



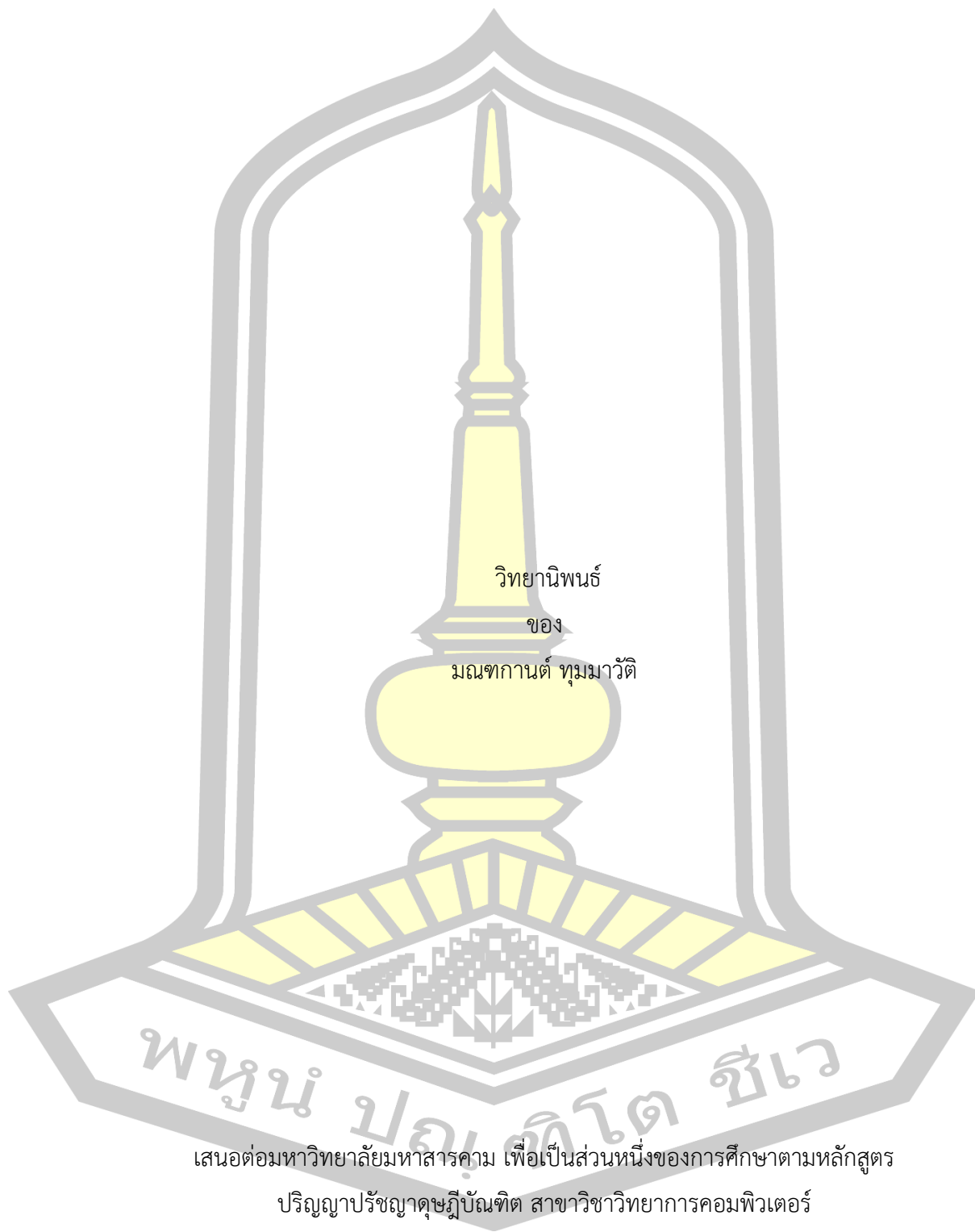
การจำแนกความคิดเห็นของลูกค้าโรงแรมเพื่อใช้ในการธุรกิจ

วิทยานิพนธ์
ของ
มณฑกานต์ ทุมมาวัติ

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
มิถุนายน 2564

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การจำแนกความคิดเห็นของลูกค้าโรงแรมเพื่อใช้ในงานธุรกิจ



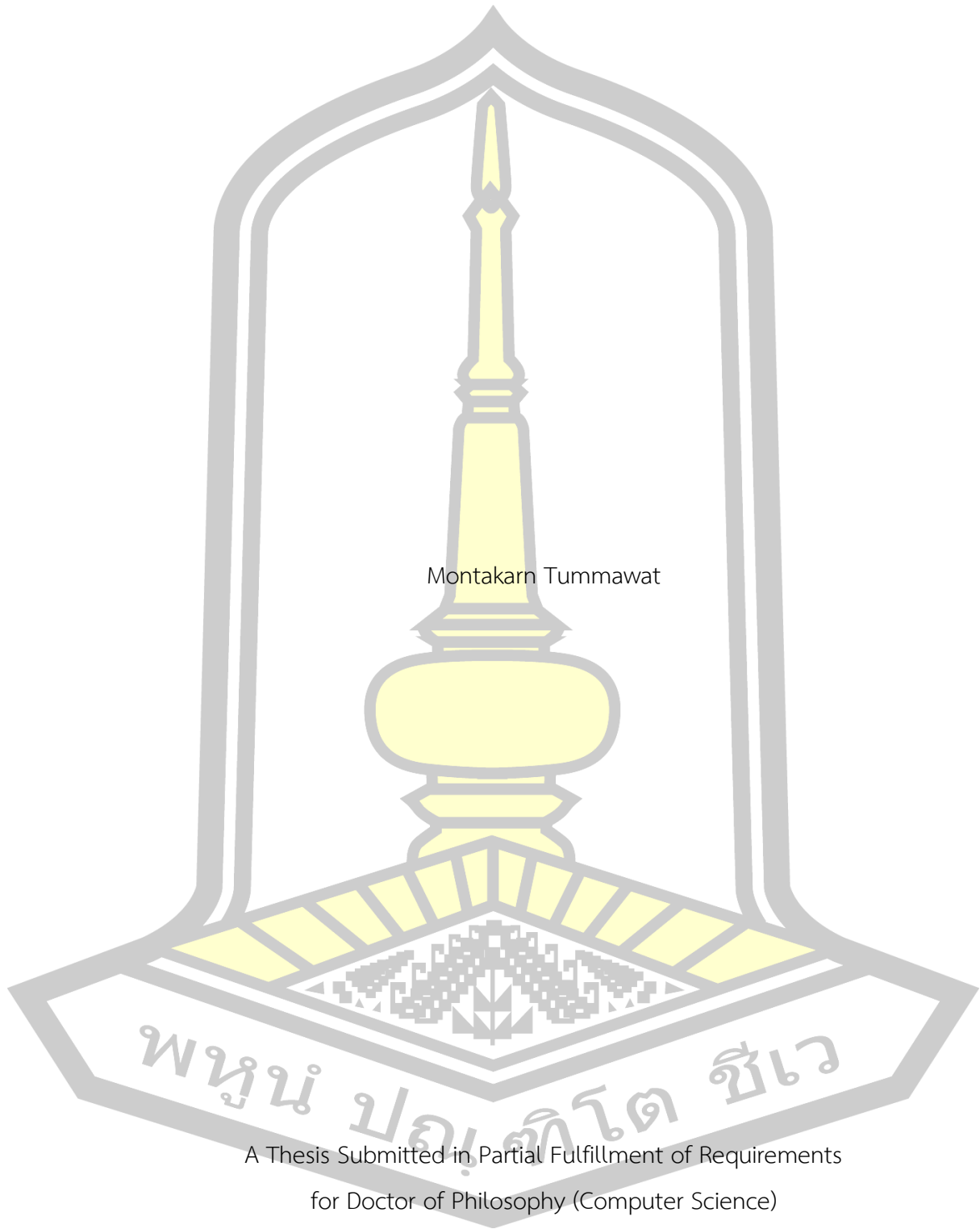
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

มิถุนายน 2564

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Opinion Classification of Hotel customer's Reviews for Business

Montakarn Tummawat



A Thesis Submitted in Partial Fulfillment of Requirements
for Doctor of Philosophy (Computer Science)

June 2021

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวมณฑกานต์ ทุมมา
วดี แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รศ. ดร. กฤษณพงศ์ สมสุข)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. ฉัตรเกล้า เจริญผล)

.....กรรมการ

(ผศ. ดร. พัฒนพงษ์ ชมภูวิเศษ)

.....กรรมการ

(ผศ. ดร. พนิดา ทรงรัมย์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ ของมหาวิทยาลัยมหาสารคาม

.....
(ผศ. ศศิธร แก้วมัน)

คณบดีคณะวิทยาการสารสนเทศ

.....
(รศ. ดร. กริสน์ ชัยมูล)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การจำแนกความคิดเห็นของลูกค้าโรงแรมเพื่อใช้ในงานธุรกิจ
 ผู้วิจัย มณฑกานต์ ทุมมาวัติ
 อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. ฉัตรเกล้า เจริญผล
 ปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชา วิทยาการคอมพิวเตอร์
 มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม ปีที่พิมพ์ 2564

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยนี้เป็นการนำเสนอการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรมต่อคุณลักษณะเด่นของโรงแรม เพื่อช่วยปรับปรุงกลยุทธ์การบริการทางธุรกิจของโรงแรมให้ตอบสนองความต้องการของผู้เข้าพัก วิธีที่นำเสนอในงานวิจัยนี้ถูกทดสอบประสิทธิภาพกับกลุ่มตัวอย่างจำนวน 30 โรงแรม ขั้นตอนแรก ทำการรวบรวมความคิดเห็นของผู้เข้าพักและคำอธิบายเกี่ยวกับโรงแรม ขั้นตอนที่สองกระบวนการเตรียมข้อมูล (Data Pre-processing) ถูกใช้สำหรับสกัดคำอธิบายคุณลักษณะของโรงแรมและความคิดเห็นของผู้เข้าพัก ซึ่งในกระบวนการเตรียมข้อมูลประกอบด้วยกระบวนการกำกับชนิดของคำ (Path-of-Speech Tagging) โดยในงานวิจัยนี้ Penn Treebank Tagset ถูกนำมาใช้เลือกคำกริยา (Verb) คำคุณศัพท์ (ADJ) และคำกริยาวิเศษณ์ (ADV) ในการประมวลผล จากนั้นกระบวนการวิเคราะห์ความคิดเห็นของลูกค้าถูกนำมาใช้เพื่อตัดสินใจในการระบุข้อความความคิดเห็นของแต่ละคุณลักษณะ สุดท้ายกระบวนการจับคู่คุณลักษณะและข้อความความคิดเห็น ถูกใช้เพื่อตรวจสอบความคิดเห็นของลูกค้าและจุดเด่นของโรงแรม ผลการวิจัยพบว่าจำนวนความคิดเห็นทั้งหมด 600 - 700 ข้อความ มีประสิทธิภาพโดยรวม เท่ากับ 0.78 และจำนวนความคิดเห็นอยู่ระหว่าง 280 - 400 ข้อความ พบว่ามีประสิทธิภาพโดยรวม ไม่เกิน 0.72 ฝ่ายบริหารของโรงแรมสามารถนำผลการวิจัยนี้ไปใช้เป็นแนวทางในการปรับปรุงกลยุทธ์การตลาดของโรงแรมได้

คำสำคัญ : การวิเคราะห์ความคิดเห็น, การวิเคราะห์ความรู้สึก

สำนักพิมพ์ ปณ. สก. โด. ช. เว

TITLE	Opinion Classification of Hotel customer's Reviews for Business		
AUTHOR	Montakarn Tummawat		
ADVISORS	Assistant Professor Chatklaw Jareanpon , Ph.D.		
DEGREE	Doctor of Philosophy	MAJOR	Computer Science
UNIVERSITY	Maharakham University	YEAR	2021

ABSTRACT

The objective of this paper is to present the opinion analysis of customers on hotel features to improve the hotel business service strategy in order to respond customer requirement. In this paper, the proposed methodology was tested with 30 hotels. Firstly, the reviews and description of hotels were collected. Secondly, data pre-processing was used for feature of hotel description and the hotel reviews. In data pre-processing, Penn Treebank Tagset, one of Path-of-Speech Tagging process, was used for selecting a verb, an adjective and an adverb in data pre-processing. After that, the opinion analysis was used for polarity identification of each feature. Finally, matching feature and polarity process were utilized to verify the customer review and hotel description. The results showed that the proposed method can provide the overall performance of 0.78 for the data sets of 600 to 700 comments, and the overall performance of 0.72 the data sets of 280 to 400 comments. Based on this study, hotel managers can use the results in this case for improving the hotel's marketing strategy.

Keyword : Opinion Mining, Sentiment analysis

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จสมบูรณ์ได้ด้วยความกรุณาเป็นอภัยสูงจาก ผู้ช่วยศาสตราจารย์ ดร. ฉัตรเกล้า เจริญผล อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ที่ได้กรุณารับเป็นอาจารย์ที่ปรึกษา ท่านได้เมตตาให้ความรู้ ช่วยเหลือและให้แนวคิด ให้กำลังใจตลอดจนคำแนะนำในการปรับปรุงแก้ไขข้อบกพร่องต่างๆ ผู้วิจัยทราบซึ่งในความกรุณา และขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.กฤษณ์พงศ์ สมสุข ประธานกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.พัฒน์พงษ์ ชมภูวิเศษ และ ผู้ช่วยศาสตราจารย์ ดร.พินดา ทรงรัมย์ กรรมการควบคุมวิทยานิพนธ์ ที่กรุณาให้คำปรึกษา คำแนะนำและแนวคิด ตลอดจนแนวทางแก้ไข ปัญหาข้อบกพร่องในการทำวิทยานิพนธ์

ขอกราบขอบพระคุณคณาจารย์ทุกท่าน แม้นมิได้เอ่ยนาม ที่ได้อบรมสั่งสอน ให้ความรู้ทางด้านวิชาการ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้

ขอขอบพระคุณพี่ๆ เพื่อนๆ น้องๆ สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม เพื่อนร่วมงานสาขาวิชาเทคโนโลยีดิจิทัล คณะบริหารศาสตร์ มหาวิทยาลัยกาฬสินธุ์ ที่คอยช่วยเหลือและเป็นกำลังใจให้กับผู้วิจัยมาโดยตลอด

ขอขอบพระคุณมหาวิทยาลัยกาฬสินธุ์ที่ได้กรุณามอบทุนในการศึกษาต่อระดับปริญญาเอกแก่ผู้วิจัยในครั้งนี้

สุดท้ายนี้ขอกราบขอบพระคุณบิดา มารดา คู่ชีวิต และ ญาติพี่น้องและเพื่อนๆ ทุกคน ที่เป็นแรงผลักดันและเป็นกำลังใจสำคัญให้ตลอดเวลาในการศึกษาเล่าเรียน หากวิทยานิพนธ์ฉบับนี้มีประโยชน์และคุณค่าทางการศึกษา ผู้เขียนขอยกความดีทั้งหมดให้กับบุคคลที่ผู้วิจัยได้กล่าวถึง

มณฑกานต์ ทุมมาวัตี

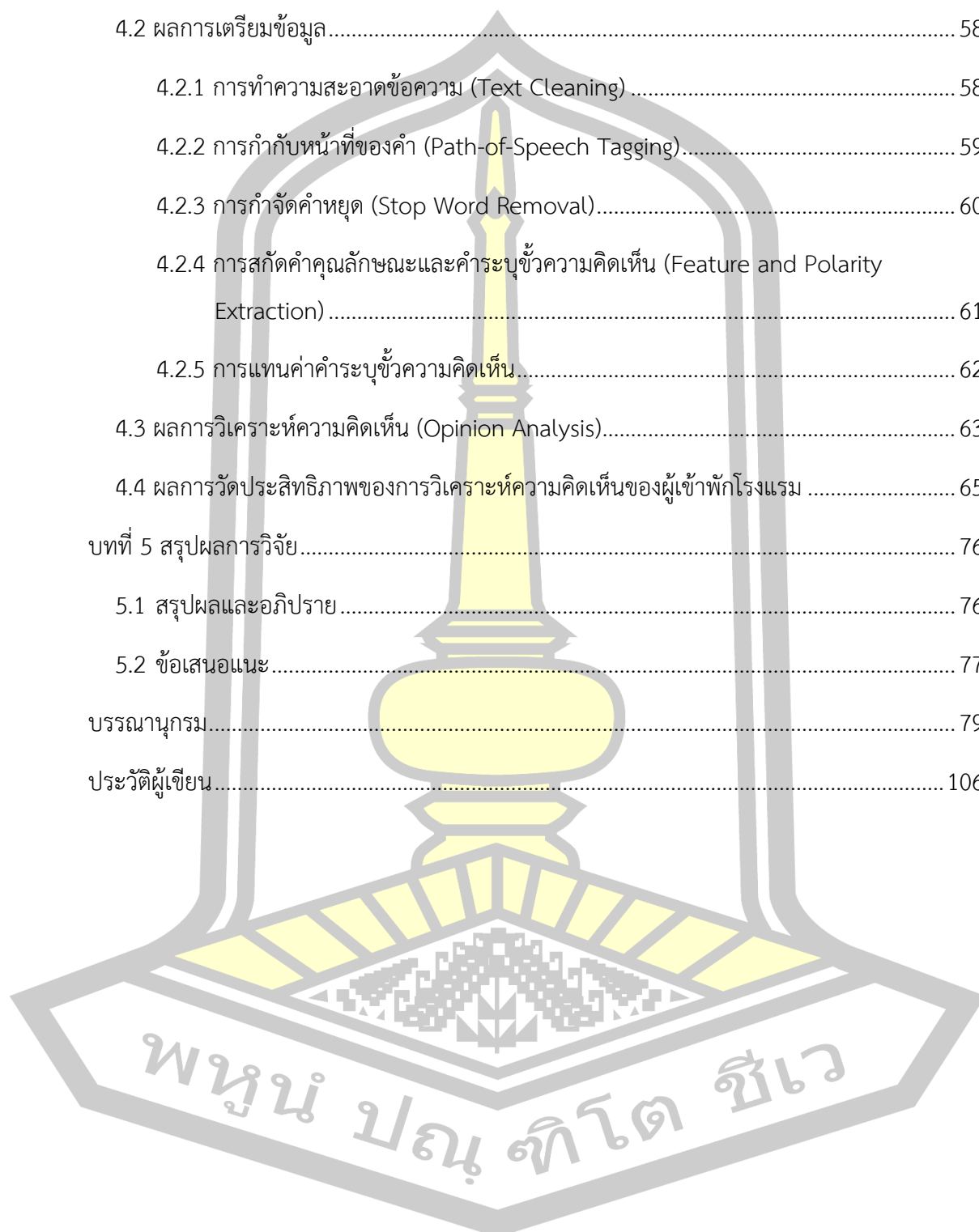
พูน ปณ สิโต ชีเว

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 หลักการและเหตุผล.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	2
1.3 ความสำคัญของการวิจัย.....	2
1.4 ขอบเขตของการวิจัย.....	2
1.5 นิยามศัพท์เฉพาะ.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	3
2.1 การบริหารลูกค้าสัมพันธ์.....	3
2.1.1 แนวคิดพื้นฐานของการบริหารลูกค้าสัมพันธ์.....	3
2.1.2 การบริหารประสบการณ์ลูกค้า.....	4
2.2 การประมวลผลภาษาทางธรรมชาติ.....	5
2.3 การทำเหมืองข้อมูลใน CRM.....	6
2.3.1 การทำเหมืองความคิดเห็น.....	7
2.4 ขั้นตอนการเตรียมข้อมูล.....	8
2.5 วิธีการจำแนกประเภทความคิดเห็น (Opinion Classifier Algorithm).....	11

2.5.1	วิธีการเลิกชิคอนเบส.....	11
2.5.2	วิธีการแมชชีนเลิร์นนิง	12
2.6	การวัดประสิทธิภาพในการจำแนกข้อมูล	14
2.6.1	การแบ่งข้อมูล	14
2.6.2	การวัดประสิทธิภาพการจำแนก	16
2.7	งานวิจัยที่เกี่ยวข้อง	18
2.7.1	งานวิจัยที่ใช้วิธีการเลิกชิคอนเบส	18
บทที่ 3	วิธีดำเนินการวิจัย.....	35
3.1	การเก็บรวบรวมข้อมูล (Data Collection)	36
3.2	กระบวนการเตรียมข้อมูล (Data Preprocessing)	39
3.2.1	การทำความสะอาดข้อความ (Text Cleaning)	39
3.2.2	การกำกับชนิดของคำ (Path-of-Speech Tagging).....	40
3.2.3	การกำจัดคำหยุด (Stop Word Removal).....	41
3.2.4	การสกัดคุณลักษณะและการสกัดคำระบุข้อความความเห็น (Feature and Polarity ...	43
	Extraction).....	43
3.2.5	การแทนค่าคำระบุข้อความความเห็น (Polarity Representation).....	45
3.3	การวิเคราะห์ความคิดเห็น (Opinion Analysis).....	47
3.4	การวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม (Evaluation).....	48
3.5	การสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม (Summarization).....	50
3.6	การเปรียบเทียบผลการวิเคราะห์ข้อความความเห็นกับคำอธิบายของโรงแรม	51
บทที่ 4	ผลการดำเนินงาน	52
4.1	เครื่องมือและข้อมูลที่ใช้ในการทดลอง.....	52
4.1.1	เครื่องมือที่ใช้ในการทดลอง	52
4.1.2	การรวบรวมข้อมูล (Data Collection)	52

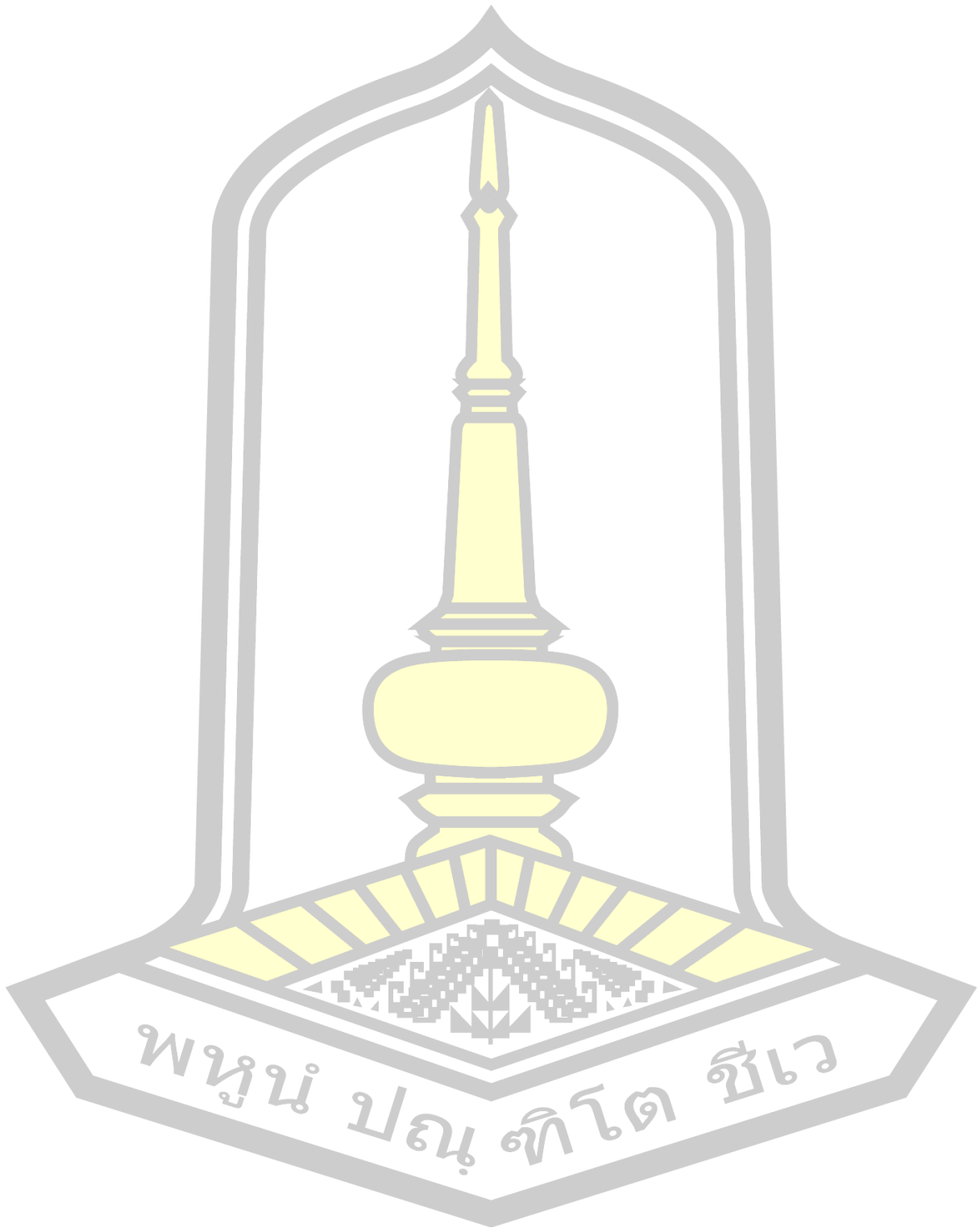
4.1.3 การระบุผลเฉลยโดยผู้เชี่ยวชาญ.....	57
4.2 ผลการเตรียมข้อมูล.....	58
4.2.1 การทำความสะอาดข้อความ (Text Cleaning)	58
4.2.2 การกำกับหน้าที่ของคำ (Path-of-Speech Tagging).....	59
4.2.3 การกำจัดคำหยุด (Stop Word Removal).....	60
4.2.4 การสกัดคำคุณลักษณะและคำระบุขั้วความคิดเห็น (Feature and Polarity Extraction)	61
4.2.5 การแทนค่าคำระบุขั้วความคิดเห็น.....	62
4.3 ผลการวิเคราะห์ความคิดเห็น (Opinion Analysis).....	63
4.4 ผลการวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม	65
บทที่ 5 สรุปผลการวิจัย.....	76
5.1 สรุปผลและอภิปราย	76
5.2 ข้อเสนอแนะ.....	77
บรรณานุกรม.....	79
ประวัติผู้เขียน.....	106



สารบัญตาราง

	หน้า
ตารางที่ 1 ตัวอย่างการตัดคำภาษาอังกฤษ.....	9
ตารางที่ 2 ตัวอย่างคำหยุดภาษาอังกฤษ.....	10
ตารางที่ 3 ตัวอย่างคำหยุดภาษาไทย.....	10
ตารางที่ 4 Confusion Matrix.....	16
ตารางที่ 5 ผลการทดลองชุดข้อมูล 4 ชุด	24
ตารางที่ 6 การวัดประสิทธิภาพการสกัดคำคุณลักษณะ ค่าความแม่นยำและค่าความระลึก.....	25
ตารางที่ 7 เปรียบเทียบประสิทธิภาพการจัดกลุ่มคำด้วยเทคนิควิธี K-Mean และ Enhance	25
ตารางที่ 8 ตัวอย่างข้อความความคิดเห็นภาษาต่างๆ.....	40
ตารางที่ 9 คลังคำหยุดมาตรฐาน จาก Stanford Stopword List	42
ตารางที่ 10 ตารางคำ Synonym.....	43
ตารางที่ 11 ชนิดของคำที่นำมาวิเคราะห์	44
ตารางที่ 12 ตัวอย่างคำระบุข้อความความคิดเห็นจากคลังคำ SentiWordNet	45
ตารางที่ 13 ตัวอย่างค่าคะแนนเชิงบวกและเชิงลบของคำศัพท์ Good ใน SentiWordNet	46
ตารางที่ 14 ตัวอย่างการวิเคราะห์ความคิดเห็น	48
ตารางที่ 15 ตัวอย่างตารางผลลัพธ์ที่ได้จากการวิเคราะห์เทียบกับผลเฉลย.....	49
ตารางที่ 16 ตารางแสดงจำนวนข้อความความคิดเห็นทั้งหมด(ทุกภาษา).....	54
ตารางที่ 17 ตารางแสดงจำนวนข้อความความคิดเห็นที่ได้ทำการคัดเลือก (เลือกเฉพาะภาษาอังกฤษ).....	55
ตารางที่ 18 แสดงตัวอย่างรายชื่อโรงแรมและคำอธิบายคุณลักษณะโรงแรม.....	56
ตารางที่ 19 การตรวจสอบผลเฉลยของผู้เชี่ยวชาญ.....	57
ตารางที่ 20 ตารางแสดงจำนวนข้อความความคิดเห็นทั้งหมด (ทุกภาษา)	58
ตารางที่ 21 ตัวอย่างจำนวนคำคุณลักษณะที่เป็นคำหลักและคำเหมือนที่ปรากฏในข้อความความคิดเห็น	62

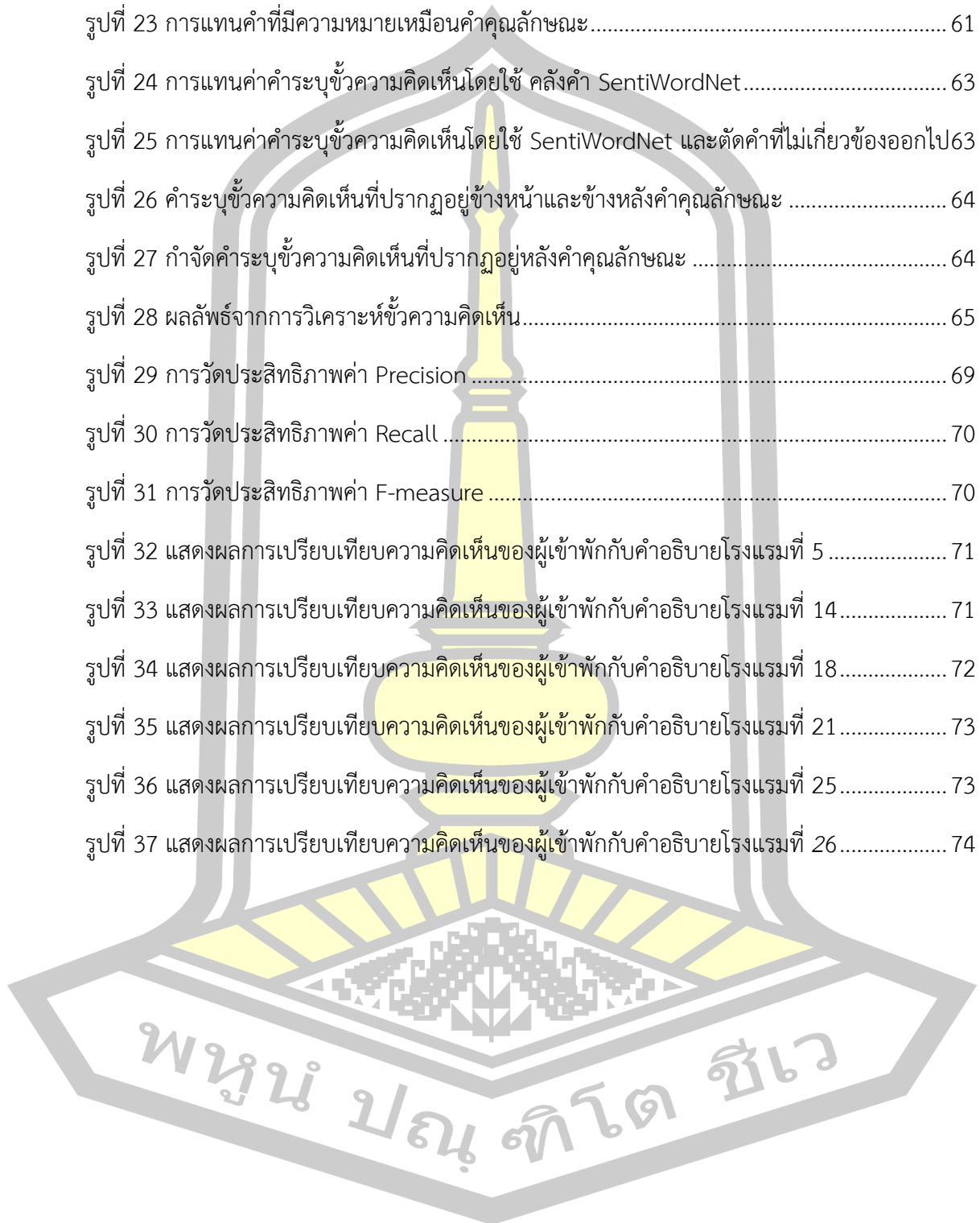
ตารางที่ 22 การวัดประสิทธิภาพค่า Precision Recall F-measure 65



สารบัญภาพ

	หน้า
รูปที่ 1 ขั้นตอนการเตรียมข้อมูลอยู่รูปแบบที่ไม่มีโครงสร้าง.....	9
รูปที่ 2 ตัวอย่างการแบ่งข้อมูลแบบ 5 – Fold Cross Validation ที่มา : [20].....	15
รูปที่ 3 กรอบแนวคิดในการวิเคราะห์ความคิดเห็นแบบปริยาย ที่มา : 20	
รูปที่ 4 รูปแบบแนวคิดใหม่สำหรับมุมมองตามการจัดประเภทสำหรับสถานที่ท่องเที่ยว	21
รูปที่ 5 สรุปความคิดเห็นจากการโรงแรมที่อยู่ในสถานที่ที่ผู้ใช้สนใจ ที่มา : [28].....	22
รูปที่ 6 ข้อความแสดงความคิดเห็นที่นำมาสรุปเป็นข้อมูลทางสถิติ ที่มา : [28].....	23
รูปที่ 7 โมเดลแสดงกระบวนการจำแนกความคิดเห็นของลูกค้าโรงแรมเพื่อใช้ในเชิงธุรกิจ	35
รูปที่ 8 โรงแรม Ibis Styles Bangkok Khaosan Viengtai ในเว็บไซต์ www.tripadvisor.com ...	36
รูปที่ 9 ข้อความคิดเห็นของผู้เข้าพักโรงแรม Ibis Styles Bangkok Khaosan Viengtai.....	37
รูปที่ 10 คำอธิบายรายละเอียดของโรงแรม Ibis Styles Bangkok Khaosan Viengtai	37
รูปที่ 11 เว็บไซต์ Import.io.....	38
รูปที่ 12 ลักษณะของเว็บไซต์ Import.io ที่ใช้ในการเก็บรวบรวมข้อความคิดเห็น	38
รูปที่ 13 ตัวอย่างข้อความคิดเห็นภาษาต่างๆ.....	39
รูปที่ 14 POS Tagging (Penn Treebank Tagset)	41
รูปที่ 15 ขั้นตอนการสกัดคำระบุขั้วความคิดเห็น (Polarity Extraction)	45
รูปที่ 16 ตัวอย่างการสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม.....	50
รูปที่ 17 ตัวอย่างการเปรียบเทียบผลการวิเคราะห์ข้อความคิดเห็นกับคำอธิบายของโรงแรม	51
รูปที่ 18 ลักษณะของข้อความคิดเห็นที่เก็บไว้ในไฟล์ Excel.....	53
รูปที่ 19 ลักษณะของคำอธิบายรายละเอียดของโรงแรมที่เก็บไว้ในไฟล์ Excel	53
รูปที่ 20 ข้อมูลที่ได้จากขั้นตอนการทำ POS Tagging	59
รูปที่ 21 ตัวอย่างข้อมูลที่ได้จากขั้นตอนการกำจัดคำหยุด.....	60

รูปที่ 22 ก่อนการแทนค่าที่มีความหมายเหมือนคำคุณลักษณะ	61
รูปที่ 23 การแทนค่าที่มีความหมายเหมือนคำคุณลักษณะ	61
รูปที่ 24 การแทนค่าคำระบุชี้วัดความคิดเห็นโดยใช้ คลังคำ SentiWordNet	63
รูปที่ 25 การแทนค่าคำระบุชี้วัดความคิดเห็นโดยใช้ SentiWordNet และตัดคำที่ไม่เกี่ยวข้องออกไป	63
รูปที่ 26 คำระบุชี้วัดความคิดเห็นที่ปรากฏอยู่ข้างหน้าและข้างหลังคำคุณลักษณะ	64
รูปที่ 27 กำจัดคำระบุชี้วัดความคิดเห็นที่ปรากฏอยู่หลังคำคุณลักษณะ	64
รูปที่ 28 ผลลัพธ์จากการวิเคราะห์ชี้วัดความคิดเห็น	65
รูปที่ 29 การวัดประสิทธิภาพค่า Precision	69
รูปที่ 30 การวัดประสิทธิภาพค่า Recall	70
รูปที่ 31 การวัดประสิทธิภาพค่า F-measure	70
รูปที่ 32 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 5	71
รูปที่ 33 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 14	71
รูปที่ 34 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 18	72
รูปที่ 35 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 21	73
รูปที่ 36 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 25	73
รูปที่ 37 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 26	74



บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

การบริหารลูกค้าสัมพันธ์ (Customer Relationship Management : CRM) เป็นกลยุทธ์ทางธุรกิจเพื่อสร้างความสัมพันธ์ระยะยาวกับลูกค้า เรียนรู้ความต้องการที่แตกต่างกันของลูกค้า เพื่อตอบสนองความต้องการของลูกค้าด้วยสินค้าหรือบริการที่เหมาะสมกับลูกค้าแต่ละคน และยังสร้างความพึงพอใจให้กับลูกค้าเหล่านั้น และยังช่วยให้เกิดระบบที่มีประสิทธิภาพกับแหล่งวัตถุดิบ (supplier) [1] [2] ในปัจจุบันการบริหารลูกค้าสัมพันธ์ยังใช้อุปกรณ์เทคโนโลยีสารสนเทศรูปแบบใหม่ในการมีปฏิสัมพันธ์ เช่น สมาร์ทโฟน Tablet เป็นต้น โดย [3] เรียกว่า (Electronic Customer Relationship Management : eCRM) เพื่อให้ผู้ใช้บริการได้รับการบริการที่ประทับใจ ซึ่งมุ่งให้ลูกค้ามีประสบการณ์เชิงบวกกับองค์กร และสร้างความจงรักภักดีต่อสินค้าและบริการ เรียกว่าการบริหารประสบการณ์ลูกค้า (Customer Experience Management :CEM) [4] ซึ่งเป็นแหล่งพบปะติดต่อสื่อสารและแสดงความคิดเห็นเกี่ยวกับภาพรวมของโรงแรม เช่น “โรงแรมนี้ดี” “โรงแรมนี้พอใช้” “โรงแรมนี้แย่” เป็นต้น นอกจากนี้ยังมีความคิดเห็นที่กล่าวถึงคุณสมบัติในด้านต่างๆ ของโรงแรม เช่น “ห้องพักสะอาด” “เดินทางมาได้สะดวก” “การบริการดี” “อาหารสดใหม่” “ราคาย่อมเยา” โดยความคิดเห็นเหล่านี้ทำให้เกิดข้อมูลจำนวนมาก และเป็นแหล่งวิเคราะห์ถึงความต้องการ แนวคิด ทศนคติของผู้ใช้บริการโรงแรม ซึ่งทางโรงแรมสามารถนำข้อมูลมาปรับปรุงและพัฒนาโรงแรมให้มีประสิทธิภาพยิ่งขึ้น และยังเป็นแหล่งข้อมูลที่สำคัญสำหรับนักท่องเที่ยว เพื่อให้ได้เข้าพักโรงแรมที่ตรงตามความต้องการของแต่ละประเภท ดังนั้นจะต้องพิจารณาปัจจัยคุณสมบัติของโรงแรมหลายๆ อย่าง เช่น ห้องพัก การบริการ สถานที่ตั้ง อาหาร ความสะอาด [5] เป็นต้น โดยวิเคราะห์ความคิดเห็นจากที่เคยเข้าพักหรือใช้บริการ ซึ่งจะช่วยให้การพัฒนาระบบ CRM ในการระบุว่าผู้เข้าพักที่สนใจคุณสมบัติต่างๆ ของโรงแรม ซึ่งในงานวิจัยด้าน CRM ในปัจจุบันได้มีความต้องการการพัฒนาระบบ CRM เพื่อตอบสนองความต้องการในการเข้าพักของผู้เข้าพักโรงแรม แต่การวิเคราะห์ความคิดเห็นของผู้เข้าพักที่มีจำนวนมากและมีรูปแบบการเขียนที่อาจไม่ถูกต้องตามโครงสร้างไวยากรณ์ ทำให้ยากที่จะอ่านและวิเคราะห์ความคิดเห็น ทำให้ความแม่นยำในการระบุความต้องการการเข้าพักยังไม่เพียงพอ [6] นำเสนอวิธีการใหม่ในการจำแนกความคิดเห็นในระดับประโยค และจำแนกข้อความความคิดเห็น แบ่งเป็นความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ [7] นำเสนอวิธีการในการวิเคราะห์ความคิดเห็นเกี่ยวกับโรงแรมของลูกค้า เพื่อนำมาตรวจสอบปัจจัยที่มีผลต่อการเลือกโรงแรมของลูกค้า [8] นำเสนอวิธีการสกัดคุณลักษณะ [9]ทำการวิเคราะห์และทำนายคะแนนค่าความพึงพอใจจากบทวิจารณ์ของลูกค้า [10] ทำการจำแนกความคิดเห็นของผู้ใช้บริการโรงแรม ต่อบริการของโรงแรม

จากงานวิจัยที่ศึกษามาข้างต้นพบว่า พบว่างานวิจัยส่วนมากเป็นการวิเคราะห์ความคิดเห็นของลูกค้าต่อสินค้า หรือบริการ เพื่อสกัดให้ได้ความคิดเห็นออกมาในเชิงบวก และความคิดเห็นในเชิง

ลบ เท่านั้น ดังนั้นในการวิจัยนี้จึงนำเสนอวิธีการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรมร่วมกับ รายละเอียดของโรงแรม เพื่อธุรกิจโรงแรมทราบความต้องการของผู้เข้าพักโรงแรม และสามารถนำ ข้อความคิดเห็นที่เป็นเชิงบวก และความคิดเห็นที่เป็นเชิงลบ เป็นสารสนเทศที่สามารถนำไปปรับปรุง และประยุกต์ใช้ในการบริหารงานอย่างมีประสิทธิภาพ เพื่อตอบสนองความต้องการของผู้เข้าพักได้ ตรงต่อความต้องการของผู้เข้าพัก เป็นอย่างดี และเพิ่มส่วนแบ่งการตลาดให้กับธุรกิจโรงแรมได้

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อจำแนกความคิดเห็นของผู้เข้าพักเปรียบเทียบกับคุณลักษณะเด่นของโรงแรม

1.3 ความสำคัญของการวิจัย

จำแนกความคิดเห็นบนเว็บไซต์ของผู้เข้าพักโรงแรมเปรียบเทียบกับคุณลักษณะเด่นของ โรงแรม เพื่อตอบสนองความต้องการของผู้เข้าพักโรงแรม และพัฒนากลยุทธ์การบริการธุรกิจโรงแรม จากการประยุกต์ใช้การวิเคราะห์ข้อความคิดเห็น

1.4 ขอบเขตของการวิจัย

1. ข้อความคิดเห็นที่นำไปวิเคราะห์มาจากเว็บไซต์ www.tripAdvisor.com จำนวน 12,644 ข้อความคิดเห็น จาก 30 โรงแรม
2. วิเคราะห์ความคิดเห็นภาษาอังกฤษของผู้เข้าพักโรงแรม
3. จำแนกความคิดเห็นออกเป็นเชิงบวก เชิงลบและเปรียบเทียบกับคุณลักษณะเด่นของ โรงแรม
4. ใช้วิธีการวิเคราะห์ความคิดเห็นจากพจนานุกรม SentiWordNet

1.5 นิยามศัพท์เฉพาะ

1. การบริหารลูกค้าสัมพันธ์ (Customer Relationship Management : CRM) คือ การ สร้างความสัมพันธ์กับลูกค้า
2. การบริหารประสบการณ์ลูกค้า (Customer Experience Management : CEM) คือ มุ่ง ให้ความสำคัญประสบการณ์เชิงบวกกับองค์กร เพื่อสร้างความจงรักภักดีต่อสินค้าและบริการ

พหุ มณ สิโต ชีเว

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยเรื่องการจำแนกความคิดเห็นของผู้เข้าพักโรงแรมเปรียบเทียบกับคุณลักษณะเด่นของโรงแรม ในงานวิจัยนี้ได้รวบรวมข้อมูลเอกสาร หลักการทฤษฎี และงานวิจัยที่เกี่ยวข้องต่างๆ ดังต่อไปนี้

2.1 การบริหารลูกค้าสัมพันธ์

ในปัจจุบันการดำเนินธุรกิจขององค์กร จะประสบผลสำเร็จมากน้อยเพียงใด ลูกค้ามีส่วนสำคัญเป็นอย่างมาก การรักษาลูกค้าให้อยู่ได้นานเป็นสิ่งสำคัญที่องค์กรต่างให้ความสนใจ นอกจากนี้ทุกบริษัท ต้องรวบรวมข้อมูลต่างๆ ที่เกี่ยวกับลูกค้าให้มากที่สุดเพื่อให้องค์กรสามารถสร้างและรักษาความจงรักภักดีของลูกค้าที่มีต่อธุรกิจ เนื่องจากลูกค้ามีความสำคัญเป็นอย่างยิ่งต่อธุรกิจและเป็นผู้สร้างกำไรในระยะยาวให้กับองค์กร การบริหารลูกค้าสัมพันธ์ (Customer Relationship Management : CRM) [6] จึงเป็นเครื่องมือหนึ่งที่องค์กรหลายๆ องค์กรนำมาใช้เป็นอย่างมาก

การบริหารลูกค้าสัมพันธ์ เป็นกลยุทธ์ทางธุรกิจใช้เพื่อสร้างความสัมพันธ์ระยะยาวกับลูกค้า และเป็นกระบวนการที่ทำให้ธุรกิจเข้าใจถึงพฤติกรรมของลูกค้า โดยการจัดเก็บและวิเคราะห์ข้อมูลลูกค้าขนาดใหญ่ ซึ่งทำให้ธุรกิจมีความสามารถในการวางกลยุทธ์ที่แตกต่างกันเพื่อตอบสนองความต้องการของลูกค้าที่หลากหลาย และใช้เทคโนโลยีสารสนเทศให้เกิดประโยชน์โดยการสร้างการมีปฏิสัมพันธ์กับลูกค้า CRM เป็นทั้งกลยุทธ์และเครื่องมือขององค์กร และของพนักงานในองค์กรที่จะให้การบริการแก่ลูกค้า ซึ่งกลยุทธ์ไม่ได้หมายความว่าถึงเทคโนโลยีที่มีราคาแพง แต่หมายถึงการประยุกต์ใช้เทคโนโลยีที่มีอยู่ได้อย่างมีประสิทธิภาพสูงสุด ทั้งนี้เพื่อจะช่วยลดต้นทุน ในการพัฒนาองค์กร ดังนั้นองค์กรควรให้ความสำคัญกับการกำหนดกลยุทธ์ CRM ในการกำหนดกลยุทธ์ CRM อย่างมีประสิทธิภาพ ดังนั้นทุกคนในองค์กรควรมีส่วนร่วมในการเป็นส่วนหนึ่งของการกำหนดกลยุทธ์ ซึ่งกลยุทธ์ทาง CRM จะสร้างความได้เปรียบทางธุรกิจให้กับองค์กร และยังสามารถช่วยให้องค์กรคาดการณ์ส่วนแบ่งทางการตลาดได้

2.1.1 แนวคิดพื้นฐานของการบริหารลูกค้าสัมพันธ์

การพัฒนาระบบการจัดการลูกค้าสัมพันธ์ที่มีประสิทธิภาพมีสิ่งสำคัญที่เป็น หลักของการบริหาร คือ การตระหนักถึงความสำคัญของลูกค้าแต่ละราย ว่าลูกค้าแต่ละรายนั้นมีความสำคัญไม่เท่ากัน การที่องค์กรสามารถทำให้ลูกค้าจงรักภักดีต่อองค์กรได้นั้น เป็นหัวใจหลักในการนำองค์กรไปสู่ความสำเร็จในการดำเนินธุรกิจระยะยาวซึ่ง CRM เป็นเครื่องมือทางการบริหารจัดการซึ่งถูกนำมาใช้มีวัตถุประสงค์เพื่อช่วยให้องค์กรสามารถจัดการกระบวนการต่าง ๆ ภายในองค์กรให้

สามารถตอบสนองความต้องการของลูกค้าได้ และให้ลูกค้าเกิดความพึงพอใจสูงสุดในองค์กร การบริหาร CRM จะประสบความสำเร็จได้นั้นมีขั้นตอนที่สำคัญ ดังนี้ คือ

1. การร่วมมือกันอย่างทันท่วงทีในการดำเนินกลยุทธ์ CRM ของบุคลากรทุกระดับในองค์กร
2. พนักงานทุกระดับและทุกหน่วยเก็บข้อมูลเพื่อสนับสนุนระบบ CRM อย่างถูกต้อง
3. เครื่องมือ CRM จะต้องสอดคล้องกับตัวระบบการบริหารเพื่อให้พนักงานและลูกค้ามีความสะดวกในการใช้งาน
4. ใช้ข้อมูลรายงาน CRM ที่จำเป็นและมีการแบ่งปันไปสู่ทีมงาน
5. การดำเนินกลยุทธ์ CRM นั้นไม่ใช่การมุ่งเน้นการนำเทคโนโลยีราคาแพงเป็นหัวใจสำคัญ แต่ต้องการสามารถใช้เทคโนโลยีที่มีอยู่แล้ว ถึงแม้ว่าจะเป็นเทคโนโลยีราคาถูกแต่ต้องการสามารถใช้ให้เกิดประสิทธิภาพได้ หากเปรียบเทียบกับ การนำเทคโนโลยี ไฮ-เทคเข้ามาใช้แล้วทำให้เกิดความวุ่นวาย และเพิ่มต้นทุนมหาศาล การใช้เทคโนโลยีที่มีอยู่ก็จะก่อให้เกิดคุณค่ามากกว่า
6. การทำ CRM จะเป็นตัวช่วยบอกองค์กรว่าควรจจะรักษาลูกค้าประเภทใดแนวความคิดเกี่ยวกับการเก็บรักษาลูกค้าให้นาน ๆ นั้นจะช่วยลดต้นทุน เนื่องจากขั้นตอนในการทำงานเท่าเดิม องค์กรไม่ต้องเริ่มกระบวนการทำงานใหม่บ่อย ๆ ถ้าหากลูกค้าเข้า ๆ ออก ๆ จะทำให้เสียต้นทุนและไม่เกิดโอกาสในการทำกำไร ซึ่งโอกาสในการทำกำไรนั้นส่วนหนึ่งมาจาก การทำ Cross Selling และ Up Selling

