



การสร้างแบบจำลองการแบ่งกลุ่มลักษณะลูกค้าที่ใช้บริการที่พักแบบการท่องเที่ยวเชิงนิเวศ
ในประเทศไทย ด้วยเทคนิคเหมืองข้อมูล

**Consumer characteristic Modeling in Ecotourism Accommodations in Thailand by
Using Data Mining**

นาย ณัฐดนัย เขียววาท

Mr. Natdanai Kiewwath

คณะบริหารธุรกิจ มหาวิทยาลัยแม่โจ้ จังหวัดเชียงใหม่ 50290

บทคัดย่อ

งานวิจัยในครั้งนี้ได้พัฒนาแบบจำลองข้อมูลลักษณะกลุ่มลูกค้าที่ใช้บริการที่พักแบบท่องเที่ยวเชิงนิเวศในประเทศไทยโดยเทคนิคเหมืองข้อมูล เทคนิคที่ทำการสร้างแบบจำลองข้อมูลเลือกใช้การวิเคราะห์ข้อมูลแบบต้นไม้ตัดสินใจและการจัดกลุ่ม K-Mean เพื่อค้นหาความสัมพันธ์ระหว่างตัวแปรของข้อมูลและจำแนกกลุ่มของผู้ที่ใช้บริการที่พักเชิงนิเวศในประเทศไทย เมื่อการสำรวจข้อมูลจากผู้ใช้บริการที่พักเชิงนิเวศกระจายทุกภูมิภาคของประเทศไทยโดยมีการออกแบบสำรวจเพื่อจัดเก็บตัวแปรหลายด้านทั้งข้อมูลพื้นฐานของบุคคล ด้านระบบ Online Review ข้อมูลแหล่งที่พัก ด้านการประเมินระบบในแง่ต่างๆ ประกอบไปด้วย ประโยชน์ของการใช้งานระบบ Online Review ความง่ายในการใช้งานระบบ Online Review ความเป็นส่วนตัวในการใช้งานระบบ Online Review สถานะทางสังคมที่มีต่อการใช้งานระบบ Online Review ความเชื่อมั่นในการใช้งานระบบ Online Review ความปลอดภัยของระบบในการใช้งานระบบ Online Review ความเต็มใจที่จะใช้งานระบบ Online Review ความพึงพอใจในการใช้งานระบบ Online Review โดยการสำรวจข้อมูลเป็นการสำรวจข้อมูลบนระบบออนไลน์ ประเด็นคือการวิเคราะห์ข้อมูลของกลุ่มลูกค้าที่ใช้บริการที่พักแบบท่องเที่ยวเชิงนิเวศต่อการใช้งานระบบ Online Review

จากต้นแบบกลุ่มลูกค้าที่ใช้บริการที่พักแบบท่องเที่ยวเชิงนิเวศโดยใช้เทคนิคเหมืองข้อมูลแบบต้นไม้การตัดสินใจจะได้ปัจจัยที่มีผลต่อการเลือกที่พักผ่านระบบ Online Review ลำดับความสำคัญของปัจจัย การจัดกลุ่มเพื่อหากกลุ่มของลูกค้าเพื่อได้ลักษณะของลูกค้ากลุ่มหลักที่จะใช้บริการที่พักผ่านระบบงาน

Abstract

This research has developed a data model of customer groups that use eco-tourism services in Thailand using data mining techniques. The technique used for data modeling was the decision tree data analysis and K-Mean grouping in order to find the relationship between data variables and classifying those using ecotourism in Thailand. When surveying data from users of ecological accommodations spread in all regions of Thailand. There is a survey design to store many variables, including basic information of the Online Review system, accommodation information. System evaluation in various aspects includes the benefits of using the Online Review system, the ease of use of the Online Review system, the privacy of the Online Review system, the social status of using the Online Review system Confidence in Online System Review System Security for Online Review System, Willingness to Use Online Review System, Satisfaction of Online Review System by surveying data as online data survey. The point is to analyze the data of the customers who use the eco-tourism accommodation to use the Online Review system.

From the prototype of the group of customers who use ecotourism accommodation using tree data mining techniques, the decision will get the factors that affect the selection of accommodations through the Online Review system. Grouping to find groups of customers in order to get the characteristics of the main groups of customers who will use the accommodation through the work system.

กิตติกรรมประกาศ

โครงการวิจัยเรื่อง การสร้างแบบจำลองการแบ่งกลุ่มลักษณะลูกค้าที่ใช้บริการที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย ด้วยเทคนิคเหมืองข้อมูล (Consumer Characteristic Modeling in Ecotourism Accommodations in Thailand by Using Data Mining) ได้จัดทำสำเร็จลุล่วง โดยได้รับทุนอุดหนุนการวิจัยจากคณะ บริหารธุรกิจ มหาวิทยาลัยแม่โจ้ ผู้วิจัยขอขอบคุณอย่างสุดซึ้งต่อหน่วยงานและบุคคลต่างๆที่เกี่ยวข้องกับการวิจัยในครั้งนี้อันได้แก่ กลุ่มผู้ประกอบการแหล่งท่องเที่ยวเชิงนิเวศประเทศไทย นักท่องเที่ยวที่ใช้บริการที่พักเชิงท่องเที่ยวในนิเวศ ผู้ช่วยศาสตราจารย์ ฉัตร ชูชื่น ที่ได้ให้ข้อมูลเพื่อทำการวิจัย ตลอดจนหลักสูตรระบบสารสนเทศทางธุรกิจ บุคลากรคณะบริหารธุรกิจ มหาวิทยาลัยแม่โจ้ ที่ช่วยอำนวยความสะดวกและสนับสนุนด้านข้อมูล สถานที่อุปกรณ์ เทคนิคในการพัฒนาระบบงานอย่างดีเสมอมา

ผู้วิจัย

สารบัญ

เรื่อง

หน้า

บทคัดย่อ

กิตติกรรมประกาศ

สารบัญ

สารบัญตาราง

สารบัญภาพ

บทที่ 1 บทนำ

ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

วัตถุประสงค์ของโครงการวิจัย

ขอบเขตของโครงการวิจัย

ประโยชน์ที่คาดว่าจะได้รับ

นิยามศัพท์

บทที่ 2 การตรวจเอกสาร

แนวคิดและทฤษฎี

งานวิจัยที่เกี่ยวข้อง

กรอบแนวคิดของการวิจัย

บทที่ 3 วิธีวิจัย

ประชากร

กลุ่มตัวอย่าง

เครื่องมือที่ใช้ในการวิจัย

วิธีการเก็บรวบรวมข้อมูล

ก

ค

ง

1

1

1

2

2

3

7

14

15

15

16

18

สารบัญ (ต่อ)

เรื่อง	หน้า
การวิเคราะห์ข้อมูล	19
บทที่ 4 ผลการวิจัย	
รวบรวมข้อมูลและการจัดทำสารสนเทศ	21
การอธิบายลักษณะทั่วไปของกลุ่มตัวอย่าง โดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ	21
การใช้งานระบบการรีวิวออนไลน์ โดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ	29
แสดงผลข้อมูลจากการทำต้นไม้ตัดสินใจและการจำแนกกลุ่มโดย k-mean	33
บทที่ 5 สรุป อภิปรายผล และ ข้อเสนอแนะ	
สรุปผลการวิจัย	39
อภิปรายผลผลการวิจัย	39
ข้อเสนอแนะในการทำวิจัยครั้งต่อไป	40
เอกสารอ้างอิง	41
ภาคผนวก ก ประวัติผู้วิจัย	43

สารบัญตาราง

ตาราง	หน้า
3.1 คะแนนเฉลี่ยและเกณฑ์การแปลความหมาย ระดับปัจจัยของการรีวิวออนไลน์ที่ส่งผลต่อความไว้วางใจในการตัดสินใจในการใช้บริการที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย	17
4.1 เพศของผู้ตอบแบบสอบถาม	21
4.2 ช่วงอายุของผู้ใช้งานบริการระบบรีวิวออนไลน์	22
4.3 ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถาม	23
4.4 อาชีพของผู้ตอบแบบสอบถาม	24
4.5 ภูมิฐานะของผู้ตอบแบบสอบถาม	25
4.6 รายได้เฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม	26
4.7 รายจ่ายเฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม	27
4.8 ประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์	28
4.9 อุปกรณ์ของการใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ	29
4.10 ความถี่ในการใช้งานระบบ Online Review ในการจองห้องพักเชิงนิเวศ	30
4.11 จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาจองห้องพักออนไลน์แต่ละครั้ง	31
4.12 จำนวนรีวิวของผู้เคยใช้บริการจองห้องพักที่จะมีผลต่อตัดสินใจจองห้องพักเชิงนิเวศ	32
4.13 การวัดประสิทธิภาพจำนวน cluster จากระยะเฉลี่ยจากค่ากลาง centroid	36
4.14 ค่ากึ่งกลางของแต่ละ Cluster	37

สารบัญภาพ

ภาพที่	หน้า
2.1 กรอบการวิจัย	14
3.1 ขั้นตอนการวิเคราะห์ข้อมูล	19
4.1 เพศของผู้ตอบแบบสอบถาม	22
4.2 ช่วงอายุของผู้ใช้งานบริการระบบรีวิวออนไลน์	23
4.3 ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถาม	24
4.4 อาชีพของผู้ตอบแบบสอบถาม	25
4.5 ภูมิฐานะของผู้ตอบแบบสอบถาม	26
4.6 รายได้เฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม	27
4.7 รายจ่ายเฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม	28
4.8 ประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์	29
4.9 อุปกรณ์ของการใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ	30
4.10 ความถี่ในการใช้งานระบบ Online Review ในการจองห้องพักเชิงนิเวศ	31
4.11 จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาจองห้องพักออนไลน์แต่ละครั้ง	32
4.12 จำนวนรีวิวของผู้เคยใช้บริการจองห้องพักที่จะมีผลต่อตัดสินใจจองห้องพักเชิง	33
4.13 ภาพการกำหนดค่าเป้าหมายและการกำหนดค่าในแถวกลุ่มทดลอง	33
4.14 การกำหนดกระบวนการต้นไม่ตัดสินใจ	34
4.15 การกำหนดค่าก่อนการประมวลผลต้นไม่ตัดสินใจ	34
4.16 ผลลัพธ์จากการประมวลผลต้นไม่ตัดสินใจ	35
4.17 สรุปจำนวนข้อมูลใน Cluster	37
4.18 ภาพตัวอย่างข้อมูลที่ถูกแบ่ง Cluster	37

บทที่ 1

บทนำ

ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

ในปัจจุบันธุรกิจด้านการให้บริการที่พักซึ่งมีหลากหลายขนาดและหลากหลายรูปแบบ มีการแข่งขันที่สูง มีรูปแบบธุรกิจที่พักประเภทหนึ่งซึ่งเป็นการให้บริการที่พักแบบท่องเที่ยวเชิงนิเวศน์ซึ่งเป็นรูปแบบของที่พักแบบใหม่ที่กำลังได้รับความนิยมมากขึ้นในประเทศไทย เพื่อให้กลุ่มของผู้ประกอบการที่พักเชิงนิเวศสามารถที่จะวิเคราะห์และจำแนกกลุ่มของลูกค้าว่าลูกค้าที่เข้ามาใช้บริการเป็นลูกค้าที่เป็นลูกค้าที่มีคุณลักษณะแบบใดเพื่อที่จะสามารถเข้าใจลักษณะลูกค้าและสามารถเข้าถึงลูกค้าได้อย่างตรงกลุ่ม สามารถรักษากลุ่มลูกค้าเดิมได้อย่างเข้าใจ พัฒนาเพิ่มเติมในรูปแบบของการท่องเที่ยวเชิงนิเวศน์ให้ดึงดูดกลุ่มลูกค้าใหม่ที่มีลักษณะใกล้เคียงกันได้ในอนาคต ดังนั้นจึงมีการนำเอาเทคโนโลยีสารสนเทศเข้ามาช่วยในการวิเคราะห์และจำแนกจัดกลุ่มลูกค้าของที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย เพื่อหาลักษณะของลูกค้าโดยใช้เทคนิคการทำเหมืองข้อมูลตามกรอบการทำเหมืองข้อมูล CRISP-DM เพื่อหาความสัมพันธ์ของลักษณะของลูกค้า ขั้นตอนวิธี Apriori และการแบ่งกลุ่ม (Cluster) โดยอ้างอิงจากข้อมูลการสำรวจลูกค้าที่เข้าพัก เพื่อให้กลุ่มเจ้าของธุรกิจและผู้ให้บริการที่พักแบบการท่องเที่ยวเชิงนิเวศของไทยได้นำเอาผลลัพธ์ที่ได้นั้นก็คือได้ทราบถึงลักษณะของกลุ่มลูกค้าเป้าหมาย เพื่อการรักษากลุ่มลูกค้าและการเข้าถึงกลุ่มลูกค้าได้อย่างรวดเร็วในอนาคต

วัตถุประสงค์ของโครงการวิจัย

- 1) เพื่อศึกษารวบรวมจำแนกและจัดกลุ่มจัดกลุ่มลูกค้าที่ใช้บริการที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทยด้วยเทคนิคการทำเหมืองข้อมูล
- 2) เพื่อนำเสนอแบบจำลองการแบ่งกลุ่มลูกค้าที่เข้าพักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทยสำหรับการเข้าถึงกลุ่มลูกค้าได้อย่างรวดเร็วและตรงกลุ่มเป้าหมาย

ขอบเขตของโครงการวิจัย

1. ขอบเขตด้านเนื้อหา

การศึกษาปัจจัยและลักษณะลูกค้าที่ใช้บริการในธุรกิจที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย ประกอบด้วยปัจจัยการศึกษา 4 ปัจจัย คือ

- 1) Generation เป็นการแบ่งกลุ่มของบุคคลตามช่วงอายุ โดยแบ่งออกเป็น 4 กลุ่ม คือ Gen B Gen X Gen Y Gen Z
- 2) ระดับพฤติกรรมการใช้โทรศัพท์มือถือ อินเทอร์เน็ต และโซเชียลเน็ตเวิร์ก
- 3) ข้อมูลพื้นฐานทั่วไป ข้อมูลอาชีพ ช่วงรายได้ ลักษณะที่อยู่อาศัยและจังหวัดที่อยู่อาศัย
- 4) รูปแบบและลักษณะของการท่องเที่ยว

2. ขอบเขตด้านประชากร

ประชากรและกลุ่มตัวอย่างในการศึกษาที่ใช้ในการศึกษาวิจัยในครั้งนี้ คือ กลุ่มผู้ใช้บริการหรือลูกค้าที่มีประสบการณ์ในการใช้ที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย จำนวน 200 คน

3. ขอบเขตด้านระยะเวลา

ระยะเวลาที่ใช้ในการเก็บรวบรวมข้อมูลครั้งนี้ เริ่มตั้งแต่เดือนมกราคม 2562 – เดือนสิงหาคม 2562

4. ขอบเขตด้านสถานที่

ศึกษาและเก็บข้อมูลจากกลุ่มผู้ใช้บริการหรือลูกค้าที่มีประสบการณ์ในการใช้ที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย

ประโยชน์ที่คาดว่าจะได้รับ

- 1) เพื่อทราบถึงลักษณะของกลุ่มลูกค้าที่ใช้บริการที่พักรูปแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย
- 2) เพื่อได้แบบจำลองการแบ่งกลุ่มลูกค้าที่แสดงถึงความสำคัญของแต่ละปัจจัยที่มีผลต่อการใช้บริการในธุรกิจที่พักรูปแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย
- 3) นำผลการศึกษาที่ได้เพื่อไปเผยแพร่ในการประชุมวิชาการหรือวารสารทางวิชาการระดับชาติหรือนานาชาติ

นิยามศัพท์

การทำเหมืองข้อมูล (Data Mining)

การจำแนกและการทำนายข้อมูล Classification & Prediction

การแบ่งกลุ่มข้อมูลแบบเคมีน (k-means clustering)

รีวิวนออนไลน์

การท่องเที่ยวเชิงนิเวศ

บทที่ 2

การตรวจเอกสาร

แนวคิดและทฤษฎี

การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไป ประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้าน วิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม เทคนิคต่างๆในการจำแนกประเภท ของข้อมูล อาทิ เช่น การจำแนก ข้อมูลด้วยการสร้างต้นไม้ตัดสินใจ (Decision tree classifier) การ จำแนกข้อมูลด้วยเบย์เซียนและเบย์เซียนบิลิฟเน็ตเวิร์ค (Bayesian classifier and Bayesian belief networks) การจำแนกข้อมูลด้วยกฎ (Rulebased classifiers) การจำแนกข้อมูลด้วยโครงข่าย ประสาทเทียมและการส่งค่าย้อนกลับ (Neural network and backpropagation) การจำแนกข้อมูลจาก กฎความสัมพันธ์ของข้อมูล (Classification based on association rule mining) การค้นหาเพื่อนบ้าน ใกล้ที่สุด k อันดับ (k-nearest-neighbor) และทำการศึกษา เกี่ยวกับการทำนายข้อมูล ที่จะประกอบไป ด้วยการถดถอยเชิงเส้นตรง (linear regression) และการถดถอยที่ ไม่เป็นเส้นตรง (Nonlinear regression) การทำเหมืองข้อมูล มีการใช้เทคนิคหลายแบบในการวิเคราะห์ข้อมูล โดยมากใช้ 3 เทคนิคใหญ่ๆ คือ Classification โดยสร้างโมเดลจากข้อมูลที่มีอยู่เพื่อทำนายเหตุการณ์ที่จะเกิดขึ้น ในอนาคต Clustering แบ่งข้อมูลเป็นหลายๆ กลุ่มอาศัยความคล้ายคลึงกันของข้อมูล Association rules อาศัยความสัมพันธ์ของข้อมูลที่เกิดร่วมกัน สร้างเป็นกฎความสัมพันธ์ เช่น “ซื้อเบียร์แล้วจะ ซื้อผ้าอ้อมไปด้วย”

