การวิเคราะห์ข้อมูลสะท้อนกลับด้านธุรกิจโรงแรม ผ่านบทวิจารณ์ออนไลน์ โดยใช้เทคนิคการวิเคราะห์ความรู้สึก

Analysis of Feedback on Hotel Business via Online Reviews using Sentiment Analysis Techniques

ณรงค์ฤทธิ์ บุญสมาน

Narongrit Bunsaman

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science
Prince of Songkla University

2565

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

การวิเคราะห์ข้อมูลสะท้อนกลับด้านธุรกิจโรงแรม ผ่านบทวิจารณ์ออนไลน์ โดยใช้เทคนิคการวิเคราะห์ความรู้สึก

Analysis of Feedback on Hotel Business via Online Reviews using Sentiment Analysis Techniques

ณรงค์ฤทธิ์ บุญสมาน

Narongrit Bunsaman

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Science
Prince of Songkla University

2565

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ การวิเคราะห์ข้อมูลสะท้อนกลับด้านธุรกิจโรงแรม ผ่านบทวิจารณ์ออนไลน์ โดยใช้เทคนิคการวิเคราะห์ความรู้สึก

ผู้เขียน นาย ณรงค์ฤทธิ์ บุญสมาน

สาขาวิชา วิทยาการข้อมูล

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....

(ผู้ช่วยศาสตราจารย์ ดร. กิตย์ศิริ ช่อเจี้ยง)

คณะกรรมการสอบ

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ชิดชนก โชคสุชาติ)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. กฤติยา ดวงมณี)

.....กรรมการ

(รองศาสตราจารย์ ดร. นวลวรรณ สุนทรภิษัช)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. กิตย์ศิริ ช่อเจี้ยง)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

.....

(ศาสตราจารย์ ดร. ดำรงศักดิ์ ฟ้ารุ่งสาง)

คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ .....
(ผู้ช่วยศาสตราจารย์ ดร. กิตย์ศิริ ช่อเจี้ยง)
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ลงชื่อ .....
(นาย ณรงค์ฤทธิ์ บุญสมาน)
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ .....
(นาย ณรงค์ฤทธิ์ บุญสมาน)
นักศึกษา

ชื่อวิทยานิพนธ์ การวิเคราะห์ข้อมูลสะท้อนกลับด้านธุรกิจโรงแรม ผ่านบทวิจารณ์ออนไลน์ โดยใช้เทคนิคการวิเคราะห์ความรู้สึก
ผู้เขียน นาย ณรงค์ฤทธิ์ บุญสมาน
สาขาวิชา วิทยาการข้อมูล
ปีการศึกษา 2564

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์หาความสัมพันธ์พฤติกรรมของนักท่องเที่ยว และการวิเคราะห์ความรู้สึกจากบทวิจารณ์ออนไลน์รวมไปถึงจัดหมวดหมู่บทวิจารณ์และคำศัพท์ต่าง ๆ โดยข้อมูลได้รวบรวมจากเว็บไซต์ Agoda โดยคัดเลือกเฉพาะข้อมูลโรงแรมใน 3 จังหวัด ที่เป็นที่นิยมในการท่องเที่ยวและเข้าร่วมโครงการ “เราเที่ยวด้วยกัน” ได้แก่ จังหวัดภูเก็ต กระบี่ พังงา จำนวนของบทวิจารณ์ทั้งหมด 246,532 บทวิจารณ์ ในกระบวนการวิเคราะห์ทำการแบ่งเป็น 2 กระบวนการหลัก โดยกระบวนการแรกทำการวิเคราะห์โดยใช้ การวิเคราะห์ความสัมพันธ์ (Association rule) กระบวนการที่สองเป็นการวิเคราะห์ความรู้สึกจากบทวิจารณ์ออนไลน์ โดยการวิเคราะห์ความสัมพันธ์มีวัตถุประสงค์ เพื่อวิเคราะห์พฤติกรรมของนักท่องเที่ยวและเพื่อเปรียบเทียบพฤติกรรมของนักท่องเที่ยวชาวไทยในกรณีเข้าร่วมโครงการ “เราเที่ยวด้วยกัน” โดยทำการวิเคราะห์ 2 ข้อ ได้แก่ ข้อแรกเป็นการวิเคราะห์เพื่อดูพฤติกรรมการเข้าพักโรงแรมของนักท่องเที่ยวในหลายมิติ ซึ่งจะมีการเปรียบเทียบโดยใช้อัลกอริทึม Apriori และ FP-Growth ข้อที่สองเป็นการวิเคราะห์เปรียบเทียบพฤติกรรมนักท่องเที่ยวชาวไทยก่อนเกิดสถานการณ์ COVID-19 และ ขณะเกิดสถานการณ์ COVID-19 โดยทั้งสองขั้นตอนผลลัพธ์กฎความสัมพันธ์จำนวนเท่ากันคือ 295 กฎ แต่ระยะเวลาการประมวลผลของ FP-Growth เร็วกว่าระยะเวลาในการประมวลผลของ Apriori algorithm ถึง 4.79 เท่า ในส่วนของการวิเคราะห์ความรู้สึกจากบทวิจารณ์ออนไลน์ ทำการวิเคราะห์ด้วย การวิเคราะห์ความรู้สึก (Sentiment Analysis) จากนั้นนำผลลัพธ์ที่ได้มาจัดหมวดหมู่แบบจำลองหัวข้อ (Topic Modelling) และนำผลลัพธ์ที่ได้ไปแสดงในรูปแบบ Dashboard

Thesis Analysis of Feedback on Hotel Business via Online Reviews using Sentiment Analysis Techniques
Author Mr. Narongrit Bunsaman
Major Program Data Science
Academic 2021

ABSTRACT

The purpose of this research was to analyze the relationship between tourists' behavior and the sentiment analysis of online reviews, including categorizing reviews and words. The information is collected from the Agoda website, selecting only hotel information in 3 provinces that are popular for tourism and participating in the project “Rao tiew duay kun” including Phuket, Krabi, Phang Nga. There are 246,532 reviews. The analysis process is divided into 2 main processes. The first, process was conducted using Association rule. The second, process is conducting the sentiment analysis from online reviews. The first process aims to find the tourists' behavior and to compare with the Thai tourists behavior who participating in the project “Rao tiew duay kun”. There are two steps. The first step is to analyze the hotel stay behavior of tourists in many perspectives. They were compared using the Apriori and FP-Growth algorithms. The second step is a comparative analysis of the behavior of Thai tourists before the COVID-19 situation and during the COVID-19 situation. Both algorithms produce a total of 295 correlation rules. However, the execution time of FP-Growth, is 4.79 times faster than that of Apriori algorithm. In terms of sentiment analysis from online reviews, sentiment analysis was performed and the results were categorized (Topic Modeling) and display the results in Dashboard format.

กิตติกรรมประกาศ

วิทยานิพนธ์เรื่อง การวิเคราะห์ข้อมูลสะท้อนกลับด้านธุรกิจโรงแรม ผ่านบทวิจารณ์ออนไลน์ โดยใช้เทคนิคการวิเคราะห์ความรู้สึก สำเร็จลุล่วงไปได้ด้วยความกรุณาอย่างสูง และความอนุเคราะห์จาก ผู้ช่วยศาสตราจารย์ ดร.กิตย์ศิริ ช่อเจี้ยง เป็นอาจารย์ที่ปรึกษา ซึ่งได้ให้ความรู้ คำแนะนำและเป็นที่ปรึกษา รวมไปถึงการชี้แนะแนวทางในการจัดทำ และบอกถึงจุดบกพร่องในการทำวิทยานิพนธ์ ด้วยความเอาใจใส่และความทุ่มเทที่ทางอาจารย์ที่ปรึกษาได้มอบให้ ทำให้ผู้วิจัยสามารถจัดทำวิทยานิพนธ์ไปได้สำเร็จอย่างลุล่วงด้วยดี และสามารถนำวิชาต่าง ๆ ที่ได้เรียนมาประยุกต์ใช้ในการทำวิทยานิพนธ์ครั้งนี้ จึงขอกราบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณ รศ.ดร.นวลวรรณ สุนทรภิษัช ดร.ชิดชนก โชคสุชาติ และดร.กฤติยา ดวงมณี เป็นคณะกรรมการในการสอบวิทยานิพนธ์ในครั้งนี้ ซึ่งได้รับคำแนะนำ และแสดงความคิดเห็นถึงจุดบกพร่องต่าง ๆ ในการทำวิทยานิพนธ์ได้อย่างทั่วถึง เพื่อให้วิทยานิพนธ์ครั้งนี้ สามารถแก้ไขและสำเร็จลุล่วงไปด้วยดี โดยผู้วิจัยหวังว่าจะมีประโยชน์ต่อผู้อ่านไม่มากก็น้อย ผู้วิจัยขอกราบขอบพระคุณบิดา มารดาและผู้มีพระคุณทุกท่าน รวมไปถึงครอบครัวที่เป็นแรงบันดาลใจทำให้วิทยานิพนธ์สำเร็จลุล่วงไปด้วยดี หากเกิดข้อบกพร่องประการใดทางผู้จัดทำขออภัยอย่างสูงมา ณ ที่นี้ด้วย

นาย ณรงค์ฤทธิ์ บุญสมาน

สารบัญ

สารบัญภาพ (11)

สารบัญตาราง (14)

บทที่1 บทนำ 1

1.1 ความสำคัญและปัญหา 1

1.2 นิยามความหมาย 3

1.3 วัตถุประสงค์ 4

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 4

1.5 ขอบเขตการดำเนินงานวิจัย 5

บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 6

2.1 Sentiment Analysis 6

2.2 Association Rules 8

บทที่ 3 กรอบการดำเนินงานวิจัย 14

3.1 การนิยามปัญหา 14

3.2 กรอบการวิเคราะห์บทวิจารณ์ออนไลน์ 16

3.2.1 รวบรวมรายชื่อโรงแรมที่เข้าร่วมโครงการ “เราเที่ยวด้วยกัน” 18

3.2.2 รวบรวมรีวิวจากเว็บไซต์ Agoda 18

3.2.3 กระบวนการเตรียมข้อมูล 19

3.2.4 กระบวนการวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูลและการวัดประสิทธิภาพความเร็วในการประมวลผล 21

3.2.4.1 การแปลงชุดข้อมูลที่ใช้ในการทดลอง 21

3.2.4.2 การเปรียบเทียบประสิทธิภาพความเร็วในการประมวลผลระหว่าง Rapid Miner และ R-Studio 26

3.2.4.3 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม Apriori โดยการใช้ R-Studio 29

3.2.4.4 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม FP-Growth โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent 29

สารบัญ (ต่อ)

3.2.4.5 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA2 จากอัลกอริทึม FP-Growth และ Apriori โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent 30

3.2.5 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ 31

3.2.6 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ 31

3.2.6.1 การวิเคราะห์อารมณ์และความรู้สึกของประโยคทั้งหมดในแต่ละบทวิจารณ์ 32

3.2.6.2 การวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์ 32

3.2.6.3 การจัดกลุ่มคำศัพท์ 33

3.2.6.4 การแสดงผลลัพธ์ 36

บทที่ 4 ผลดำเนินงานวิจัย 39

4.1 การวิเคราะห์กฎความสัมพันธ์ของข้อมูล 39

4.1.1 ข้อมูลเชิงลึกของบทวิจารณ์ออนไลน์ที่ได้รวบรวม 39

4.1.2 ผลการทดลองการวิเคราะห์บทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020 (อ้างอิงจากข้อ 3.2.4) 40

4.2 ผลการทดลองการวิเคราะห์หาพฤติกรรมนักท่องเที่ยวชาวไทยในช่วงก่อนการเกิดสถานการณ์ COVID-19 และขณะเกิด COVID-19 (อ้างอิงจากข้อ 3.2.5) 41

4.2.1 การหากฎความสัมพันธ์ที่เกี่ยวข้อง 41

4.2.2 การแปรผลลัพธ์ที่ได้จากการหากฎความสัมพันธ์ 42

4.3 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ 46

4.3.1 ข้อมูลเชิงลึกของบทวิจารณ์ออนไลน์ที่ได้รวบรวม 46

4.3.2 ผลการทดลองการวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ (อ้างอิงจากข้อ 3.2.6.1) 47

4.3.3 ผลการทดลองการวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์ (อ้างอิงจากข้อ 3.2.6.2) 50

สารบัญ (ต่อ)

	4.3.4 ผลการทดลองการจัดกลุ่มคำศัพท์ 2 วิธีเพื่อทำการเปรียบเทียบผลลัพธ์และนำผลลัพธ์ที่เหมาะสมที่สุดมาใช้ (อ้างอิงจากข้อ 3.2.6.3)	50
	4.3.5 ผลการแสดงผลลัพธ์ (อ้างอิงจากข้อ 3.2.6.4)	51
บทที่ 5 สรุปผล		57
อ้างอิง		
ภาคผนวก		

สารบัญภาพ

รูปที่ 1 : การแบ่งแยกจำนวนบทวิจารณ์ในแต่ละปี ตั้งแต่ปี 2009-2020 15

รูปที่ 2 : ตัวอย่างรายละเอียดบทวิจารณ์ออนไลน์ที่สนใจ 15

รูปที่ 3 : System Architecture 16

รูปที่ 4 : ตัวอย่างผลลัพธ์ข้อมูลที่ได้จากการจัดกลุ่มข้อมูล 19

รูปที่ 5 : การแปลงคอลัมน์ให้อยู่ในรูปแบบแถว 20

รูปที่ 6 : กราฟแสดงจำนวนบทวิจารณ์ก่อนและหลังกระบวนการเตรียมข้อมูล 20

รูปที่ 7 แสดงชุดข้อมูลในรูปแบบของ binominal (DATA1) 21

รูปที่ 8 แสดงชุดข้อมูลในรูปแบบของ transaction (DATA2) 22

รูปที่ 9 : Process Model ที่ใช้ในการทำสร้าง Association Rule 23

รูปที่ 10 : ข้อมูล ExampleSet จากการประมวลผลจากโมเดล FP-Growth 23

รูปที่ 11 : ข้อมูล Frequency Item Sets จากการประมวลผลจากโมเดล Create Association Rules 24

รูปที่ 12 : ข้อมูล Association Rules จากการประมวลผลจากโมเดล Create Association Rules 24

รูปที่ 13 : ข้อมูล Description Association Rules (support = 0.01, confidence = 0.5) จากการประมวลผลจากโมเดล Create Association Rules 25

รูปที่ 14 : วิธีการวิเคราะห์ Association Rules โดยใช้อัลกอริทึม Apriori 26

รูปที่ 15 : Process Model ที่ใช้ในการวัดประสิทธิภาพความเร็วในการประมวลผล 27

รูปที่ 16 : โค้ดที่ใช้ในการประมวลผลของโปรแกรม R Studio 28

รูปที่ 17 : แสดงตัวอย่างโค้ดที่มีการระบุ consequent = “continent” 30

รูปที่ 18 : แสดงตัวอย่างโค้ดที่มีการระบุส่วนของ continent 31

รูปที่ 19 : วิธีการแยกคำและหาชนิดของคำ 32

รูปที่ 20 : ตัวอย่างโค้ดวิธีการแยกคำ วิเคราะห์อารมณ์และความรู้สึกของคำ และหาชนิดของคำ 33

สารบัญภาพ (ต่อ)

รูปที่ 21 : วิธีการจัดกลุ่มคำโดยใช้ Model LDA 34

รูปที่ 22 : ผลลัพธ์ที่ได้จากการจัดกลุ่มโดยใช้ Model LDA รูปแบบ wordcloud 35

รูปที่ 23 : ตัวอย่างผลลัพธ์ 20 คำแรกที่ได้จากการจัดกลุ่มแบบ Manual 36

รูปที่ 24 : ผลลัพธ์ 5 อันดับโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุดของ 3 จังหวัด 37

รูปที่ 25 : ตารางสรุปผลลัพธ์ที่ได้ของชุดข้อมูลบทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020 40

รูปที่ 26 : ตารางสรุปผลลัพธ์ที่ได้จากการหาความสัมพันธ์ช่วงก่อนการระบาด COVID-19 42

รูปที่ 27 : ตารางสรุปผลลัพธ์ที่ได้จากการหาความสัมพันธ์ช่วงหตุการระบาด COVID-19 44

รูปที่ 28 : ชุดข้อมูลกฎความสัมพันธ์ที่เกิดขึ้นเหมือนกันจาก ผลลัพธ์ข้อมูลชุดที่ 1 และ ผลลัพธ์ข้อมูลชุดที่ 2 45

รูปที่ 29 : 10 อันดับแรกของประเทศที่มีจำนวนบทวิจารณ์เชิงบวกและเชิงลบเยอะที่สุด 47

รูปที่ 30 : จำนวนบทวิจารณ์เชิงบวกและเชิงลบทั้งหมดโดยแบ่งเป็น 3 จังหวัด 47

รูปที่ 31 : จำนวนบทวิจารณ์เชิงบวกและเชิงลบเฉพาะประเทศไทยทั้งหมดโดยแบ่งเป็น 3 จังหวัด 48

รูปที่ 32 : จำนวนบทวิจารณ์ทั้งหมดเชิงบวก และเชิงลบทั้งหมด 3 จังหวัดในแต่ละปี 49

รูปที่ 33 : ผลลัพธ์ที่ได้จากการแยกแยกคำศัพท์และหาชนิดของคำ 50

รูปที่ 34 : การแสดงผลลัพธ์ทั้งหมดในรูปแบบ Dashboard โดยใช้ Data Studio 51

รูปที่ 35 : การแสดงผลลัพธ์ทั้งหมดในรูปแบบ Dashboard แบ่งออกเป็น 3 ส่วน 51

รูปที่ 36 : การแสดงผลลัพธ์คำศัพท์ที่ได้จากบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและรูปแบบ Word cloud 53

รูปที่ 37 : การแสดงประโยคบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและผลลัพธ์ค่าของ Sentiment 53

รูปที่ 38 : การแสดงผลลัพธ์ประเทศของนักท่องเที่ยว 54

รูปที่ 39 : การแสดงผลลัพธ์รูปแบบแผนภูมิวงกลมและรูปแบบแผนภูมิแท่ง 55

สารบัญภาพ (ต่อ)

รูปที่ 40 : การแสดงผลลัพธ์จำนวนบทวิจารณ์รายเดือนและรายปีแสดงผลลัพธ์ในรูปแบบกราฟเส้น 55
รูปที่ 41 : การแสดงผลลัพธ์โดยการกรองข้อมูลในส่วนของชื่อโรงแรมและหมวดหมู่ 55

สารบัญตาราง

ตารางที่ 1 แสดงจำนวนบทวิจารณ์และจำนวนโรงแรมของ 3 จังหวัด 14

ตารางที่ 2 ตัวอย่างรายละเอียดตัวแปรในชุดข้อมูล 18

ตารางที่ 3 รายละเอียดความสัมพันธ์ระหว่างชุดข้อมูล 21

ตารางที่ 4 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA 1 23

ตารางที่ 5 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2จากอัลกอริทึม Apriori โดยการใช้ R-Studio 29

ตารางที่ 6 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม FP-Growth โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent 29

ตารางที่ 7 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA2 จาก อัลกอริทึม FP-Growth และ Apriori โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent 30

บทที่ 1
บทนำ

1.1 ความสำคัญและปัญหา

“การท่องเที่ยว” เป็นกิจกรรมอย่างหนึ่งเป็นที่นิยมของคนทั่วโลก การท่องเที่ยวมีทั้งท่องเที่ยวในประเทศและท่องเที่ยวนอกประเทศ จุดประสงค์ของการท่องเที่ยว คือ การพักผ่อน หรือสร้างความสนุกสนานและคลายเครียด การท่องเที่ยวสามารถไปแบบลักษณะกลุ่มเพื่อน ครอบครัว หรือสามารถไปท่องเที่ยวคนเดียว สิ่งสำคัญของการท่องเที่ยว คือ การหาที่พักที่เหมาะสมกับความต้องการตนเอง ที่พักที่มีสิ่งอำนวยความสะดวกครบครัน จะส่งผลให้นักท่องเที่ยวผ่อนคลาย และมีความสุข สร้างความประทับใจให้กับผู้เข้าพัก ทางที่พักสามารถนำเสนอจุดเด่นของตัวที่พักเพื่อดึงดูดนักท่องเที่ยวผ่านเว็บไซต์ต่าง ๆ หรือสื่อชนิดอื่น ๆ การที่จะเข้าพักตามที่พักต่าง ๆ นักท่องเที่ยวควรค้นหาข้อมูลของที่พักผ่านสื่อต่าง ๆ และอ่านบทวิจารณ์เพื่อเป็นตัวตัดสินใจว่าควรเข้าพักสถานที่ใดที่จะทำให้ตัวเราเกิดความผ่อนคลายและประทับใจ

ปัจจุบันสื่อออนไลน์เข้ามามีบทบาทในด้านการท่องเที่ยวเป็นอย่างมาก [1] เนื่องจากการท่องเที่ยวโดยส่วนใหญ่ เป็นในลักษณะของการท่องเที่ยวเพื่อเปิดโลกทัศน์ท่องเที่ยวในสถานที่ใหม่ ๆ เพื่อเพิ่มพูนประสบการณ์ จึงเป็นสาเหตุให้นักท่องเที่ยวยังไม่เคยเดินทางเข้าพักในสถานที่นั้นจริง ๆ ดังนั้นเพื่อความปลอดภัยของนักท่องเที่ยวเอง นักท่องเที่ยวจึงจำเป็นต้องศึกษารีวิวจากนักท่องเที่ยวที่เคยเข้าพักก่อนหน้า โดยปัจจุบันเครือข่ายสังคมออนไลน์นั้น มีการเปิดให้แสดงความคิดเห็นมากมาย รวมทั้งการรีวิวประวัติการเข้าพัก ซึ่งจะมีรายละเอียดเกี่ยวกับจุดเด่นและข้อควรปรับปรุงของสถานที่พักนั้น ๆ โดยความคิดเห็นเหล่านี้มีอิทธิพลเป็นอย่างมากต่อการตัดสินใจในการเข้าพักโรงแรมของนักท่องเที่ยว [2], [3] นอกจากนี้ในมุมมองของผู้ประกอบการและหน่วยงานที่เกี่ยวข้อง ทั้งระดับของตำบล อำเภอ และจังหวัด ข้อมูลความคิดเห็นจำนวนมากเหล่านี้เปรียบเสมือนได้กับขุมทรัพย์ที่มีประโยชน์อย่างมาก แต่ในปัจจุบันการนำข้อมูลส่วนนี้มาก่อให้เกิดประโยชน์ จะมีการวิเคราะห์ข้อมูลเฉพาะโรงแรมที่มีงบประมาณจำนวนมาก หรือการวิเคราะห์เฉพาะโรงแรมเฉพาะกลุ่มเท่านั้น จากที่ได้กล่าวมาแล้วข้างต้น ส่งผลให้หน่วยงานที่เกี่ยวข้องไม่สามารถมองเห็นภาพรวมของนักท่องเที่ยวที่มารับบริการเข้าพักโรงแรมในจังหวัดที่ต้องการได้

ในปัจจุบันอุตสาหกรรมการท่องเที่ยวซบเซาจากวิกฤตการณ์ระบาดของ COVID-19 ทำให้ประชาชนทั่วโลกได้รับผลกระทบรวมไปถึงประเทศไทยเช่นกัน การระบาดของ COVID-19 ส่งผลกระทบต่อเศรษฐกิจไทยอย่างรุนแรง โดยเฉพาะอย่างยิ่งการเดินทางท่องเที่ยวระหว่างประเทศส่งผลให้จำนวนนักท่องเที่ยวต่างชาติลดน้อยลงจนกระทั่งเป็นศูนย์เมื่อมีนโยบายปิดประเทศ [4] อีกทั้งยังส่งผลให้คนไทยเกิดการว่างงานเพิ่มมากขึ้น โดยเฉพาะแรงงานจากภาคธุรกิจด้านการท่องเที่ยวที่ปิดตัวลง ดังนั้นจังหวัดที่เป็นจุดหมายของนักท่องเที่ยวโดยเฉพาะจังหวัดภูเก็ต กระบี่ และพังงาโดยสถานที่ท่องเที่ยวต่าง ๆ มากมายในจังหวัดได้รับรางวัลในระดับนานาชาติ [5] ได้รับผลกระทบอย่างมหาศาล สถานที่ต่าง ๆ ไร้นักท่องเที่ยว นอกจากนี้ธุรกิจที่พึ่งพาสถานที่ท่องเที่ยวเหล่านี้ เช่น ร้านอาหาร ผับ บาร์ โรงแรม และรีสอร์ตต่าง ๆ รวมไปถึงธุรกิจของฝากและของที่ระลึกซบเซาตามไปด้วย จากผลกระทบดังกล่าวรัฐบาลจึงหาแนวทางแก้ไขและช่วยเหลือประชาชนที่เกี่ยวข้องทุกฝ่าย จึงได้มีโครงการ “เราเที่ยวด้วยกัน” เพื่อมาช่วยกระตุ้นเศรษฐกิจภาคการท่องเที่ยวให้ดีขึ้น โครงการเราเที่ยวด้วยกันมีการสนับสนุนส่วนลดค่าโรงแรมที่พัก สนับสนุนส่วนลดค่าอาหาร ค่าเข้าสถานที่ท่องเที่ยวและสนับสนุนค่าเดินทางโดยเครื่องบิน

“โครงการเราเที่ยวด้วยกัน” เป็นโครงการที่ช่วยให้นักท่องเที่ยว ชำระค่าห้องพัก ค่าอาหาร เพียง 60% ของค่าใช้จ่ายทั้งหมด ด้วยเหตุนี้ทำให้โอกาสหรือขอบเขตงบประมาณในการเลือกเข้าพักโรงแรมขยายมากยิ่งขึ้น นักท่องเที่ยวได้ตัวเลือกเข้าพักในโรงแรมได้มากขึ้น เจ้าของธุรกิจโรงแรมสามารถเสนอขายหรือผลิตโปรโมชั่นได้มากขึ้น ซึ่งเป็นโอกาสที่ดีในการนำเสนอโรงแรมให้เป็นที่รู้จักได้อีกทางหนึ่ง การเข้าพักโรงแรมที่อยู่ใกล้เคียงกับสถานที่ที่เลือกท่องเที่ยวนั้น เป็นหนึ่งทางเลือกสำคัญที่นักท่องเที่ยวใช้เป็นตัวเลือกในการตัดสินใจเลือกเข้าพักโรงแรมใดโรงแรมหนึ่ง นักท่องเที่ยวมักจะสำรวจและค้นหาข้อมูลที่เกี่ยวข้อง เช่น การค้นหาข้อมูลโรงแรม การดูรูปภาพประกอบ และการอ่านรีวิวของโรงแรม เป็นต้น เพื่อเป็นการตัดสินใจในการเข้าพัก [6] ในปัจจุบันผู้ประกอบการคิดค้นวิธีการต่าง ๆ ในการดึงดูดนักท่องเที่ยว การแชร์ข้อมูลต่าง ๆ ลงในโซเชียลมีเดียหรือเว็บไซต์เพื่อเข้าถึงนักท่องเที่ยว ทำให้นักท่องเที่ยวมองเห็นข้อมูลและภาพของสถานที่นั้น ๆ ได้จริง ซึ่งส่งผลต่อการตัดสินใจในการเดินทางมาเข้าพัก [7] นักวิจัยสามารถวิเคราะห์พฤติกรรมและความชอบของนักท่องเที่ยวได้จากประสบการณ์ของนักท่องเที่ยวที่เผยแพร่ลงในเว็บไซต์หรือโซเชียลมีเดีย [8] ความคิดเห็นต่าง ๆ ของผู้บริโภคที่เขียนลงในเว็บไซต์ต่าง ๆ เป็นสิ่งสำคัญต่อการตัดสินใจของนักท่องเที่ยว รวมทั้งลักษณะการสื่อสาร การใช้คำเรียบเรียงประโยค ใช้คำให้ดูน่าสนใจเป็นอีกสิ่งหนึ่งที่มีผลต่อการ

ตัดสินใจเพื่อดึงดูดนักท่องเที่ยว [9] และทำให้นักท่องเที่ยวเห็นความน่าเชื่อถือของสถานที่นั้น ๆ [10], [11] ในปัจจุบันสื่อโซเชียลมีเดียมีการเติบโตขึ้นอย่างมาก ทำให้ทุกคนรวมถึงผู้ประกอบการหันมาใช้โซเชียลมีเดียในการทำธุรกิจต่าง ๆ โดยการใส่ข้อมูลเชิงท่องเที่ยวเพื่อทำให้ผู้คนสนใจ และเพื่อกระตุ้นการตัดสินใจ ผู้ประกอบการได้สร้างช่องทางในการให้ผู้บริโภคสามารถแสดงความคิดเห็นในด้านที่ตนเองประทับใจ สิ่งเหล่านี้สามารถทำให้สถานที่เหล่านั้นดูมีความน่าเชื่อถือเพิ่มมากขึ้น [12]