2.1.2 การบริหารประสบการณ์ลูกค้า

การบริหารประสบการณ์ลูกค้า (Customer Experience Management : CEM) [4] คือ การสร้างประสบการณ์แบบองค์รวม การสร้างความจงรักภักดีในสินค้าและบริการ เป็นการทำให้ลูกค้ามีประสบการณ์เชิงบวกกับองค์กร องค์ประกอบที่จะสร้างให้เกิดความจงรักภักดีกับสินค้าและบริการ มีดังนี้

1. การสร้างตราสินค้า เป็นการสร้างมาตรฐานให้กับสินค้าและบริการ เพื่อให้ผู้บริโภคมีความเชื่อมั่นต่อสินค้าและบริการ นอกจากนี้ยังเกิดความพึงพอใจต่อสินค้าและบริการ
2. กำหนดกลุ่มลูกค้าเป้าหมาย มีการกำหนดกลุ่มลูกค้าเป้าหมายให้ชัดเจน หากเราทราบกลุ่มลูกค้าเป้าหมายที่ชัดเจนแน่นอนแล้วนั้น ทำให้สามารถผลิตสินค้าหรือบริการที่ตรงตามความต้องการของผู้บริโภค
3. สร้างความพึงพอใจในสินค้าหรือบริการให้ตรงตามความต้องการของผู้บริโภค ทำให้เกิดประสบการณ์ที่ดีและเกิดการซื้อซ้ำในสินค้าหรือบริการนั้นๆ
4. การใช้นวัตกรรมใหม่ในสินค้าหรือบริการ เป็นการนำจุดเด่นหรือเอกลักษณ์ในสินค้าและบริการไปต่อยอดเพื่อให้เกิดการพัฒนาผลิตภัณฑ์ เป็นการสร้างมูลค่าเพิ่มให้กับสินค้าหรือบริการ และ

เป็นการกระตุ้นให้เกิดความต้องการซื้อสินค้าหรือบริการ เป็นการรักษาสถานะผู้บริโภคเดิม และเป็นการขยายฐานผู้บริโภคใหม่ให้กับองค์กรหรือธุรกิจ

2.2 การประมวลผลภาษาทางธรรมชาติ

ภาษาเป็นเครื่องมือที่ใช้สื่อสารระหว่างมนุษย์ด้วยกัน หรือแม้กระทั่งสื่อสารกับคอมพิวเตอร์ แต่ภาษาที่ใช้มีรูปแบบที่แตกต่างกันไป ส่วนภาษาที่มนุษย์ใช้สื่อสารกับคอมพิวเตอร์ เป็น ภาษาที่มีโครงสร้างที่แน่นอน คอมพิวเตอร์สามารถนำไปประมวลผลได้ทันที เช่น php, java หรือ C++ แต่สำหรับภาษาที่มนุษย์ใช้สื่อสารกันเองเป็นภาษาที่ไม่มีโครงสร้างหรือรูปแบบเฉพาะตัว และเป็นไปตามธรรมชาติของการเรียนรู้ในสมองของมนุษย์แต่ละคน ซึ่งมีลักษณะที่แตกต่างกันไป เรียกว่า “ภาษาธรรมชาติ” เป็นภาษาไม่มีโครงสร้างที่แน่นอน การที่คอมพิวเตอร์พยายามทำความเข้าใจกับภาษาธรรมชาติที่มนุษย์ใช้สื่อสารกันนั้น สามารถทำได้ด้วยวิธีการแทนความรู้ การสร้างกฎเกณฑ์ และการประเมินค่าเพื่อหาความหมายของภาษา [11] ดังนั้นการที่คอมพิวเตอร์จะเข้าใจภาษาธรรมชาติได้ดีเพียงไรนั้นขึ้นอยู่กับ 2 แนวทางคือ 1) พัฒนาการทางด้านปัญญาประดิษฐ์ ซึ่งเป็นวิธีการแทนความรู้ (Knowledge representation) และ 2) การศึกษาและเข้าใจโครงสร้างทางภาษาศาสตร์แบบมีโครงสร้าง ซึ่งทั้งสองกระบวนการดังกล่าว ถูกเรียกว่า การประมวลผลภาษาธรรมชาติ ซึ่งได้แบ่งระดับขั้นการวิเคราะห์ ดังนี้

1. การวิเคราะห์ระดับวิจิภาค (Morphological analysis) เป็นการวิเคราะห์ระดับคำ
2. การวิเคราะห์ระดับวากยสัมพันธ์ (Syntactic analysis) เป็นการวิเคราะห์คำตามหน้าที่ของคำ (Part-of-Speech) เพื่อเป็นข้อมูลพื้นฐานในการตรวจสอบโครงสร้างทางไวยากรณ์ เกี่ยวกับการวางตำแหน่งของคำ กลุ่มคำประเภทต่างๆ ที่รวมกันเป็นประโยค
3. การวิเคราะห์ระดับความหมาย (Semantic analysis) เป็นการวิเคราะห์เพื่อให้ทราบความหมายของคำแต่ละคำในประโยค
4. วิเคราะห์ระดับวาทนิพจน์ (Discourse integration) เป็นการพิจารณาความหมาย ของประโยคโดยดูจากประโยคข้างเคียงร่วมด้วย
5. การวิเคราะห์ระดับปฏิบัติ (Pragmatic analysis) เป็นการแปลความหมายของประโยคถึงสิ่งที่ผู้พูดต้องการสื่อความหมายถึง

กระบวนการวิเคราะห์ภาษาจะเริ่มต้นที่ระดับต่ำสุดก่อน คือการวิเคราะห์ระดับวิจิภาค หรือ คำ ไปจนถึงระดับวากยสัมพันธ์ ที่สามารถอธิบายได้ด้วยโครงสร้างของภาษาที่ ประกอบด้วยหน่วยต่างๆ ดังนี้

1. คำ (Word) คือ หน่วยหนึ่งที่เปล่งเป็นเสียงออกมาจะเป็นอิสระหรือไม่ก็ได้

2. หน่วยคำ (Morpheme) คือ ส่วนประกอบที่มีนัยสำคัญที่แท้จริงทางภาษา คือ คำ และ กลุ่มคำ หรือหมายถึงหน่วยคำที่เล็กที่มีความหมาย ความแตกต่างระหว่าง หน่วยคำ กับ คำ คือ หน่วยคำอาจเกิดขึ้นอิสระหรือไม่อิสระก็ได้ แต่คำต้องเป็นหน่วยอิสระเสมอ เช่น นักเรียน “นัก” เป็นหน่วยคำ และ “นักเรียน” จึงถือเป็นคำ

3. วลี (Phrase) คือ กลุ่มของคำ หรือคำๆ เดียวก็ได้ ที่เป็นส่วนประกอบของประโยค (วลีเป็นส่วนประกอบของประโยค แต่คำไม่ใช่ส่วนประกอบของประโยค) ดังนั้นไม่ว่าจะเป็นคำๆ เดียว หรือ กลุ่มที่ประกอบด้วยหลายๆ คำ เมื่อเป็นส่วนประกอบของประโยคเราเรียกเป็น วลี

4. นามวลี (Noun phrase) คือ วลีที่ทำหน้าที่เป็นส่วนหนึ่งของประโยคที่เห็นหน่วย ประธาน หน่วยกรรม นามวลี

5. กริยาวลี (Verb phrase) คือ วลีที่ทำหน้าที่เป็นภาคแสดงของประโยค

6. ประโยค (Sentence) คือ คำหลายคำเรียงกันในการพูดหรือ เขียน เพื่อเห็นการแสดง ความคิด 1 ความคิด อย่างสมบูรณ์ทางไวยากรณ์ โดยปกติมักประกอบไปด้วยประธานและ ภาคแสดง

7. ไวยากรณ์ (Grammar) คือ ภาษาที่มีกฎเกณฑ์และเป็นไปตามระเบียบวิธีการ ประกอบรูป คำให้เป็นประโยค การวิเคราะห์ในระดับที่สูงขึ้นกว่าระดับวากยสัมพันธ์ หรือโครงสร้างของภาษา จำเป็นต้องอาศัยฐานความรู้ที่สูงขึ้น เช่น ฐานความรู้เครือข่ายคำ (WordNet) สำหรับกระบวนการ วิเคราะห์ภาษาธรรมชาติ ในภาษาไทยส่วนใหญ่ เป็นการวิเคราะห์ภาษาในระดับจิตวิภาค และดับ วากยสัมพันธ์เท่านั้น เนื่องจากเครือข่ายคำไทย (Thai WordNet) ค่อนข้างมีจำนวนจำกัดและยังอยู่ในขั้นพัฒนา การวิเคราะห์ภาษาในระดับจิตวิภาคและวากยสัมพันธ์สำหรับภาษาไทย มีความยุ่งยาก กว่าภาษาอังกฤษเล็กน้อย เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีการแบ่งวรรคตอนระหว่างประโยค หรือ ระหว่างคำที่แน่นอน มักเขียนติดกันไปทั้งย่อหน้า ดังนั้นการวิเคราะห์ในระดับต่ำสุดหรือคำ จึง จำเป็นต้องแบ่งเอกสารตั้งแต่หน่วยใหญ่สุดคือประโยค ให้มีหน่วยเล็กสุดที่มีความหมายก่อน ซึ่ง หมายถึงคำ แล้วจึงนำคำแต่ละคำมาประกอบเป็นโครงสร้างประโยค ดังนั้นการวิเคราะห์ภาษาสำหรับ ภาษาไทยต้องอาศัย 3 กระบวนการพื้นฐาน คือ (1) กระบวนการตัดคำ (Tokenization) เพื่อให้มีความสามารถในการวิเคราะห์ระดับจิตวิภาคได้ (2) กระบวนการคำคำตามหน้าที่ของคำ Part-of-Speech tagging) และ (3) กระบวนการวิเคราะห์โครงสร้างภาษา (Syntactic analysis) เพื่อให้สามารถนำคำแต่ละคำมาประกอบเป็นโครงสร้างประโยคได้

2.3 การทำเหมืองข้อมูลใน CRM

เทคนิคการสำรวจ และได้พบข้อได้เปรียบของการใช้การทำเหมืองข้อมูลโดยเฉพาะอย่างยิ่งใน คลังข้อมูล CRM เป็นดังต่อไปนี้

1. ความถูกต้องที่ดีขึ้นของข้อมูลและการเข้าถึงข้อมูลได้เร็วขึ้น
2. คุณภาพที่ดีขึ้นของข้อมูลเนื่องจากข้อมูลที่ไม่ดีและที่ซ้ำกันจะถูกตัดออก
3. สารสกัดและจัดการข้อมูลได้รวดเร็วยิ่งขึ้นสำหรับการวิเคราะห์
4. ผลการวิเคราะห์แสดงในรายงาน
5. ความสามารถในการคำนวณและประเมินมูลค่าในอนาคต

ข้อได้เปรียบเหล่านี้ของคลังข้อมูลเป็นเหตุผลหลักที่ใช้คลังข้อมูลซึ่งเป็นเทคนิคในการ

ดำเนินการวิจัย ใช้เวลาน้อยลงในการเข้าถึงข้อมูล การวิเคราะห์ข้อมูล ส่วนปัญหาสำคัญในพื้นที่ธุรกิจเนื่องจาก เวลาของลูกค้าที่มีคุณค่ามาก ของคลังข้อมูลเพื่อสร้างรายงานที่ดีจากการวิเคราะห์จะช่วยให้องค์กร และลูกค้า ซึ่งวิธีดังกล่าวเป็นวิธีที่สะดวกในการทำมาเข้าใจสิ่งที่พวกเขาต้องการ กลยุทธ์ CRM ที่มีประสิทธิภาพสามารถผลิตโดยเทคโนโลยีการทำเหมืองข้อมูลที่มีประสิทธิภาพ การวิเคราะห์รูปแบบการทำธุรกรรมของลูกค้ากับองค์กร ความคิดเห็นต่อสินค้าของลูกค้าจากเว็บไซต์ โดยเฉพาะกลยุทธ์ที่ดีโดย องค์กร จะนำไปสู่ความสัมพันธ์ที่ดีกับลูกค้า นอกจากนี้องค์กรสามารถเพิ่มความพึงพอใจของลูกค้าและการทำกำไรให้กับองค์กรซึ่งลูกค้าเหล่านี้จะส่งเสริมภาพลักษณ์ขององค์กร Web2.0 เช่น ฟอรัม อินเทอร์เน็ตและ Facebook เป็นต้น

2.3.1 การทำเหมืองความคิดเห็น

เหมืองข้อความแสดงความคิดเห็น (Opinion Mining) หรือ การวิเคราะห์ความคิดเห็น (Sentiment analysis) เป็นศาสตร์แขนงหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) โดยเน้นการผสมผสานเทคนิคของการสืบค้นข้อมูลเข้ากับการประมวลผลทางภาษา เพื่อพัฒนาเป็นระบบที่ผู้ใช้สามารถสืบค้นและเข้าถึงข้อมูลจำนวนมากบนเครือข่าย โดยเฉพาะข้อความความคิดเห็นไม่ว่าจะอยู่ในบล็อก ฟอรัม หรือเว็บไซต์ต่างๆ ได้แก่ ข้อมูลเกี่ยวกับองค์กรและบริษัท ข้อมูลเกี่ยวกับสินค้าและบริการ หรือรายงานข่าว รวมถึงข้อความแสดงความคิดเห็นในกระดาน เว็บบอร์ด บล็อก หรือข้อความวิจารณ์ทั่วไป (Reviews and Comments) ซึ่งข้อมูลทั่วไปสามารถสืบค้นได้ด้วยเทคโนโลยีการสืบค้น (Search Engine) ที่มีการพัฒนาและใช้อยู่ในปัจจุบัน โดยผู้ใช้สามารถระบุการค้นหาด้วยคำสำคัญ (Keywords) ที่ตรงกับหัวข้อ และระบบจะสืบค้นและนำเสนอผลลัพธ์ที่ตรงตามความต้องการในการสืบค้นครั้งนั้นๆ

การทำเหมืองความคิดเห็นมีวัตถุประสงค์ในการสกัดข้อความที่เป็นข้อคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งที่สนใจ เพื่อนำไปสกัดคุณลักษณะเพื่อพิจารณาว่าแสดงไปในเชิงบวก เชิงลบ หรือเป็นกลาง จากนั้นทำการสรุปข้อมูลที่ต้องการ ตามการวิเคราะห์ความคิดเห็น แบ่งออกเป็น 3 ระดับ คือ

1. การวิเคราะห์ความคิดเห็นในระดับเอกสาร (Document-based sentiment analysis)

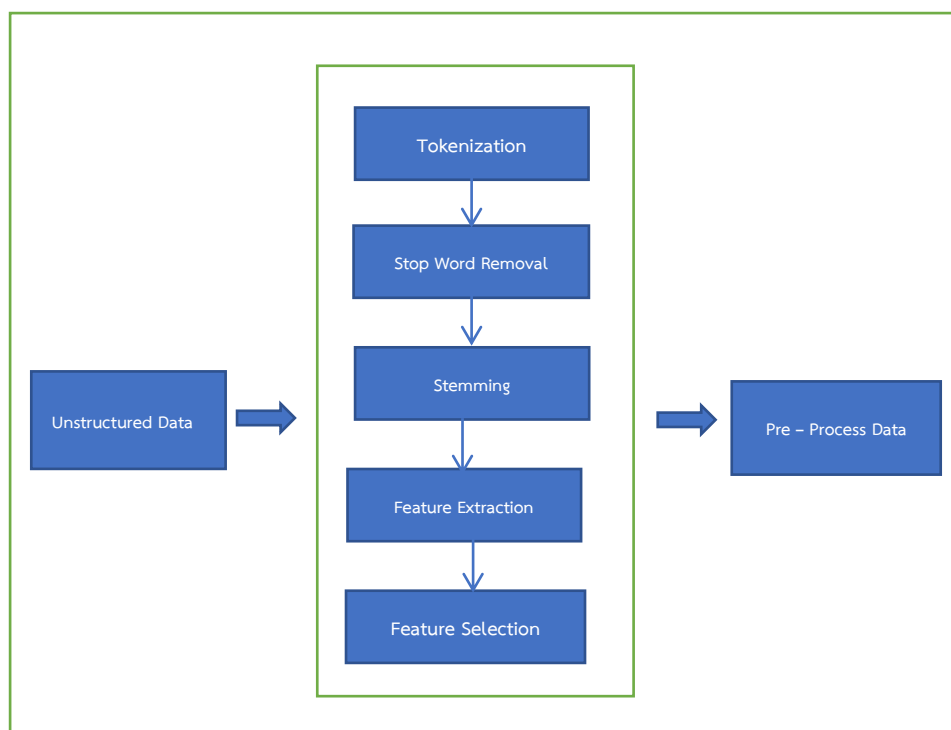
เป็นการวิเคราะห์ข้อความแสดงความคิดเห็นทั้งหมดจากเอกสารและนำมาสรุปทั้งเอกสาร เพื่อทำการระบุข้อความความคิดเห็น (polarity word) ว่าเป็นความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง

2. การวิเคราะห์ความคิดเห็นในระดับประโยค (Sentence-based sentiment analysis) ในกรณีที่ในเอกสารประกอบด้วยประโยคที่มีความคิดเห็นที่แตกต่างกัน การวิเคราะห์ความคิดเห็นในระดับเอกสารจะไม่สามารถทำการสรุปข้อความความคิดเห็นได้ จึงจำเป็นต้องทำการวิเคราะห์ความคิดเห็นในระดับประโยคโดยการแบ่งเป็นประโยคและนำมาวิเคราะห์เพื่อสรุป ว่าเป็นความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง

3. การวิเคราะห์ความคิดเห็นในระดับคำหรือวลี (Word or Phrase analysis) ในกรณีที่ในความคิดเห็นเดียวกันประกอบด้วยความคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งแตกต่างกัน การวิเคราะห์ความคิดเห็นในระดับเอกสารและระดับประโยคจะไม่สามารถนำมาใช้ได้เนื่องจากไม่มีความละเอียดมากพอ จึงต้องทำการวิเคราะห์ความคิดเห็นในระดับคุณลักษณะของวัตถุ เพื่อแยกคุณลักษณะที่สนใจของวัตถุ (Feature) ในประโยคมาทำการระบุข้อความความคิดเห็นว่าเป็นความคิดเห็นเชิงบวก เชิงลบหรือเป็นกลาง

2.4 ขั้นตอนการเตรียมข้อมูล

เนื่องจากข้อความความคิดเห็นเป็นข้อมูลที่อยู่รูปแบบที่ไม่มีโครงสร้าง ก่อนที่จะดำเนินการใดๆ จะต้องมีการเตรียมข้อมูลก่อน เพราะข้อความความคิดเห็นมีข้อมูลไม่ชัดเจน (Inconsistent) และมีข้อมูลรบกวน (Noisy) ค่อนข้างมาก ซึ่งถ้าข้อมูลไม่ชัดเจนหรือมีข้อมูลรบกวนมาก เป็นไปได้ว่าจะนำไปสู่ความ สับสนและผลลัพธ์ที่ผิดพลาด การเตรียมข้อมูลจะช่วยให้ข้อมูลมีความชัดเจน (Consistent) และ ถูกต้อง (Accuracy) วิธีการเตรียมข้อมูล [12] [13] ประกอบด้วย 5 กระบวนการ ได้แก่ กระบวนการตัดคำ (Tokenization) กำจัดคำหยุด (Stop Word Removal) การรากคำศัพท์ (Stemming) การสกัด คุณลักษณะ (Feature Extraction) และการเลือกคุณลักษณะ (Feature Selection) ดังรูปที่ 1



รูปที่ 1 ขั้นตอนการเตรียมข้อมูลอยู่รูปแบบที่ไม่มีโครงสร้าง

1. การตัดคำ

การตัดคำ คือ กระบวนการนำเอกสารข้อความนำมาแบ่งออกเป็นประโยค (Sentence) หรือแบ่งออกเป็นคำ (Word) ขึ้นอยู่กับวัตถุประสงค์ในการวิเคราะห์ โดยทั่วไปการตัดคำในข้อความภาษาอังกฤษนิยมใช้ ช่องว่าง (White Space) คอมา (Comma) จุดทศนิยม (Point) หรือสัญลักษณ์ต่างๆ [14] กระบวนการตัดคำ เริ่มจากสแกนข้อความทั้งหมดเพื่อหาขอบเขตของคำ และขอบเขตของประโยคซึ่งในคำภาษาอังกฤษจะใช้ช่องว่างในการแบ่งคำ และใช้จุดทศนิยมหลังประโยค [15] ดังแสดงตัวอย่างในตารางที่ 1

ตารางที่ 1 ตัวอย่างการตัดคำภาษาอังกฤษ

INPUT	OUTPUT
Very nice hotel in a very good area.	Very nice hotel in a very good area
Room clean wifi Internet connection slow perfect location	Room clean wifi Internet connection slow perfect location
Nice and clean place. Great value for money!	Nice and clean place Great value for money !

2. การกำจัดคำหยุด

การกำจัดคำหยุด คือ การตัดคำที่ไม่สื่อความหมายออกจากเอกสารหรือข้อความคิดเห็น ซึ่งคำหยุดจะพบมากในเอกสาร [16] คำที่ปรากฏบ่อยมากในข้อความ เช่น a, an, the, and, about, for ในภาษาอังกฤษ หรือคำว่า คือ เป็น และ หรือ ในภาษาไทย เป็นต้น สำหรับ กระบวนการนี้มีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพในการประมวลผล เพราะคำหยุดเป็นคำที่ไม่สื่อความหมาย จัดเป็นคำสิ้นเปลืองในการทำเหมืองข้อความ หากมีเป็นจำนวนมากจะทำให้การประมวลผลช้า ไม่มีประสิทธิภาพ การทำเหมืองข้อความภาษาอังกฤษ มีคลังคำหยุดที่ถูกรวบรวมและเผยแพร่ให้ นักวิจัยหรือองค์กรนำไปใช้งานอย่างแพร่หลายโดยกระบวนการนี้จะนำข้อความที่ได้ทำการตัดคำ แล้วมาตรวจสอบว่ามีคำที่ตรงกับคลังคำหยุดหรือไม่ ถ้ามีก็จะลบคำเหล่านั้นออกไป ตัวอย่างคำหยุด เช่น คำบุพบท (Prepositions) คำสรรพนาม (Pronouns) คำเชื่อม (Conjunctions) คำนำหน้านาม (Articles) [17] [18] ตัวอย่างคำหยุดภาษาอังกฤษและภาษไทยดังตารางที่ 2 และ ตารางที่ 3 ตามลำดับ

ตารางที่ 2 ตัวอย่างคำหยุดภาษาอังกฤษ

Prepositions	On	In	At	Since	For	To
Pronouns	He	She	It	Her	We	They
Conjunctions	And	Or	But	Because	So	Although
Articles	A	And	The			

ตารางที่ 3 ตัวอย่างคำหยุดภาษาไทย

คำบุพบท	กับ	ของ	ซึ่ง	ด้วย	โดย	สำหรับ
คำสรรพนาม	กระผม	ข้าพเจ้า	เธอ	ฉัน	นาย	คุณ
คำเชื่อม	กว่า	คือ	จึง	แต่	ถ้า	และ
คำวิเศษ	ใกล้	ครับ	ด้วย	ที่สุด		

3. การหารากคำศัพท์

การหารากคำศัพท์ [16] เป็นวิธีการลดจำนวนคำ โดยการจัดกลุ่มคำที่มีความหมายเหมือนกันไว้ด้วยกัน เช่น Beautiful, Beautifully, Beautify, Beautifies หรือคำที่มีความหมาย

ใกล้เคียงกัน จะถูกจัดไว้ในกลุ่มเดียวกัน จากคำศัพท์ตัวอย่างจะได้รากคำศัพท์คือ Beauty รวมทั้งตรวจสอบและแก้ไขคำที่สะกดผิดให้ถูกต้อง แก้ไขคำที่ไม่สมบูรณ์ให้เป็นคำที่สมบูรณ์ เช่น “This flower is Beautifulllll” จากข้อความข้างต้น คำว่า “Beautifulllll” จะถูกแก้ไขเป็น “Beautiful” ในการหารากคำศัพท์โดยส่วนมากจะใช้คลังคำเข้ามาช่วย ประโยชน์ของกระบวนการนี้คือ จะทำให้จำนวนคุณลักษณะลดลง ช่วยเพิ่มความเร็วในประมวลผล

4. การสกัดคุณลักษณะ

การสกัดคุณลักษณะ เป็นขั้นตอนดึงคุณลักษณะของข้อความออกมา ซึ่งการดึงคุณลักษณะนั้นจะต้องมีการกำหนดประเภทของคุณลักษณะที่จะใช้เป็นตัวแทน จากงานวิจัยที่ผ่านมาพบว่าประเภทคุณลักษณะที่นิยมใช้ในการทำเหมืองความคิดเห็น ได้แก่ การแทนข้อความด้วยถุงคำ (Bag-of-Word) การแทนข้อความด้วยค่าความถี่การเกิดคำ (Term Frequency: TF) และการแทนข้อความด้วยค่าความถี่ผกผัน (Term Frequency – Inverse Document Frequency: TF-IDF) [19]

5. การเลือกคุณลักษณะ

การเลือกคุณลักษณะ เป็นวิธีการเบื้องต้นในการลดขนาดของคุณลักษณะที่ได้จากขั้นตอนสกัดคุณลักษณะ เนื่องจากคุณลักษณะแต่ละตัวมีความสำคัญไม่เท่ากันแต่ด้วยปริมาณข้อมูลที่มีอยู่เป็นจำนวนมากจึงทำให้จำนวนคุณลักษณะมีมาก จะส่งผลต่อประสิทธิภาพในการประมวลผล เนื่องจากอัลกอริทึมในการสร้างแบบจำลองการวิเคราะห์ข้อมูลโดยทั่วไป ไม่สามารถรองรับการทำงานกับจำนวนคุณลักษณะที่มีจำนวนมากได้ดี การเลือกเทคนิคสำหรับคัดเลือกคุณลักษณะจึงเป็นสิ่งสำคัญในกระบวนการนี้ เพื่อให้ได้คุณลักษณะที่ใช้เป็นตัวแทนของข้อมูลส่วนใหญ่ได้ เทคนิคการเลือกคุณลักษณะที่นิยมคือ (Information Gain: IG)

2.5 วิธีการจำแนกประเภทความคิดเห็น (Opinion Classifier Algorithm)

มี 3 วิธี ได้แก่ วิธีการเลกซิคอนเบส วิธีการแมชชีนเลิร์นนิง และวิธีการผสมผสาน

2.5.1 วิธีการเลกซิคอนเบส

วิธีการเลกซิคอนเบสเป็นการจำแนกความคิดเห็นโดยการใช้พจนานุกรมคำ ระบุข้อความความคิดเห็นที่ถูกกำหนดคำในแต่ละโดเมนไว้แล้ว พจนานุกรมที่ใช้ในการทำเหมืองความคิดเห็นส่วนมากจะมีค่าแสดงความรู้สึก 2 ด้าน คือ ค่าแสดงความรู้สึกเชิงบวก (Positive Words) และ ค่าแสดงความรู้สึกเชิงลบ (Negative Words) วิธีการรวบรวมค่าแสดงความคิดเห็น มี 3 วิธีการคือ Manual Approach, Dictionary-based Approach และ Corpus-based Approach ซึ่งมีวิธีการดังนี้

1. Manual Approach เป็นวิธีรวบรวมคำศัพท์โดยวิธีการทำเอง ซึ่งต้องใช้

แรงงานจำนวนมากและใช้เวลานานมาก งานวิจัยส่วนมากจะใช้วิธีนี้ร่วมกับวิธีอื่นที่เป็นอัตโนมัติ เช่น การรวบรวมคำศัพท์พื้นฐานในขั้นตอนแรก หรือการตรวจสอบคำศัพท์ในขั้นตอนสุดท้าย วิธีการอัตโนมัติอาจมีข้อผิดพลาดได้

2. Dictionary-based Approach เป็นวิธีการรวมศัพท์พื้นฐาน โดยไปยังคำพ้องเสียง (Synonyms) และคำตรงกันข้าม (Antonyms) โดยรวบรวมจากพจนานุกรมคำที่ได้รับการยอมรับ เช่น WordNet เพราะพจนานุกรมคำเหล่านั้นจะมีรายการคำพ้อง และคำตรงข้าม วิธีนี้มีขั้นตอนการทำงานดังนี้

2.1) รวบรวมคำที่แสดงความรู้สึกพื้นฐานที่บอกได้ว่าเป็นคำแสดงทัศนคติในเชิงบวกหรือเชิงลบ

2.2) เพิ่มคุณสมบัติของคำโดยการค้นหาใน WordNet หรือ พจนานุกรมออนไลน์อื่นๆ เพื่อค้นหาคำพ้องเสียงและคำตรงข้าม

2.3) เพิ่มคำใหม่ที่ค้นพบลงในชุดคำสั่งที่สร้างขึ้นในกระบวนการแรก

2.4) ทำซ้ำทุกกระบวนการจนไม่พบคำใหม่

2.5) ตรวจสอบความถูกต้องของคำทั้งหมดและข้อผิดพลาดอื่นๆ

3. Corpus-based Approach เป็นวิธีที่นำมาใช้เพื่อแก้ปัญหาของคำที่ไม่มีในพจนานุกรม เช่น คำเฉพาะ คำกำกวม เป็นต้น เป็นวิธีการที่แก้ปัญหา 2 วิธีแรก แต่มีข้อจำกัด คือ ฐานข้อมูลต้องใหญ่มาก เพื่อให้ได้สถิติข้อมูลที่น่าเชื่อถือ

การจำแนกความคิดเห็นด้วยวิธีการเลกซิคอนเบส มี 3 กระบวนการหลัก ได้แก่ 1.การสกัดความรู้สึกจากคลังคำความคิดเห็น 2.คำนวณข้อผิดพลาด 3.ระบุข้อผิดพลาด

2.5.2 วิธีการแมชชีนเลิร์นนิง

วิธีการแมชชีนเลิร์นนิง เป็นวิธีกระบวนการเรียนรู้เพื่อให้เครื่องคอมพิวเตอร์สามารถทำงานได้อย่างมีประสิทธิภาพ โดยอาศัยการเรียนรู้จากข้อมูลอดีตที่รับเข้าไป เป็นได้ 2 เทคนิค คือ

1. การเรียนรู้แบบมีผู้สอน [20] (Supervised Learning) เป็นเทคนิคที่เน้นการเรียนรู้จากข้อมูลที่มีในอดีต หรือข้อมูลชุดสอน (Training Set) สำหรับสร้างโมเดลเรียนรู้ เพื่อวิเคราะห์หาคำตอบให้กับข้อมูลชุดใหม่ (Test Set) การใช้ข้อมูลสอนจำนวนมาก มีข้อดีคือ ทำให้การจำแนกความคิดเห็นมีความถูกต้องสูง แต่ก็ใช้เวลาเป็นเวลานาน

1.1) นาอิวเบย์ (Naïve Bayes) เป็นวิธีการเรียนรู้แบบมีผู้สอนวิธีหนึ่ง ที่ได้รับความนิยมอย่างมาก ในการจำแนกความคิดเห็น โดยใช้หลักการความน่าจะเป็น ดังสมการที่ (2.1)

$$P(\text{Class}) = \frac{\text{Count}(W|\text{Class})}{(\text{Count}(W))} \quad (2.1)$$

โดยที่ $P(Class)$	คือ กลุ่มข้อมูล เช่น กลุ่มความคิดเห็นเชิงบวก (Positive Class)
$(Class)$	คือ กลุ่มความคิดเห็นเชิงลบ (Negative Class)
$(W Class)$	คือ จำนวนค่าที่อยู่ในคลาส
W	คือ จำนวนค่าที่อยู่ในเอกสาร

ข้อดีของวิธีนี้ออฟเบย์ คือ มีวิธีการคำนวณง่าย และสะดวกต่อการนำไปใช้งาน ได้ผลลัพธ์ที่สามารถนำไปประยุกต์ใช้ได้ดี แต่มีข้อเสีย คือ ใช้ได้กับแอทริบิวที่เป็นอิสระกันเท่านั้น

1.2) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) เป็นวิธีการเรียนรู้เชิงสถิติ และเป็นวิธีการเรียนรู้แบบมีผู้สอน (Supervised Learning Technique) นิยมใช้ในการจำแนกข้อมูล (Classification) เป็นวิธีการจำแนกข้อมูลออกเป็น 2 กลุ่ม โดยใช้เส้นระนาบแบ่งข้อมูล (Hyperplane) ซึ่งข้อมูลจะถูกแบ่งด้วยเส้นไฮเปอร์เพลน มากกว่า 2 เส้น เพื่อหาเส้นที่เหมาะสมที่สุด ซัพพอร์ตเวกเตอร์แมชชีนแบบเชิงเส้น (Linear Support Vector Machine) เป็นวิธีการแบ่งกลุ่มข้อมูลด้วยเส้นไฮเปอร์เพลน โดยพิจารณาจากระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่มที่ใกล้กันมากที่สุด เช่น เมื่อพิจารณาข้อมูลที่ประกอบด้วยข้อมูล 2 กลุ่ม ดังสมการที่ (2.2)

$$D = ((x_1, y_1), \dots, (x_n, y_n))$$

$$x_i \in \mathbb{R}$$

$$y_i \in (-1, 1)$$
(2.2)

โดยที่ n คือ จำนวนของข้อมูล

i คือ ข้อมูลลำดับที่ $1, 2, \dots, n$

x คือ จำนวนจริงใดๆ

y มีค่า $-1, 1$

เมื่อมีข้อมูลนำเข้า x และต้องการจะจำแนกข้อมูล x ว่าอยู่ในกลุ่มใด พิจารณาได้จากเงื่อนไข ดังสมการที่ (2.3)

$$\begin{aligned} \text{ถ้า } \bar{w} * \bar{x} + b &> 0 && \text{แสดงว่า } x \text{ เป็นกลุ่ม } 1, (y = +1) \\ \text{ถ้า } \bar{w} * \bar{x} + b &< 0 && \text{แสดงว่า } x \text{ เป็นกลุ่ม } 2, (y = -1) \end{aligned}$$
(2.3)

โดยที่ $\bar{w}$ คือ เวกเตอร์น้ำหนัก

$\bar{x}$ คือ เวกเตอร์ข้อมูล

b คือ ค่าโน้มน้าเอียง (bias)

1.3) เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: K-NN) การจำแนกประเภทข้อมูล ด้วยวิธีเพื่อนบ้านใกล้ที่สุด จะขึ้นกับข้อมูลที่อยู่ใกล้เคียง K ตัว หลักการทำงานคือ ทำการวัดระยะห่างระหว่างข้อมูลที่ต้องการทำนายกับข้อมูลที่อยู่ใกล้เคียง จำนวน K ตัว คำตอบที่ทำนายได้ คือ คลาสที่พบมากที่สุดของข้อมูลที่อยู่ใกล้กันทั้งหมด K ตัว การนำวิธีการ K-NN มาใช้กับการจำแนกความคิดเห็น จะกำหนดค่า K ให้เป็นเลขคี่ วิธีการนี้ เหมาะสำหรับข้อมูลที่เป็นตัวเลข และไม่เหมาะสมกับข้อมูลที่มีแอตทริบิวจำนวนมาก เนื่องจากเสียเวลาในการคำนวณหาค่าความห่างเพื่อแบ่งกลุ่ม โดยทั่วไปจะใช้วิธีการวัดระยะทางแบบ Euclidean Distance คือ การหาค่ารากที่สองของผลต่างระหว่างแอตทริบิวแต่ละตัวยกกำลังสอง ดังสมการที่ (2.4)

$$D_{Euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_L - y_L)^2} \quad (2.4)$$

โดยที่ x_1 คือ แอตทริบิวที่ 1 ของข้อมูลชุดที่ 1

y_1 คือ แอตทริบิวที่ 1 ของข้อมูลเพื่อนบ้านที่นำมาพิจารณา

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) หรือเรียกว่า การสร้างตัวแบบในการบรรยาย (Descriptive Model) เป็นเทคนิคที่เน้นการนำข้อมูลที่มีอยู่มาศึกษาเพื่อพิจารณาหาความสัมพันธ์ของข้อมูลเป็นหลัก เทคนิคการเรียนรู้แบบไม่มีผู้สอนประกอบด้วย การแบ่งกลุ่มข้อมูล (Clustering) และการหากฎความสัมพันธ์ (Association Rule)

3. วิธีการผสมผสาน (Hybrids Methodology) การจำแนกความคิดเห็นด้วยวิธีการผสมผสาน เป็นการผสมผสานระหว่างวิธีการเลกซิคอนเบลส กับวิธีการแมชชีนเลิร์นนิง ซึ่งเป็นวิธีการที่มีประสิทธิภาพสูงสุด โดยใช้วิธีการสกัดคำตัวแทนคุณลักษณะ จากคำในเอกสารและคลังคำความคิดเห็นที่มีอยู่ เพื่อจำแนกความคิดเห็นด้วยวิธีการแมชชีนเลิร์นนิง

2.6 การวัดประสิทธิภาพในการจำแนกข้อมูล

การวัดประสิทธิภาพในงานวิจัยด้านเหมืองความคิดเห็น โดยทั่วไปจะหาค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยประสิทธิภาพโดยรวม (F-measure) ในการวิจัยต้องแบ่งข้อมูลออกเป็น 2 ส่วน เพื่อทำการวัดประสิทธิภาพ คือ ข้อมูลชุดสอน กับข้อมูลชุดทดสอบ

2.6.1 การแบ่งข้อมูล

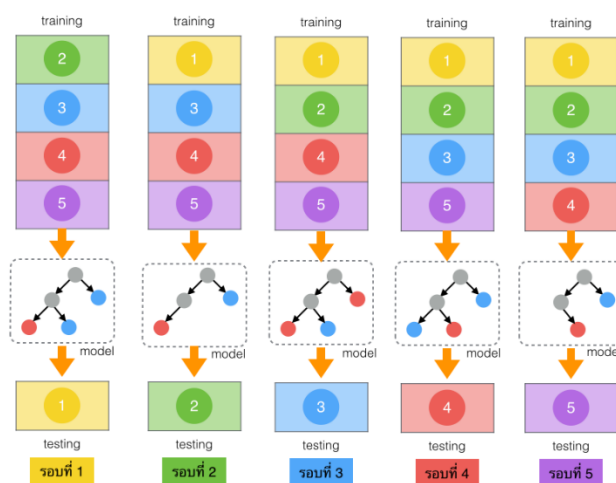
การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพการจำแนก แบ่งเป็น 3 วิธีการหลัก [20] ได้แก่

1) วิธี Self - Consistency Test หรือ Use Training Set เป็นวิธีการแบ่งที่ง่ายที่สุด คือ ข้อมูลที่ใช้ในการสร้างและทดสอบโมเดล เป็นข้อมูลชุดเดียวกัน เป็นวิธีที่ได้ผลการวัดประสิทธิภาพค่อนข้างสูง แต่ไม่เป็นที่ยอมรับและไม่เหมาะที่จะนำไปใช้ในงานวิจัยต่างๆ เนื่องจากการวัด

ประสิทธิภาพใช้ชุดข้อมูลชุดเดียวกันในการสร้างตัวจำแนก และตัวทดสอบ ทำให้เกิดความลำเอียงเกิดขึ้น

2) วิธี Split Test เป็นวิธีการแบ่งข้อมูลด้วยการสุ่ม ซึ่งทำการแบ่งข้อมูลออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 60% ต่อ 40% โดยข้อมูลส่วนที่หนึ่ง (70% หรือ 60%) ใช้ในการสร้างโมเดลและข้อมูลส่วนที่สอง (30% หรือ 40%) ใช้ในการทดสอบประสิทธิภาพของโมเดล ในการทดสอบแบบนี้ทำการสุ่มข้อมูลเพียงครั้งเดียวซึ่งในบาง ครั้งถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะคล้ายกับข้อมูลที่ใช้สร้างโมเดลจะทำให้ผลการวัดประสิทธิภาพได้ออกมาดี แต่ในทางตรงข้าม ถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะแตกต่างกับข้อมูลที่ใช้สร้างโมเดลมาก มีผลการวัดประสิทธิภาพได้ออกมาไม่ดี ดังนั้นจึงควรใช้วิธี Split Test นี้หรือทำการสุ่ม หลายครั้ง ส่วนข้อดีของวิธีนี้คือใช้เวลาในการสร้างโมเดลน้อย จึงเหมาะกับชุดข้อมูลที่มีขนาดใหญ่

3) วิธี Cross - validation Test วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัย เป็นวิธีทดสอบประสิทธิภาพของโมเดลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธีนี้ จะทำการแบ่งข้อมูลออกเป็นส่วนๆ ละเท่าๆ กัน (k ชุด) เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปจนครบจำนวนที่แบ่งไว้ เช่น การทดสอบด้วยวิธี 5-fold cross-validation ดังแสดงใน รูปที่ 2



รูปที่ 2 ตัวอย่างการแบ่งข้อมูลแบบ 5 – Fold Cross Validation ที่มา : [20]

2.6.2 การวัดประสิทธิภาพการจำแนก

การวัดประสิทธิภาพในงานวิจัยด้านเหมืองความคิดเห็น โดยทั่วไปจะวัดประสิทธิภาพ ได้แก่ [20] ค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยประสิทธิภาพโดยรวม (F-measure) อธิบายโดยใช้ตาราง Confusion Matrix คือ ตารางแบบจัตุรัส โดยมีจำนวนแถวเท่ากับจำนวนคอลัมน์และเท่ากับ จำนวนคลาส เช่น มีคลาสคำตอบอยู่ 2 ค่า คือ ความคิดเห็นด้านบวก (Positive) และ ความคิดเห็นด้านลบ (Negative) ดังนั้นตาราง confusion matrix นี้จะสร้างได้เป็นตารางขนาด 2x2 โดยข้อมูลด้านคอลัมน์คือ คลาสที่อยู่ในข้อมูลชุดสอน (Actual) และข้อมูลในแนวแถว คือ คลาสที่ทำนายได้ (Predicted) ดังตารางที่ 4

ตารางที่ 4 Confusion Matrix

Actual \ Predicted	Positive	Negative
Positive	a	b
Negative	c	d

โดยที่

a (True Positive) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาส Positive

b (False Negative) คือ จำนวนข้อมูลที่ทำนายว่าเป็นคลาส Negative แต่คำตอบคือ Positive

c (False Positive) คือ จำนวนข้อมูลที่ทำนายว่าเป็นคลาส Positive แต่คำตอบคือ Negative

d (True Negative) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาส Negative

1) การวัดค่าความถูกต้อง (Accuracy)

การวัดค่าความถูกต้องของการจำแนกความคิดเห็น เป็นการวัดความถูกต้องของวิธีการหรือรูปแบบการจำแนกประเภทข้อมูลโดยจะพิจารณาทุกคลาส คำนวณจากผลรวมของค่าที่ทำนายคลาสได้ถูกต้องหารด้วยผลรวมของค่าที่ทำนายทั้งหมด ดังสมการที่ (2.5)

$$Accuracy = \frac{a + d}{a + b + c + d} \quad (2.5)$$

2) การวัดค่าความแม่นยำ (Precision)

การวัดค่าความแม่นยำของการจำแนกความคิดเป็นเป็นการวัดประสิทธิภาพของวิธีการหรือรูปแบบการจำแนกประเภทข้อมูล โดยพิจารณาแยกทีละคลาส โดยคำนวณจากค่าที่ทำนายถูกต้อง ว่าเป็นคลาสที่พิจารณาหารด้วยผลรวมของค่าที่ทำนายถูกต้องว่าเป็นคลาสที่พิจารณา และค่าที่ทำนายว่าเป็นคลาสอื่น แต่ในความเป็นจริงอยู่ในคลาสที่พิจารณา ดังสมการที่ (2.6), (2.7) ตามลำดับ

$$\text{Precision}_{\text{positive}} = \frac{a}{a+c} \quad (2.6)$$

$$\text{Precision}_{\text{negative}} = \frac{d}{b+d} \quad (2.7)$$

3) การวัดค่าความระลึก (Recall)

การวัดค่าความระลึกของวิธีการจำแนกความคิดเห็น เป็นการวัดความถูกต้องของวิธีการ โดยจะพิจารณาแยกทีละคลาส คำนวณจากค่าที่ทำนายถูกต้องว่าเป็นคลาสที่พิจารณา หารด้วยผลรวมของค่าที่ทำนายถูกต้องว่าเป็นคลาสที่พิจารณา และค่าที่ทำนายว่าเป็นคลาสที่พิจารณาแต่คำตอบอยู่ในคลาสอื่น ดังสมการที่ (2.8), (2.9) ตามลำดับ

$$\text{Recall}_{\text{positive}} = \frac{a}{a+b} \quad (2.8)$$

$$\text{Recall}_{\text{negative}} = \frac{d}{c+d} \quad (2.9)$$

4) ค่าเฉลี่ยประสิทธิภาพโดยรวม (F-measure)

ค่าเฉลี่ยประสิทธิภาพโดยรวม เป็นการนำค่าความระลึกและค่าความแม่นยำมาพิจารณา รวมกัน ระบบที่มีประสิทธิภาพที่ดี จะต้องมีความระลึกและค่าความแม่นยำสูงใกล้เคียงกัน ดังสมการที่ (2.10) และ สมการที่ (2.11)

$$\text{F - measure}_{\text{positive}} = \frac{2 \times \text{Precision}_{\text{positive}} \times \text{Recall}_{\text{positive}}}{\text{Precision}_{\text{positive}} + \text{Recall}_{\text{positive}}} \quad (2.10)$$

$$\text{F - measure}_{\text{negative}} = \frac{2 \times \text{Precision}_{\text{negative}} \times \text{Recall}_{\text{negative}}}{\text{Precision}_{\text{negative}} + \text{Recall}_{\text{negative}}} \quad (2.11)$$

2.7 งานวิจัยที่เกี่ยวข้อง

2.7.1 งานวิจัยที่ใช้วิธีการเลิกชิคอนเบส

Khan และคณะ [21] นำเสนอการจำแนกความรู้สึกในระดับประโยคโดยใช้ความหมายของคำศัพท์เฉพาะ (Lexical) จากความคิดเห็นของลูกค้าออนไลน์ ซึ่งใช้ไดเมนชันในการจำแนกความรู้สึก ของคำศัพท์บริบทโครงสร้างประโยค โดยนำเสนอวิธีการจำแนกประโยคความคิดเห็นในระดับประโยคออกเป็นเชิงบวก เชิงลบ และเป็นกลาง โดยมีขั้นตอนต่างๆ ดังนี้ 1. แบ่งข้อความความคิดเห็นออกเป็นประโยคและเก็บไว้เป็นคลังประโยคความคิดเห็น 2. กำจัดและแก้ไขคำที่เขียนผิดในประโยคความคิดเห็น กำกับหน้าที่ของคำโดยใช้ POS Tagging ในแต่ละคำของประโยคความคิดเห็นที่เก็บไว้ในคลังคำ ซึ่งจะช่วยให้ทราบถึงตำแหน่งของแต่ละคำในประโยค 3. เลือกใช้ตำแหน่งของคำในประโยคความคิดเห็นมาแทนเป็นคุณลักษณะซึ่งจะแปลงคุณลักษณะเหล่านี้เป็นเวกเตอร์คุณลักษณะ 4. แบ่งประโยคความคิดเห็นออกเป็น 2 รูปแบบ คือ ประโยคที่เป็นความจริง (Objective) และประโยคที่เป็นความคิดเห็น (Subjective) และจำแนกความคิดเห็นออกเป็นเชิงบวก เชิงลบ และเป็นกลางด้วยวิธีการ (Lexical Dictionary) รวมถึงการตรวจสอบบริบทของคำในประโยคเพื่อใช้ในการระบุข้อความความคิดเห็น ในการจำแนกความคิดเห็นกับชุดข้อมูล Movie Reviews ด้วยวิธีการหาค่าถ่วงน้ำหนักในระดับประโยคที่นำเสนอมีความถูกต้อง 86.60% Hotel Reviews 80% และความคิดเห็น Customer Review ในระดับประโยคในการหาความหมายของคำได้ค่าความถูกต้อง 86.6% ซึ่งสูงกว่าการจำแนกในระดับประโยคโดยการหาความหมายของ Hu และ Lui (2004) คือ 84.2%

