

Table 5 Confusion matrix with random forest

		Actual					Precision
		Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	
Predicted	Strategy 1	303	5	1	5	0	96%
	Strategy 2	0	179	2	1	1	97%
	Strategy 3	0	0	50	0	0	100%
	Strategy 4	2	8	1	90	1	88%
	Strategy 5	0	0	1	0	33	97%

จากตารางที่ 5 (Table 5) แสดงผลการคำนวณ Confusion matrix ด้วยวิธี Random forest ผลจากการทำนายในแต่ละคลาสพบว่าความแม่นยำในแต่ละคลาส คือ 96% (Strategy 1) 97% (Strategy 2) 100% (Strategy 3) 88% (Strategy 4) และ 97% (Strategy 5) โดยมีค่าเฉลี่ยความแม่นยำเท่ากับ 95%

Table 6 Confusion matrix with multi-layer perceptron

		Actual					Precision
		Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5	
Predicted	Strategy 1	173	60	17	39	27	54%
	Strategy 2	0	147	2	17	17	80%
	Strategy 3	4	0	46	0	0	92%
	Strategy 4	2	8	1	90	1	88%
	Strategy 5	0	0	0	0	34	100%

จากตารางที่ 6 (Table 6) แสดงผลการคำนวณ Confusion matrix ด้วยวิธี Multi-layer perceptron ผลจากการทำนายในแต่ละคลาสพบว่าความแม่นยำในแต่ละคลาส คือ 54% (Strategy 1) 80% (Strategy 2) 92% (Strategy 3) 88% (Strategy 4) และ 100% (Strategy 5) โดยมีค่าเฉลี่ยความแม่นยำเท่ากับ 83%

อภิปรายผล

เทคนิคเหมืองข้อมูลจำแนกประเภทที่ใช้วิธีการเรียนรู้แบบมีผู้สอน (supervised learning) โดยสร้างแบบจำลองจากข้อมูลที่เราเรียกว่า Training data ที่ประกอบด้วยตัวแปรอิสระต่าง ๆ และตัวแปรตาม โดยแบบจำลองที่ได้จะบอกถึงความสัมพันธ์ของตัวแปรต้นและตัวแปรตาม โดยผลการวิจัยแสดงให้เห็นว่าวิธี Decision tree มีประสิทธิภาพการทำนายดีที่สุด จากการประเมินพารามิเตอร์การประเมินผลจากความถูกต้องและความแม่นยำ จากผลการคำนวณความแม่นยำการทำนายในตาราง Confusion matrix พบว่า วิธี Decision tree มีค่าเฉลี่ยความแม่นยำในการทำนายสูงสุดคือ 98% โดย Songpan (2017) ได้ให้เหตุผลของ Decision tree ว่าเป็นเทคนิคที่ให้ผลเร็วเมื่อเทียบกับเทคนิคอื่น ผลลัพธ์ที่ได้สามารถนำไปใช้งาน และสามารถแปลงเป็นกฎได้ สามารถนำไปประยุกต์ใช้กับการค้นหาข้อมูล SQL ได้ง่าย และ Decision tree มีความแม่นยำสูง การทำนายกลุ่มของนักท่องเที่ยวที่ใช้วิธี Decision tree เช่นเดียวกับการวิจัยของ Kittidachanupap (2015) พบว่าวิธี Decision tree มีกระบวนการสร้างโมเดลไม่ยุ่งยาก และซับซ้อนเหมือนกับการทำนายกลุ่มของนักท่องเที่ยวที่ใช้วิธี PCA-NN และได้ประสิทธิภาพของแบบจำลองการแนะนำสถานที่ท่องเที่ยวรายบุคคลที่ไม่แตกต่างจากการใช้วิธี Decision tree ดังนั้นการสร้างแบบจำลองการทำนายกลุ่มของนักท่องเที่ยวตามลักษณะส่วนบุคคลด้วยวิธี Decision tree จึงมีความเหมาะสมสำหรับนำไปใช้ในการพัฒนาแบบจำลองการแนะนำสถานที่ท่องเที่ยวรายบุคคล