จากการศึกษาในหัวข้อดังกล่าวผู้วิจัยได้นำข้อมูลมาหาความสัมพันธ์โดยใช้การหาความสัมพันธ์ของข้อมูล โดยใช้กฎความสัมพันธ์ (Association Rules) ด้วยอัลกอริทึม Apriori และ FP-Growth เพื่อค้นหาไอเทมเซ็ตที่ปรากฏร่วมกันบ่อย รวมทั้งการวัดประสิทธิภาพความเร็วในการทำงานของทั้งสองอัลกอริทึม หลังจากนั้นทำการวิเคราะห์เปรียบเทียบพฤติกรรมนักท่องเที่ยวชาวไทยก่อนเกิดสถานการณ์ COVID-19 และ ขณะเกิดสถานการณ์ COVID-19 และทำการวิเคราะห์ข้อมูลความรู้สึกจากบทวิจารณ์โดยใช้ การวิเคราะห์ความรู้สึก (Sentiment Analysis) เพื่อบอกความรู้สึกจากบทวิจารณ์ออนไลน์ [13] และแยกคำศัพท์ในบทวิจารณ์เพื่อนำคำศัพท์ที่ได้จัดหมวดหมู่เพื่อใช้เป็นคลังข้อมูลในการจัดหมวดหมู่บทวิจารณ์เพื่อการวิเคราะห์ข้อมูลโดยทำการเปลี่ยนเทียบ 2 วิธี การสร้างแบบจำลองหัวข้อ (Topic Modelling) [xx] และวิธีการป้อนข้อมูลโดยผู้วิจัย (Manual) และคัดเลือกผลลัพธ์วิธีที่ที่เหมาะสมสำหรับงานวิจัยครั้งนี้ และนำผลลัพธ์ไปแสดงในรูปแบบ Dashboard บทวิจัยนี้จะสามารถช่วยให้ผู้ที่สนใจหรือหน่วยงานที่เกี่ยวข้อง สามารถนำไปใช้ประโยชน์ในการพัฒนาวางแผนรับมือกับสถานการณ์ในอนาคตได้ [14]

1.2 นิยามความหมาย

บทวิจารณ์ออนไลน์ (Reviews) คือ ข้อคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งที่นำมาใช้เพื่อประเมินความชอบ ไม่ชอบ หรือผลสำเร็จของสิ่งนั้นเป็นส่วนหนึ่งของผู้ใช้บริการ โดยที่ได้รับหลังจากการใช้บริการโรงแรมซึ่งประกอบด้วยข้อเท็จจริง ความคิดเห็นและข้อเสนอแนะที่แสดงอารมณ์ ความรู้สึก บทวิจารณ์ออนไลน์เป็นกระบวนการหนึ่งในการบ่งบอกถึงสิ่งที่เกิดขึ้น และมีการสะท้อนกลับซึ่งช่วยให้เกิดการพัฒนา ปรับปรุงและเรียนรู้ที่จะทำให้ดีขึ้น

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการวิเคราะห์อารมณ์และความรู้สึกจากข้อความ เพื่อบอกความรู้สึกของผู้คนที่มีต่อบางสิ่งบางอย่าง และแบ่งได้เป็น ความรู้สึกดี (Positive) ความรู้สึกที่ไม่ดี (Negative) ซึ่งสามารถนำมาช่วยในการสรุปความหมายของบทวิจารณ์ออนไลน์ได้ว่า

มีความคิดเห็นไปในทางไหน เพื่อเป็นแนวทางให้ปฏิบัติ หรือการชี้ให้เห็นข้อบกพร่องพร้อมทั้งเสนอแนวทางแก้ไข ดังนั้นการวิเคราะห์บทวิจารณ์ออนไลน์ทำให้ทราบว่าผู้ใช้บริการให้ความสำคัญกับเรื่องใด มีข้อเสนอแนะหรือความต้องการให้ปรับปรุงในส่วนใด

กฎความสัมพันธ์ (Association Rule) เป็นการหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายในกลุ่มข้อมูลขนาดใหญ่ เพื่อนำผลลัพธ์ที่ได้จากความสัมพันธ์ มาใช้ในการพัฒนาบริการและการวางแผนรับมือกับสถานการณ์ในอนาคตได้

1.3 วัตถุประสงค์

1.3.1 เพื่อสร้างคลังข้อมูลคำศัพท์ (Corpus) และประโยคบอกความรู้สึกในการเข้าพักโรงแรมทั้งเชิงบวกและเชิงลบ

1.3.2 เพื่อศึกษาแนวโน้มและวิเคราะห์ความสัมพันธ์ของนักท่องเที่ยวที่เข้าพัก แยกตามทวีป หรือโซน รวมทั้งรูปแบบการเดินทางของนักท่องเที่ยว สิ่งอำนวยความสะดวก ช่วงเวลา จำนวนคืนเข้าพัก เป็นต้น

1.3.3 เพื่อเปรียบเทียบและหาความสัมพันธ์ของนักท่องเที่ยวชาวไทยในกรณีเข้าร่วมโครงการ “เราเที่ยวด้วยกัน”

1.3.4 เพื่อจัดหมวดหมู่ของบทวิจารณ์และคำศัพท์ต่างๆ จากบทวิจารณ์ผ่านการวิเคราะห์ความรู้สึก

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1.4.1 คลังข้อมูลคำศัพท์ (Corpus) บอกความรู้สึกทั้งเชิงบวกและเชิงลบจากข้อมูลรีวิว

1.4.2 แนวโน้มและความสัมพันธ์ของนักท่องเที่ยวชาวไทยที่เข้าพักก่อนและหลังเกิดเหตุการณ์ COVID-19 โครงการ “เราเที่ยวด้วยกัน”

1.4.3 บทวิจารณ์ที่ได้จากการวิเคราะห์ความรู้สึกโดยแบ่งเป็นหมวดหมู่

1.4.4 Dashboard แนวโน้มความสัมพันธ์ของนักท่องเที่ยวที่เข้าพักโรงแรม 11 ปีย้อนหลัง

1.5 ขอบเขตการดำเนินงานวิจัย

1.5.1 ข้อมูลที่ใช้ในงานวิจัย เป็นข้อมูลที่ได้จากการใช้เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping Techniques) ในภาษา R

1.5.2 ข้อมูลที่เก็บรวบรวมเป็นข้อมูลของโรงแรมที่เข้าร่วมโครงการ “เราเที่ยวด้วยกัน” จาก www.agoda.com ในพื้นที่จังหวัดภูเก็ต พังงา และกระบี่

1.5.3 การสร้างชุดข้อมูล ในการวิจัยจะเป็นการสร้างชุดข้อมูลโดย ข้อมูลที่เก็บอยู่ในช่วงปี 2009 ถึง 2020

1.5.4 บทวิจารณ์ที่ใช้ในการวิจัย คัดเลือกบทวิจารณ์เฉพาะที่เป็นภาษาอังกฤษและภาษาไทย

บทที่ 2
แนวคิดทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีหลักการ

งานวิจัยนี้ได้รวบรวมข้อมูลรายชื่อโรงแรมจากเว็บไซต์ “เราเที่ยวด้วยกัน” โดยรวบรวมเฉพาะรายชื่อโรงแรมที่อยู่ในจังหวัดภูเก็ต กระบี่และพังงา รวบรวมข้อมูลบทวิจารณ์ออนไลน์ในเว็บไซต์ Agoda ตามรายชื่อที่รวบรวมจากเว็บไซต์ “เราเที่ยวด้วยกัน” โดยใช้ภาษา R ในการรวบรวมโดยใช้เทคนิคการขูดข้อมูล (Web Scraping) โดยมีรายชื่อโรงแรมจำนวน 427 โรงแรม และบทวิจารณ์จำนวน 246,532 บทวิจารย์ได้รวบรวมตั้งแต่ปี 2009-2020

2.1 Sentiment Analysis เป็นการวิเคราะห์ความรู้สึก วิเคราะห์อารมณ์จากข้อความซึ่งเป็นสาขาหนึ่งของงานด้าน NLP (Natural Language Processing) การใช้ Sentiment Analysis มาช่วยในการวิเคราะห์ความรู้สึกของคน เพื่อช่วยให้เข้าใจในพฤติกรรม และสามารถนำผลที่ได้รับมาปรับให้เข้ากับธุรกิจ เพื่อตอบสนองความต้องการของคนเหล่านั้นได้ตรงจุดและมีประสิทธิภาพ [15]

คนเราสามารถมีความรู้สึกจากหลาย ๆ สิ่งที่มากระตุ้นให้เกิดความรู้สึก เช่น ข้อความ สื่อต่าง ๆ ที่สามารถทำให้เรามีความรู้สึกต่าง ๆ ทิศทางความรู้สึก มีอยู่ 3 ประเภท ได้แก่

1.ทิศทางบวก : Positive

2.ทิศทางลบ : Negative

3.เป็นกลาง : Neutral

การวิเคราะห์ความรู้สึกช่วยให้เราเข้าใจในพฤติกรรมของผู้ที่ใช้บริการและสามารถนำปรับปรุงให้เข้ากับธุรกิจได้ เช่น ถ้าผู้ใช้บริการรู้สึกประทับใจในการเข้าพักโรงแรม แสดงว่า การบริการดีเยี่ยม แต่ถ้าผู้เข้าพักรู้สึกแย่ แสดงว่าการบริการไม่ได้มาตรฐาน จึงจำเป็นต้องนำไปปรับปรุงเพื่อให้การบริการดีขึ้น ซึ่งขั้นตอนในการทำ Sentiment Analysis มีขั้นตอนคร่าว ๆ 3 ขั้นตอน ได้แก่

1. Data Collection เป็นการรวบรวมข้อมูลความคิดเห็นต่าง ๆ จากแหล่งข้อมูลที่สนใจ เช่น วิจัยเล่มนี้ เกี่ยวกับบทวิจารณ์โรงแรม ผู้วิจัยได้รวบรวมความคิดเห็นจากเว็บไซต์ Agoda ข้อมูลที่เก็บอยู่ในช่วงปี 2009 ถึง 2020

2. Data Pre-Processing เป็นการเตรียมข้อมูลก่อนนำไปวิเคราะห์ เช่น การทำความสะอาดข้อมูลให้พร้อมนำไปใช้ โดยการจัดการเตรียมข้อมูลให้มีความเหมาะสมต่อการวิเคราะห์ มีการคัดเลือกเฉพาะข้อคิดเห็นที่มีความสมบูรณ์ในทุกคอลัมน์ และการจัดการข้อมูลที่ซ้ำออกไป

3. Sentiment Analysis นำข้อมูลที่ผ่านการ pre-process แล้วมาทำการวิเคราะห์ความรู้สึกอารมณ์จากประโยคและคำที่ได้จากการเตรียมข้อมูล เพื่อจำแนกข้อความในเชิงบวกหรือเชิงลบ (Positive / Negative)

จากการค้นคว้าข้อมูล พบว่ามีหลายงานวิจัยที่ให้ความสำคัญต่อการ วิเคราะห์ความรู้สึก กับข้อมูลที่มีความหลากหลาย ยกตัวอย่างดังนี้

K.Arun, K.Srinagesh, และ M.Ramesh (2017) ได้ศึกษาเกี่ยวกับ *"การวิเคราะห์ความรู้สึกในทวิตเตอร์จากการทวีตเกี่ยวกับการเลิกใช้เงินตราของอินเดียโดยใช้ภาษา R"* พบว่า มีการเลิกใช้ค่าเงินตรา (Demonetization) ทำให้มีชาวอินเดียได้ออกมาทวีตเกี่ยวกับความรู้สึกและแบ่งปันความคิดเห็นบนทวิตเตอร์เป็นจำนวนมาก โดยแยกได้ทั้งเชิงบวกและเชิงลบ โดยการวิเคราะห์ความรู้สึกโดยใช้ R Language ข้อมูลที่นำมาวิเคราะห์ได้จากการใช้ Twitter API หลังจากได้ข้อมูลจึงได้ทำการทำความสะอาดข้อมูล (data cleaning) และลบคำหยุด (stop words) ตามด้วยการจำแนกทวีตเป็นเชิงบวกและเชิงลบ ขั้นตอนสุดท้าย คือ การสร้าง word cloud เพื่อเปรียบเทียบเชิงบวกและเชิงลบ และความคิดเห็นของประชาชนในปัจจุบัน สรุปผลได้ว่า ความคิดเห็นเชิงลบคิดเป็นเปอร์เซ็นจำนวน 38% และความคิดเห็นเชิงบวก 23% และความคิดเห็นเป็นกลาง 39% [16]

Maurizio Naldi (2019) ได้ศึกษาเกี่ยวกับ *"บทวิจารณ์ของวิธีคำนวณความรู้สึกด้วยแพ็คเกจ R"* พบว่า ได้ทำการวิเคราะห์แพ็คเกจ จำนวน 4 แพ็คเกจ ได้แก่ syuzhet, Rsentiment, SentimentR, และ Sentiment Analysis ภาษาที่เลือกใช้สำหรับการวิเคราะห์ คือ Python Language และ R Language ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลที่ปรากฏบนเว็บไซต์ Amazon หรือ Trip Advisor หลังจากการวิเคราะห์ทั้ง 4 แพ็คเกจ ผลสรุป คือ SentimentR ที่ถูกพัฒนาโดย Tyler Rinker เป็นแพคเกจที่ได้รับเลือกให้เป็นแพ็คเกจที่ได้รับเลือกใช้ เพราะเป็นแพ็คเกจที่สามารถนำไปใช้ได้หลายบริบท [17]

Indy Wijngaards, Martijn Burger, และ Job van Exel (2021) ได้ศึกษาเกี่ยวกับ *"เปิดกล่องคำถามเกี่ยวกับความพึงพอใจในการทำงานแบบกึ่งเปิดเชิงปริมาณและเชิงปริมาณผ่านการวิเคราะห์ความรู้สึกโดยใช้คอมพิวเตอร์"* พบว่า ผู้วิจัยได้ยกตัวอย่างพนักงานจำนวน 395 คน ในการ

ทำแบบสอบถามเกี่ยวกับความพึงพอใจในการทำงาน ในการวิเคราะห์ใช้ 2 แพ็คเกจ ได้แก่ SentimentR และ LIWC2015 โดยวัดจากคำถามแบบปิดและแบบกึ่งเปิด ผลลัพธ์พบว่าคำถามแบบกึ่งเปิดมีความถูกต้องและมีการแยกแยะเป็นอย่างดี ทำให้การวิเคราะห์ความรู้สึกโดยใช้คอมพิวเตอร์ในการเข้ามาช่วยให้มีความง่ายขึ้น [18]

Sumera Mehboob and others (2020) ได้ศึกษาเกี่ยวกับ *"การจำแนกอารมณ์ตามความรู้สึกทวีตของผู้ที่มีชื่อเสียงโดยใช้ภาษา R"* พบว่า วิจัยนี้มุ่งเน้นในการศึกษาด้านอารมณ์ในทวิตเตอร์ ผู้วิจัยจึงทำการใช้ R Language ในการจำแนกอารมณ์ และใช้เทคนิค twitter scraper ในการรวบรวมข้อมูลจากบัญชีทวิตเตอร์ของดนดัง โดยใช้ R Language ในการค้นหาข้อมูลเชิงลึกเกี่ยวกับบุคลิกภาพ เชิงบวกและเชิงลบ อิโมติคอน ผลลัพธ์พบว่า ถ้าในทวิตเตอร์ มีผู้ใช้ทวีตในด้านเชิงลบมากกว่าด้านเชิงบวก แสดงให้เห็นว่าบุคคลนั้น อาจจะมีความเศร้า โกรธ หวาดกลัว หรือวิตกกังวล แต่ถ้ามีทวีตในด้านเชิงบวกมากกว่า แสดงว่าผู้ทวีตนั้นมีความสุข ดังนั้นวิจัยฉบับนี้เป็นวิจัยที่มีบทบาทในด้านการระบุอารมณ์ของผู้ใช้ทวิตเตอร์ [19]

Vaibhav Singh, Aayushi Mahajan และ Deepanshi Chaudhary (2020) ได้ศึกษาเกี่ยวกับ *"การวิเคราะห์ความรู้สึกของบทวิจารณ์โรงแรมจาก TripAdvisor"* พบว่า มีการรวบรวมความคิดเห็นจากเว็บไซต์ TripAdvisor ซึ่งความคิดเห็นแบ่งเป็น 3 ประเภท ได้แก่ เชิงบวก เชิงลบ และเป็นกลาง ความคิดเห็นเหล่านี้มีความสำคัญต่อการบ่งบอกถึงคุณภาพของโรงแรม มีการจัดหมวดหมู่โดยใช้ Naïve Bayes Classifier ซึ่งเป็นไลบรารีของ Text Blob เพื่อคำนวณจิตวิสัยและความตรงกันข้าม ในการแยกความถี่ของคำจากความคิดเห็น โดยใช้แนวทาง Term Frequency - Inverse Document Frequency (TFIDF) [20]

2.2 Association Rules คือ เทคนิคการวิเคราะห์เหมืองข้อมูลที่ใช้หาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายในกลุ่มข้อมูลที่มีขนาดใหญ่ Associations Rules เป็นที่นิยมในการหาความสัมพันธ์เกี่ยวกับธุรกิจในด้านการค้าปลีก เช่น ซุปเปอร์มาเก็ต หรือร้านสะดวกซื้อ เป็นต้น เหตุผลที่กฎความสัมพันธ์เป็นที่นิยมในหมู่ธุรกิจร้านค้าปลีก คือ เพื่อศึกษาพฤติกรรมในการซื้อสินค้าของลูกค้าและหาความสัมพันธ์เกี่ยวกับสินค้าที่ลูกค้าซื้อ เมื่อได้ผลลัพธ์ก็จะสามารถจัดวางสินค้าได้ตรงตามความต้องการของลูกค้าที่ต้องการซื้อสินค้าและมีความสะดวกในการหยิบจ่ายสินค้า ในการหาความสัมพันธ์นั้นมีอยู่หลายวิธี แต่สำหรับบทวิจัยนี้วิธีที่สนใจ คือ Apriori Algorithm และ FP-Growth Algorithm

Apriori เป็นขั้นตอนหนึ่งของ Association Rules และเป็นอัลกอริทึมพื้นฐานที่นิยมในการหา frequent itemset และเป็นขั้นตอนที่เป็นรู้จักกันอย่างแพร่หลาย โดยอัลกอริทึมนี้ถูกนำเสนอโดย R. Agrawal และ R. Srikant ในปี ค.ศ. 1994 ซึ่งในการทำ Apriori จะมีอยู่ 2 ขั้นตอนใหญ่ ๆ ได้แก่

1. การหา frequent itemset คือ การหารูปแบบของข้อมูลที่เกิดร่วมขึ้นกันบ่อย ๆ ในฐานข้อมูล หรือมากกว่าค่า minimum support ที่ผู้ใช้กำหนด ในขั้นตอนนี้สามารถแบ่งได้อีก 2 ขั้นตอนย่อย ได้แก่

1.1 การสร้างรูปแบบของ itemset ที่มีค่ามากกว่า minimum support มาทำการสร้างรูปแบบของ itemset ที่มีขนาดยาวมากขึ้นทีละหนึ่งขั้นไปเรื่อย ๆ

1.2 การนับค่า support (count) หลังจากที่สร้างรูปแบบของ itemset ได้แล้ว ขั้นตอนถัดมาคือการ คำนวณค่า support ที่เกิดขึ้น โดยที่ support คือจำนวนเปอร์เซ็นต์ที่พบ itemset ในฐานข้อมูล

2. การสร้าง association rule หลังจากขั้นตอนการหา frequent itemset ได้แล้วจะนำรูปแบบที่หาได้มาสร้างเป็นกฎความสัมพันธ์

<u>ยกตัวอย่างเช่น</u>

Apple > Cereal หมายความว่า เมื่อลูกค้าซื้อ Apple แล้วลูกค้าจะซื้อ Cereal ร่วมไปด้วย

FP-Growth เป็นขั้นตอนที่ปรับปรุงมาจาก Apriori ซึ่งกำจัดข้อเสียของวิธีแบบ Apriori โดยจะทำงานได้เร็วกว่าและไม่มีการดึงข้อมูลหลายรอบ และนิยมใช้กันอย่างแพร่หลายสำหรับการค้นหาข้อมูลที่ปรากฏร่วมกันบ่อย ๆ ซึ่งใช้เป็นกระบวนการในการวิเคราะห์ที่สามารถค้นหารูปแบบหรือการเชื่อมโยงจากชุดข้อมูลที่ปรากฏร่วมกันบ่อย โดยมีขั้นตอนหลัก 2 ขั้นตอนคือ

1.สร้าง Compact data structure หรือ FP-tree

1.1 นำรูปแบบของ item แต่ละชิ้นมาหาค่า support

1.2 นำข้อมูลดังกล่าวมาเรียงตามค่า support จากมากไปหาน้อย

2.สร้างรูปแบบของ item ที่เกิดขึ้นร่วมกันบ่อย ๆ (frequent itemset) จาก FP-tree และกำหนดค่า minimum support

2.1 ทำการตัดรายการ item ที่มีค่า support ต่ำกว่าเกณฑ์ที่กำหนด

2.2 หาค่า support ของ item ต่าง ๆ ในรูปแบบ itemset ทั้งแบบ 1 item ไปจนถึง 4 item

จากการค้นคว้าข้อมูล พบว่ามีหลายงานวิจัยที่ให้ความสำคัญต่อการวิเคราะห์ความสัมพันธ์ของข้อมูล (Association Rules) มีดังนี้

สัญญา พันธุ์แดง (2020) ได้ศึกษาเกี่ยวกับ *"การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับการพยากรณ์การศึกษาต่อนักศึกษาใหม่ระดับปริญญาตรี ในมหาวิทยาลัยราชภัฏเชียงใหม่"* พบว่า มีการใช้เทคนิคการทำเหมืองข้อมูลประชากรและกลุ่มตัวอย่างเป็นข้อมูลผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีในหมาวิทยาราชภัฏเชียงใหม่ เครื่องมือที่ใช้ในการวิจัยเป็นหลัก มีทั้งหมด 2 องค์ประกอบ ได้แก่ด้านฮาร์ดแวร์และซอฟต์แวร์ โดยใช้โปรแกรม Rapid Miner Studio วิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้เทคนิคการทำเหมืองข้อมูลแบบจำลองการหาความสัมพันธ์ด้วยวิธีการ Apriori เพื่อใช้ในการค้นหากฎความสัมพันธ์ของพฤติกรรมการเลือกสมัครของนักศึกษา จากข้อมูลกฎความสัมพันธ์ที่มีได้ทั้งหมด 340 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์และพิจารณาโดย นำประเภทการสมัครมาแบ่งกลุ่มในการพิจารณาและพิจารณ์การสอดคล้องของผลสรุป โดยจะพิจารณาจากค่าสนับสนุน (Support) = 0.01 ค่าความเชื่อมั่น (Confidence) = 95 % และค่าความสอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 89 กฎ [21]

สุภาพรรณ คงมณีพรรณ (2019) ได้ศึกษาเกี่ยวกับ *"การหากฎความสัมพันธ์จากฐานข้อมูลการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ มายเฮลท์ ของลูกค้าโดยใช้อัลกอริทึม เอฟพีโกรท และการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้ามายเฮลท์ โดยเทคนิคอาร์เอฟเอ็มด้วยโปรแกรม เรพพิดไมเนอร์ : กรณีศึกษาของร้านยาเชนแห่งหนึ่ง"* พบว่า กฎความสัมพันธ์ของการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ MYHEALTH จำนวน 7 กฎ ซึ่งความสัมพันธ์เหล่านี้สามารถช่วยในการจัดทำแผนการตลาด เพื่อตอบสนองความต้องการของลูกค้าและส่งเสริมให้เกิดการซื้ออย่างต่อเนื่องของลูกค้าได้ และยังทำให้บริษัทสามารถสร้างแผนจัดตำแหน่งสินค้า เพื่อเพิ่มยอดขายให้กับธุรกิจได้และรวมถึงการแบ่งกลุ่มลูกค้าตามลักษณะของลูกค้าเพื่อส่งเสริมพฤติกรรมการซื้อของลูกค้าได้ [22]

ธนะพร ฮองกุล และคนอื่น ๆ (2017) ได้ศึกษาเกี่ยวกับ *"การหาความสัมพันธ์ของราคาน้ำมันดิบดับบลิวทีไอ (WTI Crude) ในตลาดไนเม็กซ์ (NYMEX) ระหว่าง ราคาหุ้นกลุ่มพลังงานที่เกี่ยวข้องกับธุรกิจบริการสถานีน้ำมัน ในตลาดหลักทรัพย์แห่งประเทศไทย ด้วยเทคนิคการทำเหมือง*

ข้อมูล” พบว่า มีการหาค่าความสัมพันธ์ว่าราคาน้ำมันดิบมีการเปลี่ยนแปลงราคาปรับเพิ่มขึ้น จะมีความสัมพันธ์กับราคาหุ้นในกลุ่มพลังงานที่เกี่ยวข้องกับธุรกิจบริการสถานีน้ำมันปรับเพิ่มขึ้นหรือไม่ โดยใช้เทคนิค FP-Growth Algorithm ในการค้นหากฎความสัมพันธ์ของข้อมูล ผลจากการวิจัยพบว่าสามารถสร้างความสัมพันธ์ได้ทั้งหมด 31 กฎ กฎความสัมพันธ์ที่ได้รับการวิจัยสูงสุด คือ ถ้าราคาน้ำมันดิบเมื่อปิดการซื้อขายแล้วมีการปรับราคาเพิ่มขึ้น ก็แสดงให้เห็นว่า เมื่อปิดการซื้อขายในตลาดหลักทรัพย์แห่งประเทศไทยแล้ว ราคาหุ้น PTG มีการปรับราคาเพิ่มขึ้นด้วย โดยที่มีค่าความเชื่อมั่นสูงสุดที่ 0.672 ค่าสนับสนุนที่ 0.35 และมีค่าสหสัมพันธ์ที่ 1.05 [23]

จากผลการศึกษางานวิจัยที่เกี่ยวข้องกับการหาความสัมพันธ์ของข้อมูลสองชุดขึ้นไปภายในกลุ่มข้อมูลที่มีขนาดใหญ่ ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาความสัมพันธ์หลากหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและมีการใช้อย่างแพร่หลายคือ Apriori และ FP-Growth วิธีเหล่านี้เป็นเทคนิคการวิเคราะห์เหมืองข้อมูล โดยการค้นหากฎความสัมพันธ์ของข้อมูลมีการวัดความแม่นยำของกฎความสัมพันธ์ด้วยค่าสนับสนุน (Support) และค่าความมั่นใจ (Confidence) เป็นเกณฑ์ในการคัดเลือกรูปแบบการค้นหากฎความสัมพันธ์ [24], [25] ของข้อมูลในฐานข้อมูล

การประเมินประสิทธิภาพของอัลกอริทึม Apriori และ FP-Growth

การประเมินประสิทธิภาพของระบบของอัลกอริทึมของ Apriori และ FP-Growth การประเมินประสิทธิภาพของความแม่นยำในการสร้างกฎความสัมพันธ์ด้วยอัลกอริทึมอพริโอริ (Apriori) จะมีอยู่ 2 ขั้นตอน คือ

1. การหาเซตไอเทม (itemset) เป็นการหารูปแบบของข้อมูลที่เกิดขึ้นบ่อย ๆ หรือมากกว่าค่า Minimum Support ที่กำหนด ในขั้นตอนนี้แบ่งได้ 2 ข้อย่อยคือ

    1.1. การสร้างรูปแบบของเซตไอเทม (itemset) โดยการ join โดยที่มีค่ามากกว่า Minimum Support ที่กำหนดมาทำการสร้างรูปแบบของเซตไอเท็ม (itemset) ที่มีขนาดยาวขึ้นทีละขั้น

    1.2. การนับค่า Support หลังจากที่สร้างเซตไอเทม (itemset) จากนั้นทำการคำนวณค่า Support ที่เกิดขึ้น โดยที่ค่า Support เท่ากับจำนวนเปอร์เซ็นต์ที่เจอเซตไอเทม (itemset)

2. สร้างกฎความสัมพันธ์ การค้นหากฎความสัมพันธ์มีเกณฑ์ในการวัด 3 แบบ ได้แก่

2.1. การหาค่าสนับสนุน (Support) แสดงให้เห็นถึงเปอร์เซ็นต์ของข้อมูลทั้งหมดที่เป็นไปได้ดังสมการ (1)

$$\text{support(LHS, RHS)} = \frac{\text{frequent LHS,RHS}}{\text{number of data}} \quad (1)$$

2.2. ค่าความมั่นใจ (Confidence) แสดงให้เห็นเห็นถึงความน่าเชื่อถือของกฎค่าความมั่นใจ (Confidence) เมื่อรูปแบบเซตไอเทม ด้านซ้ายของกฎความสัมพันธ์ (Left Hand Side: LHS) เกิดขึ้นจะมีโอกาสเกิดรูปแบบเซตไอเทม (itemset) ด้านขวาของกฎความสัมพันธ์ (Right Hand Side: RHS) มากน้อยจะมีค่าอยู่ระหว่าง 0 – 1 ถ้ามีค่าใกล้เคียง 1 หมายความว่าความเชื่อมั่นความสัมพันธ์มาก ดังสมการ (2)

$$\text{confidence(LHS => RHS)} = \frac{\text{support(LHS,RHS)}}{\text{support(LHS)}} \quad (2)$$

3. ค่าสหสัมพันธ์ (Lift) แสดงให้เห็นการเกิดรูปแบบเซตไอเทมด้านซ้าย (LHS) และขวา (RHS) ของกฎความสัมพันธ์ ถ้าค่าสหสัมพันธ์มีค่าเท่ากับ 1 แสดงให้เห็นว่ารูปแบบด้านซ้าย (LHS) และขวา (RHS) ไม่ขึ้นต่อกัน (Independent) แต่ถ้ามากกว่า 1 แสดงให้เห็นว่ากฎทั้งสองมีความสัมพันธ์กันมาก ดังสมการ (3)

$$\text{lift(LHS => RHS)} = \frac{\text{support(LHS, RHS)}}{\text{support(LHS)} \times \text{support(RHS)}} \quad (3)$$

โดยจะพิจารณาเฉพาะความสัมพันธ์ที่มีค่าสนับสนุนและค่าความมั่นใจสูงกว่าค่าสนับสนุนต่ำสุด (Minimum Support) และค่าความมั่นใจต่ำสุด (Minimum confidence)

ในส่วนของการประเมินประสิทธิภาพของความแม่นยำในการสร้างกฎความสัมพันธ์ด้วย FP-Growth มีอยู่ 2 ขั้นตอน คือ