Trilla และ Alias [22] นำเสนอการวิเคราะห์ความรู้สึกจากความคิดเห็นที่ใช้สำหรับแสดงออก ข้อความเป็นคำพูดการวิจัยเพื่อปรับปรุงสถานะของ Text-to-Speech ศิลปะ (TTS) การศึกษาการสังเคราะห์ทั้งการประมวลผลของข้อความที่นำเข้าและความสามารถที่จะทำให้คำพูดที่แสดงออกธรรมชาติ มุ่งเน้นไปที่อดีตเป็นงาน Front-end ในการผลิตของการพูดสังเคราะห์บทความนี้สำรวจการปรับตัวที่เหมาะสมของขั้นตอนการวิเคราะห์ความรู้สึกเป็น เชิงบวก เชิงลบ เป็นกลาง นั้นสามารถใช้เป็นคุณลักษณะการป้อนข้อมูลสำหรับการพูดที่แสดงออก synthesis.To เหตุนี้เราประเมินแตกต่างกันของคุณสมบัติและการจำแนก เพื่อตรวจสอบขั้นตอนการปรับตัวที่เหมาะสมที่สุดประสิทธิภาพของโครงการนี้สำหรับการวิเคราะห์ความรู้สึกได้รับการประเมินโดยข้อมูลที่ใช้คือ Semeval 2,007 ชุด และคลังทวิตเตอร์ ในระดับประโยคซึ่งเป็นที่เหมาะสมสำหรับสถานการณ์ที่แสดงออก TTS การทดลองดำเนินการตรวจสอบขั้นตอนการเสนอที่เกี่ยวกับการวิเคราะห์ความรู้สึกในงานนี้แสดงให้เห็นว่าการพิจารณา unigrams ที่เกี่ยวข้องมากที่สุดเพียงอย่างเดียว (ด้วยวิธีการถ่วงน้ำหนัก) ส่งผลให้ประสิทธิภาพการจัดหมวดหมู่ที่ดีกว่าเมื่อเทียบกับการใช้คุณลักษณะเพิ่มเติม เช่น

bigrams, แท็ก POS ถึงแม้ว่าผลที่ได้รับการแสดงผลแนวโน้มประสิทธิภาพที่คล้ายกัน สำหรับการกำหนดค่าต่างๆที่มีประสิทธิภาพระดับที่แตกต่างกันมีการปฏิบัติตามจำนวนที่ส่งหมวดหมู่และปริมาณของข้อมูลที่มีอยู่ สำหรับปัญหาที่เกิดขึ้นโดยเฉพาะอย่างยิ่งการจัดการ ในงานนี้ กลยุทธ์ที่ประสบความสำเร็จการจัดหมวดหมู่ให้อัตราประสิทธิภาพที่คล้ายกันของ 56% กับ Semeval 2,007 ชุด, และ 73%

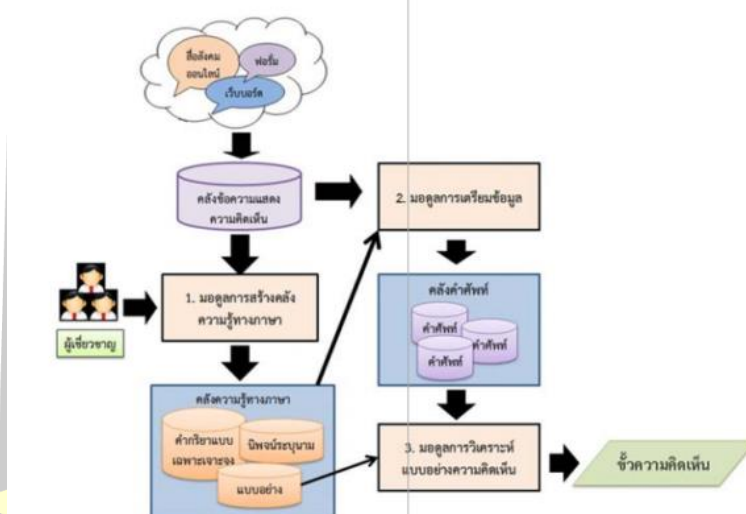
Sharma และคณะ [23] นำเสนอวิธีการจำแนกข้อความความคิดเห็นของประโยคแบบไม่มีผู้สอน จำแนกเป็น 3 ข้อคือ บวก, ลบ และ กลาง ด้วยวิธีการใช้พจนานุกรม โดยข้อมูลที่ใช้ในการวิเคราะห์เป็นการแสดงความคิดเห็นของลูกค้าต่อโทรศัพท์เคลื่อนที่จากเว็บไซต์ Amazon.com มีวิธีการทดลองคือ 1. ขั้นตอนการเตรียมการประมวลผล ระบุค่าที่เป็นความคิดเห็นในชุดข้อมูลด้วย POS tagger 2. ใช้พจนานุกรม Wordnet เพื่อระบุค่าที่เป็นความคิดเห็น 3. สกัดค่าที่เป็นความคิดเห็นโดยวิธีจับคู่กับคำที่อยู่ใน Seed list 4. กำหนดข้อของความเห็นด้วยจำนวนของคำบวกถ้าในประโยคมีคำเชิงบวกมากกว่าคำเชิงลบแสดงว่าประโยคนั้นมีความเชื่อมั่นเชิงบวก การตรวจสอบผลทดลองจะทำการอ่านประโยคและระบุข้อความความคิดเห็นด้วยคนแล้วนำมาเปรียบเทียบกับผลลัพธ์ที่ได้จากระบบ พบว่า Accuracy เป็น 74%, Precision เป็น 81% และ recall เป็น 78 %

กานดา แผ้ววัฒนากุล [19] นำเสนอกระบวนการจำแนกข้อเสนอแนะออกจากบทวิจารณ์ประเภทอื่น โดยเปรียบเทียบผลลัพธ์ของอัลกอริทึมต้นไม้ตัดสินใจ นาอ์ฟเบย์ และซัพพอร์เวกเตอร์แมชชีน เพื่อหาอัลกอริทึมที่เหมาะสมที่สุด ส่วนกระบวนการในหาจำแนกประเภทข้อเสนอแนะ แบ่งออกเป็น 4 ประเภท ได้แก่ ข้อเสนอแนะทางตรง ข้อเสนอแนะเชิงขอร้อง ข้อเสนอแนะเชิงคำถาม และข้อเสนอแนะเชิงเงื่อนไข การทดลองใช้บทวิจารณ์ทั้งหมด 2,561 ประโยค พบว่าอัลกอริทึมซัพพอร์เวกเตอร์แมชชีนแบบเคอร์เนลโพลีโนเมียล ที่มีอินพุตเวกเตอร์ประกอบด้วย คำ กับการกำกับคำที่เกิดขึ้นร่วมกันบ่อย ได้ผลลัพธ์การจำแนกข้อเสนอแนะดีที่สุด มีค่าความแม่นยำ 87.75 % ค่าความระลึก 93.62 % และค่าถ่วงดุล 89.51 % จากนั้นจำแนกประเภทข้อเสนอแนะและวัดประสิทธิภาพด้วยค่าเฉลี่ยแบบให้น้ำหนักทุกประเภทเท่ากัน (Micro Averaging) ค่าความแม่นยำ 94.94 % ความระลึก 94.94 % กระบวนการที่นำเสนอถือว่ามีความถูกต้องสูงสำหรับข้อเสนอแนะที่ไม่มีความกำกวม ช่วยลดระยะเวลาในการอ่านบทวิจารณ์และข้อเสนอแนะลงได้

ฐิติมา เกษมศรีธนาวัฒน์ และ ธนสนี เพียรตระกูล [24] นำเสนอวิธีการจำแนกความคิดเห็นตามทัศนคติ โดยใช้ตัวจำแนกแบบเบย์ร่วมกับการเลือกคุณลักษณะด้วยอัลกอริทึมรีลีฟ ว่าเป็นทัศนคติที่ดีหรือไม่ดีต่อสินค้าหรือบริการ ข้อมูลที่นำมาทดลองมาจากความคิดเห็นของผู้ใช้หนังสือคอมพิวเตอร์ ซึ่งเขียนเป็นภาษาอังกฤษ ของเว็บไซต์ amazon.com โดยเลือกจากหมวดการเขียนโปรแกรมจาวา ทั้งหมด 100 ความเห็น ซึ่งประกอบด้วยทัศนคติที่ดี 57 ความเห็น และทัศนคติที่ไม่ดี 43 ความเห็น ในขั้นตอนการเตรียมข้อมูลโดยคัดเลือกคุณลักษณะก่อนที่จะนำไปทดลองกับ Learning

Method ด้วยการใช้ Feature Selection Algorithm 2 ชนิด คือ Principal Components Analysis (PCA) และ Relief Algorithm เพื่อเพิ่มการทำงานในช่วงสอนให้มีประสิทธิภาพและทำการเปรียบเทียบประสิทธิภาพของการจำแนกความคิดเห็นตามทัศนคติโดยใช้ Learning Method 3 ชนิด ได้แก่ Naïve Bayes, Decision Tree (48) และ Multi – Layer Perceptron และประเมินความถูกต้องของการจำแนกความคิดเห็นโดยใช้วิธีการทดสอบแบบ 5-Fold Cross – Validation เมื่อทำการเปรียบเทียบผลการจำแนกแต่ละวิธีพบว่า Naïve Bayes ให้ความถูกต้องในการจำแนกสูงที่สุด คือ 68 %

มาสวีร์ มาศดิสรโชติ [25] นำเสนอการทำเหมืองความคิดเห็นภาษาไทย โดยแนวคิดและหลักการการทำเหมืองความคิดเห็นที่มุ่งเน้นประเด็นการวิเคราะห์ความคิดเห็นแบบปริยายสำหรับข้อความภาษาไทย เพื่อหาข้อสรุปว่าความคิดเห็นเหล่านั้นแสดงความรู้สึกในเชิงบวกหรือเชิงลบ จึงได้นำเสนอแนะแนวทางการวิเคราะห์ความคิดเห็นแบบปริยายสำหรับข้อความภาษาไทย โดยการสกัดคำหรือวลีในระดับคุณลักษณะของวัตถุร่วมกับการพิจารณาบริบทในบริเวณใกล้เคียง แสดงในรูปที่ 3

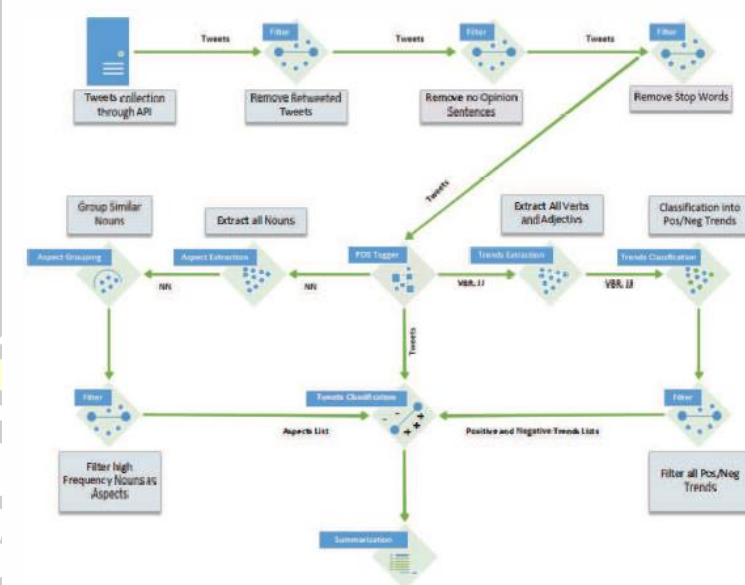


รูปที่ 3 กรอบแนวคิดในการวิเคราะห์ความคิดเห็นแบบปริยาย ที่มา :

อัจฉรา ภูระย้า และ มาลีรัตน์ โสตานิล [5] งานวิจัยนี้ เป็นการนำเสนอโมเดลทัศนคติการให้บริการของโรงแรมระดับสี่ดาว จากบทวิจารณ์ภาษาไทยและภาษาอังกฤษโดยการวิเคราะห์ด้วยนาอ็ฟ เบย์ ต้นไม้ตัดสินใจ และซัพพอร์ตเวกเตอร์ แมชชีน เพื่อการเปรียบเทียบหาวิธีที่มีค่าความถูกต้องสูงสุดของการวิเคราะห์ทัศนคติเกี่ยวกับคุณสมบัติ 5 ด้าน ได้แก่สภาพห้องพัก ที่ตั้งสภาพแวดล้อม การให้บริการ สิ่งอำนวยความสะดวกและราคา พบว่า ซัพพอร์ตเวกเตอร์แมชชีน ได้ค่าความถูกต้องสูงสุด ผู้วิจัยวิเคราะห์บทวิจารณ์ ถึงคุณสมบัติย่อย ซึ่งประกอบในคุณสมบัติหลักแต่

ละด้าน ซึ่งจำแนกเป็นการวิจารณ์ด้านบวก หรือการวิจารณ์ด้านลบของคุณสมบัติหลัก โดยการคำนวณค่าน้ำหนักนำคุณสมบัติด้านสภาพห้องพัก มาเป็น Training Set แล้วเข้าสู่กระบวนการ Text Processing ซึ่งมี 4 ขั้นตอน คือ (1) Tokenization (ตัดคำ) (2) Filter Token (ตัดคำที่มีความยาวไม่ตรงกับที่ต้องการ) (3) Stem (ทำคำรากศัพท์) (4) Stop Word Filtering (ตัด stop word) แล้วทำการเรียนรู้ด้วย Machine Learning Algorithm ได้แก่ นาอ์ฟเบย์ ต้นไม้ตัดสินใจ และซัพพอร์ตเวกเตอร์ แมชชีน โดยคุณสมบัติด้านอื่น จะนำเข้าสู่กระบวนการ เช่นกัน ส่วน Testing Set ที่ผ่านกระบวนการ Text Processing แล้ว จะทดสอบกับโมเดลของคุณสมบัติหลักแต่ละด้าน เพื่อแสดงว่าเป็นบทวิจารณ์ด้านบวก หรือบทวิจารณ์ด้านลบ

Afzaal และคณะ [26] ได้นำเสนอแนวคิดใหม่สำหรับมุมมองตามการจัดประเภทสำหรับสถานที่ท่องเที่ยว ที่สกัดจากข้อมูลเกี่ยวกับการท่องเที่ยวจาก Twitter ที่นำเสนอประกอบด้วยหกขั้นตอน 1) การรวบรวมข้อมูลผ่านกราฟ API ทวิตเตอร์ 2) การเตรียมกรองข้อมูล 3) การกำกับหน้าที่ของคำ 4) การสกัดมุมมอง 5) การสกัดแนวโน้ม 6) การจำแนกข้อมูลทวิตเตอร์ นักท่องเที่ยวสามารถอ่านความคิดเห็นของผู้เข้าชมก่อนหน้านี้ ซึ่งจะได้รับข้อมูลเกี่ยวกับสถานที่ท่องเที่ยว ที่นักท่องเที่ยวที่ต้องการท่องเที่ยว และจะเป็นประโยชน์ในการตัดสินใจเกี่ยวกับการท่องเที่ยวไปยังสถานที่ท่องเที่ยวใด ๆ แสดงในรูปที่ 4



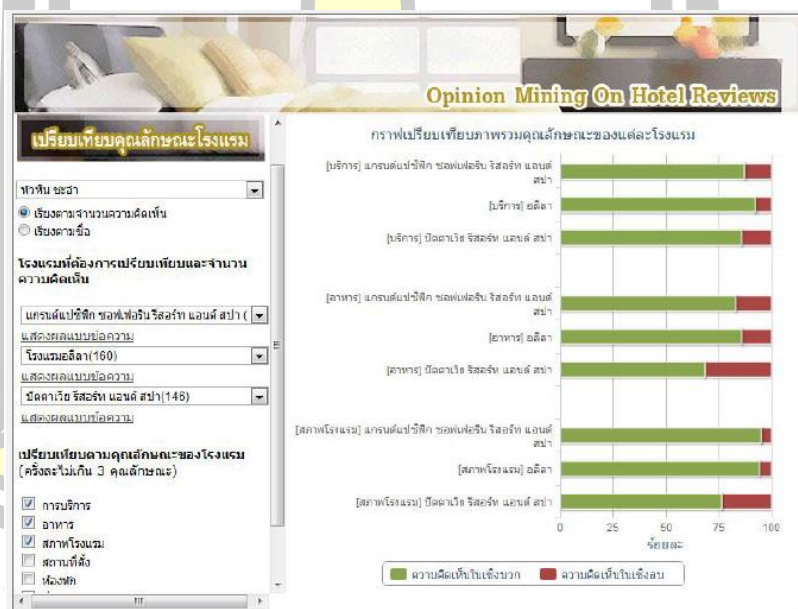
รูปที่ 4 รูปแบบแนวคิดใหม่สำหรับมุมมองตามการจัดประเภทสำหรับสถานที่ท่องเที่ยว

ที่มา : [26]

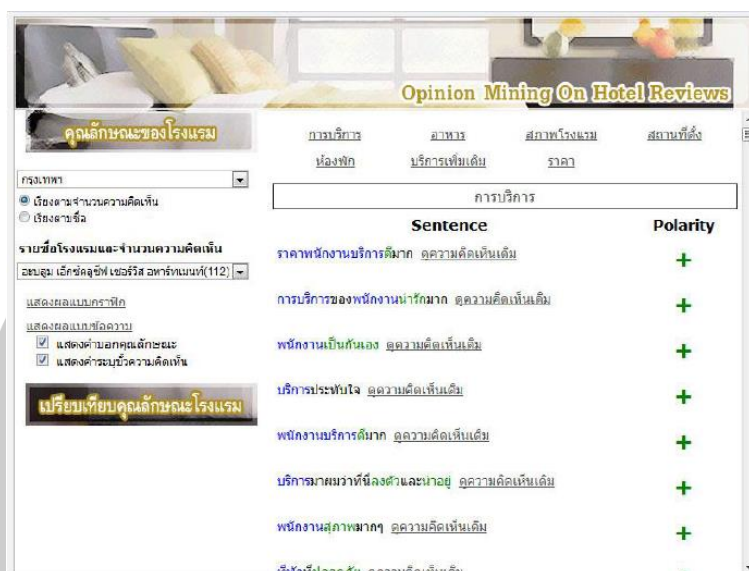
Kanzutaka และคณะ [27] นำเสนอเทคโนโลยีพื้นฐานในการพัฒนาระบบวิเคราะห์ข้อมูลการท่องเที่ยว โดยใช้ข้อมูลบนทวิตเตอร์ วิธีการดึงทวิต (ประโยคโพสต์บริการ) ที่เกี่ยวข้องกับ

สถานที่เป้าหมายและกิจกรรมการท่องเที่ยวในการสกัดใช้คำสั้งพื้นฐานและคำที่เกี่ยวข้องโดยใช้ Okapi-BM25 สกัดประโยค และวิธีการเรียนรู้ของเครื่องอยู่บนพื้นฐานของ Bayes คำถูกต้องเพิ่มขึ้น 89% - 92 % จากนั้นทำการวิเคราะห์ข้อความความคิดเห็นของทวิตเตอร์ ว่าเป็นความคิดเห็นเชิงบวกหรือความคิดเห็นเชิงลบ

วรัญญา วรณศรี ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม เป็นระบบที่สามารถวิเคราะห์ข้อมูลที่อยู่ในรูปแบบการแสดงความคิดเห็นด้วยภาษาธรรมชาติ โดยอาศัยหลักการการวิเคราะห์และสรุปผลจากทัศนคติในระดับคุณลักษณะ (Feature-based sentiment analysis and summarization) ทำให้ข้อมูลการแสดงความคิดเห็นอยู่ในรูปแบบที่เป็นข้อมูลสรุปให้เข้าใจได้ง่ายในรูปแบบกราฟิก โดยสามารถเลือกกลับไปดูข้อความแสดงความคิดเห็นที่ถูกนำมาวิเคราะห์ในหัวข้อของมุมมองที่สนใจได้ โดยสามารถแสดงผลเชิงสถิติและในรูปแบบกราฟิก เพื่อเพิ่มความสะดวกให้แก่นักท่องเที่ยวสำหรับค้นหาข้อมูลความคิดเห็นเกี่ยวกับโรงแรมทางอินเทอร์เน็ตในการเลือกโรงแรมที่ตรงตามความต้องการของตนเอง และลดระยะเวลาในการอ่านข้อความแสดงความคิดเห็นเกี่ยวกับโรงแรม ด้วยข้อมูลสรุปในรูปแบบกราฟิก ดังรูปที่ 5 และ รูปที่ 6 ตามลำดับ



รูปที่ 5 สรุปความคิดเห็นจากการโรงแรมที่อยู่ในสถานที่ที่ผู้ใช้สนใจ ที่มา : [28]



รูปที่ 6 ข้อความแสดงความคิดเห็นที่นำมาสรุปเป็นข้อมูลทางสถิติ ที่มา : [28]

Bucur [29] นำเสนอเทคนิคการทำเหมืองข้อมูลเกี่ยวกับการท่องเที่ยว งานวิจัยนี้เป็น การจำแนกความคิดเห็นเกี่ยวกับโรงแรม โดยจำแนกคุณตามความรู้สึกของผู้ใช้ที่มีต่อคุณสมบัติของ โรงแรม ในงานวิจัยนี้ได้แยกขั้นตอนในการทำงานเป็น 2 ขั้นตอน หลักคือ 1. การรวบรวมความ คิดเห็น 2. การวิเคราะห์ข้อมูล ในการรวบรวมข้อความความคิดเห็น ได้ใช้ชุดข้อมูลของเว็บไซต์ TripAdvisor.com ประกอบไปด้วย ความคิดเห็นที่เป็นเชิงบวก จำนวน 1,500 ความคิดเห็น แยกเป็น ประโยคได้ 15,377 ประโยค และเชิงลบ 1,500 ความคิดเห็น แยกเป็นประโยคได้ 16,551 ประโยค รวมทั้งหมด 31,928 ประโยค โดยใช้เครื่องมือ Crawler ซึ่งได้ถูกรวบรวมไว้ โดย Enrique Valles Balaguer and Paola Rosso (<http://users.dsic.upv.es/nle/?file=kop4.php>) ขั้นตอนการ วิเคราะห์ข้อมูล ได้ทำการวิเคราะห์ข้อมูล 3 ระดับ ได้แก่ ระดับคำ ระดับประโยค และระดับเอกสาร ในขั้นตอนการวิเคราะห์ประกอบด้วย การแบ่งคำ และใช้การกำกับหน้าที่ของคำโดย Eric Brill Algorithm เป็นภาษา PHP และในขั้นตอนการวิเคราะห์ข้อมูลแบบ Unsupervised ในการระบุค่าชี้วัด ความคิดเห็น จะใช้ SentiWordNet ซึ่งถูกกำหนดค่าตัวเลขกำกับ คะแนนอยู่ระหว่าง 0-1 และ รวม คะแนนในแต่ละประโยค โดยการระบุค่าของแต่ละคำถูกระบุโดย Adam Westerski (<http://www.gi2mo.org>) ในการคำนวณประโยคความคิดเห็นจะใช้ ค่า threshold 0.2 ในการแบ่ง ว่าประโยคนั้นเป็น เชิงบวก หรือ เชิงลบ ถ้าประโยคนั้นมีความคิดเห็นเป็นนั้นมีค่า น้อยกว่า -0.2 ถือ ว่าเป็นเชิงลบ มีค่ามากกว่า 0.2 เป็นเชิงบวก ถ้ามีค่าอยู่ระหว่าง -0.2 ถึง 0.2 ถือว่าประโยคนั้น ประกอบไปด้วย คำ Objective หรือ มีค่าเป็นกลาง ในการทดลองได้คำนวณหาค่า Precision

accuracy recall F-measure โดยใช้ตาราง confusion matrix ในการระบุผลการทำนาย ในการทดลองได้แบ่งข้อมูลในการทดลองออกเป็น 4 ชุด ความคิดเห็น โดยมีผลลัพธ์ดังตารางที่ 5

ตารางที่ 5 ผลการทดลองชุดข้อมูล 4 ชุด

จำนวนความคิดเห็น	ค่าความแม่นยำ	ค่าความระลึก	ค่าความถูกต้อง	ค่าประสิทธิภาพรวม
100	75.51%	74%	72%	74.74%
1000	78.05%	86.8%	76.5%	82.19%
2000	76.23%	86.6%	74.5%	81.06%
3000	73.66%	85.6%	72.2%	79.18%

Liu และคณะ [30] นำเสนอวิธีการทำเหมืองความคิดเห็นในระดับคุณลักษณะ โดยมีขั้นตอนในการดำเนินงานวิจัย 3 ขั้นตอนคือ 1. สกัดคำคุณลักษณะและคำระบุข้อความความคิดเห็น 2. จัดกลุ่มคำคุณลักษณะ 3. ปรับทิศทางความคิดเห็นของคำคุณลักษณะ ในขั้นตอนที่ 1 ประกอบไปด้วยการสกัดคุณลักษณะจากความคิดเห็น คือ สกัดเอาคำนาม (Noun) นามวลี (Noun Phrase) คำกริยา (Verb) คำกริยาวลี (Verb Phrase) ส่วนคำระบุข้อความความคิดเห็นได้สกัดเอาเฉพาะคำคุณศัพท์ (Adjective) ซึ่งจากศึกษาการเขียนประโยคของภาษาจีนเบื้องต้นระบุว่า ประโยคส่วนใหญ่จะเขียนเอาคำระบุข้อความความคิดเห็นไว้ต่อจากคำคุณลักษณะ ซึ่งหมายความว่าคำระบุข้อความความคิดเห็นที่อยู่ด้านขวาจะเป็นของคำคุณลักษณะนั้น ในขั้นตอนนี้จะทำการลบคำลบกวน Noise ออกจากประโยคก่อน ซึ่งระบุว่าคำที่มีความถี่ต่ำถือว่าเป็น Noise ซึ่งปัญหาของการสกัดคำคุณลักษณะของงานวิจัยนี้ระบุว่า มีบางประโยคที่เขียนเป็นคำคุณลักษณะโดยนัย เช่น "ถูก" "แพง" เป็นต้น ซึ่งคำเหล่านี้เป็นมีความถึงคำคุณลักษณะ คือ ราคา ดังนั้นจึงทำการระบุคำคุณลักษณะที่เป็นคุณลักษณะโดยนัยจะมาจากคำคุณศัพท์ (Adjective) เนื่องจาก คำระบุข้อความความคิดเห็นบางคำเป็นคำที่เป็น Synonym หรือ Antonym กัน ดังนั้นจึงใช้การจัดกลุ่มคำเหล่านี้ด้วยพจนานุกรม POSTagging ส่วนคำที่เป็นคำคุณลักษณะโดยนัยที่ตรงกับคำกลุ่มคำระบุข้อความความคิดเห็นจะถูกนำไปถ่วงน้ำหนักคำเพื่อนำมาเป็นคำคุณลักษณะ ในขั้นตอนที่ 2 การจัดกลุ่มคำคุณลักษณะ เนื่องจากการเขียนถึงคำคุณลักษณะของผู้เขียน ความคิดเห็นมีวิธีการเขียนที่แตกต่างกัน เช่น "Price" "Value" เป็นต้น ในงานวิจัยนี้ได้จัดกลุ่มคำที่มีความหมายเหมือนคำคุณลักษณะด้วยวิธีการ จัดกลุ่มการเพิ่มพูน (Clustering enhancement) ข้อมูลที่ใช้ในการทดลองประกอบไปด้วย ความคิดเห็นของ คอมพิวเตอร์ที่ 1 มี 500 ความคิดเห็น 1459 ประโยค คอมพิวเตอร์ที่ 2 มี 1000 ความคิดเห็น 2798 ประโยค โทรศัพท์ มี 1000 ความคิดเห็น 3067 ประโยค กล้องมี 1000 ความคิดเห็น 2674 ประโยค จากนั้นได้ทำการ

ทดลองเพื่อวัดประสิทธิภาพการสกัดคำคุณลักษณะ ซึ่งเปรียบเทียบกับวิธีการสกัดคุณลักษณะแบบกระจายเป็นคู่ (Double Propagation) ซึ่งผลลัพธ์ที่ได้ดังตารางที่ 6

ตารางที่ 6 การวัดประสิทธิภาพการสกัดคำคุณลักษณะ ค่าความแม่นยำและค่าความระลึก

Data sets		Computer 1	Computer 2	Phone	Camera
Precision	Ours	57%	69%	71%	64%
	Double Propagation	59%	65%	71%	65%
Recall	Ours	52%	60%	62%	58%
	Double Propagation	55%	58%	62%	57%

การวัดประสิทธิภาพการสกัดคำคุณลักษณะโดยนัยมีผลลัพธ์ดังนี้ ค่าความแม่นยำและค่าความระลึก Computer ที่ 1 65% และ 56% Computer ที่ 2 72% 67% Phone 79% 0.70% Camera 74% 0.65% ตามลำดับ

เปรียบเทียบประสิทธิภาพการจัดกลุ่มคำด้วยเทคนิควิธี K-Mean และ Enhance ซึ่งได้ผลลัพธ์ดังตารางที่ 7

ตารางที่ 7 เปรียบเทียบประสิทธิภาพการจัดกลุ่มคำด้วยเทคนิควิธี K-Mean และ Enhance

Data sets		Computer1	Computer2	Phone	Camera
K		23	33	25	35
Precision	K-Means	54%	64%	63%	72%
	Enhance	65%	70%	67%	79%
Recall	K-Means	43%	52%	58%	55%
	Enhance	53%	62%	65%	64%

Agarwal และคณะ [31] ได้นำเสนอรูปแบบการวิเคราะห์ความรู้สึกบนพื้นฐานความรู้สามัญสำนึก โดยสร้างพจนานุกรมข้อบริบท (contextual sentiment lexicon) ด้วยออนโทโลยีโดยใช้ชุดข้อมูลสำหรับการวิเคราะห์ 3 ชุดข้อมูล คือ Restaurant review dataset จำนวน 4,488 รีวิว,

Movie review dataset จำนวน 2,000 รีวิว และ Software review dataset คิดเห็นเชิงบวก 1,000 รีวิว และคิดเห็นเชิงลบ 915 รีวิว) ใช้พจนานุกรม ConceptNet เพื่อสร้างออนโทโลยีโดเมน สำหรับความคิดเห็นของสินค้า ใช้พจนานุกรม Wordnet เพื่อกำหนดขอบเขตคุณลักษณะของสินค้า ความหมายของประโยคที่สอดคล้องกันจะถูกกำหนดด้วยพจนานุกรมข้อบ่งชี้ ผลการทดลอง พบว่า ขั้นตอนวิธีวิเคราะห์ความรู้สึกบนพื้นฐานความรู้สามัญสำนึกด้วยใช้สร้างพจนานุกรมข้อบ่งชี้ของชุดข้อมูลทั้ง 3 พบว่า ความถูกต้อง (Accuracy) ของ Software dataset เป็น 80.1%, Movie dataset เป็น 78.9% และ Restaurant dataset เป็น 79.4%

khan และคณะ [32] นำเสนอวิธีวัดความรู้สึกเชิงบวก เชิงลบ และเป็นกลางบนทวิตเตอร์ จำแนก 3 ขั้นตอน 1. จำแนกอารมณ์จากสัญลักษณ์แสดงความรู้สึก 2. จำแนกจากข้อความที่แสดงความรู้สึกโดยเปรียบเทียบจากพจนานุกรม 3. จำแนกข้อความที่แสดงความรู้สึกด้วยการจำแนก 3 ขั้นตอน โดยรวบรวมข้อมูลจาก Twitter ด้วย Twitter Streaming API บนเวลาจริง โดยข้อมูลที่รวบรวมนั้นเป็นข้อความภาษาอังกฤษ โดยใช้ชุดข้อมูล 6 ชุด จากการสุ่ม 2,115 ทวิต หาความหมายของศัพท์โดยใช้ดิกชันนารี 3 อย่างคือ (WordNet, Spell Check, jSpell) และทำการแทนที่ด้วย Netlgo และ msm ดิกชันนารี ตรวจสอบคำสะกดและแทนที่รากศัพท์ด้วยคำที่สมบูรณ์ด้วย Jazzy Checker, jSpell และ Snowball ตัดคำหยุดด้วย Stanford, Wiki และ Textifier ตัด URL, UserName, Hashtag และอักขระพิเศษออกยกเว้นสัญลักษณ์แสดงอารมณ์ การจำแนกข้อความคิดเห็นแบ่งออกเป็น 3 ขั้นตอน คือ จำแนกความคิดเห็นเชิงบวก/ลบ ด้วยสัญลักษณ์แสดงอารมณ์ จำแนกข้อความรู้สึกด้วย Bag of Word จำแนกความรู้สึกด้วยด้วยการคำนวณน้ำหนักความรู้สึกจากคำศัพท์ความคิดเห็น (SentiWordNet) ว่าเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง ผลการวิจัยพบว่าความถูกต้อง (Accuracy) ของการจำแนกเป็น 85.7% ความสามารถในการที่จะจัดข้อความที่ไม่เกี่ยวข้องออกไป (Precision) 85.3%, ความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา 82.2%

Mostafa [33] นำเสนอการทำเหมืองข้อความสำหรับวิเคราะห์ความรู้สึกของต่อสินค้าที่หือดัง ได้แก่ Nokia, T-Mobile, IBM, KLM และ DHL โดยรวบรวมข้อความในเว็บไซต์ Tweeter มา ประเมินความรู้สึกของผู้บริโภค ว่ามีความคิดเห็นต่อสินค้าแต่ละยี่ห้อในเชิงบวก หรือในเชิงลบ โดยใช้พจนานุกรมคำแสดงความคิดเห็น จำนวน 6,800 คำ ซึ่งมีทั้งความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ กำหนดให้ความคิดเห็นเชิงบวกมีน้ำหนัก +1 และ กำหนดให้ความคิดเห็นเชิงลบมีน้ำหนัก -1 แล้วประเมินความคิดเห็นจากคำที่พบในประโยค และสรุปผล แสดงในรูปแบบกราฟ โดยใช้ StreamGraph Software เพื่อแสดงให้เห็นความรู้สึกในภาพรวมของผู้บริโภคต่อสินค้ายี่ห้อเหล่านั้น

Liu และคณะ [34] ได้นำเสนอการวิเคราะห์ความรู้สึกระดับประโยคที่มีกริยาด้วยเทคนิค Support vector machine โดยใช้ข้อมูลในการวิเคราะห์เป็นการแสดงความคิดเห็นของลูกค้านั้นเป็นประโยคที่มีกริยา ซึ่งมีวิธีการทดลองโดยการสร้าง Vector ของคุณลักษณะเป็น 2 กลุ่ม

1. กลุ่มที่มีคุณสมบัติของภาษาทั่วไป และ 2. กลุ่มที่มีคุณสมบัติที่จะเกี่ยวข้องโดยตรงกับการวิเคราะห์กิริยา ผลการทดลองพบว่าความถูกต้องเป็น 50%

Jose และ Losada [35] นำเสนอวิธีการใหม่ในการจำแนกความคิดเห็นในระดับประโยค เพื่อนำไปใช้ในการวิเคราะห์ความคิดเห็นให้เกิดประโยชน์อย่างแท้จริง โดยใช้ข้อมูลที่มีความหลากหลาย แบ่งเป็น 2 ประเด็นหลัก คือ จำแนกข้อความความคิดเห็น (Subjective) กับข้อเท็จจริง (Objective) และจำแนกขั้วความคิดเห็น (Polarity Classification) แบ่งเป็นความคิดเห็นเชิงบวก (Positive) และความคิดเห็นเชิงลบ (Negative) การทดลองผู้วิจัยใช้คุณลักษณะที่หลากหลาย ได้แก่ คำศัพท์ (Vocabulary Feature) ตำแหน่งของคำ (Positional Feature) หน้าที่ของคำ (Part of Speech Feature) คุณลักษณะประโยค (Syntactic Patterns) คลังคำความคิดเห็น (Sentiment Lexicon Feature) เป็นต้น ข้อมูลที่ใช้ในการทดสอบ ผู้วิจัยใช้ข้อมูลจาก 4 แหล่ง ได้แก่ 1) Multilingual Opinion Analysis Test Collection (MOAT) ซึ่งเป็นข้อมูลจากแหล่งข่าวจำนวน 14 หัวข้อ ที่ความแตกต่างกัน ประกอบด้วยความคิดเห็นเชิงบวกจำนวน 179 ความคิดเห็น และความคิดเห็นเชิงลบ จำนวน 417 ความคิดเห็น 2) Finegrained Sentiment Dataset (FSD) เป็นข้อมูลวิจารณ์ผลิตภัณฑ์ที่รวบรวมจากหลากหลายโดเมน ประกอบด้วยความคิดเห็นเชิงบวก 923 ความคิดเห็น และความคิดเห็นเชิงลบ 1,320 ความคิดเห็น 3) Multi-Perspective Question Answering Dataset (MPQA) เป็นข้อความความคิดเห็นที่ประกอบข้อความความคิดเห็นส่วนบุคคล เกี่ยวกับความเชื่อต่าง ๆ ประกอบด้วย ความคิดเห็นเชิงบวก 1,626 ความคิดเห็น และความคิดเห็นเชิงลบ 3,255 ความคิดเห็น และ 4) Pang and Lee Subjectivity Dataset (PL) เป็นคลังความคิดเห็นเชิงประโยคที่รวบรวมจากข้อมูลวิจารณ์หนังสือ ใช้วิธีการซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) ในการจำแนกความคิดเห็น แบ่งข้อมูลสำหรับการจำแนกด้วยวิธีการสุ่ม (Split Test) โดยแบ่งเป็นข้อมูลชุดสอน 75% และข้อมูลชุดทดสอบ 25% จากการทดสอบพบว่า การใช้ Unigram/Bigram ร่วมกับคุณลักษณะที่เป็นคลังความคิดเห็น ให้ประสิทธิภาพในการจำแนกดีที่สุด

Natchiar และ Baulkani [36] นำเสนอวิธีการเลือกคุณลักษณะใหม่เพื่อแก้ไขปัญหาคำซ้อนซ้ำซ้อน ข้อมูลความผิดปกติ ข้อมูลซ้ำซ้อนและไม่เกี่ยวข้อง รูปแบบที่แตกต่างกันในข้อมูล CRM ที่กำหนดด้วยคุณลักษณะที่เกี่ยวข้องโดยการใช้มาตรการเหมือนข้อมูลที่มียุทธศาสตร์เทคนิคในการปรับปรุงข้อมูลที่มีคุณภาพและมีความเกี่ยวข้องหลังจากก่อนการประมวลผล วิธีการที่นำเสนอมีการทดสอบโดยใช้ชุดข้อมูล KDD 2009 จำนวน 50,000 ชุดข้อมูล 230 คุณลักษณะ ข้อมูลเชิงบวกจำนวน 49,110 ชุด ข้อมูลเชิงลบจำนวน 890 ชุด และมีค่าสูญหาย 90 % ซึ่งมีขั้นตอนกระบวนการเลือกแอตทริบิว คือ 1. หาค่าแต่ละแอตทริบิวโดยการเพิ่มของข้อมูล (Gain) 2. ใช้ Ranker เลือกคุณลักษณะที่มีค่าเพิ่มของข้อมูลมากที่สุด 3. ใช้ข้อมูลชุดสอนทั้งหมดสำหรับเลือกแอตทริบิว นำข้อมูลที่ไม่วุ่นวายและมีข้อมูลรบกวนเยอะ (Noisy) ทำการประเมินความถูกต้องของการจำแนกความคิดเห็นโดยใช้วิธีการทดสอบ

แบบ 10-Fold Cross – Validation และทำการเปรียบเทียบประสิทธิภาพของการจำแนกความคิดเห็นตามทัศนคติโดยใช้ Learning Method 4 ชนิด ได้แก่ Naïve Bayes, Decision Tree (J48), SVM และ KNN แล้ว เมื่อทำการเปรียบเทียบผลการจำแนกแต่ละวิธีพบว่า Naïve Bayes ให้ความถูกต้องในการจำแนกสูงที่สุด

Kongthon และคณะ [7] นำเสนอวิธีการในการวิเคราะห์ความคิดเห็นเกี่ยวกับโรงแรมของลูกค้า เพื่อนำมาตรวจสอบปัจจัยที่มีผลต่อการเลือกโรงแรมของลูกค้า และเพื่อให้ธุรกิจเข้าใจความคิดเห็นของลูกค้าที่มีต่อโรงแรม งานวิจัยนี้ใช้ข้อมูลภาษาไทยจากเว็บไซต์โกด้า (Agoda) โดยรวบรวมความคิดเห็นที่มีต่อโรงแรมจำนวน 275 แห่งในประเทศไทย และข้อความความคิดเห็นจากลูกค้าทั้งหมด 8,500 ความคิดเห็น วิธีการทดลอง ประกอบด้วย 2 โมดูล คือ การสร้างทรัพยากรภาษา (Language Resource Construction Module) และ การทำเหมืองความคิดเห็น (Opinion Mining) สำหรับการสร้างทรัพยากรภาษา มีขั้นตอนคือ 1) รวบรวมข้อมูลจากเว็บไซต์ จากนั้นนำความคิดเห็นมาตัดแท็กความคิดเห็นด้วยตนเอง แบ่งเป็นความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ แล้วนำไปเก็บไว้ในคลังข้อมูลความคิดเห็น (Opinion Corpus) คลังความคิดเห็น ประกอบด้วยชุดข้อมูล 3 ชุดได้แก่ วลีความคิดเห็น (Sentiment Phrases) คุณลักษณะที่เป็นกลุ่มคำที่มีความหมายใกล้เคียงกัน (Feature Synsets) และ กลุ่มข้อความความคิดเห็นของคำ (Polar Word Data Set) จากนั้นทำการสร้างรูปแบบการวิเคราะห์ความคิดเห็น และนำข้อมูลไปสรุปในรูปแบบของสถิติเพื่อเจ้าของธุรกิจสามารถนำข้อมูลไปใช้ในการปรับปรุงความสัมพันธ์กับลูกค้าและพัฒนาความพึงพอใจให้ลูกค้าได้อย่างสะดวกและรวดเร็ว

Ennaji และคณะ [37] นำเสนอกรอบแนวคิดอัจฉริยะ (Social Intelligence Framework) ที่สามารถแยกแยะและรวบรวมความคิดเห็นที่แสดงออกทางสื่อสังคมออนไลน์ เพื่อช่วยให้ผู้ประกอบการทราบข้อมูลเพิ่มเติมเกี่ยวกับความคิดเห็นของลูกค้าที่มีต่อผลิตภัณฑ์ต่างๆ ได้ ข้อมูลที่ใช้รวบรวมจากเว็บไซต์ทวิตเตอร์ โดยใช้ Twitter4J สำหรับกรอบแนวคิดนี้ เป็นการสกัดและวิเคราะห์ความคิดเห็นของประชาชนเกี่ยวกับผลิตภัณฑ์จากเว็บไซต์เครือข่ายสังคมออนไลน์ ประกอบด้วย 1) โมดูลการสกัดข้อมูล เป็นการสกัดข้อมูลจากเว็บไซต์สื่อสังคมออนไลน์ 2) โมดูลการวิเคราะห์ข้อมูล เป็นการวิเคราะห์ข้อมูลที่สกัดจากเว็บไซต์สื่อสังคม 3) ข้อมูลสื่อสังคมออนไลน์ เป็นพื้นที่เก็บข้อมูล ซึ่งพื้นที่เก็บข้อมูลนี้เป็นแหล่งที่มีคุณค่าในการไหลคลังข้อมูลและระบบ CRM 4) โมดูลการวิเคราะห์ เป็นโมดูลที่จะบอกผลการวิเคราะห์ข้อมูล กรอบแนวคิดของงานวิจัย คือ ทำการพัฒนากลไกอัจฉริยะที่สามารถระบุความคิดเห็นของผู้คนในเครือข่ายสังคมออนไลน์ได้ เพื่อให้บริษัทได้ทราบข้อมูลความต้องการของผู้ใช้และทราบข้อมูลการจัดอันดับผลิตภัณฑ์เปรียบเทียบกับคู่แข่ง โดยทำการเก็บรวบรวมจุดแข็งและจุดอ่อนของผลิตภัณฑ์ต่างๆ ขั้นตอนการสกัดความคิดเห็น เริ่มจากการรวบรวมข้อมูล จากนั้นกำจัดข้อความที่เป็นสแปม (Spam Detection) ได้แก่ ข้อความที่

ประกอบด้วย URLs, ข้อความที่เป็นข้อความการตอบคำถามหรือกล่าวถึง (Replies/Mentions) เช่น ข้อความที่ขึ้นต้นด้วย @ ตามด้วย USERNAMES จากนั้นทำการระบุข้อความคิดเห็น โดยแบ่งเป็น ความคิดเห็นเชิงบวก ความคิดเห็นเชิงลบ และความคิดเห็นที่เป็นกลาง แล้วสรุปเป็นรูปแบบกราฟ สำหรับแต่ละผลิตภัณฑ์ จากการศึกษาพบว่า งานวิจัยนี้เป็นการสรุปความคิดเห็นโดยรวม ซึ่งการนำไปใช้อาจจะไม่ตรงตามกลุ่มลูกค้า ควรจะมีการนำข้อมูลโปรไฟล์ของลูกค้ามาเพิ่มในการวิเคราะห์ และสรุปเป็นความคิดเห็นตามกลุ่มอายุ เพศ และอาชีพ เป็นต้น

Yaakub และคณะ [6] นำเสนอวิธีการใหม่ในการทำเหมืองความคิดเห็น โดยใช้รูปแบบหลายมิติ เพื่อรวบรวมลักษณะของลูกค้ากับความคิดเห็นเกี่ยวกับผลิตภัณฑ์และนำมาสร้างเป็นรูปแบบธุรกิจ (Business Model) โดยแปลงความคิดเห็นเก็บไว้ในรูปแบบตารางความจริง (Fact Table) เช่น ลูกค้า (Customer) ผลิตภัณฑ์ (Product) เวลา (Time) และความคิดเห็น (Opinion) มีขั้นตอนการดำเนินการ คือ นำความคิดเห็นมาสกัดหาคุณลักษณะตามที่ได้ตั้งไว้ในตารางความจริง ในการแปลงข้อความคิดเห็นลงในตาราง ผู้วิจัยตั้งค่าคะแนนไว้ 7 ระดับ ระหว่าง -3 ถึง 3 คะแนน ได้แก่ ระดับ 3 คะแนน หมายถึง ดีเยี่ยม (Excellent) ระดับ 2 คะแนน หมายถึง เห็นความแตกต่าง (Distinguish) ระดับ 1 คะแนน หมายถึง ยอมรับ (Accept) ระดับ 0 คะแนน หมายถึง เป็นกลาง (Neutral) ระดับ -1 คะแนน หมายถึง ปฏิเสธเล็กน้อย ระดับ -2 คะแนน หมายถึง ไม่ยอมรับ ระดับ -3 คะแนน หมายถึง แย่ ในส่วนของการสกัดคุณลักษณะแปลงลงในตารางผลิตภัณฑ์ ซึ่งมีรูปแบบของแอตทริบิวต์แล้ว จะต้องตรวจสอบว่าในความคิดเห็นมีแอตทริบิวต์นั้นหรือไม่ การตรวจสอบแบ่งเป็น 3 กรณี คือ ตรงกัน ตรงกันบางส่วน และไม่ตรงกันเลย ในการรวมข้อมูลไปไว้ในรูปแบบคลังข้อมูลความคิดเห็น หลังจากสร้างคลังข้อมูลความคิดเห็นไว้แล้ว ทำการวิเคราะห์ความคิดเห็น แล้วสร้างกลุ่มข้อมูลในรูปแบบ 3 มิติ ซึ่งประกอบด้วย ข้อมูลประเภทผลิตภัณฑ์ กลุ่มอายุของลูกค้า และข้อความคิดเห็นที่มีต่อผลิตภัณฑ์ ประโยชน์ของงานวิจัยนี้ คือ สามารถอธิบายความคิดเห็นของลูกค้าแต่ละกลุ่มที่มีต่อคุณลักษณะของผลิตภัณฑ์ต่างได้อย่างสะดวกและรวดเร็วขึ้น

Eirinaki [38] งานวิจัยนี้นำเสนอระบบในการทำเหมืองความคิดเห็นและการจัดลำดับคะแนนความคิดเห็นของแสดงความคิดเห็น ประกอบไปด้วย 1. ขั้นตอนการรวบรวมข้อมูล 2. การเตรียมข้อมูล 3. การทำเหมืองความคิดเห็น 4. การจัดลำดับคะแนนความคิดเห็น ในขั้นตอนการรวบรวมความคิดเห็นได้ใช้ชุดข้อมูลจากงานวิจัยของ Bing Lui ประกอบไปด้วย Camera 80 ความคิดเห็น DVD Player 390 ความคิดเห็น Vacuum 40 ความคิดเห็น ในขั้นตอนการเตรียมข้อมูลได้นำความคิดเห็นต่างๆ มาตัดออกเป็นคำ ในขั้นตอนการทำเหมืองความคิดเห็น ได้ทำการกำกับหน้าที่ของคำ POStagging และทำการคำนวณหาระยะห่างระหว่างคำคุณลักษณะจากชนิดคำนาม (Noun) และคำระบุข้อความคิดเห็น จากชนิดคำ คุณศัพท์ (Adjective) โดยระบุว่า คำคุณศัพท์คำใด ที่อยู่ใกล้คำคุณลักษณะที่สุด จะถือว่าคำระบุข้อความคิดเห็นนั้นเป็นของคำคุณลักษณะดังกล่าว ในขั้นตอน