ขั้นตอนการทำเหมืองข้อมูล (Process of Data Mining)

ขั้นตอนในการวิเคราะห์ข้อมูลได้ใช้หลักการของกระบวนการหาความรู้แบบ Cross Industry Standard Process for Data Mining (CRISP-DM) ซึ่งเป็นแนวทางในการดำเนินงาน CRISP-DM ประกอบไปด้วย 6 ขั้นตอน

1) ความเข้าใจทางธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางการค้า ไม่นับรวมทั้งวางแผนในการดำเนินการคร่าวๆ

2) ความเข้าใจเกี่ยวกับข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาเพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3) การเตรียมข้อมูล (Data Preparation) ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็ขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM

4) การพัฒนาแบบจำลอง (Modeling) ขั้นตอนนี้จะเป็ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางการค้า ไม่นับ ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูล หรือ การแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปขั้นตอน Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย เทคนิคในการวิเคราะห์ข้อมูลต่างๆ เช่น

- การแบ่งกลุ่มข้อมูล (Clustering)
- การหากฎความสัมพันธ์ (Association Rules)
- การจำแนกประเภทข้อมูล (Classification) ตัวอย่างเช่น
 - เทคนิค Decision Tree
 - เทคนิค Naive Bayes
 - เทคนิค Neural Network
 - เทคนิค Support Vector Machines (SVM)

5) การทดสอบแบบจำลอง (Evaluation) ในขั้นตอนนี้เราจะได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทางการค้า ไม่นับแล้วแต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้ สำหรับการสร้างโมเดลด้วยเทคนิค Classification มีการทดสอบประสิทธิภาพของโมเดลอยู่ 3 แบบใหญ่ คือ

- Self-consistency test
- Split test
- Cross-validation test

6) การนำแบบจำลองไปใช้ (Deployment) ในกระบวนการทำงานของ CRISP-DM นั้นไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคทางค่า ไม่นิ่งเท่านั้น แม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่มีประโยชน์ แต่จะต้องนำองค์ความรู้ที่ได้เหล่านี้ไปใช้ได้จริง ในองค์กรหรือบริษัท ตัวอย่างเช่น การสร้างรายงานเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจได้ง่าย และสามารถนำไปออกโปรโมชันได้ เป็นต้น

การจำแนกและทำนายข้อมูล Classification & Prediction

Classification เป็นแขนงหนึ่งในการทำ Data mining ซึ่งเป็นศาสตร์ที่ใกล้เคียงกับการเรียนรู้ ด้วยเครื่องจักร (Machine Learning) โดยใช้ข้อมูลฝึกฝน (Training set) นำมาทำให้เครื่องจักรเรียนรู้ และสร้างแบบจำลองเพื่อนำไปใช้ในการทำนายผลข้อมูล โดยส่วนที่ต้องการทำนายจะถูกเรียกว่า Class label or Class การทำ Data Mining เป็นวิธีการนำข้อมูลฝึกฝน (Training Set) มาจำแนกโดยอาศัยคุณลักษณะ (Attribute), ประเภท หรือ คลาส (Class) ของแต่ละข้อมูลนำมาสร้างแบบจำลอง (Model) ด้วยกระบวนการเรียนรู้ด้วยเครื่อง (Machine Learning) จุดประสงค์ของการจัดจำแนกประเภท คือ สามารถนำแบบจำลองที่สร้างขึ้นมาทำนายข้อมูลที่ไม่เคยพบมาก่อน (ข้อมูลในอนาคต) แล้วได้ผลลัพธ์ถูกต้องแม่นยำเป็นที่น่าพอใจ

การแบ่งกลุ่มข้อมูล (data clustering)

เป็นวิธีการวิเคราะห์ข้อมูล ซึ่งใช้ในการเรียนรู้ของเครื่อง การทำเหมืองข้อมูลโดยจะแบ่งชุดข้อมูล (มักจะเป็นเวกเตอร์) ออกเป็นกลุ่ม (cluster) นำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือ คล้ายกันจัดไว้ในกลุ่มเดียวกัน ขั้นตอนวิธีที่ใช้ในการแบ่งกลุ่มจะอาศัยความเหมือน (similarity) หรือ ความใกล้ชิด (proximity) โดยคำนวณจากการวัดระยะระหว่างเวกเตอร์ของข้อมูลเข้า โดยใช้การวัดระยะแบบต่าง ๆ เช่น การวัดระยะแบบยูคลิด (Euclidean distance) การวัดระยะแบบแมนฮัตตัน (Manhattan distance) การวัดระยะแบบเชบิเชฟ (Chebychev distance)

การแบ่งกลุ่มข้อมูลจะแตกต่างจากการแบ่งประเภทข้อมูล (classification) โดยจะแบ่งกลุ่มข้อมูลจากความคล้าย โดยไม่มีการกำหนดประเภทของข้อมูลไว้ก่อน จึงกล่าวได้ว่าการแบ่งกลุ่มข้อมูล เป็นการเรียนรู้แบบไม่มีผู้สอน

ขั้นตอนวิธีการแบ่งกลุ่ม ได้แก่ k-means clustering, hierarchical clustering, self-organizing map (som)

การแบ่งกลุ่มข้อมูลอาจใช้เป็นขั้นตอนเบื้องต้นของการวิเคราะห์ข้อมูล เพื่อช่วยในการลดขนาดข้อมูล (แยกเป็นหลาย ๆ กลุ่มและคัดเฉพาะบางกลุ่มเพื่อทำการวิเคราะห์ต่อไป หรือแยกการวิเคราะห์ออกเป็นสำหรับแต่ละกลุ่ม) ก่อนที่จะนำไปวิเคราะห์ด้วยวิธีการอื่นต่อไป

ขั้นตอนวิธีในการแบ่งกลุ่มข้อมูล โดยทั่วไปแบ่งได้เป็น 2 ประเภทใหญ่ๆ คือ การแบ่งแบบเป็นลำดับขั้น (hierarchical) และ การแบ่งแบบตัดเป็นส่วน (partitional) การแบ่งแบบเป็นลำดับขั้นนั้น จะมีทำการแบ่งกลุ่มจากกลุ่มย่อยที่ถูกแบ่งไว้ก่อนหน้านั้นซ้ำหลายครั้ง ส่วนการแบ่งแบบตัดเป็นส่วนนั้น การแบ่งจะทำเพียงครั้งเดียว การแบ่งแบบเป็นลำดับขั้น จะมี 2 ลักษณะ คือ แบบล่างขึ้นบน(bottom-up) หรือ เป็นการแบ่งแบบรวมกลุ่มจากกลุ่มย่อยให้ใหญ่ขึ้นไปเรื่อยๆ โดยเริ่มจากกลุ่มเล็กสุดคือในแต่ละกลุ่มมีข้อมูลเพียงตัวเดียว และ แบบบนลงล่าง (top-down) หรือ เป็นการแบ่งแบบกลุ่มจากกลุ่มใหญ่ให้ย่อยไปเรื่อยๆ โดยเริ่มจากกลุ่มใหญ่ที่สุด คือกลุ่มเดียวมีข้อมูลทุกตัวอยู่ในกลุ่ม

การแบ่งกลุ่มข้อมูลแบบเคมีน (k-means clustering)

เป็นวิธีหนึ่งในวิธีการแบ่งเวกเตอร์ (vector quantization) ที่มีรากฐานมาจากการประมวลผลสัญญาณ วิธีนี้เป็นที่นิยมสำหรับการแบ่งกลุ่มข้อมูล (cluster analysis) ในการทำเหมืองข้อมูล (data mining) การแบ่งกลุ่มข้อมูลแบบเคมีน ใช้สำหรับการแบ่งการสังเกตจำนวน n สิ่งเป็น k กลุ่ม โดยแต่ละการสังเกตจะอยู่ในกลุ่มที่มีค่าเฉลี่ย(ที่ใช้เป็นแม่แบบ)ใกล้เคียงกันที่สุด โดยวิธีนี้จะเป็นการแบ่งพื้นที่ข้อมูลไปเป็นแผนภาพไวโรนอย

วิธีการจัดกลุ่มนี้อยู่ในกลุ่มความซับซ้อนของปัญหาเอ็นพีแบบยาก (NP-hard) แต่อย่างไรเราสามารถนำขั้นตอนวิธีแบบศึกษาสำนึก (heuristic algorithm) มาใช้หาจุดศูนย์กลางของกลุ่มข้อมูลจากการลู่อเข้าได้อย่างมีประสิทธิภาพ ซึ่งจะเหมือนกับขั้นตอนวิธีหาค่าคาดหวังสูงสุด (expectation-maximization algorithm) สำหรับโมเดลแบบผสม (Mixture Model) ของการแจกแจงปรกติ (Gaussian distribution) เนื่องจากทั้งสองขั้นตอนวิธีจะใช้แนวทางกระทำซ้ำการกลั่นกรอง (iterative refinement approach) นอกจากนี้ ทั้งสองขั้นตอนวิธียังใช้จุดศูนย์กลางของคลัสเตอร์สร้างแบบจำลองข้อมูล อย่างไรก็ตาม การแบ่งกลุ่มข้อมูลแบบเคมีนมีแนวโน้มจะได้คลัสเตอร์ผลลัพธ์ที่มีตำแหน่งขอบเขตใกล้เคียงกัน ในขณะที่ขั้นตอนวิธีหาค่าคาดหวังสูงสุดนั้นยอมให้คลัสเตอร์ผลลัพธ์มีรูปร่างที่แตกต่างกันได้ ขั้นตอนวิธีนี้ไม่มีอะไรเกี่ยวข้องกับวิธีการค้นหาเพื่อนบ้านใกล้สุด (k-nearest neighbor) ซึ่งเป็นเทคนิคการเรียนรู้ของเครื่อง (machine learning) ที่เป็นที่นิยมอีกอย่างหนึ่ง

ระบบการรีวิวออนไลน์ (Online Customer Reviews)

ในปัจจุบันเนื่องจากความนิยมของสื่อสังคมออนไลน์ ที่มีผู้ใช้งานจำนวนมากเข้ามาแลกเปลี่ยนข้อมูลต่างๆ ซึ่งกันและกัน ทำให้การใช้งานระบบการรีวิวออนไลน์ (Online Customer Reviews) กลายเป็น ช่องทางสำคัญในการหาข้อมูลของลูกค้ายุคใหม่ก่อนการตัดสินใจซื้อหรือใช้บริการ เนื่องจากความเป็นอิสระของผู้วิจารณ์ที่ไม่มีส่วนได้เสียในตราสินค้า และเพราะความนิยม รวมถึงการเข้าถึงอินเทอร์เน็ตที่ง่ายดายมากยิ่งขึ้น ทำให้ ระบบการรีวิวออนไลน์มีจำนวนเพิ่มมากขึ้นเรื่อยๆ ในทุกเว็บไซต์ที่มีการเปิดให้มีการเขียนการรีวิวออนไลน์ ทำให้ผู้ใช้งานเผชิญหน้ากับสถานการณ์ที่ข้อมูลมีมากเกินไปจนเกิดความจำเป็น ที่เรียกว่า ปรากฏการณ์ Information Overloading (Malhotra, 1984) โดยยังมีการวิจัยพบว่าความสามารถในการรับรู้ข้อมูลจะลดลงเมื่อมีข้อมูลที่มากเกินไป ซึ่งจะส่งผลกระทบต่อคุณภาพของการตัดสินใจในท้ายสุด (Jacoby et al., 1974) โดยยังพบว่า การที่มีข้อมูลจำนวนมากเกินไปทำให้ผู้บริโภคจำเป็นต้องมีการเลือกที่จะรับข้อมูล โดยมีการอ่านแบบผ่านๆ และรับรู้เฉพาะข้อมูลแบบองค์รวม ซึ่งในระหว่างกระบวนการนั้นผู้บริโภคจะยังคงมีความกังวลต่อรายละเอียดที่อ่านข้ามไป ทำให้เกิดเป็นความสับสน ไม่แน่ใจ ไม่ใช่ว่าการมีข้อมูลเยอะ จะเป็นผลเสียแต่เพียงอย่างเดียว แต่ผู้บริโภคสามารถรับรู้ได้ว่าสินค้าหรือบริการนั้น ได้รับความนิยมนี่ไม่ว่าจะมีข้อมูลในทิศทางบวกหรือทางลบ (Park and Lee, 2007)

งานวิจัยที่เกี่ยวข้อง

Catherine Bounsaythip and Esa Rinta-Runsala (2001) ได้วิจัยเกี่ยวกับลูกค้าสัมพันธ์ (CRM: Customer Relationship Management) โดยเฉพาะการจัดกลุ่มลูกค้าและการเก็บรวบรวมข้อมูลลูกค้า และการใช้การทำเหมืองข้อมูลในการสนับสนุนการตัดสินใจ เริ่มด้วยการอธิบายถึงการทำนายพฤติกรรมลูกค้า เช่น การเก็บข้อมูลและการเตรียมข้อมูล การแบ่งกลุ่มลูกค้าและการจำลองข้อมูลลูกค้า จากนั้นจะเสนอภาพรวมโดยทั่วไปของวิธีการทำเหมืองข้อมูลรวมถึง Cluster Discovery, Decision Tree, Neural Networks, Association Rule และ Sequential Pattern Discovery ยังครอบคลุมไปถึง Web Mining ที่แยกออกมาเนื่องจากเป็นที่นิยมหัวข้อสุดท้ายจะกล่าวถึงแนวทางในการเลือก Data Mining

บวร น้อยแสง(2549) ได้ศึกษาและพัฒนาคลังข้อมูลของระบบการขายและประยุกต์ขั้นตอนวิธีการทำเหมืองข้อมูลกับคลังข้อมูลที่พัฒนาขึ้นสำหรับช่วยในการวิเคราะห์การขาย ซึ่งในการทำ

เหมือนข้อมูลจะเป็นการจัดกลุ่มลูกค้าโดยวิธีแบบ K-Means ใช้ระยะทางแบบยุคลิด ในการรวมกลุ่มลูกค้า

ส่วนการพัฒนาคลังข้อมูลใช้รูปแบบ Schema คือ Star Schema ประกอบด้วยตัววัดคือ จำนวนรวมของสินค้าที่ขาย ราคาสินค้าเฉลี่ย ต้นทุนเฉลี่ยของสินค้า มูลค่ารวมในการขายสินค้า และมีมิติ (Dimension) อยู่ 4 มิติคือ เวลา สินค้า ลูกค้า พนักงาน โดยนำเสนอสารสนเทศของการวิเคราะห์การขายอยู่ในรูปของระบบ OLAP (On-Line Analytic Processing) โดยใช้โปรแกรม Microsoft SQL Server 2005 ซึ่งเป็นโปรแกรมจัดการฐานข้อมูลในการดำเนินงาน การใช้งานระบบ OLAP ช่วยให้การวิเคราะห์การขายมีความรวดเร็วในการเรียกดูข้อมูลหรือสารสนเทศและยังสามารถเรียกดูสารสนเทศในเชิงมิติต่างๆ ที่มีความซับซ้อนได้ เพื่อช่วยสนับสนุนในการตัดสินใจในการบริหารงานด้านการขาย

ส่วนการทำเหมืองข้อมูล ได้ใช้ข้อมูลที่ได้จากคลังข้อมูลซึ่งมีการกรองข้อมูลให้มีความถูกต้องของข้อมูลมากขึ้นเนื่องจากขั้นตอนวิธีในการทำเหมืองข้อมูลนั้นมีมากมายไม่ว่าจะเป็นการใช้ขั้นตอนวิธีทางคณิตศาสตร์ ทางสถิติ หรือทางคอมพิวเตอร์เข้ามาช่วยจัดการข้อมูลที่มีขนาดใหญ่ เพื่อให้ได้สารสนเทศหรือสามารถแก้ปัญหาที่ต้องการได้ ในการศึกษาครั้งนี้จึงใช้ขั้นตอนวิธีการรวมกลุ่มแบบ K-Means เพื่อรวมกลุ่มลูกค้าเพื่อช่วยให้สามารถดำเนินนโยบายทางการขาย การส่งเสริมการขายการสร้างความสัมพันธ์กับลูกค้า หรือ การบริหารงานด้านการขายอื่นๆ ได้ถูกกลุ่มเป้าหมายมากขึ้นในการทำเหมืองข้อมูลครั้งนี้ใช้แบบจำลองสำหรับการดำเนินงานแบบ CRISP-DM และในวิเคราะห์การรวมกลุ่มแบบ K-Means