1. สร้าง compact data structure ที่เรียกว่า FP-tree การสร้าง FP-Tree โดยอ่านข้อมูลจากฐานข้อมูลเพียง 2 ครั้งเท่านั้น จากนั้นทำการหาค่า Support ของรูปแบบไอเทม ที่มีความยาว 1 ไอเทม ดังสมการ (4)

$$\text{support x} = \frac{\text{frequent X}}{\text{number of data X}} \quad (4)$$

2. สร้างรูปแบบของความถี่ไอเทมที่เกิดขึ้นบ่อย (frequent itemset) จาก FP-Tree โดยกำหนดค่า Minimum Support = 1% (0.01) เนื่องจากข้อมูลมีการกระจายกลุ่มข้อมูลจำนวนมากและมีปริมาณที่สูงเกินไปทำให้การตั้งค่า Minimum Support ที่มีค่ามากกว่า 0.01 ทำให้ไม่สามารถหากลุ่ม Itemset และกฎความสัมพันธ์ที่ดีที่สุดได้ โดยใช้สมการ (1) (2)

บทที่ 3

กรอบการดำเนินงานวิจัย

3.1 การนิยามปัญหา

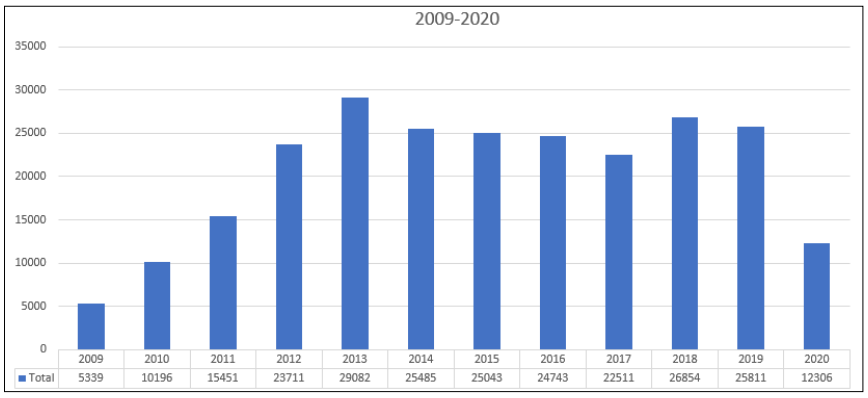
งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาข้อมูลบทวิจารณ์ออนไลน์ และนำเสนอการวิเคราะห์แนวโน้มและความสัมพันธ์ของนักท่องเที่ยว ซึ่งจากการศึกษาข้อมูลบทวิจารณ์ออนไลน์พบว่าข้อมูลสามารถบ่งบอกอะไรได้หลายอย่างที่เคยเกิดขึ้นในอดีต เช่น ความถี่ของนักท่องเที่ยวที่เข้าพัก แยกตามทวีป หรือโซน รวมทั้งรูปแบบการเดินทางของนักท่องเที่ยว สิ่งอำนวยความสะดวกที่ให้บริการ ช่วงเวลา จำนวนคืนเข้าพัก เป็นต้น ซึ่งข้อมูลเหล่านี้สามารถนำมาวิเคราะห์เพื่อหาแนวโน้มและความสัมพันธ์เพื่อนำผลลัพธ์ไปใช้ในการวางแผนเพื่อเป็นแนวทางในการพัฒนาปรับปรุงบริการให้ดียิ่งขึ้น

จำนวนข้อมูลบทวิจารณ์ออนไลน์และจำนวนรายชื่อโรงแรมจากเว็บไซต์ “เราเที่ยวด้วยกัน” โดยการรวบรวมข้อมูลบทวิจารณ์ออนไลน์ในเว็บไซต์ Agoda โดยมีจำนวนโรงแรม 427 โรงแรม และบทวิจารณ์จำนวน 246,532 บทวิจารณ์ได้รวบรวมตั้งแต่ปี 2009-2020 โดยแบ่งแยกตามตารางที่ 1 และตารางที่ 2 ดังนี้

ตารางที่ 1 แสดงจำนวนบทวิจารณ์และจำนวนโรงแรมของ 3 จังหวัด

จังหวัด	จำนวนบทวิจารณ์	จำนวนโรงแรม
กระบี่	97,212	158
พังงา	11,847	41
ภูเก็ต	137,473	228
รวม	246,532	427

จากตารางที่ 1 มีจำนวนบทวิจารณ์ทั้งสิ้น 246,532 บทวิจารณ์ และมีจำนวนโรงแรมทั้งหมด 427 โรงแรม โดยมีรายละเอียดแบ่งเป็น กระบี่ 97,212 บทวิจารณ์ มาจาก 158 โรงแรม ในส่วนของพังงาน 11,847 บทวิจารณ์ มาจาก 41 โรงแรม และในส่วนของภูเก็ต 137,473 มาจาก 228 โรงแรม



รูปที่ 1 : การแบ่งแยกจำนวนบทวิจารณ์ในแต่ละปี ตั้งแต่ปี 2009-2020

จากรูปที่ 1 แสดงให้เห็นว่าในปี 2009 เป็นปีที่มีบทวิจารณ์จำนวนน้อยที่สุด คือ 5,339 บทวิจารณ์ และในปี 2013 เป็นปีที่มีบทวิจารณ์มากที่สุด คือ 29,082 บทวิจารณ์ และมีแนวโน้มลดลงเฉลี่ย 22.6% เมื่อเปรียบเทียบระหว่างปี 2013 และ 2017 ในปี 2018 เริ่มมีแน้วโน้มเพิ่มขึ้นและลดลงเรื่อย ๆ ในปีถัดมา และลดลงอย่างเห็นได้ชัดในปี 2020 ประมาณ 52% ของนักท่องเที่ยวในปี 2019 เนื่องจากผลกระทบจากการระบาดของ COVID-19

ในส่วนของบทวิจารณ์ออนไลน์ที่ทำการรวบรวมประกอบด้วยหัวข้อต่าง ๆ ที่สนใจแสดงดังรูปที่ 2

รูปที่ 2 : ตัวอย่างรายละเอียดบทวิจารณ์ออนไลน์ที่สนใจ

จากรูปที่ 2 แสดงรายละเอียดบทวิจารณ์ออนไลน์ที่สนใจดังนี้

1.หัวข้อรีวิว

2.รายละเอียดในรีวิว

3.วัน/เดือน/ปี ที่รีวิว

4.คะแนนรีวิว

5.ประเทศผู้รีวิว

6.ชนิดกลุ่มนักท่องเที่ยว 6 ชนิด

6.1 Business traveler

6.2 Group

6.3 Couple

6.4 Family with teens

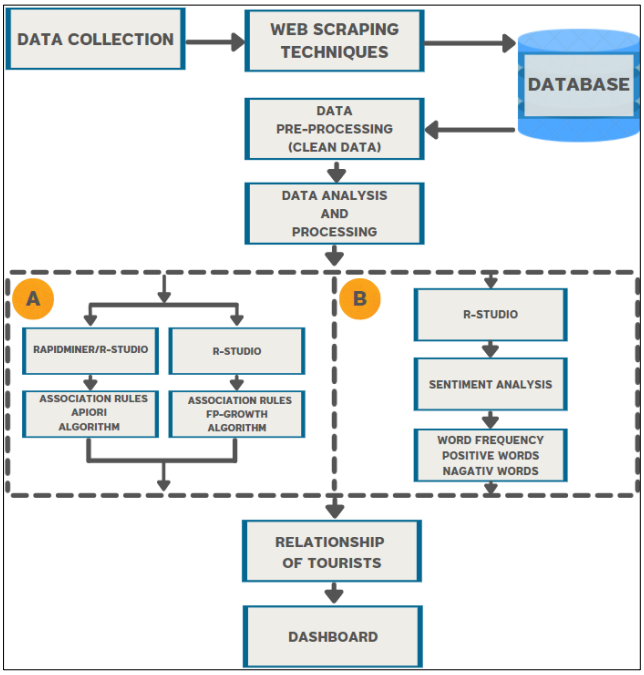
6.5 Family with young children

6.6 Solo traveler

7.จำนวนวันที่เข้าพัก

3.2 กรอบการวิเคราะห์บทวิจารณ์ออนไลน์

ในงานวิจัยนี้ กระบวนการในการวิเคราะห์บทวิจารณ์ออนไลน์ โดยมีรูปแบบการทำงานของระบบแสดงในรูปที่ 3



รูปที่ 3 : System Architecture

จากรูปที่ 3 จากการศึกษาทฤษฎีและวรรณกรรมที่เกี่ยวข้องกับการวิเคราะห์บทวิจารณ์ออนไลน์ ทำให้ผู้วิจัยทราบถึงแนวทางและได้วางแผนการบวนการวิเคราะห์ดังภาพออกมาในรูปแบบ System Architecture เพื่อกำหนดเป็นกรอบการดำเนินงานวิจัยตามวัตถุประสงค์ของงานวิจัยนี้ ซึ่งกระบวนการมีดังนี้

1. รวบรวมรายชื่อโรงแรมที่เข้าร่วมโครงการ “เราเที่ยวด้วยกัน”

2. รวบรวมรีวิวจากเว็บไซต์ Agoda

3. กระบวนการเตรียมข้อมูล

4. กระบวนการวิเคราะห์แบ่งเป็น 2 กระบวนการหลักดังนี้

4.1 การวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูล (Association Rules)

4.2 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยค (Sentiment Analysis)

5. ประเมินผลการวิเคราะห์

งานวิจัยนี้ได้นำเสนอกระบวนการวิเคราะห์บทวิจารณ์ออนไลน์โดยแบ่งกระบวนการวิเคราะห์แบ่งเป็น 2 กระบวนการหลักรายละเอียดมีดังนี้

1.การวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูล (Association Rules)

1.1 การวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูล (Association Rules) เป็นการวิเคราะห์เพื่อดูพฤติกรรมการเข้าพักโรงแรมของนักท่องเที่ยวในมุมมองกว้าง ซึ่งจะมีการเปรียบเทียบโดยใช้อัลกอริทึม Apriori และ FP-Growth

1.2 การวิเคราะห์เปรียบเทียบพฤติกรรมนักท่องเที่ยวชาวไทยก่อนเกิดสถานการณ์ COVID-19 และ ขณะเกิดสถานการณ์ COVID-19

2. การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยค (Sentiment Analysis)

2.1 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยค (Sentiment Analysis) เพื่อบอกความรู้สึกของผู้คนที่มีต่อบางสิ่งบางอย่าง และแบ่งได้เป็น ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative)

2.2 การจัดการบทวิจารณ์และคำศัพท์ต่างๆ ที่ได้จากการวิเคราะห์อารมณ์และความรู้สึก และนำคำศัพท์ที่ได้ทำการจัดให้เป็นหมวดหมู่เพื่อเปรียบเทียบค่าความถี่ของแนวโน้มในแต่ละหมวดหมู่

กระบวนการวิเคราะห์บทวิจารณ์ออนไลน์ มีรายละเอียดดังนี้

3.2.1 รวบรวมรายชื่อโรงแรมที่เข้าร่วมโครงการ “เราเที่ยวด้วยกัน”

งานวิจัยฉบับนี้มีการรวบรวมข้อมูลจากเว็บไซต์ Agoda โดยรวบรวมเฉพาะโรงแรมที่เข้าร่วมโครงการ “เราเที่ยวด้วยกัน” จากเว็บไซต์ เราเที่ยวด้วยกัน โดยรวบรวมเฉพาะ 3 จังหวัดภาคใต้ ได้แก่ ภูเก็ต กระบี่ และพังงา เพื่อเป็นตัวอย่างข้อมูล โดยใช้เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping Techniques) โดยมีจำนวนรายชื่อโรงแรมที่รวบรวมทั้งหมดจำนวน 427 โรงแรม

3.2.2 รวบรวมรีวิวจากเว็บไซต์ Agoda

รวบรวมบทวิจารณ์ออนไลน์จากเว็บไซต์ Agoda โดยรวบรวมตามรายชื่อโรงแรมในข้อ 3.2.1 โดยใช้เทคนิคการดึงข้อมูลจากเว็บไซต์ (Web Scraping Techniques) ในภาษา R (R Programming) มีจำนวนข้อมูลทั้งสิ้น 246,532 บทวิจารณ์ โดยมีการรวบรวมข้อมูลจำนวน 11 ปีย้อนหลังตั้งแต่ ค.ศ. 2009 – 2020 ข้อมูลที่ทำการศึกษานั้นเป็นไฟล์ขนาด 210 MB โดยมีตัวอย่างรายละเอียดแสดงข้อมูลดังตารางที่ 2

ตารางที่ 2 ตัวอย่างรายละเอียดตัวแปรในชุดข้อมูล

ตัวแปร	ตัวอย่างความหมาย	ลักษณะข้อมูล
Hotel_id	ไอดีโรงแรม	Numeric
Hotel_name	ชื่อโรงแรม	Text
Hotel_star	จำนวนดาวโรงแรม	Numeric
Review_title	หัวข้อรีวิว	Text
Review_comment	รายละเอียดในรีวิว	Text
Reviewer_country	ประเทศผู้รีวิว	Text
Review_month	เดือนที่รีวิว	Text
Review_day	วันที่รีวิว	Numeric
Review_year	ปีที่รีวิว	Numeric
Review_score	คะแนนรีวิว	Numeric
Group_name	ชนิดกลุ่มนักท่องเที่ยว	Text
Stay_detail	จำนวนวันที่เข้าพัก	Numeric
Hotel_location	จังหวัดโรงแรม	Text

3.2.3 กระบวนการเตรียมข้อมูล

กระบวนการเตรียมข้อมูล เป็นกระบวนการนำข้อมูลบทวิจารณ์ที่ได้รับ มาผ่านกระบวนการเตรียมข้อมูล โดยการทำความสะอาดข้อมูลให้พร้อมนำไปใช้ โดยการจัดการเตรียมข้อมูลให้มีความเหมาะสมต่อการวิเคราะห์ มีการคัดเลือกเฉพาะข้อคิดเห็นที่มีความสมบูรณ์ในทุกคอลัมน์ เริ่มต้นด้วยการจัดการข้อมูลที่ซ้ำออกไป คัดเลือกเฉพาะบทวิจารณ์ภาษาอังกฤษและภาษาไทย จากนั้นทำการแปลภาษาไทยให้เป็นภาษาอังกฤษทั้งหมดโดยใช้ google translate api จากผลการศึกษางานวิจัยที่เกี่ยวข้องกับเครื่องมือในการแปลภาษาพบว่า google translate มีประสิทธิภาพเหมาะสมที่สุดในการแปลเกี่ยวกับบทวิจารณ์ออนไลน์โรงแรมถึง 89.33% [26] โดยจำนวนบทวิจารณ์ภาษาไทยที่นำไปแปลทั้งหมดมีจำนวน 22,874 บทวิจารณ์ จากการตรวจสอบบทวิจารณ์ภาษาไทยพบว่าภาษาที่ใช้ในการรีวิวโดยส่วนใหญ่ คำที่ใช้จะเป็นคำทับศัพท์ภาษาอังกฤษเกี่ยวกับ คำศัพท์โรงแรมและภาษาโรงแรมที่นิยมใช้กัน เช่น คำว่า วอล์คอิน อัพเกรด เช็คอิน เช็คเอาท์ เซอร์วิส แท็กซี่ คอนเฟิร์ม บุคกิ้ง บิล และจากุชชี่ เป็นต้น หลังจากนั้นทำการแบ่งข้อความให้เหมาะสมตามรูปแบบข้อมูลที่ต้องการ และตัดข้อความให้เหลือเฉพาะส่วนที่สมบูรณ์ จากนั้นทำการเพิ่มคอลัมน์ทวีป (Reviewer_continent) โดยการจัดกลุ่มจากข้อมูลประเทศผู้รีวิว (Reviewer_country) แสดงตัวอย่างดังรูปที่ 4

Hotel_id	Hotel_name	Hotel_star	Review_title	Reviewer_country	Reviewer_continent
360	Palm Kaew Resort Krabi	0	Passable"	Thailand	Asia
280	K Guesthouse	1.5	Friendly and h	Thailand	Asia
280	K Guesthouse	1.5	Xxxxxx"	Thailand	Asia
280	K Guesthouse	1.5	Muffies Zimme	Thailand	Asia
280	K Guesthouse	1.5	ทำเลใจกลางเมื	Thailand	Asia
280	K Guesthouse	1.5	very nice gues	Thailand	Asia
281	J. Mansion	2	Amazing locati	Thailand	Asia
281	J. Mansion	2	กลางๆครับ"	Thailand	Asia
281	J. Mansion	2	ห้องถูกดี"	Thailand	Asia

รูปที่ 4 : ตัวอย่างผลลัพธ์ข้อมูลที่ได้จากการจัดกลุ่มข้อมูล

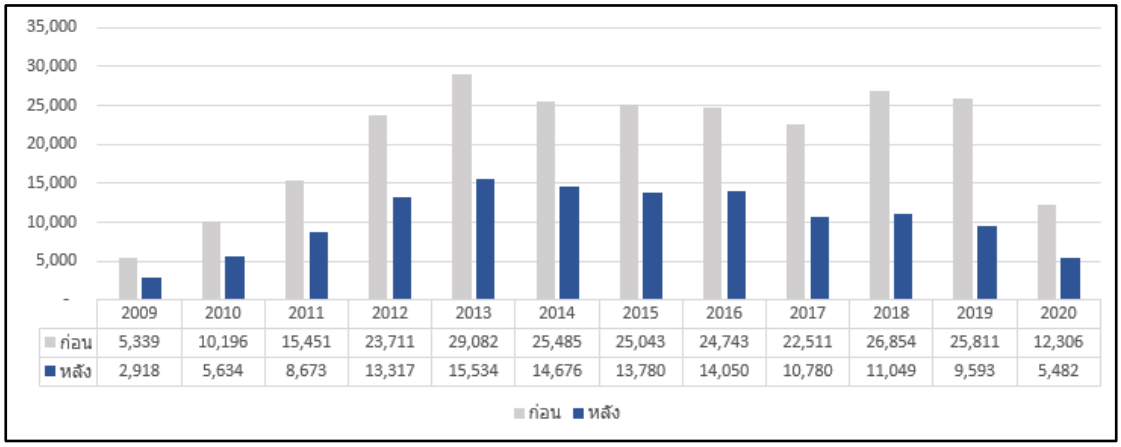
จากรูปที่ 4 แสดงให้เห็นถึงผลลัพธ์ที่ได้จากการจัดกลุ่มข้อมูลจากข้อมูลประเทศผู้รีวิวให้เห็นทวีปโดยใช้ภาษา R ในการจัดกลุ่มข้อมูล จากนั้นทำการเตรียมข้อมูลเพื่อแปลงข้อมูลสำหรับการวิเคราะห์เพื่อหาความสัมพันธ์โดยการแปลงข้อมูลใช้ภาษา R แสดงผลลัพธ์ดังตัวอย่างรูปที่ 5

Group_name
Group
Family with young children
Couple
Couple
Couple
Family with young children
Family with teens
Couple
Couple
Family with young children

id	Business traveler	Couple	Family with teens	Family with young children	Group	Solo traveler
1	0	0	0	0	1	0
2	0	0	0	1	0	0
3	0	1	0	0	0	0
4	0	1	0	0	0	0
5	0	1	0	0	0	0
6	0	0	0	1	0	0
7	0	0	1	0	0	0
8	0	1	0	0	0	0
9	0	1	0	0	0	0
10	0	0	0	1	0	0

รูปที่ 5 : การแปลงคอลัมน์ให้อยู่ในรูปแบบแถว

จากรูปที่ 5 แสดงให้เห็นถึงการแปลงคอลัมน์ให้อยู่ในรูปแบบแถวและให้ข้อมูลอยู่ในรูปแบบ 0 หรือ 1 โดยเลือกเฉพาะคอลัมน์ที่ต้องการหาความสัมพันธ์คอลัมน์ที่แปลงมีดังนี้ Hotel_star, Review_month, Review_year, Group_name, Stay_detail และ Reviewer_continent ซึ่งผลลัพธ์ที่ได้จะออกมาในรูปแบบไฟล์ CSV เพิ่มเตรียมเข้าสู่กระบวนการวิเคราะห์ในลำดับถัดไปโดยรายละเอียดข้อมูลที่ได้จากกระบวนการเตรียมข้อมูลแสดงดังรูป



	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
ก่อน	5,339	10,196	15,451	23,711	29,082	25,485	25,043	24,743	22,511	26,854	25,811	12,306
หลัง	2,918	5,634	8,673	13,317	15,534	14,676	13,780	14,050	10,780	11,049	9,593	5,482

รูปที่ 6 : กราฟแสดงจำนวนบทวิจารณ์ก่อนและหลังกระบวนการเตรียมข้อมูล

จากรูปที่ 6 แสดงให้เห็นถึงจำนวนข้อมูลที่ได้จากกระบวนการเตรียมข้อมูลก่อนนำเข้ากระบวนการและหลังนำเข้ากระบวนการโดยข้อมูลในปี 2009-2020 ทั้งหมดลดลงจากจำนวน 246,532 บทวิจารณ์ลดลงเหลือ 125,486 บทวิจารณ์โดยคิดเป็นเปอเซ็นเฉลี่ยอยู่ที่ 48.69 %

3.2.4 กระบวนการวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูลและการวัดประสิทธิภาพความเร็วในการประมวลผล

กระบวนการวิเคราะห์จะนำข้อมูลไปวิเคราะห์ในโปรแกรม Rapid Miner และ R Studio โดยใช้ Association Rule เพื่อหาความสัมพันธ์ของข้อมูลและการเปรียบเทียบประสิทธิภาพความเร็วในการประมวนผลโดยใช้อัลกอริทึม Apriori และ FP-Growth โดยกำหนดค่า Minimum Support เท่ากับ 1% (0.01) เนื่องจากข้อมูลมีการกระจายกลุ่มข้อมูลจำนวนมาก และมีปริมาณที่สูงเกินไปทำให้การตั้งค่า Minimum Support ที่มีค่ามากกว่า 0.01 ทำให้ไม่สามารถหากลุ่มเซตไอเทม (itemset) และกฎความสัมพันธ์ที่ดีที่สุดได้ และ ตั้งค่า Minimum Confidence เท่ากับ 50% (0.5)

ชุดข้อมูลที่ใช้ : จำนวนบทวิจารณ์ทั้งหมด ได้จากการเตรียมข้อมูลในข้อ 3.2.2 จำนวน 125,486 บทวิจารณ์ ปี 2009-2020

3.2.4.1 การแปลงชุดข้อมูลที่ใช้ในการทดลอง

จากข้อมูลในข้อ 3.2 ที่ได้กล่าวมาแล้วข้างต้นนั้น ทางผู้วิจัยได้จัดทำใน 2 รูปแบบเพื่อให้รองรับในการใช้อัลกอริทึมของ FP-Growth และ Apriori จากโปรแกรม Rapid Miner และ R-Studio คือ ชุดข้อมูล DATA1 จะเป็นการแปลงข้อมูลแบบ binominal ซึ่งคือการแปลงข้อมูลตัวเลขให้เป็น binominal หรือค่า true, false และ ชุดข้อมูล DATA2 จะเป็นการแปลงข้อมูลแบบ transaction ซึ่งเป็นการเขียนได้ในรูปของเซ็ตของ item ที่เป็นเหตุ ไปสู่เซ็ตของ item ที่เป็นผล โดยแต่ละ transaction จะมีหมายเลข transaction id ที่ไม่ซ้ำกัน ดังรูปที่ 7 และ 8 ตามลำดับ

Business.traveler	Couple	Family.with.teens	Family.with.young.children	Group	Solo.traveler
FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
FALSE	FALSE	FALSE	FALSE	TRUE	FALSE

รูปที่ 7 แสดงชุดข้อมูลในรูปแบบของ binominal (DATA1)

items	transactionID
{X.ปฟstar=4.5,month=November,year=2015,Gname=Couple,Stay=2,continent=Asia}	1
{X.ปฟstar=4.5,month=April,year=2016,Gname=Family with teens,Stay=4,continent=Oceania}	2
{X.ปฟstar=4.5,month=August,year=2016,Gname=Couple,Stay=2,continent=Asia}	3
{X.ปฟstar=4.5,month=February,year=2019,Gname=Family with young children,Stay=4,continent=Asia}	4
{X.ปฟstar=4.5,month=August,year=2013,Gname=Couple,Stay=5,continent=Europe}	5
{X.ปฟstar=4.5,month=January,year=2017,Gname=Family with teens,Stay=4,continent=Oceania}	6
{X.ปฟstar=4.5,month=September,year=2015,Gname=Couple,Stay=4,continent=Asia}	7
{X.ปฟstar=4.5,month=June,year=2015,Gname=Solo traveler,Stay=3,continent=Asia}	8
{X.ปฟstar=4.5,month=December,year=2012,Gname=Family with teens,Stay=5,continent=Asia}	9
{X.ปฟstar=4.5,month=May,year=2015,Gname=Family with young children,Stay=3,continent=Asia}	10

รูปที่ 8 แสดงชุดข้อมูลในรูปแบบของ transaction (DATA2)

โดยในการประมวลผลของชุดข้อมูลทั้ง 2 นี้ มีความเชื่อมโยงกับอัลกอริทึมและโปรแกรมที่ใช้ รายละเอียดดังตารางที่ 3

ตารางที่ 3 รายละเอียดความสัมพันธ์ระหว่างชุดข้อมูล

โปรแกรม	อัลกอริทึม	ชุด DATA1	ชุด DATA2
R-Studio	Apriori	P	P
R-Studio	FP-Growth	O	O
Rapid Miner	FP-Growth	P	O

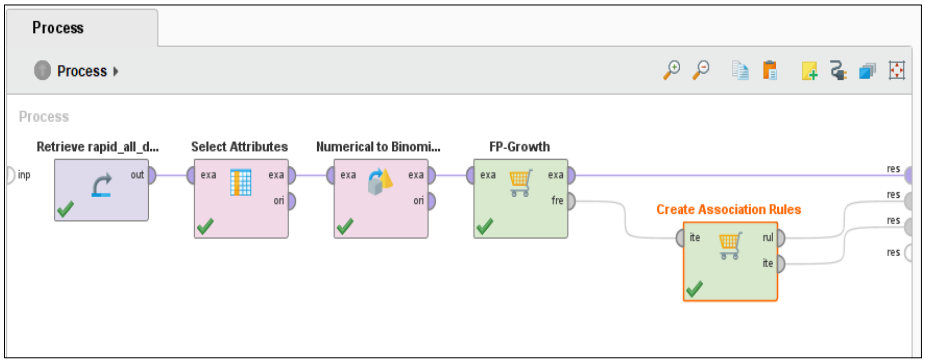
3.2.4.2 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA 1 ระหว่าง FP-Growth และ Apriori โดยการใช้ Rapid Miner และ R-Studio

ตารางที่ 4 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA 1

โปรแกรม	อัลกอริทึม	ชุด DATA1	ชุด DATA2
R-Studio	Apriori	P	P
R-Studio	FP-Growth		O
Rapid Miner	FP-Growth	P	O

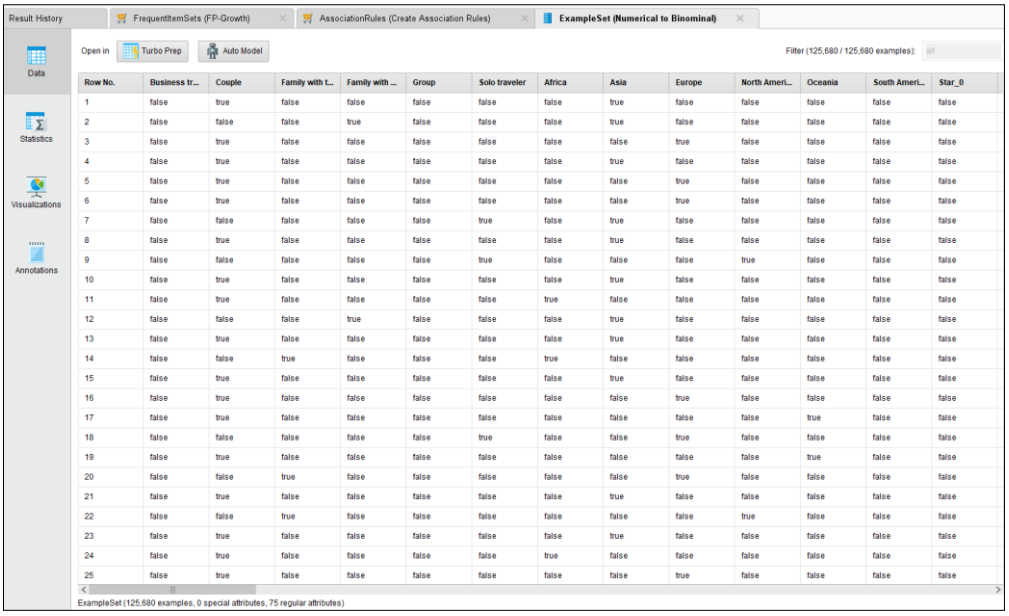
Rapid Miner (FP-Growth)