และจากการวิจัยของ Nuankaew et al. (2017) ใช้การวิเคราะห์สถิติพื้นฐานร่วมกับใช้ Decision tree ทดสอบความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวกับพฤติกรรมและวิถีชีวิตของนักท่องเที่ยว ตามทฤษฎีการเรียนรู้ของกฎความสัมพันธ์ พบว่า ค่าความถูกต้องของโมเดลอยู่ในระดับสูง และด้านการพิจารณาจากการประเมินประสิทธิภาพของโมเดล เช่นเดียวกับ Rukpukdee & Morachart (2018) ที่ได้ใช้ Decision tree เพื่อพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทางธรรมชาติในจังหวัดอุบลราชธานี ซึ่งได้ผลการวิเคราะห์ในระดับดีมาก

อย่างไรก็ตามเป้าหมายของ Decision tree สำหรับการจำแนกประเภทข้อมูลจากข้อมูล Training data ในการสร้างโหนดราก แล้วใช้โหนดรากเป็นโหนดปัจจุบันที่มีข้อมูลเป็นข้อมูล Training data ทั้งหมด ดังนั้นการจำแนกประเภทข้อมูลอาจเขียนในรูป If-then ไปยังโหนดอื่น โดยการเขียนคำตอบ (class) เป็นค่าเป้าหมาย จากผลการวิจัยนี้ได้ใช้แบบจำลองจำแนกประเภทหลายวิธี ซึ่งแต่ละวิธีมีข้อจำกัด และเงื่อนไขในการเลือกใช้แตกต่างกันไป แต่ในภาพรวมการสร้างแบบจำลองจำแนกประเภทข้อมูลจะพยายามให้เกิดความผิดพลาดให้น้อยที่สุดบนข้อมูล Training data เมื่อนำแบบจำลองไปใช้จริงกับข้อมูลที่ไม่ทราบประเภท แบบจำลองอาจจำแนกข้อมูลได้ไม่แม่นยำ ดังนั้นจึงต้องมีการใช้ข้อมูลอีกชุดหนึ่งที่ทราบประเภทของข้อมูล มาใช้ในการตรวจสอบความสามารถหรือความแม่นยำของข้อมูลชุดนี้คือ Testing data จากนั้นจึงทำการวัดประสิทธิภาพของแบบจำลองโดยวิธี Confusion matrix (Elatia et al., 2016; Auwatanamongkoll, 2018)

การพิจารณาการเลือกแบบจำลองที่เหมาะสมได้พิจารณาจากค่าพารามิเตอร์การวัดดังตารางที่ 3 (Table 3) นั้น แบบจำลอง Decision tree มีค่าประสิทธิภาพของความถูกต้อง ความแม่นยำ สัดส่วนของการทายถูกต้องการทายผิด และค่า F - measure มีแนวโน้มที่มีค่าดีกว่าวิธีอื่น ๆ และเมื่อมาพิจารณาตาราง Confusion matrix ก็พบว่า จำนวนของการทายถูกในทุกคลาสตามแนวทแยงนั้น แบบจำลอง Decision tree มีค่าสูงกว่าวิธี Random forest และ Multi-layer perceptron สำหรับ Random forest เป็น Model ประเภทหนึ่งของ Machine learning ถูกพัฒนาขึ้นจาก Decision tree ต่างกันที่ Random forest เป็นการเพิ่มจำนวน Tree เป็น Tree หลายๆ ต้น ทำให้ประสิทธิภาพในการทำงานสูงขึ้น แม่นยำมากขึ้น ซึ่งโมเดล Random forest เป็นโมเดลที่ได้รับความนิยมไปอย่างมากในการใช้ Machine learning จากผลการวิจัยจะเห็นว่าค่าประสิทธิภาพของวิธี Random forest จะมีความใกล้เคียงกับ Decision tree และ Multi-layer perceptron คือแบบจำลองในกลุ่มของโครงข่ายประสาทเทียม (Neuron network) โดยมีวิธีการคำนวณแบบไม่เชิงเส้น (Non-linear) เหมือนกับวิธี Logistic regression ซึ่งใช้แก้ปัญหาตัวแปรหลายชั้นที่ไม่ชัดเจน (Larose & Larose., 2015; Witten et al., 2017; Sanguansat, 2019) ในขณะที่วิธีการของ Bayes network classifier และ Naive bayes จะใช้วิธีการคำนวณโดยอาศัยหลักการทางสถิติคือ ความน่าจะเป็น ในการจำแนกกลุ่มของข้อมูล ส่วน Sequential minimal optimization เป็นหนึ่งในวิธีการของ Support vector machine