บุษยรัตน์ ต้นสุทธิเวสส์ (2549) ทำการค้นคว้าข้อมูลที่เกี่ยวข้องจัดเก็บไว้แล้วมาสร้างแบบจำลองพฤติกรรมผู้บริโภคโดยการทำเหมืองข้อมูลและการคำนวณทางสถิติซึ่งดำเนินการตามกระบวนการค้นหาข้อมูลที่มีประโยชน์ (KDD Process) ที่ซ่อนอยู่ในระบบฐานข้อมูล โดยผลที่ได้จากการทำเหมืองข้อมูลนี้คือ แบบจำลองพฤติกรรมผู้บริโภคในลักษณะต่างๆ กันไปตามวิธีการทำเหมืองข้อมูล ซึ่งแบบจำลองดังกล่าวจะเป็นข้อมูลที่แสดงในรูปแบบความสัมพันธ์กันระหว่างพฤติกรรมผู้บริโภคในอนาคต และสามารถแบ่งกลุ่มผู้บริโภคจากสมบัติหรือพฤติกรรมผู้บริโภคนั้นคือ ผู้บริโภคที่มีสมบัติหรือพฤติกรรมคล้ายกันจะอยู่กลุ่มเดียวกัน และผู้บริโภคที่มีสมบัติหรือพฤติกรรมต่างกันจะอยู่คนละกลุ่มกัน นอกจากนี้ยังสามารถทำนายสมบัติหรือพฤติกรรมของผู้บริโภคที่ทำให้เป็นลูกค้าชั้นดีได้จากข้อมูลของผู้บริโภคเท่าที่มีอยู่

หลังจากการวัดผลที่ได้จากแบบจำลองเป็นที่น่าพอใจแล้ว จึงนำแบบจำลองนี้มาประยุกต์ใช้เพื่อเสนอแนวทางในการทำกิจกรรมส่งเสริมการขายการตลาดตั้งแต่ขั้นตอนการหาลูกค้า

ที่มีโอกาสเป็นลูกค้าชั้นดี การแบ่งกลุ่มลูกค้า การจัดกิจกรรมส่งเสริมการขายเพื่อเป็นส่วนช่วยให้ลูกค้ามีความภักดีกับองค์กร รวมถึงการกระตุ้นการซื้อของลูกค้า โดยการนำเสนอสินค้าเมื่อลูกค้าซื้อสินค้าที่จูงขายสินค้า (Cross-Selling) การนำเสนอรายการสินค้าที่เหมาะสมแก่ลูกค้าแต่ละราย (Personalized Shopping List) การจัดทำรายการส่งเสริมการขายให้กลุ่มลูกค้าแต่ละกลุ่ม

Qi Luo (2008) กล่าวว่า การค้นหาความรู้และการทำเหมืองข้อมูลได้กลายมาเป็นส่วนสำคัญเนื่องจากความต้องการมีมากขึ้น ซึ่งรวมถึงการใช้ในการเรียนรู้ของเครื่อง (Machine Learning), ฐานข้อมูล (Databases), สถิติ (Statistics), การจัดหาความรู้ (Knowledge Acquisition), ข้อมูลจินตทัศน์ (Data Visualization) และ การประมวลผลคุณภาพสูง (High Performance Computing) การค้นหาความรู้ และการทำเหมืองข้อมูลมีประโยชน์อย่างมากในหลายๆ ด้านของปัญญาประดิษฐ์ (Artificial Intelligence) เช่น อุตสาหกรรม พาณิชยกรรม การศึกษา เป็นต้น ซึ่งความสัมพันธ์ระหว่างความรู้กับเหมืองข้อมูลและกระบวนการค้นหาความรู้ ฐานข้อมูล (KDD) ทฤษฎีการทำเหมืองข้อมูล ขั้นตอนการทำเหมืองข้อมูล เทคโนโลยีเหมืองข้อมูล การใช้เหมืองข้อมูล

เป้าหมายสำคัญของการค้นหาความรู้และการทำเหมืองข้อมูลคือ “เปลี่ยนข้อมูลให้เป็นความรู้” ยกตัวอย่างเช่น ความรู้ที่ได้จากฐานข้อมูลทางการแพทย์ อาจนำไปตีพิมพ์ในวารสารทางการแพทย์ ความรู้ที่ได้จากการวิเคราะห์ฐานข้อมูลทางการเงินหรือการตลาดอาจนำไปปรับปรุงการปฏิบัติงานในธุรกิจและกำหนดหลักสูตรการบริหาร

การทำเหมืองข้อมูลเป็นส่วนหลักในการค้นหาความรู้ในฐานข้อมูล (KDD) ซึ่งมีกระบวนการ ซึ่งอาจ มีขั้นตอนดังนี้ การเลือกข้อมูล การกรองข้อมูล การแปรรูปข้อมูล การค้นหา รูปแบบข้อมูล (เหมืองข้อมูล) การหาสิ่งที่จะนำเสนอ การตีความ การประเมินผล บ่อยครั้งที่การทำเหมืองข้อมูลและ KDD สามารถใช้แทนกันได้เนื่องจากการทำเหมืองข้อมูลเป็นกุญแจสำคัญในกระบวนการ KDD

พรรณธิดา เพชรบุญมี, ดวงกมล โพธิ์นาค และมนต์ชัย เทียนทอง(2556)การพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคลปี ตัวแบบที่พัฒนาขึ้น โดยใช้กฎการจำแนกเทคนิคต้นไม้การตัดสินใจ ถูกสร้างและทดสอบตัวต้นแบบด้วยวิธีการตรวจสอบไขว้(K-fold crossvalidation) และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (percentage split) ทำการวิเคราะห์ข้อมูลด้วย โปรแกรม WEKA ผลการทดลองพบว่า ได้กฎการจำแนกข้อมูลทั้งสิ้นจำนวน 8 เมื่อวัดค่าประสิทธิภาพของตัวแบบ การพัฒนาตัวต้นแบบด้วยวิธีการแบ่งข้อมูลการพัฒนาตัวแบบด้วยวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 66.00 จะมีค่าความถูกต้องเท่ากับร้อยละ 85 ค่า

ความแม่นยำ เท่ากับร้อยละ 88.20 ค่าความระลึกลับเท่ากับร้อยละ 85.00 และค่าความถ่วงดุลเท่ากับร้อยละ 85.10 ซึ่งประสิทธิภาพทุกค่าสูงกว่าวิธีการอื่น โดยงานวิจัยนี้สามารถนำไปใช้ในการพัฒนาตัวแบบการพยากรณ์รูปแบบ การเรียนรู้ตามประสบการณ์ของเดวิด โคล์ป โดยใช้กฎการจำแนกเทคนิคต้นไม้การตัดสินใจที่มีความถูกต้องและ แม่นยำในการทำนายรูปแบบการเรียนรู้ได้เป็นอย่างดี

ปรีชา ลิ่มตระกูล, วิชา เจริญกัณท์ และวิทยา พรพัชรพงศ์. (2559) การใช้ข้อมูลปัจจัยการผลิตมันสำปะหลังจากสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อสร้างตัวแบบ พยากรณ์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) โดยจะใช้ อัลกอริทึม จำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes, และ LADTree แล้ว ทำการทดสอบตัวแบบการพยากรณ์ด้วยวิธี Cross-validation Test พบว่ายังให้ค่าแม่นยำเพียง 70.96% ซึ่ง ผู้วิจัยเห็นว่ายังไม่ดีพอ จึงได้ปรับปรุงวิธีการทดสอบ โดยการแบ่งชุดข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) จำนวน 5 ชุด แล้วทำการสร้างตัวแบบการพยากรณ์ด้วยอัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และ LADTree อีกครั้งพบว่า ค่าความแม่นยำสูงขึ้นกว่าวิธี Cross-validation Test ในทุกอัลกอริทึม ดังนั้นผู้วิจัยจึงเลือกตัวแบบการพยากรณ์ที่ให้ค่าความแม่นยำสูงสุด ในส่วนของข้อมูลทดสอบ (Test Set) อัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่ 80.12% และ อัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่ 89.55%

ฝนทิพย์ คุณแก้ว. (2555) การทำเหมืองข้อมูลมีบทบาทในการวิเคราะห์ข้อมูลและสารสนเทศขององค์กรและ หน่วยงานทั้งภาครัฐกิจ การแพทย์วิศวกรรมและอื่นๆ จุดเด่นของการทำเหมืองข้อมูล คือ ช่วย สนับสนุนเทคนิคการวิเคราะห์ข้อมูลที่สามารถค้นพบความสัมพันธ์ภายในกลุ่มข้อมูล ในงานนี้ได้เห็นจุดเด่นของการจำแนกหาลักษณะข้อมูลในแต่ละกลุ่ม เพื่องานทางการแพทย์ที่มีประโยชน์ ต่อแพทย์และผู้ป่วยที่อาจจะสื่อสารกัน ได้ผ่านทางโมเดลที่จัดทำขึ้นมาจากการทำเหมืองข้อมูล โมเดลเพื่อการจำแนกนี้สามารถแสดงผลออกมาได้ในรูปแบบ โครงสร้างต้นไม้การตัดสินใจแต่เมื่อใดที่เราอยากรู้ข้อมูลเพียงบางส่วน ผลลัพธ์ของการทำเหมืองข้อมูลตามปกติจะไม่เอื้อต่อการตอบคำถามที่มีลักษณะเจาะจง เช่น นี้ผู้วิจัยได้เห็นความสำคัญ ในการพัฒนาวิธีการจำแนกข้อมูลตาม ข้อกำหนดที่ผู้ใช้ระบุ โดยงานวิจัยใช้วิธีการเขียน โปรแกรมเชิงตรรกะเพื่อพัฒนาฟังก์ชันการทำงานของการทำงานการทำเหมืองข้อมูลแบบจำแนก เพื่อให้ ได้ซึ่งแพทเทิร์นของข้อมูลที่นำมาจาก แหล่งข้อมูลจริง และได้ผลในรูปแบบของกฎ ถ้า แล้ว และเพิ่มการเขียน โปรแกรมเชิงตรรกะในรูปแบบที่สร้างเงื่อนไขให้กับขั้นตอนการการแสดงผลลัพธ์ซึ่งจะช่วยในการคัดเลือกเฉพาะบาง

กฎที่ผู้ใช้สนใจ เพื่อเพิ่มประสิทธิภาพในการค้นหาคำตอบที่ตรงกับความต้องการของผู้ใช้และได้ผลของแพทเทิร์นที่กระชับ ไม่มีข้อมูลที่เกินความจำเป็นที่ผู้ใช้ไม่ต้องการอีกทั้งยังช่วยลดเวลาในการค้นหาทุกทั้งหมดแทนที่การค้นหาจะต้องค้นหาทุกทั้งหมดที่เป็นไปได้เนื่องจากกฎที่ได้จากการจำแนกข้อมูลแล้ว ได้ถูกบันทึกไว้ในหน่วยความจำซึ่งสามารถค้นหาด้วยเงื่อนไขแบบใดก็ได้

ฉิมจารย์ และ คณะ (2015) ได้ทำการวิจัยเรื่อง ปัจจัย ที่ส่งผลต่อความพึงพอใจในการใช้ Cloud Computing ของนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีราชมงคล อีสาน ศูนย์กลาง โดยมีวัตถุประสงค์ 1) เพื่อศึกษาระดับความพึงพอใจของนักศึกษาที่เข้ามาใช้งาน คลาวด์คอมพิวเตอร์ 2) เพื่อเปรียบเทียบความพึงพอใจของนักศึกษาที่มีต่อการใช้คลาวด์คอมพิวเตอร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นนักศึกษาหลักสูตรวิทยาการ คอมพิวเตอร์ สาขาวิชา เทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยี ราชมงคลอีสาน ศูนย์กลาง ที่ลงทะเบียนเรียนในรายวิชา Operating System, File Processing และ รายวิชา Microcomputer System and Interfacing ดำเนินการวิจัยโดยใช้แบบสอบถามในการศึกษา ความพึงพอใจในการใช้คลาวด์คอมพิวเตอร์ โดยใช้ความสามารถของคลาวด์รูปแบบ SaaS (Software as a Service) ผ่านระบบสารสนเทศ ได้แก่ ระบบคอมพิวเตอร์เสมือน Facebook และ Dropbox ผลการวิจัยพบว่า เพศที่แตกต่างกันมีความพึงพอใจในการใช้คลาวด์คอมพิวเตอร์ไม่แตกต่างกัน ส่วน ความพึงพอใจใน รายวิชา Operating System, File Processing และรายวิชา Microcomputer System and Interfacing พบว่ามีความพึงพอใจแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

นวรรตน์ มินุชนาถ (2012) ได้ทำการศึกษาเรื่องการยอมรับการใช้งานระบบจัดการเอกสาร อิเล็กทรอนิกส์ของการเคหะแห่งชาติ มีวัตถุประสงค์เพื่อ เพื่อศึกษาระดับการรับรู้ประโยชน์และ ความง่ายต่อการใช้งาน ระบบจัดการเอกสารอิเล็กทรอนิกส์ของบุคลากรการเคหะแห่งชาติ กลุ่ม ตัวอย่างที่ใช้ในการศึกษา คือ เจ้าหน้าที่ธุรการของการเคหะแห่งชาติ จำนวน 120 คน วิธีการศึกษา ได้ทำการเก็บข้อมูลกลุ่มตัวอย่างจากการใช้แบบสอบถาม การวิเคราะห์ข้อมูลใช้สถิติเชิง พรรณนา โดยการหาค่าร้อยละ ค่าความถี่ ส่วนเบี่ยงเบนมาตรฐาน การหาค่าเฉลี่ยเพื่ออธิบายข้อมูล เบื้องต้น เกี่ยวกับกลุ่มตัวอย่าง ใช้การวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA) ใช้การ ทดสอบ กับกลุ่มตัวอย่างที่ มากกว่า 2 กลุ่ม ถ้าพบความแตกต่างจะทดสอบความแตกต่างเป็นรายคู่ โดยใช้วิธี Least Significant Difference (LSD) และใช้สถิติวิเคราะห์การถดถอย (Regression) เพื่อทดสอบ สมมติฐาน ผลการศึกษาพบว่า ผู้ตอบแบบสอบถามส่วนใหญ่เป็นเพศชาย มีอายุระหว่าง 21-30 ปี ระดับการศึกษาปริญญาตรี ประสบการณ์ในตำแหน่งหน้าที่ ที่ ได้รับอยู่ระหว่าง 5-10 ปีการยอมรับ ใน ภาพรวมทั้ง 3 ชั้น จัดอยู่ในระดับมาก เพศ อายุ ระดับการศึกษา ประสบการณ์ในตำแหน่งหน้าที่

ที่ได้รับมีผลต่อการรับรู้ประโยชน์กับการรับรู้ความง่ายต่อการใช้งาน การรับรู้ความง่ายต่อการใช้งานมี ผลต่อการรับรู้ประโยชน์ และการรับรู้ประโยชน์กับการรับรู้ความง่ายต่อการใช้งานมีผลต่อการยอมรับระบบ