Process Model ที่ใช้ในโปรแกรม Rapid Miner แสดงดังตัวอย่างรูปที่ 9



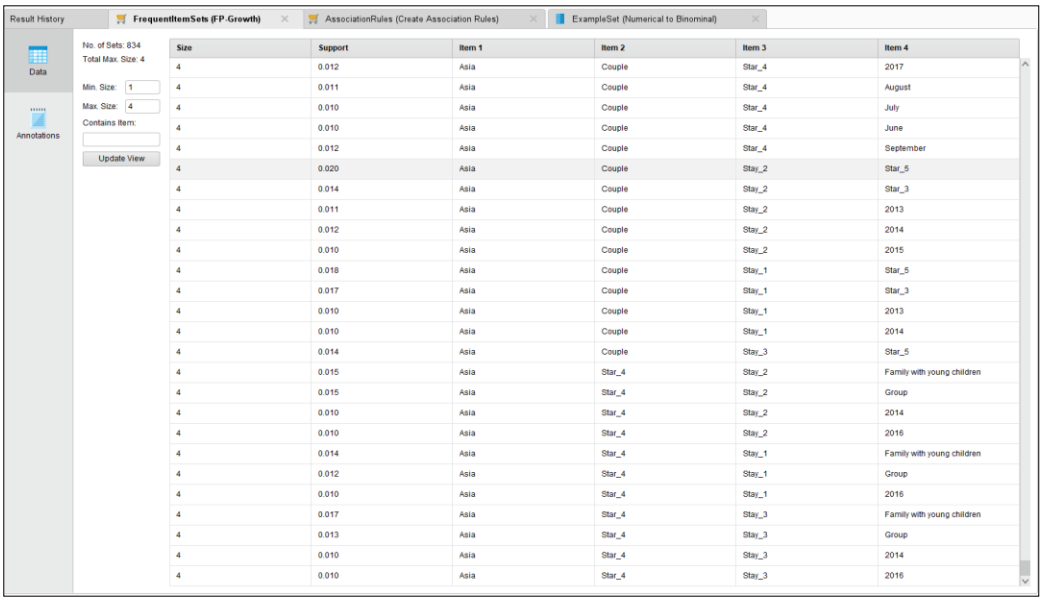
รูปที่ 9 : Process Model ที่ใช้ในการทำสร้าง Association Rule

จากรูปที่ 9 แสดงให้เห็นถึงกระบวนการสร้าง Operator ในการหาความสัมพันธ์ เริ่มต้นด้วยการนำไฟล์เข้า จากนั้นทำการใช้ Operator “Select Attribute” เพื่อเลือกเฉพาะ Attribute ที่ใช้สำหรับการวิเคราะห์ จากนั้นทำการเพิ่ม Operator “Numerical to Binominal” เพื่อแปลงข้อมูลตัวเลขจาก 0 หรือ 1 ให้เป็น True หรือ False จากนั้นทำการเพิ่ม Operator “FP-Growth” เพื่อนำข้อมูลเข้าสู่กระบวนการวิเคราะห์ด้วยอัลกอริทึม FP-Growth ขั้นตอนสุดท้ายทำการเพิ่ม Operator “Create Association Rule” เพื่อทำการสร้างกฎความสัมพันธ์จากข้อมูลที่ได้จากการทำ FP-Growth แสดงผลลัพธ์ดังตัวอย่างรูปที่ 10, 11, 12 และ 13



รูปที่ 10 : ข้อมูล ExampleSet จากการประมวลผลจากโมเดล FP-Growth

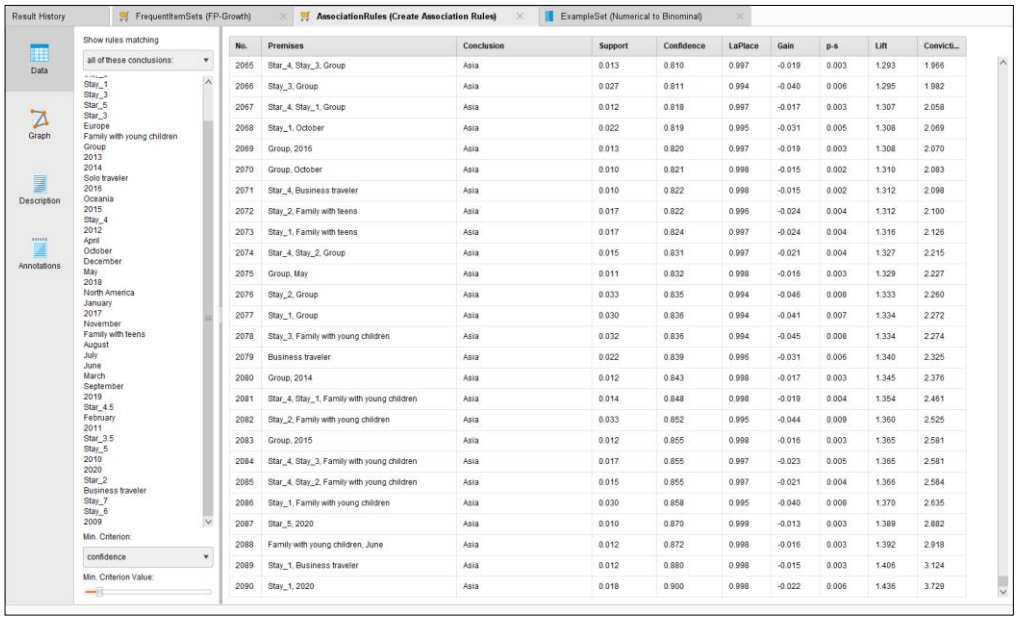
จากรูปที่ 10 คือ ตัวอย่างข้อมูลที่ทำการแปลงข้อมูลจาก Numerical ให้เป็น Binominal เพื่อแปลงข้อมูลตัวเลขจาก 0 หรือ 1 ให้เป็น True หรือ False



Size	Support	Item 1	Item 2	Item 3	Item 4
4	0.012	Asia	Couple	Star_4	2017
4	0.011	Asia	Couple	Star_4	August
4	0.010	Asia	Couple	Star_4	July
4	0.010	Asia	Couple	Star_4	June
4	0.012	Asia	Couple	Star_4	September
4	0.020	Asia	Couple	Stay_2	Star_5
4	0.014	Asia	Couple	Stay_2	Star_3
4	0.011	Asia	Couple	Stay_2	2013
4	0.012	Asia	Couple	Stay_2	2014
4	0.010	Asia	Couple	Stay_2	2015
4	0.018	Asia	Couple	Stay_1	Star_5
4	0.017	Asia	Couple	Stay_1	Star_3
4	0.010	Asia	Couple	Stay_1	2013
4	0.010	Asia	Couple	Stay_1	2014
4	0.014	Asia	Couple	Stay_3	Star_5
4	0.015	Asia	Star_4	Stay_2	Family with young children
4	0.015	Asia	Star_4	Stay_2	Group
4	0.010	Asia	Star_4	Stay_2	2014
4	0.010	Asia	Star_4	Stay_2	2016
4	0.014	Asia	Star_4	Stay_1	Family with young children
4	0.012	Asia	Star_4	Stay_1	Group
4	0.010	Asia	Star_4	Stay_1	2016
4	0.017	Asia	Star_4	Stay_3	Family with young children
4	0.013	Asia	Star_4	Stay_3	Group
4	0.010	Asia	Star_4	Stay_3	2014
4	0.010	Asia	Star_4	Stay_3	2016

รูปที่ 11 : ข้อมูล Frequency Item Sets จากการประมวลผลจากโมเดล Create Association Rules

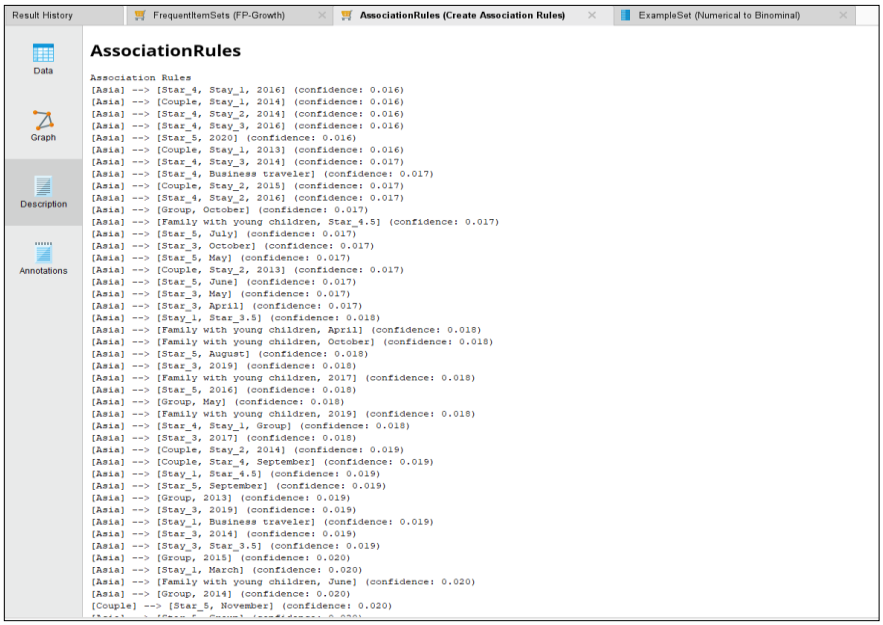
จากรูปที่ 11 แสดงให้เห็นถึงผลลัพธ์จากการหาค่า support ของ item ต่าง ๆ ในรูปแบบ itemset ทั้งแบบ 1 itemset ไปจนถึง 4 itemset จากการประมวลผลด้วย Operator “FP-Growth”



No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convicti...
2065	Star_4, Stay_3, Group	Asia	0.013	0.810	0.997	-0.019	0.003	1.293	1.966
2066	Stay_3, Group	Asia	0.027	0.811	0.994	-0.040	0.006	1.295	1.982
2067	Star_4, Stay_1, Group	Asia	0.012	0.818	0.997	-0.017	0.003	1.307	2.058
2068	Stay_1, October	Asia	0.022	0.819	0.995	-0.031	0.005	1.308	2.069
2069	Group, 2016	Asia	0.013	0.820	0.997	-0.019	0.003	1.308	2.070
2070	Group, October	Asia	0.010	0.821	0.998	-0.015	0.002	1.310	2.083
2071	Star_4, Business traveler	Asia	0.010	0.822	0.998	-0.015	0.002	1.312	2.098
2072	Stay_2, Family with teens	Asia	0.017	0.822	0.996	-0.024	0.004	1.312	2.100
2073	Stay_1, Family with teens	Asia	0.017	0.824	0.997	-0.024	0.004	1.316	2.126
2074	Star_4, Stay_2, Group	Asia	0.015	0.831	0.997	-0.021	0.004	1.327	2.215
2075	Group, May	Asia	0.011	0.832	0.998	-0.016	0.003	1.329	2.227
2076	Stay_2, Group	Asia	0.033	0.835	0.994	-0.046	0.008	1.333	2.260
2077	Stay_1, Group	Asia	0.030	0.836	0.994	-0.041	0.007	1.334	2.272
2078	Stay_3, Family with young children	Asia	0.032	0.836	0.994	-0.045	0.008	1.334	2.274
2079	Business traveler	Asia	0.022	0.839	0.996	-0.031	0.006	1.340	2.325
2080	Group, 2014	Asia	0.012	0.843	0.998	-0.017	0.003	1.345	2.376
2081	Star_4, Stay_1, Family with young children	Asia	0.014	0.848	0.998	-0.019	0.004	1.354	2.461
2082	Stay_2, Family with young children	Asia	0.033	0.852	0.995	-0.044	0.009	1.360	2.525
2083	Group, 2015	Asia	0.012	0.855	0.998	-0.016	0.003	1.365	2.581
2084	Star_4, Stay_3, Family with young children	Asia	0.017	0.855	0.997	-0.023	0.005	1.365	2.581
2085	Star_4, Stay_2, Family with young children	Asia	0.015	0.855	0.997	-0.021	0.004	1.366	2.584
2086	Stay_1, Family with young children	Asia	0.030	0.858	0.995	-0.040	0.008	1.370	2.635
2087	Star_5, 2020	Asia	0.010	0.870	0.999	-0.013	0.003	1.389	2.882
2088	Family with young children, June	Asia	0.012	0.872	0.998	-0.016	0.003	1.392	2.918
2089	Stay_1, Business traveler	Asia	0.012	0.880	0.998	-0.015	0.003	1.406	3.124
2090	Stay_1, 2020	Asia	0.018	0.900	0.998	-0.022	0.006	1.436	3.729

รูปที่ 12 : ข้อมูล Association Rules จากการประมวลผลจากโมเดล Create Association Rules

จากรูปที่ 12 แสดงให้เห็นถึงผลลัพธ์จากการประมวลผลด้วย Operator “Create Association Rules” แสดงผลลัพธ์ออกมาในรูปแบบตารางแสดงความสัมพันธ์ของข้อมูลในแต่ละ itemset และแสดงผลลัพธ์ค่า support, confidence และค่า Lift

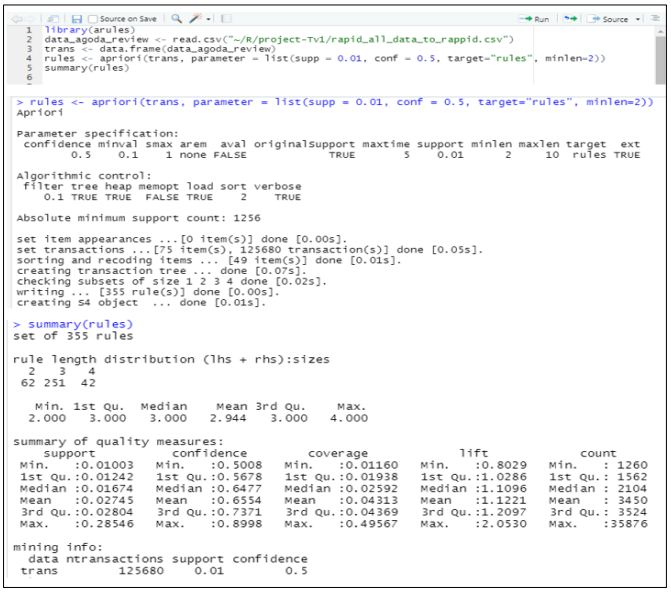


รูปที่ 13 : ข้อมูล Description Association Rules (support = 0.01, confidence = 0.5) จากการประมวลผลจากโมเดล Create Association Rules

จากรูปที่ 13 กฎความสัมพันธ์ข้างต้นมีทั้งหมด 355 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์โดยพิจารณาจากค่าสนับสนุน (Support) = 0.01 ค่าความเชื่อมั่น (Confidence) = 0.5 และค่าความ สอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 295 กฎ ซึ่งพบว่ามีกฎที่ค่าสอดคล้อง (Lift) ที่มีค่าน้อยกว่าเท่ากับ 1 อยู่ 60 กฎ

R Studio (Apriori Algorithm)

การวิเคราะห์ในโปรแกรม R Studio โดยใช้อัลกอริทึม Apriori วิเคราะห์ข้อมูลโดยใช้แพ็คเกจ arules วิธีการทำและผลลัพธ์แสดงดังรูปที่ 14



```
library(arules)
data_agoda_review <- read.csv("~/R/project-Tv1/rapid_all_data_to_rappid.csv")
trans <- data.frame(data_agoda_review)
rules <- apriori(trans, parameter = list(supp = 0.01, conf = 0.5, target="rules", minlen=2))
summary(rules)

> rules <- apriori(trans, parameter = list(supp = 0.01, conf = 0.5, target="rules", minlen=2))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target  ext
        0.5    0.1    1 none FALSE            TRUE       5    0.01      2     10  rules TRUE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 1256

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[75 item(s), 125680 transaction(s)] done [0.05s].
sorting and recoding items ... [49 item(s)] done [0.01s].
creating transaction tree ... done [0.07s].
checking subsets of size 1 2 3 4 done [0.02s].
writing ... [355 rule(s)] done [0.00s].
creating S4 object  ... done [0.01s].

> summary(rules)
set of 355 rules

rule length distribution (lhs + rhs):sizes
  2   3   4
 62 251  42

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   3.000   3.000   2.944   3.000   4.000

summary of quality measures:
    support         confidence        coverage            lift            count
 Min.   :0.01003   Min.   :0.5008   Min.   :0.01160   Min.   :0.8029   Min.   : 1260
 1st Qu.:0.01242   1st Qu.:0.5678   1st Qu.:0.01938   1st Qu.:1.0286   1st Qu.: 1562
 Median :0.01674   Median :0.6477   Median :0.02592   Median :1.1096   Median : 2104
 Mean   :0.02745   Mean   :0.6554   Mean   :0.04313   Mean   :1.1221   Mean   : 3450
 3rd Qu.:0.02804   3rd Qu.:0.7371   3rd Qu.:0.04369   3rd Qu.:1.2097   3rd Qu.: 3524
 Max.   :0.28546   Max.   :0.8998   Max.   :0.49567   Max.   :2.0530   Max.   :35876

mining info:
  data ntransactions support confidence
 trans        125680    0.01        0.5
```

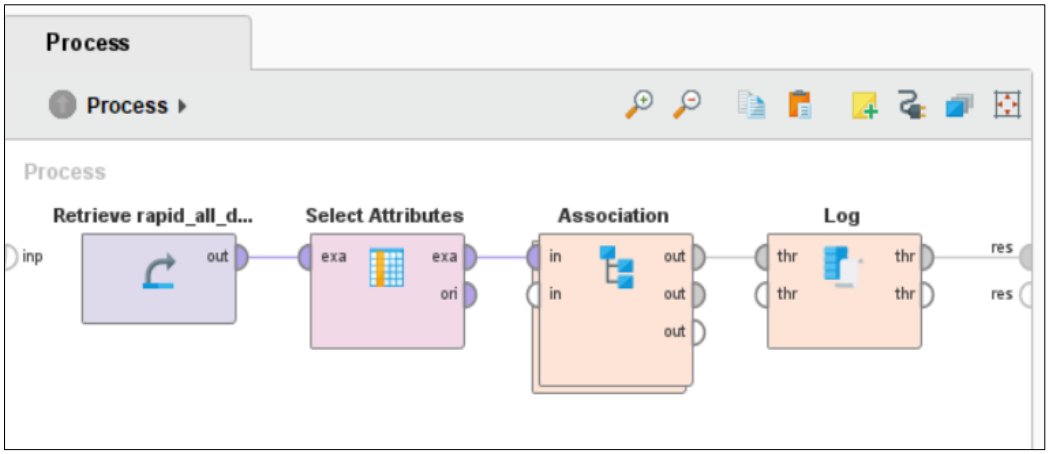
รูปที่ 14 : วิธีการวิเคราะห์ Association Rules โดยใช้อัลกอริทึม Apriori

จากรูปที่ 1 กฎความสัมพันธ์ข้างต้นที่มีได้ทั้งหมด 355 กฎ จากการประมวลผลโปรแกรม R Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์โดยพิจารณาจากค่าสนับสนุน (Support) = 0.01 ค่าความเชื่อมั่น (Confidence) = 0.5 และค่าความ สอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 295 กฎ ซึ่งจะเห็นได้ว่าผลลัพธ์จำนวนกฎที่ได้จากทั้ง 2 อัลกอริทึม มีจำนวนกฎที่เท่ากัน และกฎเหล่านั้นเป็นกฎเดียวกัน

3.2.4.2 การเปรียบเทียบประสิทธิภาพความเร็วในการประมวลผลระหว่าง Rapid Miner และ R-Studio

Rapid Miner (FP-Growth Algorithm)

การวัดประสิทธิภาพความเร็วในการประมวลผลของโปรแกรม Rapid Miner โดยใช้อัลกอริทึม FP-Growth โดยมี Process Model ในการทำ Association rule ที่ใช้แสดงดังรูปที่ 15



รูปที่ 15 : Process Model ที่ใช้ในการวัดประสิทธิภาพความเร็วในการประมวลผล

จากรูปที่ 15 แสดงให้เห็นถึงกระบวนการสร้าง Operator ในการหาความสัมพันธ์เพื่อวัดประสิทธิภาพความเร็วในการประมวลผล เริ่มต้นด้วยการนำไฟล์เข้า จากนั้นทำการใช้ Operator “Select Attribute” เพื่อเลือกเฉพาะ Attribute ที่ใช้สำหรับการวิเคราะห์ จากนั้นทำการเพิ่ม Object “Association” ข้างใน Object ก็จะประกอบไปด้วย Operator “Numerical to Binominal” เพื่อแปลงข้อมูลตัวเลขจาก 0 หรือ 1 ให้เป็น True หรือ False จากนั้นนำ Operator “FP-Growth” เพื่อนำข้อมูลเข้าสู่กระบวนการวิเคราะห์ด้วยอัลกอริทึม FP-Growth และ Operator “Create Association Rule” เพื่อทำการสร้างกฎความสัมพันธ์จากข้อมูลที่ได้จากการทำ FP-Growth สุดท้ายทำการเพิ่ม Operator “Log” เพื่อใช้ในการเก็บข้อมูลระยะเวลาในการประมวลผล

ผลการวัดประสิทธิภาพความเร็วในการประมวลผลทั้งหมด 3 รอบ

Round 1 - 645 * 0.001 = 0.645 seconds

Round 2 – 614 * 0.001 = 0.614 seconds

Round 3 – 611 * 0.001 = 0.611 seconds

เฉลี่ยใช้เวลาในการประมวลผลประมาณ 0.623 วินาที

R Studio (Apriori Algorithm)

การวัดประสิทธิภาพความเร็วในการประมวลผลของโปรแกรม R Studio โดยใช้อัลกอริทึม Apriori เพื่อหา Association rule วิธีทำแสดงดังรูปที่ 16

```
> start_time <- Sys.time ()
> library (dplyr)
> library (arules)
> data_agoda_review <- read.csv ("~/R/project-Tv1/rapid_all_data_to_rappid2.csv")
> trans <- data.frame (data_agoda_review)
> data_tran <- trans [2:76] %>% mutate_if (is.numeric,as.logical)
> rules <- apriori (data_tran, parameter = list (supp = 0.01, conf = 0.5,
target="rules", minlen=2) )
> finish_time <- Sys.time ()
> end_time <- finish_time-start_time
> summary (rules)
> end_time
```

รูปที่ 16 : โค้ดที่ใช้ในการประมวลผลของโปรแกรม R Studio

จากรูปที่ 16 เป็นตัวอย่างของโค้ดที่ใช้ในการวัดประสิทธิภาพความเร็วในการประมวลผลของโปรแกรม R Studio โดยใช้แพ็คเกจ (arules) อัลกอริทึม Apriori เพื่อหา Association rule

ผลการวัดประสิทธิภาพความเร็วในการประมวลผลทั้งหมด 3 รอบ

Round 1 - Time difference of 3.061 seconds

Round 2 - Time difference of 2.882 seconds

Round 3 - Time difference of 3.006 seconds

เฉลี่ยใช้เวลาในการประมวลผลประมาณ 2.983 วินาที

จากการเปรียบเทียบประสิทธิภาพความเร็วในการประมวลผลชุดข้อมูลระหว่าง Rapid Miner ซึ่งรันด้วย FP-Growth Algorithm และ R-Studio ซึ่งรันด้วย Apriori Algorithm พบว่า Rapid Miner ใช้เวลาในการประมวลผลเพียง 0.623 วินาที ซึ่งมีความเร็วมากกว่า R-Studio ที่ใช้เวลาในการประมวลถึง 2.983 วินาที ผลต่างของระยะเวลาคิดเป็น 2.36 วินาที หรือ อาจสรุปได้ว่า Rapid Miner ประมวลผลได้เร็วกว่า R-Studio ประมาณ 4.79 เท่า

3.2.4.3 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม Apriori โดยการใช้ R-Studio

ตารางที่ 5 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2จากอัลกอริทึม Apriori โดยการใช้ R-Studio

โปรแกรม	อัลกอริทึม	ชุด DATA1	ชุด DATA2
R-Studio	Apriori	P	P
R-Studio	FP-Growth	O	O
Rapid Miner	FP-Growth	P	O

ทางผู้วิจัยได้ทำการประมวลผลใน 2 ประเด็น คือ จำนวนกฎความสัมพันธ์ที่ได้ และระยะเวลาที่ใช้ พบว่า ทั้งชุด DATA1 (binominal) และ DATA2 (transaction) ได้จำนวนกฎที่เท่ากัน นั่นคือ 355 กฎ ซึ่งมีจำนวนกฎเท่ากันกับ FP-Growth ที่ทดสอบด้วยโปรแกรม Rapid miner เช่นกัน

ส่วนของระยะเวลาที่ใช้ในการประมวลผลนั้น พบว่า ชุด DATA2 ประมวลผลได้เร็วกว่า ชุด DATA1 ถึง 4.58 เท่า โดยชุด DATA2 ใช้เวลาเพียง 0.65 วินาที และ ชุด DATA1 ใช้เวลาทั้งสิ้น 2.98 วินาที ซึ่งอาจเป็นเพราะชุด DATA2 นั้น ข้อมูลอยู่ในลักษณะของ transaction ซึ่งเป็นการแปลงข้อมูลเป็นเหตุและผลต่อกันอยู่แล้ว

3.2.4.4 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม FP-Growth โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent

ตารางที่ 6 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA1 และ DATA2 จากอัลกอริทึม FP-Growth โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent

โปรแกรม	อัลกอริทึม	ชุด DATA1	ชุด DATA2	หมายเหตุ
R-Studio	Apriori	P	P	-
R-Studio	FP-Growth	O	P	มีการเพิ่มค่า consequent
Rapid Miner	FP-Growth	P	O	-

การประมวลผลด้วยอัลกอริทึม FP-Growth ของ R-Studio นั้น โดยปกติแล้วไม่สามารถประมวลผลได้ แต่หากทางผู้วิจัยได้ระบุ consequent (ผลที่จะตามมา) ให้กับโปรแกรม จะสามารถประมวลผลได้ ตัวอย่างแสดงดังรูปที่ 17

```
rules = fpgrowth(txns, support = 0.01, confidence = 0.5, maxLength = 10,consequent="continent")
```

รูปที่ 17 : แสดงตัวอย่างโค้ดที่มีการระบุ consequent = “continent”

จากรูปที่ 17 ผู้วิจัยได้ทำการกำหนด consequent มีค่าเท่ากับทวีป หรือ continent เพื่อหากฎความสัมพันธ์ที่เกี่ยวข้องกับทวีปของนักท่องเที่ยว และพบว่าจำนวนกฎความสัมพันธ์ทั้งหมดที่ได้คือ 257 กฎ ซึ่ง 257 กฎเหล่านี้เป็น subset ภายใน 355 กฎของความสัมพันธ์ที่ได้จากข้อ 3.2.4.2 และ 3.2.4.3 โดยกฎความสัมพันธ์อีก 98 กฎนั้น (355-257) ไม่ได้มีความเกี่ยวข้องกับทวีปของนักท่องเที่ยว

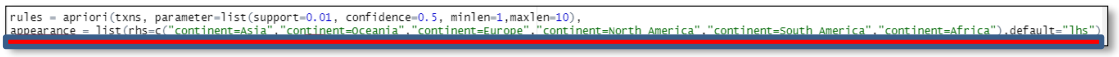
จากกฎความสัมพันธ์ที่ได้ 257 กฎนั้น โปรแกรม R-Studio ใช้เวลาในการประมวลผลทั้งสิ้น 70.3 วินาที ซึ่งถือว่าใช้ระยะเวลานานมาก

3.2.4.5 การเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA2 จากอัลกอริทึม FP-Growth และ Apriori โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent

ตารางที่ 7 รายละเอียดการเปรียบเทียบการหาความสัมพันธ์ของข้อมูล DATA2 จากอัลกอริทึม FP-Growth และ Apriori โดยการใช้ R-Studio แบบมีการเพิ่มค่า consequent

โปรแกรม	อัลกอริทึม	ชุด DATA1	ชุด DATA2	หมายเหตุ
R-Studio	Apriori	P	P	มีการเพิ่มค่า continent
R-Studio	FP-Growth	O	P	มีการเพิ่มค่า consequent
Rapid Miner	FP-Growth	P	O	-

จากข้อที่ 3.2.4.4 เมื่อประมวลผล FP-Growth ด้วย R-Studio แบบเจาะจง Continent แล้ว ทางผู้วิจัยจึงประมวลผล Apriori ด้วย R-Studio โดยเจาะจงส่วนของ continent เช่นกันเพื่อเปรียบเทียบจำนวนกฎความสัมพันธ์และระยะเวลาที่ใช้ในการประมวลผล ตัวอย่างแสดงในรูปที่ 18



รูปที่ 18 : แสดงตัวอย่างโค้ดที่มีการระบุส่วนของ continent

ผลลัพธ์ที่ได้จากการระบุส่วนของ continent นั้นพบว่า จำนวนกฎความสัมพันธ์มีจำนวนเท่ากันคือ 257 กฎ แต่ระยะเวลาที่ใช้ในการประมวลผลวิธีนี้ ใช้เวลาทั้งสิ้น 3.2 วินาที ซึ่งเร็วกว่าแบบ FP-Growth จากข้อ 3.2.4.4 ถึง 21.97 เท่า

3.2.5 การวิเคราะห์หาพฤติกรรมนักท่องเที่ยวชาวไทยในช่วงก่อนการเกิดสถานการณ์ COVID-19 และขณะเกิด COVID-19

จากการวิเคราะห์ด้วย FP-Growth Algorithm และ Apriori Algorithm แสดงผลลัพธ์ได้ว่ามีกฎความสัมพันธ์ที่เหมือนกัน แต่ระยะเวลาในการประมวลแตกต่างกัน โดยสรุปได้ว่า FP-Growth Algorithm จากโปรแกรม Rapid Miner ใช้ระยะเวลาประมวลผลรวดเร็วมากที่สุด ดังนั้นในการวิเคราะห์พฤติกรรมนักท่องเที่ยวของชาวไทยสำหรับสถานการณ์ก่อนเกิด COVID-19 และ ขณะเกิด COVID-19 จึงมีการประมวลผลด้วยโปรแกรม Rapid Miner โดย FP-Growth Algorithm

ส่วนของชุดข้อมูลในการวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูล

ชุดข้อมูลที่ใช้ในการวิเคราะห์หาพฤติกรรมของนักท่องเที่ยวชาวไทย แบ่งออกเป็น 2 ชุดข้อมูล