การจัดลำดับคะแนนความคิดเห็นจะให้คะแนนจากคุณลักษณะที่ได้ทำการสกัดคุณลักษณะ รวมถึงการสรุปคะแนนในประโยคความคิดเห็นว่าประโยคความคิดเห็นดังกล่าวควรจะเป็นขั้วบวก ขั้วลบ หรือเป็นกลาง ในขั้นตอนนี้ประกอบไปด้วย 2 อัลกอริทึม ได้แก่ 1. The High Adjective Count algorithm จะกำหนดคะแนนคำนาม เป็นค่า 0 ถ้าคำนามใดมีคำ Adjective อยู่ใกล้ จะถูกเพิ่มค่าขึ้น 1 ค่า ซึ่งถ้าค่าของคำนามที่ถูกเพิ่มขึ้นจะถูกนำค่า threshold มากำหนด สำหรับหาค่าความถี่ของคำนามที่จะนำมาเป็นคุณลักษณะในอนาคต 2. The max Opinion Score algorithm ในขั้นตอนนี้จะกำหนดค่าคำระบุขั้วความคิดเห็น (Adjective) โดยคำระบุขั้วความคิดเห็นที่เป็นเชิงบวกจะกำหนดเป็นค่า 4 และคำระบุขั้วความคิดเห็นที่เป็นเชิงลบจะกำหนดเป็น -4 ส่วนคำระบุขั้วความคิดเห็นมีค่าปฏิเสธอยู่ด้านหน้า จะถูกนำไปคูณด้วย -1 เพื่อเปลี่ยนความหมายของคำให้เป็นตรงกันข้ามและทำการเพิ่มคุณลักษณะจากการหาค่าความถี่และความถี่ผกผัน ของคุณลักษณะจาก The High Adjective Count algorithm จากนั้นได้คำนวณว่าคุณลักษณะดังกล่าวจะเป็นเชิงบวกหรือเชิงลบ โดยนำคะแนนจากค่าของคำระบุขั้วความคิดเห็นมารวมกัน ถ้าต้องการทราบว่าความคิดเห็นดังกล่าวเป็นขั้วบวกหรือขั้วลบจะนำค่าของแต่ละคุณลักษณะมาเฉลี่ยกันระหว่างเชิงบวกและเชิงลบ จากการทดลองพบว่า ทุกข้อมูลชุดทดสอบมีความถูกต้องในการจำแนกมากกว่า 87%

Tribhuvan [39] นำเสนอวิจัยขั้นตอนของการทำเหมืองความคิดเห็นแบบ ระดับคุณลักษณะ (Feature Based) ของเหมืองความคิดเห็น และการสรุปความคิดเห็น โดยกล่าวว่า การทำเหมืองความคิดเห็นประกอบด้วย 3 ระดับคือ ระดับเอกสาร ระดับประโยค ระดับคุณลักษณะ 1. ระดับเอกสาร คือ การจำแนกความคิดเห็นทั้งเอกสารเป็นเชิงบวกหรือเชิงลบจากคำสำคัญต่างๆ ที่ปรากฏในเอกสาร 2. ระดับประโยค คือ การตัดข้อความความคิดเห็นในเอกสารออกเป็นประโยคและจำแนกความคิดเห็นแต่ละประโยคเป็นเชิงบวกหรือเชิงลบ ส่วนการทำเหมืองความคิดเห็นในระดับคุณลักษณะ คือ การสรุปแต่คุณสมบัติของสินค้าเป็นเชิงบวกหรือเชิงลบ โดยงานวิจัยนี้มุ่งเน้นไปที่การทำเหมืองความคิดเห็นและสรุปขั้นตอนในการทำ feature based เป็น 7 ขั้นตอนดังนี้ ขั้นตอนที่ 1 การรวบรวมข้อมูลความคิดเห็นและขั้นตอนการเตรียมข้อมูลต่างๆ เช่น การแก้ไขคำผิด การตัดคำ เป็นต้น โดยกล่าวว่า การรวบรวมข้อมูลความคิดเห็นจะต้องเกิดขึ้นจากความไม่ลำเอียง ดังนั้นจะต้องมีขั้นตอนที่เป็นมาตรฐานในการรวบรวมข้อมูลด้วย โดยนำเสนอ โปรแกรมพื้นฐานในการเก็บข้อมูล ได้แก่ web crawlers, spiders, robots ขั้นตอนที่ 2 การกำหนดที่ของคำ (Path-of-speech Tagging หรือ POS Tagging) คือการนำข้อความที่ผ่านการตัดคำมาหาชนิดของคำ เช่น คำนาม คำคุณลักษณะ คำขยายนาม เป็นต้น โดยกล่าวว่า คำที่เป็นคุณลักษณะส่วนใหญ่คือคำนาม และนำเสนอตัวอย่างดังนี้ “This is good phone” ผลลัพธ์ที่ได้คือ This/Phone is/V good/ADJ Phone/N เป็นต้น ขั้นตอนที่ 3 การสกัดคุณลักษณะ คือ การสกัดเอาเฉพาะคำนามของทุกประโยคความคิดเห็นมาใช้เป็นคุณลักษณะ ขั้นตอนที่ 4 การสกัดคำระบุขั้วความคิดเห็น โดยสกัดเอาเฉพาะคำขยาย

ประธาน เนื่องจากคำแสดงความคิดเห็นส่วนใหญ่จะเป็นคำขยายประธาน ขั้นตอนที่ 5 การระบุคำแสดงความคิดเห็น คือ การระบุคำที่สกัดได้เป็นคำแสดงความคิดเห็นในเชิงบวกหรือเชิงลบ จะใช้การหาความหมายของคำในการระบุ เช่น ความหมายตามพจนานุกรม เป็นต้น ขั้นตอนที่ 6 เป็นการระบุประโยคความคิดเห็น คือ การระบุประโยคความคิดเห็นนั้นมีความคิดเห็นเป็นอย่างไรในแต่ละคุณลักษณะ เช่น แบตเตอรี่ทน แต่กล้องถ่ายรูปแย่มาก เป็นต้น ซึ่งประโยคนี้มีความคิดเห็นเป็นเชิงบวกต่อคุณลักษณะแบตเตอรี่ และเชิงลบต่อกล้อง ในขั้นตอนนี้ ยังรวมถึงการสร้างกฎต่างๆ ขึ้นมาใช้ในการระบุความคิดเห็นด้วย เช่น ถ้าเจอคำนี้แสดงอยู่หน้าคำใด คำแสดงความคิดเห็นก็จะมีมีความหมายในทางตรงกันข้าม เป็นต้น ขั้นตอนที่ 6 การสรุปความคิดเห็นทั้งหมด คือ การสรุปว่าในแต่ละคุณลักษณะมีความคิดเห็นเป็นเชิงบวกหรือเชิงลบจำนวนเท่าไร ด้านการวัดประสิทธิภาพงานประเภท feature based จะทำการวัดค่า Recall และ Precision จากคำที่สกัดได้ว่ามีความถูกต้องหรือไม่

Milović [40] การทำ eCRM (Electronic Customer Relationship Management) เป็นเทคโนโลยีใหม่ที่ทางบริษัทได้ทำการปรับปรุงการตลาดรูปแบบใหม่ และความสามารถทางธุรกิจ ซึ่งจะช่วยให้เข้าใจลูกค้า กิจกรรม ความเป็นส่วนตัว ปัจจัยที่เกี่ยวข้องกับลูกค้า รวมถึงการพัฒนาธุรกิจให้เหนือกว่าคู่แข่งผ่านทางอุปกรณ์อิเล็กทรอนิกส์ ซึ่งเป็นกลยุทธ์ใหม่ในการติดต่อสื่อสารกับลูกค้า ผู้ที่เกี่ยวข้องอื่นๆ เช่น แหล่งวัตถุดิบ ผู้สนับสนุน เป็นต้น ซึ่งหลักๆ แล้ว eCRM จะเน้นไปที่การสื่อสารหรือประชาสัมพันธ์กับลูกค้าเพื่อผลประโยชน์ที่ลูกค้าจะได้รับเอง โดยเฉพาะในเว็บไซต์โฆษณาโรงแรม จำเป็นต้องมีการปรับปรุงแก้ไขข้อมูล รูปภาพ และให้ความสะดวกกับการจองโรงแรมของผู้ที่จะเข้าพักอย่างง่ายดาย และมีประสิทธิภาพ และยังรวมไปถึงภาษาที่ใช้ในการให้ข้อมูลโรงแรมควรมีหลากหลายภาษา ที่ใช้ในการให้ข้อมูลโรงแรมควรมีหลายภาษา เพื่ออำนวยความสะดวกด้านการสื่อสารระหว่างโรงแรมและผู้เข้าพัก การนำเสนอห้องพักที่สามารถเจาะจงตรงกับความต้องการของผู้เข้าพักได้ ก็จะทำให้ลูกค้ามีความสนใจและยอมรับในการบริการ การค้นหาโรงแรมโดยดูตัวช่วยในการค้นหา แล้วสามารถแสดงขึ้นในลำดับแรกๆ รวมถึงความรวดเร็วในการแสดงหน้าเว็บเหล่านี้ล้วนเป็นกลยุทธ์ในการดึงดูดลูกค้า รวมถึงดึงดูดผู้ร่วมลงทุนด้วย เนื่องจากผู้ที่ต้องการร่วมลงทุนได้มองเห็นโอกาสในการสร้างผลกำไรจากปัจจัยต่างๆ จากการเจริญเติบโตของยอดจองผู้เข้าพักออนไลน์ จะสร้างบอร์ด หรือฟอร์ม สำหรับถาม ตอบ หรือ เพื่อรับฟังความคิดเห็นของลูกค้าในเว็บไซต์ก็เป็นสิ่งสำคัญที่จะช่วยให้การพัฒนาในด้านต่างๆ ตรงตามความต้องการของลูกค้าและ 2 ใน 3 ของบริษัทมีแผนในการพัฒนาการทำการตลาดผ่านทางสื่อสังคมออนไลน์ เนื่องจาก 40% ของนักท่องเที่ยวมีการเข้าใช้เว็บไซต์เครือข่ายสังคมออนไลน์ ซึ่งอิทธิพลต่อการเข้าพักโรงแรม และจากการสำรวจนักท่องเที่ยว 1000 คน ของอังกฤษ ของอังกฤษ พบว่า มีการจองผ่านโซเชียลมีเดีย 42% เลือกที่พักที่แตกต่าง 35% เปลี่ยนรีสอร์ทผ่านสังคมออนไลน์ 15% ดังนั้นสรุปได้ว่า การนำ eCRM เข้ามาใช้ในการธุรกิจจะ

สามารถช่วยให้ผู้ที่เข้าพักระยะยาวและได้รับประโยชน์มากขึ้น รวมถึงเจ้าของอุตสาหกรรมโรงแรมในระยะสั้น และระยะยาวด้วย

Kasper และคณะ [41] งานวิจัยนี้นำเสนอว่า ข้อความความคิดเห็นของผู้ใช้หรือผู้เข้าพักไม่ควรมีการระบุมาเป็นตัวเลขระดับของความคิดเห็น เนื่องจากมันขัดแย้งกับความเป็นจริงจากข้อความความคิดเห็นที่ผู้แสดงความคิดเห็นได้แสดงความคิดเห็นผ่านข้อความ และงานวิจัยนี้ได้นำเสนอวิธีการรวบรวมความคิดเห็นต่างๆ บนเว็บไซต์วิเคราะห์ และจำแนกที่เป็นลักษณะข้อความ และนำเสนอผลลัพธ์ในรูปแบบที่กระชับ ซึ่งเรียกว่า BSAHOT โดยมีขั้นตอนต่างๆ เริ่มจากรวบรวมข้อความความคิดเห็นเกี่ยวกับโรงแรมทั้งหมดจากนั้น กรองเอาข้อความความคิดเห็นเฉพาะภาษาเยอรมัน และทำการแบ่งข้อความความคิดเห็นออกเป็นประโยค และทำการจำแนกคำระบุข้อความทางสถิติ และทำการสกัดองค์ประกอบข้อมูลทางภาษาเพื่อวิเคราะห์เนื้อหา และคำนวณข้อความความคิดเห็นอย่างละเอียด จะมีการกำหนดว่าคำระบุข้อความความคิดเห็น ในการแบ่งข้อความ โดยค่าทางสถิติและการจำแนกทางภาษาได้เข้าร่วมกับค่าที่เป็นคำระบุข้อความความคิดเห็นที่ใช้สากล จากนั้นทำการวิเคราะห์ผลลัพธ์และเก็บรวบรวมข้อความความคิดเห็นที่เป็นประโยคข้อความ โดยแยกไว้เป็นฐานข้อมูลผลลัพธ์เฉพาะเพื่อนำเสนอให้กับผู้ใช้ ซึ่ง BSAHOT มีเป้าหมายที่จะให้ผู้เข้าพักได้เห็นมุมมองที่เป็นจริงในรูปแบบผลสรุปจากข้อความความคิดเห็นบนเว็บ ซึ่งเป็นภาษาเยอรมัน โดยมีขั้นตอนต่างๆ ดังนี้ 1.การรวบรวมข้อความความคิดเห็นโดยดึงจากแท็กใน และทำการกรองรหัสทางภาษาโดยเลือกเอาเฉพาะภาษาเยอรมัน เก็บไว้ในฐานข้อมูล 2.ตัดข้อความความคิดเห็นออกเป็นประโยค และแบ่งตามหัวข้อ 3. วิเคราะห์เพิ่มเติมว่าเป็นหัวข้อหรือลบด้วยการจำแนกสถิติของคำระบุข้อความความคิดเห็นและการสกัดองค์ประกอบข้อมูลทางภาษาเพื่อให้การวิเคราะห์เนื้อหาละเอียดยิ่งขึ้น ซึ่งค่าของคำระบุข้อความความคิดเห็นจะถูกกำหนดไว้ก่อน จากนั้นนำประโยคข้อความความคิดเห็นที่ถูกแบ่ง มาตัวเป็นคำกำกับหน้าที่ของคำ เช็คคำถูกคำผิดเพื่อแก้ไขคำ หาคำที่มีความหมายใกล้เคียงกัน สกัดเอาเฉพาะคำสำคัญในประโยค จากนั้นจะเก็บผลลัพธ์ไว้เป็นฐานข้อมูลและสามารถดึงผลลัพธ์ขึ้นใช้ได้ตามต้องการ ในการทดลองวิเคราะห์ระบบ จะใช้ความคิดเห็น 1,559 ความคิดเห็นซึ่งได้จากการรวบรวม แบ่งออกเป็น 4,792 ประโยค สำหรับการทดลอง ซึ่งในประโยคสำหรับการทดลอง ซึ่งในประโยคเหล่านั้นมีความคิดเห็นที่เป็นเชิงบวก 2,240 ประโยค เชิงลบ 1,183 ประโยค และเป็นกลาง 938 ประโยค และเรื่องอื่นๆ ที่ไม่เกี่ยวกับโรงแรม 431 ประโยค จำแนกด้วยวิธีทางสถิติ สามารถทำนายได้ถูกต้อง 3,145 ประโยค ทำนายผิด 705 ประโยค หรือค่าความถูกต้อง 0.82 วิธีการสกัดข้อมูล ทำนายได้ถูกต้อง 2,604 ประโยค ทำนายผิด 486 ประโยค หรือค่าความถูกต้อง 0.68 เมื่อนำทั้งสองวิธีมารวมกัน ทำนายได้ถูกต้อง 3,208 ประโยค ทำนายผิด 646 ประโยค หรือค่าความถูกต้อง 0.83

Tsai และคณะ [8] นำเสนอวิธีการสกัดคุณลักษณะเพื่อเพิ่มประสิทธิภาพการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม ขั้นตอนแรกเป็นการสร้างตัวจำแนกเพื่อระบุข้อความความคิดเห็นที่เป็น

ประโยชน์ ซึ่งตัวจำแนกที่ใช้ในงานวิจัยนี้ ได้แก่ Decision Tree, Random Forest, Logistic Regression และ Support Vector Machine (SVM) จากนั้นทำการแบ่งประโยคเพื่อวิเคราะห์ความคิดเห็นตามคุณลักษณะ ซึ่งประกอบด้วย Location, Sleep Quality, Rooms, Service, Value และ Cleanliness ขั้นตอนต่อมาเป็นการสรุปข้อความความคิดเห็นของแต่ละประโยคเพื่อนำมาสรุปภาพรวมของแต่ละความคิดเห็น โดยกระบวนการวิเคราะห์ ประกอบด้วย 4 ขั้นตอน ได้แก่ 1) การระบุข้อความความคิดเห็นของแต่ละประโยค (Identify Sentence Polarity) 2) การคำนวณค่าคะแนนของแต่ละประโยค (Calculate Sentence Score) 3) การจัดกลุ่มประโยค (Group Sentences into Clusters) และ 4) เลือกประโยคจากคลัสเตอร์ (Select Sentences from Clusters) ผลการทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอมีประสิทธิภาพที่ดี

Khalid และคณะ [42] นำเสนอวิธีการจำแนกความคิดเห็นจากข้อความความคิดเห็นที่ไม่มีโครงสร้างโดยใช้ตัวจำแนกแบบเอนเซมเบิล (Ensemble Classifier) เพื่อนำผลการจำแนกไปพัฒนากลยุทธ์ทางการตลาดในการปรับปรุงคุณภาพของการบริการและความพึงพอใจของลูกค้า เริ่มจากการประเมินประสิทธิภาพของโมเดลการเรียนรู้ของเครื่อง (Machine Learning Models) โดยใช้โมเดลที่นิยมใช้ในวรรณกรรม แล้วนำเสนอวิธีการ Gradient Boosted Support Vector Machine (GBSVM) เพื่อหาตัวจำแนกที่ดีที่สุด ผลการทดลองโดยใช้ชุดข้อมูล 2 ชุด โดยใช้วิธี Term Frequency (TF) และ Term Frequency-Inverse Document Frequency ซึ่งผลการทดลองแสดงให้เห็นว่า วิธีการที่นำเสนอมีประสิทธิภาพที่ดีเมื่อเทียบกับวิธีการอื่น

Songpan [9] ทำการวิเคราะห์และทำนายคะแนนความพึงพอใจ (Rating) จากบทวิจารณ์ของลูกค้า (Customer Reviews) ซึ่งเว็บไซต์บทวิจารณ์ทั่วไปจะเปิดให้ลูกค้าแสดงความคิดเห็นทั้งแบบข้อความและเป็นค่าคะแนนความพึงพอใจ แต่ปัญหาที่พบคือ ลูกค้าบางรายให้คะแนนตรงกันข้ามกับความคิดเห็น ดังนั้น งานวิจัยนี้จึงได้ทำการวิเคราะห์และทำนายคะแนนความพึงพอใจของลูกค้าเพื่อให้สอดคล้องกับบทวิจารณ์ของลูกค้า โดยใช้แบบจำลองการจำแนกประเภทความน่าจะเป็น (Probability's Classifier Model) ข้อมูลที่ใช้ในการวิจัย คือ บทวิจารณ์ของลูกค้าที่เข้าพักรั้วโรงแรมซึ่งประกอบด้วย ข้อความความคิดเห็นเชิงบวก (Positive) และความคิดเห็นเชิงลบ (Negative) ในการวิเคราะห์และทำนายค่าคะแนนความพึงพอใจใช้วิธีการนาอ์ฟเบย์ และวิธีการต้นไม้ตัดสินใจ พบว่า วิธีการนาอ์ฟเบย์มีประสิทธิภาพความถูกต้องในการให้ค่าคะแนนความพึงพอใจ เท่ากับ 94.37% ซึ่งสูงกว่าวิธีการต้นไม้ตัดสินใจ

Calheiros และคณะ [10] ทำการจำแนกความคิดเห็นของผู้ใช้บริการโรงแรม ด้วยวิธีการเหมืองข้อความและ วิธีการ Latent Dirichlet Allocation (LDA) โดยรวบรวมข้อมูลบทวิจารณ์โรงแรมจำนวน 400 ข้อความ เข้าสู่กระบวนการประเมินความคิดเห็นของผู้เข้าพักโดยใช้วิธีการเหมืองข้อความ (Text Mining) และการสร้างแบบจำลองหัวข้อ (Topic Modeling) ผลการทดลองแสดงให้เห็นถึงความรู้สึกละเอียด (Topic) ของลูกค้าที่มีต่อบริการของโรงแรม ได้แก่ Food, Different, Hospitality, Location, Romance และ Site ซึ่งสามารถนำไปใช้วิเคราะห์จุด

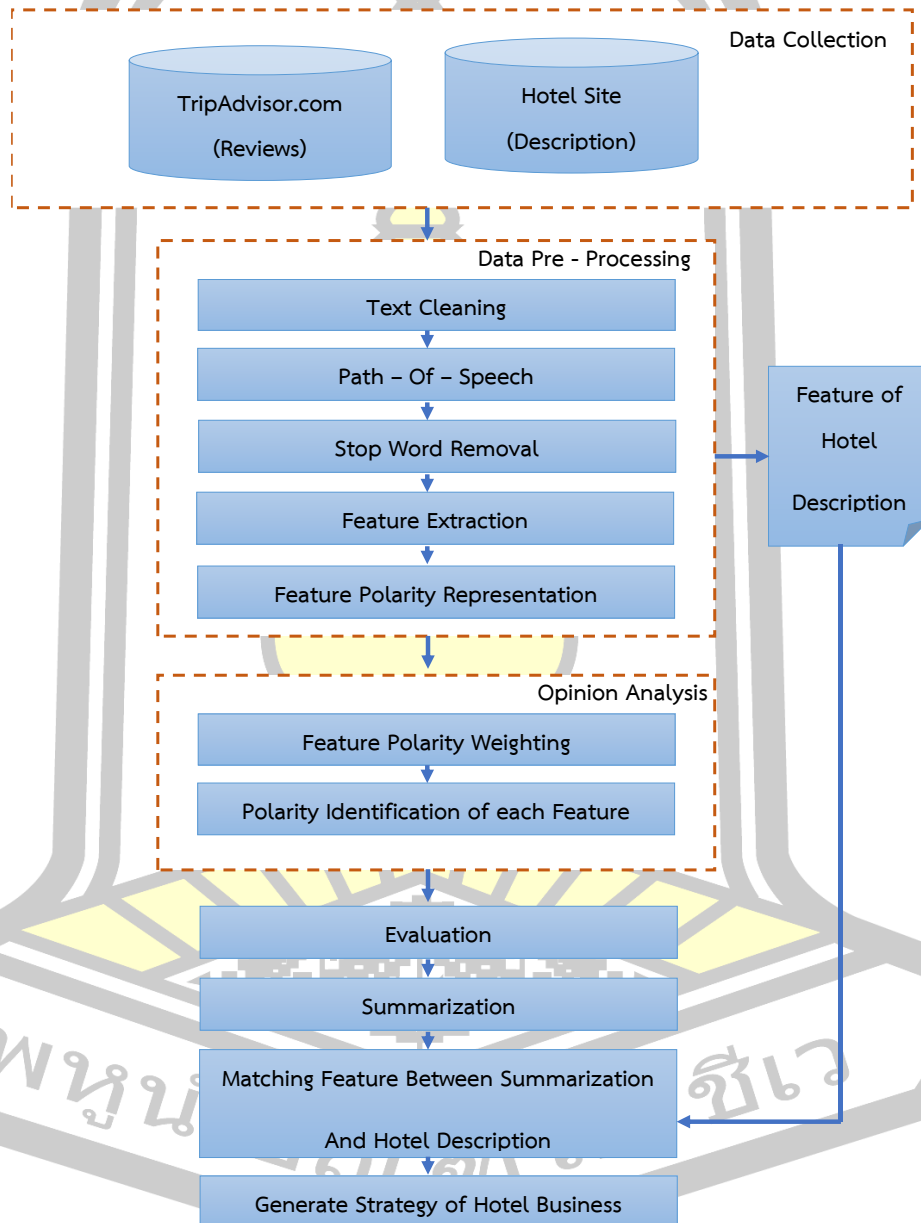
แข็งและจุดอ่อนของโรงแรมเพื่อนำไปปรับปรุงธุรกิจได้ เช่น ผลการจำแนกความคิดเห็นพบว่ามียบทวิจารณ์จำนวน 95 ข้อความ มีความคิดเห็นในเชิงบวกที่ชัดเจนเกี่ยวกับด้าน Location และบทวิจารณ์จำนวน 45 ข้อความ มีความคิดเห็นในเชิงบวกที่ชัดเจนเกี่ยวกับด้าน Romance ฝ่ายบริหารของโรงแรมสามารถนำผลการจำแนกความคิดเห็นจากบทวิจารณ์นี้ไปใช้เป็นแนวทางในการพัฒนากลยุทธ์การตลาดของโรงแรมได้

จากการศึกษางานวิจัยต่างๆ พบว่า ขั้นตอนการพัฒนาการจำแนกความคิดเห็น และการวิเคราะห์ข้อมูลคุณสมบัติต่างๆ ของผู้แสดงความคิดเห็น เช่น เพศ อายุ อาชีพ รายได้ การศึกษา เหล่านี้เป็นต้น ส่วนช่วยพัฒนาธุรกิจและองค์กร และช่วยตอบสนองความต้องการของผู้ใช้บริการ ซึ่งจะเป็นตัวช่วยให้ธุรกิจหรือองค์กร เข้าใจถึงความต้องการของผู้ใช้บริการ รวมถึงการบริหารจัดการลูกค้าสัมพันธ์ซึ่งเป็นส่วนหนึ่งของ CRM นั่นเอง แหล่งข้อมูลที่น่าสนใจในการวิเคราะห์จากเว็บไซต์เครือข่ายสังคมออนไลน์ ได้แก่ ในเพจเฟซบุ๊ก ทวิตเตอร์ บล็อก จากเว็บไซต์ของโรงแรม เช่น Agoda (www.agoda.com) เว็บไซต์ TripAdvisor (www.tripadvisor.com) และเว็บไซต์ venere (www.venere.com) จากฐานข้อมูล UCI โรงแรม และจากงานวิจัยอื่น เป็นต้น โดยมีเครื่องมือในการรวบรวมข้อมูลดังนี้ กราฟ API เฟสบุ๊ค ทวิตเตอร์ เครื่องมือ PageFaceger Import.IO รวมถึงซอฟต์แวร์ที่ดึงจากแท็ก HTML ของเว็บไซต์ การพัฒนาการจำแนกความคิดเห็นร่วมกับโปรไฟล์ของผู้แสดงความคิดเห็น โดยทั่วไปมีขั้นตอนต่างๆ ดังนี้ 1) การตัดคำ ประกอบด้วย การตัดตามช่องว่าง การแบ่งเป็นประโยค การตัดโดยใช้พจนานุกรม การตัดโดยใช้คลังคำข้อมูล 2) การกำจัดคำหยุด คือ การลบคำที่ไม่จำเป็นในการจำแนกข้อความคิดเห็นออกจากเอกสาร คำหยุดเหล่านี้ได้แก่ คำบุพบท คำสรรพนาม คำเชื่อม และคำนำหน้านาม เป็นต้น 3) การหารากคำศัพท์ คือ การหาคำเดิมของคำศัพท์ 4) การสกัดคุณลักษณะ คือ การดึงคุณลักษณะของข้อความออกมา ได้แก่ การถอนข้อความด้วยถุงคำ (Bag of Word) การแทนข้อความด้วยค่าความถี่การเกิดคำ (Term Frequency : TF) และการแทนข้อความด้วยค่าความถี่ผกผัน (Term Frequency – Inverse Document Frequency : TF-IDF) 5) การเลือกคุณลักษณะ เช่น Information Gain หรือคัดเลือกจากคุณลักษณะของโดเมนที่สนใจ 6) การจำแนก สามารถจำแนกด้วย เครื่องเรียนรู้ การจัดกลุ่ม เป็นต้น 7) การวัดประสิทธิภาพงานวิจัย การหาค่า Precision Recall F-Measure Accuracy เป็นต้น

บทที่ 3

วิธีดำเนินการวิจัย

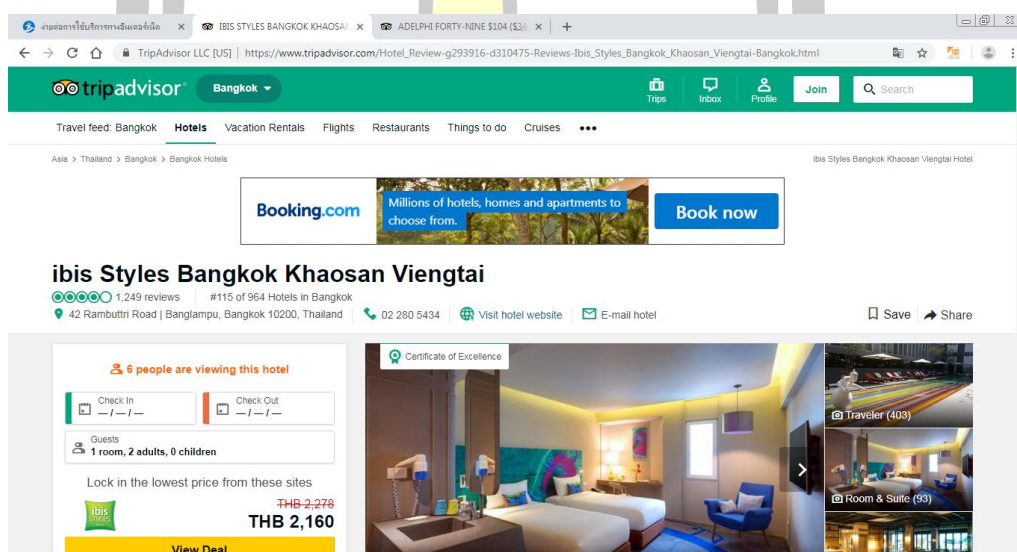
ในบทนี้จะกล่าวถึงขั้นตอนต่างๆ ในการจำแนกความคิดเห็นของโรงแรมเพื่อใช้ในเชิงธุรกิจ โดยมีขั้นตอนต่างๆ ดังรูปที่ 7



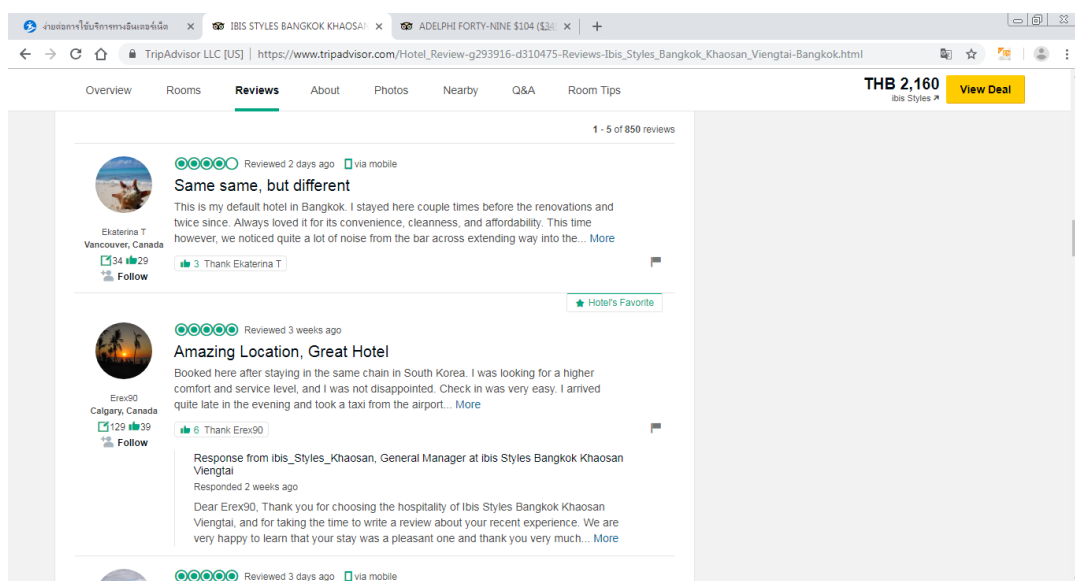
รูปที่ 7 โมเดลแสดงกระบวนการจำแนกความคิดเห็นของลูกค้าโรงแรมเพื่อใช้ในเชิงธุรกิจ

3.1 การเก็บรวบรวมข้อมูล (Data Collection)

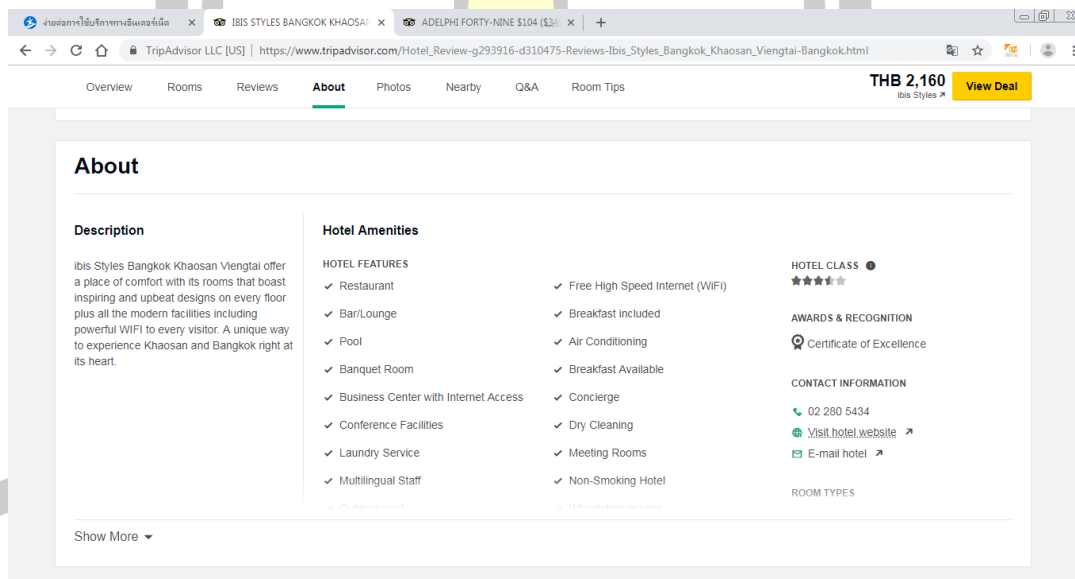
ข้อมูลที่ใช้ในงานวิจัยนี้ ผู้วิจัยได้รวบรวมข้อความความคิดเห็นเกี่ยวกับโรงแรม ซึ่งเป็นรูปแบบของข้อความ (Text) ที่ถูกเขียนขึ้นจากประสบการณ์ของผู้เข้าพักโรงแรม ประกอบไปด้วยความคิดเห็นหลากหลาย ทั้งความคิดเห็นเชิงบวกและความคิดเห็นเชิงลบที่มีต่อคุณลักษณะที่สนใจ ความคิดเห็นจากผู้เข้าพักเหล่านี้สามารถบ่งบอกถึงความรู้สึก หรือทัศนคติของผู้เข้าพักที่มีต่อโรงแรมได้ ผู้วิจัยได้รวบรวมความคิดเห็นของผู้เข้าพักโรงแรม จำนวน 30 แห่ง จากเว็บไซต์ www.tripadvisor.com โดยใช้ Import.io APIs ในการรวบรวมความคิดเห็นของผู้เข้าพัก ตั้งแต่ปี ค.ศ. 2004 - 2019 ตัวอย่างข้อความความคิดเห็น เช่น “The room sure is tiny, yet very clean and comfy.”, “This hotel ok for it 's price.”, “Great hotel great location excellent breakfast but twin beds and room size is very small.” ดังรูปที่ 8 ถึง รูปที่ 10



รูปที่ 8 โรงแรม Ibis Styles Bangkok Khaosan Viengtai ในเว็บไซต์ www.tripadvisor.com

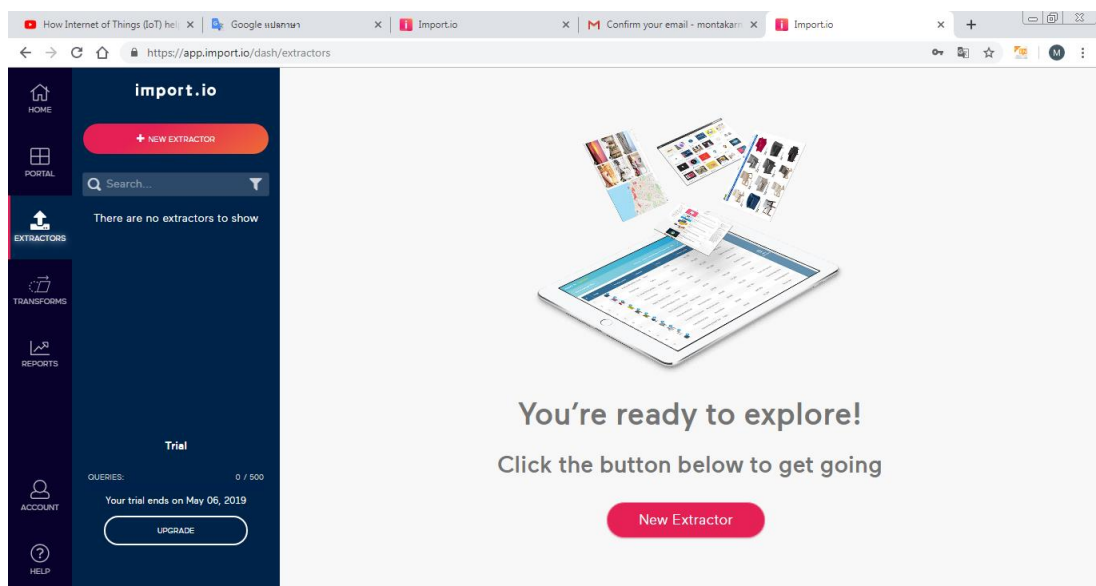


รูปที่ 9 ข้อความคิดเห็นของผู้เข้าพักโรงแรม Ibis Styles Bangkok Khaosan Viengtai

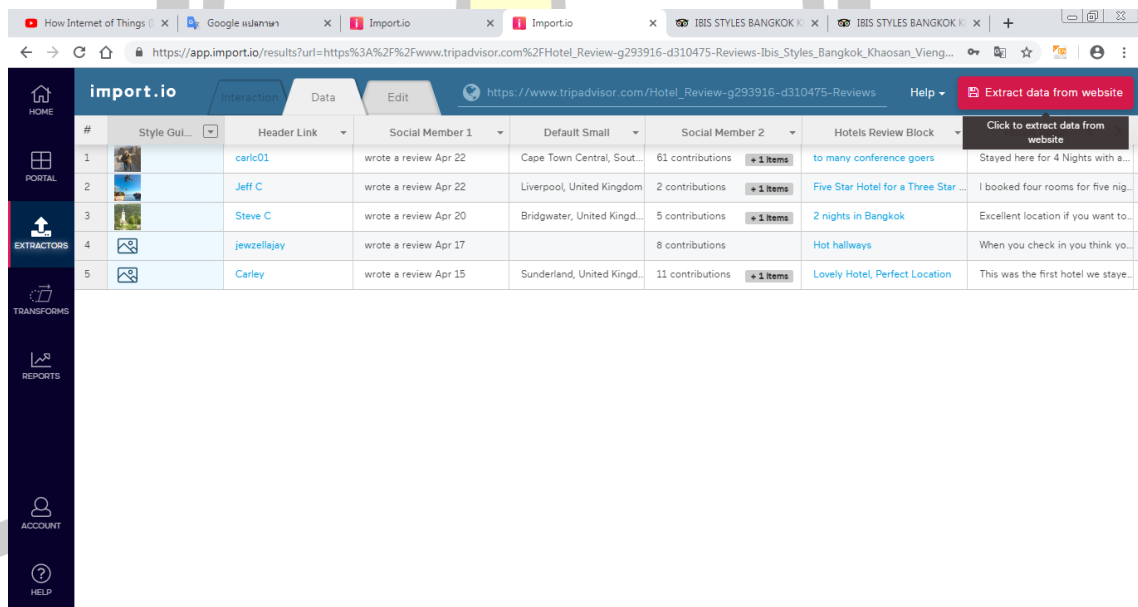


รูปที่ 10 คำอธิบายรายละเอียดของโรงแรม Ibis Styles Bangkok Khaosan Viengtai

ในการเก็บรวบรวมข้อความคิดเห็นของผู้เข้าพักโรงแรม และคำอธิบายรายละเอียดของโรงแรม โดยใช้เว็บไซต์ Import.io ในการเก็บรวบรวมข้อมูล ดังรูปที่ 11 ดังรูปที่ 12



รูปที่ 11 เว็บไซต์ Import.io



รูปที่ 12 ลักษณะของเว็บไซต์ Import.io ที่ใช้ในการเก็บรวบรวมข้อความความคิดเห็น

ในการคัดเลือกข้อความความคิดเห็นที่ใช้ในการทดลองวิเคราะห์จะคัดเลือกเอาเฉพาะข้อความที่เป็นภาษาอังกฤษ และข้อความความคิดเห็นที่มีคำคุณลักษณะ (Feature) 5 คุณลักษณะ คือ Location Service Food Rooms Price รวมทั้งคำระบุข้อความความคิดเห็น ที่ปรากฏอยู่ในข้อความความคิดเห็นนั้น ถ้าข้อความความคิดเห็นใดไม่มีคุณลักษณะและคำระบุข้อความความคิดเห็นจะไม่คัดเลือกมาทดลอง เนื่องจากไม่

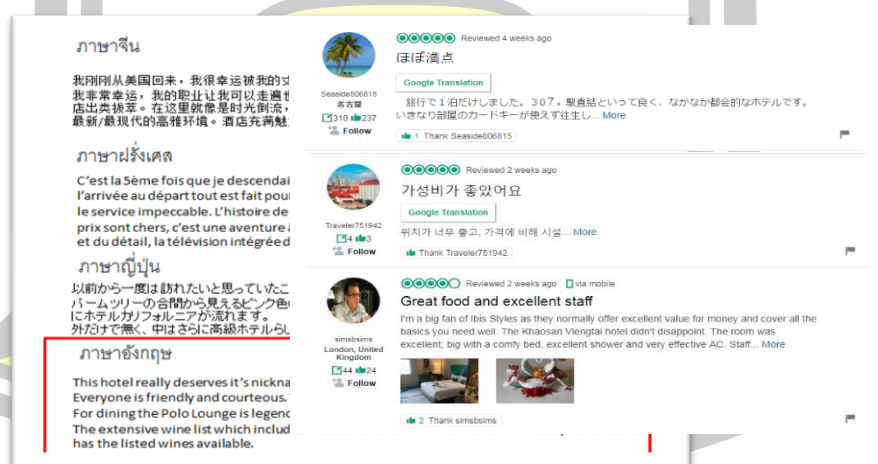
สามารถระบุได้ว่าค่าแสดงความรู้สึกนั้นเป็นของคุณลักษณะใด และในการระบุผลเฉลี่ย ผู้วิจัยได้คัดเลือกเอาเฉพาะข้อความความคิดเห็นที่ผู้เชี่ยวชาญมีความเห็นตรงกันมากกว่า 3 คน

โดยวิธีการระบุผลเฉลี่ยจะให้ผู้เชี่ยวชาญ อ่านและระบุว่าในแต่ละข้อความความคิดเห็นกล่าวถึงคุณลักษณะใดในเชิงบวกหรือเชิงลบ ซึ่งคุณสมบัติของผู้เชี่ยวชาญจะต้องมีความรู้ความสามารถในการอ่านข้อความภาษาอังกฤษ ดังนั้นงานวิจัยนี้จะใช้นักศึกษาคณะการท่องเที่ยวและการโรงแรม และนักศึกษาคณะมนุษยสัมพันธ์ ปีที่ 4 จำนวน 5 คน ที่มีความสามารถด้านภาษาอังกฤษ เป็นผู้เชี่ยวชาญในการอ่านระบุผลเฉลี่ย

3.2 กระบวนการเตรียมข้อมูล (Data Preprocessing)

3.2.1 การทำความสะอาดข้อความ (Text Cleaning)

เนื่องจากผู้เข้าพักโรงแรมมีหลากหลายเชื้อชาติ ข้อความความคิดเห็นของผู้เข้าพักโรงแรมจึงมีหลากหลายภาษา ซึ่งมีรูปแบบอักขระที่ใช้แตกต่างกันไปและทำให้ไม่สามารถนำไปวิเคราะห์ได้ งานวิจัยนี้ ผู้วิจัยได้ทำการคัดเลือกเฉพาะข้อความความคิดเห็นที่เขียนด้วยภาษาอังกฤษ มาใช้ในการวิเคราะห์ ตัวอย่างข้อความความคิดเห็นภาษาต่างๆ ที่รวบรวมจากเว็บไซต์แสดงดังรูปที่ 13 และตารางที่ 8



รูปที่ 13 ตัวอย่างข้อความความคิดเห็นภาษาต่างๆ

ตารางที่ 8 ตัวอย่างข้อความคิดเห็นภาษาต่างๆ

ภาษาจีน 我刚刚从美国回来，我很幸运被我的丈夫戴维带到了比佛利山庄的酒店。哇哦。这就是我所能说的。我非常幸运，我的职业让我可以走遍世界各地，在我这期间，我们可以住在很好的酒店。这个酒店出类拔萃。在这里就像是时光倒流，直到玛琳黛德丽来到这一天...但它拥有所有2018年最新/最现代的高雅环境。酒店充满魅力和历！	คัดออก
ภาษาฝรั่งเศส C'est la 5ème fois que je descendais dans cet hôtel qui est simplement magique. De l'arrivée au départ tout est fait pour vous simplifier la vie. Les chambres sont splendides et le service impeccable. L'histoire de l'hôtel rend le lieu encore plus magique et même si les prix sont chers, c'est une aventure à vivre absolument au moins une fois. Symbole du luxe et du détail,	
ภาษาญี่ปุ่น 以前から一度は訪れたいと思っていたこのホテル。残念ながら宿泊では無く、観光に行ってきた。パームツリーの合間から見えるピンク色の外観はイーグルスのレコードジャケットそのまま。頭の中にホテルカリフォルニアが流れます。外だけでなく、中はさらに高級ホテルらしい装い。	
ภาษาอังกฤษ This hotel really deserves it's nickname The Pink Palace as you are treated like royalty. Everyone is friendly and courteous. The rooms are well appointed and spotless. For dining the Polo Lounge is legendary and deserves its reputation for excellent cuisine. The extensive wine list which includes the best of local and international wines up to date and has the listed wines available.	นำมาใช้เป็นชุดข้อมูล

3.2.2 การกำกับชนิดของคำ (Path-of-Speech Tagging)

การกำกับชนิดของคำหรือประเภทของคำ เป็นการระบุหน้าที่ของแต่ละคำในข้อความคิดเห็น ซึ่งมีทั้งหมด 8 ชนิด คือ 1) คำนาม (Noun) 2) คำสรรพนาม (Pronoun) 3) คำกริยา (Verb) 4)

คำกริยาวิเศษณ์ (Adverb) 5) คำคุณศัพท์ (Adjective) 6) คำบุพบท (Preposition) 7) คำสันธาน (Conjunction) และ 8) คำอุทาน (Interjection) โดยคำแต่ละประเภทมีหน้าที่แตกต่างกันออกไป เช่น คำนาม ทำหน้าที่เป็นประธานและกรรมของประโยค คำกริยา ทำหน้าที่บ่งบอกการกระทำของประธาน ส่วนคำบุพบท ทำหน้าที่เชื่อมคำ เป็นต้น ชุดแท็ก (Tags) มาตรฐานที่ใช้ระบุหน้าที่ของคำในภาษาอังกฤษที่นิยมใช้ ได้แก่ Penn treebank, Bown corpus tag-set และ British national corpus โดย Penn Treebank มีจำนวน 48 แท็ก, Bown corpus tag-set มีจำนวน 192 แท็ก และ British national corpus มีจำนวน 61 แท็ก งานวิจัยนี้ใช้ Penn Treebank Tagset ในการกำกับชนิดของคำ เนื่องจาก มีจำนวนแท็กไม่เยอะจนเกินไป ซึ่งมีรายละเอียดแท็กดังแสดงในรูปที่ 14

CC	Coordinating conj.	TO	infinitival <i>to</i>
CD	Cardinal number	UH	Interjection
DT	Determiner	VB	Verb, base form
EX	Existential there	VBD	Verb, past tense
FW	Foreign word	VBG	Verb, gerund/present pple
IN	Preposition	VBN	Verb, past participle
JJ	Adjective	VBP	Verb, non-3rd ps. sg. present
JJR	Adjective, comparative	VBZ	Verb, 3rd ps. sg. present
JJS	Adjective, superlative	WDT	Wh-determiner
LS	List item marker	WP	Wh-pronoun
MD	Modal	WPS	Possessive <i>wh</i> -pronoun
NN	Noun, singular or mass	WRB	Wh-adverb
NNS	Noun, plural	#	Pound sign
NNP	Proper noun, singular	\$	Dollar sign
NNPS	Proper noun, plural	.	Sentence-final punctuation
PDT	Predeterminer	,	Comma
POS	Possessive ending	:	Colon, semi-colon
PRP	Personal pronoun	(	Left bracket character
PPS	Possessive pronoun	)	Right bracket character
RB	Adverb	"	Straight double quote
RBR	Adverb, comparative	'	Left open single quote
RBS	Adverb, superlative	"	Left open double quote
RP	Particle	'	Right close single quote
SYM	Symbol	"	Right close double quote

รูปที่ 14 POS Tagging (Penn Treebank Tagset)

3.2.3 การกำจัดคำหยุด (Stop Word Removal)

คำหยุดเป็นคำที่ไม่มีความหมายที่ปรากฏในข้อความคิดเห็น ซึ่งเป็นคำที่ไม่ควรนำไปวิเคราะห์ เนื่องจากไม่ได้สื่อความหมายและยังทำให้การวิเคราะห์ข้อมูลทำงานช้าลงด้วย คำหยุดที่มักพบบ่อยในข้อความภาษาอังกฤษ ได้แก่ “a”, “and”, “the”, “you”, “are”, “I” เป็นต้น ตัวอย่างข้อความคิดเห็น เช่น “The rooms the staff and service are all wonderful a very special place to visit I couldn't fault it and could have stayed a lot longer.” มีคำหยุดในประโยคคือ “The”, “and”, “are”, “all”, “a”, “to”, “I”, “couldn't”, “it”, “could”, “have” ในงานวิจัยนี้ทำการลบคำหยุดออกจากข้อความคิดเห็นโดยใช้คลังคำหยุดมาตรฐาน จาก Stanford Stopword List ซึ่งอยู่ในรูปแบบเอกสาร (Text file) จำนวน 257 คำ ดังแสดงตารางที่ 9 ขั้นตอนนี้จะทำหลังจากที่ทำการกำกับหน้าที่ของคำแล้ว เพราะถ้าหากกำจัดคำหยุดก่อน อาจทำให้คำบางคำที่เป็นคำนาม ถูกตัดออกไป