หรือ SVM เป็นเทคนิคหนึ่งในโครงข่ายประสาทเทียมเช่นเดียวกัน โดยใช้หลักการสร้างระนาบหรือเส้นในการตัดสินใจแล้วแบ่งข้อมูลออกเป็น 2 กลุ่ม โดยพยายามสร้างเส้นแบ่งให้อยู่กึ่งกลางระหว่างสองกลุ่มมากที่สุด อาศัยการหาระยะทางจากเส้นไปยังข้อมูลที่อยู่ใกล้เส้นแบ่งมากที่สุดของข้อมูลแต่ละประเภท (Han et al., 2012)

จากผลการวิจัยในภาพรวมทั้งหมดที่กล่าวมานั้น จะเห็นได้ว่าผลจากการทำนายของนักท่องเที่ยวดต่อการจำแนกกลุ่มที่ให้ความสนใจกิจกรรมในการพัฒนาแหล่งท่องเที่ยวเข้าคือ ในทุกแบบจำลองที่นำมาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูลแสดงถึงประสิทธิภาพของข้อมูลที่สามารถนำมาใช้ในการสร้างระบบสารสนเทศด้านการท่องเที่ยว โดยการประยุกต์ใช้ข้อมูลต่าง ๆ จากการออกแบบการเก็บข้อมูลหรือนำข้อมูลในคลังข้อมูลที่มีอยู่เดิมสามารถนำมาค้นหาคำตอบที่ซ่อนอยู่ในเหมืองข้อมูลนั้นได้ ดังนั้น หน่วยงานที่สามารถนำผลการวิจัยนี้ไปใช้ประโยชน์ ได้แก่ องค์การบริหารส่วนจังหวัดเพชรบูรณ์ หรือ สำนักงานการท่องเที่ยวและกีฬาจังหวัดเพชรบูรณ์ จะนำไปพัฒนาต่อยอดการสร้างระบบสารสนเทศการทำนายพฤติกรรมของนักท่องเที่ยวที่มีลักษณะเปลี่ยนแปลงอยู่ตลอดเวลาเมื่อการทำนายขาดความแม่นยำจะทำให้ทราบทันทีว่าควรต้องมีการปรับปรุงและเปลี่ยนแปลง พัฒนาแหล่งท่องเที่ยวให้สอดคล้องกับความต้องการของนักท่องเที่ยวต่อไปในอนาคต