ชุดข้อมูลชุดที่ 1 : เป็นชุดข้อมูลบทวิจารณ์ของชาวไทยที่รวบรวมได้ตั้งแต่ปี 2009-2019 รวมระยะเวลา 10 ปี ซึ่งรวบรวมจากโรงแรมที่เข้าร่วมโครงการเราเที่ยวด้วยกันในจังหวัดภูเก็ต พังงา และกระบี่ โดยมีบทวิจารณ์ทั้งสิ้น 27,084 บทวิจารณ์

ชุดข้อมูลชุดที่ 2 : เป็นชุดข้อมูลบทวิจารณ์ของชาวไทยที่รวบรวมระหว่างเดือนกรกฎาคม ถึงธันวาคมปี 2020 ซึ่งรวบรวมจากโรงแรมที่เข้าร่วมโครงการเราเที่ยวด้วยกันในจังหวัดภูเก็ต พังงา และกระบี่ โดยมีบทวิจารณ์ทั้งสิ้น 3,047 บทวิจารณ์ โดยทางผู้วิจัยรวบรวมตั้งแต่เดือนกรกฎาคมเป็นต้นมา เนื่องจากเป็นเดือนแรกที่โครงการเราเที่ยวด้วยกันเริ่มใช้นโยบายในการส่งเสริมการท่องเที่ยวนี้

3.2.6 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์

การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์เริ่มโดยการนำบทวิจารณ์ทั้งหมดที่ได้หลังจากกระบวนการเตรียมข้อมูล ข้อ 3.2.3

โดยทำการวิเคราะห์ด้วยภาษา R และ ไลบรารี sentimentr ในโปรแกรม R-Studio โดยแบ่งกระบวนการออกเป็น 4 ส่วน

3.2.6.1 การวิเคราะห์อารมณ์และความรู้สึกของประโยคทั้งหมดในแต่ละบทวิจารณ์

การวิเคราะห์อารมณ์และความรู้สึกของประโยคทั้งหมดในแต่ละบทวิจารณ์เริ่มต้นด้วยการนำบทวิจารณ์ที่ได้จากการเตรียมข้อมูลในข้อ 3.2.3 มาทำวิเคราะห์อารมณ์และความรู้สึกในแต่ละบทวิจารณ์โดยทำการวิเคราะห์ด้วยภาษา R และ วิเคราะห์โดยใช้ไลบรารี sentimentr ในโปรแกรม R-Studio วิธีการทำแสดงดังรูปที่ 19

```
library(sentimentr)
SENTIMENTfn <- function(xtext){
  data_sen <- sentiment_by(xtext, by = NULL)
  return(data_sen$ave_sentiment)
 }
comm <- read.csv("~/R/sentiment/book.csv")
alldata <- data.frame()
for (i in 1:nrow(comm)) {
  data_sen <- SENTIMENTfn(comm$Translate_comment[i], by = NULL)
  alldata <- rbind(alldata,data_sen)
}
```

รูปที่ 19 : วิธีการแยกคำและหาชนิดของคำ

จากรูปที่ 19 เป็นตัวอย่างของโค้ดที่ใช้ในวิเคราะห์อารมณ์และความรู้สึกของประโยคทั้งหมดในแต่ละบทวิจารณ์โดยใช้ โดยใช้แพ็คเกจ sentimentr ในการวิเคราะห์

3.2.6.2 การวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์

การวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์เริ่มต้นด้วยการนำบทวิจารณ์ที่ได้จากการเตรียมข้อมูลในข้อ 3.2.3 มาทำการแยกคำ วิเคราะห์อารมณ์และความรู้สึกของคำ และหาชนิดของคำ (Part of Speech) ในแต่ละบทวิจารณ์โดยทำการวิเคราะห์ด้วยภาษา R และ วิเคราะห์อารมณ์และความรู้สึกของคำโดยใช้ไลบรารี sentimentr และ POStagObject ในโปรแกรม R-Studio วิธีการทำแสดงดังรูปที่ 20

```
source("https://s1cladal.github.io/rscripts/POStagObject.r")
library(sentimentr)
comm <- read.csv("~/R/R/02-all_data_hotel_review_clean_2009-2020.csv")
SENTIMENTfn <- function(xtext){
 # require("sentimentr")
  data_sen <- sentiment_by(xtext, by = NULL)
  return(data_sen$ave_sentiment)
}
POStag <- function(object){
  # define paths to corpus files
  corpus.tmp <- object
  # define sentence annotator
  sent_token_annotator <- Maxent_Sent_Token_Annotator()
  # define word annotator
  word_token_annotator <- Maxent_Word_Token_Annotator()
  # define pos annotator
  pos_tag_annotator <- Maxent_POS_Tag_Annotator(language = "en", probs = FALSE,
  model = "C:\\Users\xx\Documents\\R\\win-library\\4.0\\openNLPdata\\models\\en-pos-maxent.bin")
  # convert all file content to strings
  Corpus <- lapply(corpus.tmp, function(x){
    x <- as.String(x)  }  )
  # loop over file contents
  lapply(Corpus, function(x){
    y1 <- NLP::annotate(x, list(sent_token_annotator, word_token_annotator))
    y2<- NLP::annotate(x, pos_tag_annotator, y1)
    y2w <- subset(y2, type == "word")
    tags <- sapply(y2w$features, '[[', "POS")
    getpointSenti <- SENTIMENTfn(x[y2w])
    r1 <- sprintf("%s,%s", x[y2w], tags)
    r2 <- paste("",r1,",",getpointSenti,"/", collapse = "")
    return(r2)  }  )
}
comm <- read.csv("~/R/R/2009-2020.csv")
comm$id <- 1:nrow(comm)
finaldata_sentiment = data.frame()
for (i in 1:nrow(comm)) {
  Sys.sleep(0.1)
  finaldata <- data.frame()
  #data and clean data
  text <- comm$Translate_comment[i]
  # clean data
  text <- text %>%
    str_squish()
  text <- gsub("@\\w+", "", text)
  text <- gsub("https?://.+", "", text)
  text <- gsub("\\d+\\w*\\d*", "", text)
  text <- gsub("#\\w+", "", text)
  text <- gsub("[^\x01-\x7F]", "", text)
  text <- gsub("[[:punct:]]", " ", text)
  # Remove spaces and newlines
  text <- gsub("\n", " ", text)
  text <- gsub("^\\s+", "", text)
  text <- gsub("\\s+$", "", text)
  text <- gsub("[ |\t]+", " ", text)
  textpos <- POStag(object = text)

  textpos <- as.list(strsplit(as.character(textpos) , "/")[[1]])

  datasss <- unlist(textpos)
  df <- data.frame(x = c(datasss))
  finaldata<- data.frame(df %>% extract(x, c("word", "pos","sentiment"), "([^,]+),([^,]+), ([^)]+)"))
  finaldata <- data.frame(finaldata,comm$X.uHotel_id[i],comm$Hotel_location[i],comm$id[i])
  finaldata$word_no <- 1:nrow(finaldata)
  finaldata_sentiment  <-  rbind(finaldata_sentiment,finaldata)
}
```

รูปที่ 20 : ตัวอย่างโค้ดวิธีการแยกคำ วิเคราะห์อารมณ์และความรู้สึกของคำ และหาชนิดของคำ

จากรูปที่ 20 เป็นตัวอย่างของโค้ดที่ใช้ในการแยกคำ วิเคราะห์อารมณ์และความรู้สึกของคำ และหาชนิดของคำ โดยใช้แพ็คเกจ sentimentr และ POStagObject

3.2.6.3 การจัดกลุ่มคำศัพท์

การจัดกลุ่มคำศัพท์เริ่มต้นด้วยการนำข้อมูลผลลัพธ์ที่ได้ในข้อ 3.2.6.2 มาทำการรวมคำศัพท์ที่เหมือนกันโดยใช้ภาษา R โดยผลลัพธ์ที่ได้จากการรวมคำศัพท์มีจำนวน 66,832 คำ จากนั้นทำการจัดหมวดหมู่ให้ในแต่ละคำเพื่อเป็นคลังข้อมูลในการจัดกลุ่มคำศัพท์ โดยมีการกำหนดชื่อหัวข้อหมวดหมู่อ้างอิงมาจากหัวข้อผลรวมของคะแนนรีวิวบนเว็บไซต์ Agoda โดยมีชื่อหมวดหมู่ดังนี้

1.ความสะอาด (Cleanliness)

2.สิ่งอำนวยความสะดวก (Facilities)

3.ทำเลที่ตั้ง (Location)

4.ความสะดวกสบายและคุณภาพของห้องพัก (Room comfort and quality)

5.การให้บริการของพนักงาน (Service)

6.ความคุ้มค่าของเงิน (Value for money)

ในการจัดกลุ่มคำศัพท์ทางผู้วิจัยได้จัดทำ 2 วิธีเพื่อทำการเปรียบเทียบผลลัพธ์และนำผลลัพธ์ที่เหมาะสมที่สุดมาใช้ในการวิเคราะห์ในขั้นตอนถัดไป โดยทำการจัดกลุ่มโดยใช้ Model LDA (Latent Dirichlet Allocation) ในภาษา R และจัดกลุ่มแบบ Manual

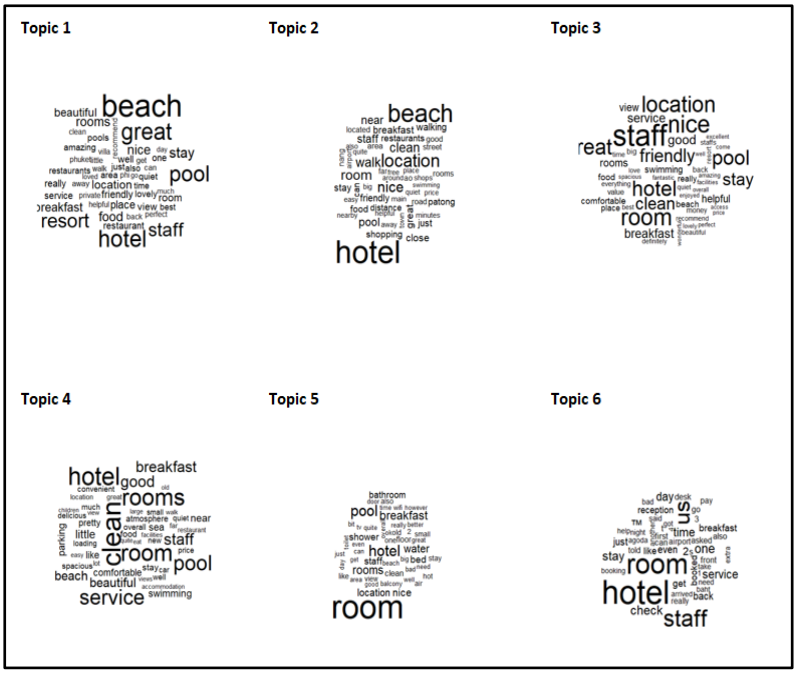
การจัดกลุ่มรูปแบบที่ 1 Model LDA (Latent Dirichlet Allocation)

การจัดกลุ่มคำโดยใช้ Model LDA (Topicmodels) เพื่อจัดกลุ่มของคำศัพท์ โดยใช้ภาษา R และไลบรารี Topicmodels วิธีการทำและผลลัพธ์แสดงดังรูปที่ 21 และ 22

```
library(topicmodels)
dtm = convert(dfm, to = "topicmodels")
set.seed(1)
m = LDA(dtm, method = "Gibbs", k = 6,  control = list(alpha = 0.1))
#m
terms(m, 5)
result <- data.frame(terms)
topic = 6
words = posterior(m)$terms[topic, ]
topwords = head(sort(words, decreasing = T), n=50)
head(topwords)
library(wordcloud)
wordcloud(names(topwords), topwords)
```

รูปที่ 21 : วิธีการจัดกลุ่มคำโดยใช้ Model LDA

จากรูปที่ 21 เป็นตัวอย่างของโค้ดที่ใช้ในการจัดกลุ่มโดยใช้ภาษา R และไลบรารี Topicmodels โดยทำการกำหนดค่าผลลัพธ์ออกมา 6 Topics และให้แสดงผลลัพธ์ออกมาในรูปแบบ Wordcloud



รูปที่ 22 : ผลลัพธ์ที่ได้จากการจัดกลุ่มโดยใช้ Model LDA รูปแบบ wordcloud

จากรูปที่ 22 เป็นผลลัพธ์ที่ได้จากการจัดกลุ่มโดยใช้ Model LDA จากผลลัพธ์หากนำชื่อหัวข้อหมวดหมู่ที่กำหนดไว้มากำหนดให้แต่ละหมวดหมู่ซึ่งผลลัพธ์ที่ยังไม่เหมาะสมและไม่ตรงตามความต้องการ โดยผลลัพธ์ที่ได้แต่ละกลุ่มคำศัพท์ที่เห็นได้ชัดคือคำว่า “beach great pool resort hotel location staff nice clean room” หากนำคำเหล่านี้ไปจัดกลุ่มตามหัวข้อหมวดหมู่กลุ่มคำศัพท์ จากข้อที่ 3.2.6.3 พบว่ามีความเป็นไปได้ที่คำศัพท์กลุ่มนี้จะอยู่ในหมวดหมู่ ทำเลที่ตั้ง (Location) เนื่องจากพบคำว่า “beach resort hotel location” และ การให้บริการของพนักงาน (Service) เนื่องจากพบคำว่า “great staff nice clean room” ทำให้พบว่าผลลัพธ์ที่ได้ในแต่ละกลุ่มคำศัพท์เจอคำซ้ำมากเกินไปทำให้จำแนกไม่ครบทุกหมวดหมู่

การจัดกลุ่มรูปแบบที่ 2 การป้อนข้อมูลโดยผู้วิจัย (Manual)

การจัดกลุ่มคำแบบ Manual เป็นการป้อนข้อมูลโดยผู้วิจัยโดยคำนึงถึงความหมายของคำศัพท์และความเป็นไปได้ เพื่อให้ผลลัพธ์ตรงกับหมวดหมู่ที่กำหนดมากที่สุด เพื่อใช้ผลลัพธ์เป็นคลังข้อมูลในการจัดกลุ่มในการวิเคราะห์ในขั้นตอนถัดไป วิธีที่ใช้จัดการข้อมูลจัดการด้วยโปรแกรม Excel ในการจัดการข้อมูลแสดงรายละเอียดดังรูปที่ 23

ID	Term	Role	Cleanliness	Facilities	Location	Room_comfort_and_quality	Service	Value_for_money
1	aa	Noun	0	0	0	0	0	0
2	ab	Noun	0	0	0	0	0	0
3	abandon	Verb	0	0	0	0	0	0
4	abate	Verb	0	0	0	0	0	0
5	abbreviate	Verb	0	0	0	0	0	0
6	abhor	Verb	0	0	0	0	0	0
7	ability	Noun	0	0	0	0	0	0
8	ablution	Noun	1	0	0	1	0	1
9	abound	Verb	0	0	0	0	0	0
10	abrasion	Noun	1	0	0	1	0	1
11	absorb	Verb	1	0	0	0	0	0
12	absorber	Noun	1	1	0	1	0	1
13	abuse	Verb	0	0	0	0	0	0
14	abut	Verb	0	0	0	0	0	0
15	accede	Verb	0	0	0	0	0	0
16	accelerate	Verb	0	0	0	0	0	0
17	accent	Noun	0	0	0	0	1	0
18	accentuate	Verb	0	0	0	0	1	0
19	accept	Verb	0	0	0	1	1	1
20	access	Verb	0	0	1	0	0	0

รูปที่ 23 : ตัวอย่างผลลัพธ์ 20 คำแรกที่ได้จากการจัดกลุ่มแบบ Manual

จากรูปที่ 23 เป็นตัวอย่างของผลลัพธ์ที่ได้จากการจัดกลุ่มแบบ Manual โดยทำการกำหนดค่าในแต่ละหมวดหมู่ให้เป็น 0 หรือ 1 โดย 0 มีความหมายคือ ไม่ได้อยู่ในหมวดหมู่กลุ่มคำศัพท์ที่กำหนด จากข้อ3.2.6.3 และ 1 มีความหมายเกี่ยวข้องตามหัวข้อหมวดหมู่กลุ่มคำศัพท์ที่กำหนดจากข้อ 3.2.6.3

3.2.6.4 การแสดงผลลัพธ์

หลังจากได้ผลลัพธ์จากข้อ 3.2.6.1 – 3.2.6.3 ทำการนำผลลัพธ์ทั้งหมดที่ได้มาแสดงผลลัพธ์แนวโน้มของการวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ในรูปแบบ Dashboard โดยใช้ Data Studio ในการแสดงผลลัพธ์ จากนั้นทำการคัดเลือกข้อมูลโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุด 5 อันดับแรกในแต่ละจังหวัดมาแสดงผลลัพธ์ โดยข้อมูลโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุด 5 อันดับแรกแสดงผลลัพธ์ดังรูปที่ 24

	id	name_hotel	จำนวน
phuket	118	The Royal Paradise Hotel & Spa	3,277
	181	Sleep with Me Hotel Design Hotel at Pa	2,036
	165	Patong Resort Hotel	1,849
	99	Rayaburi Hotel Patong	1,796
	177	Wyndham Sea Pearl Resort Phuket	1,600
phangnga	269	Aleenta Resort	571
	238	Seaview Resort Khao Lak	386
	241	The Briza Beach Resort	384
	272	Ayara Villas Hotel	363
	257	Khaolak Emerald Beach Resort & Spa	281
krabi	420	Areetara Resort	3,154
	431	Aonang Orchid Resort	2,287
	286	Centara Anda Dhevi Resort and Spa	1,886
	275	Deevana Plaza Krabi Aonang	1,744
	313	Golden Beach Resort	1,712

รูปที่ 24 : ผลลัพธ์ 5 อันดับโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุดของ 3 จังหวัด

จากรูปที่ 24 เป็นผลลัพธ์ที่ได้จากการจัดอันดับโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุดของ 3 จังหวัด ตามลำดับดังนี้

จังหวัดภูเก็ต

1. The Royal Paradise Hotel & Spa
2. Sleep with me Hotel Design Hotel at Spa
3. Patong Resort Hotel
4. Rayaburi Hotel Patong
5. Wyndham Sea Pearl Resort Phuket

จังหวัดพังงา

1. Aleenta Resort
2. Seaview Resort Khao lak
3. The Briza Beach Resort
4. Araya Villas Hotel
5. Khaolak Emerald Beach Resort & Spa

จังหวัดกระบี่

1. Areetara Resort
2. Aonang Orchid Resort
3. Centara Anda Dhevi Resort and Spa
4. Deevana Plaza Krabi Aonang
5. Golden Beach Resort.

บทที่ 4

ผลดำเนินงานวิจัย

บทวิจัยฉบับนี้ได้ทำการวิเคราะห์บทวิจารณ์ออนไลน์เพื่อวิเคราะห์ความสัมพันธ์ของข้อมูล (Association Rules) เป็นการวิเคราะห์เพื่อดูพฤติกรรมการเข้าพักโรงแรมของนักท่องเที่ยวในมุมมองกว้าง ซึ่งจะมีการเปรียบเทียบโดยใช้อัลกอริทึม Apriori และ FP-Growth และเปรียบเทียบประสิทธิภาพความเร็วในการวิเคราะห์ อีกทั้งมีการวิเคราะห์เปรียบเทียบพฤติกรรมนักท่องเที่ยวชาวไทยก่อนเกิดสถานการณ์ COVID-19 และ หลังการเกิดสถานการณ์ COVID-19 และทำการวิเคราะห์ข้อมูลความรู้สึกจากบทวิจารณ์โดยใช้ การวิเคราะห์ความรู้สึก (Sentiment Analysis) เพื่อบอกความรู้สึกจากบทวิจารณ์ออนไลน์ และแยกคำศัพท์ในบทวิจารณ์เพื่อนำคำศัพท์ที่ได้จัดหมวดหมู่เพื่อใช้เป็นคลังข้อมูลในการจัดหมวดหมู่บทวิจารณ์เพื่อการวิเคราะห์ข้อมูลโดยทำการเปลี่ยนเทียบ 2 วิธีการสร้างแบบจำลองหัวข้อ (Topic Modelling) และวิธีการป้อนข้อมูลโดยผู้วิจัย (Manual) และคัดเลือกผลลัพธ์วิธีที่ที่เหมาะสมสำหรับงานวิจัยครั้งนี้ และนำผลลัพธ์ไปแสดงในรูปแบบ Dashboard

4.1 การวิเคราะห์กฎความสัมพันธ์ของข้อมูล

4.1.1 ข้อมูลเชิงลึกของบทวิจารณ์ออนไลน์ที่ได้รวบรวม

ข้อมูลบทวิจารณ์ออนไลน์ในเว็บไซต์ Agoda โดยมีจำนวนโรงแรม 427 โรงแรม และบทวิจารณ์จำนวน 246,532 บทวิจารณ์ได้รวบรวมตั้งแต่ปี 2009-2020 หลังจากผ่านกระบวนการเตรียมข้อมูลในข้อ 3.2.3 จำนวนบทวิจารณ์มีจำนวนทั้งสิ้น 125,486 บทวิจารณ์

การวิเคราะห์บทวิจารณ์ออนไลน์เพื่อหาความสัมพันธ์ของข้อมูลโดยใช้โปรแกรม Rapid Miner โดยใช้ Model Association Rule เพื่อหาความสัมพันธ์ของข้อมูลและการเปรียบเทียบประสิทธิภาพความเร็วในการประมวนผลโดยใช้อัลกอริทึม FP-Growth โดยกำหนดค่า Minimum Support = 1% (0.01) เนื่องจากข้อมูลมีการกระจายกลุ่มข้อมูลจำนวนมาก และมีปริมาณที่สูงเกินไปทำให้การตั้งค่า Minimum Support ที่มีค่ามากกว่า 0.01 ทำให้ไม่สามารถหากลุ่มเซตไอเทม (itemset) และกฎความสัมพันธ์ที่ดีที่สุดได้ และ ตั้งค่า Minimum Confidence = 50% (0.5)

4.1.2 ผลการทดลองการวิเคราะห์บทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020 (อ้างอิงจากข้อ 3.2.4)

ผลลัพธ์ของชุดข้อมูลบทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020 กฎความสัมพันธ์ได้ทั้งหมด 355 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์โดยพิจารณาจากค่าสนับสนุนตามที่กำหนด และค่าความ สอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 295 กฎ โดยทำการคัดเลือกเฉพาะกลุ่ม itemset ที่มีความสัมพันธ์เกี่ยวกับชนิดกลุ่มนักท่องเที่ยวผลลัพธ์กฎความสัมพันธ์ที่สนใจจำนวน 40 กฎ ผลลัพธ์แสดงดังรูปที่ 25

Premises	Conclusion
Star_4, Business traveler	Asia
Stay_1, Business traveler	Asia
Star_4, Family with teens	Asia
Stay_1, Family with teens	Asia
Stay_2, Family with teens	Asia
Stay_3, Family with teens	Asia
Group, May	Asia
Star_4, Stay_1, Group	Asia
Star_4, Stay_2, Group	Asia
Star_4, Stay_3, Group	Asia
Family with young children, April	Asia
Family with young children, December	Asia
Family with young children, June	Asia
Family with young children, October	Asia
Star_3, Family with young children	Asia
Star_4, Stay_1, Family with young children	Asia
Star_4, Stay_2, Family with young children	Asia
Star_4, Stay_3, Family with young children	Asia
Star_5, Family with young children	Asia
Asia, Stay_1, Star_5	Couple
Asia, Stay_2, Star_5	Couple
Asia, Stay_3, Star_5	Couple
Couple, Star_4, Stay_1	Asia
Couple, Star_4, Stay_2	Asia
Couple, Star_4, Stay_3	Asia
Couple, Stay_1, Star_3	Asia
Couple, Stay_1, Star_5	Asia
Couple, Stay_2, Star_5	Asia
Star_4, North America	Couple
Couple, Oceania	Star_4
Star_5, Oceania	Couple
Stay_2, Oceania	Couple
Stay_3, Oceania	Couple
Europe, February	Couple
Europe, January	Couple
Europe, March	Couple
Europe, Stay_4	Couple
Star_4, Europe	Couple
Star_5, Europe	Couple
Stay_2, Europe	Couple

รูปที่ 25 : ตารางสรุปผลลัพธ์ที่ได้ของชุดข้อมูลบทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020

จากรูปที่ 24 แสดงผลลัพธ์ได้ว่ากลุ่มนักท่องเที่ยวทวีป Asia รูปแบบ Business traveler, Family with teens และ Group มีการเข้าพักโรงแรมระดับ 4 ดาว และเข้าพักเป็นจำนวน 1-3 คืน หากเป็นกลุ่มนักท่องเที่ยวรูปแบบ Business traveler จะมีการเข้าพักเพียงจำนวน 1 คืน เท่านั้น ในส่วนของกลุ่มนักท่องเที่ยวทวีป Asia รูปแบบ Family with young children และ Couple มีการเข้าพักโรงแรมระดับ 3-5 ดาว และเข้าพักเป็นจำนวน 1-3 คืน และกลุ่มนักท่องเที่ยวรูปแบบ Family with young children นี้จะมีการเข้าพักในช่วงเดือน เมษายน มิถุนายน ตุลาคม และธันวาคม เนื่องจาก ในเดือนเมษายนและธันวาคมเป็นช่วงที่มีวันหยุดประจำปีและเหมาะแก่การเที่ยวทะเล ส่วนในเดือนมิถุนายนเป็นช่วงโลว์ซีซั่นจึงทำให้ไม่มีนักท่องเที่ยวพลุกพล่านมากนัก จึงทำให้ทางโรงแรมมีการเสนอโปรโมชั่นเพื่อดึงดูดนักท่องเที่ยว และในเดือนตุลาคมเป็นเดือนแห่งเทศกาลท้องถิ่นจึงทำให้นักท่องเที่ยวนิยมมาท่องเที่ยวเพื่อชมเทศกาลของประเทศไทย เช่นในภูเก็ต จะมีเทศกาลถือศีลทานผัก (ทานเจ) เป็นต้น [27], [28], [29] หากเป็นกลุ่มนักท่องเที่ยวทวีป North America มีการเข้าพักโรงแรมระดับ 4 ดาว ในส่วนของกลุ่มนักท่องเที่ยวทวีป Oceania รูปแบบ Couple มีการเข้าพักโรงแรมระดับ 4-5 ดาว และเข้าพักเป็นจำนวน 2-3 คืน และในส่วนของกลุ่มนักท่องเที่ยวทวีป Europe รูปแบบ Couple พบว่ามีการเข้าพักโรงแรมระดับ 4-5 ดาว เข้าพักเป็นจำนวน 2-4 คืน และมีการเข้าพักในช่วงเดือนมกราคม กุมภาพันธ์ และมีนาคม เนื่องจากทางทวีปยุโรปมีอากาศหนาวมาก ทำให้ไม่เหมาะในการท่องเที่ยวในประเทศตนเอง อีกทั้งนักท่องเที่ยวส่วนใหญ่ชอบการอาบแดด (ผิวสีแทน)

4.2 ผลการทดลองการวิเคราะห์หาพฤติกรรมนักท่องเที่ยวชาวไทยในช่วงก่อนการเกิดสถานการณ์ COVID-19 และขณะเกิด COVID-19 (อ้างอิงจากข้อ 3.2.5)

4.2.1 การหากฎความสัมพันธ์ที่เกี่ยวข้อง

การนำบทวิจารณ์ทั้ง 2 ชุดข้อมูล มาจัดเตรียมข้อมูลใหม่และทำการประมวลผลเพื่อหากฎความสัมพันธ์ที่เกี่ยวข้องโดยการใช้ FP-Growth Algorithm โดยกำหนดค่า Minimum Support = 1% (0.01) และ ตั้งค่า Minimum Confidence = 50% (0.5)

ผลลัพธ์ชุดข้อมูลในการวิเคราะห์เพื่อหาความสัมพันธ์ของข้อมูล

ผลลัพธ์ข้อมูลชุดที่ 1

ผลลัพธ์ของชุดข้อมูลก่อนการเกิดสถานการณ์ COVID-19 กฎความสัมพันธ์ได้ทั้งหมด 83 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์โดยพิจารณาจากค่าสนับสนุนตามที่กำหนด และค่าความ สอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 73 กฎ

ผลลัพธ์ข้อมูลชุดที่ 2

ผลลัพธ์ของชุดข้อมูลขณะเกิดสถานการณ์ COVID-19 กฎความสัมพันธ์ได้ทั้งหมด 89 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules ได้นำมาวิเคราะห์โดยพิจารณาจากค่าสนับสนุนตามที่กำหนด และค่าความ สอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 52 กฎ

4.2.2 การแปรผลลัพธ์ที่ได้จากการหากฎความสัมพันธ์

การแปรผลลัพธ์ข้อมูลชุดที่ 1

กฎความสัมพันธ์ที่มีค่าความสอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 73 กฎ จากผลลัพธ์กฎความสัมพันธ์ทำการคัดเลือกเฉพาะกลุ่ม itemset ที่มีความสัมพันธ์เกี่ยวกับชนิดกลุ่มนักท่องเที่ยวผลลัพธ์กฎความสัมพันธ์ที่ได้ 22 กฎความสัมพันธ์ ผลลัพธ์แสดงดังรูปที่ 26