ตารางที่ 9 คลังคำหยุดมาตรฐาน จาก Stanford Stopword List

!!	-	and	doesn't	herself	my	should	to	whom	Hasn't
?!	(	any	doing	him	myself	shouldn't	too	why	Haven't
??	)	are	don't	himself	no	so	under	why's	hes
!?	&	aren't	down	his	nor	some	until	with	heres
`	%	as	during	how	not	such	up	won't	hows
``	\$	at	each	how's	of	than	very	would	im
"	@	be	few	i	off	that	was	wouldn't	Isn't
-lrb-	!	because	for	i'd	on	that's	wasn't	you	its
-rrb-	^	been	from	i'll	once	the	we	you'd	lets
-lsb-	#	before	further	i'm	only	their	we'd	you'll	Mustn't
-rsb-	*	being	had	i've	or	theirs	we'll	you're	Shan't
,	..	below	hadn't	if	other	them	we're	you've	shes
.	...	between	has	in	ought	themselves	we've	your	Shouldn't
:	'll	both	hasn't	into	our	then	were	yours	thats
;	's	but	have	is	ours	there	weren't	yourself	theres
"	'm	by	haven't	isn't	ourselves	there's	what	yourselves	theyll
'	a	can	having	it	out	these	what's	###	theyre
?	about	can't	he	it's	over	they	when	return	theyve
<	above	cannot	he'd	its	own	they'd	when's	Aren't	Wasn't
>	after	could	he'll	itself	same	they'll	where	Can't	were
{	again	couldn't	he's	let's	shan't	they're	where's	Couldn't	Weren't
}	against	did	her	me	she	they've	which	Didn't	whats
[	all	didn't	here	more	she'd	this	while	Doesn't	whens
]	am	do	here's	most	she'll	those	who	Don't	wheres
+	an	does	hers	mustn't	she's	through	who's	Hadn't	whos

3.2.4 การสกัดคุณลักษณะและการสกัดคำระบุข้อความคิดเห็น (Feature and Polarity Extraction)

หลังจากทำการกำจัดคำหยุดเรียบร้อยแล้ว ผู้วิจัยได้ทำการสกัดคำคุณลักษณะและคำระบุข้อความคิดเห็น เพื่อเลือกคำคุณลักษณะสำคัญที่ใช้ในการระบุคุณสมบัติและคำระบุข้อความคิดเห็นมาใช้ในการวิเคราะห์ความรู้สึกของสิ่งที่สนใจ [5] [38] [39] [43] [44] ซึ่งประกอบไปด้วย ทำเลที่ตั้ง (Location) ห้อง (Rooms) การบริการ (Service) อาหารและเครื่องดื่ม (Food) และราคาห้องพัก (Price) โดยคำคุณลักษณะเหล่านี้ ถูกรวบรวมจากงานวิจัยที่ได้ศึกษามาก่อนหน้า ซึ่งนอกเหนือจากคำเหล่านี้ยังมีคำที่มีความหมายเหมือนคำคุณลักษณะเหล่านี้ปรากฏอยู่ ซึ่งอยู่ในประเภทของคำนาม ดังนั้น จึงสกัดเอาคำนามที่มีอยู่ในประโยคความคิดเห็นมาช่วยในการวิเคราะห์ โดยคำนามเหล่านี้จะต้องเป็นคำที่มีความหมายเหมือนหรือใกล้เคียงกับคำคุณลักษณะข้างต้น (Synonym) ซึ่งในการหาคำที่มีความหมายเหมือนคำคุณลักษณะใช้พจนานุกรมคำ Synonym ตัวอย่างคำที่มีความหมายเหมือนคำคุณลักษณะดังตารางที่ 10

ตารางที่ 10 ตารางคำ Synonym

ลำดับ	คำคุณลักษณะ	คำที่มีความหมายเหมือนคำคุณลักษณะ
1	Location	Area, City, Street, Metro, Station, Train, Bus, Airport, Distance, Place
2	Service	Staff, Reception, Front desk, Wi-Fi, Pool, Spa, Gym
3	Food	Breakfast, Bar, Coffee, Buffet, Dinner, Fruit, Bakery, Restaurant
4	Room	Bathroom, Floor, Shower, Size, Window, Door, View, Building, room
5	Price	Value, Cost, Amount, Discount, Payment, Bill, Cheap, Expensive, Worth

หลังจากทำการสกัดคำคุณลักษณะจากข้อความคิดเห็นของผู้เข้าพักโรงแรมแล้ว ผู้วิจัยได้ทำการสกัดคำระบุข้อความคิดเห็นตามหน้าที่ของคำ โดยเลือกคำที่ทำหน้าที่เป็น คำกริยา (Verb) คำคุณศัพท์ (ADJ) และคำกริยาวิเศษณ์ (ADV) เนื่องจากคำประเภทนี้สามารถบ่งบอกถึงความรู้สึกได้ ดังแสดงในตารางที่ 11 ตัวอย่างวิธีการสกัดคำระบุข้อความคิดเห็นแสดงดังรูปที่ 15

ตารางที่ 11 ชนิดของคำที่นำมาวิเคราะห์

ตัวย่อ	คำอธิบาย	ใช้ระบุประเภทคำ
NN	Noun	Feature
NNS	Noun	Feature
NNP	Noun	Feature
NNPS	Noun	Feature
JJ	Adjective, comparative	Polarity
JJR	Adjective, superlative	Polarity
JJS	Adjective, base form	Polarity
VB	Verb, base form	Polarity
VBD	Verb, past tense	Polarity
VBG	Verb, gerund	Polarity
VCN	Verb, past	Polarity
VBP	Verb, non-3d	Polarity
VBZ	Verb 3d	Polarity
RB	Adverb	Polarity
RBR	Adverb, comparative	Polarity
RBS	Adverb, superlative	Polarity

จากตารางที่ 11 การคัดเลือกชนิดของคำที่จะแทนคำคุณลักษณะ (Feature) ได้แก่ คำนาม (Noun) โดยมีตัวย่อ ได้แก่ NN, NNS, NNP, NNPS และคัดเลือกชนิดของคำที่จะแทนคำระบุข้อความคิดเห็น (Polarity) จากชนิดของคำคุณศัพท์ (Adjective) โดยมีตัวย่อ ได้แก่ JJ, JJR, JJS คำกริยา (Verb) โดยมีตัวย่อ ได้แก่ VB, VBD, VBG, VBN, VBP, VBZ และคำขยายกริยา (Adverb) โดยมีตัวย่อ ได้แก่ RB, RBR, RBS เป็นต้น

พหุ ประสิทธิภาพ



Result

Rooms [NNS] staff [NN] service [NNS] wonderful [VBP] very [RB] special [JJ] place [NN] visit [VB] I [JJ] couldn [NN]'[NNP]...



wonderful [VBP] very [RB] special [JJ] visit [VB] I [JJ]

รูปที่ 15 ขั้นตอนการสกัดคำระบุข้อความคิดเห็น (Polarity Extraction)

จากรูปที่ 15 แสดงขั้นตอนการสกัดคำระบุข้อความคิดเห็น โดยใช้ Pen Treebank Tagset ในการกำกับหน้าที่ของคำ จากนั้นเลือกคำที่มีหน้าที่เป็นคำกริยา (Verb) คำคุณศัพท์ (ADJ) และ คำกริยาวิเศษณ์ (ADV) เพื่อนำไปใช้ในการระบุข้อความคิดเห็นในขั้นตอนถัดไป

3.2.5 การแทนค่าคำระบุข้อความคิดเห็น (Polarity Representation)

หลังจากทำการสกัดคำคุณลักษณะและคำระบุข้อความคิดเห็นเรียบร้อยแล้ว ขั้นตอนต่อมา ทำการแทนค่าคำระบุข้อความคิดเห็นโดยใช้คลังคำมาตรฐาน SentiWordNet ซึ่งประกอบด้วย คำศัพท์และค่าคะแนนของคำ โดยงานวิจัยนี้ใช้หลักการพิจารณาค่าคะแนนที่ปรากฏในช่องคะแนน ถ้าปรากฏค่าคะแนนในช่องคำเชิงบวกแสดงว่าคำศัพท์นั้นเป็นคำเชิงบวก ถ้าปรากฏค่าคะแนนในช่อง คำเชิงลบแสดงว่าคำศัพท์นั้นเป็นคำเชิงลบ แต่ถ้าค่านั้นปรากฏในหลายส่วนของ SentiWordNet แล้ว ปรากฏค่าคะแนนทั้งในช่องคำเชิงบวกและคำเชิงลบ จะพิจารณาจากจำนวนของคำที่ปรากฏในแต่ละ ด้าน ตัวอย่างคำระบุข้อความคิดเห็นที่ได้จากคลังคำ SentiWordNet แสดงดังตารางที่ 12

ตารางที่ 12 ตัวอย่างคำระบุข้อความคิดเห็นจากคลังคำ SentiWordNet

ลำดับ	คำศัพท์	คะแนนคำเชิงบวก	คะแนนคำเชิงลบ	ผลการสรุป
1	Clean	0.5	0	บวก
2	pretty	0.5	0	บวก
3	ugly	0	0.375	ลบ
4	lovely	0.625	0	บวก
5	basic	0	0.25	ลบ
6	bad	0	0.625	ลบ
7	good	0.75	0	บวก

กรณี que คำศัพท์ปรากฏในหลายส่วนของ SentiWordNet ซึ่งขึ้นอยู่กับบริบทและหน้าที่ของคำ งานวิจัยนี้ได้นำเอาคำศัพท์ในแต่ละบริบทมาวิเคราะห์ร่วมกัน ถ้าคำศัพท์นั้นปรากฏเป็นคำเชิงบวก มากกว่า จะสรุปว่าคำศัพท์นั้นเป็นคำเชิงบวก และถ้าคำศัพท์นั้นเป็นคำเชิงลบที่ปรากฏมากกว่าจะ สรุปว่าคำศัพท์นั้นเป็นคำเชิงลบ ตัวอย่าง เช่น คำว่า Good มีคำศัพท์ที่ปรากฏใน SentiWordNet ทั้งหมดคือ 27 คำ พบว่า มีคำที่ให้ความหมายในเชิงบวก จำนวน 23 คำ และให้ความหมายในเชิงลบ จำนวน 1 คำ ดังนั้น จึงสรุปว่าคำศัพท์ Good เป็นคำที่แสดงความรู้สึกในเชิงบวก ตัวอย่างการ พิจารณาคำศัพท์ Good ดัง

ตารางที่ 13

ตารางที่ 13 ตัวอย่างค่าคะแนนเชิงบวกและเชิงลบของคำศัพท์ Good ใน SentiWordNet

ลำดับ	คำศัพท์	คะแนนคำเชิงบวก	คะแนนคำเชิงลบ
1	Good	0.625	0
2	Good	0	0
3	Good	0.5	0
4	Good	0.625	0
5	Good	0	0
6	Good	0.75	0
7	Good	0	0
8	Good	0.625	0
9	Good	0.375	0.125
10	Good	0.5	0
11	Good	0.75	0
12	Good	1	0
13	Good	0.875	0
14	Good	0.75	0
15	Good	0.75	0
16	Good	1	0
17	Good	0.375	0
18	Good	0.65	0
19	Good	1	0
20	Good	0.75	0
21	Good	0.625	0

22	Good	0.5	0
23	Good	0.875	0
24	Good	0.625	0
25	Good	0.5	0
26	Good	0.375	0
27	Good	0	0
จำนวนคำศัพท์ที่ให้ค่าคะแนน		เชิงบวก 23 คำ	เชิงลบ 1 คำ

ในการจัดการคลังคำความรู้สึกใน SentiWordNet ที่มีคำศัพท์เป็นจำนวน 49,578 คำ และแต่ละคำมีระดับของการแสดงออกถึงความรู้สึกที่แตกต่างกัน เช่น คำว่า “Good” มีการแสดงความรู้สึกทั้งเชิงบวกและเชิงลบ และการแสดงความรู้สึกของแต่ละคำทั้งในเชิงบวกและเชิงลบนั้นถูกแบ่งเป็นหลายระดับความรู้สึกตามค่าคะแนน ส่งผลให้การเลือกใช้คำแสดงความรู้สึกเพื่อไประบุข้อความคิดเห็นเป็นเรื่องที่ยุ่งยากและไม่ชัดเจน ซึ่งการระบุความรู้สึกในแต่ละความคิดเห็นไม่ควรนำค่าคะแนนไปหักล้างกันเนื่องจากคำที่เป็นเชิงบวกไม่สามารถหักล้างด้วยค่าเชิงลบในระดับค่าคะแนนได้ ดังนั้น งานวิจัยนี้ได้จัดการคำแสดงความรู้สึกใน SentiWordNet ให้เป็นที่ชัดเจนว่าเป็นข้อความคิดเห็นในเชิงบวกหรือในเชิงลบ โดยการเปรียบเทียบปริมาณการแสดงความรู้สึกในแต่ละด้าน นอกจากนี้ ผู้วิจัยได้ทำการตรวจสอบคำศัพท์ใน SentiWordNet พบว่า มีคำที่แสดงออกถึงความรู้สึกที่ไม่ซ้ำซ้อนกัน จำนวน 38,318 คำ แต่ในคำที่ไม่ซ้ำซ้อนกันนี้พบว่ามีคำที่มีความหมายคลุมเครือหรือคำที่มีปริมาณค่าเชิงบวกค่าเชิงลบเท่ากัน ผู้วิจัยได้ทำการตัดคำศัพท์ที่มีความหมายคลุมเครือออก ซึ่งหลังจากทำการตัดคำที่คลุมเครือออกแล้ว มีจำนวนคำศัพท์คงเหลือ จำนวน 36,872 คำ

3.3 การวิเคราะห์ความคิดเห็น (Opinion Analysis)

หลังจากดำเนินการเตรียมข้อมูลเรียบร้อยแล้ว ผู้วิจัยได้ทำการวิเคราะห์ และสรุปความคิดเห็นในแต่ละคุณลักษณะว่าเป็นความคิดเห็นเชิงบวกหรือเชิงลบ โดยตรวจสอบจำนวนของคำบริบทที่แสดงออกถึงความรู้สึกที่ปรากฏด้านหน้าของคุณลักษณะแต่ละคำ โดยเขียนเป็นกฎ ได้ดังนี้ 1) ถ้าปรากฏคำแสดงความรู้สึกในเชิงบวกมากกว่าเชิงลบ แสดงว่าคุณลักษณะนั้นเป็นเชิงบวก 2) ถ้าคำที่ปรากฏอยู่ด้านหน้าของคุณลักษณะเป็นคำเชิงลบจำนวนมากกว่าแสดงว่าคุณลักษณะนั้นเป็นคำเชิงลบ 3) ถ้ามีจำนวนเท่ากันจะใช้คำบริบทที่อยู่ด้านหลังที่แสดงถึงความรู้สึกคำแรกมาเป็นตัวตัดสินในการระบุข้อความคิดเห็นของคุณลักษณะนั้น ตัวอย่างประโยคเช่น “Room clean wifi Internet connection slow perfect location” คุณลักษณะ “Room” ไม่มีคำระบุข้อความคิดเห็นอยู่ด้านหน้าของคำศัพท์ ดังนั้น จึงนำเอาคำที่อยู่ด้านหลังของคำศัพท์มาเป็นตัวตัดสิน พบว่ามีคำว่า “Clean” ซึ่งเป็นคำเชิงบวก ดังนั้นจึงสรุปว่าคุณลักษณะ “Room” เป็นเชิงบวก คุณลักษณะ

“wifi internet connection” อยู่ในกลุ่มของคุณลักษณะ “Service” จากการตรวจสอบคำระบุข้อความความคิดเห็นที่ปรากฏอยู่ด้านหน้าของคำคุณลักษณะ พบว่ามีคำว่า “clean” ซึ่งเป็นคำเชิงบวก ปรากฏอยู่ ดังนั้นจึงสรุปได้ว่าคุณลักษณะ “Service” เป็นเชิงบวก และคุณลักษณะ “Location” มีคำบริบทที่แสดงออกถึงความรู้สึกอยู่ด้านหน้าคือ “perfect” ซึ่งเป็นคำเชิงบวก ดังนั้นจะสรุปได้ว่าคุณลักษณะ “Location” เป็นเชิงบวก แต่ไม่ถูกต้องเสมอไปเนื่องจากโครงสร้างของการเขียนประโยคภาษาอังกฤษสามารถเขียนคำคุณศัพท์หรือคำที่บ่งบอกถึงความรู้สึกไว้ด้านหน้าหรือด้านหลังของคำคุณลักษณะก็ได้ เช่น Service+ เนื่องจากมีคำเชิงบวกอยู่ด้านหน้าของคุณลักษณะ Service คือ คำว่า Clean ซึ่งจากความคิดเห็นนี้ผู้เข้าพักได้แสดงความคิดเห็นต่อคุณลักษณะ Service เป็นเชิงลบ แต่เขียนไว้ด้านหลังคือคำว่า slow ซึ่งทำให้ผลการวิเคราะห์ความรู้สึกเกิดความผิดพลาด แต่ส่วนใหญ่แล้วผู้เขียนความคิดเห็นจะเขียนไว้ด้านหน้าของคำคุณลักษณะ ดัง ตารางที่ 14

ตารางที่ 14 ตัวอย่างการวิเคราะห์ความคิดเห็น

ตัวอย่าง ความคิดเห็น	Room	clean	wifi	Internet	connection	slow	perfect	location
แทนค่า คุณลักษณะ และระบุข้อความ ความคิดเห็น	Room	+	Service	Service		-	+	Location
วิเคราะห์ ความคิดเห็น	Room+		Service+					Location+

3.4 การวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม (Evaluation)

หลังจากทำการวิเคราะห์ความคิดเห็นเรียบร้อยแล้ว ผู้วิจัยได้นำผลที่ได้จากการทดลองไปเปรียบเทียบกับผลเฉลยที่ได้จากผู้เชี่ยวชาญ เพื่อทำการวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม ซึ่งงานวิจัยนี้ใช้ค่าความแม่นยำ (Precision) และค่าระลึก (Recall) ในการวัดประสิทธิภาพ

1) การวัดค่าความแม่นยำ (Precision)

การวัดค่าความแม่นยำของการวิเคราะห์ความคิดเห็น เป็นการตรวจสอบความถูกต้องของการทำนายโดยพิจารณาจากค่าที่ทำนายถูกต้องหารด้วยค่าที่ทำนายทั้งหมด ดังสมการ (3.1)

$$Precision = \frac{\text{จำนวนข้อมูลที่ทำนายถูกต้อง}}{\text{จำนวนของการทำนายทั้งหมด}} \quad (3.1)$$

2) การวัดค่าความระลึก (Recall)

การวัดค่าความระลึกของวิธีการจำแนกความคิดเห็นเป็นการวัดความถูกต้องของการทำนาย โดยจะพิจารณาจากค่าที่ทำนายถูกต้องหารด้วยค่าจำนวนความคิดเห็นทั้งหมดในข้อมูลที่น่ามาวิเคราะห์ ดังสมการ (3.2)

$$Recall = \frac{\text{จำนวนข้อมูลที่ทำนายถูกต้อง}}{\text{จำนวนความคิดเห็นทั้งหมดในข้อมูลที่น่ามาวิเคราะห์}} \quad (3.2)$$

3) การวัดค่าประสิทธิภาพโดยรวม (F-measure)

การวัดค่าประสิทธิภาพโดยรวม เป็นการพิจารณาค่าความระลึกรวมกับค่าความแม่นยำ ดังสมการ (3.3)

$$F - \text{measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3.3)$$

หลังจากที่ผ่านการประเมินประสิทธิภาพการจำแนกแล้ว การประเมินประสิทธิภาพโดยรวมของโมเดล ให้ผู้เชี่ยวชาญทางด้านโรงแรม (ผู้บริหารระดับสูงของโรงแรม) เป็นผู้พิจารณาผลที่ได้จากการจำแนกข้อความความคิดเห็นในเชิงบวกเชิงลบของแต่ละคุณลักษณะและทำการเปรียบเทียบกับคุณลักษณะที่ถูกระบุใน Description Hotel

ตารางที่ 15 ตัวอย่างตารางผลลัพธ์ที่ได้จากการวิเคราะห์เทียบกับผลเฉลย

ลำดับ ความ คิดเห็น	ผลเฉลย	ผลลัพธ์ที่ได้จากการวิเคราะห์	ค่าความระลึก ถูก=1 ผิด=0	ความแม่นยำ ในการทำนาย ถูก=1 ผิด=0
1	Location+ Room+	Location+ Room+	1, 1	1,1
2	Location+ Room-	Location+ Room+	1, 0	1,0
3	Location- Room-	Location+ Room+	0, 0	0,0
4	Location+ Room-	Location+	1, 0	1
5	Location+ Room-	Location- Room-	0,1	0,1
6	Location+ Room-	Location+ Room+ Service+	1, 0	1, 0, 0

จากตารางที่ 15 หากมีการทำนายคุณลักษณะและชี้ระบุชี้ความคิดเห็นตรงกันในแต่ละความคิดเห็น ถือว่า ทำนายถูกต้อง แต่ถ้ามักมีการทำนายคุณลักษณะผิด หรือคำระบุชี้ความคิดเห็นผิด ก็จะถือว่า ทำนายผิดพลาด ดังนั้นค่าประสิทธิภาพ Precision, Recall, F-Measure

$$\text{Precision} = \frac{6}{12}$$

$$\text{Precision} = 0.5$$

$$\text{Recall} = \frac{6}{12}$$

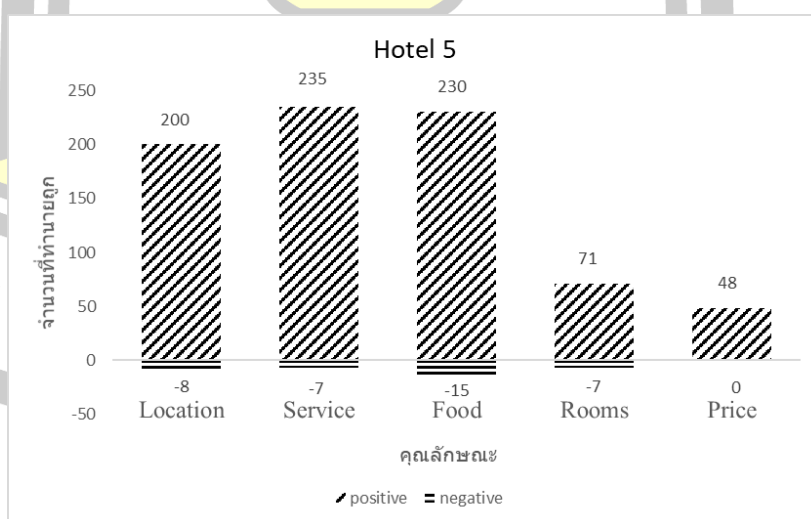
$$\text{Recall} = 0.5$$

$$\text{F - measure} = \frac{2 \times (0.5 \times 0.5)}{(0.5+0.5)}$$

$$\text{F - measure} = 0.5$$

3.5 การสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม (Summarization)

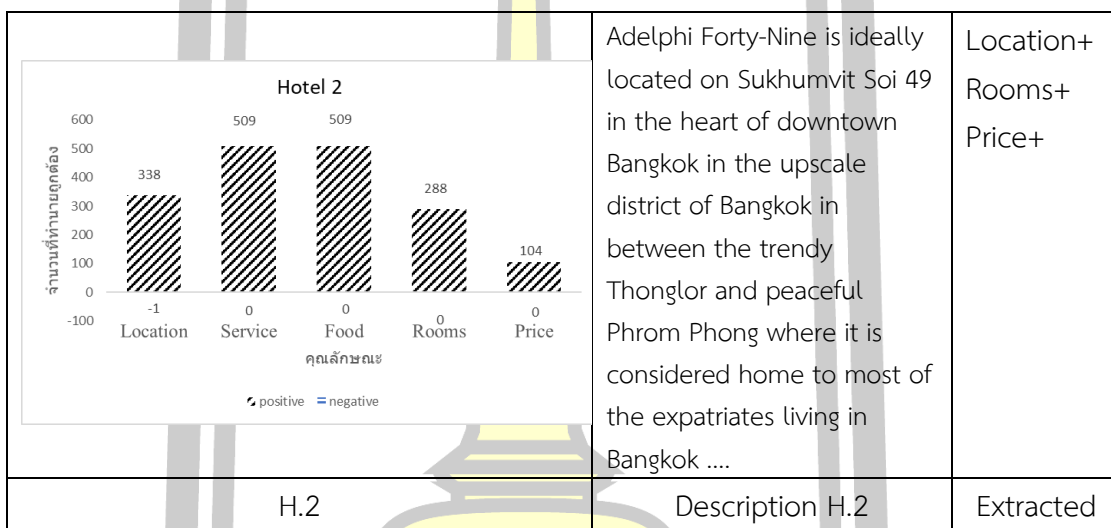
ในการสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม ผู้วิจัยจัดทำเป็นรูปแบบของกราฟแสดงจำนวนที่ทำนายถูกต้องแบ่งตามคุณลักษณะที่สนใจ ทั้ง 5 ลักษณะ ดังแสดงในรูปที่ 16



รูปที่ 16 ตัวอย่างการสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม

3.6 การเปรียบเทียบผลการวิเคราะห์ข้อความคิดเห็นกับคำอธิบายของโรงแรม

หลังจากทำการสรุปผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักแล้ว ทำการเปรียบเทียบว่า ข้อความคิดเห็นของผู้เข้าพักโรงแรม มีความสอดคล้องกับคำอธิบายของโรงแรมใน 5 คุณลักษณะ คือ Location, Service, Food, Room และ Price หรือไม่ จำนวนเท่าใด ถ้าแสดงความคิดเห็นไปในทิศทางเดียวกันกับคำอธิบายของโรงแรมแสดงว่าโรงแรมนั้นมีการเขียนคำอธิบายของโรงแรม สอดคล้องกันกับความคิดเห็นของผู้เข้าพักโรงแรม ดังแสดงในรูปที่ 17



รูปที่ 17 ตัวอย่างการเปรียบเทียบผลการวิเคราะห์ข้อความคิดเห็นกับคำอธิบายของโรงแรม

จากรูปที่ 17 พบว่าผู้เข้าพักมีความคิดเห็นเชิงบวกในด้าน Service, Food, Location, Room และ Price ตามลำดับ ส่วนโรงแรมได้เขียนคำอธิบายจุดเด่นไว้ซึ่งได้สกัดออกเป็น 3 ด้าน คือ Location, Rooms และ Price

พหุ ประสิทธิภาพ

บทที่ 4

ผลการดำเนินงาน

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาขั้นตอนวิธีในการลดคุณลักษณะและแก้ไขปัญหาคูณลักษณะซ้ำซ้อนสำหรับการจำแนกความคิดเห็นที่อยู่บนเครือข่ายสังคมออนไลน์ ประกอบด้วย การเลือกคุณลักษณะ การตรวจสอบคุณลักษณะที่ซ้ำซ้อน การจำแนกความคิดเห็นและทดสอบประสิทธิภาพ การนำเสนอผลการวิจัยแบ่งเป็น 3 ส่วน ได้แก่ ส่วนที่ 1. เครื่องมือและข้อมูลต่าง ๆ ที่ใช้ในการทดลอง ส่วนที่ 2. วิธีการทดลอง ส่วนที่ 3. ผลการประเมินประสิทธิภาพและการอภิปรายผลการทดลอง มีรายละเอียดดังต่อไปนี้

4.1 เครื่องมือและข้อมูลที่ใช้ในการทดลอง

4.1.1 เครื่องมือที่ใช้ในการทดลอง

เครื่องมือและข้อมูลต่าง ๆ ที่ใช้ในการทดลองในงานวิจัย ได้แก่ ด้านฮาร์ดแวร์ ประกอบด้วย เครื่องคอมพิวเตอร์แบบพกพา ASUS ซีพียู AMD Ryzen7 4700U แรม 16.00 GB ด้านซอฟต์แวร์ และภาษาที่ใช้ในการเขียนการทดลอง ได้แก่ ระบบปฏิบัติการ Windows 10 Pro 64-bit พัฒนาซอฟต์แวร์โดยใช้โปรแกรม Matlab

4.1.2 การรวบรวมข้อมูล (Data Collection)

ในการรวบรวมข้อมูลได้แบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อความความคิดเห็นของผู้เข้าพักโรงแรมและคำอธิบายรายละเอียดของโรงแรม ซึ่งโรงแรมที่ได้ทำการเก็บรวบรวมข้อความความคิดเห็นมีระดับคุณภาพตั้งแต่ 2 ดาว ถึง 5 ดาว และเป็นโรงแรมที่อยู่ในประเทศไทยในเขตกรุงเทพมหานคร จำนวน 30 โรงแรม โดยมีตัวอย่างจากเว็บไซต์ www.Tripadvisor.com ดังรูปที่ 18 ถึง

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Font: Tahoma, 11, Bold, Italic, Underline, Text Color, Background Color, Font Face, Font Size, Bold, Italic, Underline, Text Color, Background Color, Font Face, Font Size

Alignment: Wrap Text, Merge & Center, Left, Center, Right, Justify, Indent, Decrease Indent, Increase Indent, Decrease Indent, Increase Indent

Number: General, Percentage, Decimal, Fraction, Text, Scientific, Accounting, Custom Number Format

Styles: Conditional Formatting, Format as Table, Cell Styles

Cells: Insert, Delete, Format, Sum, Fill, Clear, Sort & Filter, Find & Select, Editing

E20

A	B	C	D
Header Link	Social Member	Hotels Review text	Date Of Stay
1	19 contributions; 12 helpful v	My first stay in an Ibis hotel and I am floored by the interior designs. Minimalist room decor and what I re	April 2019
2	5 contributions; 1 helpful v	We are currently staying in this hotel. Our late check in at 3am was smooth and quick. After a long and t	April 2019
3	3 contributions; 1 helpful v	This hotel is situated on one of the busiest shopping streets in Bangkok, but... we never heard a single	March 2019
4	437 contributions; 107 helpf	Having slept in some pretty grim places over the last week, it was such a pleasant experience to park my h	April 2019
5	9 contributions; 1 helpful v	Relatively a new property. Nice location for a happening area. Very supportive staff especially at "Streets"	April 2019
6	sarahvaughn2415	13 contributions; 7 helpful v	We stayed here for a couple nights, while in Bangkok. The rooms and beds were amazing. We had plenty
7	an a	26 contributions; 9 helpful v	cs really nice, I like this hotel, friendly & helpful staffs. Nice clean spacious room, will be back. Its safe
8	dodgeite	55 helpful v	cs 3 nights here...great hotel, right in the middle of all the entertainment...walk out the door and bang e
9	VeeVee007	90 contributions; 85 helpful v	Located in a convenient spot for a short walk to the main attractions this is a fun and comfortable hotel. Fu
10	sasha	14 contributions; 10 helpful v	Located with tremendous service. on a busy road street, but definitely more calm than the nearby khao
11	Paul C	Weymouth, United Kingdom	This modern hotel is a calming sanctuary in the very busy and noisy Khaosan district. The rooms are very
12	EDaya2016	Yorkshire, United Kingdom	This hotel is in a great location situated on the street next to Kho San. Our room was facing Kho San and w
13	Tafari Stevenson-Howar	Detroit, Michigan	ibis Styles Bangkok Khaosan Vientai is not only in a great bustling district its also a very comfortable and s
14	अश्विनी a		My room is located the backside and didn't face for noisy problem. The rainbow pool is beautiful. The brea
15	Grant	Taree, Australia	Currently staying in this great hotel which is situated in a central location. The rooms are newly refurbis
16	sarahvaughn2415	13 contributions; 7 helpful v	We stayed here for a couple nights, while in Bangkok. The rooms and beds were amazing. We had plenty
17	an a	26 contributions; 9 helpful v	cs really nice, I like this hotel, friendly & helpful staffs. Nice clean spacious room, will be back. Its safe
18	dodgeite	55 helpful v	cs 3 nights here...great hotel, right in the middle of all the entertainment...walk out the door and bang e
19	VeeVee007	90 contributions; 85 helpful v	Located in a convenient spot for a short walk to the main attractions this is a fun and comfortable hotel. Fu

Sheet1 Description

รูปที่ 18 ลักษณะของข้อความคิดเห็นที่เก็บไว้ในไฟล์ Excel

The screenshot shows the Microsoft Excel interface. The title bar indicates the file is 'ibis-hotel1-(Crawl-Run)---2019-05-01T073723Z-mix - Excel'. The 'Home' ribbon is active, displaying options for Font, Paragraph, Styles, Cells, and Editing. The spreadsheet area shows a table with columns A through R and rows 1 through 19. The text 'Ibis Styles Bangkok Khaosan Viengtai Description' is entered in cell B2. The text 'Ibis Styles Bangkok Khaosan Viengtai is situated in the action-packed Khaosan area surrounded by all the vibrant life that makes it the must-go-to place in Bangkok....' is entered in cell B3.

รูปที่ 19 ลักษณะของคำอธิบายรายละเอียดของโรงแรมที่เก็บไว้ในไฟล์ Excel

จากรูปที่ 18 คือ การรวบรวมข้อความความคิดเห็นของผู้เข้าพักโรงแรมจากเว็บไซต์ www.tripadvisor.com และในการรวบรวมความคิดเห็นของผู้เข้าพัก ใช้เว็บไซต์ Import.io APIs ซึ่งมี คอลัมน์ ลิงก์ส่วนหัว (Header Link) สมาชิกโซเชียล (Social Member) ข้อความรีวิวโรงแรม (Hotels Review text) และ วันที่เข้าพัก (Date Of Stay) โดยที่ข้อความเหล่านี้เกิดจากบุคคลทั่วไปที่มีประสบการณ์ในการเข้าพักหรือมีความสนใจในโรงแรมนั้น โดยอาจจะแสดงความรู้สึกในเชิงบวก

หรือเชิงลบต่อโรงแรมเหล่านั้น ส่วนรูปที่ 19 คำอธิบายรายละเอียดของโรงแรม Ibis Styles Bangkok Khaosan Viengtai คือ ส่วนคำอธิบายรายละเอียดของโรงแรมจะอธิบายถึงลักษณะเด่น ข้อดี จุดเด่น ข้อได้เปรียบ เช่น ใกล้แหล่งท่องเที่ยว การให้บริการเสริมของโรงแรมนั้น จากการรวบรวมข้อความคิดเห็นได้ความคิดเห็นทั้งหมดจาก 30 โรงแรมดังนี้

ตารางที่ 16 ตารางแสดงจำนวนข้อความคิดเห็นทั้งหมด(ทุกภาษา)

ลำดับ	ชื่อโรงแรม	จำนวนข้อความคิดเห็น
1	Ibis Styles Bangkok Khaosan Viengtai	1,249
2	Adelphi Forty-Nine Hotel	946
3	S Sukhumvit Suites	990
4	Phranakorn_Nornlen_Hotel	2,534
5	Metro Pratunam Boutique Hotel	441
6	CityPoint Hotel	1,392
7	Smart Suites	1,047
8	Citrus Sukhumvit 11 by Compass Hospitality	718
9	Tints of Blue Hotel	587
10	Holiday Inn Express Bangkok Sukhumvit 11	673
11	Bally Suite Sukhumvit	800
12	Majestic Suites	651
13	Citrus Sukhumvit 13 by Compass Hospitality	766
14	Salil Hotel Sukhumvit Soi 8	747
15	The Key Bangkok Hotel by Compass Hospitality	760
16	Citichic Hotel Sukhumvit by Compass Hospitality	543
17	Hotel Clover Asoke	858
18	Sam's Lodge	484
19	Vince Hotel	1,207
20	Citadines Sukhumvit 11 Bangkok	939
21	Paragon Inn	844
22	Solitaire Bangkok	989
23	Siri Sathorn Bangkok by UHG	758
24	Shama Sukhumvit Bangkok	714

25	The Phoenix Hotel Bangkok	747
26	Grand Mercure Bangkok Asoke Residence	783
27	The Okura Prestige Bangkok	714
28	Bangkok Marriott Hotel The Surawongse	1,347
29	Centre Point Hotel Pratunam	1,678
30	Radisson Blu Plaza Bangkok	898

ตารางที่ 17 ตารางแสดงจำนวนข้อความคิดเห็นที่ได้ทำการคัดเลือก (เลือกเฉพาะภาษาอังกฤษ)

ลำดับ	ชื่อโรงแรม	จำนวนข้อความคิดเห็น
1	Ibis Styles Bangkok Khaosan Viengtai	696
2	Adelphi Forty-Nine Hotel	682
3	S Sukhumvit Suites	702
4	Phranakorn_Nornlen_Hotel	715
5	Metro Pratunam Boutique Hotel	294
6	CityPoint Hotel	398
7	Smart Suites	375
8	Citrus Sukhumvit 11 by Compass Hospitality	398
9	Tints of Blue Hotel	309
10	Holiday Inn Express Bangkok Sukhumvit 11	368
11	Bally Suite Sukhumvit	391
12	Majestic Suites	379
13	Citrus Sukhumvit 13 by Compass Hospitality	400
14	Salil Hotel Sukhumvit Soi 8	400
15	The Key Bangkok Hotel by Compass Hospitality	285
16	Citichic Hotel Sukhumvit by Compass Hospitality	285
17	Hotel Clover Asoke	399
18	Sam's Lodge	413
19	Vince Hotel	400
20	Citadines Sukhumvit 11 Bangkok	405
21	Paragon Inn	376

22	Solitaire Bangkok	420
23	Siri Sathorn Bangkok by UHG	400
24	Shama Sukhumvit Bangkok	403
25	The Phoenix Hotel Bangkok	399
26	Grand Mercure Bangkok Asoke Residence	389
27	The Okura Prestige Bangkok	384
28	Bangkok Marriott Hotel The Surawongse	383
29	Centre Point Hotel Pratunam	400
30	Radisson Blu Plaza Bangkok	396

ตารางที่ 18 แสดงตัวอย่างรายชื่อโรงแรมและคำอธิบายคุณลักษณะโรงแรม

ลำดับ ที่	ชื่อโรงแรม	คำอธิบายคุณลักษณะโรงแรม
1	Ibis Styles Bangkok Khaosan Viengtai	Ibis Styles Bangkok Khaosan Viengtai is situated in the heart of the action-packed Khaosan area surrounded by all the vibrant life that makes it the must-go-to place in Bangkok.....
2	Adelphi Forty-Nine Hotel	Adelphi Forty-Nine is ideally located on Sukhumvit Soi 49 in the heart of downtown Bangkok in the upscale district of Bangkok in between the trendy Thonglor and peaceful Phrom Phong where it is considered home to most of the expatriates living in Bangkok
3	S Sukhumvit Suites	S Sukhumvit Suites is a stylish city hotel in the heart of Sukhumvit area, ideally located right on the main Sukhumvit Road, between Sukhumvit 21 (Asoke Intersection) and Sukhumvit 23.....

4.1.3 การระบุผลเฉลยโดยผู้เชี่ยวชาญ

ในการระบุผลเฉลยโดยผู้เชี่ยวชาญจำนวน 5 คน เพื่อระบุว่าแต่ละความคิดเห็นกล่าวถึงคุณลักษณะที่สนใจเป็นเชิงบวกหรือเชิงลบ ผู้วิจัยได้ส่งตารางคำคุณลักษณะและคำที่มีความหมายเหมือนคำคุณลักษณะให้แก่ผู้เชี่ยวชาญด้วย ในการระบุข้อความความเห็นสำหรับแต่ละคำคุณลักษณะผู้เชี่ยวชาญพิจารณาจากนิยามของคำแสดงความรู้สึกในเชิงบวกและเชิงลบ เช่น คำเชิงบวก คือ คำที่แสดงถึงความรู้สึกดี หรือความรู้สึกด้านบวก คำเชิงลบ คือ คำที่แสดงถึงความรู้สึก ไม่ดี หรือความรู้สึกด้านลบ เป็นต้น หลังจากผู้เชี่ยวชาญทั้ง 5 ท่าน ทำการระบุผลเฉลยเรียบร้อยแล้ว ผู้วิจัยได้คัดเลือกเอาเฉพาะข้อความความเห็นที่ผู้เชี่ยวชาญมีความเห็นตรงกันมากกว่า 3 คน การตรวจสอบผลเฉลยจากความคิดเห็นของผู้เชี่ยวชาญแสดงดังตารางที่ 19

ตารางที่ 19 การตรวจสอบผลเฉลยของผู้เชี่ยวชาญ

ลำดับ	ชื่อโรงแรม	จำนวน ข้อความ ความเห็น ภาษาอังกฤษ	จำนวน ความเห็นที่ ถูกต้อง	จำนวน ข้อความ ความเห็นที่ ตรงกัน
1	Ibis Styles Bangkok Khaosan Viengtai	695	1	694
2	Adelphi Forty-Nine Hotel	682	0	682
3	S Sukhumvit Suites	701	1	700
4	Phranakorn Nornlen Hotel	711	4	707
5	Metro Pratunam Boutique Hotel	294	0	294
6	CityPoint Hotel	398	1	397
7	Smart Suites	375	2	373
8	Citrus Sukhumvit 11 by Compass Hospitality	398	5	393
9	Tints of Blue Hotel	309	6	303
10	Holiday Inn Express Bangkok Sukhumvit 11	368	3	365
11	Bally Suite Sukhumvit	391	5	386
12	Majestic Suites	379	3	376
13	Citrus Sukhumvit 13 by Compass Hospitality	400	1	399
14	Salil Hotel Sukhumvit Soi 8	400	5	395
15	The Key Bangkok Hotel by Compass Hospitality	285	0	285
16	Citichic Hotel Sukhumvit by Compass Hospitality	285	0	285
17	Hotel Clover Asoke	399	0	399

18	Sam's Lodge	413	0	413
19	Vince Hotel	400	0	400
20	Citadines Sukhumvit 11 Bangkok	405	0	405
21	Paragon Inn	376	0	376
22	Solitaire Bangkok	420	0	420
23	Siri Sathorn Bangkok by UHG	400	0	400
24	Shama Sukhumvit Bangkok	403	0	403
25	The Phoenix Hotel Bangkok	399	1	398
26	Grand Mercure Bangkok Asoke Residence	389	0	389
27	The Okura Prestige Bangkok	384	0	384
28	Bangkok Marriott Hotel The Surawongse	383	0	383
29	Centre Point Hotel Pratunam	400	0	400
30	Radisson Blu Plaza Bangkok	396	0	396

4.2 ผลการเตรียมข้อมูล

4.2.1 การทำความสะอาดข้อความ (Text Cleaning)

ในขั้นตอนการคัดเลือกข้อความคิดเห็น ได้คัดเลือกเอาเฉพาะข้อความภาษาอังกฤษ จากข้อความคิดเห็นทั้งหมดที่มีอยู่ในแต่ละโรงแรมดังตารางที่ 20

ตารางที่ 20 ตารางแสดงจำนวนข้อความคิดเห็นทั้งหมด (ทุกภาษา)

ลำดับ	ชื่อโรงแรม	ข้อความคิดเห็นทั้งหมด	จำนวนข้อความคิดเห็นภาษาอังกฤษ
1	Ibis Styles Bangkok Khaosan Viengtai	1,249	696
2	Adelphi Forty-Nine Hotel	946	682
3	S Sukhumvit Suites	990	702
4	Phranakorn Nornlen Hotel	2534	715
5	Metro Pratunam Boutique Hotel	441	294
6	CityPoint Hotel	1,392	398
7	Smart Suites	1,047	375
8	Citrus Sukhumvit 11 by Compass Hospitality	718	398
9	Tints of Blue Hotel	587	309
10	Holiday Inn Express Bangkok Sukhumvit 11	673	368
11	Bally Suite Sukhumvit	800	391

12	Majestic Suites	651	379
13	Citrus Sukhumvit 13 by Compass Hospitality	766	400
14	Salil Hotel Sukhumvit Soi 8	747	400
15	The Key Bangkok Hotel by Compass Hospitality	760	285
16	Citichic Hotel Sukhumvit by Compass Hospitality	543	285
17	Hotel Clover Asoke	858	399
18	Sam's Lodge	484	413
19	Vince Hotel	1,207	400
20	Citadines Sukhumvit 11 Bangkok	939	405
21	Paragon Inn	844	376
22	Solitaire Bangkok	989	420
23	Siri Sathorn Bangkok by UHG	758	400
24	Shama Sukhumvit Bangkok	714	403
25	The Phoenix Hotel Bangkok	747	399
26	Grand Mercure Bangkok Asoke Residence	783	389
27	The Okura Prestige Bangkok	714	384
28	Bangkok Marriott Hotel The Surawongse	1,347	383
29	Centre Point Hotel Pratunam	1,678	400
30	Radisson Blu Plaza Bangkok	898	396

4.2.2 การกำกับหน้าที่ของคำ (Path-of-Speech Tagging)

การกำกับหน้าที่ของคำโดยใช้ Penn Treebank Tagset ในการกำกับหน้าที่ของคำ แสดงได้ดังรูปที่ 20

ข้อมูลที่ได้จากขั้นตอนการทำ POS Tagging											
sometimes[RB]	when[WRB]	you[PRP]	stay[VBP]	it[IN]	an[DT]	iconic[JJ]	hotel[NN]	;	it[PRP]	does[VBZ]	not[RB]
we[PRP]	just[RB]	returned[VBD]	home[RB]	after[IN]	a[DT]	six-night[JJ]	stay[NN]	at[IN]	the[DT]	beverly[JJ]	hills[NN]
[JJ]	stayed[VBD]	at[IN]	this[DT]	hotel[NN]	for[IN]	my[PRP\$]	birthday[JJ]	weekend[NN]	and[CC]	it[PRP]	was[VB]
we[PRP]	have[VBP]	stayed[VBN]	at[IN]	the[DT]	hotel[NN]	in[IN]	the[DT]	past[NN]	...	what[WP]	draws[VB]
can[MD]	[VB]	it[JJ]	wait[NN]	to[TO]	go[VB]	back[RB]	...	it[PRP]	[VBD]	s[NNS]	like[IN]
it[PRP]	was[VBD]	our[PRP\$]	first[JJ]	time[NN]	at[IN]	the[DT]	bbh[NN]	and[CC]	we[PRP]	had[VBD]	a[DT]
[JJ]	really[RB]	love[VB]	this[DT]	hotel[NN]	...	it[PRP]	has[VBZ]	a[DT]	beautiful[JJ]	old[JJ]	hollywo
[JJ]	've[VBP]	been[VBN]	waiting[VBG]	to[TO]	have[VB]	brunch[NN]	in[IN]	the[DT]	polo[NN]	lounge[NN]	for[IN]
[JJ]	stayed[NN]	with[IN]	my[PRP\$]	family[NN]	for[IN]	the[DT]	first[JJ]	time[NN]	in[IN]	the[DT]	beverly
[NN]	was[VBD]	lucky[JJ]	enough[RB]	to[TO]	travel[VB]	with[IN]	work[NN]	to[TO]	this[DT]	amazing[JJ]	hotel[N
this[DT]	is[VBZ]	the[DT]	first[JJ]	time[NN]	[NN]	have[VBP]	stayed[VBN]	here[RB]	and[CC]	it[PRP]	was[VB]
the[DT]	was[VBD]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]	the[DT]

รูปที่ 20 ข้อมูลที่ได้จากขั้นตอนการทำ POS Tagging

4.2.3 การกำจัดคำหยุด (Stop Word Removal)

การกำจัดคำหยุด ใช้คลังคำหยุดมาตรฐาน จาก Stanford Stopword List ซึ่งอยู่ในรูปแบบเอกสาร (Text file) จำนวน 257 คำ ข้อมูลที่ได้จากขั้นตอนการกำจัดคำหยุด แสดงดังรูปที่ 21

1

'sometimes'	'when'	'you'	'stay'	'in'	'an'	'iconic'	'hotel'	'	'it'	'does'	'not'
''''	'we'	'just'	'returned'	'home'	'after'	'a'	'six-night'	'stay'	'at'	'the'	'beverly'
''''	'i'	'stayed'	'at'	'this'	'hotel'	'for'	'my'	'birthday'	'weekend'	'and'	'it'
''''	'we'	'have'	'stayed'	'at'	'the'	'hotel'	'in'	'the'	'past'	'	'what'
'can'	''	't'	'wait'	'to'	'go'	'back'	'	'it'	''	's'	'like'
'it'	'was'	'our'	'first'	'time'	'at'	'the'	'bhh'	'and'	'we'	'had'	'a'
'i'	'really'	'love'	'this'	'hotel'	'	'it'	'has'	'a'	'beautiful'	'old'	'hollywood'
'i'	''''ve'	'been'	'waiting'	'to'	'have'	'brunch'	'in'	'the'	'polo'	'lounge'	'for'
''''	'i'	'stayed'	'with'	'my'	'family'	'for'	'the'	'first'	'time'	'in'	'the'
'i'	'was'	'lucky'	'enough'	'to'	'travel'	'with'	'work'	'to'	'this'	'amazing'	'hotel'
'this'	'is'	'the'	'first'	'time'	'i'	'have'	'stayed'	'here'	'and'	'it'	'was'
''''	'this'	'was'	'one'	'of'	'six'	'hotels'	'throughout'	'california'	'that'	'we'	'stayed'
''''	'a+'	'rating'	'for'	'this'	'hotel'	'!	'everything'	'at'	'this'	'hotel'	'from'

2

'sometimes'	[]	[]	'stay'	[]	[]	'iconic'	'hotel'	[]	[]	[]	[]
[]	[]	'just'	'returned'	'home'	[]	[]	'six-night'	'stay'	[]	[]	'beverly'
[]	[]	'stayed'	'at'	[]	'hotel'	[]	'birthday'	'weekend'	[]	[]	[]
[]	[]	[]	'stayed'	[]	[]	'hotel'	[]	[]	'past'	[]	[]
[]	''	't'	'wait'	[]	'go'	'back'	[]	[]	''	's'	'like'
[]	[]	[]	'first'	'time'	[]	[]	'bhh'	[]	[]	[]	[]
[]	'really'	'love'	[]	'hotel'	[]	[]	[]	[]	'beautiful'	'old'	'hollywood'
[]	''''ve'	[]	'waiting'	[]	[]	'brunch'	[]	[]	'polo'	'lounge'	[]
[]	[]	'stayed'	[]	[]	'family'	[]	[]	'first'	'time'	[]	[]
[]	[]	'lucky'	'enough'	[]	'travel'	[]	'work'	[]	[]	'amazing'	'hotel'
[]	[]	[]	'first'	'time'	[]	[]	'stayed'	[]	[]	[]	[]
[]	[]	[]	'one'	[]	'six'	'hotels'	'throughout'	'california'	[]	[]	'stayed'
[]	'a+'	'rating'	[]	[]	'hotel'	[]	'everything'	[]	[]	'hotel'	[]