สรุปผลการวิจัย

การทำเหมืองข้อมูลมีบทบาทต่อการใช้งานสารสนเทศที่มีอยู่แล้วในองค์กรหรือการเก็บข้อมูลขึ้นมาใหม่ เพื่อนำความรู้ที่ซ่อนอยู่ในข้อมูลเหล่านั้นมาวิเคราะห์และแปรผลให้เกิดรูปแบบสารสนเทศและนำมาใช้ให้เกิดประโยชน์ ผลการวิจัยนี้แสดงให้เห็นว่าการเลือกใช้แบบจำลอง Decision tree มีความเหมาะสมกับข้อมูลชุดการพัฒนาการท่องเที่ยวอำเภอเขาแก้ว จังหวัดเพชรบูรณ์ การเลือกแบบจำลองที่เหมาะสมกับประเภทของข้อมูลและวัตถุประสงค์ของการหาคำตอบเป็นสิ่งสำคัญที่ผู้ใช้ต้องคำนึงถึงและใช้กระบวนการทดสอบและตรวจสอบประสิทธิภาพของแบบจำลองเหล่านั้น ซึ่งในแต่ละแบบจำลองจะมีอัลกอริธึมที่แตกต่างกัน และเป็นสิ่งสำคัญที่ผู้วิจัยจะต้องทราบข้อดี ข้อเสีย และข้อจำกัดของแต่ละอัลกอริธึมเพื่อให้เกิดการนำแบบจำลองไปใช้งานให้ได้ประโยชน์สูงสุด งานวิจัยนี้สามารถนำไปใช้สร้างนโยบายสาธารณะ วางแผนการท่องเที่ยวและพัฒนาสารสนเทศด้านการท่องเที่ยว อำเภอเขาแก้ว จังหวัดเพชรบูรณ์ต่อไป

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัยจากงบประมาณแผ่นดินประจำปี 2562 มหาวิทยาลัยราชภัฏเพชรบูรณ์ ภายใต้แผนการวิจัยการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อพัฒนาการท่องเที่ยวและเศรษฐกิจเชิงสร้างสรรค์อำเภอเขาแก้ว จังหวัดเพชรบูรณ์ คณะผู้วิจัยขอขอบพระคุณมา ณ โอกาสนี้

เอกสารอ้างอิง

- Auwatanamongkoll S. *Data Mining*. Bangkok: CU Press, 2018.
- Caichompoo S, Caichompoo B, Preechayakul P. Tourism management information system of Khung Bangkachoa area, Phrapadaeng District, Samutprakran Province. *Phranakhon Rajabhat Research Journal (Humanities and Social Science)*. 2017; 12(1): 90-99.
- Elatia S, Ipperciel D, ZAIANE OR. *Data Mining and Learning Analytics Applications in Educational Research*. New Jersey: John Wiley & Sons, Inc., 2016.
- Han J, Kamber M, Pei J. *Data Mining Concepts and Techniques*. 3^{ed}. USA: Elsevier Inc. Morgan Kaufmann Publishers, 2012.
- Kittidachanupap N. *The Development of a Personalized Model for Tourist Attraction*. Thesis. Suranaree University of Technology. Nakhon Ratchasima, 2015.
- Kosinanon M. *Neural Network System for the Tourism Forecasting System*. Thesis. King Mongkut's Institute of Technology North Bangkok, 2006.
- Lake P. *Forecasting International Tourist Arrival to Thailand by Nationality Using Time Series Data Mining Techniques*. KU SRC 1st National Conference. 2016, 453-463.
- Larose DT, Larose CD. *Data Mining and Predictive Analytics*. New Jersey: John Wiley & Sons, Inc., 2015.
- Nuankaew W, Nuankaew P, Bussaman S. et al. A comparative model of cultural tourism in the north and the northeast of Thailand. *Journal of Information Technology Management and Innovation*. 2017; 4(2): 27-38.
- Rukpukdee K, Morachart W. Database Development of Natural Travel Attractions in Ubon Ratchathani Province Using Data Mining Techniques from Tourist's Requirement and Behavior. *Graduate School Journal*. 2018; 11(3): 165-177.
- Sanguansat P. *Artificial Intelligence with Machine Learning*. IDC Premier: Nonthaburi, 2019.
- Songpan W. *Data Mining*. Khon Kaen University: Khon Kaen, 2017.
- Witten IH, Frank E, Hall MA. et al. *Data Mining Practical Machine Learning Tools and Techniques, Fourth Edition*. USA: Elsevier Inc. Morgan Kaufmann Publishers, 2017.
- Yotsawat W, Purisangkaha S, Kittirakpunya W. et al. Development of Domestic Tourists Model Using Clustering and Association Rule Techniques Case Study: Phra Nakhon Si Ayutthaya Province. *The Journal of King Mongkut's University of Technology North Bangkok*. 2017; 27(4): 829-841.