Premises	Conclusion	Support	Confidence	Lift
Business traveler	Stay_1	0.04360508	0.613187954	1.201967
Star_3, Business traveler	Stay_1	0.0136612	0.651408451	1.276887
Star_4, Business traveler	Stay_1	0.01842416	0.589834515	1.15619
Couple	Stay_1	0.21101019	0.513062214	1.005701
Couple, August	Stay_1	0.0173165	0.531143828	1.041145
Couple, July	Stay_1	0.01849801	0.52681388	1.032657
Couple, June	Stay_1	0.01990105	0.557394002	1.0926
Couple, November	Stay_1	0.01587653	0.519951632	1.019206
Couple, September	Stay_1	0.01816571	0.590636255	1.157762
Couple, Star_2	Stay_1	0.01019052	0.584745763	1.146215
Couple, Star_3	Stay_1	0.04626348	0.541486603	1.061419
Couple, Star_5	Stay_1	0.04024516	0.516587678	1.012612
Stay_1, Star_5	Couple	0.04024516	0.513424399	1.248369
Family with young children, May	Stay_1	0.01181509	0.52892562	1.036797
Family with young children, October	Stay_1	0.01203663	0.546063652	1.070391
Family with young children, October	Star_4	0.01103973	0.500837521	1.167759
Star_3, Family with young children	Stay_1	0.01620883	0.58768407	1.151975
Star_3, Family with teens	Stay_1	0.01236893	0.52672956	1.032492
Solo traveler	Stay_1	0.04135283	0.558325025	1.094425
Star_3, Solo traveler	Stay_1	0.01369812	0.564687976	1.106898
Star_4, Solo traveler	Stay_1	0.01510117	0.549731183	1.07758
Star_3, Group	Stay_1	0.02466401	0.511094109	1.001844

รูปที่ 26 : ตารางสรุปผลลัพธ์ที่ได้จากการหาความสัมพันธ์ช่วงก่อนการระบาด COVID-19

จากรูปที่ 26 แสดงผลลัพธ์ได้ว่าในกลุ่มนักท่องเที่ยวรูปแบบ Business traveler มีการเข้าพักเฉพาะโรงแรม 3 ดาว และ 4 ดาว และเข้าพักเป็นจำนวน 1 คืน ส่วนของกลุ่มนักท่องเที่ยวในรูปแบบ Couple มีการเข้าพักในช่วงเดือนมิถุนายน กรกฎาคม สิงหาคม กันยายน และพฤศจิกายน โดยเข้าพักโรงแรมระดับ 2 ดาว 3 ดาว และ 5 ดาวจำนวน 1 คืน นักท่องเที่ยวกลุ่มคู่รักเข้าพักในช่วงเดือนดังกล่าว เพราะในช่วงเดือนดังกล่าวเป็นช่วงก่อนหน้าไฮซีซั่น สถานที่ท่องเที่ยวเริ่มคึกคัก และบางคู่เลือกที่จะเข้าพักโรงแรมระดับ 2-3 ดาว และ 5 ดาว ส่วนของกลุ่มนักท่องเที่ยวรูปแบบ Group นักท่องเที่ยวแบบกลุ่มมักจะเลือกเข้าพักโรงแรมระดับ 3 ดาว จำนวน 1 คืน ส่วนของกลุ่มนักท่องเที่ยวในรูปแบบ Family with teens, Family with young children มีการเข้าพักเฉพาะโรงแรม 3 ดาว จำนวน 1 คืน ส่วนของกลุ่มกลุ่มนักท่องเที่ยวในรูปแบบ Solo traveler มีการเข้าพักเฉพาะโรงแรม 3 ดาว และ 4 ดาว จำนวน 1 คืน จากรายละเอียดข้างต้นจะพบว่าก่อนเกิดเหตุการณ์ COVID-19 นักท่องเที่ยวเข้าพักตั้งแต่โรงแรมระดับ 2 ดาว ขึ้นไป เนื่องโรงแรมและค่าครองชีพในจังหวัดภูเก็ตราคาค่อนข้างสูง ดังนั้น นักท่องเที่ยวจะแบ่งเป็น 2 กลุ่ม ได้แก่

1.กลุ่มที่พักโรงแรมระดับ 2-3 ดาว

2.กลุ่มที่พักโรงแรมระดับ (สำหรับนักท่องเที่ยวรายได้สูง) 4-5 ดาว

การแปรผลลัพธ์ข้อมูลชุดที่ 2

กฎความสัมพันธ์ที่มีค่าความสอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ผลตามกฎความสัมพันธ์ 52 กฎ จากผลลัพธ์กฎความสัมพันธ์ทำการคัดเลือกเฉพาะกลุ่ม itemset ที่มีความสัมพันธ์เกี่ยวกับชนิดกลุ่มนักท่องเที่ยวผลลัพธ์กฎความสัมพันธ์ที่ได้ 32 กฎความสัมพันธ์ ผลลัพธ์แสดงดังรูปที่ 26

Premises	Conclusion	Support	Confidence	Lift
Business traveler	Stay_1	0.02166065	0.640776699	1.073953
Star_4, Business traveler	Stay_1	0.010502133	0.64	1.072651
Couple, July	Stay_1	0.01772235	0.642857143	1.07744
Couple, November	Stay_1	0.048900558	0.628691983	1.053699
Couple, October	Stay_1	0.061043649	0.607843137	1.018756
Couple, September	Stay_1	0.035116508	0.614942529	1.030655
Couple, Star_3	Stay_1	0.032162783	0.604938272	1.013887
Couple, Star_4	Stay_1	0.094191008	0.614561028	1.030015
Couple, Star_4, November	Stay_1	0.016081392	0.604938272	1.013887
Couple, Star_4, October	Stay_1	0.025270758	0.652542373	1.093673
Couple, Star_4, September	Stay_1	0.013127667	0.64516129	1.081302
Couple, Star_5, November	Stay_1	0.021988841	0.598214286	1.002618
Couple, Stay_2, November	Star_5	0.010173942	0.574074074	1.610685
Star_5, August	Couple	0.015425008	0.546511628	1.369425
Stay_1, Star_5, December	Couple	0.021332458	0.503875969	1.262591
Stay_2, August	Couple	0.012471283	0.535211268	1.341109
Star_3, Family with young children	Stay_1	0.013455858	0.706896552	1.184771
Family with young children	Stay_1	0.087298983	0.604545455	1.013229
Family with young children, July	Stay_1	0.010173942	0.673913043	1.12949
November, Family with young children	Stay_1	0.022973416	0.630630631	1.056948
Star_4, Family with young children	Stay_1	0.037085658	0.672619048	1.127321
Stay_2, Family with young children	Star_5	0.020347883	0.521008403	1.461798
December, Family with teens	Star_4	0.014440433	0.5	1.258051
Star_4, Family with teens	Stay_1	0.025927141	0.637096774	1.067785
Stay_2, Family with teens	Star_5	0.016081392	0.538461538	1.510766
November, Solo traveler	Stay_1	0.015425008	0.671428571	1.125326
October, Solo traveler	Star_4	0.016081392	0.521276596	1.311585
September, Solo traveler	Stay_1	0.010830325	0.6	1.005611
Star_3, Solo traveler	Stay_1	0.018706925	0.626373626	1.049813
Star_5, Solo traveler	Stay_1	0.01378405	0.666666667	1.117345
Group, Star_3	Stay_1	0.02559895	0.666666667	1.117345
November, Group	Stay_1	0.034131933	0.675324675	1.131856

รูปที่ 27 : ตารางสรุปผลลัพธ์ที่ได้จากการหาความสัมพันธ์ช่วงหตุการระบาด COVID-19

จากรูปที่ 27 แสดงผลลัพธ์ได้ว่าในกลุ่มนักท่องเที่ยวรูปแบบ Couple, Family with young children, Family with teens, Solo traveler เนื่องจากอยู่ในช่วงรัฐบาลมีโครงการเราเที่ยวด้วยกัน ระหว่างเดือน กรกฎาคม 2020 - เมษายน 2021 ซึ่งเป็นโครงการที่ช่วยออกค่าใช้จ่าย จึงทำให้มีการเข้าพักในรูปแบบที่ยกระดับมากขึ้นโดยเข้าพักในโรงแรมระดับ 3-5 ดาว จำนวนคืนที่เข้าพักเฉลี่ย 1-2 คืน และในส่วนของกลุ่มนักท่องเที่ยวรูปแบบ Group มีการเข้าพักในเดือน พฤศจิกายน เข้าพักในโรงแรมระดับ 3 ดาว จำนวนคืนที่เข้าพักเพียงแค่ 1 คืน ในเดือนพฤษจิกายน – ธันวาคม โดยที่ประชุมคณะรัฐมนตรี หรือ ครม. เห็นชอบประกาศวันหยุดเพื่อกระตุ้นเศรษฐกิจ ให้ประชาชนออกไปท่องเที่ยว โดยแบ่งเป็น 2 ช่วง รวมวันหยุดเสาร์ – อาทิตย์ คือ วันที่ 19 – 22 พฤศจิกายน 2020 และวันที่ 10 – 13 ธันวาคม 2020 อีกทั้งในเดือน ธันวาคมเป็นเดือนสุดท้ายของปี ทั้งนี้ในช่วงวันหยุดชดเชยวันที่ 11 ธันวาคม จะตรงกับวันศุกร์ ส่งผลให้เป็นวันหยุดยาว 4 วัน หยุดติดกับ วันที่พฤหัสบดี 10 ธันวาคม วันรัฐธรรมนูญ และวันหยุดประจำสัปดาห์ คือ วันเสาร์ที่ 12 ธันวาคม และวันอาทิตย์ที่

13 ธันวาคม ทำให้พบว่ากลุ่มนักท่องเที่ยวเกิดขึ้นในหลากหลายรูปแบบโดยมีกลุ่มนักท่องเที่ยวรูปแบบ Couple, , Family with young children, Family with teens และ Solo traveler

การเปรียบเทียบผลลัพธ์ของสองชุดข้อมูล (พฤติกรรมนักท่องเที่ยวชาวไทย)

ผลลัพธ์ของชุดข้อมูลก่อนการเกิดสถานการณ์ COVID-19 (ผลลัพธ์ข้อมูลชุดที่ 1) กฎความสัมพันธ์ที่มีค่าความสอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ทั้งหมด 73 กฎ โดยทำการคัดเลือกเฉพาะกลุ่ม itemset ที่มีความสัมพันธ์เกี่ยวกับชนิดกลุ่มนักท่องเที่ยวผลลัพธ์กฎความสัมพันธ์ที่ได้ 22 กฎ และผลลัพธ์ของชุดข้อมูลขณะเกิด COVID-19 (ผลลัพธ์ข้อมูลชุดที่ 2) กฎความสัมพันธ์ที่มีค่าความสอดคล้อง (Lift) ที่มีค่ามากกว่า 1 ได้ทั้งหมด 52 กฎ โดยทำการคัดเลือกเฉพาะกลุ่ม itemset ที่มีความสัมพันธ์เกี่ยวกับชนิดกลุ่มนักท่องเที่ยวผลลัพธ์กฎความสัมพันธ์ที่ได้ 32 กฎ จากผลลัพธ์ทั้งสองชุดข้อมูลพบว่ามีกฎความสัมพันธ์ที่เกิดขึ้นเหมือนกันทั้งหมด 9 กฎ แสดงดังรูปที่ 27

Premises	Conclusion
Business traveler	Stay_1
Star_4, Business traveler	Stay_1
Couple, July	Stay_1
Couple, November	Stay_1
Couple, September	Stay_1
Couple, Star_3	Stay_1
Star_3, Family with young children	Stay_1
Star_3, Solo traveler	Stay_1
Group, Star_3	Stay_1

รูปที่ 28 : ชุดข้อมูลกฎความสัมพันธ์ที่เกิดขึ้นเหมือนกันจาก ผลลัพธ์ข้อมูลชุดที่ 1 และผลลัพธ์ข้อมูลชุดที่ 2

จากรูปที่ 28 แสดงผลลัพธ์ได้ว่ากฎความสัมพันธ์ที่เกิดขึ้นเหมือนกันสามารถบ่งบอกได้ถึงพฤติกรรมของนักท่องเที่ยวได้ชัดเจนว่าเป็นพฤติกรรมที่จะเกิดขึ้นแน่นอน โดยพบว่ากลุ่มนักท่องเที่ยวกลุ่ม Business travelers มีพฤติกรรมที่จะเลือกเข้าพักในโรงแรมระดับ 4 ดาว และเลือกเข้าพักเพียง 1 คืน เพราะ Business travelers เป็นกลุ่มที่ต้องการเข้าพักโรงแรมที่ดีมีสิ่งอำนวยความสะดวกที่เหมาะต่อการทำธุรกิจ เช่น ห้องประชุม โซนในการนั่งทำงาน อินเตอร์เน็ตที่รวดเร็ว รูมเซอร์วิส เป็นต้น [30] ส่วนของกลุ่มนักท่องเที่ยวในรูปแบบ Couple นิยมเข้าพักในเดือน กรกฎาคม กันยายน และพฤศจิกายน และจะเลือกเข้าพักในโรงแรมระดับ 3 ดาว เพราะเมื่อพูดถึงการเที่ยวทะเล ทะเลของภาคใต้มีความสวยงามมากที่สุดในประเทศ และในช่วงเดือน กรกฎาคม กันยายน และพฤศจิกายน สภาพอากาศ ของภาคเหนือและภาคกลาง จะอยู่ในช่วงฤดูหนาวและฤดูฝน เมื่ออากาศหนาวทำให้

ผู้คนอยากจะมาเที่ยวทะเลที่ภาคใต้ เพราะสภาพอากาศของภาคใต้จะอยู่ในช่วงฤดูร้อน เหมาะแก่การเที่ยวทะเล ส่วนของกลุ่ม Family with young children, Solo traveler และ Group นิยมเข้าพักในโรงแรมระดับ 3 ดาว เช่นกัน เพราะโรงแรมระดับ 3 ดาว ถือว่าเป็นโรงแรมราคาประหยัดที่มีสิ่งอำนวยความสะดวกครบครันและเป็นที่นิยมในหมู่นักท่องเที่ยวที่ต้องการท่องเที่ยวแล้วต้องการประหยัดค่าใช้จ่าย

4.3 การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์

4.3.1 ข้อมูลเชิงลึกของบทวิจารณ์ออนไลน์ที่ได้รวบรวม

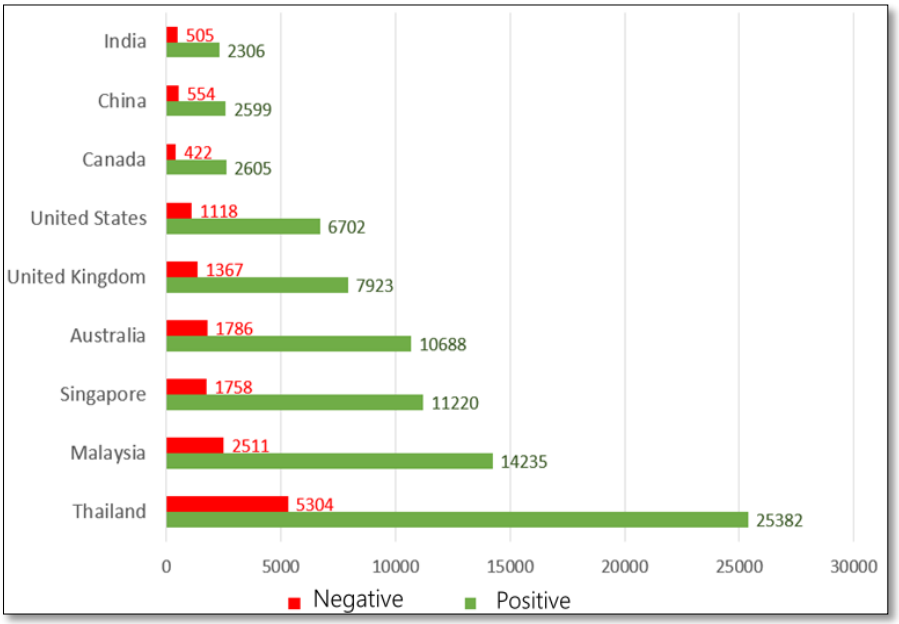
ข้อมูลบทวิจารณ์ออนไลน์ในเว็บไซต์ Agoda โดยมีจำนวนโรงแรม 427 โรงแรม และบทวิจารณ์จำนวน 246,532 บทวิจารณ์ได้รวบรวมตั้งแต่ปี 2009-2020

การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ใช้ภาษา R ในโปรแกรม R Studio โดยใช้ไลบรารี เพื่อวิเคราะห์อารมณ์และความรู้สึกของข้อมูลและการวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์รวมทั้งการทดสอบประสิทธิภาพของ Model LDA ว่ามีความเหมาะสมที่จะนำมาใช้ในขั้นตอนการจัดกลุ่มคำศัพท์ในการวิจัยครั้งนี้หรือไม่และนำผลลัพธ์ทั้งหมดไปแสดงในรูปแบบ Dashboard โดยวิธีทำการแบ่งเป็น 4 ข้อ ดังนี้

1. การวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ (อ้างอิงจากข้อ 3.2.6.1)
2. การวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์ (อ้างอิงจากข้อ 3.2.6.2)
3. การจัดกลุ่มคำศัพท์ 2 วิธีเพื่อทำการเปรียบเทียบผลลัพธ์และนำผลลัพธ์ที่เหมาะสมที่สุดมาใช้ (อ้างอิงจากข้อ 3.2.6.3)
4. การแสดงผลลัพธ์ (อ้างอิงจากข้อ 3.2.6.4)

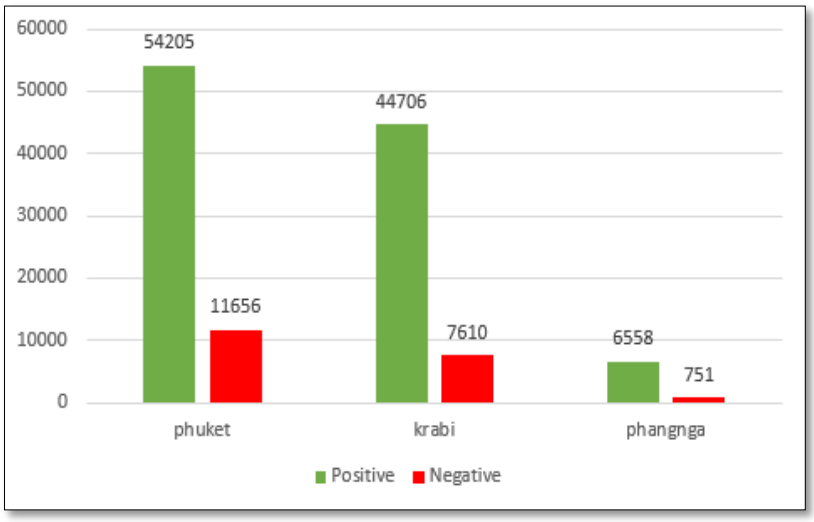
4.3.2 ผลการทดลองการวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ (อ้างอิงจากข้อ 3.2.6.1)

จากวิเคราะห์ด้วย ไลบรารี setimentr ในภาษา R เพื่อวิเคราะห์อารมณ์และความรู้สึกของประโยคทั้งหมดในแต่ละบทวิจารณ์ ผลลัพธ์แสดงดังรูปที่ 28 29 30 และ 31



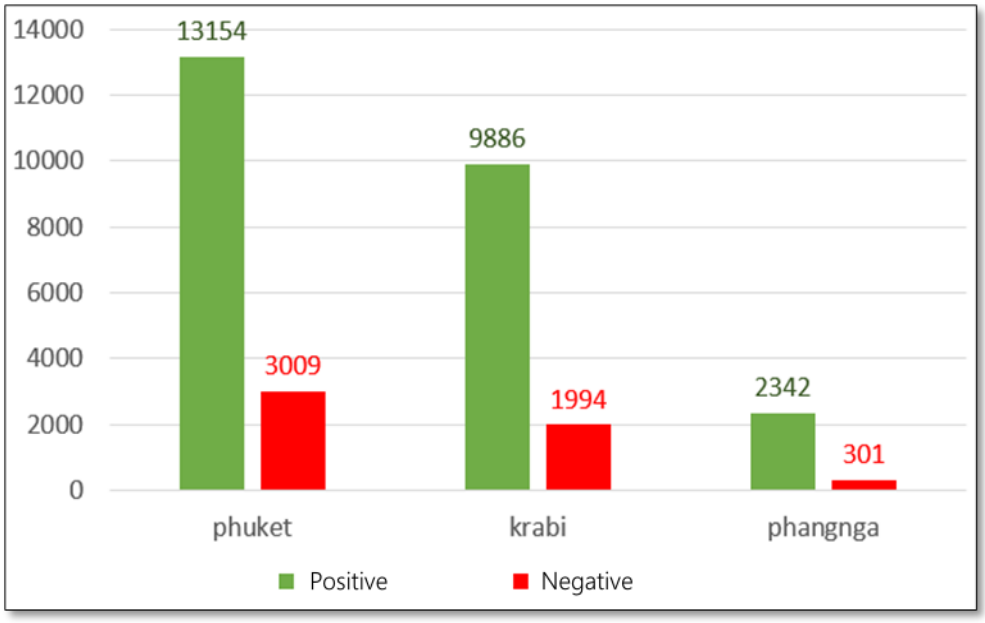
รูปที่ 29 : 10 อันดับแรกของประเทศที่มีจำนวนบทวิจารณ์เชิงบวกและเชิงลบเยอะที่สุด

จากรูปที่ 29 แสดงให้เห็นถึงจำนวนบทวิจารณ์เชิงบวก (Positive) และเชิงลบ (Negative) ของทุกประเทศโดยมีประเทศ Thailand เป็นอันดับแรก มีจำนวนบทวิจารณ์เชิงบวกจำนวน 25,382 และเชิงลบ จำนวน 5,304 อันดับสองเป็นประเทศ Malaysia มีบทวิจารณ์เชิงบวกจำนวน 14,235 และเชิงลบจำนวน 2,511 อันดับที่สามเป็นประเทศ Singapore มีจำนวนบทวิจารณ์เชิงบวก 11,220 และเชิงลบจำนวน 1,758 และประเทศที่มีจำนวนน้อยที่สุดคือประเทศ India มีจำนวนบทวิจารณ์เชิงบวกจำนวน 2,306 และเชิงลบจำนวน 505



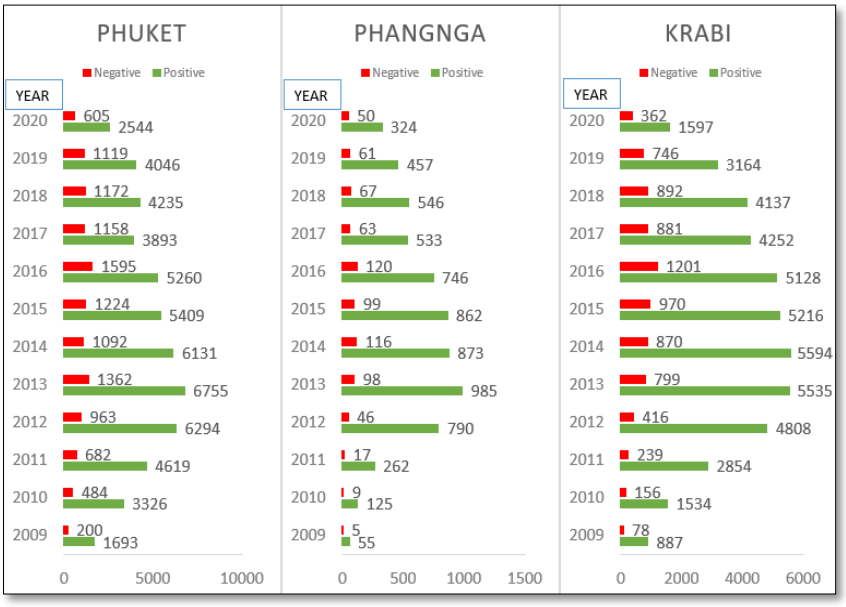
รูปที่ 30 : จำนวนบทวิจารณ์เชิงบวกและเชิงลบทั้งหมดโดยแบ่งเป็น 3 จังหวัด

จากรูปที่ 30 แสดงให้เห็นถึงผมลัพธ์ของจำนวนบทวิจารณ์เชิงบวกและเชิงลบทั้งหมดโดยแบ่งเป็น 3 จังหวัด พบว่าจังหวัดภูเก็ตมีจำนวนเยอะที่สุดจำนวนบทวิจารณ์เชิงบวกจำนวน 54,205 และบทวิจารณ์เชิงลบจำนวน 11,656 รองลงมาคือจังหวัดกะบี่มีจำนวนบทวิจารณ์เชิงบวกจำนวน 44,706 และบทวิจารณ์เชิงลบจำนวน 7,610 น้อยที่สุดคือจังหวัดพังงามีจำนวนบทวิจารณ์เชิงบวกจำนวน 6,558 และบทวิจารณ์เชิงลบจำนวน 751



รูปที่ 31 : จำนวนบทวิจารณ์เชิงบวกและเชิงลบเฉพาะประเทศไทยทั้งหมดโดยแบ่งเป็น 3 จังหวัด

จากรูปที่ 31 แสดงให้เห็นถึงผมลัพธ์ของจำนวนบทวิจารณ์เชิงบวกและเชิงลบเฉพาะประเทศไทยโดยแบ่งเป็น 3 จังหวัด พบว่าจังหวัดภูเก็ตมีจำนวนเยอะที่สุดจำนวนบทวิจารณ์เชิงบวกจำนวน 13,154 และบทวิจารณ์เชิงลบจำนวน 3,009 รองลงมาคือจังหวัดกะบี่มีจำนวนบทวิจารณ์เชิงบวกจำนวน 9,886 และบทวิจารณ์เชิงลบจำนวน 1,994 น้อยที่สุดคือจังหวัดพังงามีจำนวนบทวิจารณ์เชิงบวกจำนวน 2,342 และบทวิจารณ์เชิงลบจำนวน 301



รูปที่ 32 : จำนวนบทวิจารณ์ทั้งหมดเชิงบวก และเชิงลบทั้งหมด 3 จังหวัดในแต่ละปี

จากรูปที่ 32 แสดงให้เห็นถึงผมลัพธ์ของจำนวนบทวิจารณ์เชิงบวกและเชิงลบทั้งหมดใน 3 จังหวัดในแต่ละปี พบว่าจังหวัดภูเก็ตปีที่มีจำนวนบทวิจารณ์เชิงบวกมากที่สุดในปี 2013 จำนวน 6,755 บทวิจารณ์ จำนวนบทวิจารณ์เชิงลบมากที่สุดในปี 2016 และในทุก ๆ ปีสังเกตได้ว่าจะมีบทวิจารณ์เชิงลบเฉลี่ยอยู่ที่ 21.35% จังหวัดพังงาปีที่มีจำนวนบทวิจารณ์เชิงบวกมากที่สุดในปี 2013 จำนวน 985 บทวิจารณ์ จำนวนบทวิจารณ์เชิงลบมากที่สุดในปี 2016 จำนวน 120 บทวิจารณ์ และในทุก ๆ ปีสังเกตได้ว่าจะมีจำนวนบทวิจารณ์เชิงลบเฉลี่ยอยู่ที่ 10.91% จังหวัดกระบี่ที่มีจำนวนบทวิจารณ์เชิงบวกมากที่สุดในปี 2014 จำนวน 5,594 บทวิจารณ์ จำนวนบทวิตารณ์เชิงลบจำนวนมากที่สุดในปี 1,201 บทวิจารณ์ และในทุก ๆ ปีสังเกตได้ว่าจะมีจำนวนบทวิจารณ์เชิงลบฉลี่ยอยู่ที่ 16.38%

4.3.3 ผลการทดลองการวิเคราะห์อารมณ์และความรู้สึกของคำศัพท์จากบทวิจารณ์ (อ้างอิงจากข้อ 3.2.6.2)

ผลลัพธ์ที่ได้จากการแยกคำ วิเคราะห์อารมณ์และความรู้สึกของคำ และหาชนิดของคำ โดยใช้ โดยใช้แพ็คเกจ setimentr และ POStagObject แสดงผลลัพธ์ดังรูปที่ 33

word_i	word	pos	hotel_id	hotel	comment_i	word_n	ave_sentimen	g_pos
1	The	DT	275	krabi	1	1	0	Det
2	Room	NNP	275	krabi	1	2	0	Noun
3	the	DT	275	krabi	1	3	0	Det
4	private	JJ	275	krabi	1	4	0	Adj
5	access	NN	275	krabi	1	5	0	Noun
6	to	TO	275	krabi	1	6	0	To
7	the	DT	275	krabi	1	7	0	Det
8	pool	NN	275	krabi	1	8	0	Noun
9	the	DT	275	krabi	1	9	0	Det
10	best	JJS	275	krabi	1	10	0.5	Adj
11	free	JJ	275	krabi	1	11	0.5	Adj
12	breakfast	NN	275	krabi	1	12	0.6	Noun
13	I	PRP	275	krabi	1	13	0	Pron
14	get	VBP	275	krabi	1	14	0	Verb
15	it	PRP	275	krabi	1	15	0	Pron

รูปที่ 33 : ผลลัพธ์ที่ได้จากการแยกแยกคำศัพท์และหาชนิดของคำ

จากรูปที่ 33 เป็นตัวอย่างผลลัพธ์ที่ได้จากการแยกคำศัพท์และหาชนิดของคำจากบทวิจารณ์ 15 คำแรกซึ่งผลลัพธ์ทั้งหมดที่ได้มีจำนวน 6,365,012 คำ จากนั้นเตรียมข้อมูลคำศัพท์เข้าสู่ขั้นตอนการจัดหมวดหมู่โดยนำคำศัพท์ทั้งหมดมาทำการรวมคำศัพท์ที่เหมือนกันโดยใช้ภาษา R โดยผลลัพธ์ที่ได้จากการรวมคำศัพท์มีจำนวน 66,832 คำ