พัชรินทร์ พุ่มลำเจียก (2013) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาระดับของปัจจัย ความสัมพันธ์ และอิทธิพลของความไว้วางใจของเว็บไซต์ การรับรู้ว่าใช้งานง่าย การรับรู้ถึง ประโยชน์ ทักษะคติ และพฤติกรรมของผู้ใช้หนังสืออิเล็กทรอนิกส์ในกรุงเทพมหานครที่ ผู้วิจัยพัฒนา ขึ้นกับข้อมูลเชิงประจักษ์ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือ ผู้ที่เคยดาวน์โหลด หนังสืออิเล็กทรอนิกส์แบบเสียเงินผ่านทางเว็บไซต์ อาศัยอยู่ในกรุงเทพมหานคร มีอายุระหว่าง 16-35 ปี จำนวน 330 คน โดยใช้แบบสอบถามที่ผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหา ค่าดัชนี ความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์อยู่ระหว่าง 0.75 ถึง 1 ค่าความเชื่อมั่นทั้งหมด เท่ากับ 0.93 การวิเคราะห์ข้อมูลใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน และการตรวจสอบความ สอดคล้องของแบบจำลองพฤติกรรมของผู้ใช้หนังสือ อิเล็กทรอนิกส์ ที่ผู้วิจัยสร้างขึ้นกับข้อมูลเชิง ประจักษ์ด้วยการวิเคราะห์สมการ โครงสร้าง ผลการวิจัยพบว่า ระดับความคิดเห็นของปัจจัยความ น่าเชื่อถือของเว็บไซต์ การรับรู้ว่าใช้งานง่าย การรับรู้ถึงประโยชน์ ทักษะคติที่มีต่อการ ใช้ และ พฤติกรรมของผู้ใช้เกือบทั้งหมดอยู่ในระดับมากที่สุดความสัมพันธ์ของ ปัจจัย พบว่า ความสัมพันธ์ ของตัวแปรทั้งหมดมีค่าอยู่ระหว่าง 0.011 ถึง 0.629 ส่วนแบบจำลองพฤติกรรมของผู้ใช้หนังสือ อิเล็กทรอนิกส์มีความสอดคล้องกับ ข้อมูลเชิงประจักษ์ โดยค่าไคสแควร์ (X2) เท่ากับ 141.960 ค่า X2/df เท่ากับ 1.164 ค่าองศาอิสระ (df) เท่ากับ 122 ค่า P-Value เท่ากับ 0.105 ค่า RMSEA เท่ากับ 0.022 และดัชนีวัดระดับความกลมกลืน (GFI) มีค่าเท่ากับ 0.958 พฤติกรรมของผู้ใช้ได้รับอิทธิพล ทางตรงจากทักษะคติที่มีต่อการใช้งาน โดยมีค่า สัมประสิทธิ์อิทธิพลเท่ากับ 0.94 อย่างมีนัยสำคัญทาง สถิติที่ระดับ 0.001 และได้รับอิทธิพลทางอ้อมมากที่สุดจากปัจจัยการรับรู้ที่ใช้งานง่าย รองลงมาคือ ปัจจัยการรับรู้ถึงประโยชน์และปัจจัยความน่าเชื่อถือของเว็บไซต์ โดยมีค่าสัมประสิทธิ์อิทธิพล เท่ากับ 0.65, 0.22 และ 0.20 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ตัวแปรในแบบจำลอง สามารถอธิบายความแปรปรวนของแบบจำลองได้ร้อยละ 89

Lu, Chang and Chang (2014) ได้ มี การศึกษาวิจัยในเรื่อง “ Consumer attitudes toward blogger’s sponsored recommendations and purchase intention: The effect of sponsorship type, product type, and brand awareness” โดยใช้ตัวแปรด้าน 1) รูปแบบการสนับสนุน 2) ชนิดของสินค้า 3) ความเป็นที่รู้จักของตราสินค้า ผลลัพธ์แสดงให้เห็นว่า หากสินค้าเป็นสินค้าที่ง่ายต่อการพิจารณา ก่อนการซื้อ (Search Good) และมี Brand Awareness ที่สูง ผู้บริโภคมีแนวโน้มจะไปทิศทางที่ดีกับ

Sponsored Recommendation Posts โดยจะเพิ่ม ความตั้งใจในสินค้าตัวนั้นๆ การได้รับการสนับสนุนไม่ว่าจะทางตรงหรือทางอ้อมกับเจ้าของ Blog ไม่ได้ส่งผลอย่างมีนัยสำคัญกับทัศนคติของผู้อ่าน การมีบทความให้อ่านกลับช่วยให้มีความน่าเชื่อถือ และความเชื่อมั่นให้กับผู้บริโภค เนื่องด้วยกฎหมายที่บังคับให้มีการเปิดเผยหากได้รับการสนับสนุน จากผู้ประกอบการทำให้ผู้บริโภครู้สึกว่ายังมีการเปิดเผยว่ามีตราสินค้าสนับสนุนทำให้ Sponsored Recommendation Post มีความซื่อสัตย์ บริสุทธิ์ใจ และยังคงเป็นเครื่องมือทางการตลาดที่สำคัญ

Ballantine and Au Yeung (2014) ได้มีการศึกษาวิจัยในเรื่อง “The effects of review valence in organic versus sponsored blog sites on perceived credibility, brand attitude, and behavioral intentions” โดยใช้การวิจัยแบบ Quantitative ผ่านวิธีการทดลองใน ช่องทางออนไลน์ (Online Experiment) จากตัวแปร 1) แหล่งที่มาของบทความ 2) ผลลัพธ์ของรีวิว (Review Valance) ผลลัพธ์แสดงให้เห็นว่าบทความรีวิวทั้ง CR และ SR นั้นต่างก็มีความน่าเชื่อถือ โดยหากพบความคิดเห็นในด้านลบก็ส่งผลต่อทัศนคติและการซื้อสินค้าในด้านลบ และให้ผลกลับกัน กับความคิดเห็นในด้านบวก แต่จากการวิจัยกลับพบว่าความคิดเห็นที่ดูเป็นกลางนั้นส่งผลให้บทความ ดูน่าเชื่อถือและมีความเชี่ยวชาญมากกว่า

Pongsatorn Tantrabundit (2015) ได้มีการศึกษาวิจัย “How online review and rating systems affect online consumer buying behavior in the hotel industry” โดย ใช้ส่วนประกอบของ eWOM เป็นตัวแปรสำคัญ ประกอบไปด้วย 1) ผลลัพธ์ของรีวิว (Review 18 Valence) 2) เนื้อหาที่เป็นตัวอักษร 3) จำนวนของรีวิว 4) ความน่าเชื่อถือของแหล่งที่มา 5) ระดับ ประโยชน์ของรีวิว 6) การตอบสนองของตราสินค้า ซึ่งผลลัพธ์ออกมาว่าในทุกส่วนประกอบของ eWOM ต่างส่งผลเป็นไปในทิศทางเดียวกันกับพฤติกรรมของผู้บริโภค และยังเสริมว่านอกจาก ส่วนประกอบเหล่านั้น เนื้อหาของบทความที่ระบุเหตุผลก็ส่งผลมากกว่าบทความที่เขียนด้านอารมณ์

Heikkinen (2012) ได้มีการศึกษาวิจัย “Electronic word-of-mouth: an experimental analysis of online consumer reviews of smartphones” จากการศึกษาด้วย วิธี Quantitative Research พบว่าด้านความน่าเชื่อถือของข้อมูลมีปัจจัยที่ส่งผลให้ข้อมูลมีความน่าเชื่อถือมากที่สุด ประกอบไปด้วยความมีน้ำหนักของข้อมูลและความเกี่ยวข้องของข้อมูลต่อสินค้า

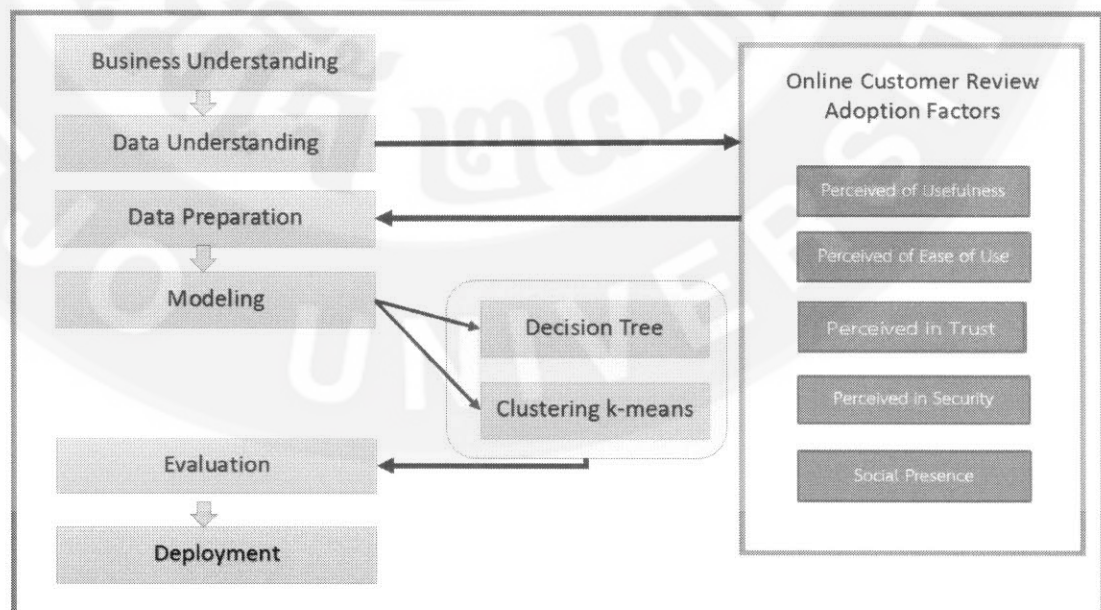
Zhang et al. (2010) ได้มีการศึกษาวิจัย “The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews” โดย ได้มีการศึกษาโดยใช้วิธีการ Quantitative Research และการเก็บข้อมูล Secondary Data ซึ่งได้ผลลัพธ์

ออกมาว่า การเขียนรีวิวของผู้บริโภค (consumer reviews) ซึ่งมีการ การบรรยายถึงรสชาติอาหาร บรรยากาศ บริการ และ จำนวนของรีวิวต่างส่งผลต่อความนิยมใน ร้านอาหาร ในขณะที่เดียวกันความ คิดเห็นจากผู้เชี่ยวชาญ (Editor's Choice) กลับมีความสัมพันธ์ด้าน ลบกับการตัดสินใจใช้บริการ ของผู้บริโภค

กรอบแนวคิดการวิจัย

โดยหลักการแล้วการพัฒนาสารสนเทศเป็นการแสดงข้อมูลจากข้อมูลที่ถูกรวบรวม ไว้ในรูปแบบที่เป็นไปตามหลักของการออกแบบฐานข้อมูล การเก็บข้อมูลในการวิจัยถูกออกแบบ ตามลักษณะความต้องการในเชิงวิเคราะห์พิจารณาจากปัจจัยแวดล้อมต่างๆที่ต้องการ เมื่อทำการ รวบรวมข้อมูลต่างๆไม่ว่าจะด้วยวิธีการออนไลน์หรือออฟไลน์ก็จะมีกระบวนการบันทึกข้อมูลบนระบบงาน คอมพิวเตอร์ การวิเคราะห์ข้อมูลนั้นมีหลายระบบจากข้อมูลที่ได้สามารถวิเคราะห์หา ความเชื่อมั่น และความพึงพอใจต่อการใช้ระบบ review online ในการให้ข้อมูลสำหรับการเลือกที่พักเชิงนิเวศ โดยวิธีการทางสถิติ และสามารถผลิตสารสนเทศในเชิงสรุปในด้านข้อมูลต่างๆได้ แต่การค้นหา ความสัมพันธ์ระหว่างปัจจัยและการจำแนกกลุ่มของลูกค้ายังมีเทคนิคที่เข้ามาค้นหาข้อมูลดังกล่าว ได้โดยใช้เทคนิคเหมืองข้อมูลซึ่งมีหลายเทคนิคที่นิยมใช้ โดยขึ้นอยู่กับความเหมาะสมของข้อมูลว่า ต้องการคำตอบใดและข้อมูลถูกจัดเก็บมาแบบไหน

ดังนั้นจากที่มาความสำคัญและปัญหารวมถึงทฤษฎีข้างต้นนั้นสามารถสรุปได้เป็นกรอบ แนวความคิดของโครงการวิจัยดังรูปที่ 2.1



ภาพที่ 2.1 กรอบแนวคิดการวิจัย (Research Conceptual Framework)

บทที่ 3

วิธีวิจัย

ประชากร

ประชากรในการศึกษาที่ใช้ในการศึกษาวิจัยในครั้งนี้ คือ กลุ่มลูกค้าของที่พักรังเชิงนิเวศที่เคยมีประสบการณ์ในการจองที่พักแบบออนไลน์และเคยใช้งานระบบการรีวิวออนไลน์ประกอบการตัดสินใจจองมาก่อน โดยต้องเป็นผู้ใช้งานที่มีอายุตั้งแต่ 18 ปีขึ้นไป และพักอาศัยในประเทศไทย โดยแบ่งพื้นที่ตามภูมิภาคต่างๆ ในประเทศไทย

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ กลุ่มลูกค้าของที่พักรังเชิงนิเวศที่เคยมีประสบการณ์ในการจองที่พักแบบออนไลน์และเคยใช้งานระบบการรีวิวออนไลน์ประกอบการตัดสินใจจองมาก่อนในประเทศไทยที่ได้กล่าวถึงข้างต้น โดยได้ทำการคำนวณขนาดของกลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ใช้สูตรการคำนวณกลุ่มตัวอย่างของทาโร ยามาเน่ (1973) ที่ระดับความเชื่อมั่นร้อยละ 95 ดังนี้

$$\text{สูตร} \quad n = \frac{N}{1 + Ne^2}$$

โดยกำหนดให้ N = จำนวนประชากรทั้งหมด

n = ขนาดกลุ่มตัวอย่าง

e = ระดับที่ยอมรับให้เกิดความผิดพลาด

กลุ่มตัวอย่างในการศึกษาที่ใช้ในการศึกษาวิจัยในครั้งนี้ คือ กลุ่มลูกค้าของที่พักรังเชิงนิเวศที่เคยมีประสบการณ์ในการจองที่พักแบบออนไลน์และเคยใช้งานระบบการรีวิวออนไลน์ประกอบการตัดสินใจที่อาศัยในประเทศไทย โดยแบ่งตามภูมิภาคต่างๆ จำนวนทั้งสิ้น 350 ราย

การศึกษาเรื่อง ปัจจัยของการรีวิวออนไลน์ที่ส่งผลต่อความไว้วางใจในการตัดสินใจในการซื้อบริการที่พักแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย ผู้วิจัยใช้วิธีการสุ่มตัวอย่างวิธีไม่ใช้ความน่าจะเป็น (Non Probability Sampling) แบบเจาะจง (Purposive Sampling)

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้แบ่งออกเป็น 2 ประเภท คือ

1. เครื่องมือที่ใช้ในการเก็บข้อมูล คือ การใช้แบบสอบถาม
2. เครื่องมือที่ใช้ในการวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล
 - 2.1 โปรแกรม Microsoft Excel 2014 ใช้ในการจัดเก็บข้อมูลที่ได้จากแบบสอบถาม
 - 2.2 RapidMiner Studio ใช้ในการวิเคราะห์และแสดงผลข้อมูลด้วยเทคนิคเหมืองข้อมูล คือ ต้นไม้ตัดสินใจและการแบ่งกลุ่มด้วย k-mean

เครื่องมือที่ใช้ในการเก็บข้อมูล

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลของการศึกษาครั้งนี้ คือ การใช้แบบสอบถาม ซึ่งแบบสอบถามที่ใช้ในการเก็บรวบรวมข้อมูลครั้งนี้ ประกอบด้วยคำถามประเภทต่างๆ โดยแบ่งคำถามออกได้เป็น 4 ส่วน ดังนี้

ส่วนที่ 1 เป็นส่วนที่ประกอบด้วยข้อคำถามแบบตรวจสอบรายการ (Check list) ซึ่งเป็นคำถามเกี่ยวกับปัจจัยส่วนบุคคลของผู้ใช้งานระบบการรีวิวออนไลน์ในการตัดสินใจของที่พักเชิงนิเวศ

ส่วนที่ 2 เป็นส่วนที่ประกอบด้วยข้อคำถามแบบตรวจสอบรายการ (Check list) ซึ่งเป็นคำถามเกี่ยวกับปัจจัยเกี่ยวกับการใช้การใช้นระบบการรีวิวออนไลน์ในการตัดสินใจของที่พักเชิงนิเวศ

ส่วนที่ 3 เป็นส่วนที่ประกอบด้วยข้อคำถามแบบมาตราส่วนประมาณค่า (Rating scale) ชนิด 5 ตัวเลือก ผู้ศึกษาได้กำหนดเกณฑ์การให้คะแนน คือ

ตอบเห็นด้วยมากที่สุด	5	คะแนน
ตอบเห็นด้วยมาก	4	คะแนน
ตอบเห็นด้วยปานกลาง	3	คะแนน
ตอบเห็นด้วยน้อย	2	คะแนน
ตอบเห็นด้วยน้อยที่สุด	1	คะแนน

และได้กำหนดระดับการแปลผลระดับปัจจัยที่มีผลต่อการตัดสินใจในการใช้บริการระบบการรีวิวออนไลน์ในการตัดสินใจของที่พักเชิงนิเวศ โดยใช้สูตรการคำนวณ ดังนี้

$$\text{อันดับภาคชั้น} = \frac{\text{ค่าพิสัย}}{\text{จำนวนชั้น}}$$

$$\text{ค่าพิสัย} = \frac{\text{คะแนนสูงสุด} - \text{คะแนนต่ำสุด}}{\text{จำนวนชั้น}}$$

จากผลการคำนวณดังกล่าว ได้นำมาเป็นเกณฑ์วัดค่าเฉลี่ยระดับปัจจัยของการรีวิวออนไลน์ที่ส่งผลกระทบต่อความไว้วางใจในการตัดสินใจในการซื้อบริการที่פקแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย ดังต่อไปนี้

$$\frac{\text{คะแนนสูงสุด} - \text{คะแนนต่ำสุด}}{\text{จำนวนชั้น}} = \frac{5 - 1}{5} = 0.80$$

โดยนำคะแนนที่ได้มาวิเคราะห์ค่าเฉลี่ย การแปลความหมายของระดับคะแนนเฉลี่ย โดยยึดเกณฑ์ตามค่าที่ได้ จากสูตรคำนวณของระดับชั้น = 0.80 ดังนี้