3

'sometimes'	'stay'	'iconic'	'hotel'	'live'	'reputation'	'however'	'case'	'beverly'	'hills'	'hotel'	'perfection'	'sleek'
'just'	'returned'	'home'	'six-night'	'stay'	'beverly'	'hills'	'hotel'	'stayed'	'hotel'	'several'	'times'	'every'
'stayed'	'hotel'	'birthday'	'weekend'	'amazing'	'quickly'	'staff'	'learned'	'name'	'even'	'though'	'first'	'time'
'stayed'	'hotel'	'past'	'draws'	'us'	'back'	'history'	'beauty'	'ambience'	'switched'	'rooms'	'times'	'casita'
''	't'	'wait'	'go'	'back'	''	's'	'like'	'stepping'	'back'	'time'	'rooms'	'staff'
'first'	'time'	'bhh'	'ground'	'floor'	'room'	'patio'	'3'	'people'	'one'	'sofabed'	'what'	'loved'
'really'	'love'	'hotel'	'beautiful'	'old'	'hollywood'	'charm'	'rooms'	'furnished'	'gorgeous'	'fabrics'	'make'	'feel'
''''ve'	'waiting'	'brunch'	'polo'	'lounge'	'years'	'try'	'famous'	'mccarthy'	'salad'	'well'	'finally'	'visited'
'stayed'	'family'	'first'	'time'	'beverly'	'hills'	'hotel'	'august'	'checked'	'room'	'air-con'	'leaking'	'terrible'
'lucky'	'enough'	'travel'	'work'	'amazing'	'hotel'	'blown'	'away'	'food'	'amazing'	'service'	'insane'	'able'
'first'	'time'	'stayed'	'really'	'fantastic'	'arrival'	'experience'	'rooms'	'food'	'pool'	'area'	'really'	'great'
'one'	'six'	'hotels'	'throughout'	'california'	'stayed'	'3week'	'visit'	'state'	'sadly'	'say'	'worst'	'expected'
'a+'	'rating'	'hotel'	'everything'	'hotel'	'front'	'desk'	'housekeepi...	'valet'	'cafe'	'personnel'	'excellent'	'polite'

รูปที่ 21 ตัวอย่างข้อมูลที่ได้จากขั้นตอนการกำจัดคำหยุด

จากรูปที่ 21 แสดงตัวอย่างข้อมูลที่ได้จากขั้นตอนการกำจัดคำหยุด ส่วนที่ 1 แสดงคำที่ปรากฏในประโยคความคิดเห็น ส่วนที่ 2 แสดงการตัดคำที่เป็นคำหยุดออก และในตารางที่ 3 แสดงประโยคที่ผ่านการตัดคำหยุดออกเรียบร้อยแล้ว

4.2.4 การสกัดคำคุณลักษณะและคำระบุชี้วัดความคิดเห็น (Feature and Polarity Extraction)

ในการสกัดคุณลักษณะ ผู้วิจัยได้นำคำทั้งหมดไปค้นหาคำที่มีความหมายเหมือนคำคุณลักษณะ ถ้าตรงกับคำที่มีความหมายเหมือนคำคุณลักษณะให้แทนด้วยคำคำนั้น ด้วยคำคุณลักษณะที่กำหนด ซึ่งช่วยให้การจัดกลุ่มคำคุณลักษณะง่ายขึ้นและสามารถนำไปวิเคราะห์ได้อย่างมีประสิทธิภาพ ผลลัพธ์ที่ได้แสดงดังรูปที่ 22 และ รูปที่ 23

	1	2	3	4	5	6	7	8	9	10	11	1
1	'second'	'time'	'hotel'	'liked'	'lot'	'center'	'super'	'crowdy'	'khao'	'san'	'just'	'step'
2	'really'	'love'	'hotel'	'like'	'thank'	'much'	'tommy'	'wonderful'	'hospitality'	'helpful'	'nice'	'food'
3	'large'	'hotel'	'road'	'khao'	'san'	'road'	'heart'	'nightlife.the'	'rooms'	'large'	'clean'	'swimr'
4	'family'	'really'	'loves'	'hotel'	'location'	'perfect'	'pool'	'decent'	'rooms'	'spacious'	'clean'	'breakf'
5	'gold'	'member'	'member'	'check'	'children'	'stay'	'free'	'room'	'hotel'	'means'	'share'	'parent'
6	'loved'	'stay'	'hotel'	'right'	'middle'	'lovely'	'bangkok'	'streets'	'massages'	'modern'	'peaceful'	'come'
7	'stayed'	'kids'	'years'	'old'	'asia'	'first'	'time'	'hotel'	'clean'	'family'	'room'	'really'
8	'clean'	'hotel'	'friendly'	'staff'	'close'	'loads'	'amenities'	'sights'	'stay'	'good'	'breakfast'	'swimr'
9	'loved'	'hotel'	'great'	'pool'	'good'	'breakfast'	'best'	'part'	'location'	'street'	'lots'	'restau'
10	'really'	'really'	'good'	'experience'	'hotel'	'behaviour'	'hotel'	'staffs'	'good'	'food'	'good'	'love'
11	'booked'	'hotel'	'night'	'stop'	'way'	'australia'	'husband'	'visited'	'thailand'	'first'	'time'	'yo'
12	'great'	'stay'	'place'	'visit'	'bangkok'	'stayed'	'family'	'room'	'perfect'	'parents'	'grown'	'offspr'
13	'staff'	'helpful'	'polite'	'shared'	'information'	'places'	'interest'	'visit'	'checkout'	'bags'	'kept'	'access'
14	'stayed'	'old'	'viengtai'	'hotel'	'made'	'perfect'	'sense'	'stay'	'newly'	'refurbished'	'ibis.the'	'hotel'
15	'stayed'	'hotel'	'days'	'hotel'	'clean'	'comfortable'	'minor'	'troubles'	'stayed'	'mother'	'rooms'	'card'
16	'stayed'	'end'	'march'	'hotel'	'perfect'	'familv'	'comfortable'	'lono'	'flight'	'thought'	'd'	'booke'

รูปที่ 22 ก่อนการแทนคำที่มีความหมายเหมือนคำคุณลักษณะ

	1	2	3	4	5	6	7	8	9	10	11	1
1	'second'	'time'	'hotel'	'liked'	'lot'	'center'	'super'	'crowdy'	'khao'	'san'	'just'	'step'
2	'really'	'love'	'hotel'	'like'	'thank'	'much'	'tommy'	'wonderful'	'hospitality'	'helpful'	'nice'	'food'
3	'large'	'hotel'	'road'	'khao'	'san'	'road'	'heart'	'nightlife.the'	'rooms'	'large'	'clean'	'swimr'
4	'family'	'really'	'loves'	'hotel'	'location'	'perfect'	'service'	'decent'	'rooms'	'spacious'	'clean'	'food'
5	'gold'	'member'	'member'	'check'	'children'	'stay'	'free'	'rooms'	'hotel'	'means'	'share'	'parent'
6	'loved'	'stay'	'hotel'	'right'	'middle'	'lovely'	'bangkok'	'streets'	'massages'	'modern'	'peaceful'	'come'
7	'stayed'	'kids'	'years'	'old'	'asia'	'first'	'time'	'hotel'	'clean'	'family'	'rooms'	'really'
8	'clean'	'hotel'	'friendly'	'service'	'close'	'loads'	'amenities'	'sights'	'stay'	'good'	'food'	'swimr'
9	'loved'	'hotel'	'great'	'service'	'good'	'food'	'best'	'part'	'location'	'location'	'lots'	'restau'
10	'really'	'really'	'good'	'experience'	'hotel'	'behaviour'	'hotel'	'staffs'	'good'	'food'	'good'	'love'
11	'booked'	'hotel'	'night'	'stop'	'way'	'australia'	'husband'	'visited'	'thailand'	'first'	'time'	'yo'
12	'great'	'location'	'stay'	'visit'	'bangkok'	'stayed'	'family'	'rooms'	'perfect'	'parents'	'grown'	'offspr'
13	'service'	'helpful'	'polite'	'shared'	'information'	'places'	'interest'	'visit'	'checkout'	'bags'	'kept'	'access'
14	'stayed'	'old'	'viengtai'	'hotel'	'made'	'perfect'	'sense'	'stay'	'newly'	'refurbished'	'ibis.the'	'hotel'
15	'stayed'	'hotel'	'days'	'hotel'	'clean'	'comfortable'	'minor'	'troubles'	'stayed'	'mother'	'rooms'	'card'
16	'stayed'	'end'	'march'	'hotel'	'perfect'	'familv'	'comfortable'	'lono'	'flight'	'thought'	'd'	'booke'

รูปที่ 23 การแทนคำที่มีความหมายเหมือนคำคุณลักษณะ

หลังจากทำการแทนคำที่มีความหมายเหมือนคำคุณลักษณะ พบว่าในแต่ละชุดข้อความความคิดเห็นได้กล่าวถึงแต่ละคุณลักษณะโดยใช้คำเหมือนมากกว่าการกล่าวถึงคุณลักษณะนั้นโดยตรง ซึ่งแสดงตัวอย่างได้ดังตารางที่ 21

ตารางที่ 21 ตัวอย่างจำนวนคำคุณลักษณะที่เป็นคำหลักและคำเหมือนที่ปรากฏในข้อความคิดเห็น

	F.1		F.2		F.3		F.4		F.5		Sum.
	Main	Synonym	Main	Synonym	Main	Synonym	Main	Synonym	Main	Synonym	
H.1	347	785	102	847	161	2,739	360	1,126	116	219	6,802
H.2	196	770	168	849	145	590	199	1,041	65	130	4,153
H.3	450	1,097	140	588	205	493	294	4,014	117	161	7,559
H.4	138	903	82	543	216	573	243	400	22	78	3,198

อธิบายตารางที่ 21 โดยที่ H.1 คือ โรงแรมที่ 1, H.2 คือ โรงแรมที่ 2, H.3 คือ โรงแรมที่ 3 และ H.4 คือ โรงแรมที่ 4 ส่วน F.1 คือ Feature Location, F.2 คือ Feature Service, F.3 คือ Feature Food, F.4 คือ Feature Rooms และ F.5 คือ Feature Price

ในข้อมูล H.1 มีจำนวนคำคุณลักษณะที่ปรากฏอยู่จำนวน 6,802 คำ ซึ่งประกอบไปด้วย คำคุณลักษณะที่มากที่สุด คือ คุณลักษณะ F.4 จำนวน 360 คำ คำคุณลักษณะที่น้อยที่สุด คือ คุณลักษณะ F.2 จำนวน 102 คำ คำที่มีความหมายเหมือนคำคุณลักษณะมากที่สุดคือ F.3 จำนวน 2,739 คำ คำที่มีความหมายเหมือนคำคุณลักษณะน้อยที่สุดคือ F.5 จำนวน 219 คำ

ในข้อมูล H.2 มีจำนวนคำคุณลักษณะที่ปรากฏอยู่จำนวน 4,153 คำ ซึ่งประกอบไปด้วย คำคุณลักษณะที่มากที่สุด คือ คุณลักษณะ F.4 จำนวน 199 คำ คำคุณลักษณะที่น้อยที่สุด คือ คุณลักษณะ F.5 จำนวน 65 คำ คำที่มีความหมายเหมือนคำคุณลักษณะมากที่สุดคือ F.4 จำนวน 1,041 คำ คำที่มีความหมายเหมือนคำคุณลักษณะน้อยที่สุดคือ F.5 จำนวน 130 คำ

ในข้อมูล H.3 มีจำนวนคำคุณลักษณะที่ปรากฏอยู่จำนวน 7,559 คำ ซึ่งประกอบไปด้วย คำคุณลักษณะที่มากที่สุด คือ คุณลักษณะ F.1 จำนวน 450 คำ คำคุณลักษณะที่น้อยที่สุด คือ คุณลักษณะ F.5 จำนวน 117 คำ คำที่มีความหมายเหมือนคำคุณลักษณะมากที่สุดคือ F.4 จำนวน 4,014 คำ คำที่มีความหมายเหมือนคำคุณลักษณะน้อยที่สุดคือ F.5 จำนวน 161 คำ

ในข้อมูล H.4 มีจำนวนคำคุณลักษณะที่ปรากฏอยู่จำนวน 3,198 คำ ซึ่งประกอบไปด้วย คำคุณลักษณะที่มากที่สุด คือ คุณลักษณะ F.4 จำนวน 243 คำ คำคุณลักษณะที่น้อยที่สุด คือ คุณลักษณะ F.5 จำนวน 22 คำ คำที่มีความหมายเหมือนคำคุณลักษณะมากที่สุดคือ F.1 จำนวน 903 คำ คำที่มีความหมายเหมือนคำคุณลักษณะน้อยที่สุดคือ F.5 จำนวน 78 คำ

4.2.5 การแทนค่าคำระบุข้อความคิดเห็น

งานวิจัยนี้ใช้คลังคำ SentiWordNet ในการแทนค่าคำระบุข้อความคิดเห็น ในข้อความคิดเห็น โดยเลือกใช้เฉพาะคำที่ถูกกำกับหน้าที่เป็น JJ, JJR, JJS, VB, VBD, VBG, VBN, VBP, VBZ, RB, RBR และ RBS จากนั้นทำการวิเคราะห์ค่าของคำระบุข้อความคิดเห็น โดยพิจารณาจากค่าคะแนนที่ปรากฏใน SentiWordNet โดยการเปรียบเทียบปริมาณการแสดงความรู้สึกในแต่ละด้าน

10 เครื่องหมาย และมีค่าเชิงลบโดยแทนด้วยเครื่องหมาย - อยู่ทางด้านหน้า ค่าคุณลักษณะ rooms จำนวน 1 เครื่องหมาย จากนั้นเปรียบเทียบปริมาณการแสดงความรู้สึก พิจารณาแล้วพบว่าค่าเชิงบวก + มากกว่าค่าเชิงลบ - แสดงว่า คำศัพท์นั้นเป็นค่าเชิงบวก +

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	9 'rooms'			4									
2	8 'food'		0 'food'			3							
3	2 'rooms'		3 'service'		1 'rooms'		0 'food'		4 'rooms'		1 'rooms'		11
4	1 'location'		1 'service'		1 'rooms'		1 'food'		13 'food'		0 'service'		8
5	2 'rooms'		4 'rooms'		14 'food'		0 'food'		1 'location'			3	
6	8 'service'			1									
7	4 'rooms'		11 'location'		1 'service'			1					
8	2 'service'		4 'food'		1 'service'		3 'location'			2			
9	2 'service'		1 'food'		1 'location'		0 'location'			1			
10	6 'food'			4									
11	8 'service'		4 'rooms'		1 'location'		2 'location'			3			
12	1 'location'		2 'rooms'		2 'food'		1 'service'			3			
13	0 'service'		6 'service'		0 'rooms'		0 'location'		0 'location'		2 'food'		0
14	7 'food'		1 'rooms'		3 'rooms'		0 'rooms'				1 'rooms'		0
15	3 'rooms'		3 'service'		5 'rooms'		0 'rooms'		0 'rooms'		8 'service'		-1
16	7 'rooms'		-1 'rooms'		1 'service'		2 'food'		0 'location'		4 'rooms'		
17	1 'location'		0 'location'		1 'location'		0 'food'		1 'price'		2 'location'		1
18	0 'rooms'		0 'rooms'		3 'service'		0 'food'		2 'location'		3 'location'		0

รูปที่ 26 ค่าระบุข้อความคิดเห็นที่ปรากฏอยู่ข้างหน้าและข้างหลังคำคุณลักษณะ

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	9 'rooms'												
2			8 'food'										
3	0 'rooms'		0 'food'		3 'service'		2 'price'		16 'location'				
4	5 'rooms'		14 'food'		4 'service'				39 'location'				
5	4 'rooms'		14 'food'						1 'location'				
6					8 'service'								
7	4 'rooms'				1 'service'				11 'location'				
8			4 'food'		3 'service'				3 'location'				
9			1 'food'		2 'service'				1 'location'				
10			6 'food'										
11	4 'rooms'				8 'service'				3 'location'				
12	2 'rooms'		2 'food'		1 'service'				1 'location'				
13	0 'rooms'		2 'food'		6 'service'				0 'location'				
14	5 'rooms'		7 'food'		6 'service'				4 'location'				
15	-1 'rooms'				17 'service'				6 'location'				
16	4 'rooms'		2 'food'		1 'service'				0 'location'				
17			0 'food'		1 'service'		1 'price'		4 'location'				
18	0 'rooms'		0 'food'		3 'service'		0 'price'		5 'location'				

รูปที่ 27 การจัดค่าระบุข้อความคิดเห็นที่ปรากฏอยู่หลังคำคุณลักษณะ

ในการคัดเลือกกว่าค่าระบุข้อความคิดเห็นแต่ละคำเป็นของคำคุณลักษณะใด จะใช้การคัดเลือกโดยถ้าค่าระบุข้อความคิดเห็นใดปรากฏอยู่ด้านหน้าของคำคุณลักษณะใด จะถือว่าค่าระบุข้อความคิดเห็นเป็นของคำคุณลักษณะนั้นและกำหนดสัญลักษณ์แทนค่าเชิงบวกด้วยเครื่องหมายบวก (+) และกำหนดสัญลักษณ์แทนค่าเชิงลบด้วยเครื่องหมายลบ (-) รูปที่ 28

1	rooms+'				
2	food+'				
3	rooms+'	food+'	service+'	price+'	location+'
4	rooms+'	food+'	service+'	location+'	
5	rooms+'	food+'	location+'		
6	service+'				
7	rooms+'	service+'	location+'		
8	food+'	service+'	location+'		
9	food+'	service+'	location+'		
10	food+'				
11	rooms+'	service+'	location+'		
12	rooms+'	food+'	service+'		location+'
13	rooms+'	food+'	service+'		location+'
14	rooms+'	food+'	service+'		location+'
15	rooms-'	service+'	location+'		
16	rooms+'	food+'	service+'	location+'	
17	food+'	service+'	price+'	location+'	
18	rooms+'	food+'	service+'	price+'	location+'

รูปที่ 28 ผลลัพธ์จากการวิเคราะห์ข้อความคิดเห็น

จากรูปที่ 28 ในบรรทัดที่ 1 คำคุณลักษณะ rooms+ เมื่อวิเคราะห์ข้อความคิดเห็นเรียบร้อยแล้ว พบว่าเป็นเชิงบวก + หากดูจากรูปที่ 26 จะพบว่า คำคุณลักษณะ Room มีค่าตัวเลข 9 ระบุอยู่ข้างหน้า และมีค่าตัวเลข 4 ระบุอยู่ทางด้านหลัง มีวิธีการคัดเลือก คือ ถ้าคำระบุข้อความคิดเห็นใดปรากฏอยู่ด้านหน้า [44] ของคำคุณลักษณะใด จะถือว่าคำระบุข้อความคิดเห็นเป็นของคำคุณลักษณะนั้นๆ

4.4 ผลการวัดประสิทธิภาพของการวิเคราะห์ข้อความคิดเห็นของผู้เข้าพักโรงแรม

หลังจากวิเคราะห์ข้อความคิดเห็นของผู้เข้าพักที่มีต่อแต่ละคุณลักษณะ ผู้วิจัยได้นำผลลัพธ์สุดท้ายที่ได้จากการทดลองไปเทียบกับผลเฉลี่ยของผู้เชี่ยวชาญตามลำดับข้อมูล การวัดประสิทธิภาพในงานวิจัยนี้ ใช้วิธีการวัดค่า Precision ค่า Recall และค่า F-measure ซึ่งมีผลลัพธ์ที่ได้จากการทดลองดังตารางที่ 22 และ Error! Reference source not found. - รูปที่ 31

ตารางที่ 22 การวัดประสิทธิภาพค่า Precision Recall F-measure

ชื่อ โรงแรม	จำนวน ความ คิดเห็น ทั้งหมด	จำนวนที่ ทำนายถูก	จำนวนที่ ทำนาย ทั้งหมด	จำนวนผล เฉลี่ย ทั้งหมด	ค่าความ แม่นยำ (Precision)	ค่าความ ระลึก (Recall)	ค่าประสิทธิภาพ โดยรวม (F-measure)
H.1	694	1,613	2,248	2,186	0.69	0.82	0.75
H.2	682	1,749	2,001	2,136	0.85	0.81	0.83
H.3	700	788	1,442	2,006	0.82	0.81	0.81
H.4	707	1,477	1,613	1,847	0.74	0.79	0.76
H.5	294	524	686	821	0.76	0.63	0.69

H.6	397	470	663	795	0.70	0.59	0.64
H.7	373	397	600	687	0.64	0.56	0.6
H.8	393	792	1,200	1,094	0.66	0.72	0.69
H.9	303	713	1,023	927	0.69	0.76	0.72
H.10	365	597	862	920	0.69	0.64	0.66
H.11	386	522	1,030	753	0.50	0.69	0.58
H.12	376	583	906	944	0.64	0.61	0.62
H.13	399	593	1,006	1,015	0.58	0.58	0.58
H.14	395	533	861	1,031	0.62	0.51	0.56
H.15	285	347	581	730	0.59	0.47	0.52
H.16	285	348	581	675	0.59	0.51	0.55
H.17	399	535	775	820	0.69	0.65	0.67
H.18	413	479	932	1,318	0.51	0.36	0.42
H.19	400	575	732	930	0.78	0.61	0.68
H.20	405	415	1,046	843	0.39	0.49	0.43
H.21	376	808	1,230	1,559	0.65	0.51	0.57
H.22	420	350	680	1,000	0.51	0.35	0.42
H.23	400	317	1,003	686	0.31	0.46	0.37
H.24	403	329	1,009	700	0.32	0.47	0.38
H.25	398	222	876	736	0.25	0.30	0.27
H.26	389	419	1,054	798	0.39	0.52	0.45
H.27	384	376	1,035	778	0.36	0.48	0.41
H.28	383	451	907	820	0.49	0.55	0.52
H.29	400	328	1,104	780	0.29	0.42	0.34
H.30	396	417	965	848	0.43	0.49	0.46
ค่าเฉลี่ย	420	602.23	1,021.7	1039.43	0.57	0.57	0.56

อธิบายตารางที่ 22 โดยที่ H.1 คือ โรงแรมที่ 1 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 694 ข้อความ จำนวนที่ทำนายถูกต้อง 1,613 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 2,248 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 2,186 ครั้ง ค่าความแม่นยำ 69 % ค่าความระลึก 82 % และค่าประสิทธิภาพโดยรวม 75 %

H.2 คือ โรงแรมที่ 2 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 682 ข้อความ จำนวนที่ทำนายถูกต้อง 1,749 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 2,001 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 2,136 ครั้ง ค่าความแม่นยำ 85 % ค่าความระลึก 81 % และค่าประสิทธิภาพโดยรวม 83 %

H.14 คือ โรงแรมที่ 14 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 395 ข้อความ จำนวนที่
ทำนายถูกต้อง 533 ครั้ง จำนวนที่ทำกรทำนายทั้งหมด 861 ครั้ง จำนวนผลเฉลยทั้งหมด 1,031
ครั้ง ค่าความแม่นยำ 62 % ค่าความระลึก 51 % และค่าประสิทธิภาพโดยรวม 56 %

H.26 คือ โรงแรมที่ 26 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 389 ข้อความ จำนวนที่
ทำนายถูกต้อง 419 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 1,054 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 798
ครั้ง ค่าความแม่นยำ 35 % ค่าความระลึกลับ 52 % และค่าประสิทธิภาพโดยรวม 45 %

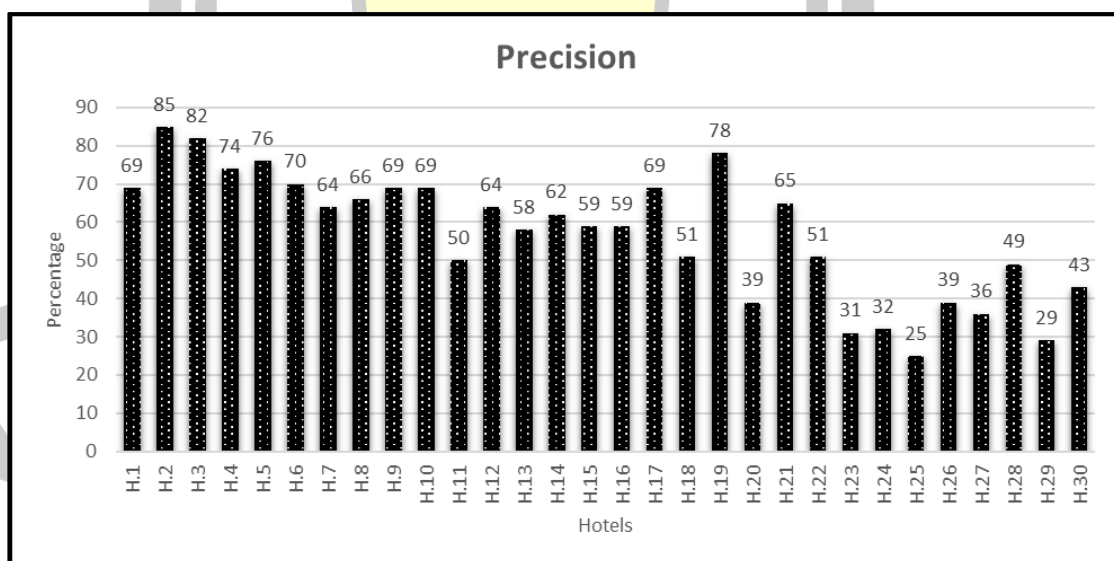
H.27 คือ โรงแรมที่ 27 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 384 ข้อความ จำนวนที่
ทำนายถูกต้อง 376 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 1,035 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 778
ครั้ง ค่าความแม่นยำ 36 % ค่าความระลึกลับ 48 % และค่าประสิทธิภาพโดยรวม 41 %

H.28 คือ โรงแรมที่ 28 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 383 ข้อความ จำนวนที่
ทำนายถูกต้อง 451 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 907 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 820 ครั้ง
ค่าความแม่นยำ 49 % ค่าความระลึกลับ 55 % และค่าประสิทธิภาพโดยรวม 52 %

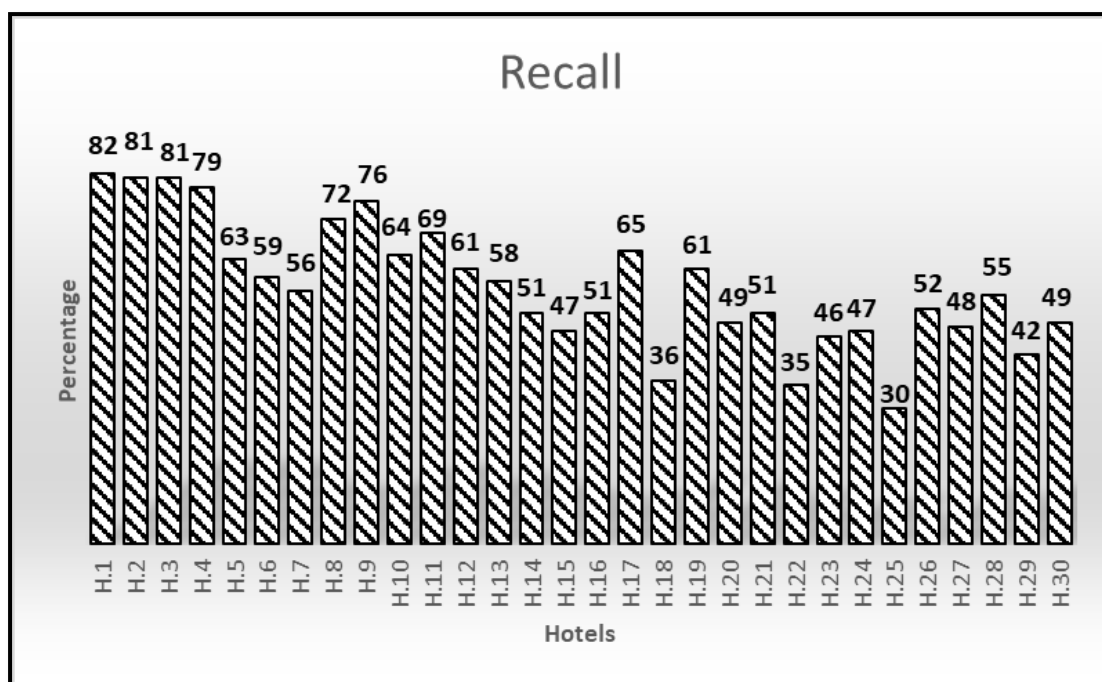
H.29 คือ โรงแรมที่ 29 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 400 ข้อความ จำนวนที่
ทำนายถูกต้อง 328 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 1,104 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 780
ครั้ง ค่าความแม่นยำ 29 % ค่าความระลึกลับ 42 % และค่าประสิทธิภาพโดยรวม 34 %

H.30 คือ โรงแรมที่ 30 มีจำนวนความคิดเห็นที่นำมาวิเคราะห์ 396 ข้อความ จำนวนที่
ทำนายถูกต้อง 417 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 965 ครั้ง จำนวนผลเฉลี่ยทั้งหมด 848 ครั้ง
ค่าความแม่นยำ 43 % ค่าความระลึกลับ 49 % และค่าประสิทธิภาพโดยรวม 46 %

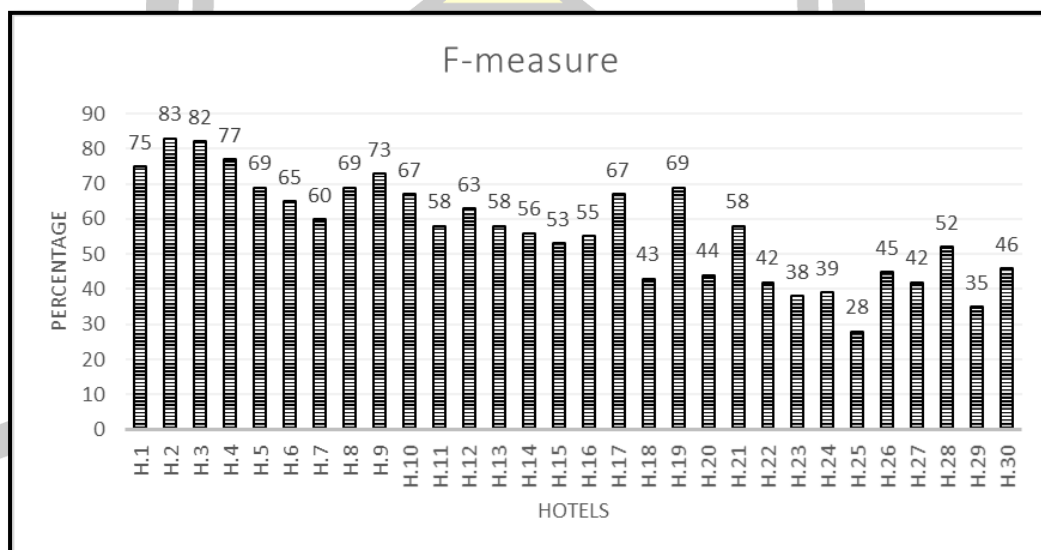
ค่าเฉลี่ยรวมของทั้ง 30 โรงแรม คือ จำนวนความคิดเห็นที่นำมาวิเคราะห์ 420 ข้อความ
จำนวนที่ทำนายถูกต้อง 602 ครั้ง จำนวนที่ทำการทำนายทั้งหมด 1,021 ครั้ง จำนวนผลเฉลี่ยทั้งหมด
1,039 ครั้ง ค่าความแม่นยำ 57 % ค่าความระลึกลับ 57 % และค่าประสิทธิภาพโดยรวม 56 %



รูปที่ 29 การวัดประสิทธิภาพค่า Precision

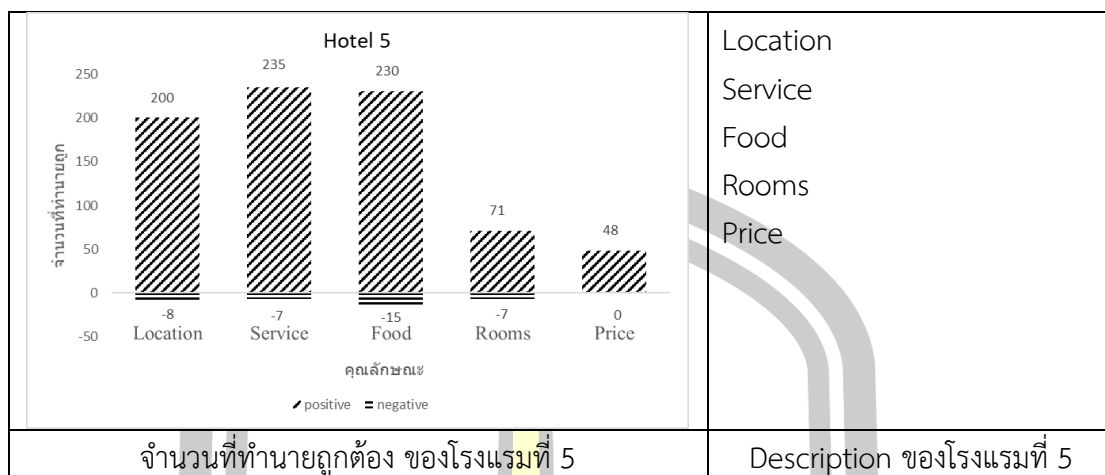


รูปที่ 30 การวัดประสิทธิภาพค่า Recall



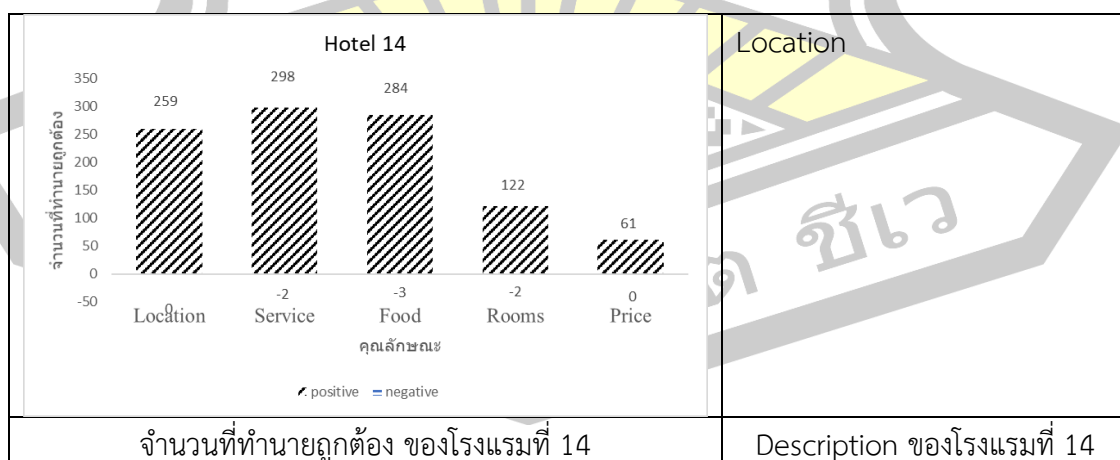
รูปที่ 31 การวัดประสิทธิภาพค่า F-measure

ผู้วิจัยได้ทำการเปรียบเทียบว่าข้อความคิดเห็นของผู้เข้าพักโรงแรม มีความสอดคล้องกับคำอธิบายของโรงแรมหรือไม่ โดยใช้กราฟแสดงผลเปรียบเทียบระหว่างความคิดเห็นของผู้เข้าพัก (Hotel Reviews) และคุณลักษณะของแต่ละโรงแรมที่ได้จากผลเฉลยของคำอธิบายโรงแรม (Hotel Description) ดังแสดงในรูปที่ 32 ถึง รูปที่ 37



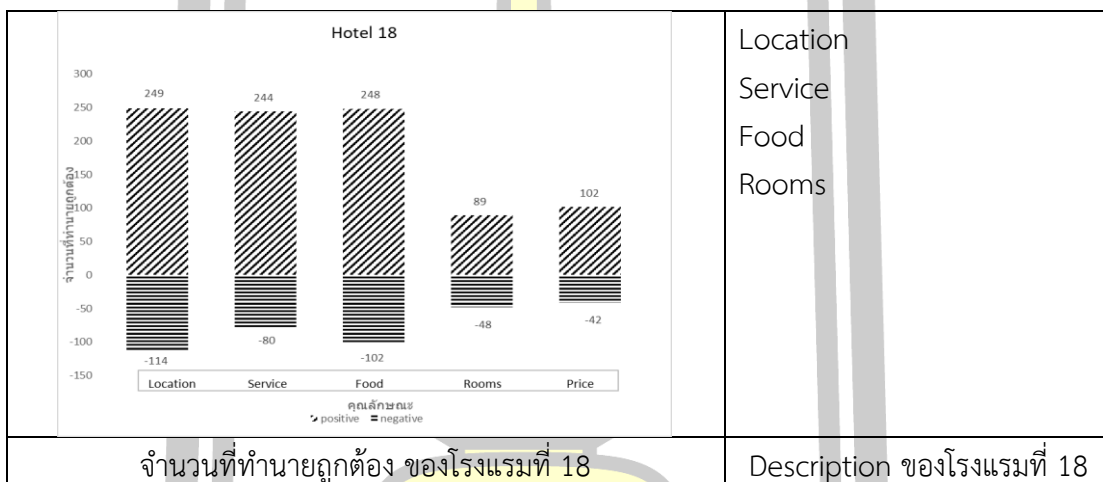
รูปที่ 32 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 5

จากรูปที่ 32 พบว่าผู้เข้าพักมีความคิดเห็นเชิงบวกในด้าน Service, Food, Location, Room และ Price ตามลำดับ โรงแรมได้เขียนคำอธิบายจุดเด่นไว้ซึ่งได้สกัดออกเป็น 5 ด้าน คือ Location, Service, Food, Rooms และ Price อธิบายผลการเปรียบเทียบระหว่างการทํานายถูกต้องกับคำอธิบายโรงแรมได้ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักส่วนใหญ่แสดงความคิดเห็นเป็นเชิงบวกจำนวนมากต่อ 3 คุณลักษณะได้แก่ Service Food Location และผู้เข้าพักได้แสดงความคิดเห็นเป็นเชิงบวกจำนวนน้อยต่อคุณลักษณะด้าน Rooms และ Price ดังนั้น เจ้าของธุรกิจควรปรับปรุงหรือสร้างกลยุทธ์ใหม่เพื่อให้คุณลักษณะ Rooms และ Price เป็นคุณลักษณะที่มีจุดเด่นเทียบเท่าคุณลักษณะอื่นๆ นอกจากนี้ทั้ง 4 คุณลักษณะได้แก่ Service Food Location และ Room มีผู้แสดงความคิดเห็นในเชิงลบ เจ้าของธุรกิจควรปรับปรุงการให้บริการทั้ง 4 ด้านนี้ เพื่อให้ตรงตามความต้องการของผู้เข้าพักและสร้างความจงรักภักดีต่อการให้บริการของโรงแรม



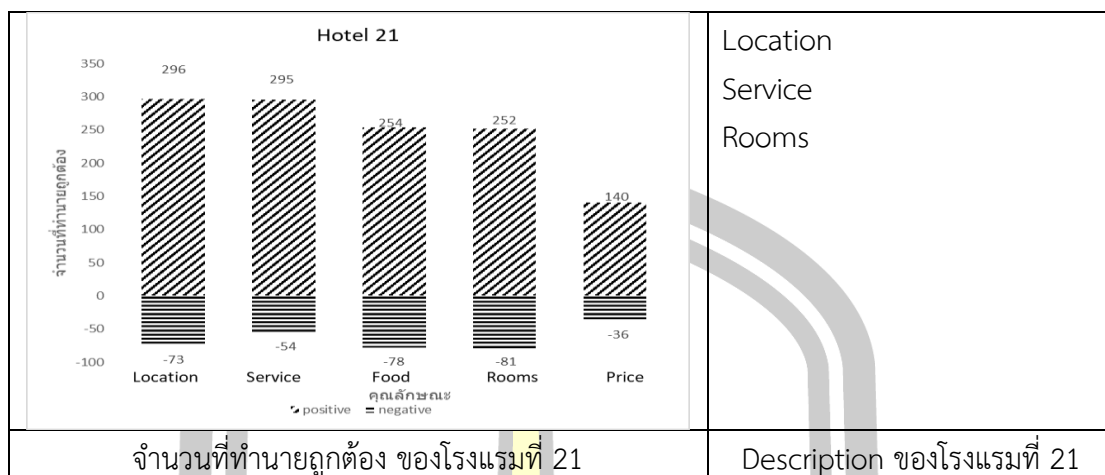
รูปที่ 33 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 14

จากรูปที่ 33 อธิบายผลได้ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักได้แสดงความคิดเห็นเป็นเชิงบวกต่อคุณลักษณะ Service, Food, Location, Room และ Price ตามลำดับ ส่วนโรงแรมได้เขียนคำอธิบายจุดเด่นของโรงแรม 1 ด้าน คือ Location ดังนั้น ในการเขียนคำอธิบายโรงแรมควรเขียนให้ครอบคลุมกับคุณลักษณะเด่นจากความคิดเห็นของผู้เข้าพัก เนื่องจากผู้เข้าพักบางคนอาจต้องการได้รับบริการด้านอื่นๆ ของโรงแรม ซึ่งโรงแรมนี้ไม่ได้กล่าวถึง อาจทำให้ไม่ได้รับความสนใจจากผู้เข้าพัก เจ้าของธุรกิจควรเขียนคำอธิบายโรงแรมเพื่อโฆษณาประชาสัมพันธ์ทั้ง 5 ด้านของคุณลักษณะ โดยเน้นคุณลักษณะด้าน Service ซึ่งเป็นจุดเด่นที่มีผู้เข้าพักแสดงความคิดเห็นเป็นเชิงบวกมากที่สุด เพื่อให้เป็นที่สนใจของผู้เข้าพักที่กำลังมองหาคุณลักษณะที่ตรงตามความต้องการตัดสินใจเข้าพัก



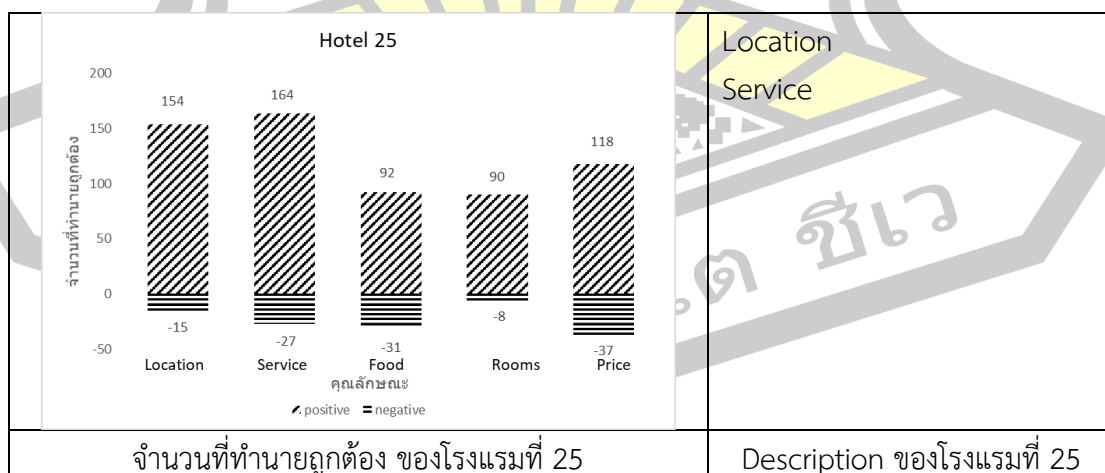
รูปที่ 34 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 18

จากรูปที่ 34 อธิบายผลการเปรียบเทียบระหว่างการทำนายถูกต้องกับคำอธิบายโรงแรมได้ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักส่วนใหญ่แสดงความคิดเห็นเป็นเชิงบวกต่อ 5 คุณลักษณะได้แก่ Location Food Service Price และ Room ตามลำดับ ส่วนโรงแรมได้เขียนคำอธิบายจุดเด่นของโรงแรม 4 ด้าน คือ Location Service Food และ Rooms แต่ยังมีผู้เข้าพักได้แสดงความคิดเห็นเป็นเชิงลบเป็นจำนวนมากต่อทั้ง 5 ด้านของคุณลักษณะ ซึ่งแสดงให้เห็นว่าเจ้าของธุรกิจควรเอาข้อคิดเห็นเชิงลบไปปรับปรุงในแต่ละด้านของคุณลักษณะโดยเฉพาะด้าน Location Food และ Service ที่มีผู้แสดงความคิดเห็นเป็นเชิงลบจำนวนมาก เมื่อปรับปรุงแล้วเจ้าของธุรกิจจึงควรเขียนคำอธิบายโรงแรมเพื่อโฆษณาประชาสัมพันธ์ทั้ง 5 ด้านของคุณลักษณะ เพื่อให้เป็นที่สนใจของผู้ที่กำลังจะเข้าพักที่กำลังมองหาคุณลักษณะที่ตรงตามความต้องการ



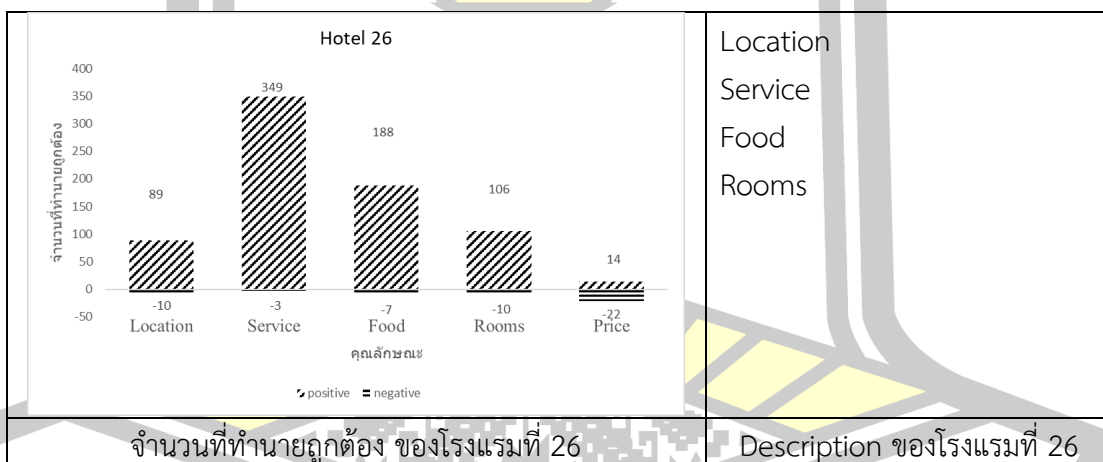
รูปที่ 35 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 21

จากรูปที่ 35 อธิบายผลการเปรียบเทียบระหว่างการทำนายถูกต้องกับคำอธิบายโรงแรมได้ ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักส่วนใหญ่แสดงความคิดเห็นเป็นเชิงบวกต่อ 5 คุณลักษณะ ได้แก่ Location Service Room Food และ Price ตามลำดับ ส่วนโรงแรมได้เขียนคำอธิบายจุดเด่นของ โรงแรม 3 ด้าน คือ Location Service และ Rooms แต่ยังมีผู้เข้าพักจำนวนมากเช่นเดียวกัน ได้แสดงความคิดเห็นเป็นเชิงลบต่อทั้ง 5 ด้านของคุณลักษณะ เรียงตามลำดับได้ดังนี้ Room Food Location Service และ Price ซึ่งแสดงให้เห็นว่า ผู้เข้าพักโรงแรม มีความคิดเห็นเป็นเชิงบวกและความคิดเห็นเป็นเชิงลบทั้ง 5 ด้าน ส่วนคำอธิบายโรงแรม ได้กล่าวถึงจุดเด่นไว้เพียง 3 ด้านเท่านั้น ดังนั้นเจ้าของธุรกิจควรเอาข้อคิดเห็นเชิงลบไปใช้ในปรับปรุง เมื่อปรับปรุงแล้วเจ้าของธุรกิจจึงควรเขียนคำอธิบายโรงแรมเพื่อโฆษณาประชาสัมพันธ์ทั้ง 5 ด้านของคุณลักษณะ ตามที่ผู้เข้าพักได้กล่าวถึงในส่วนที่เป็นเชิงบวก เพื่อทำให้เป็นที่สนใจ และตัดสินใจเลือกเข้าพัก ของผู้ที่กำลังจะเข้าพักที่กำลังมองหาคุณลักษณะที่ตรงตามความต้องการ