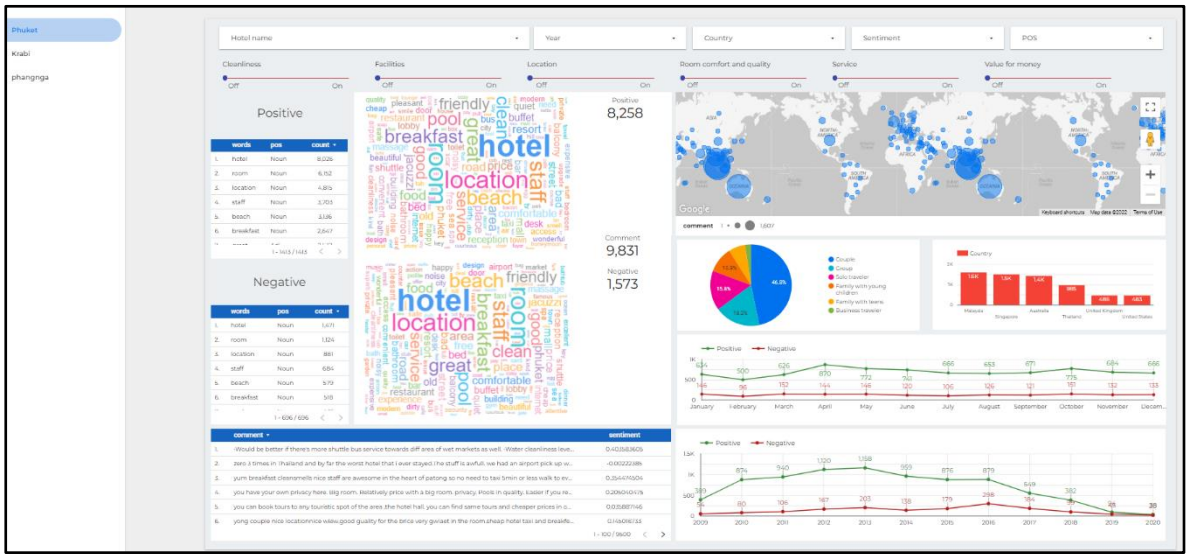
4.3.4 ผลการทดลองการจัดกลุ่มคำศัพท์ 2 วิธีเพื่อทำการเปรียบเทียบผลลัพธ์และนำผลลัพธ์ที่เหมาะสมที่สุดมาใช้ (อ้างอิงจากข้อ 3.2.6.3)

ผลลัพธ์ของการจัดกลุ่มคำศัพท์ผู้วิจัยได้ทำการจัดรูปแบบออกเป็น 2 แบบ

จากการผลการทดลองพบว่าการจัดกลุ่มรูปแบบที่ 1 Model LDA ไม่เหมาะสมที่จะนำมาใช้ในการจัดกลุ่มในการวิจัยครั้งนี้เนื่องจากผลลัพธ์ที่ได้ในแต่ละกลุ่มมีคำศัพท์ที่คล้ายคลึงกันมากเกินไปทำให้มองไม่เห็นภาพหัวข้อที่ต้องการมากำหนดชื่อในแต่ละกลุ่มทางผู้วิจัยจึงเลือกใช้วิธีการจัดกลุ่มรูปแบบที่ 2 การป้อนข้อมูลโดยผู้วิจัย (Manual) โดยการป้อนข้อมูลครั้งนี้ผู้วิจัยคำนึงถึงความหมายของคำศัพท์และความเป็นไปได้ที่เหมาะสมที่สุด เพื่อใช้ผลลัพธ์เป็นคลังข้อมูลในการจัดกลุ่มแสดงผลลัพธ์ในการวิเคราะห์

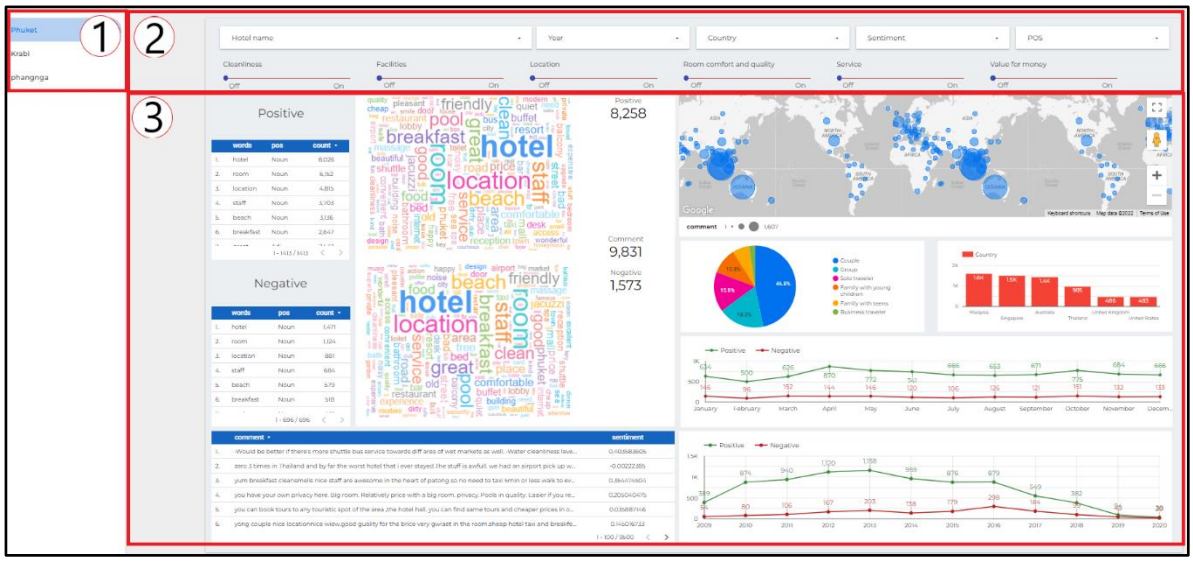
4.3.5 ผลการแสดงผลลัพธ์ (อ้างอิงจากข้อ 3.2.6.4)

ผลลัพธ์ทั้งหมดที่ได้มาแสดงผลลัพธ์แนวโน้มของการวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์ในรูปแบบ Dashboard โดยใช้ Data Studio ในการแสดงผลลัพธ์ จากนั้นทำการคัดเลือกข้อมูลโรงแรมที่มีจำนวนบทวิจารณ์มากที่สุด 5 อันดับแรกในแต่ละจังหวัดมาแสดงผลลัพธ์ โดยแสดงผลลพธ์ดังรูปที่ 34



รูปที่ 34 : การแสดงผลลัพธ์ทั้งหมดในรูปแบบ Dashboard โดยใช้ Data Studio

จากรูปที่ 34 เป็นผลลัพธ์ที่ได้นำมาแสดงในรูปแบบ Dashboard โดยรูปแบบของการแสดงผลลัพธ์ แบ่งออกเป็น 3 ส่วนดังรูปที่ 35



รูปที่ 35 : การแสดงผลลัพธ์ทั้งหมดในรูปแบบ Dashboard แบ่งออกเป็น 3 ส่วน

จากรูปที่ 35 เป็นรูปแบบของการหน้าจอแสดงผลลัพธ์ โดยแบ่งออกเป็น 3 ส่วนมีรายละเอียดแต่ละส่วนดังนี้ดังนี้

1. ปุ่มสำหรับแสดงผลลัพธ์ของแต่ละจังหวัดโดยมีให้เลือก 3 จังหวัด ภูเก็ต (Phuket) กระบี่ (Krabi) พังงา (Phangnga)
2. ปุ่มสำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลที่ต้องการโดยมีรายละเอียดดังนี้
   2.1 ปุ่ม Hotel name สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลของชื่อโรงแรมที่เลือก
   2.2 ปุ่ม Year สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลของปีที่ต้องการ
   2.3 ปุ่ม Country สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลของประเทศที่ต้องการ
   2.4 ปุ่ม Sentiment สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลของบทวิจารณ์ที่มีความรู้สึกเป็นเชิงบวก (Positive) หรือเชิงลบ (Negative)
   2.5 ปุ่ม POS สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลชนิดของคำ (Part of speech)
   2.6 ปุ่ม Switch Off และ Switch On สำหรับกรองข้อมูลให้แสดงเฉพาะข้อมูลที่เปิดหรือปิด มีทั้งหมด 6 ปุ่มได้แก้
      2.6.1 ปุ่ม Cleanliness เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับความสะอาด
      2.6.2 ปุ่ม Facilities เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับสิ่งอำนวยความสะดวก
      2.6.3 ปุ่ม Location เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับตำแหน่งหรือพื้นที่ใกล้เคียง
      2.6.4 ปุ่ม Room comfort and quality เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับความสะดวกสบายและคุณภาพของห้องพัก
      2.6.5 ปุ่ม Service เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับการบริการ
      2.6.6 ปุ่ม Value for money เปิด-ปิด สำหรับแสดงบทวิจารณ์ที่เกี่ยวข้องกับความคุ้มค่าของเงิน
3. หน้าจอสำหรับแสดงค่าผลลัพธ์ออกมาในรูปแบบต่าง ๆ โดยมีรายละเอียดดังนี้
   3.1 แสดงผลลัพธ์คำศัพท์ที่ได้จากบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและรูปแบบ Word cloud ดังรูปที่ 36

รูปที่ 36 : การแสดงผลลัพธ์คำศัพท์ที่ได้จากบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและรูปแบบ Word cloud

จากรูปที่ 36 เป็นการแสดงผลลัพธ์คำศัพท์ที่ได้จากบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและรูปแบบ Word cloud รวมไปถึงจำนวนคำศัพท์ ชนิดของคำ จำนวนบทวิจารณ์เชิงบวกเชิงลบ และจำนวน

3.2 บทวิจารณ์ทั้งหมดแสดงผลลัพธ์บทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและผลลัพธ์ของค่า Sentiment ที่ได้จากการวิเคราะห์บทวิจารณ์ แสดงดังรูปที่ 37

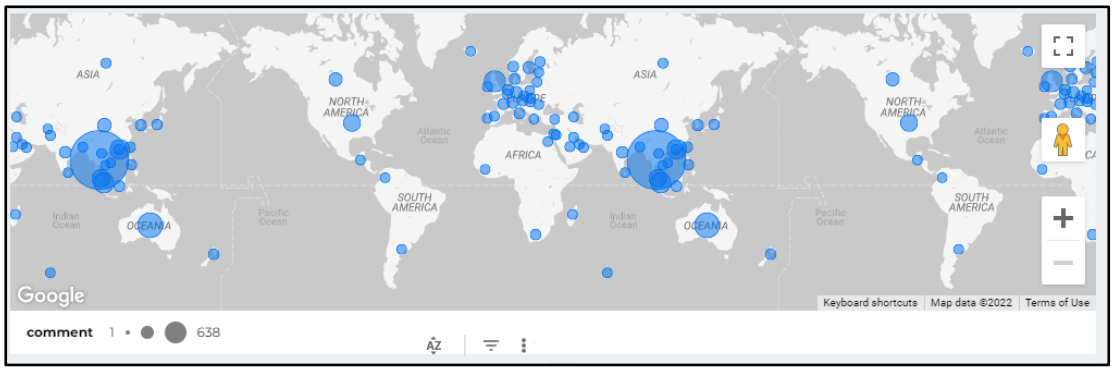
	comment	sentiment
1.	worth to stay the beachfront suite is same as the picture,it's really nice!but u've to share the living room,kitchen & t...	0.451911513
2.	wonderful room, great location, good service, nice beach if asked to use one word to describe this hotel, i will say "...	0.329506096
3.	we had serious problems with mosquitos. I know that this is Thailand, but whenever we returned the room was ful...	-0.131105073
4.	villa design, private plunge pool, great location We enjoyed our stay at Briza and loved our room very much, especi...	0.581078535
5.	view, staff, clean the stay was perfect, staff was very helpful during the stay. definitely worth what we pay for.	0.61967252
6.	view, spacious, nice staff i stayed in the 3-bedrooms beach front villa with my family, they were all very happy with i...	0.421689439

1 - 100 / 1920

รูปที่ 37 : การแสดงประโยคบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและผลลัพธ์ค่าของ Sentiment

จากรูปที่ 37 เป็นการแสดงประโยคบทวิจารณ์ทั้งเชิงบวกและเชิงลบในรูปแบบตารางและผลลัพธ์ของค่า Sentiment ที่ได้จากการวิเคราะห์บทวิจารณ์

3.3 แผนที่แสดงประเทศของนักท่องเที่ยว แสดงดังรูปที่ 38



รูปที่ 38 : การแสดงผลลัพธ์ประเทศของนักท่องเที่ยว

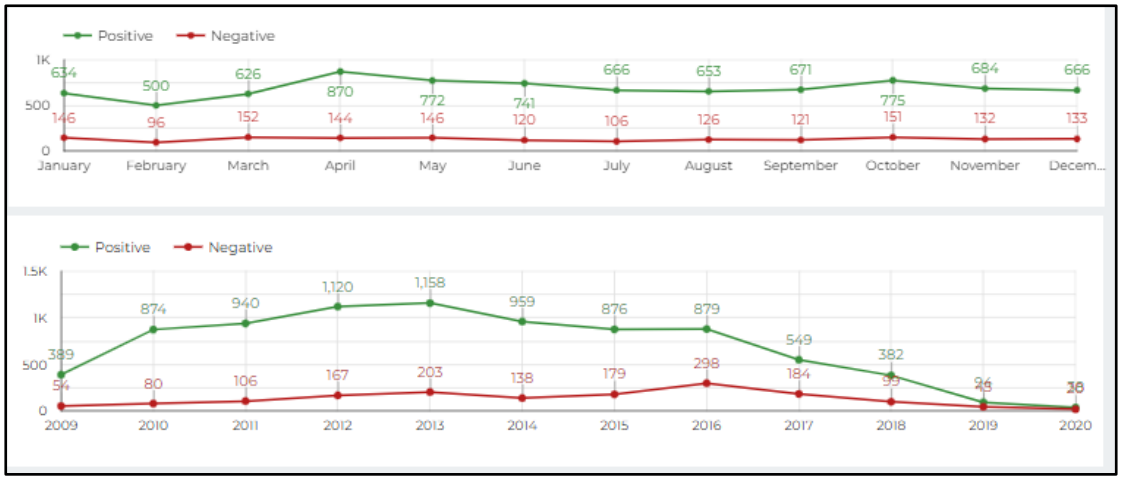
จากรูปที่ 38 เป็นการแสดงผลลัพธ์ของประเทศของนักท่องเที่ยวรูปแบบแผนที่โดยขนาดของวงกลมจะบ่งบอกถึงจำนวนปริมาณของนักท่องเที่ยวที่มาท่องเที่ยวประเทศไทย

3.4 จำนวนบทวิจารณ์ทั้งหมดแสดงผลลัพธ์บในรูปแบบแผนภูมิวงกลม (Pie Chart) และผลลัพธ์ของจำนวนบทวิจารณ์โดยแบ่งตามรูปแบบนักท่องเที่ยวแสดงผลลัพธ์ในรูปแบบแผนภูมิแท่ง (Bar Charts) แสดงดังรูปที่ 39

รูปที่ 39 : การแสดงผลลัพธ์รูปแบบแผนภูมิวงกลมและรูปแบบแผนภูมิแท่ง

จากรูปที่ 39 เป็นการแสดงผลลัพธ์ของจำนวนบทวิจารณ์โดยแบ่งตามรูปแบบนักท่องเที่ยวแสดงผลลัพธ์บในรูปแบบแผนภูมิวงกลมโดยมีค่าเป็นเปอร์เซ็นและผลลัพธ์ของ และจำนวนบทวิจารณ์ทั้งหมดแสดงผลลัพธ์ในรูปแบบแผนภูมิแท่ง

3.5 จำนวนบทวิจารณ์เชิงบวกและเชิงลบรายเดือนและรายปีแสดงผลลัพธ์ในรูปแบบกราฟเส้น แสดงดังรูปที่ 40



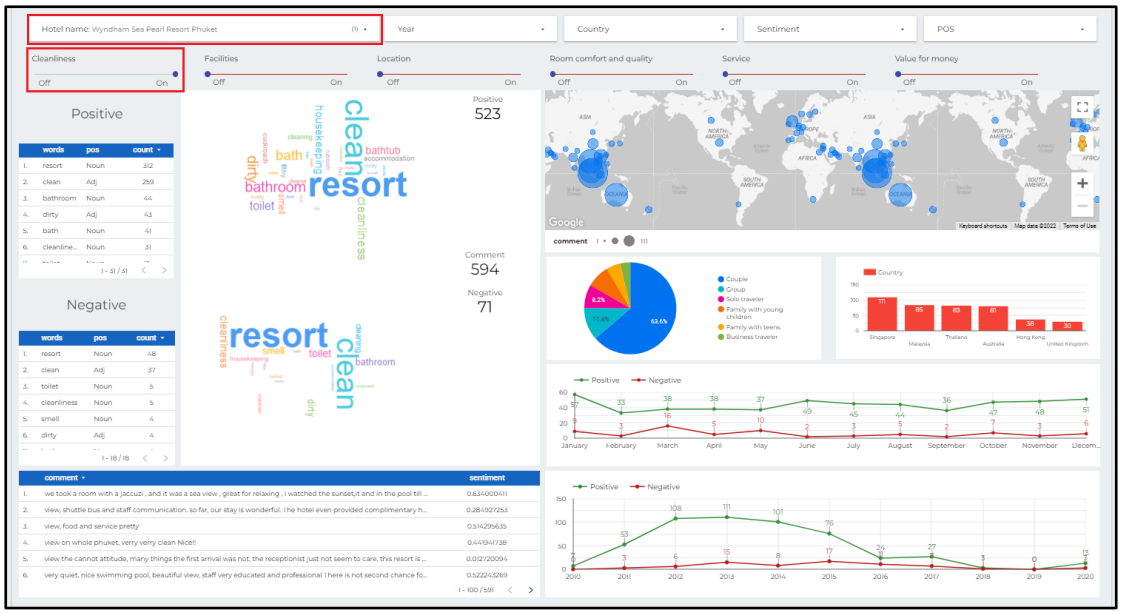
รูปที่ 40 : การแสดงผลลัพธ์จำนวนบทวิจารณ์รายเดือนและรายปีแสดงผลลัพธ์ในรูปแบบกราฟเส้น

จากรูปที่ 40 เป็นการแสดงผลลัพธ์จำนวนบทวิจารณ์เชิงบวกและเชิงลบรายเดือนและรายปีแสดงผลลัพธ์ในรูปแบบกราฟเส้น

ตัวอย่างผลลัพธ์จาก Dashboard

ผลลัพธ์ที่ได้จากการวิเคราะห์อารมณ์และความรู้สึกจากข้อความและประโยคจากบทวิจารณ์ออนไลน์และการจัดหมวดหมู่ของบทวิจารณ์และคำศัพท์ต่างๆ จากบทวิจารณ์ที่ผ่านการวิเคราะห์ความรู้สึก แสดงผลลัพธ์ในรูปแบบ Dashboard เพื่อแสดงผลลัพธ์ทั้งหมดที่ได้

ตัวอย่าง การกรองข้อมูลให้แสดงเฉพาะข้อมูลโรงแรมที่ต้องการและกรองข้อมูลหมวดหมู่ที่ต้องการแสดงข้อมูล แสดงดังรูปที่ 41



รูปที่ 41 : การแสดงผลลัพธ์โดยการกรองข้อมูลในส่วนของชื่อโรงแรมและหมวดหมู่

จากรูปที่ 41 เป็นการแสดงผลลัพธ์ที่ได้จากการกรองข้อมูลในส่วนของชื่อโรงแรมให้แสดงเฉพาะข้อมูลโรงแรมที่เลือก และทำการกรองข้อมูลโดยการปรับปุ่มเป็น Switch ในหมวดหมู่ Cleanliness ให้อยู่ในรูปแบบ On ผลลัพธ์ที่ได้พบว่ารูบแบบนักท่องเที่ยวที่มากที่สุดคือรูปแบบ Couple จำนวน 63.6% ประเทศที่มากที่สุดคือ Singapore จำนวน 111 บทวิจารณ์ และมีจำนวนบทวิจารณ์ทั้งหมดจำนวน 594 บทวิจารณ์ แบ่งได้เป็นบทวิจารณ์เชิงบวกจำนวน 523 บทวิจารณ์ และบทวิจารณ์เชิงลบจำนวน 71 บทวิจารณ์ ในส่วนของคำศัพท์เชิงบวกที่เด่นชัดคือคำว่า resort clean bathtub housekeeping bathroom และ toilet คำศัพท์เหล่านี้สามารถบ่งบอกถึงความรู้สึกเชิงบวกของผู้วิจารณ์ที่มีต่อโรงแรมว่าเป็นโรงแรมที่มีสะอาดในส่วนห้องน้ำ (bathroom) อ่างอาบน้ำ (bathtub) สุขา (toilet) และรวมไปถึงแม่บ้านทำความสะอาด (housekeeping) ในส่วนของคำศัพท์เชิงลบที่เด่นชัดคือคำว่า clean toilet smells bad และ dirty คำศัพท์เหล่านี้สามารถบ่งบอกความรู้สึกในเชิงลบเหมือนกันแต่มีการพูดถึงในปริมาณที่น้อยกว่าเชิงบวกเป็นอย่างมากทำให้ทราบว่าสิ่งที่ถูกพูดถึงในบทวิจารณ์เชิงบวกก็มีส่วนบทวิจารณ์เชิงลบที่ไม่ชอบเหมือนกัน

บทที่ 5

สรุปผล

งานวิจัยฉบับนี้มีจุดประสงค์ คือ เพื่อศึกษาแนวโน้มความรู้สึกจากบทวิจารณ์ออนไลน์และวิเคราะห์ความสัมพันธ์ของพฤติกรรมนักท่องเที่ยวที่เข้าพักในช่วง 12 ปีที่ผ่านมา โดยรวบรวมข้อมูลจากเว็บไซต์ agoda สำหรับโรงแรมที่เข้าร่วมโครงการเราเที่ยวด้วยกันพบว่า บทวิจารณ์ทั้งหมดใน 3 จังหวัดในในช่วงปี 2013-2014 มีจำนวนบทวิจารณ์มากที่สุด และเริ่มลดลงตั้งแต่ปี 2014 ไปจนถึงปี 2019 เนื่องจากในช่วงปีตั้งแต่ 2019 เกิดสถานการณ์ Covid-19 ระบาดทำให้ลดลงอย่างเห็นได้ชัด และผลลัพธ์ความสัมพันธ์ของพฤติกรรมนักท่องเที่ยวผลลัพธ์จากทั้ง 2 algorithm จะได้กฎความสัมพันธ์เท่ากันคือทั้งหมด 295 กฎ แต่ระยะเวลาจาก FP-Growth ซึ่งประมวลด้วย RapidMiner เร็วกว่า Apriori algorithm ซึ่งประมวลผลด้วย R Studio ถึง 4.79 เท่า จากนั้นวิเคราะห์ข้อมูลที่เกี่ยวข้องเฉพาะกลุ่มของนักท่องเที่ยวพบว่า นักท่องเที่ยวต่างชาติที่เด่นชัดจะเป็น Oceania, Europe และ Asia โดยประเภทการท่องเที่ยวที่พบส่วนใหญ่ จะมี 5 ประเภท คือ business traveler, group, couple, family with teens และ family with young children ซึ่งลักษณะการท่องเที่ยวแบบ solo traveler หรือ backpacker ยังไม่เด่นชัดในการท่องเที่ยวซ้ำหรือปริมาณที่มาก นอกจากนี้นักท่องเที่ยวส่วนใหญ่จะพักในโรงแรมระดับ 4-5 ดาว และจำนวนคืนที่พักจะอยู่ระหว่าง 1-3 คืน โดยกลุ่มของนักท่องเที่ยวจากแต่ละกลุ่มประเทศนั้น จะเข้าพักในช่วงเดือนที่แตกต่างกัน ซึ่งเป็นสิ่งที่ดีที่ทำให้ประเทศไทย โดยเฉพาะจังหวัดภูเก็ต พังงาและกระบี่ มีนักท่องเที่ยวหมุนเวียนเข้ามาตลอดทั้งปี และแสดงผลลัพธ์การวิเคราะห์เพื่อหาความรู้สึกจากบทวิจารณ์ออนไลน์ของนักท่องเที่ยวและจัดหมวดหมู่บทวิจารณ์และคำศัพท์ต่าง ๆ รวมไปถึงการวิเคราะห์ความรู้สึกจากบทวิจารณ์ออนไลน์ ในการจัดหมวดหมู่กลุ่มคำศัพท์ 6 หมวดหมู่ เพื่อต้องการจัดหมวดหมู่ให้กับบทวิจารณ์โดยวิธีที่เลือกใช้ในการจัดกลุ่มหมวดหมู่คือวิธีรูปแบบที่ 2 เป็นการป้อนข้อมูลโดยผู้วิจัย (Manual) เนื่องจากวิธีรูปแบบที่ 1 การใช้ Model LDA ผลลัพธ์ที่ได้ในแต่ละ Topics ผลลัพธ์แต่ละกลุ่มมีคำศัพท์ที่คล้ายคลึงกันมากเกินไปทำให้มองไม่เห็นภาพหัวข้อที่ต้องการมากำหนดชื่อในแต่ละกลุ่มทางผู้วิจัยจึงเลือกใช้วิธีการจัดกลุ่มรูปแบบที่ 2 จากการนำข้อมูลที่ได้จัดกลุ่มนำมาใช้เป็นฐานข้อมูลที่ได้จากการวิเคราะห์ความรู้สึกของคำและประโยค โดยผลลัพธ์ที่ได้นั้นสามารถบ่งบอกถึงคำศัพท์หรือบทวิจารณ์ที่สามารถสะท้อนถึงสภาพแวดลอบแต่ละโรงแรมและโดยรวมแต่ละจังหวัด

สามารถบ่งบอกได้ว่าเด่นและด้อยในด้านไหนบ้าง ผ่านการจัดหมวดหมู่และคำศัพท์ต่าง ๆ ที่ได้จากการวิเคราะห์ความรู้สึกจากบทวิจารณ์ และสามารถบ่งบอกได้ถึงรูปแบบหรือประเทศของนักท่องเที่ยวที่นิยม และสามารถนำข้อมูลเหล่านี้ไปช่วยในการตัดสินใจในการพัฒนาหรือปรับปรุงบริการให้ดียิ่งขึ้นได้

ภาคผนวก

ภาคผนวก ก

1.ผลลัพธ์ของชุดข้อมูลบทวิจารณ์ออนไลน์ตั้งแต่ปี 2009-2020 กฎความสัมพันธ์ได้ทั้งหมด 355 กฎ จากการประมวลผลโปรแกรม Rapid Miner Studio ด้วยต้นแบบการจำลอง Associate Rules

No	Premises	Conclusion	Support	Confidence	Lift
1	Star_4, May	Couple	0.021596	0.500832	1.010521
2	September	Couple	0.038785	0.502322	1.013529
3	Stay_2, 2014	Couple	0.016233	0.502963	1.014822
4	Stay_3, Solo traveler	Asia	0.010878	0.503504	0.803606
5	Star_4, November	Couple	0.018711	0.504296	1.01751
6	Asia, Stay_3, Star_5	Couple	0.013635	0.50472	1.018366
7	Stay_1, 2012	Couple	0.013109	0.504911	1.018752
8	February	Asia	0.036219	0.505	0.805993
9	May	Couple	0.045344	0.505553	1.020048
10	Star_4, January	Couple	0.01966	0.505636	1.020216
11	2013	Couple	0.062605	0.505729	1.020404
12	Stay_2, August	Couple	0.010678	0.507576	1.024129
13	Stay_4	Couple	0.054022	0.507904	1.024792
14	Stay_5	Star_4	0.024967	0.509348	1.097497
15	Couple, Stay_5	Star_4	0.013308	0.509924	1.098738
16	Stay_3, Star_4.5	Couple	0.010368	0.510797	1.030628
17	Stay_3, Family with young children	Star_4	0.019683	0.510858	1.100752
18	Asia, Star_5	Couple	0.059544	0.510906	1.030848
19	Asia, Stay_3, 2016	Star_4	0.010089	0.511102	1.101277
20	Stay_3, May	Couple	0.010376	0.511591	1.032231
21	Business traveler	Stay_1	0.013635	0.511663	1.955607
22	Stay_2, January	Couple	0.010352	0.512426	1.033915

No	Premises	Conclusion	Support	Confidence	Lift
23	Stay_6	Star_4	0.013189	0.514135	1.107811
24	Star_4, March	Couple	0.017396	0.515223	1.039559
25	January	Couple	0.044491	0.515275	1.039663
26	Stay_3, June	Star_4	0.010041	0.516393	1.112678
27	Star_4, 2012	Couple	0.024266	0.516452	1.042038
28	November	Couple	0.043304	0.516638	1.042414
29	Stay_3, 2013	Couple	0.014599	0.51766	1.044477
30	Couple, Oceania	Star_4	0.03355	0.519112	1.118536
31	Star_4, March	Asia	0.017556	0.519943	0.829843
32	Star_4, 2012	Asia	0.024481	0.521031	0.83158
33	Stay_3, 2012	Couple	0.012479	0.522697	1.054639
No	Premises	Conclusion	Support	Confidence	Lift
34	Asia, 2011	Couple	0.017333	0.522711	1.054667
35	Asia, Stay_3, Family with young children	Star_4	0.016847	0.523008	1.126931
36	Asia, Stay_2, Star_5	Couple	0.019851	0.523539	1.056338
37	Stay_3, 2017	Star_4	0.010798	0.523975	1.129015
38	Stay_2, November	Couple	0.011252	0.525493	1.060281
39	Stay_2, March	Couple	0.010312	0.528811	1.066975
40	Asia, 2020	Stay_1	0.018177	0.529234	2.022767
41	Stay_3, 2018	Star_4	0.010208	0.529339	1.140572
42	Stay_4	Asia	0.056413	0.530381	0.846503
43	Asia, 2010	Couple	0.011714	0.530878	1.071145
44	March	Couple	0.041431	0.531976	1.073361
45	Stay_5	Couple	0.026099	0.532434	1.074285
46	Star_4, Stay_5	Couple	0.013308	0.533035	1.075499
47	2012	Asia	0.056612	0.533453	0.851406
48	2012	Couple	0.056684	0.534129	1.077706
49	Stay_6	Couple	0.013707	0.534327	1.078105