คะแนนเฉลี่ย	เกณฑ์การแปลผล
ระหว่าง 1.00 – 1.80	มีความเห็นด้วยอยู่ในระดับน้อยที่สุด
ระหว่าง 1.81 – 2.60	มีความเห็นด้วยอยู่ในระดับน้อย
ระหว่าง 2.61 – 3.40	มีความเห็นด้วยอยู่ในระดับปานกลาง
ระหว่าง 3.41 – 4.20	มีความเห็นด้วยอยู่ในระดับมาก
ระหว่าง 4.21 – 5.00	มีความเห็นด้วยอยู่ในระดับมากที่สุด

ตารางที่ 3.1 คะแนนเฉลี่ยและเกณฑ์การแปลความหมาย ระดับปัจจัยของการรีวิวออนไลน์ที่ส่งผลกระทบต่อความไว้วางใจในการตัดสินใจในการซื้อบริการที่פקแบบการท่องเที่ยวเชิงนิเวศในประเทศไทย

แบบสอบถามใช้วัดระดับความคิดเห็นของผู้ใช้งานระบบการจองที่פקเชิงนิเวศในประเทศไทยที่ส่งผลต่อการใช้งานระบบการรีวิวออนไลน์ โดยแบบสอบถามจะถูกแบ่งเป็น 7 ประเด็น ดังนี้

1. ความคิดเห็นเกี่ยวกับการรับรู้ถึงประโยชน์ที่ได้รับจากการใช้งานรีวิวออนไลน์ (Perceived usefulness in Online Customer Review)
2. ความคิดเห็นเกี่ยวกับการรับรู้ว่าการใช้งานรีวิวออนไลน์เป็นเรื่องที่ง่ายต่อการใช้งาน (Perceived ease of Use in Online Customer Review หรือ PEOU)
3. ความคิดเห็นเกี่ยวกับความไว้วางใจในการใช้งานรีวิวออนไลน์ (Perceived Trust in Online Customer Review)
4. ความคิดเห็นเกี่ยวกับความปลอดภัยต่อการใช้งานรีวิวออนไลน์ (Perceived Security in Online Customer Review)

- 5 ความคิดเห็นเกี่ยวกับสถานะทางสังคมของผู้ใช้บริการในการใช้งานรีวิวออนไลน์ (Social Presence in Online Customer Review)
- 6 ความคิดเห็นเกี่ยวกับความเต็มใจที่จะใช้งานระบบการรีวิวออนไลน์ (Intention to Purchase in Online Customer Review)
- 7 ความคิดเห็นเกี่ยวกับความพึงพอใจในการใช้งานระบบการรีวิวออนไลน์ (Satisfaction to Purchase in Online Customer Review)

ส่วนที่ 4 เป็นส่วนที่ประกอบด้วยข้อคำถามแบบปลายเปิดที่เกี่ยวกับการแสดงความคิดเห็นและข้อเสนอแนะ เพื่อปรับปรุงเกี่ยวกับ การใช้งานระบบการรีวิวออนไลน์ในการตัดสินใจจองที่พักเชิงนิเวศ

วิธีการเก็บรวบรวมข้อมูล

ในการวิจัยนี้ ผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลจาก 2 แหล่งข้อมูล ดังนี้

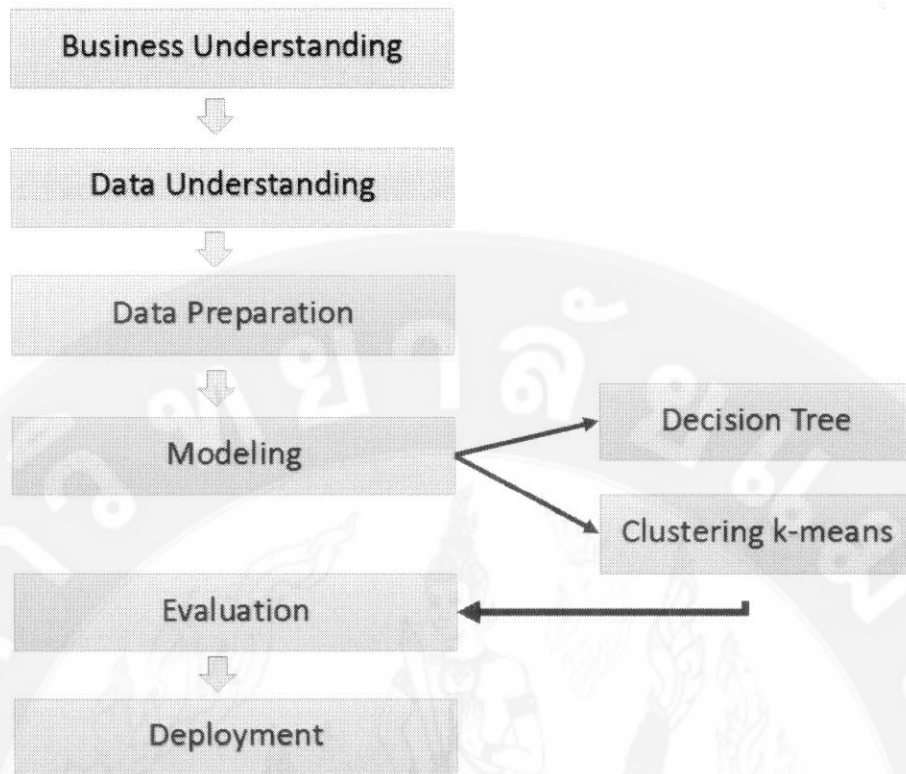
- 1) การเก็บรวบรวมข้อมูลจากเอกสาร (Documentary Research) บทความวิจัย บทความวิชาการ งานวิจัย วิทยานิพนธ์ หนังสือ ข้อมูล สถิติตัวเลข และใช้การสืบค้นจากฐานข้อมูลทางวิชาการออนไลน์ (Academic Online Database) เพื่อนำมาตอบคำถามในการศึกษาให้เป็นไปตามวัตถุประสงค์ที่ได้ตั้งไว้ และเพื่อนำมาสร้างกรอบแนวคิดที่ใช้ในการศึกษา
- 2) การเก็บรวบรวมข้อมูลภาคสนาม (Field Survey) โดยใช้แบบสอบถามเก็บรวบรวมข้อมูลจากผู้ใช้งานการจองที่พักแบบออนไลน์และเคยใช้งานระบบการรีวิวออนไลน์ประกอบการตัดสินใจในการจองมาก่อน โดยต้องเป็นผู้ใช้งานที่มีอายุตั้งแต่ 18 ปีขึ้นไป และพักอาศัยในประเทศไทย จำนวน 350 ราย โดยมีขั้นตอนการดำเนินงานดังนี้
 - จัดทำต้นแบบของแบบสอบถามเพื่อวัดความน่าเชื่อถือของแบบสอบถาม จำนวน 20 ชุด
 - แจกแบบสอบถามให้กับกลุ่มตัวอย่าง โดยกลุ่มตัวอย่างที่กรอกแบบสอบถามที่เป็นต้นแบบ จะไม่ซ้ำกับกลุ่มตัวอย่างที่กรอกแบบสอบถามฉบับจริง
 - รวบรวมข้อมูลจากแบบสอบถามเพื่อนำมาทดสอบความน่าเชื่อถือ โดยทดสอบทางสถิติ
 - ปรับปรุงและพัฒนาแบบสอบถามเพื่อนำมาใช้ในการเก็บรวบรวมข้อมูล

จัดทำแบบสอบถามผ่านทางระบบออนไลน์ โดยการใช้ Google Drive
- แจกแบบสอบถามไปยังกลุ่มเป้าหมายโดยผ่านการประชาสัมพันธ์ทางอีเมล
เฟสบุ๊ก

การวิเคราะห์ข้อมูล

ขั้นตอนในการวิเคราะห์ข้อมูลมีลำดับขั้นตอนและการใช้เครื่องมือในการวิเคราะห์ดังนี้

- 1) Business Understanding ขั้นตอนการทำความเข้าใจกับปัญหา หรือโอกาสทางธุรกิจ ระบุ output เป้าหมายที่ต้องการ ได้จากการวิเคราะห์ข้อมูลด้วย data mining
- 2) Data Understanding ขั้นตอนนี้ยังเป็นการ รวบรวมข้อมูลที่เกี่ยวข้อง ข้อมูลต้องน่าเชื่อถือ ข้อมูลที่ได้มีปริมาณที่มากพอ ข้อมูลที่ได้มีความเหมาะสมเพียงพอต่อการวิเคราะห์
- 3) Data Preparation ขั้นตอนการเตรียมข้อมูลเพื่อประมวลผล เป็นขั้นตอนที่ยาวนานที่สุด เนื่องจากโมเดลที่ได้จากการทำ data mining จะได้ผลลัพธ์ที่ถูกต้องหรือไม่ขึ้นอยู่กับคุณภาพข้อมูลโดยต้องทำดังนี้
 - Data Selection
 - Data Cleaning
 - Data Transformation
- 4) Modeling ขั้นตอนการวิเคราะห์ด้วยเทคนิค data mining ซึ่งมีสามกลุ่มคือ
 - Classification การสร้างโมเดลเพื่อทำนายข้อมูลจากข้อมูลในอดีต ซึ่งจากการวิจัยเป็นการใช้ข้อมูลจากการสำรวจเพื่อเป็น training data
 - Clustering การแบ่งกลุ่มตามความคล้ายคลึงกันของข้อมูล ในการวิจัยใช้ k-mean ในการหาความคล้ายกันของข้อมูล
 - Associate rules การหาความสัมพันธ์ของข้อมูลที่เกิดร่วมกัน ซึ่งข้อมูลที่จัดเก็บจากการวิจัยยังไม่ได้นำมาวิเคราะห์ในรูปแบบนี้
- 5) Evaluation การประเมินและการวัดผล โมเดลต่างๆที่เราใช้ในการวิเคราะห์ข้อมูล
- 6) Deployment การนำผลจากโมเดลไปใช้งานจริง



ภาพที่ 3.1 -ขั้นตอนการวิเคราะห์ข้อมูล

บทที่ 4

ผลการวิจัย

รวบรวมข้อมูลและการจัดทำสารสนเทศ

ในขั้นตอนนี้จะทำการโหลดข้อมูลจาก excel เข้ามาไว้ในโปรแกรม rapid mining สามารถผลิตสารสนเทศได้และยังสามารถมีเครื่องมือที่ใช้วิเคราะห์ด้วยเทคนิคเหมืองข้อมูล โดยการนำเสนอผลการวิจัยในส่วนที่ 1 และ 2 จะนำเสนอข้อมูลสารสนเทศ และส่วนที่ 3 จะแสดงผลข้อมูลจากการทำค้นไม้ตัดสินใจและการจำแนกกลุ่มโดย k-mean

ส่วนที่ 1 การอธิบายลักษณะทั่วไปของกลุ่มตัวอย่างโดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ

ส่วนที่ 2 การใช้งานระบบการรีวิวออนไลน์โดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ

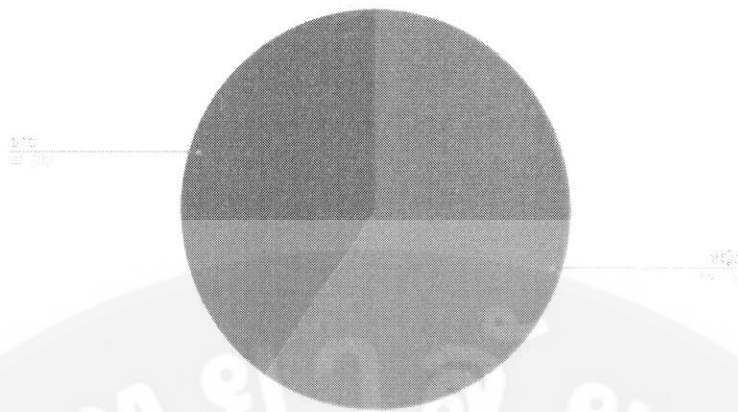
ส่วนที่ 3 จะแสดงผลข้อมูลจากการทำค้นไม้ตัดสินใจและการจำแนกกลุ่มโดย k-mean

ส่วนที่ 1 การอธิบายลักษณะทั่วไปของกลุ่มตัวอย่างโดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ

ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

ตารางที่ 4.1 เพศของผู้ตอบแบบสอบถาม

เพศ	ความถี่	ร้อยละ
ชาย	141	40.29
หญิง	209	59.71
รวม	350	100.00

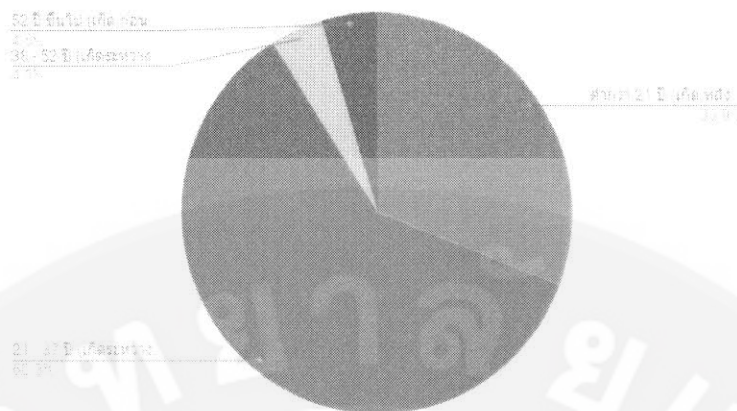


ภาพที่ 4.1 เพศของผู้ตอบแบบสอบถาม

จากตารางที่ 4.1 และภาพที่ 4.1 พบว่า เพศของผู้ตอบแบบสอบถามมากที่สุด ได้แก่ เพศหญิง คิดเป็นร้อยละ 59.71 ที่เหลือเป็นเพศชาย คิดเป็น ร้อยละ 40.29

ตารางที่ 4.2 ช่วงอายุของผู้ใช้งานบริการระบบรีวิวออนไลน์

ประเภทผู้ใช้บริการ	ความถี่	ร้อยละ
21 - 37 ปี (เกิดระหว่าง พ.ศ. 2524 - 2541)	211	60.3
38 - 52 ปี (เกิดระหว่าง พ.ศ. 2509 - 2523)	15	4.3
52 ปี ขึ้นไป (เกิด ก่อน พ.ศ. 2509)	16	4.6
ต่ำกว่า 21 ปี (เกิด หลัง พ.ศ. 2541 เป็นต้นไป)	108	30.9
รวม	350	100.00

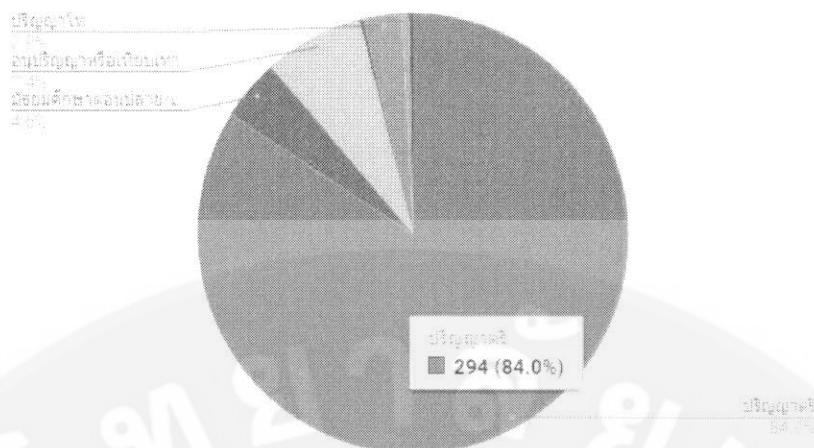


ภาพที่ 4.2 ช่วงอายุของผู้ใช้งานบริการระบบปริวิวออนไลน์

จากตารางที่ 4.2 และภาพที่ 4.2 พบว่า ช่วงอายุของผู้ใช้งานบริการระบบปริวิวออนไลน์ จำนวนกว่าร้อยละ 60.3 เป็นผู้ใช้งานที่มีช่วงอายุระหว่าง 21 - 37 ปี รองลงมาเป็นกลุ่มช่วงอายุต่ำกว่า 21 ปี จำนวนร้อยละ 30.9 ส่วนผู้ใช้งานที่มีช่วงอายุตั้งแต่ 52 ปีขึ้นไป และช่วงอายุระหว่าง 38 - 52 ปี มีจำนวนร้อยละ 4.6 และ 4.3 ตามลำดับ

ตารางที่ 4.3 ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถาม

ระดับการศึกษา	ความถี่	ร้อยละ
ประถมศึกษา	1	0.3
ปริญญาตรี	294	84.0
ปริญญาโท	10	2.9
ปริญญาเอก	2	0.6
มัธยมศึกษาตอนต้น	1	0.3
มัธยมศึกษาตอนปลาย/เทียบเท่า	16	4.6
อนุปริญญาหรือเทียบเท่า	26	7.4
รวม	350	100.00