รูปที่ 36 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 25

จากรูปที่ 36 พบว่าผู้เข้าพักมีความคิดเห็นเชิงบวกในด้าน Service, Location, Food, Price และ Rooms ตามลำดับ ในส่วนของโรงแรมได้เขียนคำอธิบายจุดเด่นของโรงแรม 2 ด้าน คือ Location และ Service อธิบายผลการเปรียบเทียบระหว่างการทำนายถูกต้องกับคำอธิบายโรงแรม ได้ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักส่วนใหญ่แสดงความคิดเห็นเป็นเชิงบวกต่อ 5 คุณลักษณะ ได้แก่ Service Location Price Food และ Room ตามลำดับ ซึ่ง 2 อันดับแรก ตรงกับคำอธิบายโรงแรม แต่ยังมีผู้เข้าพักจำนวนมากเช่นเดียวกัน ได้แสดงความคิดเห็นเป็นเชิงลบต่อทั้ง 5 ด้านของคุณลักษณะ เรียงตามลำดับได้ดังนี้ Price Room Food Service และ Location ซึ่งแสดงให้เห็นว่าผู้เข้าพักโรงแรม มีทั้งความคิดเห็นเป็นเชิงบวกและความคิดเห็นเป็นเชิงลบทั้ง 5 ด้าน ส่วนคำอธิบายโรงแรม ได้กล่าวถึงจุดเด่นไว้เพียง 2 ด้านเท่านั้น ดังนั้นเจ้าของธุรกิจควรเอาข้อคิดเห็นเชิงลบไปใช้ในการปรับปรุงจุดบกพร่องของคุณลักษณะด้านต่างๆ เมื่อปรับปรุงแล้วเจ้าของธุรกิจจึงควรเขียนคำอธิบายโรงแรมเพื่อโฆษณาประชาสัมพันธ์ทั้ง 5 ด้านของคุณลักษณะ โดยเน้นคุณลักษณะ Service และ Location ตามที่ผู้เข้าพักได้กล่าวถึงในส่วนที่เป็นเชิงบวก เพื่อทำให้เป็นที่สนใจ และตัดสินใจเลือกเข้าพัก ของผู้ที่กำลังจะเข้าพักที่กำลังมองหาคุณลักษณะ Service และ Location ที่ตรงตามความต้องการ และเขียนคำอธิบายโรงแรมเพิ่มเติมในส่วนของคุณลักษณะด้าน Price ซึ่งเป็นคุณลักษณะเด่นที่ผู้เข้าพักได้แสดงความคิดเห็นเป็นเชิงบวก จำนวนมากเป็นอันดับที่ 3 ซึ่งหากผู้เข้าพักที่ต้องการ



รูปที่ 37 แสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรมที่ 26

จากรูปที่ 37 อธิบายผลการเปรียบเทียบระหว่างการทำนายถูกต้องกับคำอธิบายโรงแรมได้ดังนี้ จากรูปด้านซ้ายพบว่าผู้เข้าพักส่วนใหญ่แสดงความคิดเห็นเป็นเชิงบวกต่อ 5 คุณลักษณะ ได้แก่ Service Food Room Location และ Price ตามลำดับ ในส่วนของโรงแรมได้เขียนคำอธิบายจุดเด่นของโรงแรม 4 ด้าน คือ Location Service Food และ Rooms แต่ยังมีผู้เข้าพักอีกจำนวนหนึ่ง ได้แสดงความคิดเห็นเป็นเชิงลบต่อทั้ง 5 ด้านของคุณลักษณะ เรียงตามลำดับได้ดังนี้ Price Room Location Food และ Service ซึ่งแสดงให้เห็นว่า ผู้เข้าพักโรงแรม มีทั้งความคิดเห็นเป็นเชิงบวกและ

ความคิดเห็นเป็นเชิงลบทั้ง 5 ด้าน ส่วนคำอธิบายโรงแรม ได้กล่าวถึงจุดเด่นไว้เพียง 4 ด้าน ดังนั้น เจ้าของธุรกิจควรเอาข้อคิดเห็นเชิงลบไปใช้ในปรับปรุง เมื่อปรับปรุงแล้วเจ้าของธุรกิจจึงควรเขียน คำอธิบายโรงแรมเพื่อโฆษณาประชาสัมพันธ์ทั้ง 5 ด้านของคุณลักษณะ ตามที่ผู้เข้าพักได้กล่าวถึงในส่วนที่เป็นเชิงบวก อาจเน้นในส่วนที่ผู้เข้าพักมีความคิดเห็นเชิงบวกเป็นอันดับที่ 1 คือ ด้าน Service เพื่อให้เป็นที่สนใจ และตัดสินใจเลือกเข้าพัก ของผู้ที่กำลังจะเข้าพักที่กำลังมองหาความต้องการ คุณลักษณะด้าน Service

จากผลการเปรียบเทียบระหว่างผลการทำนายที่ถูกต้องกับคำอธิบายของโรงแรมพบว่า (โรงแรมที่มีความคิดเห็นเป็นเชิงลบ) โรงแรมที่ 18 และ โรงแรมที่ 21 เจ้าของธุรกิจควรโฆษณาผ่าน คำอธิบายของโรงแรมเฉพาะด้านที่เป็นจุดเด่นหรือมีความคิดเห็นของผู้เข้าพักที่เป็นเชิงบวกจำนวนมากจริงๆ หรือควรทำการปรับปรุงการให้บริการแต่ละด้านของโรงแรมให้ตรงตามความต้องการของผู้เข้าพัก โดยการตามหาสาเหตุที่แท้จริงที่ทำให้ผู้เข้าพักได้แสดงความคิดเห็นเป็นเชิงลบเพื่อทำการปรับปรุงการให้บริการให้ดีขึ้น ทำให้ผู้เข้าพักเกิดความประทับใจและจงรักภักดีต่อโรงแรม แล้วทำการ โฆษณาประชาสัมพันธ์ผ่านคำอธิบายใหม่ (โรงแรมที่เขียนคำอธิบายไม่ครอบคลุม) โรงแรมที่ 14 และ โรงแรมที่ 25 ส่วนโรงแรมที่เขียนคำอธิบายไม่ครอบคลุมในคุณลักษณะด้านต่างๆ แต่มีผู้แสดงความคิดเห็นเป็นเชิงบวกต่อคุณลักษณะจำนวนมากกว่าคำอธิบายของโรงแรม ควรปรับปรุงการเขียน คำอธิบายของโรงแรมใหม่โดยอ้างอิงจากผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักเพื่อไม่ทำให้เสีย โอกาสสำหรับผู้ที่กำลังจะเข้าพักที่สนใจต่อคุณลักษณะด้านอื่นๆ ของโรงแรม



บทที่ 5

สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอวิธีการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรมร่วมกับคุณลักษณะเด่นของโรงแรม เพื่อตอบสนองความต้องการของผู้เข้าพักโรงแรม โดยรวบรวมข้อความคิดเห็นของผู้เข้าพักจากเว็บไซต์ www.tripAdvisor.com ในการระบุคุณลักษณะเด่นของโรงแรม ใช้หลักการสกัดคำคุณลักษณะที่อยู่ในข้อความคิดเห็น และคำที่มีความหมายเหมือนคำคุณลักษณะโดยใช้พจนานุกรมคำเหมือน (Synonym) การสกัดคำระบุข้อความคิดเห็น เป็นขั้นตอนในการสกัดเอาคำที่บ่งบอกถึงความรู้สึกที่เป็นเชิงบวกเชิงลบในโดเมนของโรงแรมเพื่อนำมาวิเคราะห์และสรุปว่า คำคุณลักษณะแต่ละคำเป็นเชิงบวกหรือเชิงลบ ซึ่งในการสกัดระบุข้อความคิดเห็น ใช้วิธีการสกัดตามหน้าที่ของคำ ได้แก่ คำกริยา (VERB) คำคุณสรรพ (ADJ) และ คำขยายกริยา (ADV) เนื่องจากชนิดคำเหล่านี้สามารถบ่งบอกถึงความรู้สึกได้ ผลการวิจัยสามารถสรุปผล อภิปรายผล และข้อเสนอแนะดังนี้

5.1 สรุปผลและอภิปราย

ผลการวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรมและผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักที่มีต่อโรงแรมกับคุณลักษณะเด่นที่โรงแรมใช้ในการประชาสัมพันธ์ พบว่า

1. ผลการวัดประสิทธิภาพของการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม พบว่า จำนวนความคิดเห็นมีผลต่อประสิทธิภาพในการวิเคราะห์ความคิดเห็น ซึ่งจากผลการวัดประสิทธิภาพการวิเคราะห์ความคิดเห็นโรงแรมที่ 1, 2, 3 และ 4 ซึ่งมีจำนวนความคิดเห็นทั้งหมด 694, 682, 700 และ 707 ตามลำดับ พบว่า มีประสิทธิภาพโดยรวม เท่ากับ 0.75, 0.83, 0.81 และ 0.76 ตามลำดับ ในส่วนของโรงแรมอื่นๆ ซึ่งมีจำนวนความคิดเห็นอยู่ระหว่าง 280 – 400 ข้อความ พบว่ามีประสิทธิภาพโดยรวม ไม่เกิน 0.72 ในการเปรียบเทียบความคิดเห็นของผู้เข้าพักที่มีต่อโรงแรมกับคุณลักษณะเด่นที่โรงแรมใช้ในการประชาสัมพันธ์ พบว่า การประชาสัมพันธ์จุดเด่นในรายละเอียดของโรงแรมบนเว็บไซต์ไม่ตรงกับผลการวิเคราะห์ความคิดเห็นของผู้เข้าพัก เช่น โรงแรมที่ 1 พบว่า ผู้เข้าพักโรงแรมมีความคิดเห็นเชิงบวกในด้าน Location, Food, Service และ Rooms แต่คำอธิบายของโรงแรมได้แสดงจุดเด่นไว้เพียง 2 ด้าน คือ Location และ Rooms เท่านั้น ทางโรงแรมสามารถนำเอาข้อมูลที่วิเคราะห์ความคิดเห็นจากลูกค้าที่เป็นเชิงบวกในอีก 2 คุณลักษณะเด่น คือ Service และ Food เขียนเพิ่มเติมในคำอธิบายโรงแรมว่าเรามีจุดเด่นอีก 2 ด้านดังกล่าว ส่วนโรงแรมที่ 14 พบว่า ผู้เข้าพักมีความคิดเห็นเชิงบวกในด้าน Service, Food, Location, Room และ Price ตามลำดับ โรงแรมได้

เขียนคำอธิบายจุดเด่นของโรงแรม 1 ด้าน คือ Location และจากผลการวิเคราะห์ความคิดเห็นจากผู้เข้าพักโรงแรม พบว่า มีทั้งข้อความคิดเห็นในเชิงบวกและเชิงลบ

2. ฝ่ายบริหารของโรงแรมสามารถนำผลการวิจัยไปใช้เป็นแนวทางในการพัฒนากลยุทธ์การตลาดของโรงแรมได้ เช่น ผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรมที่ 14 พบว่า ผู้เข้าพักมีความคิดเห็นเชิงบวกทั้ง 5 ด้าน แสดงให้เห็นว่าโรงแรมมีจุดเด่นในทุกด้าน แต่ในส่วนของคำอธิบายโรงแรม ได้เขียนแสดงจุดเด่นของโรงแรมไว้เพียงด้านเดียวเท่านั้น ดังนั้นโรงแรมสามารถนำผลการวิเคราะห์ความคิดเห็นจากผู้เข้าพักไปพัฒนาคำอธิบายสำหรับประชาสัมพันธ์โรงแรมเพื่อให้ครอบคลุมกับคุณลักษณะเด่นจากความคิดเห็นของผู้เข้าพักที่มีความคิดเห็นที่เป็นเชิงบวกต่อคุณลักษณะทั้ง 5 ด้านของโรงแรม เพื่อให้เป็นที่สนใจของผู้เข้าพักที่กำลังมองหาคุณลักษณะที่ตรงตามความต้องการส่วนในโรงแรมที่ 18, โรงแรมที่ 21 และ โรงแรมที่ 25 มีผลการวิเคราะห์ที่มีความคิดเห็นของผู้เข้าพักโรงแรมทั้งที่เป็นความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ ซึ่งมีจำนวนความคิดเห็นเชิงลบทั้ง 5 คุณลักษณะ เป็นจำนวนมาก ดังนั้นทางโรงแรมจึงควรนำเอาข้อความคิดเห็นที่ผู้เข้าพักแสดงความคิดเห็นในเชิงลบ ไปเป็นข้อมูลเพื่อนำไปแก้ไขข้อบกพร่องต่างๆ ให้ดีขึ้นเพื่อพัฒนากลยุทธ์การตลาดของโรงแรมต่อไป

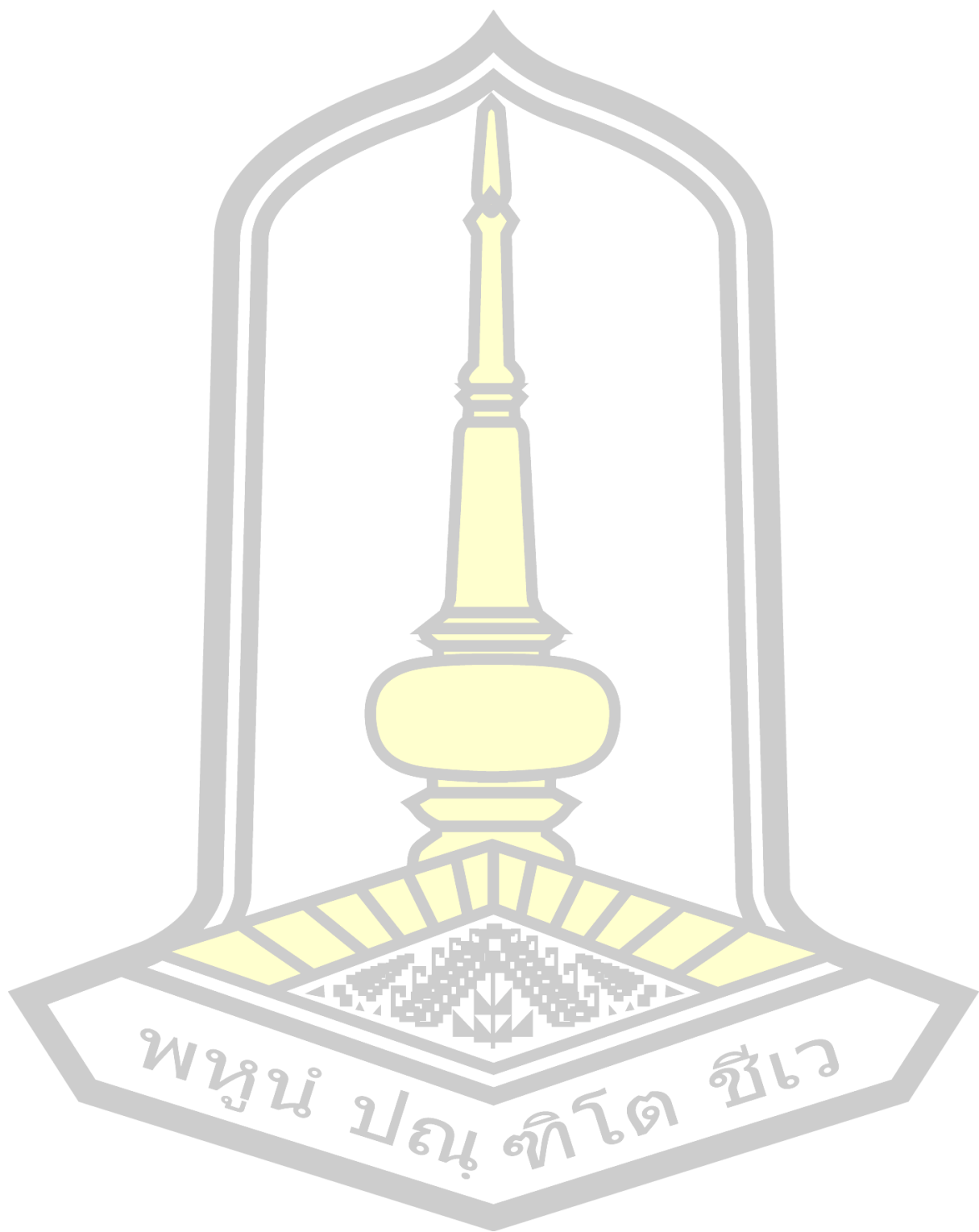
3. ผู้ที่ต้องการหาที่พักสามารถนำผลการวิเคราะห์ความคิดเห็นจากผู้เข้าพักโรงแรม ไปใช้เพื่อการค้นหาที่พักให้ตรงกับความต้องการมากขึ้น

5.2 ข้อเสนอแนะ

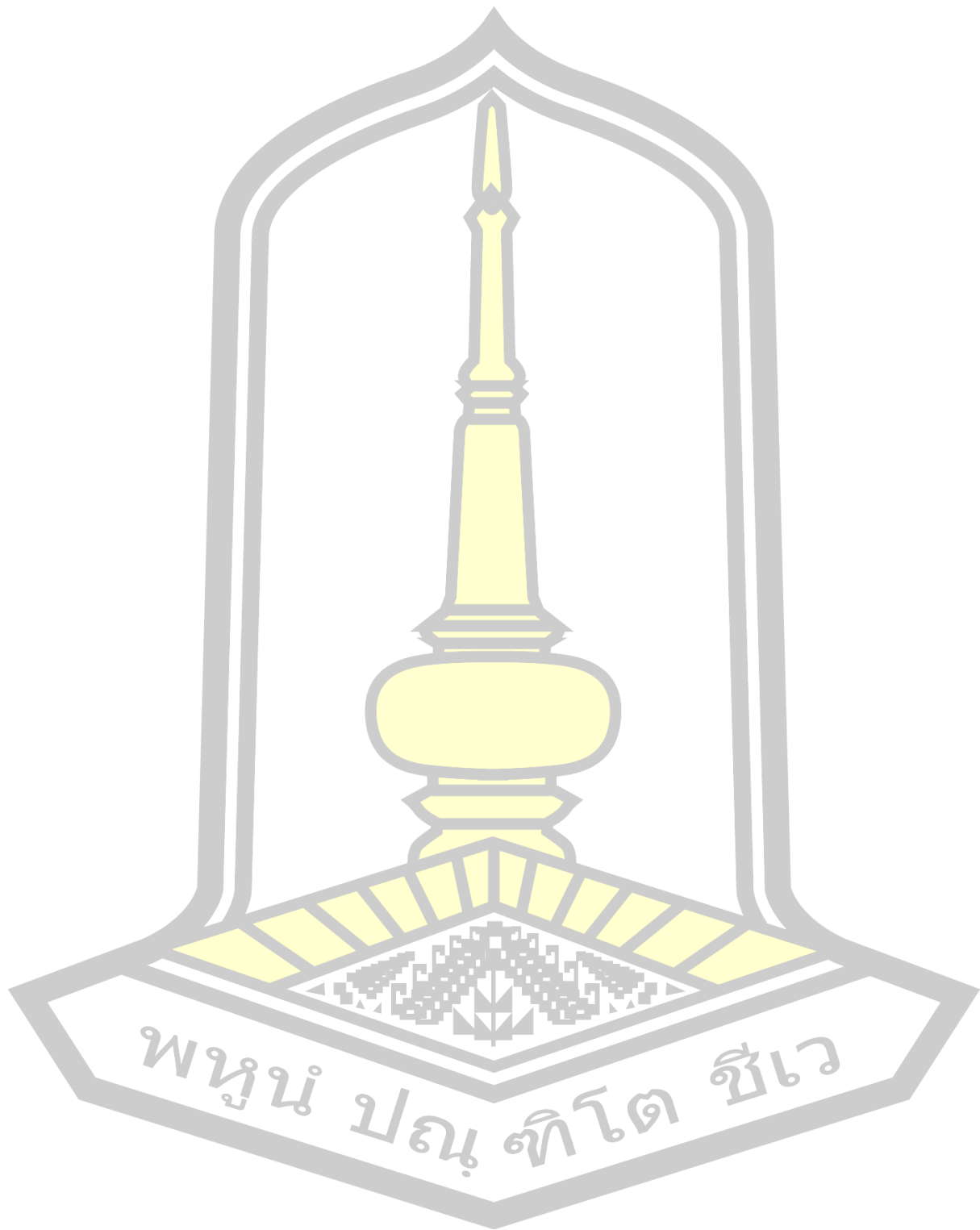
1. ผลการวิเคราะห์ความคิดเห็นของผู้เข้าพักโรงแรม พบว่า ผู้เข้าพักโรงแรมส่วนมากมีความคิดเห็นในเชิงบวก ควรเพิ่มคลาสในการจำแนกให้เห็นถึงความแตกต่างที่ชัดเจน เช่น Strong Positive , Ordinary positive, Ordinary Negative และ Strong Negative เป็นต้น เพื่อให้ฝ่ายบริหารของโรงแรมสามารถนำผลการวิจัยไปใช้เป็นแนวทางในการพัฒนากลยุทธ์การตลาดของโรงแรมได้อย่างมีประสิทธิภาพยิ่งขึ้น

2. ข้อความที่ใช้ทดสอบในงานวิจัยนี้เป็นข้อความภาษาอังกฤษ ควรนำแนวคิดขั้นตอนวิธีการที่นำเสนอไปปรับใช้กับข้อมูลที่เป็นข้อความภาษาไทยซึ่งมีความซับซ้อนแตกต่างจากภาษาอื่น

3. ควรมีการเช็คความถี่ของคำนาม (Noun) คำอื่นๆ นอกเหนือจากคำที่เป็นคำคุณลักษณะในงานวิจัยนี้ ซึ่งอาจเป็นคุณลักษณะ (Feature) แฝงที่อาจปรากฏอยู่



บรรณานุกรม



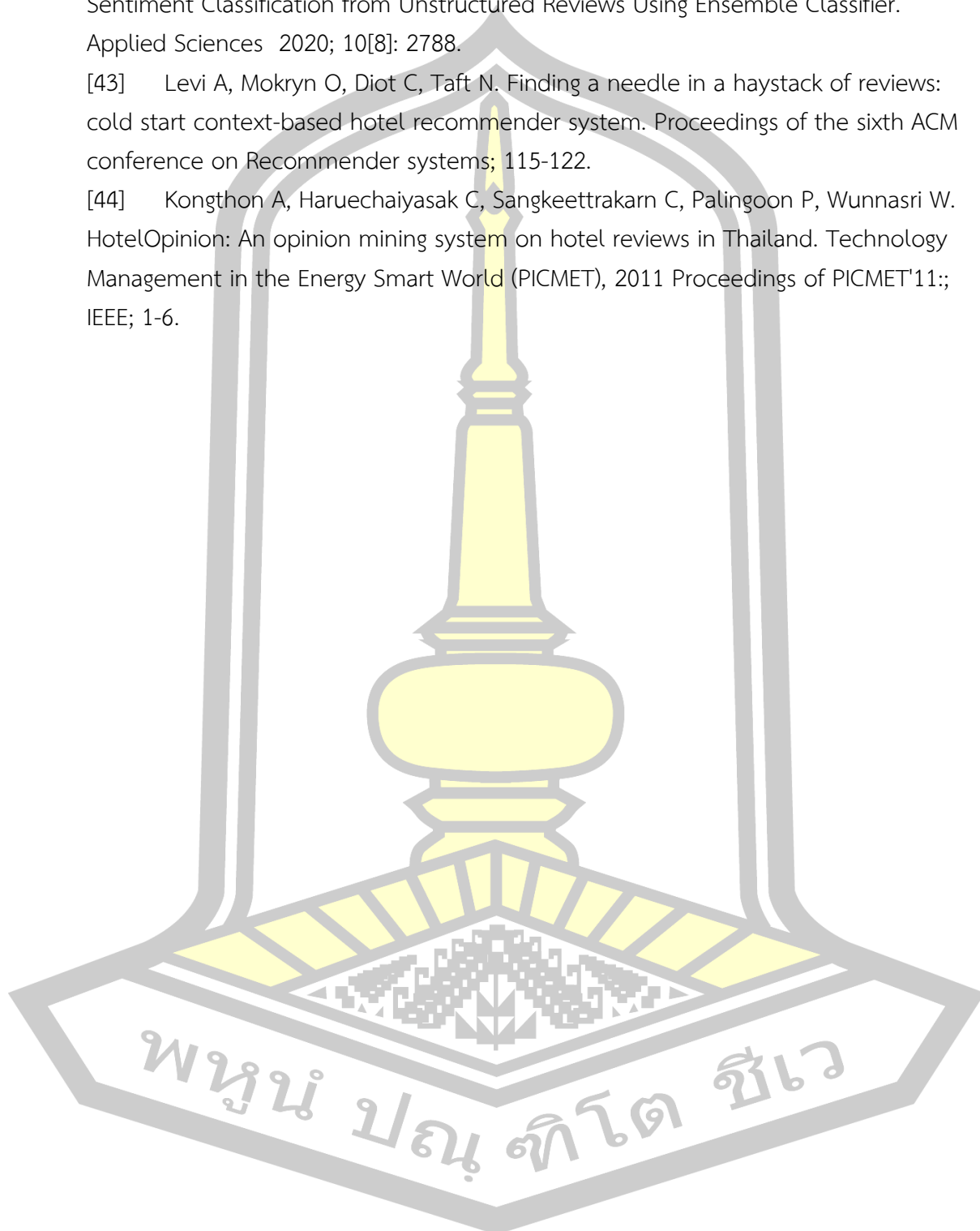
บรรณานุกรม

- [1] King SF, Burgess TF. Understanding success and failure in customer relationship management. *Industrial Marketing Management* 2008; 37[4]: 421-431.
- [2] (สวทช). GSPT4. Smart Industry. 2557;
- [3] Milović B. Social media and eCRM as a prerequisite for hotel success. 2012;
- [4] อมรรัตน์ สุขชะ และนนทิภัก เพียรโรจน์. แนวทางการบริหารประสบการณ์ลูกค้าเพื่อพัฒนาผลการดำเนินงานของธุรกิจโรงแรมในจังหวัดภูเก็ต. *วารสารบริหารธุรกิจศรีนครินทรวิโรฒ* 2558; 6[2]: 74-83.
- [5] อัจฉรา ภูระย้า และมาลีรัตน์ โสตานิล. การวิเคราะห์ทัศนคติการให้บริการของโรงแรมจากบทวิจารณ์ภาษาไทยและภาษาอังกฤษ. 2015.
- [6] Yaakub MR, Li Y, Algarni A, Peng B. Integration of opinion into customer analysis model. *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03; IEEE Computer Society*; 164-168.
- [7] Kongthong A, Angkawattanawit N, Sangkeettrakarn C, Palingoon P, Haruechaiyasak C. Using an opinion mining approach to exploit web content in order to improve customer relationship management. *PICMET 2010 Technology Management for Global Economic Growth; IEEE*; 1-6.
- [8] Tsai C-F, Chen K, Hu Y-H, Chen W-K. Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management* 2020; 80104122.
- [9] Songpan W. The analysis and prediction of customer review rating using opinion mining. *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA); IEEE*; 71-77.
- [10] Calheiros AC, Moro S, Rita P. Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management* 2017; 26[7]: 675-693.
- [11] กนกวรรณ เขียนวรรณ. การประมวลผลภาษาธรรมชาติ 1 ตุลาคม 2558; www.mbs.mut.ac.th/paper/pdf/29.pdf.
- [12] Ramasubramanian C, Ramya R. Effective pre-processing activities in text mining using improved porter's stemming algorithm. *International Journal of Advanced Research in Computer and Communication Engineering* 2013; 2[12]: 2278-1021.

- [13] Haddi E, Liu X, Shi Y. The role of text pre-processing in sentiment analysis. *Procedia Computer Science* 2013; 1726-32.
- [14] Selvam B, Abirami S. A survey on opinion mining framework. *International Journal of Advanced Research in computer and communication Engineering* 2013; 2[9]: 3544-3549.
- [15] Mumu T. Social Network Opinion and posts mining for community preference discovery. 2013;
- [16] ศุภชัย ตั้งวงศ์ศานต์. ระบบการจัดเก็บและการสืบค้นสารสนเทศด้วยคอมพิวเตอร์ (พิมพ์ครั้งที่ 2) โรงพิมพ์พิทักษ์การพิมพ์; 2553.
- [17] Kumar AA, Chandrasekhar S. Text data pre-processing and dimensionality reduction techniques for document clustering. *International Journal of Engineering Research and Technology*; ESRSA Publications;
- [18] Zhang Z. Urcf: an approach to integrating user reviews into memory-based collaborative filtering. University of Maryland, Baltimore County; 2013.
- [19] กานดา แผ้ววัฒนากุล. การวิเคราะห์เหมืองข้อเสนอแนะจากบทวิจารณ์รายการโทรทัศน์. *สถาบันบัณฑิตพัฒนบริหารศาสตร์*; 2555.
- [20] เอกสิทธิ์ พัทธวงศ์ศักดิ์. การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมนิง เบื้องต้น. กรุงเทพฯ : เอเชีย ดิจิตอลการพิมพ์; 2557.
- [21] Khan A, Baharudin B, Khan K. Sentiment classification from online customer reviews using lexical contextual sentence structure. *Software Engineering and Computer Systems*: Springer 2011:317-331.
- [22] Trilla A, Alias F. Sentence-based sentiment analysis for expressive text-to-speech. *Audio, Speech, and Language Processing, IEEE Transactions on* 2013; 21[2]: 223-233.
- [23] Sharma R, Nigam S, Jain R. Polarity detection at sentence level. *International journal of computer applications* (0975-8887) 2014; 86[11]:
- [24] รุติมา เกษมศรีธนาวัฒน์ และธนสนี เพียรตระกูล. การจำแนกความคิดเห็นโดยใช้จำแนกแบบเบย์ร่วมกับการเลือกคุณลักษณะด้วยอัลกอริทึมรีลีฟ. *ThaiLis* 2011;
- [25] มาสวีร์ มาศดิศร์โชติ. การทำเหมืองความคิดเห็นภาษาไทย. *Sripatum Review of Science and Technology* 2557; 6[1]: 120-128.
- [26] Afzaal M, Usman M. A novel framework for aspect-based opinion classification for tourist places. *Digital Information Management (ICDIM), 2015 Tenth International Conference on*; 21-23 Oct. 2015; 1-9.
- [27] Kanzutaka S, Shunsuke I, Hiroshi M, Tsutomu E. Analyzing Tourism Information on Twitter for a Local City.

- [28] วรัญญา วรณศรี. ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม. 2553; [15 ตุลาคม 2558];
<http://thailang.nectec.or.th/halloffame/images/stories/best/download/13p33c001.pdf>
- [29] Bucur C. Using opinion mining techniques in tourism. *Procedia economics and finance* 2015; 231666-1673.
- [30] Liu L, Lv Z, Wang H. Opinion mining based on feature-level. *Image and Signal Processing (CISP)*, 2012 5th International Congress on; IEEE; 1596-1600.
- [31] Agarwal B, Mittal N, Bansal P, Garg S. Sentiment analysis using common-sense and context information. *Computational intelligence and neuroscience* 2015; 201530.
- [32] Khan FH, Bashir S, Qamar U. TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision support systems* 2014; 57245-257.
- [33] Mostafa MM. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications* 2013; 40[10]: 4241-4251.
- [34] Liu Y, Yu X, Liu B, Chen Z. Sentence-level sentiment analysis in the presence of modalities. *International Conference on Intelligent Text Processing and Computational Linguistics*; Springer; 1-16.
- [35] Chenlo JM, Losada DE. An empirical study of sentence features for subjectivity and polarity classification. *Information Sciences* 2014; 280275-288.
- [36] Natchiar U, S.Baulkani. Customer Relationship Management Classification Using Data Mining Techniques *International Conference on Science, Engineering and Management Research (ICSEMR 2014)* 2014;
- [37] Ennaji FZ, El Fazziki A, Sadgal M, Benslimane D. Social intelligence framework: Extracting and analyzing opinions for social CRM. 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA); IEEE; 1-7.
- [38] Eirinaki M, Pisal S, Singh J. Feature-based opinion mining and ranking. *Journal of Computer and System Sciences* 2012; 78[4]: 1175-1184.
- [39] Tribhuvan PP, Bhirud S, Tribhuvan AP. A peer review of feature based opinion mining and summarization. *IJCSIT International Journal of Computer Science and Information Technologies* 2014; 5[1]:
- [40] Milović B. Social media and eCRM as a prerequisite for hotel success. *Management Information Systems* 2012; 7[3]: 26-31.
- [41] Kasper W, Vela M. Sentiment analysis for hotel reviews. *Computational linguistics-applications conference*; 45-52.

- [42] Khalid M, Ashraf I, Mehmood A, Ullah S, Ahmad M, Choi GS. GBSVM: Sentiment Classification from Unstructured Reviews Using Ensemble Classifier. *Applied Sciences* 2020; 10[8]: 2788.
- [43] Levi A, Mokryn O, Diot C, Taft N. Finding a needle in a haystack of reviews: cold start context-based hotel recommender system. *Proceedings of the sixth ACM conference on Recommender systems*; 115-122.
- [44] Kongthon A, Haruechaiyasak C, Sangkeettrakarn C, Palingoon P, Wunnasri W. HotelOpinion: An opinion mining system on hotel reviews in Thailand. *Technology Management in the Energy Smart World (PICMET), 2011 Proceedings of PICMET'11*; IEEE; 1-6.



ภาคผนวก ก

ตารางแสดงรายชื่อโรงแรมและคำอธิบายคุณลักษณะโรงแรม

ลำดับที่	ชื่อโรงแรม	คำอธิบายคุณลักษณะโรงแรม
1	Ibis Styles Bangkok Khaosan Viengtai	ibis Styles Bangkok Khaosan Viengtai is situated in the heart of the action-packed Khaosan area surrounded by all the vibrant life that makes it the must-go-to place in Bangkok.....
2	Adelphi Forty-Nine Hotel	Adelphi Forty-Nine is ideally located on Sukhumvit Soi 49 in the heart of downtown Bangkok in the upscale district of Bangkok in between the trendy Thonglor and peaceful Phrom Phong where it is considered home to most of the expatriates living in Bangkok
3	S Sukhumvit Suites	S Sukhumvit Suites is a stylish city hotel in the heart of Sukhumvit area, ideally located right on the main Sukhumvit Road, between Sukhumvit 21 (Asoke Intersection) and Sukhumvit 23.....
4	Phranakorn_Nornlen_Hotel	Phranakorn-Nornlen Hotel will get special offer. We are a small hotel situated within Phranakorn district (Old Bangkok), with the concept of slow life, we hope to bring your trip and rest apart from the usual hectic environment....
5	Metro Pratunam Boutique Hotel	Metro Pratunam Boutique is located in the no.1 tourist destination, Pratunam, Bangkok. It is conveniently situated in a calm and quiet environment away from the busy and bustling streets and markets of Pratunam. Major shopping malls such as Platinum, Central Chidlom, and Central World are at the near distance. BTS Chidlom can be reached within a 12 minute walk. Metro Pratunam Boutique is the perfect accommodation for family and friends who wish to stay together in the same room or individual travelers who like to

		share a bedroom with other like-minded travelers. Our stylish Boutique Hotel is furnished to a high standard with all handmade woven rattan and teakwood beds, oversized teakwood doors secured with an electric key card lock system. We have 23 bedrooms, 2 restaurants and exclusive lounge. All spacious bedrooms are air-conditioned which can accommodate from 3 to 8 person. Modern bathrooms are equipped with wash basins, high pressure showers and separate toilets. Some of our guest rooms have bathroom en suite and others have shared bathrooms.
6	CityPoint Hotel	CityPoint is a uniquely non-smoking, moderately priced hotel located in the heart of Bangkok. It is eco-friendly and is situated on a quiet street within easy walking distance to public transportation, restaurants, night life, shopping, parks, and entertainment. Citypoint's 54 smoke free rooms offer a contemporary, relaxing oasis in the midst of all the activity of Bangkok.
7	Smart Suites	Welcome to Smart suites The boutique Hotel, the hotel for both business and leisure travelers. Our friendly atmosphere is evident the moment you arrive. The friendly and courteous staff is always ready to greet you with a smile and Sawadee, Thai custom for a warm welcome. Situated in a quiet location, the Smart suites Hotel located in the heart of Bangkok. The hotel provide its guests with 26 rooms modern, bright. All rooms have windows private bathrooms with hot & cold showers, telephone with IDD, color TV with satellite, refrigerator. Our rooms also offer in-room safety deposit box and hairdryer.
8	Citrus Sukhumvit 11 by Compass Hospitality	Citrus Sukhumvit 11 is a brand-new hotel featuring 132 stylish rooms and 4 suites. The hotel is located

		near the Nana BTS Sky Train Station so shopaholics can easily reach Bangkok's major shopping malls . The bustling dining and nightlife scene of Soi 11 makes it a hotel worth trying if you want to experience the Bangkok nightlife. For medical tourism, our Citrus Sukhumvit 11 is also the perfect option as the hotel is just a 5-minute walk away from the Bumrungrad International Hospital.
9	Tints of Blue Hotel	A smart hotel right in the heart of downtown Bangkok's lively Sukhumvit area. Located in the famous Sukhumvit Soi 11, surrounded by endless options for shopping, eating (halal-friendly options) and nightlife. Just a 5 minutes walk to Nana BTS Skytrain Station, the hotel is a short distance from trendy shopping malls like Terminal 21, Emporium, Emquartier and Central Embassy, offering speedy access to other major landmarks and attractions like Bamrungrad Hospital, Indra Market, Erawan Shrine, Khaosan Road, Chatuchak Weekend Market and many more. The smart choice for savvy travellers who are looking for the best location combined with the best value, offering an affordable and hassle-free stay in Bangkok. Start the day right with our free Express Start Breakfast or Grab & Go option and stay connected with our free Wi-Fi throughout the hotel. Get the most out of your stay with our rooftop 24-hour Fitness Centre with city view, self-service laundry facility and 10-person meeting room.
10	Holiday Inn Express Bangkok Sukhumvit 11	A smart hotel right in the heart of downtown Bangkok's lively Sukhumvit area. Located in the famous Sukhumvit Soi 11, surrounded by endless options for shopping, eating (halal-friendly options) and nightlife. Just a 5 minutes walk to Nana BTS Skytrain Station, the hotel is a short distance from trendy shopping malls like Terminal 21, Emporium,

		Enquartier and Central Embassy, offering speedy access to other major landmarks and attractions like Bamrungrad Hospital, Indra Market, Erawan Shrine, Khaosan Road, Chatuchak Weekend Market and many more. The smart choice for savvy travellers who are looking for the best location combined with the best value, offering an affordable and hassle-free stay in Bangkok. Start the day right with our free Express Start Breakfast or Grab & Go option and stay connected with our free Wi-Fi throughout the hotel. Get the most out of your stay with our rooftop 24-hour Fitness Centre with city view, self-service laundry facility and 10-person meeting room.
11	Bally Suite Sukhumvit	HOTEL OVERVIEW Set right in the heart of Bangkok and within close proximity to major business, shopping and entertainment districts, Bally Suite Sukhumvit is beautifully appointed hotel providing guests with a comfortable living environment and all facilities. The rooms and suites feature Thai elements and furnishings that blend perfectly with modern conveniences for a comfortable stay. With just a short walking distance to Phrom Phong BTS station, which connects you to all part of Bangkok, Bally Suite Sukhumvit is your ideal choice of accommodation whether you are on a leisure or business trip. THE ROOMS All 78 guestrooms come with tasteful decor mixing traditional and modern styles, providing guest with a comfortable yet lively ambience. The rooms are spacious and equipped with various modern conveniences ranging from a flat-screen TV to Wi-Fi internet access for guests to make the most of their time during the stay.
12	Majestic Suites	Everyone needs a place to lay their weary head. For travelers visiting Bangkok, Majestic Suites is an excellent choice for rest and rejuvenation. Well-

		known for its charming environment and proximity to great restaurants and attractions, Majestic Suites makes it easy to enjoy the best of Bangkok. Guest rooms offer a flat screen TV, air conditioning, and a refrigerator, and Majestic Suites makes getting online easy as free wifi is available. You can also take advantage of some of the amenities offered by the hotel, including a 24 hour front desk, room service, and a concierge. In addition, guests can enjoy free breakfast during their visit. Close to some of Bangkok's most popular landmarks, such as Sukhumvit (0.4 mi) and Erawan Shrine (Thao Mahaprom Shrine) (0.9 mi), Majestic Hotel is a great destination for tourists. If you're looking for a sushi restaurant, consider a visit to Above Eleven, Nami Teppanyaki Steakhouse - at the JW Marriott Hotel Bangkok, or Tsu Japanese Restaurant - at the JW Marriott Hotel Bangkok which are all conveniently located a short distance from Majestic Suites. Bangkok is also known for some great art galleries, including TCDC Thailand Creative & Design Center, The Jam Factory Gallery, and Papaya Design Furniture & Studio, which are not too far from Majestic Hotel. Whether you're traveling for business, pleasure or both, Majestic Suites is sure to make your visit to Bangkok one worth remembering.
13	Citrus Sukhumvit 13 by Compass Hospitality	Citrus Sukhumvit 13 is a brand-new, modern, and hip hotel in Bangkok with 79 rooms ranging from the intimate Superior Room to the more spacious Citrus Suite. Located on a quiet side street off the bustling Sukhumvit Road, Citrus 13 Bangkok holds an enviable location in the heart of one of Bangkok's most vibrant areas and near the Nana BTS station. Citrus Sukhumvit 13 always believes that our guests' satisfaction is the key to our success. We have

		created a package of special benefits and privileges called the "Lucky 13 Benefits" for guests who have booked the Daily Rate with breakfast of every room type. There is no minimum stay for this package. Lucky 13 offers additional complimentary benefits ex. Late check-out at 13:00 hrs and THB 113 Credit bonus for the Citrus Cafe - excluding alcoholic drinks.
14	Salil Hotel Sukhumvit Soi 8	Nestled down a serene alley off Sukhumvit road, this 30-room hotel is filled with charm and character. As soon as you step in, you will definitely get a relaxed, cozy feel. The hotel's enchanting colonial architectural style can make you feel as if you were in Bangkok several decades ago. A warm Siamese welcome awaits you on arrival. We are passionate about making your stay with us a memorable one. Every guest room has all modern comforts that you need. On the ground floor is the Café de Salil, a place where buffet breakfast is served. Moreover, your lunch or dinner here can turn out to be a great gastronomic experience as well. Other services include a small library corner with free DVD rental, laundry, and transport & sightseeing desk, which is right at the lobby.
15	The Key Bangkok Hotel by Compass Hospitality	Keeping in mind that you're travel for business or leisure should be as comfortable as your home, our rooms were designed to welcome you with warmth for the full extent of your stay. We have all the comforts of the convenience of a hotel. Striking a harmony between Price, Service and an Ideal Address, "THE KEY" is definitely an incomparable choice. With our ambition of providing the most flexible & comfortable accommodation. "THE KEY" is a wise choice.
17	Citichic Hotel Sukhumvit by	Are you ready to experience one of Bangkok's

	Compass Hospitality	hippest settings? A modern, stylish, affordable place to stay in the heart of Bangkok awaits you. iCheck Inn is just meters away from Sukhumvit Soi 11, home to some of the area's best bars and restaurants. Our inviting, lifestyle rooms are available with the very latest amenities, including free Wi-Fi, perfect for today's modern leisure travelers. Within easy reach of Terminal 21, Bangkok's trendiest shopping and entertainment center, use our free tuk-tuk service to BTS Nana to explore the rest of this thriving metropolis. A great choice, centrally located - welcome to one of Bangkok's most exciting boutique hotels.
18	Hotel Clover Asoke	Hotel Clover Asoke, the best boutique hotel in Sukhumvit, Bangkok is now open and always welcome our value guests. Our first oversea foray, Hotel Clover in Bangkok, Thailand. Hotel Clover Asoke opened officially in June 2016. This is a 95 rooms hotel with 4 room categories. Those are Standard, Clover Style, Ladies' Floor and Executive room. Hotel Clover Asoke is the first property that has floors dedicated to the female travelers. We present with chic design and concentrate in girls friendly and solo girl's travelers concept. All rooms with different designs. Uniquely corridor themed and different style.
19	Sam's Lodge	Sam's Lodge is located in heart of Bangkok. Walking distance from the Skytrain BTS Asok Station and Underground MRT Sukhumvit Station. The Shopping malls Terminal 21 & Robinson are just minute walk. Soi cow boy is also in walking distance.
20	Vince Hotel	Vince hotel Pratunam Bangkok is a new fashion retro hotel located in the heart of Bangkok's fashion & entertainment area. Only a few minutes away from Pratunam market, Pantip Plaza, MBK, Siam Paragon,

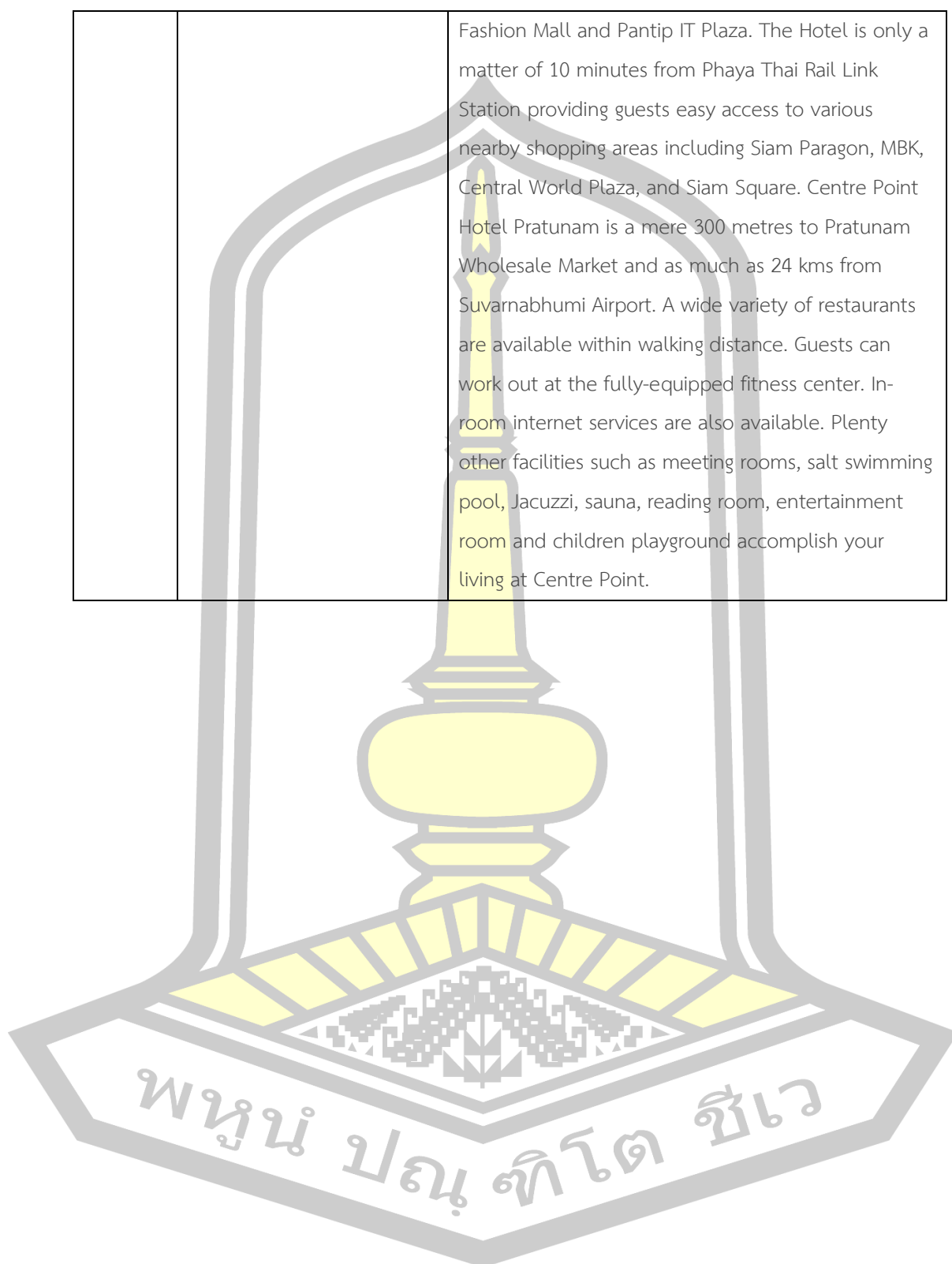
		Central World Shopping Center, BTS stations and easy access to Siam Square by sky train.
21	Citadines Sukhumvit 11 Bangkok	Citadines Sukhumvit11 Bangkok is located ideally along Sukhumvit Road, the residence is surrounded by businesses as well as the trendy shops, restaurants and pubs. With the sky train Nana station, you can explore the rest of the city easily. Vibrant 127 fully furnished apartments, comprising of a fully equipped kitchen and a home entertainment system with a sleek cable TV that slides from the living area to the bedroom.
22	Paragon Inn	Finding an ideal quiet hotel in Bangkok does not have to be difficult. Welcome to Paragon Inn, a nice option for travelers like you. While staying at Paragon Inn, visitors can check out Wat Lat Krabang (1.5 mi) and Wat Krathum Suea Pla (2.9 mi), some of Bangkok's top attractions. Guest rooms offer amenities such as a flat screen TV, air conditioning, and a refrigerator, and guests can go online with free wifi offered by the hotel. The Paragon Hotel features a concierge and room service, to help make your stay more enjoyable. The property also boasts a pool and a poolside bar. If you are driving to Paragon Inn, free parking is available. Travelers looking for steakhouses can head to Pimenton. If you're looking for something to do, Wat Lan Boon (0.3 mi) and Mai Hua Takhe Market (0.9 mi) are a nice way to spend some time, and they are all within walking distance of The Paragon Hotel. Paragon Inn looks forward to welcoming you on your visit to Bangkok.
23	Solitaire Bangkok	Tucked away on a quiet street next to Bangkok's famous Sukhumvit 11 nightlife scene, the Solitaire Hotel offers you a relaxed ambiance in a central location. Just minutes away from famous hospitals, nightlife and Nana BTS skytrain, you can easily reach

		some of the city's most popular destinations in minutes. With 319 modern and stylish rooms spread throughout four buildings, the Solitaire offers guests comfort, luxury and class within the Sukhumvit 11 enclave. Whether you are searching for short or long stay accommodation, Solitaire hotel has plenty of room choices to suit all tastes. All rooms are outfitted with a contemporary design, minibar, a smart TV and free Wi-Fi, and are located in a quiet setting to ensure our guests utmost relaxation. Solitaire offers the most contemporary design and facilities which include a swimming pool, gym, steam rooms, @Spice All Day Dining, On the Rocks lobby bar, complimentary shuttle service to the BTS skytrain and function rooms of varying size.
24	Siri Sathorn Bangkok by UHG	Siri Sathorn Bangkok by UHG is set in the heart of Silom and Sathorn CBD and is within a short distance to BTS Saladaeng and MRT Silom. The hotel offers modern contemporary accommodation for discerning tourists and business travelers. Situated in a charming residential neighborhood, Siri Sathorn Bangkok by UHG is surrounded by famous bars and restaurants, office buildings, embassies, and international hospitals. The hotel also houses Lek Massage Bangkok offering massage and spa services in authentic Thai style. Free Shuttle Service to BTS Saladaeng and MRT Silom is also provided throughout the day, making getting in and out of the hotel effortless. In addition, expressways are in close proximity allowing easy access to Suvarnabhumi and Don Muang international airports as well as all parts of the city.
25	Shama Sukhumvit Bangkok	A celebration of modern living coupled with Shama's reputable signature service and renowned Thai hospitality. Shama Sukhumvit Bangkok offers 90

		contemporary serviced apartments located in a prestigious residential area amongst fashionable restaurants, stylish shopping and the central business district with nearby access to the BTS skytrain and Bumrungrad International Hospital. This unique “resort within a city” meets the refined standards of business and leisure travellers alike. Tropical gardens awash with flora and fauna combined with contemporary furnishings, extensive clubhouse complete with rooftop swimming pool, fitness centre, kids area, yoga and meditation rooms make this property a true retreat from the bustle of the city.
26	The Phoenix Hotel Bangkok	The Phoenix Hotel Bangkok (Suvarnabhumi airport) is ideally a place for stopover travellers who has connecting flight in Bangkok. The hotel just 10 mins away from the airport by car. The hotel also offer free shuttle bus service from & to the airport 24 hours. Also, the hotel offers free shuttle bus to ARL train station nearby only 3 mins away by car. There are 3 rooms type available at our property and offer ensuite with shower, high speed internet, cable TV, parking. FYI : The meeting point for pick-up service is next to AOT counter between gate 3 & 4 (inside the hall) at the arrival hall on 2nd fl. (24 hours)
27	Grand Mercure Bangkok Asoke Residence	Conveniently located in Bangkok's trendy Sukhumvit Area, 5 minutes walking distance to Asoke BTS skytrain and Sukhumvit MRT subway stations, GRAND MERCURE Bangkok Asoke by AccorHotels, is located in the city centre. It offers an easy and quick access around Bangkok in a calm street, with bright Suites, apartment like, from 71sqm. Discover Terminal21, a famous shopping centre around the corner, then relax and enjoy the view at our rooftop swimming pool or dive into the great nightlife only steps away.