No	Premises	Conclusion	Support	Confidence	Lift
50	Star_4, North America	Couple	0.021038	0.535063	1.079589
51	Couple, Star_4, 2013	Asia	0.013468	0.535149	0.854112
52	Star_4.5	Couple	0.040315	0.535911	1.081301
53	Asia, Business traveler	Stay_1	0.012001	0.536516	2.050598
54	Asia, Stay_1, Star_5	Couple	0.018337	0.536614	1.082719
55	Star_5, 2013	Couple	0.014121	0.53697	1.083437
56	Oceania	Star_4	0.05913	0.537058	1.157205
57	Star_4, February	Couple	0.016599	0.539637	1.088819
58	Star_4, Stay_4	Asia	0.02778	0.540633	0.862864
59	Stay_7	Star_4	0.014073	0.54122	1.166172
60	Star_4, 2010	Couple	0.014583	0.54142	1.092416
61	March	Asia	0.042228	0.542208	0.865379
62	February	Couple	0.038984	0.543556	1.096725
63	Stay_2, 2012	Couple	0.014767	0.545963	1.101583
64	Couple, November	Asia	0.023644	0.546007	0.871441
65	Couple, Star_4, April	Asia	0.01122	0.547221	0.87338
66	Star_5, Stay_4	Couple	0.012065	0.547559	1.104802
67	Stay_2, Star_4.5	Couple	0.011555	0.547583	1.104851
68	Couple, December	Asia	0.023285	0.547807	0.874314
69	Star_4, 2011	Couple	0.018544	0.54895	1.10761
70	Stay_1, Star_5	Couple	0.023477	0.549627	1.108975
71	Stay_7	Couple	0.014296	0.549801	1.109326
72	Stay_3, Star_5	Couple	0.02315	0.549981	1.10969
73	Stay_3, January	Asia	0.010192	0.550108	0.877986
74	Stay_1, North America	Couple	0.011029	0.550736	1.111213
75	Couple, Star_4	Asia	0.123408	0.550829	0.879137
76	Star_5, 2015	Couple	0.011714	0.550975	1.111694
77	Star_5, October	Couple	0.010678	0.553262	1.116309
78	Couple, Star_3	Asia	0.044204	0.553427	0.883284

No	Premises	Conclusion	Support	Confidence	Lift
79	Star_5, 2014	Couple	0.014368	0.553917	1.117631
80	Star_5	Couple	0.104761	0.554379	1.118564
81	Couple, 2013	Asia	0.034753	0.555117	0.885982
82	Stay_2, Star_5	Couple	0.028473	0.556455	1.122753
83	North America	Couple	0.048595	0.556539	1.122922
84	Stay_1, Europe	Couple	0.016687	0.558698	1.127278
85	Stay_2, Solo traveler	Asia	0.015882	0.559046	0.892253
86	Stay_2, North America	Couple	0.011619	0.561417	1.132764
87	Couple	Asia	0.27882	0.562571	0.897879
88	Couple, Star_4, 2018	Asia	0.011141	0.564392	0.900785
89	Star_4, Oceania	Couple	0.03355	0.567385	1.144806
90	Couple, Star_5	Asia	0.059544	0.568386	0.907159
91	Couple, 2010	Star_4	0.014583	0.56903	1.226094
92	Couple, April	Asia	0.025373	0.569283	0.908591
93	Star_4, Europe	Couple	0.042204	0.569585	1.149244
94	Stay_3, North America	Couple	0.010926	0.569825	1.149729
95	2010	Couple	0.025628	0.57082	1.151736
96	Couple, 2018	Asia	0.023421	0.57179	0.912592
97	Couple, Star_4, 2017	Asia	0.011491	0.572222	0.913282
98	Couple, 2017	Asia	0.023644	0.574443	0.916827
99	Star_5, 2012	Couple	0.012416	0.574908	1.159984
100	Couple, Star_4.5	Asia	0.023222	0.576003	0.919316
101	Stay_1, Solo traveler	Asia	0.019038	0.57817	0.922775
102	2011	Couple	0.039965	0.578231	1.16669
103	Star_5, November	Couple	0.010049	0.57977	1.169795
104	Europe, 2013	Couple	0.011412	0.580697	1.171666
105	Europe, 2012	Couple	0.011173	0.582226	1.17475
106	Stay_2, January	Asia	0.011834	0.585799	0.934951
107	Europe	Couple	0.094807	0.585799	1.18196

No	Premises	Conclusion	Support	Confidence	Lift
108	Oceania	Couple	0.064629	0.587001	1.184383
109	Star_3, Europe	Couple	0.018823	0.588147	1.186697
110	Couple, Stay_3, Star_5	Asia	0.013635	0.588985	0.940035
111	Stay_2, Oceania	Couple	0.010057	0.589995	1.190426
112	Stay_1, January	Asia	0.011523	0.593109	0.946618
113	Europe, Stay_4	Couple	0.012719	0.593529	1.197556
114	2009	Couple	0.013826	0.594585	1.199687
115	Star_4, November	Asia	0.022082	0.595146	0.949869
116	Stay_3, Europe	Couple	0.018098	0.595594	1.201722
117	Oceania, 2012	Couple	0.010862	0.597545	1.205659
118	Star_5, North America	Couple	0.010543	0.599185	1.208967
119	2010	Star_4	0.026935	0.599929	1.292673
120	Couple, Star_4, 2014	Asia	0.014304	0.600134	0.957829
121	Europe, January	Couple	0.013515	0.600779	1.212185
122	Asia, 2010	Star_4	0.013276	0.601661	1.296406
123	Couple, 2019	Asia	0.020225	0.601992	0.960796
124	Couple, Star_4, August	Asia	0.011388	0.604484	0.964772
125	Stay_3, Oceania	Couple	0.011045	0.604976	1.220652
126	Stay_1, February	Asia	0.010081	0.605263	0.966016
127	Star_5, Europe	Couple	0.016894	0.60745	1.225644
128	Couple, Star_4, May	Asia	0.013141	0.608487	0.971162
129	Star_4, 2013	Asia	0.031326	0.608891	0.971806
130	Stay_3, March	Asia	0.010989	0.609099	0.972138
131	Star_4, April	Asia	0.026553	0.609921	0.973451
132	Couple, Star_4, June	Asia	0.01028	0.611374	0.97577
133	Couple, Star_4, July	Asia	0.010479	0.612768	0.977994
134	November	Asia	0.0514	0.613234	0.978739
135	Stay_2, 2011	Asia	0.010439	0.613583	0.979295
136	Star_4	Asia	0.285092	0.61429	0.980423

No	Premises	Conclusion	Support	Confidence	Lift
137	Stay_2, Europe	Couple	0.019572	0.61523	1.241342
138	Couple, August	Asia	0.023867	0.615748	0.982751
139	Star_5	Asia	0.116547	0.61675	0.98435
140	Stay_2, February	Asia	0.011093	0.617843	0.986094
141	Couple, Stay_2, Star_3	Asia	0.013571	0.618373	0.98694
142	2013	Asia	0.076622	0.618965	0.987884
143	Stay_3, 2012	Asia	0.014806	0.62016	0.989792
144	Couple, 2016	Asia	0.032123	0.621109	0.991307
145	Star_5, November	Asia	0.010766	0.621149	0.991371
146	Couple, Star_4, 2015	Asia	0.013245	0.621308	0.991625
147	Couple, July	Asia	0.023349	0.622081	0.992857
148	Couple, 2014	Asia	0.035789	0.622367	0.993314
149	Couple, June	Asia	0.0226	0.622476	0.993488
150	Couple, May	Asia	0.028274	0.62355	0.995203
151	Star_3, December	Asia	0.010663	0.625234	0.99789
152	Star_5, December	Asia	0.011093	0.625337	0.998054
153	Star_4, December	Asia	0.02527	0.625444	0.998225
154	April	Asia	0.058277	0.626059	0.999208
155	Couple, Stay_3	Asia	0.070295	0.626937	1.000608
156	Star_5, July	Asia	0.010599	0.627063	1.000809
157	December	Asia	0.056827	0.62795	1.002225
158	Europe, February	Couple	0.011396	0.629125	1.269376
159	Couple, 2015	Asia	0.033279	0.629864	1.00528
160	Star_5, 2013	Asia	0.016568	0.63	1.005497
161	Couple, Star_4, 2016	Asia	0.01503	0.633737	1.01146
162	Star_2	Asia	0.018209	0.634722	1.013034
163	Europe, March	Couple	0.011204	0.63505	1.281331
164	Couple, Star_4, September	Asia	0.011667	0.636799	1.016348

No	Premises	Conclusion	Support	Confidence	Lift
165	Star_4, August	Asia	0.025134	0.637172	1.016943
166	Star_5, 2016	Asia	0.011372	0.638193	1.018573
167	Star_4.5	Asia	0.048037	0.638559	1.019158
168	Star_4, 2017	Asia	0.026744	0.63863	1.01927
169	Star_4, 2018	Asia	0.02739	0.638848	1.019618
170	Couple, Star_4, October	Asia	0.012974	0.640693	1.022563
171	Star_3	Asia	0.111789	0.64087	1.022845
172	Stay_3, Star_5	Asia	0.027015	0.641802	1.024334
173	Couple, Stay_1, Star_3	Asia	0.016831	0.642922	1.026121
174	2017	Asia	0.055241	0.643043	1.026313
175	2009	Star_4	0.014974	0.643934	1.387491
176	Star_4, July	Asia	0.024377	0.645767	1.030662
177	Stay_1, 2012	Asia	0.016783	0.646409	1.031686
178	Couple, Star_4, Stay_3	Asia	0.033741	0.647896	1.034059
179	Star_3, 2017	Asia	0.011587	0.649687	1.036918
180	Stay_2, 2012	Asia	0.01758	0.649971	1.03737
181	Couple, Star_3.5	Asia	0.018934	0.650602	1.038379
182	Star_5, May	Asia	0.010742	0.650893	1.038843
183	Star_5, 2011	Couple	0.010168	0.65102	1.313555
184	Couple, September	Asia	0.025254	0.65112	1.039205
185	2018	Asia	0.057425	0.652186	1.040906
186	Stay_1, March	Asia	0.012248	0.655437	1.046095
187	Couple, Stay_2, 2013	Asia	0.010814	0.655556	1.046284
188	Couple, October	Asia	0.028641	0.656918	1.048459
189	Star_4, 2019	Asia	0.024377	0.658133	1.050397
190	July	Asia	0.052643	0.65869	1.051286
191	August	Asia	0.053106	0.66039	1.054001
192	Star_5, Oceania	Couple	0.016002	0.661833	1.335371
193	Couple, Stay_1, 2013	Asia	0.010336	0.662749	1.057765

No	Premises	Conclusion	Support	Confidence	Lift
194	Stay_3, April	Asia	0.013563	0.665104	1.061523
195	Stay_2, March	Asia	0.012974	0.665304	1.061844
196	Couple, Stay_2	Asia	0.085874	0.666049	1.063031
197	Stay_3, Star_3	Asia	0.023365	0.666061	1.063051
198	Stay_3, 2013	Asia	0.018791	0.66629	1.063416
199	Star_4, 2014	Asia	0.034028	0.667396	1.065182
200	Couple, Star_4, Stay_2	Asia	0.036402	0.667641	1.065573
201	Star_5, August	Asia	0.011157	0.668577	1.067067
202	Star_4, 2015	Asia	0.030609	0.669397	1.068375
203	Star_3, April	Asia	0.010894	0.670098	1.069494
204	Star_4, May	Asia	0.028912	0.670486	1.070114
205	Star_5, October	Asia	0.012958	0.671346	1.071486
206	Couple, Star_4, Stay_1	Asia	0.035183	0.672916	1.073991
207	2019	Asia	0.051448	0.672991	1.074111
208	Star_5, 2014	Asia	0.017484	0.67404	1.075786
209	Stay_3, November	Asia	0.013157	0.674153	1.075966
210	Star_5, June	Asia	0.010822	0.674615	1.076704
211	Star_3, 2016	Asia	0.014719	0.674826	1.077041
212	2014	Asia	0.078989	0.675388	1.077938
213	Star_4, Family with teens	Asia	0.026975	0.676459	1.079646
214	Star_3, 2015	Asia	0.013946	0.676721	1.080064
215	Star_5, 2015	Asia	0.0144	0.677286	1.080967
216	2016	Asia	0.075929	0.678149	1.082345
217	Star_3, 2018	Asia	0.013205	0.67882	1.083415
218	Star_3, 2014	Asia	0.012097	0.679499	1.084498
219	2015	Asia	0.075148	0.684325	1.092201
220	Star_4, 2016	Asia	0.036163	0.684878	1.093083
221	Star_4, October	Asia	0.029892	0.685239	1.09366
222	May	Asia	0.061648	0.687339	1.097011

No	Premises	Conclusion	Support	Confidence	Lift
223	Family with teens	Asia	0.055345	0.688169	1.098336
224	Couple, Stay_1	Asia	0.085579	0.688265	1.098489
225	Stay_3	Asia	0.157763	0.690609	1.102231
226	Star_4, June	Asia	0.026632	0.694369	1.108232
227	Star_4, September	Asia	0.025947	0.694539	1.108503
228	Couple, Stay_2, Star_5	Asia	0.019851	0.697173	1.112707
229	Stay_3, Star_4.5	Asia	0.014153	0.697291	1.112895
230	Star_5, September	Asia	0.01177	0.698345	1.114577
231	Stay_2, November	Asia	0.014982	0.699665	1.116684
232	Stay_2, December	Asia	0.016273	0.701477	1.119576
233	Star_3.5	Asia	0.041495	0.701657	1.119864
234	June	Asia	0.05501	0.701738	1.119993
235	October	Asia	0.063896	0.701979	1.120377
236	Stay_2, Star_3	Asia	0.033892	0.704372	1.124197
237	Stay_3, December	Asia	0.015436	0.70462	1.124592
238	Star_4, Stay_3	Asia	0.077626	0.706535	1.127649
239	Stay_2, Star_4.5	Asia	0.014934	0.707704	1.129514
240	September	Asia	0.054652	0.707813	1.129688
241	Star_3, 2019	Asia	0.0113	0.709	1.131583
242	Stay_3, July	Asia	0.013125	0.71022	1.13353
243	Star_5, Family with young children	Asia	0.024266	0.714118	1.139752
244	Stay_3, August	Asia	0.013237	0.71564	1.14218
245	Stay_1, 2013	Asia	0.022887	0.716388	1.143375
246	Family with young children, Stay_4	Asia	0.013348	0.716731	1.143922
247	Stay_3, 2017	Asia	0.014783	0.717324	1.144868
248	Couple, 2020	Asia	0.013731	0.717618	1.145337
249	Couple, Stay_2, 2014	Asia	0.011667	0.718704	1.147071

No	Premises	Conclusion	Support	Confidence	Lift
250	Stay_3, 2018	Asia	0.013874	0.719421	1.148216
251	Family with young children, April	Asia	0.011037	0.719481	1.14831
252	Stay_1, December	Asia	0.015842	0.721597	1.151688
253	Stay_2, 2013	Asia	0.024696	0.722041	1.152397
254	Stay_1, Star_3	Asia	0.043192	0.722859	1.153703
No	Premises	Conclusion	Support	Confidence	Lift
255	Star_4, Stay_1	Asia	0.080814	0.723582	1.154856
256	Stay_2	Asia	0.190101	0.724481	1.156292
257	Star_4, Stay_2	Asia	0.08397	0.72479	1.156784
258	Couple, Stay_2, 2015	Asia	0.010487	0.725869	1.158506
259	Couple, Stay_1, 2014	Asia	0.010049	0.726801	1.159994
260	Star_3, May	Asia	0.010838	0.727662	1.161368
261	Stay_3, May	Asia	0.014775	0.728487	1.162685
262	Stay_2, 2017	Asia	0.016217	0.729391	1.164127
263	Stay_1, November	Asia	0.016711	0.730662	1.166156
264	Star_5, Group	Asia	0.012703	0.734562	1.172381
265	Stay_3, 2019	Asia	0.011762	0.734694	1.172591
266	Stay_1, 2017	Asia	0.015795	0.735982	1.174647
267	Stay_1	Asia	0.192994	0.737634	1.177284
268	Stay_1, Star_4.5	Asia	0.011714	0.738693	1.178974
269	Stay_2, April	Asia	0.019086	0.738969	1.179415
270	Stay_2, Star_5	Asia	0.037917	0.741006	1.182665
271	Stay_2, 2018	Asia	0.016695	0.743699	1.186963
272	Stay_3, 2016	Asia	0.019739	0.745411	1.189695
273	Star_4, Group	Asia	0.044547	0.745433	1.189731
274	Stay_3, Family with teens	Asia	0.014137	0.747262	1.19265
275	Stay_2, 2016	Asia	0.022744	0.747512	1.193049

No	Premises	Conclusion	Support	Confidence	Lift
276	Stay_1, April	Asia	0.018719	0.747851	1.19359
277	Stay_3, 2014	Asia	0.02154	0.748546	1.1947
278	Stay_2, July	Asia	0.016329	0.74863	1.194833
279	Stay_2, August	Asia	0.015842	0.75303	1.201856
280	Star_4, Stay_2, 2014	Asia	0.010089	0.75447	1.204154
281	Stay_1, August	Asia	0.015978	0.754895	1.204832
282	Stay_2, 2014	Asia	0.024473	0.758272	1.210222
283	Family with young children, 2013	Asia	0.013611	0.758437	1.210486
284	Family with young children	Asia	0.115973	0.759313	1.211884
285	Star_4, Stay_2, 2016	Asia	0.010487	0.759377	1.211986
286	Stay_3, October	Asia	0.015014	0.759677	1.212465
287	Star_4, 2020	Asia	0.014193	0.760137	1.213198
288	Star_4, Family with young children	Asia	0.056556	0.76042	1.213651
289	Stay_1, 2015	Asia	0.020759	0.761473	1.215332
290	Star_3, October	Asia	0.010615	0.761578	1.215499
291	Stay_1, 2016	Asia	0.022616	0.764342	1.219909
292	Stay_3, 2015	Asia	0.020871	0.764448	1.22008
293	Star_4, Stay_3, 2016	Asia	0.010089	0.76588	1.222365
294	Stay_1, July	Asia	0.016121	0.766288	1.223016
295	Stay_1, 2018	Asia	0.018153	0.766487	1.223334
296	Group	Asia	0.100856	0.767915	1.225612
297	Stay_2, 2019	Asia	0.014464	0.768742	1.226933
298	Star_4, Stay_3, 2014	Asia	0.010408	0.770956	1.230467
299	Stay_3, September	Asia	0.013834	0.77293	1.233616
300	Stay_1, June	Asia	0.015619	0.774092	1.235471
301	Stay_2, May	Asia	0.019261	0.775674	1.237996

No	Premises	Conclusion	Support	Confidence	Lift
302	Stay_1, 2014	Asia	0.022536	0.77671	1.239649
303	Stay_2, 2015	Asia	0.023325	0.776805	1.239801
304	Familywithyoungchildren, 2014	Asia	0.014185	0.776953	1.240038
305	Stay_3, June	Asia	0.015133	0.778279	1.242154
306	Stay_2, June	Asia	0.016153	0.779915	1.244766
307	Family with young children, Star_4.5	Asia	0.010559	0.78033	1.245427
308	Star_4, Stay_1, 2016	Asia	0.010041	0.780669	1.245969
309	Family with young children, 2017	Asia	0.011244	0.780852	1.246261
310	Couple, Stay_1, Star_5	Asia	0.018337	0.781059	1.246591
311	Stay_2, Star_3.5	Asia	0.013005	0.781609	1.247469
312	Family with young children, 2016	Asia	0.014535	0.784179	1.25157
313	Star_3, Family with young children	Asia	0.015691	0.785714	1.254021
314	Group, 2013	Asia	0.01177	0.786056	1.254567
315	2020	Asia	0.034346	0.786209	1.254811
316	Stay_2, October	Asia	0.019875	0.787496	1.256865
317	Family with young children, 2018	Asia	0.013005	0.788025	1.257709
318	Stay_1, 2019	Asia	0.017675	0.789324	1.259782
319	Family with young children, December	Asia	0.013189	0.789976	1.260823
320	Star_3, Group	Asia	0.023692	0.790481	1.261629
321	Stay_2, September	Asia	0.0164	0.793369	1.266239
322	Stay_3, Star_3.5	Asia	0.012129	0.793949	1.267164
323	Stay_1, Star_3.5	Asia	0.010981	0.795153	1.269085

No	Premises	Conclusion	Support	Confidence	Lift
324	Family with young children, October	Asia	0.011045	0.796552	1.271318
325	Stay_1, September	Asia	0.017564	0.798551	1.274508
326	Stay_1, May	Asia	0.020847	0.798779	1.274872
327	Family with young children, 2015	Asia	0.014209	0.799193	1.275534
328	Stay_1, Star_5	Asia	0.034171	0.8	1.276821
No	Premises	Conclusion	Support	Confidence	Lift
329	Family with young children, 2019	Asia	0.011539	0.800885	1.278234
330	Star_4, Stay_3, Group	Asia	0.01291	0.81	1.292782
331	Stay_3, Group	Asia	0.027374	0.811481	1.295146
332	Star_4, Stay_1, Group	Asia	0.011563	0.818387	1.306167
333	Stay_1, October	Asia	0.021739	0.819219	1.307496
334	Group, 2016	Asia	0.013029	0.819549	1.308022
335	Group, October	Asia	0.010503	0.820672	1.309815
336	Stay_2, Family with teens	Asia	0.017022	0.821855	1.311702
337	Star_4, Business traveler	Asia	0.010423	0.822124	1.312133
338	Stay_1, Family with teens	Asia	0.01652	0.824254	1.315532
339	Star_4, Stay_2, Group	Asia	0.014623	0.831445	1.327009
340	Group, May	Asia	0.011428	0.832269	1.328324
341	Stay_2, Group	Asia	0.032816	0.834786	1.332341
342	Stay_1, Group	Asia	0.029557	0.835549	1.333558
343	Stay_3, Family with young children	Asia	0.032211	0.835988	1.334258
344	Business traveler	Asia	0.022369	0.839414	1.339727
345	Group, 2014	Asia	0.012472	0.842757	1.345063

No	Premises	Conclusion	Support	Confidence	Lift
346	Star_4, Stay_1, Family with young children	Asia	0.01389	0.847351	1.352394
347	Stay_2, Family with young children	Asia	0.032545	0.851898	1.359652
348	Star_4, Stay_2, Family with young children	Asia	0.015332	0.855111	1.36478
349	Group, 2015	Asia	0.01224	0.855234	1.364976
350	Star_4, Stay_3, Family with young children	Asia	0.016847	0.85587	1.365992
351	Stay_1, Family with young children	Asia	0.030298	0.857852	1.369155
352	Star_5, 2020	Asia	0.010113	0.87037	1.389134
353	Family with young children, June	Asia	0.012296	0.871751	1.391338
354	Stay_1, Business traveler	Asia	0.012001	0.880187	1.404802
355	Stay_1, 2020	Asia	0.018177	0.899448	1.435543

อ้างอิง

[1] Hajli N., (2014) . A study of the impact of social media on consumers. The Market Research Society.Vol. 56

[2] Xie K, Chen C.C and Wu S. (2015) . Online Consumer Review Factors Affecting Offline Hotel Popularity: Evidence from Tripadvisor, Journal of Travel and Tourism Marketing, Vol. 33

[3] Nilashi M, Ibrahim O, Yadegaridehkordi E, Samad S, Akbari E and Alizadeh A. (2018) . Travelers's decision making using online review in social network sites: A case on TripAdvisor. Publisher: Elsevier.Volume 28.

[4] Jui-Lung Chen, K. K. (2020) . A Case Study of Factors Affecting the Performance of the Hotel Industry of Chiang Mai in Thailand and the Impact of Covid-19: Taking De Chai Hotel Group as an Example. Advance in Management & Applied Economics.

[5] Thadanit S and Polnyotee M., (2015) . Community-Based Tourism: A Strategy for Sustainable Tourism Development of Patong Beach, Phuket Island, Thailand. Canadian Center of Science and Education. Vol. 11.

[6] Ulrike Gretzel Kyung-Hyan Yoo (2008) . Use and Impact of Online Travel Reviews. Information and Communication Technologies in Tourism

[7] Alzbeta Kiralova, Antonín Pavlicek (2014) . Development of Social Media Strategies in Tourism Destination. International Conference on Strategic Innovative Marketing, Madrid, Spain

[8] Mussab Aljahdali (2016) . Exploring TripAdvisor Online Reviews: The Case of George Eastman Museum, Rochester Institute of Technology, RIT Scholar Works

[9] Priya Rai, Ramratan Ahirwal (2018) . Tourism Review Sentiment Analysis using Lexicon Features and Machine Learning Approach.VOLUME 5

[10] J. Miguéns, R. Baggio, and C. Costa (2008) . Social media and Tourism Destinations: TripAdvisor Case Study,Advances in Tourism Research 2008, Aveiro, Portugal.

[11] Mohammad Darini, Fatemeh Khozaei (2016) . The study of factors affecting customer's satisfaction with the three-star hotels in Dubai. International Journal of Advanced Engineering, Management and Science (IJAEMS) , Vol-2.

[12] Egbert Van der Zee, Dario Bertocchi (2018) . Finding patterns in urban tourist behaviour: a social network analysis approach based on TripAdvisor reviews.

[13] Sonia Saini, Ritu Punhani, Ruchika Bathla, and other (2019) . Sentiment Analysis on Twitter Data using R. International Conference on Automation, Computational and Technology Management (ICACTM) Amity University

[14] Avadh Kishor SinghI, Ajeet Kumar, Ashish K. Maurya (2014) . An Empirical Analysis and Comparison of Apriori and FP- Growth Algorithm for Frequent Pattern Mining. International Conference on Advanced Communication, Control and Computing Technologies (ICACCCT)

[15] Ahmed Imran Kabir, Ridoan Karim and other (2018) . The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R.

[16] K.Arun, Asrinagesh and M. Ramesh (2017) . Twitter Sentiment Analysis on Demonetization Tweets in India Using R Language

[17] Maurizio Naldi (2019) . A review of sentiment computation methods with R packages

[18] Indy Wijngaards, Martijn Burger and Job van Excel (2021) . Unpacking the Quantifying and Qualifying Potential of Semi-Open Job Satisfaction Questions through Computer-Aided Sentiment Analysis

[19] Mianadeel Hassan (2020) . Sentiment Base Emotions Classification of Celebrity Tweets by Using R Language

[20] Vaibhav Singh, Aayushi Mahajan, and Deepanshi Chaudhary (2020) . Sentimental Analysis of Hotel Reviews from TripAdvisor

[21] สัญฌา พันธุ์แพง (2563) . การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับการพยากรณ์การศึกษาต่อนักศึกษาใหม่ระดับปริญญาตรี ในมหาวิทยาลัยราชภัฏเชียงใหม่

[22] สุภาพรรณ คงมณีพรรณ (2562) . การหากฎความสัมพันธ์จากฐานข้อมูลการซื้อผลิตภัณฑ์เสริมอาหารยี่ห้อ มายเฮลท์ ของลูกค้าโดยใช้อัลกอริทึม เอฟพีโกรท และการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้ามายเฮลท์ โดยเทคนิคอาร์เอฟเอ็มด้วยโปรแกรม แรพพิดไมเนอร์ : กรณีศึกษาร้านยาเชนแห่งหนึ่ง

[23] ธนะพร ฮองกุล, สายัณห์ เทพแดง และคนอื่น ๆ (2560) . การหาความสัมพัธ์ของราคาน้ำมันดิบดับบลิวทีไอ (WTI Crude) ในตลาดไนเม็กซ์ (NYMEX) ระหว่าง ราคาหุ้นกลุ่มพลังงานที่เกี่ยวข้องกับธุรกิจบริการสถานีน้ำมัน ในตลาดหลักทรัพย์แห่งประเทศไทย ด้วยเทคนิคการทำเหมืองข้อมูล

[24] Federico Antonello, Piero Baraldi, Ahmed Shokry, Enrico Zio, Ugo Gentile, Luigi Serio (2021) . Association rules extraction for the identification of functional dependencies in complex technical infrastructures: Reliability Engineering and System Safety

[25] Foxiao Zhan et al (2019) IOP Conf. Ser.: Earth Environ. Sci. Summary of Association Rules: IOP Conference Series: Earth and Environmental Science

[26] Kitsiri Chochiang, Thanakrit Thongkhamdee and Lalita Sathansat (2020) . Translation Quality Assessment of Online Translation System in Translation English to Thai on Phuket Tourism.

[27] Fanclubthailand (2021) . October in Thailand. From https://fanclubthailand.co.uk/october-in-thailand/

[28] Roy Cavanagh (2020) . Visiting Thailand in June. From https://www.thaizer.com/itineraries/visiting-thailand-in-june

[29] Kalani Stephena (2021) . When is the best time to visit Thailand? From https://www.enchantingtravels.com/destinations/asia/thailand/best-time-to-visit-thailand/

[30] JTB Business travel (2020) . What business traveler want from hotels? From https://www.jtbusabusinesstravel.com/what-business-travelers-want-in-hotels/

ประวัติผู้เขียน

ชื่อ สกุล นาย ณรงค์ฤทธิ์ บุญสมาน

รหัสประจำตัวนักศึกษา 6310025006

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิทยาศาสตรบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์	มหาวิทยาลัยสงขลานครินทร์	2561

การตีพิมพ์เผยแพร่ผลงาน

B.Narongrit, S.Pannipa And C. Kitsiri “Analysis of the relationship of tourist behavior in Andaman Coast Provinces, Southern Thailand”, 2021 25th International Computer Science and Engineering Conference (ICSEC), Thailand : pages 57-62.