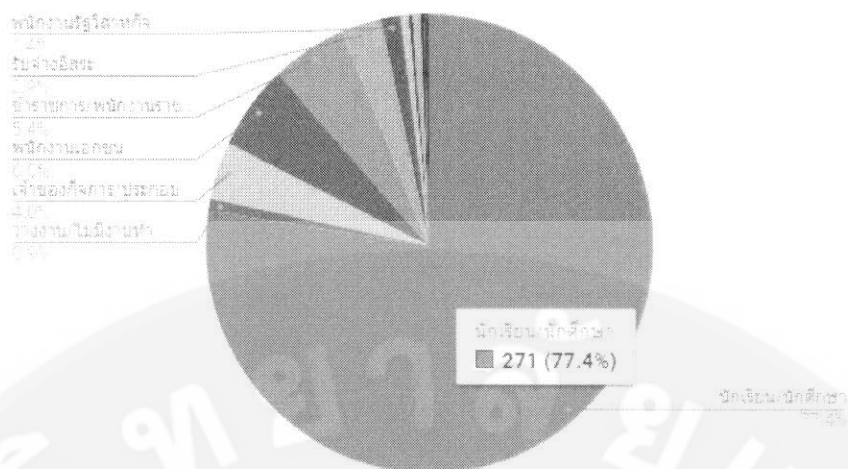


ภาพที่ 4.3 ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถาม

จากตารางที่ 4.3 และภาพที่ 4.3 พบว่า ระดับการศึกษาสูงสุดของผู้ตอบแบบสอบถามมากที่สุด ได้แก่ ระดับปริญญาตรี คิดเป็น ร้อยละ 84 รองลงมา ได้แก่ ระดับอนุปริญญา/เทียบเท่าคิดเป็นร้อยละ 7.4 ตามลำดับ

ตารางที่ 4.4 อาชีพของผู้ตอบแบบสอบถาม

อาชีพ	ความถี่	ร้อยละ
เกษียณ/ข้าราชการบำนาญ	4	1.1
ข้าราชการ/พนักงานราชการ/พนักงานมหาวิทยาลัย	19	5.4
เจ้าของกิจการ/ประกอบธุรกิจส่วนตัว/ค้าขาย	14	4.0
นักเรียน/นักศึกษา	271	77.4
พนักงานรัฐวิสาหกิจ	5	1.4
พนักงานเอกชน	21	6.0
แม่บ้าน/พ่อบ้าน	2	0.6
รับจ้างอิสระ	11	3.1
ว่างงาน/ไม่มีงานทำ	3	0.9
รวม	350	100.00

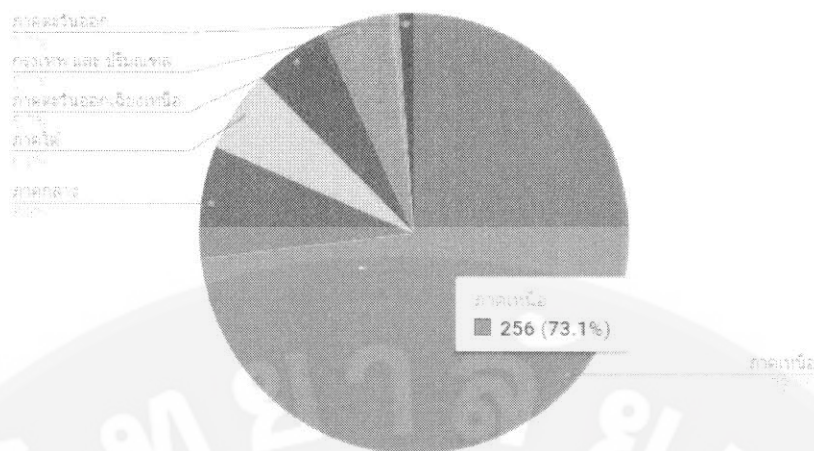


ภาพที่ 4.4 อาชีพของผู้ตอบแบบสอบถาม

จากตารางที่ 4.4 และภาพที่ 4.4 พบว่าอาชีพของผู้ตอบแบบสอบถามมากที่สุด ได้แก่ กลุ่มนักเรียน/นักศึกษา คิดเป็นร้อยละ 77.4 รองลงมา ได้แก่ กลุ่มพนักงานบริษัทเอกชน คิดเป็นร้อยละ 6.0 และกลุ่มข้าราชการ/พนักงานราชการ/พนักงานมหาวิทยาลัย คิดเป็นร้อยละ 5.4 ตามลำดับ

ตารางที่ 4.5 ภูมิภาคของผู้ตอบแบบสอบถาม

ภูมิภาค	ความถี่	ร้อยละ
กรุงเทพ และ ปริมณฑล	18	5.1
ภาคกลาง	29	8.3
ภาคตะวันตก	2	0.6
ภาคตะวันออก	4	1.1
ภาคตะวันออกเฉียงเหนือ	20	5.7
ภาคใต้	21	6.0
ภาคเหนือ	256	73.1
รวม	350	100.00

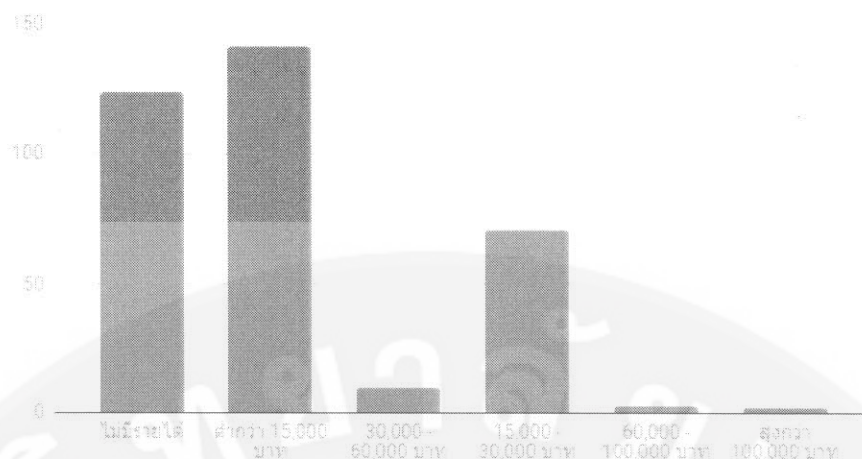


ภาพที่ 4.5 ภูมิภาคของผู้อตอบแบบสอบถาม

จากตารางที่ 4.5 และภาพที่ 4.5 พบว่าภูมิภาคของผู้อตอบแบบสอบถามมากที่สุดคือ ภาคเหนือ คิดเป็นร้อยละ 73.1 รองลงมา ได้แก่ ภาคกลาง คิดเป็นร้อยละ 8.3 และภาคใต้ คิดเป็นร้อยละ 6.0 ตามลำดับ

ตารางที่ 4.6 รายได้เฉลี่ยต่อเดือนของผู้อตอบแบบสอบถาม

รายได้เฉลี่ยต่อเดือน	ความถี่	ร้อยละ
15,000 – 30,000 บาท	71	20.3
30,000 – 60,000 บาท	10	2.9
60,000 – 100,000 บาท	3	0.9
ต่ำกว่า 15,000 บาท	141	40.3
ไม่มีรายได้	123	35.1
สูงกว่า 100,000 บาท	2	0.6
รวม	350	100.00

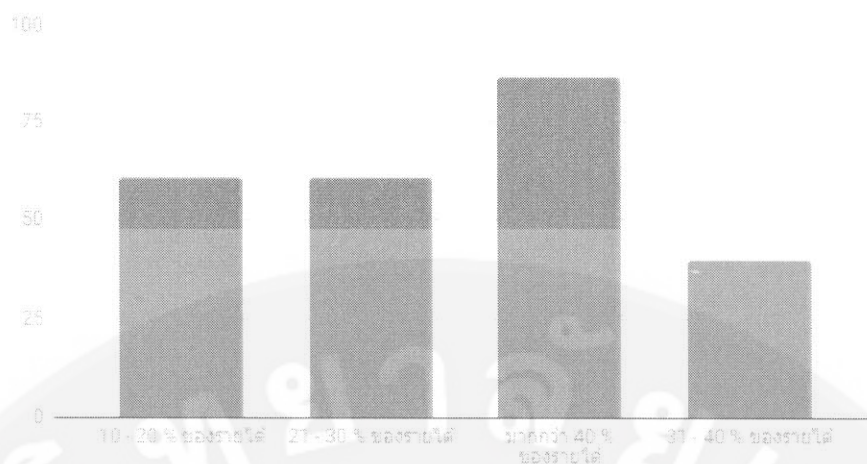


ภาพที่ 4.6 รายได้เฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม

จากตารางที่ 4.6 และภาพที่ 4.6 พบว่ารายได้เฉลี่ยต่อเดือนของผู้ตอบแบบสอบถามมากที่สุด ได้แก่ กลุ่มที่มีรายได้ต่ำกว่า 15,000 บาทต่อเดือน คิดเป็นร้อยละ 40.3 รองลงมา ได้แก่ กลุ่มที่ไม่มีรายได้ คิดเป็นร้อยละ 35.1 และ กลุ่มมีรายได้ระหว่าง 15,000 – 30,000 บาทต่อเดือน คิดเป็นกว่าร้อยละ 20.3 ตามลำดับ

ตารางที่ 4.7 รายจ่ายเฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม

รายจ่ายเฉลี่ยต่อเดือน	ความถี่	ร้อยละ
ร้อยละ 10 – 20 ของรายได้	89	25.4
ร้อยละ 21 – 30 ของรายได้	87	24.9
ร้อยละ 31 – 40 ของรายได้	55	15.7
มากกว่าร้อยละ 40 ของรายได้	119	34.0
รวม	350	100.00



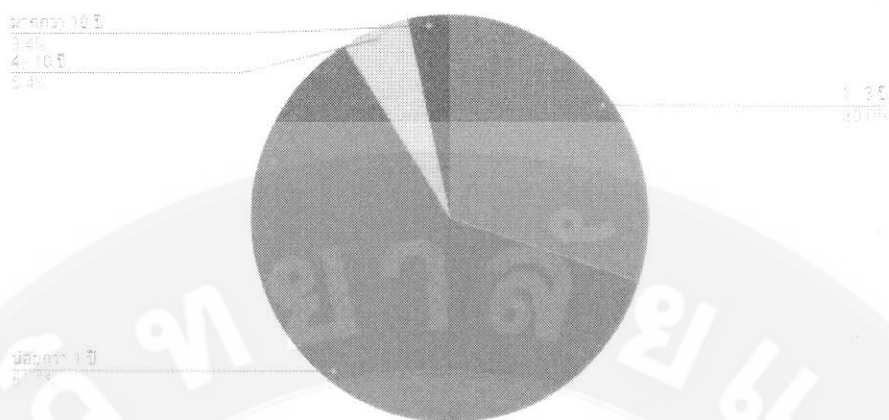
ภาพที่ 4.7 รายจ่ายเฉลี่ยต่อเดือนของผู้ตอบแบบสอบถาม

จากตารางที่ 4.7 และภาพที่ 4.7 พบว่ารายจ่ายเฉลี่ยต่อเดือนของผู้ตอบแบบสอบถามมากที่สุด ได้แก่ รายจ่ายเฉลี่ยมากกว่าร้อยละ 40 ของรายได้ต่อเดือน คิดเป็นร้อยละ 34.0 รองลงมาคือกลุ่มที่มีรายจ่ายเฉลี่ยร้อยละ 10-20 ของรายได้ต่อเดือน คิดเป็นร้อยละ 25.4 และกลุ่มที่มีรายจ่ายเฉลี่ยร้อยละ 21-30 ของรายได้ต่อเดือน คิดเป็นร้อยละ 24.9 ตามลำดับ

ตารางที่ 4.8 ประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์

ประสบการณ์ในการจองห้องพักผ่านระบบออนไลน์	ความถี่	ร้อยละ
1 - 3 ปี	105	30.0
4 - 10 ปี	19	5.4
น้อยกว่า 1 ปี	214	61.1
มากกว่า 10 ปี	12	3.4
รวม	350	100.00

ประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์ (ระบุเป็นปี)



ภาพที่ 4.8 ประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์

จากตารางที่ 4.8 และภาพที่ 4.8 พบว่าประสบการณ์ในการจองห้องพักในการท่องเที่ยวผ่านระบบออนไลน์ของผู้ตอบแบบสอบถามมากที่สุด ส่วนมากแล้วมีประสบการณ์น้อยกว่า 1 ปี คิดเป็นร้อยละ 61.1 รองลงมา มีประสบการณ์ 1 - 3 ปี คิดเป็นร้อยละ 30 ตามลำดับ

ส่วนที่ 2 การใช้งานระบบการรีวิวออนไลน์โดยการนำเสนอเป็นตารางค่าความถี่ ค่าร้อยละ และกราฟ

การใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ

1. การใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ ผ่านอุปกรณ์ของผู้ตอบแบบสอบถาม

ตารางที่ 4.9 อุปกรณ์ของการใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ

อุปกรณ์ของการใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยวเชิงนิเวศ	ความถี่	ร้อยละ
คอมพิวเตอร์แบบพกพาและคอมพิวเตอร์ส่วนบุคคล (Laptop and PC)	35	10.0
โทรศัพท์และคอมพิวเตอร์มือถือ (Smart Phone and Tablet)	315	90.0
รวม	350	100.00

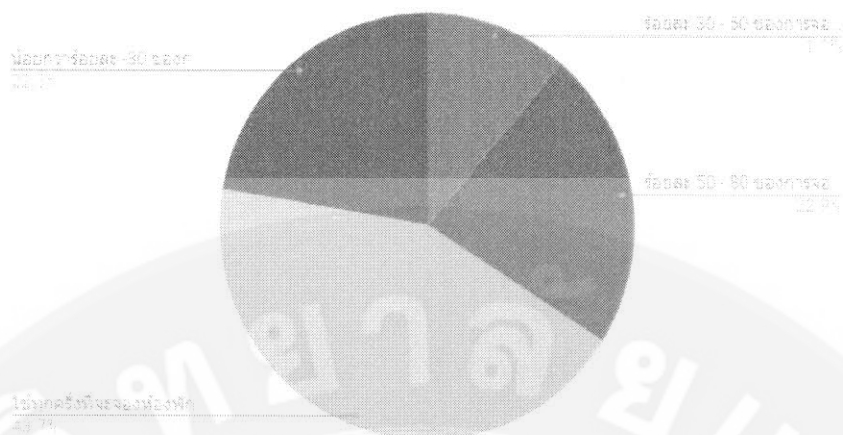
คอมพิวเตอร์แบบพกพาและ
100%โทรศัพท์และคอมพิวเตอร์
0%

ภาพที่ 4.9 อุปกรณ์ของการใช้งานระบบ Online Review ในการจองห้องพักการท่องเที่ยว
เชิงนิเวศ

2. ความถี่ในการใช้งานระบบ Online Review ในการจองห้องพักเชิงนิเวศ

ตารางที่ 4.10 ความถี่ในการใช้งานระบบ Online Review ในการจองห้องพักเชิงนิเวศ

ความถี่ในการใช้งานระบบ Online Review ในการจอง ห้องพัก	ความถี่	ร้อยละ
ใช้ทุกครั้งที่จองห้องพัก	153	43.7
น้อยกว่าร้อยละ -30 ของการจองห้องพัก	78	22.3
ร้อยละ 30 - 50 ของการจองห้องพัก	39	11.1
ร้อยละ 50 - 80 ของการจองห้องพัก	80	22.9
รวม	350	100.00

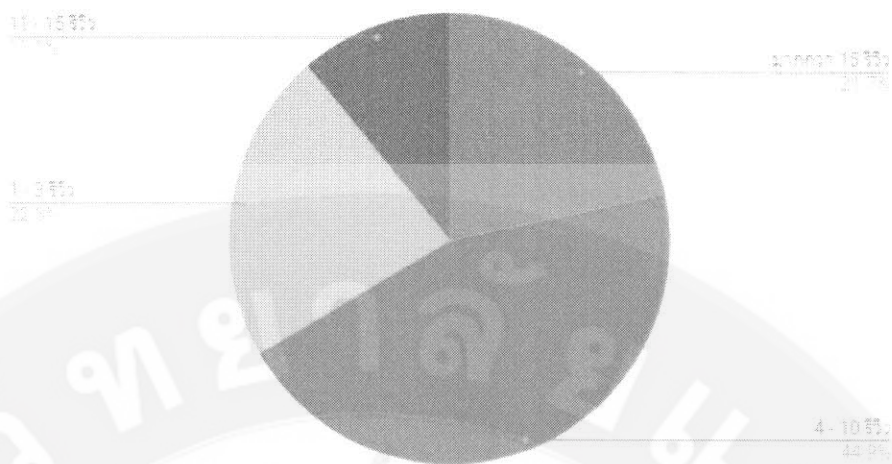


ภาพที่ 4.10 ความดีในการใช้งานระบบ Online Review ในการจองห้องพักเชิงนิเวศ

3. จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาจองห้องพักออนไลน์แต่ละครั้ง

ตารางที่ 4.11 จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาจองห้องพักออนไลน์แต่ละครั้ง

จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาจองห้องพักออนไลน์ แต่ละครั้ง	ความดี	ร้อยละ
1 - 3 รีวิว	78	22.3
11 - 15 รีวิว	39	11.1
4 - 10 รีวิว	157	44.9
มากกว่า 15 รีวิว	76	21.7
รวม	350	100.00

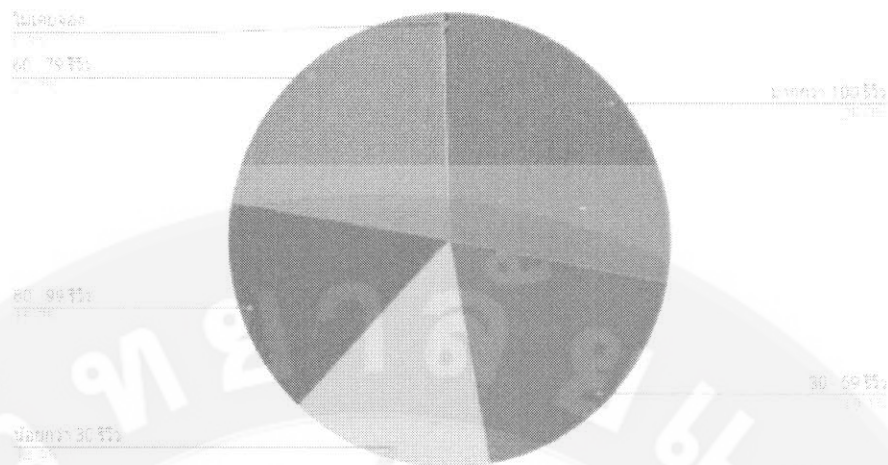


ภาพที่ 4.11 จำนวนรีวิวที่ท่านมักอ่านก่อนพิจารณาของห้องพักออนไลน์แต่ละครั้ง

4. จำนวนรีวิวของผู้เคยใช้บริการของห้องพักที่จะมีผลต่อตัดสินใจของห้องพักเชิงนิเวศ

ตารางที่ 4.12 จำนวนรีวิวของผู้เคยใช้บริการของห้องพักที่จะมีผลต่อตัดสินใจของห้องพักเชิงนิเวศ

จำนวนรีวิวของผู้เคยใช้บริการของห้องพักที่จะมีผลต่อตัดสินใจของห้องพักเชิงนิเวศ	ความถี่	ร้อยละ
30 - 59 รีวิว	67	19.1
60 - 79 รีวิว	76	21.7
80 - 99 รีวิว	55	15.7
น้อยกว่า 30 รีวิว	52	14.9
มากกว่า 100 รีวิว	98	28.0
ไม่ระบุ	2	0.6
รวม	350	100.00



ภาพที่ 4.12 จำนวนรีวิวของผู้ใช้บริการจองห้องพักที่จะมีผลต่อตัดสินใจจองห้องพักเชิง
นิเวศ

ส่วนที่ 3 แสดงผลข้อมูลจากการทำต้นไม้ตัดสินใจและการจำแนกกลุ่มโดย k-mean

ต้นไม้ตัดสินใจ

เทคนิคแรกที่ใช้คือ ต้นไม้ตัดสินใจ โดยต้องมีการกำหนดเป้าหมายในการพยากรณ์ โดยมีการกำหนดค่าดังภาพที่ 4.13 ภาพการกำหนดค่าเป้าหมายและการกำหนดค่าในแถวกลุ่มทดลอง โดยมีการกำหนดค่าเป้าหมายเป็นความพึงพอใจโดยรวมของการจองผ่านระบบ review online โดยแบ่งออกเป็นสามระดับคือ มาก ปานกลางและน้อย โดยกำหนดให้ id เป็นตัวแทนข้อมูลทั้งหมด

Row No.	recid_id	Level	Sex	age	education	occupation	homeland	experience	outcomesper	experience	booking_conf	Device	L_rate_review	agency	
1	1	ปานกลาง	หญิง	ต่ำกว่า 21 ปี	ปริญญาตรี	นักเขียนฟรี	ต่างประเทศ	ไม่มีประสบการณ์	10-20%	1-3 ปี	1-3 ครั้ง	SmartPhone	คะแนน 30	Traveloo	
2	2	สูง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ต่างประเทศ	ไม่มีประสบการณ์	21-30%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	คะแนน 50	Traveloo	
3	3	น้อย	ชาย	ต่ำกว่า 21 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	10-20%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	ไม่มีรีวิว	Traveloo	
4	4	สูง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ต่างประเทศ	ไม่มีประสบการณ์	21-30%	1-3 ปี	2-6 ครั้ง	SmartPhone	ไม่มีรีวิว	Expedia	
5	5	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	31-40%	1-3 ปี	7-12 ครั้ง	Laptop and	ไม่มีรีวิว	Traveloo	
6	6	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	มากกว่า 40%	ไม่มีประสบการณ์	1-3 ครั้ง	Laptop and	ไม่มีรีวิว	Booking
7	7	น้อย	หญิง	ต่ำกว่า 21 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	21-30%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	ไม่มีรีวิว	Agoda.cc	
8	8	ปานกลาง	ชาย	21-37 ปี	ปริญญาตรี	งานประจำ	ประเทศไทย	ไม่มีประสบการณ์	10-20%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	ไม่มีรีวิว	Agoda.cc	
9	9	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	มากกว่า 40%	4-10 ปี	1-3 ครั้ง	SmartPhone	คะแนน 50	Booking
10	10	สูง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	มากกว่า 40%	1-3 ปี	1-3 ครั้ง	SmartPhone	คะแนน 50	Agoda.cc
11	11	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	21-30%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	คะแนน 30	Expedia
12	12	ปานกลาง	ชาย	38-52 ปี	ปริญญาตรี	เจ้าของธุรกิจ	ประเทศไทย	มีประสบการณ์	30,000	มากกว่า 40%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	ไม่มีรีวิว	Booking
13	13	สูง	หญิง	ต่ำกว่า 21 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	10-20%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	คะแนน 30	Traveloo
14	14	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	นักเขียนฟรี	ประเทศไทย	ไม่มีประสบการณ์	ต่ำกว่า 10,000	21-30%	1-3 ปี	1-3 ครั้ง	SmartPhone	ไม่มีรีวิว	Expedia
15	15	ปานกลาง	ชาย	21-37 ปี	ปริญญาตรี	พนักงานออฟฟิศ	ประเทศไทย	ไม่มีประสบการณ์	15,000	มากกว่า 40%	ไม่มีประสบการณ์	1-3 ครั้ง	SmartPhone	คะแนน 30	Agoda.cc
16	16	ปานกลาง	หญิง	21-37 ปี	ปริญญาตรี	พนักงานออฟฟิศ	ประเทศไทย	ไม่มีประสบการณ์	15,000	มากกว่า 40%	1-3 ปี	1-3 ครั้ง	SmartPhone	คะแนน 40	Booking

ภาพที่ 4.13 ภาพการกำหนดค่าเป้าหมายและการกำหนดค่าในแถวกลุ่มทดลอง

ขั้นตอนถัดไปจะเป็นการเชื่อมข้อมูลเพื่อเข้ามาประมวลผลเลือกการจัดการต้นไม้ กำหนดค่าเข้า และออกกระบวนการดังภาพที่ 4.14 การกำหนดกระบวนการต้นไม้ตัดสินใจ



ภาพที่ 4.14 การกำหนดกระบวนการต้นไม้ตัดสินใจ

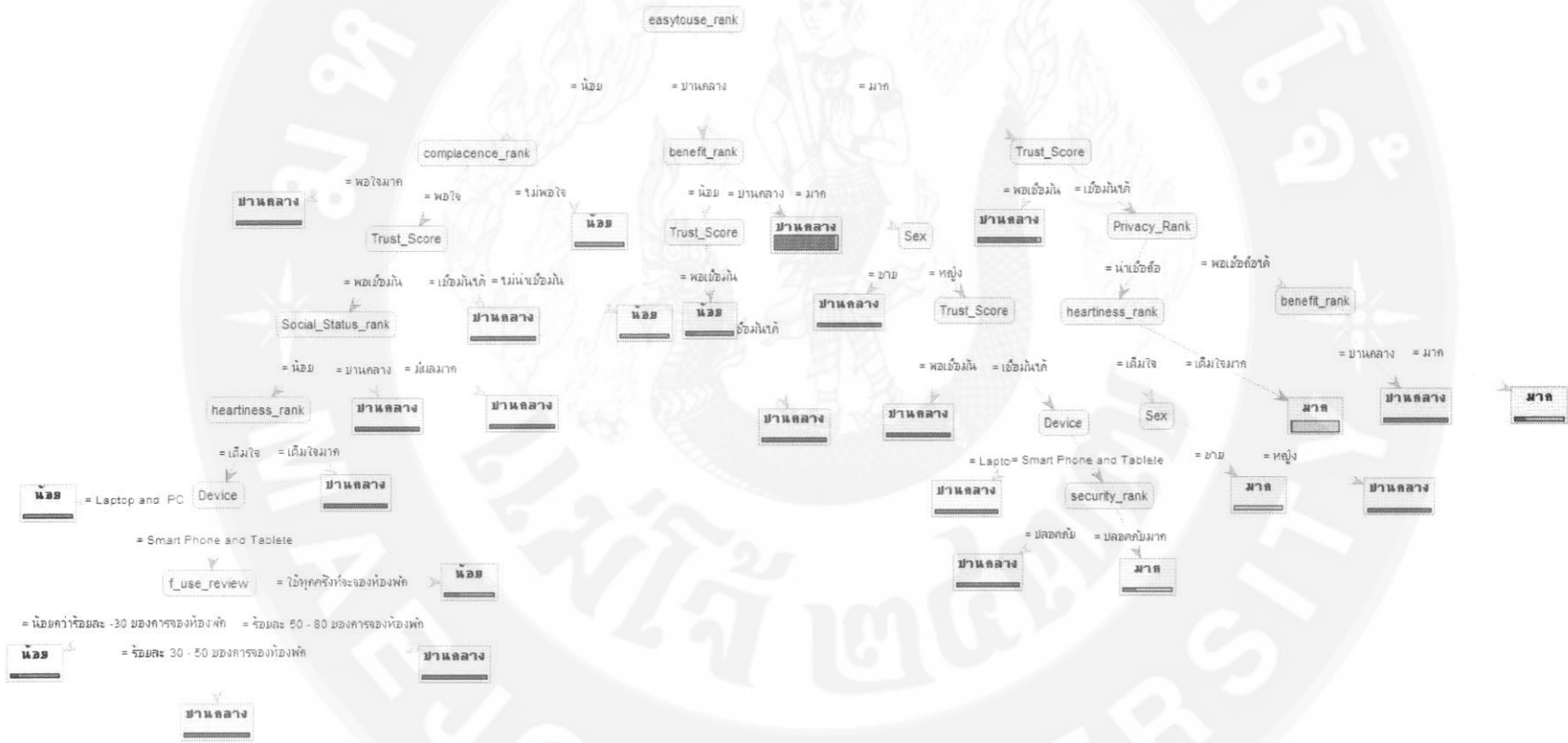
ต่อจากนั้นกำหนดค่าที่จำเป็นในการประมวลผลดังภาพที่ 4.15 การกำหนดค่าก่อนการประมวลผลต้นไม้ตัดสินใจ โดยการใส่ค่าที่จำเป็นต่างๆ เช่น การกำหนดค่า criterion เป็น gain_ratio การกำหนดความลึกของต้นไม้โดย maximal depth เท่ากับ 20 ค่า confidence ของแต่ละโหนดเป็น 0.5

The screenshot shows a 'Decision Tree' configuration window with the following parameters:

Parameter	Value
criterion	gain_ratio
maximal depth	20
apply pruning	<input checked="" type="checkbox"/>
confidence	0.5
apply prepruning	<input checked="" type="checkbox"/>
minimal gain	0.1
minimal leaf size	2
minimal size for split	4
number of prepruning alternati...	3

ภาพที่ 4.15 การกำหนดค่าก่อนการประมวลผลต้นไม้ตัดสินใจ

ทำการประมวลผลเพื่อสร้างต้นไม้ตัดสินใจหาปัจจัยหลักที่ส่งผลกับระดับความพึงพอใจในการใช้ระบบ review online โดยเมื่อประมวลผลแล้วสามารถสร้างต้นไม้ได้ดังภาพที่ 4.16 ผลลัพธ์จากการประมวลผลต้นไม้ตัดสินใจ สามารถอธิบายได้ว่า ประเด็นการใช้งานง่ายของระบบมีอิทธิพลมากที่สุดต่อความพึงพอใจโดยรวมของระบบ review online โดยแบ่งออกหลักในการพิจารณาออกเป็นสามกลุ่ม โดยคือ ใช้งานยาก พอใช้งานได้และใช้งานง่าย โดยส่งผลโดยตรงต่อความพึงพอใจกับระบบโดยรวมจากน้อยไปหามากตามลำดับ โดยปัจจัยที่ส่งผลเป็นลำดับถัดมาคือความเชื่อมั่นในความน่าเชื่อถือของระบบ โดยปัจจัยที่ส่งผลให้ไม่มีความเชื่อมั่นต่อระบบคือ complacence



ภาพที่ 4.16 ผลลัพธ์จากการประมวลผลต้นไม้ตัดสินใจ

การจำแนกกลุ่มด้วย K-Mean

ในการ classification ข้อมูลเทคนิคแรกที่ใช้เป็นการทำ cluster ข้อมูลของเกษตรกรที่ผลิตลำไยคุณภาพ จัดกลุ่มโดยใช้ปัจจัยในเรื่องของ พื้นที่ปลูก การตัดแต่งทรงพุ่ม การตัดช่อผล การใช้สาร การจัดการปุ๋ย ช่องทางการจำหน่ายผลผลิต โดยสามารถจัดกลุ่มข้อมูล โดยใช้ k-mean ได้ดังนี้

ตารางที่ 4.13 การวัดประสิทธิภาพจำนวน cluster จากระยะเฉลี่ยจากค่ากลาง centroid

จำนวน Cluster	ค่าเฉลี่ยระยะห่างของ Centroid
2	89343461.60
3	60633622.08
4	42463694.88
5	30482950.97
6	27910932.63
7	23224790.31
8	19239544.64
9	16602548.16
10	18376145.93
11	11075566.87
12	13334144.86
13	10156514.74
14	9892554.22
15	7318171.39
16	8080188.39
17	5208455.29
18	4760805.87
19	6844908.23
20	3326576.82

จากตาราง การวัดประสิทธิภาพจำนวน cluster จากระยะเฉลี่ยจากค่ากลาง centroid ทำการหาจำนวน cluster ที่เหมาะสมของข้อมูลซึ่งข้อมูลมีความคล้ายคลึงกันและเกาะกลุ่มกันได้อย่างเหมาะสม โดยวัดจากค่าเฉลี่ยจากจุดศูนย์กลาง นั่นก็คือการหา elbow point โดยเป็นจุดที่โค้งกลับของกราฟ โดยอยู่

ค่าระหว่างค่าที่กำลั้งเพิ่มและลดลง โดยโปรแกรมสามารถหาค่าออกมาได้ประมาณโดย k เท่ากับ 9 เพื่อพิจารณาจากกราฟ แต่การอ่านผลโดยการใช้การเกาะกลุ่มจากจำนวนปัจจัยและจำนวนข้อมูลพบว่าจุดเหมาะสมที่มีการเกาะกลุ่มคือ 3 Cluster โดยถือเอาจำนวนดังกล่าวไปประมวลผลข้อมูล

Cluster Model

Cluster 0: 106 items
Cluster 1: 171 items
Cluster 2: 73 items
Total number of items: 350

ภาพที่ 4.17 สรุปจำนวนข้อมูลใน Cluster

ExampleSet (350 examples, 2 special attributes, 9 regular attributes) Filter

Row No.	record_id	cluster	benefit_score	easytouse_...	Privacy_sco...	Social_Stat...	Trust_Score	security_sc...	heartiness_...	complacenc...	total score
1	1	cluster_0	20	14	13	9	19	15	17	16	123
2	2	cluster_2	30	20	20	15	25	25	25	25	185
3	3	cluster_0	18	12	15	9	15	15	15	15	114
4	4	cluster_1	28	17	20	12	20	20	21	24	162
5	5	cluster_0	18	12	16	10	20	18	16	16	130
6	6	cluster_0	22	14	13	7	15	17	20	17	125
7	7	cluster_0	19	13	16	10	13	15	13	17	118
8	8	cluster_1	22	16	20	12	20	20	18	20	148
9	9	cluster_1	30	19	15	12	15	16	16	22	146
10	10	cluster_2	29	18	22	15	25	21	25	24	179
11	11	cluster_0	20	12	14	9	16	17	19	16	125
12	12	cluster_1	25	15	20	12	20	17	20	20	149
13	13	cluster_2	29	17	20	15	24	21	25	25	176
14	14	cluster_1	24	16	20	12	20	20	20	20	152
15	15	cluster_1	24	17	21	13	20	21	21	21	158
16	16	cluster_0	23	12	16	9	18	15	20	20	133
17	17	cluster_1	24	16	20	12	20	20	20	17	149