		GRAND MERCURE Bangkok Asoke is pet friendly and our WI-FI access is complimentary. We have 40" smartTVs with Japanese, Chinese, French channels. Ideal for a family or couple stay. Fully renovated in 2017, each Suite has a full kitchen with a private balcony.
28	The Okura Prestige Bangkok	The Okura Prestige Bangkok combines Japan's rich service culture and attention to detail with Thailand's world-famous standards of hospitality. This new luxury hotel, with 240 rooms and suites, is located in the award-winning Park Ventures Ecoplex building at the heart of the city's Central Business District. The hotel offers direct access to the city's BTS skytrain network at Phloen Chit station and sits close to major shopping malls, embassies and commercial districts. Every room and suite offers impressive views across the dynamic Bangkok skyline through triple-glazed tinted panoramic windows that insulate against both heat and noise. The rooms combine comfort and luxury with the very latest technological developments designed to impress business and leisure travellers.
29	Bangkok Marriott Hotel The Surawongse	The first Marriott hotel in Bangkok, Thailand that offers a combination of 303 guest rooms, suites and extended-stay residential suites from 1 to 3 bedrooms, Bangkok Marriott Hotel The Surawongse offers two dining outlets; Praya Kitchen, an authentic Thai buffet restaurant and Yào Restaurant & Rooftop Bar, the first modern Chinese influenced multi-level hot spot in Bangkok. For meetings, the hotel in Bangkok's city centre offers 1,555 sqm meeting space with 2 ballrooms, 9 meeting rooms and a rooftop garden.
30	Centre Point Hotel Pratunam	Centre Point Hotel Pratunam is conveniently situated on Petchburi Road right across from Platinum

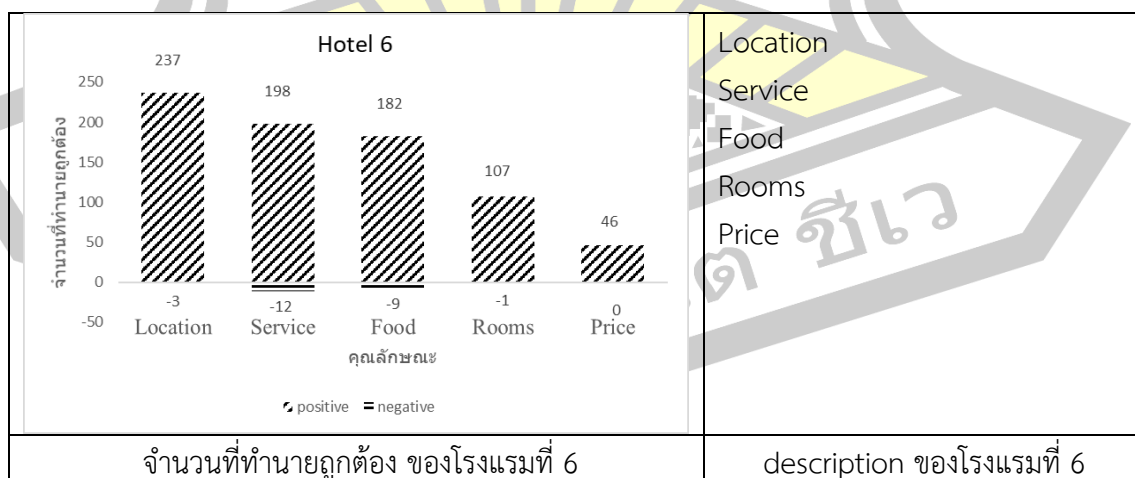
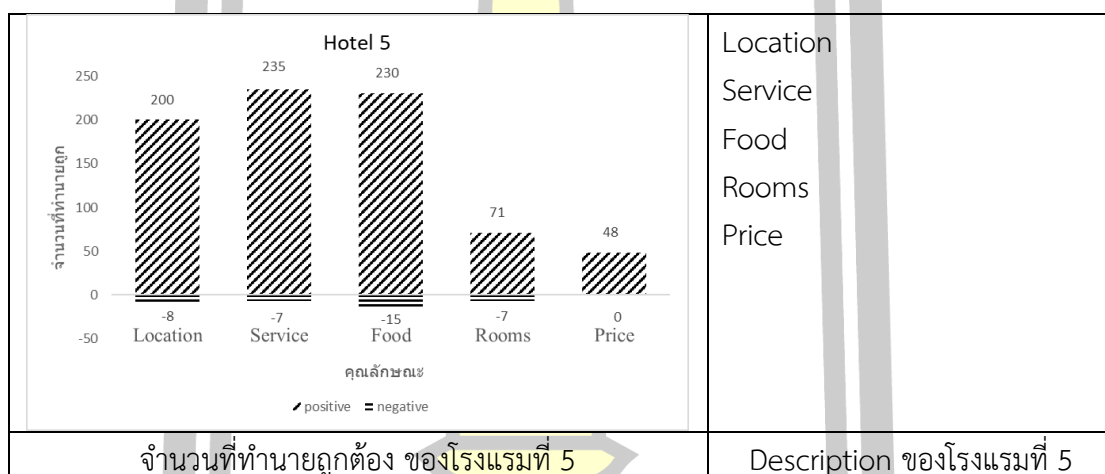
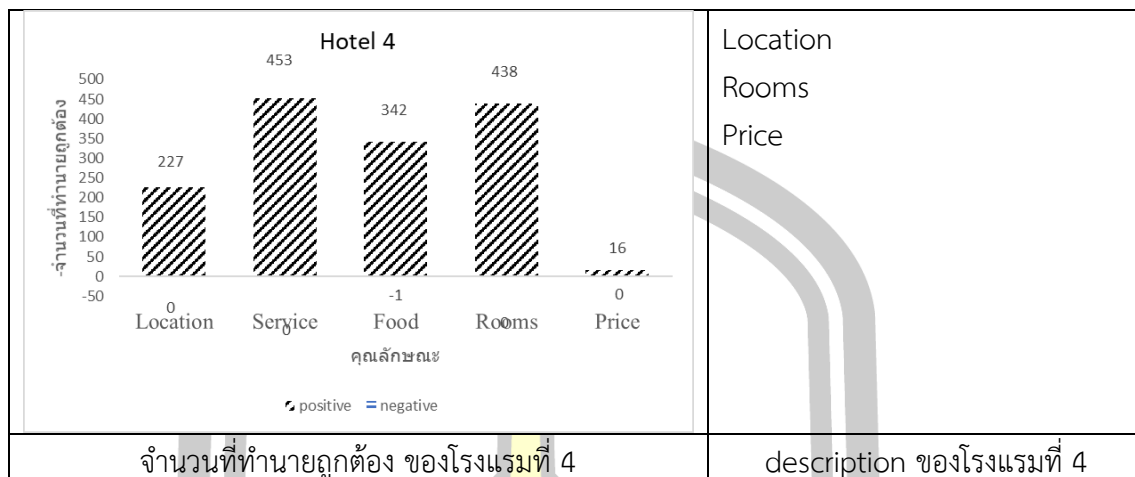
Fashion Mall and Pantip IT Plaza. The Hotel is only a matter of 10 minutes from Phaya Thai Rail Link Station providing guests easy access to various nearby shopping areas including Siam Paragon, MBK, Central World Plaza, and Siam Square. Centre Point Hotel Pratunam is a mere 300 metres to Pratunam Wholesale Market and as much as 24 kms from Suvarnabhumi Airport. A wide variety of restaurants are available within walking distance. Guests can work out at the fully-equipped fitness center. In-room internet services are also available. Plenty other facilities such as meeting rooms, salt swimming pool, Jacuzzi, sauna, reading room, entertainment room and children playground accomplish your living at Centre Point.

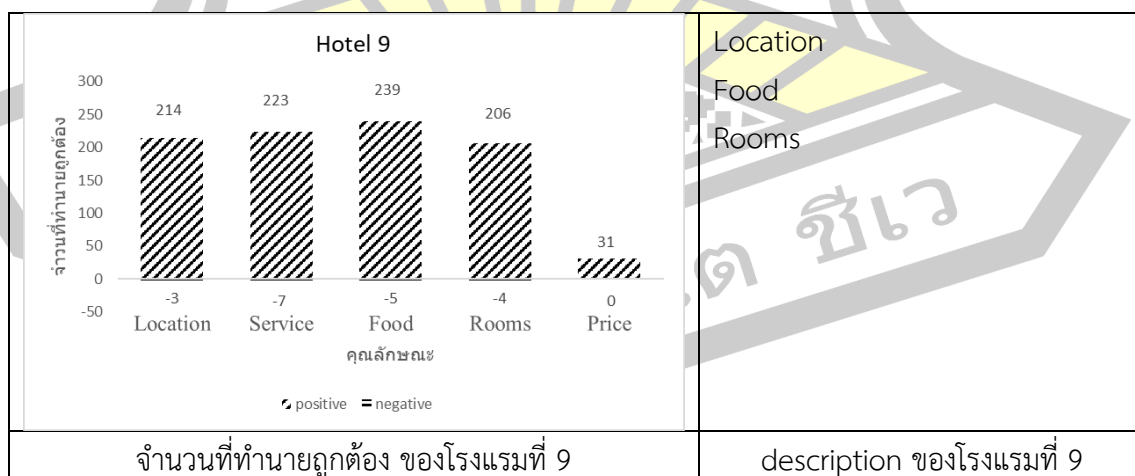
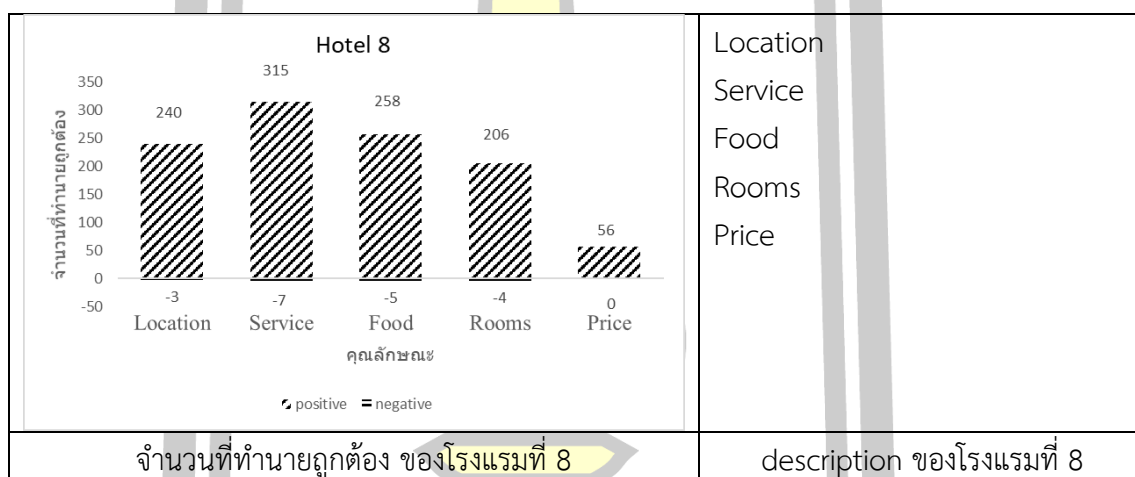
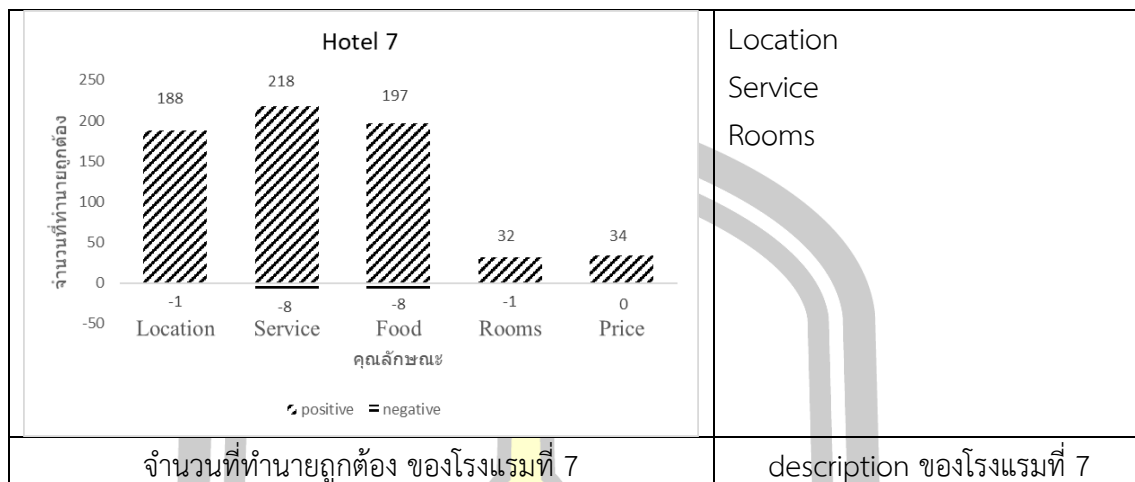


ภาคผนวก ข

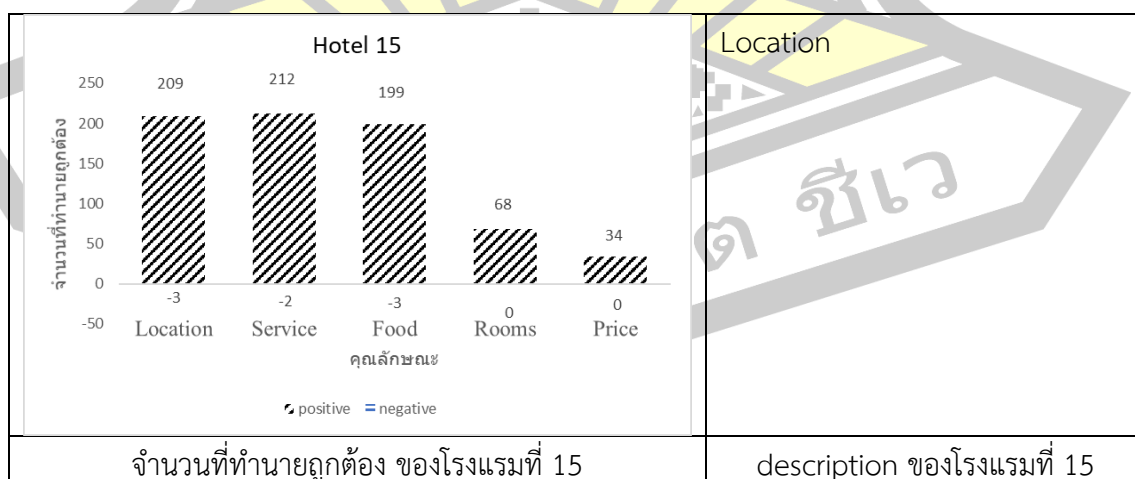
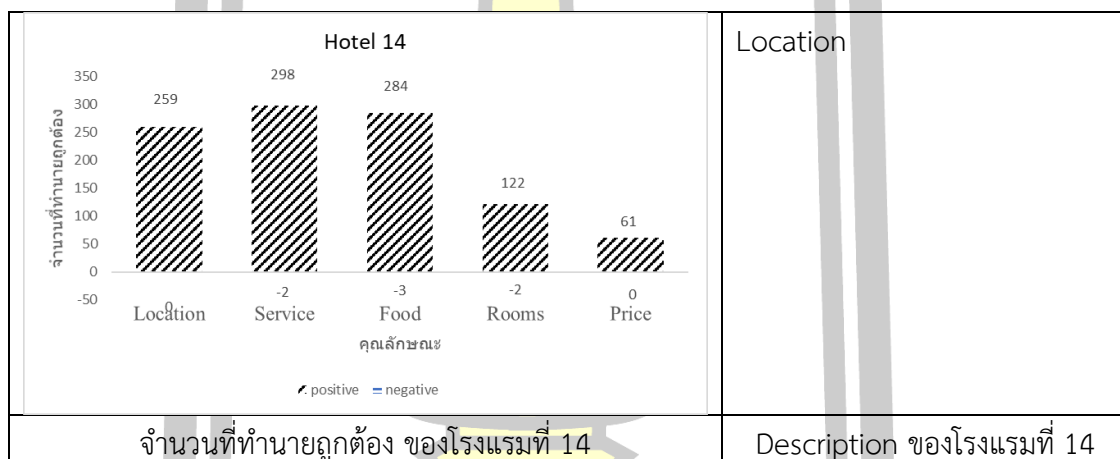
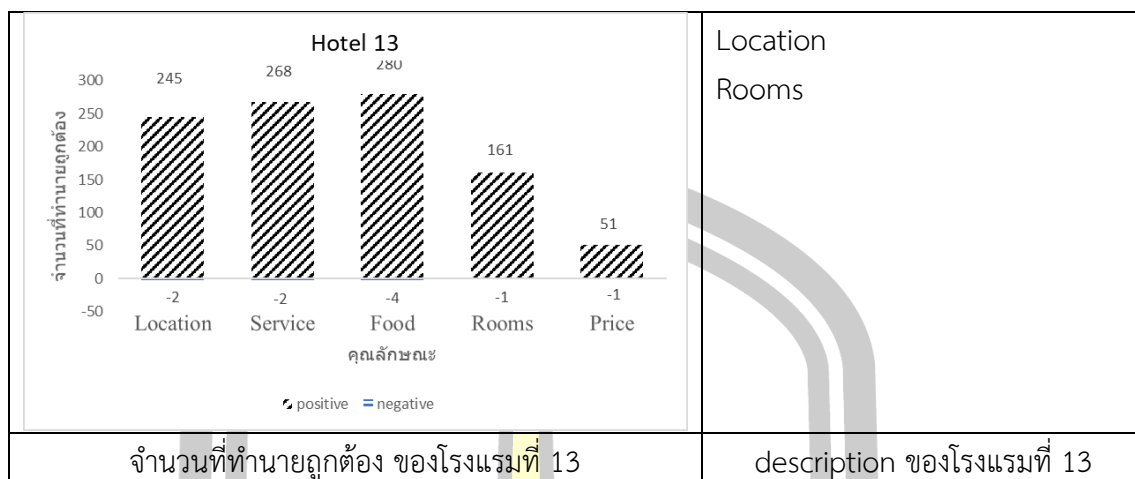
ตารางแสดงผลการเปรียบเทียบความคิดเห็นของผู้เข้าพักกับคำอธิบายโรงแรม

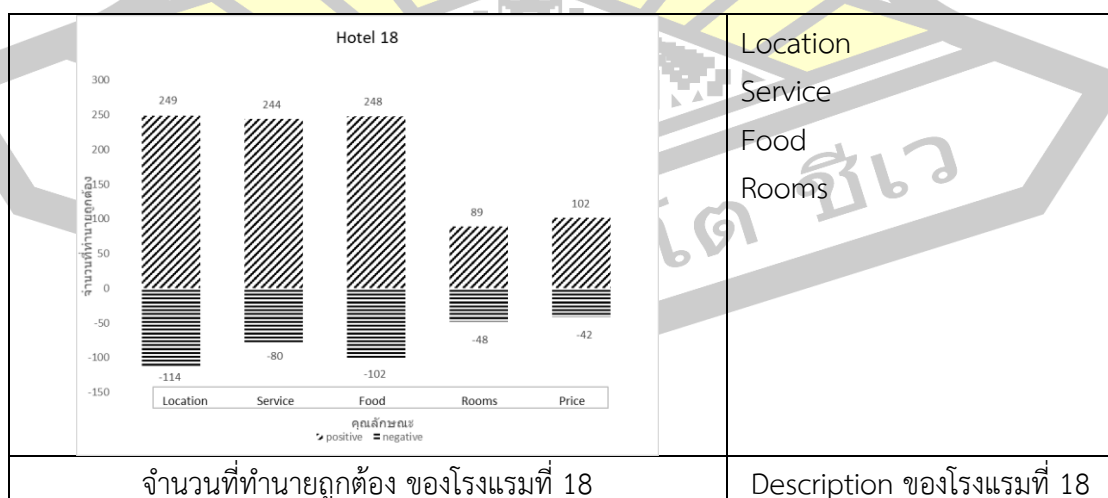
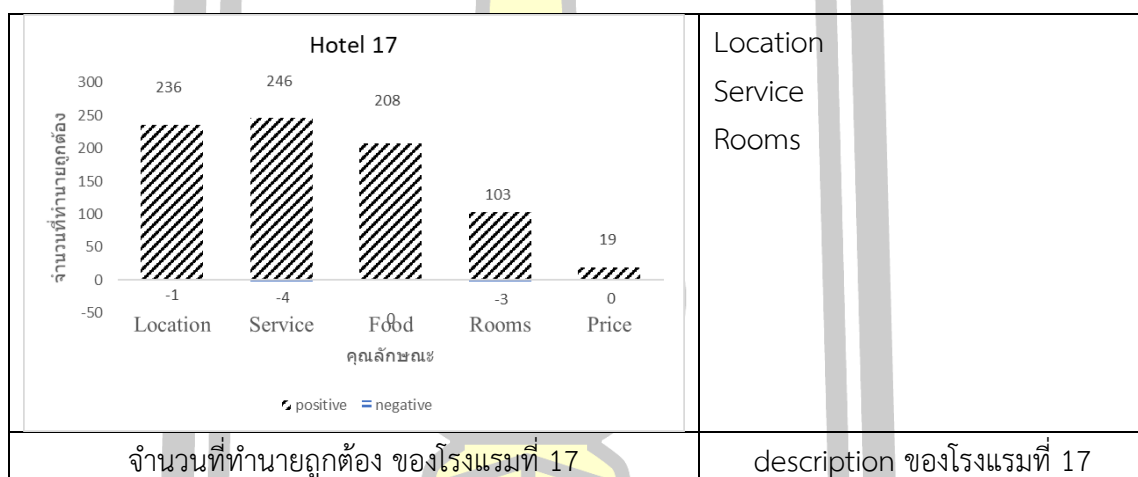
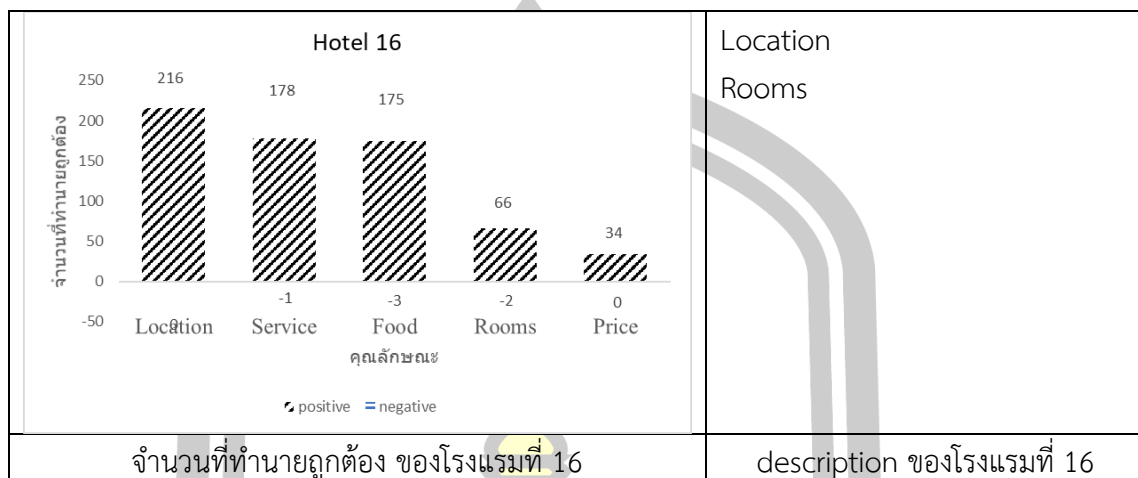
Hotel 1 จำนวนที่ทํานายถูกต้อง คุณลักษณะ positive negative	Location Rooms
จำนวนที่ทํานายถูกต้อง ของโรงแรมที่ 1	description ของโรงแรมที่ 1
Hotel 2 จำนวนที่ทํานายถูกต้อง คุณลักษณะ positive negative	Location Rooms Price
จำนวนที่ทํานายถูกต้อง ของโรงแรมที่ 2	description ของโรงแรมที่ 2
Hotel 3 จำนวนที่ทํานายถูกต้อง คุณลักษณะ positive negative	Location Service Rooms
จำนวนที่ทํานายถูกต้อง ของโรงแรมที่ 3	description ของโรงแรมที่ 3

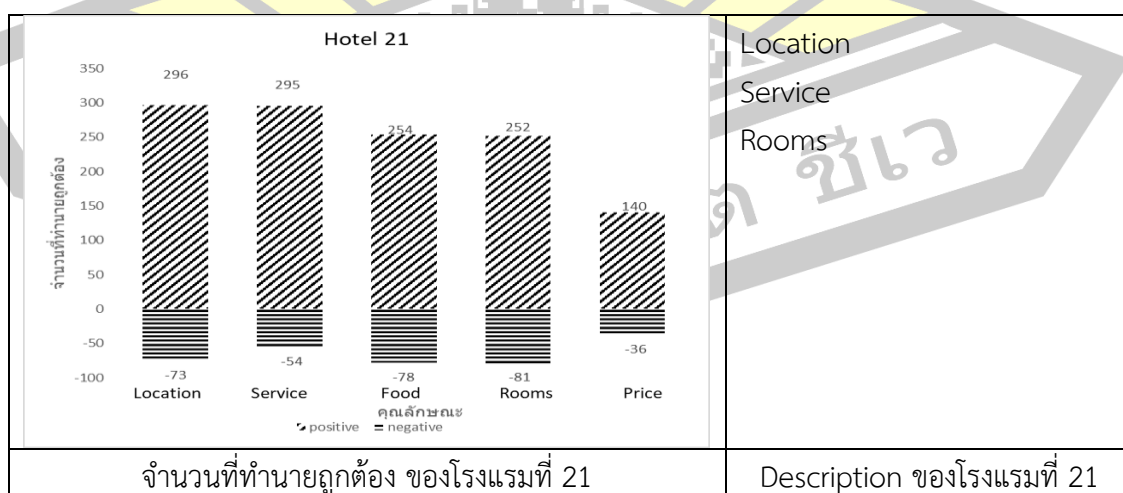
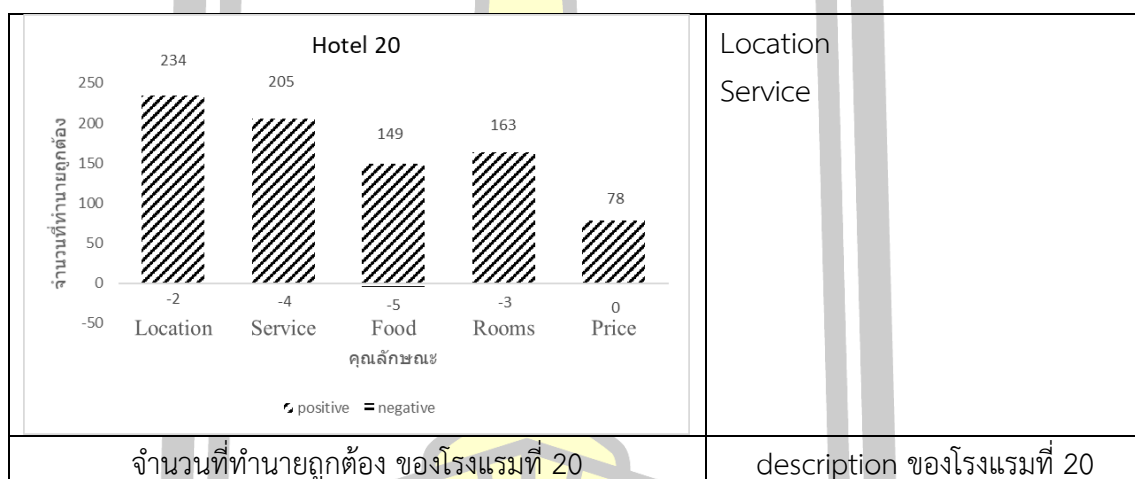
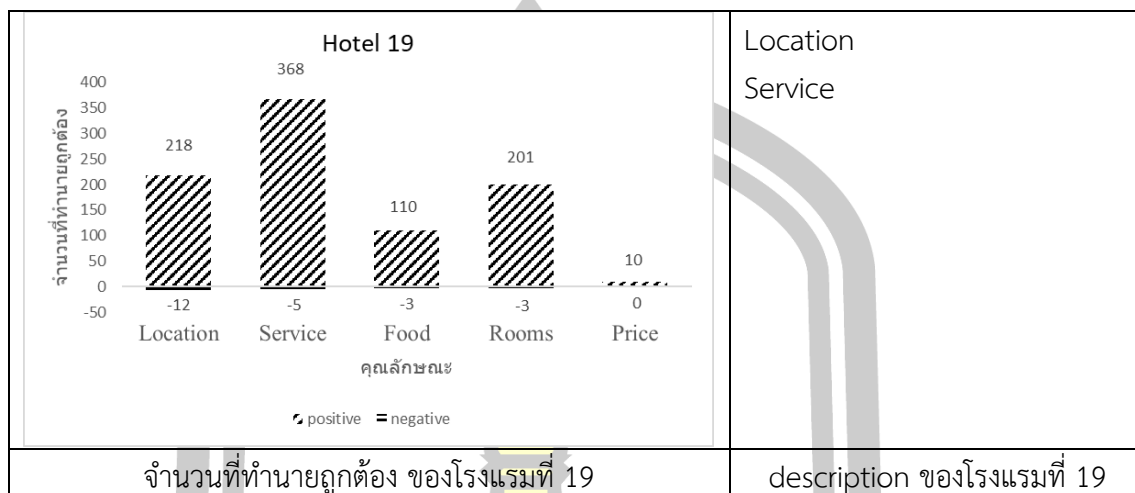


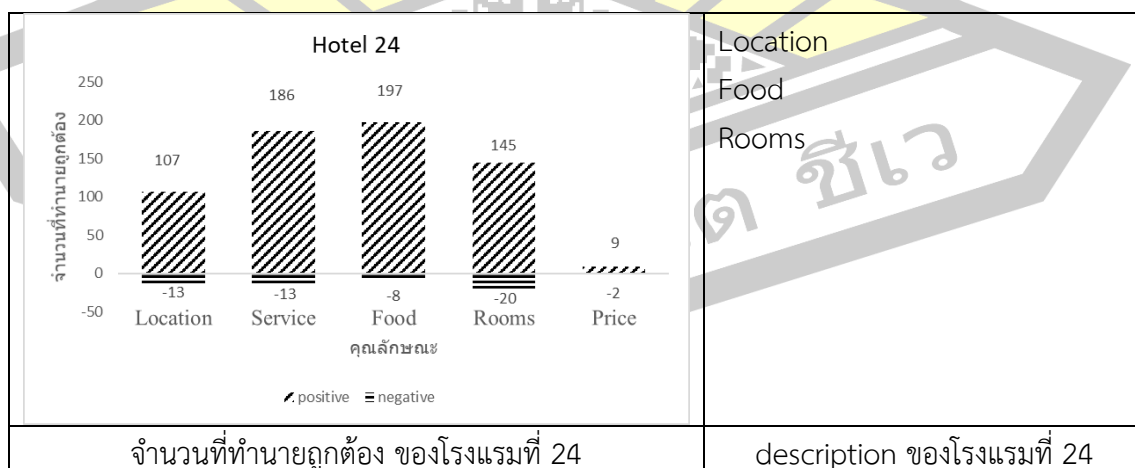
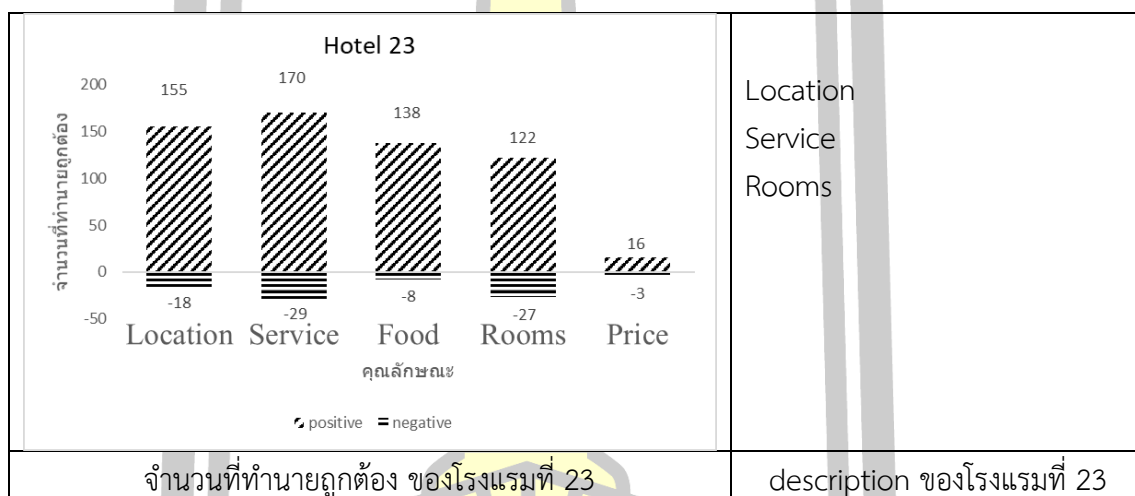
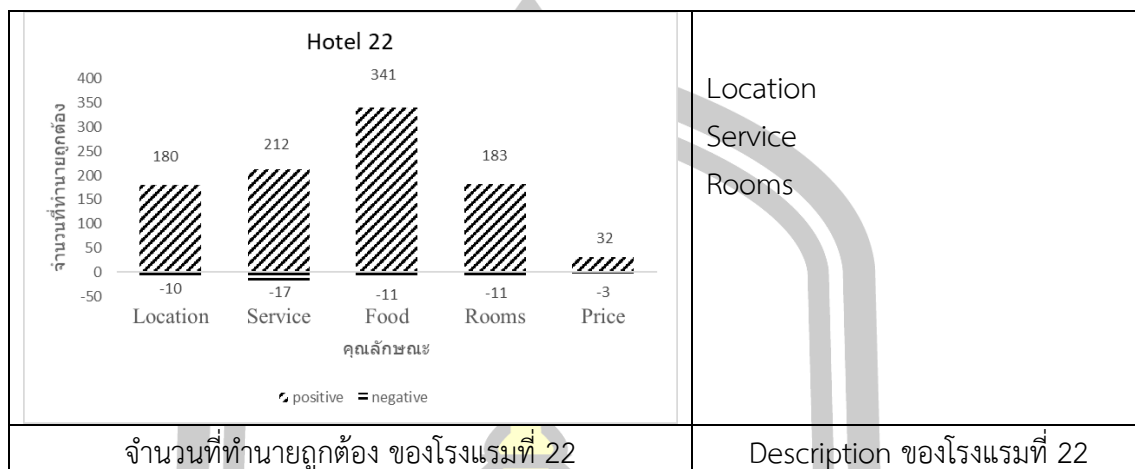


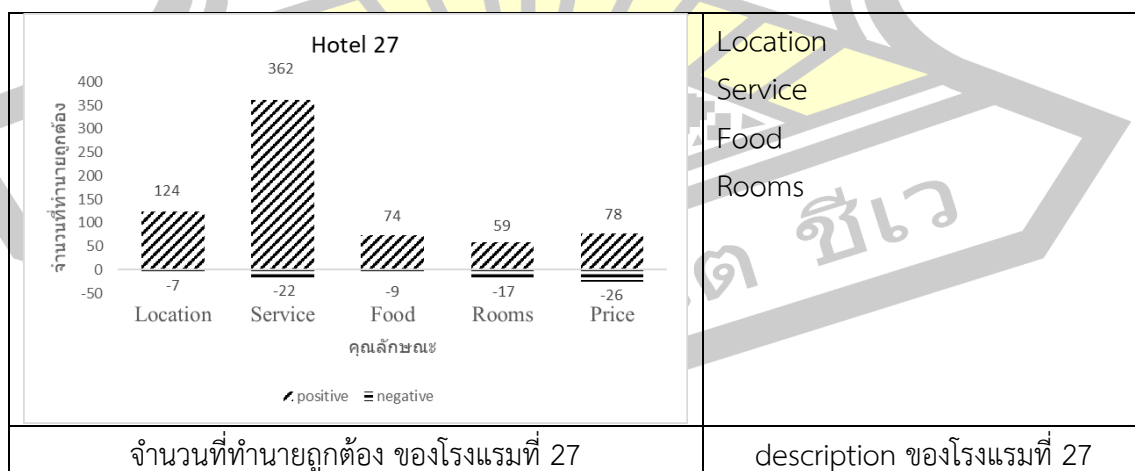
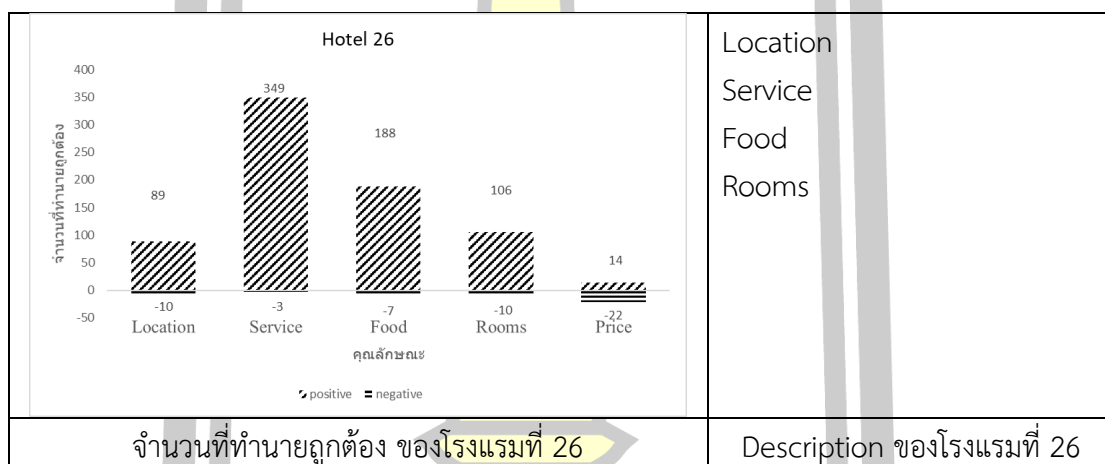
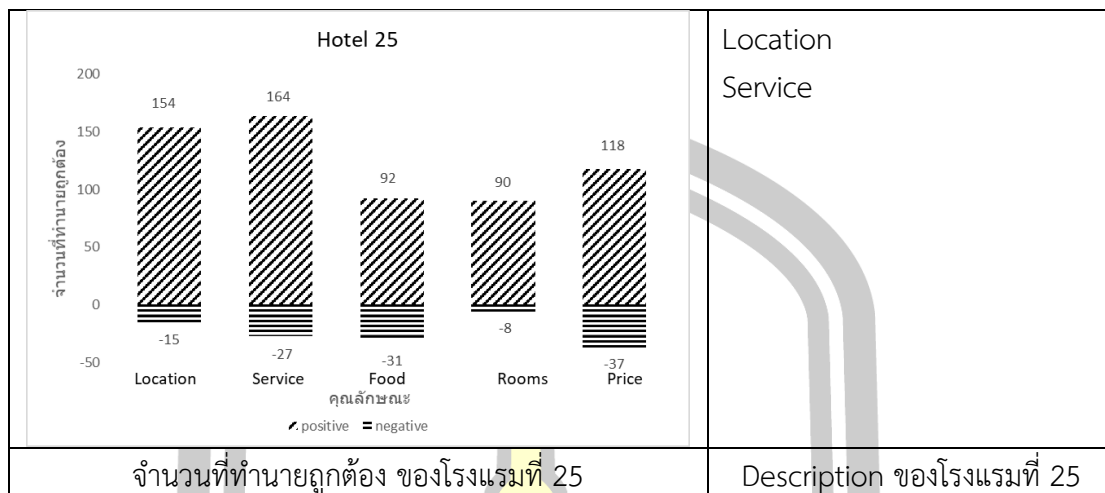
Hotel 10 จำนวนที่ทํานายถูกตอง คุณลักษณะ positive negative	Location Service Food Rooms
จำนวนที่ทํานายถูกตอง ของโรงแรมที่ 10	description ของโรงแรมที่ 10
Hotel 11 จำนวนที่ทํานายถูกตอง คุณลักษณะ positive negative	Service Rooms Price
จำนวนที่ทํานายถูกตอง ของโรงแรมที่ 11	description ของโรงแรมที่ 11
Hotel 12 จำนวนที่ทํานายถูกตอง คุณลักษณะ positive negative	Location Service Rooms
จำนวนที่ทํานายถูกตอง ของโรงแรมที่ 12	description ของโรงแรมที่ 12

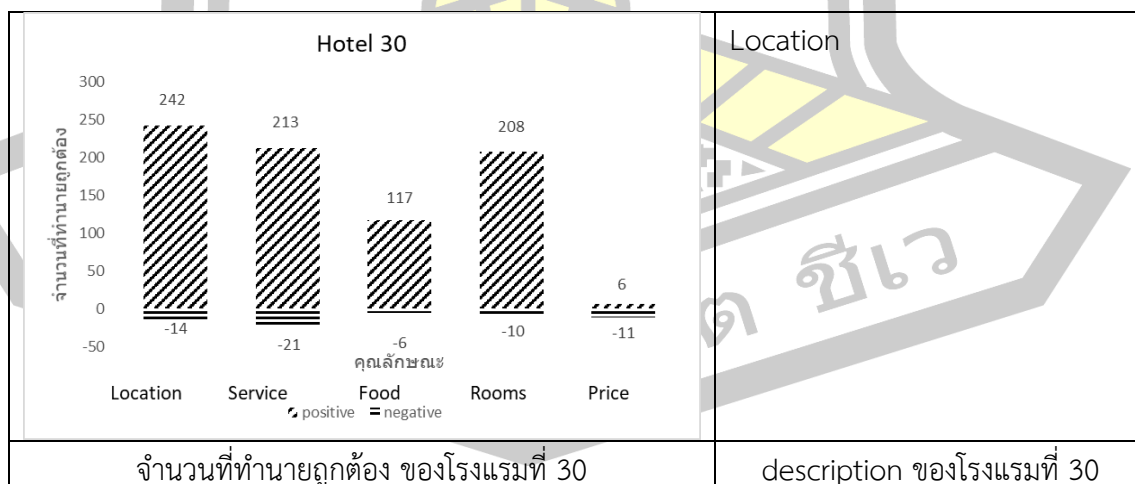
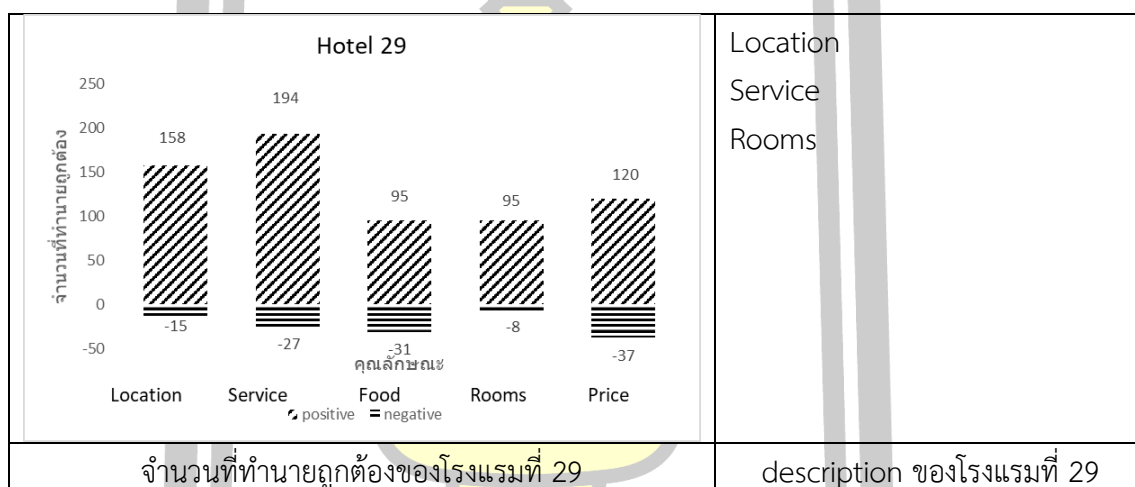
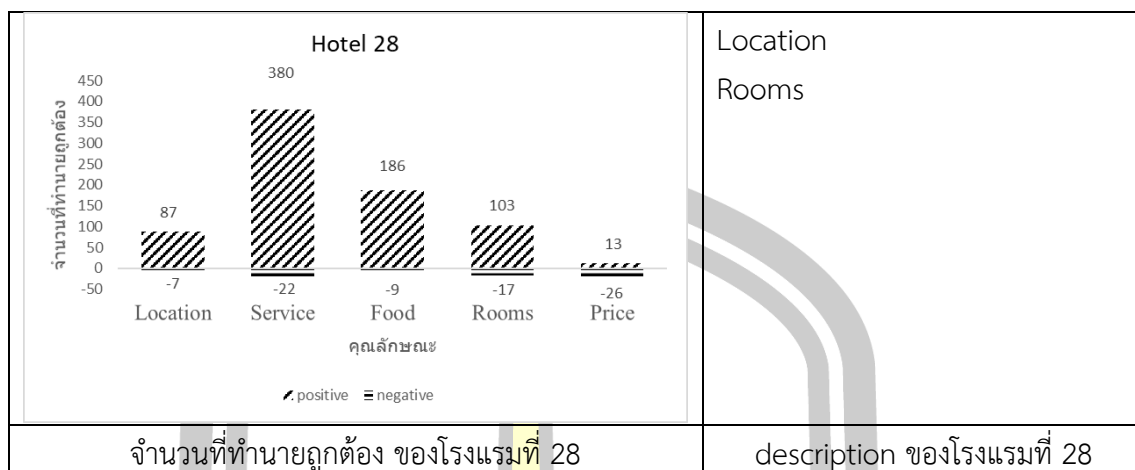












ประวัติผู้เขียน

ชื่อ	มณฑกานต์ ทุมมาวดี
วันเกิด	10 มิถุนายน 2520
สถานที่เกิด	อำเภอเมือง จังหวัดกาฬสินธุ์
สถานที่อยู่ปัจจุบัน	114 ถ.ชัยสุนทร อ.เมือง จ.กาฬสินธุ์ 46000
ตำแหน่งหน้าที่การงาน	พนักงานมหาวิทยาลัย สายวิชาการ
สถานที่ทำงานปัจจุบัน	สาขาวิชาเทคโนโลยีดิจิทัล คณะบริหารศาสตร์ มหาวิทยาลัยกาฬสินธุ์ ถนนเกษตรสมบูรณ์ อำเภอเมือง จังหวัดกาฬสินธุ์ 46000 โทร. 043-81128 โทรสาร 043-813070
ประวัติการศึกษา	พ.ศ. 2542 ปริญญาบริหารธุรกิจบัณฑิต (บธ.บ.) สาขาคอมพิวเตอร์ ธุรกิจ มหาวิทยาลัยภาคตะวันออกเฉียงเหนือ พ.ศ. 2547 ปริญญาบริหารธุรกิจมหาบัณฑิต (บธ.ม.) สาขา บริหารธุรกิจ มหาวิทยาลัยรามคำแหง พ.ศ. 2564 ปริญญาปรัชญาดุษฎีบัณฑิต (ปร.ด.) สาขาวิทยาการ คอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม

พูน ปณ ภัโต ชีเว