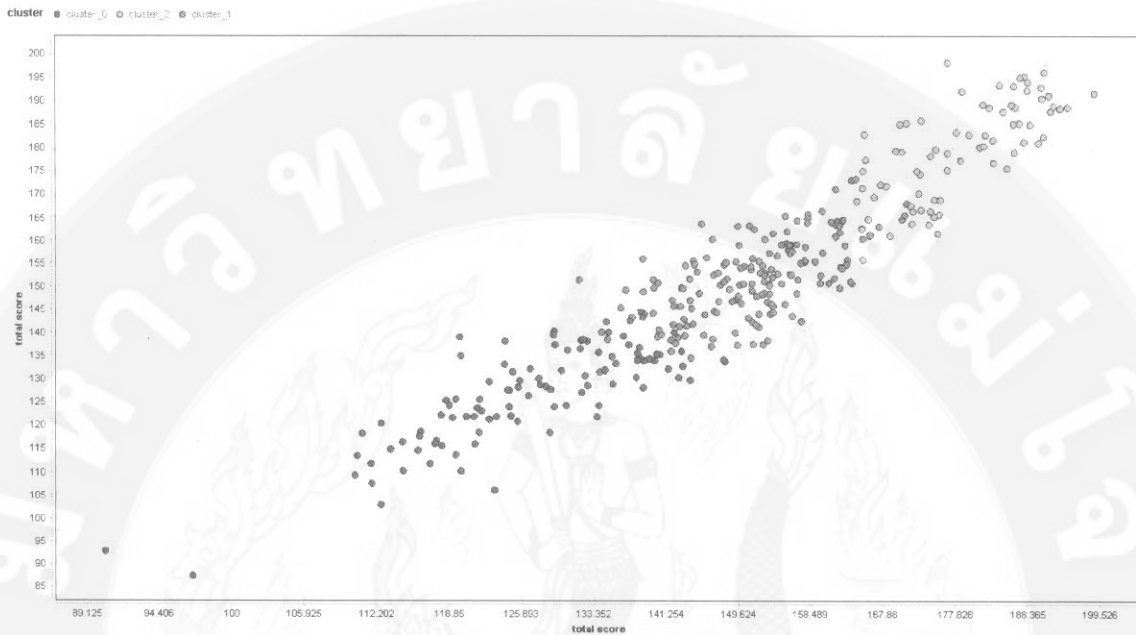
ภาพที่ 4.18 ภาพตัวอย่างข้อมูลที่ถูกแบ่ง Cluster

ตารางที่ 4.14 ค่ากึ่งกลางของแต่ละ Cluster

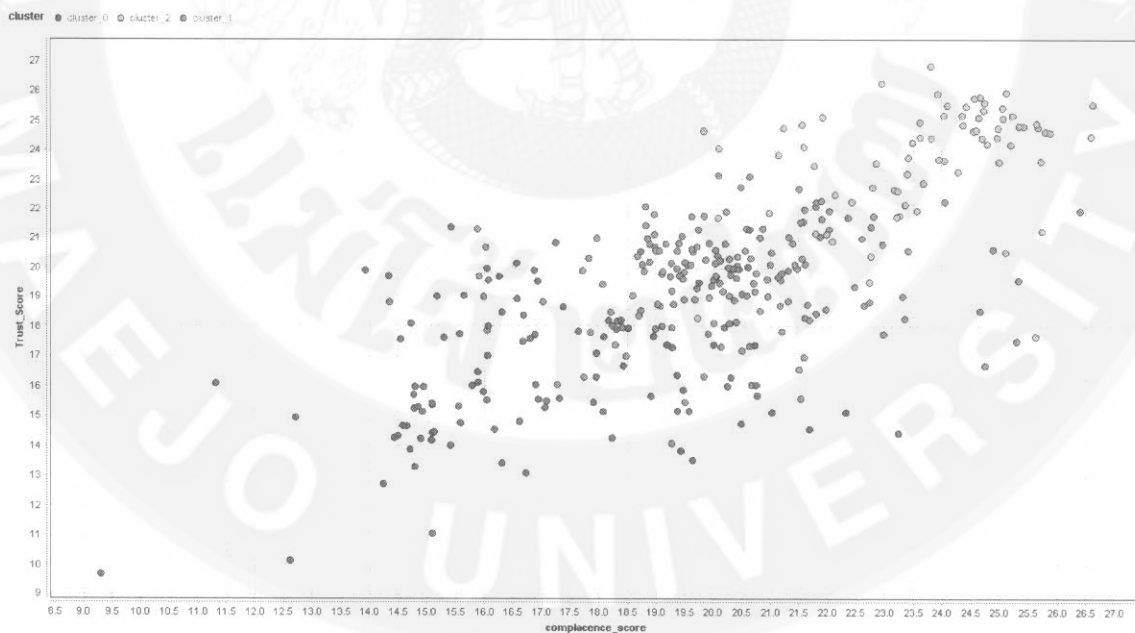
Attribute	cluster_0	cluster_1	cluster_2
benefit_score	21.802	24.906	28.562
easytouse_score	13.887	16.158	18.781
Privacy_score	15.528	19.257	23.192
Social_Status_score	9.504	11.620	13.945
Trust_Score	16.462	19.731	23.644
security_score	15.802	19.567	23.329
heartiness_score	17.142	20.404	23.822
complacence_score	16.934	20.357	23.740
total score	127.160	152	179.014

ภาพที่ 4.17 สรุปจำนวนข้อมูลใน Cluster สามารถสรุปจำนวน cluster ที่เหมาะสมได้ 3 Cluster เมื่อประมวลผล Cluster ที่ 1 คือ Cluster_0 มีจำนวน 106 ราย Cluster ที่ 2 คือ Cluster_1 มีจำนวน 171 ราย และ Cluster ที่ 3 คือ Cluster_2 มีจำนวน 73 ราย

ภาพที่ 4.18 ภาพตัวอย่างข้อมูลที่ถูกแบ่ง Cluster เป็นการแสดง Cluster ของข้อมูลแต่ละรายการ และ ตารางที่ 4.14 ค่ากึ่งกลางของแต่ละ Cluster เป็นการแสดงจุดกึ่งกลางของ Cluster เมื่อดูตามค่าปัจจัย แต่ละค่าโดยการเกาะกลุ่มของข้อมูลเกิดจาก ค่าของความพึงพอใจต่อการใช้งานระบบ review online



ภาพที่ 4.19 การเกาะกลุ่มของข้อมูลค่าความพึงพอใจต่อการใช้ระบบ review online



ภาพที่ 4.20 การเกาะกลุ่มของข้อมูลค่าความเชื่อมั่นการใช้ระบบ review online

ซึ่งจากตัวอย่างการเกาะกลุ่มกันของข้อมูลในแต่ละ Cluster โดยแสดงตัวอย่างตามค่าปัจจัยทั้ง 5 ด้านที่ไว้วิเคราะห์เป็นไปในทิศทางเดียวกัน โดยการจำแนกกลุ่มที่ชัดเจนที่สุดเป็นการใช้ค่าสรุปความพึงพอใจรวมต่อระบบ review online

บทที่ 5

สรุป อภิปรายผล และ ข้อเสนอแนะ

สรุปผลการวิจัย

การวิจัยในครั้งนี้ได้นำเอาข้อมูลกลุ่มลูกค้าของที่พักเชิงนิเวศที่เคยมีประสบการณ์ในการจองที่พักแบบออนไลน์และเคยใช้งานระบบการรีวิวออนไลน์ประกอบการตัดสินใจจองมาก่อน โดยต้องเป็นผู้ใช้งานที่มีอายุตั้งแต่ 18 ปีขึ้นไป และพักอาศัยในประเทศไทย โดยแบ่งพื้นที่ตามภูมิภาคต่างๆ ในประเทศไทย นำมาสร้างสารสนเทศจากข้อมูลสำรวจ ศึกษาปัจจัยต่างๆ ที่มีผลต่อความไว้วางใจในการตัดสินใจในการซื้อบริการที่พักแบบการท่องเที่ยวเชิงนิเวศ โดยใช้เทคนิคเหมืองข้อมูลเข้ามาช่วย 2 เทคนิคด้วยกันโดยเลือกนำเทคนิคแบบจำลองต้นไม้การตัดสินใจและเทคนิคการจำแนกกลุ่มข้อมูล k-mean

แบบจำลองต้นไม้การตัดสินใจ สามารถสร้างต้นไม้การตัดสินใจโดยมุ่งหาปัจจัยที่ส่งผลให้เกิดความพึงพอใจในการใช้งานระบบโดยรวมซึ่งจากผลการศึกษาได้ปัจจัยที่ส่งผลมากที่สุดในระดับแรกคือปัจจัยการรับรู้ว่าการใช้งานรีวิวออนไลน์เป็นเรื่องที่ง่ายต่อการใช้งาน สามารถที่จะส่งผลให้ความไว้วางใจในการตัดสินใจในการซื้อบริการที่พักแบบการท่องเที่ยวเชิงนิเวศมากที่สุด รองลงมาเป็นปัจจัยความไว้วางใจในการใช้งานรีวิวออนไลน์ และปัจจัยความปลอดภัยต่อการใช้งานรีวิวออนไลน์ตามลำดับ เทคนิคการจำแนกกลุ่มข้อมูล k-mean สามารถแบ่งกลุ่มข้อมูลได้เป็น 3 กลุ่มใหญ่ๆ โดยสามารถแบ่งกลุ่มโดยใช้ ปัจจัยความพึงพอใจในการใช้งานระบบการรีวิวออนไลน์ สามารถแบ่งกลุ่มข้อมูลได้ชัดเจนโดยสามารถแบ่งออกเป็นกลุ่มที่มีความพึงพอใจมาก กลางและน้อยตามจุดกึ่งกลางของแต่ละ Cluster

อภิปรายผลผลการวิจัย

ผลที่ได้จากการวิจัย ได้โมเดลที่เหมาะสมในการศึกษาปัจจัยต่างๆ ที่มีผลต่อความไว้วางใจในการตัดสินใจในการซื้อบริการที่พักแบบการท่องเที่ยวเชิงนิเวศ นั้นมีการวัดประสิทธิภาพของโมเดลเป็นสิ่งจำเป็นโดยขึ้นอยู่กับหลายปัจจัย สามารถสรุปการวัดประสิทธิภาพได้จากลักษณะของข้อมูลที่สำรวจและรูปแบบของคำตอบที่ต้องการ โดยการใช้เทคนิคของเหมืองข้อมูลดังที่ใช้ในการวิจัยนั้นในการที่จะเข้าถึงกลุ่มลูกค้าที่ใช้การจองออนไลน์เพื่อเข้ามาพักควรรู้แบบจำลองต้นไม้ตัดสินใจเพื่อจะได้ปัจจัยที่ส่งผลโดยตรงในการตัดสินใจเลือกจองที่พักทำให้เข้าใจถึงประเด็นที่ลูกค้าให้ความสำคัญมากที่สุด

ดังนั้นควรมุ่งเน้นไปที่การใช้งานง่ายของระบบการรีวิวออนไลน์ ส่วนอีกรูปแบบหนึ่งคือการจำแนกกลุ่มลูกค้าโดยใช้ Cluster K-Mean เป็นการแบ่งกลุ่มลูกค้าเพื่อหาลักษณะของลูกค้าที่คล้ายคลึงกันทำให้ทราบว่าปัจจัยที่สามารถแบ่งกลุ่มลูกค้าที่ชัดเจนคือ ปัจจัยความพึงพอใจในการใช้งานระบบการรีวิวออนไลน์ ดังนั้นสามารถที่จะเข้าข้อมูลไปจัดการกับลูกค้าในแต่ละกลุ่มซึ่งมีความระดับความพึงพอใจต่อระบบรีวิวออนไลน์ที่ไม่เท่ากัน

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

จากรูปแบบของโมเดลต้นแบบที่ได้สามารถนำมาขยายผล โดยต้องกำหนดเป้าหมายการศึกษา มีการรวบรวมข้อมูลจากหลายลักษณะ เนื่องจากเทคนิคเหมืองข้อมูลมีหลากหลายวิธีซึ่งเหมาะสมกับข้อมูลในแต่ละแบบที่ต่างกัน จากการวิจัยได้ทำให้เห็นว่าเทคนิคแบบต้นไม้การตัดสินใจสามารถค้นหาปัจจัยที่ส่งผลกระทบต่อการตัดสินใจของลูกค้า เทคนิค cluster k-mean สามารถแยกกลุ่มของลูกค้าได้อย่างชัดเจน โดยใช้ข้อมูลแต่ละปัจจัยมาค้นหากลุ่ม ดังนั้นเมื่อต้องการค้นหาข้อมูลที่แฝงอยู่โดยการทำสารสนเทศในรูปแบบเปรียบเทียบและสรุปข้อมูลไม่สามารถหาคำตอบได้ สามารถนำเอาเทคนิคเหมืองข้อมูลเข้ามาช่วยค้นหาความสัมพันธ์ระหว่างข้อมูล การแบ่งกลุ่มและการจำแนกข้อมูล โดยเลือกให้เหมาะสมกับคำตอบและข้อมูลที่มี

บรรณานุกรม

- กิตติ ภัคดีวัฒนะกุล. (2550). คัมภีร์ระบบสนับสนุนการตัดสินใจและระบบผู้เชี่ยวชาญ(พิมพ์ครั้งที่ 2). กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์.
- พรณิษา เพชรบุญมี, ดวงกมล โพธิ์นาค และมนต์ชัย เทียนทอง. (2556). การพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคลป์ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ. กรุงเทพมหานคร:วารสารวิทยาศาสตร์และเทคโนโลยีปีที่ 21 ฉบับที่ 6 (ฉบับพิเศษ) 2556 บวร น้อยแสง.(2549). คลังข้อมูลและเทคนิคการทำเหมืองข้อมูลสำหรับการวิเคราะห์การขาย . สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต , มหาวิทยาลัยศิลปากร, สาขาวิชา คณิตศาสตร์และเทคโนโลยีสารสนเทศ , ภาควิชาคณิตศาสตร์
- ปรีชา ลี้มตระกูล, วิภา เจริญภัณฑารักษ์ และวิทยา พรพัชรพงศ์. (2559). การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล . E-Journal, Science and Technology Silpakom University ปีที่ 3 ฉบับที่ 3 เดือนพฤษภาคม-มิถุนายน 2559.15-36
- ฝนทิพย์ คุณแก้ว. (2555). การสังเคราะห์โมเดลเพื่อการจำแนกตามข้อกำหนดของผู้ใช้. นครราชสีมา : วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยเทคโนโลยีสุรนารี
- สาธิตา สุทธิไตรรงค์ (2557). ผลกระทบของการสื่อสารแบบปากต่อปากบนอินเทอร์เน็ต (Electronic Word-of-mouth: eWOM) ต่อกระบวนการตัดสินใจซื้อในร้านอาหารของผู้บริโภค". วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. บริหารบูรณาการ. มหาวิทยาลัยธรรมศาสตร์
- Catherine Bounsaythip and Esa Rinta-Runsala. (2001). Overview of Data Mining for Customer Behavior Modeling. VTT INFORMATION TECHNOLOGY. Research Report TTE1-2001-18.
- Jiawei Han and Micheline Kamber, USA : Morgan Kaufman. (2001). Data Mining Concepts and Techniques.
- Michael J.A. Berry and Gordon S. Linoff. (2004). Data Mining Techniques Second Edition For Marketing, Sales, and Customer Relationship Management
- Pongsatorn Tantrabundit (2015), "How online review and rating systems affect online

consumer buying behavior in the hotel industry”, Doctoral dissertation , the Faculty of the Alliant School of Management. Alliant International University.

Qi Luo. (2008). Advancing Knowledge Discovery and Data Mining. 2008 Workshop on Knowledge Discovery and Data Mining. IEEE Computer Society

Rakesh Agrawal, Tomasz Imielinski, and Arun Swami, "Database Mining: A Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, Special Issue on Learning and Discovery in Knowledge-Based Databases, (to appear).

Sang Jun Lee and Keng Siau. (2001). A review of data mining techniques. Industrial Management & Data Systems 101/1[2001], pp. 41-46.





ภาคผนวก ก
ประวัติผู้วิจัย

ประวัติผู้วิจัย

ชื่อ – สกุล : นาย ญัฐดนัย เขียววาท

วัน เดือน ปีเกิด : 2 สิงหาคม 2525

ภูมิลำเนา : เชียงใหม่

ประวัติการศึกษา

- พ.ศ. 2544-2547 ปริญญาตรี สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่
- พ.ศ. 2550-2552 ปริญญาโท สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่

ประวัติการทำงาน

- พ.ศ. 2547-2557 นักวิชาการคอมพิวเตอร์
สำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่
- พ.ศ. 2557-ปัจจุบัน อาจารย์
หลักสูตรระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจ
มหาวิทยาลัยแม่โจ้