

การวิเคราะห์ความคิดเห็นจากทวิตเตอร์
สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่



ปริญญาบริหารธุรกิจมหาบัณฑิต
สาขาวิชาบริหารธุรกิจ
มหาวิทยาลัยแม่โจ้
พ.ศ. 2567

การวิเคราะห์ความคิดเห็นจากทวิตเตอร์
สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวกินเชียงใหม่



การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของความสมบูรณ์ของการศึกษาตามหลักสูตร
ปริญญาบริหารธุรกิจมหาบัณฑิต
สาขาวิชาบริหารธุรกิจ
สำนักบริหารและพัฒนามหาวิชาการ มหาวิทยาลัยแม่โจ้
พ.ศ. 2567

ลิขสิทธิ์ของมหาวิทยาลัยแม่โจ้

การวิเคราะห์ความคิดเห็นจากทวิตเตอร์
สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่

กฤติกา ตี๋พเสน

การค้นคว้าอิสระนี้ได้รับการพิจารณาอนุมัติให้เป็นส่วนหนึ่งของความสมบูรณ์ของการศึกษา
ตามหลักสูตรปริญญาบริหารธุรกิจมหาบัณฑิต
สาขาวิชาบริหารธุรกิจ

พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษา

อาจารย์ที่ปรึกษาหลัก

(อาจารย์ ดร.จักรพงษ์ สุขพันธ์)

วันที่.....เดือน.....พ.ศ.

อาจารย์ที่ปรึกษาร่วม

(อาจารย์ ดร.ประภาพร กิจดำรงธรรม)

วันที่.....เดือน.....พ.ศ.

อาจารย์ที่ปรึกษาร่วม

(ผู้ช่วยศาสตราจารย์ ดร.ดลยา ไชยวงศ์)

วันที่.....เดือน.....พ.ศ.

ประธานอาจารย์ผู้รับผิดชอบหลักสูตร

(ผู้ช่วยศาสตราจารย์ ดร.ปรีดา ศรีนฤธรรม)

วันที่.....เดือน.....พ.ศ.

สำนักบริหารและพัฒนาวิชาการรับรองแล้ว

.....

(รองศาสตราจารย์ ดร.ชัยยศ สัมฤทธิ์สกุล)

รักษาการแทนรองอธิการบดี

วันที่.....เดือน.....พ.ศ.

ชื่อเรื่อง	การวิเคราะห์ความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และรีวิวเชียงใหม่
ชื่อผู้เขียน	นางสาวกฤติกา ดิษฐ์เสนา
ชื่อปริญญา	บริหารธุรกิจมหาบัณฑิต สาขาวิชาบริหารธุรกิจ
อาจารย์ที่ปรึกษาหลัก	อาจารย์ ดร.จักรพงษ์ สุขพันธ์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึก (Sentiment Analysis) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และแฮชแท็กรีวิวเชียงใหม่ และเพื่อสร้างแบบจำลองหัวข้อ (Topic Modeling) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และแฮชแท็กรีวิวเชียงใหม่ โดยพิจารณาจากการแสดงความคิดเห็นบนทวิตเตอร์ ในรูปแบบแฮชแท็กของกินเชียงใหม่ และรีวิวเชียงใหม่ ดังนั้นในการทดสอบแบบจำลองจำแนกความคิดเห็นในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ การรวบรวมข้อมูลความคิดเห็นบนแพลตฟอร์มทวิตเตอร์ในรูปแบบแฮชแท็ก “ของกินเชียงใหม่” และ “รีวิวเชียงใหม่” จำนวน 293 ความคิดเห็น เพื่อสร้างตัวแบบในกระบวนการทดลองดังนี้ 1) การเก็บรวบรวมข้อมูล 2) การจัดเตรียมข้อมูล 3) การวิเคราะห์ความรู้สึก 4) การสร้างแบบจำลองหัวข้อ จากผลการศึกษาพบว่าสามารถวิเคราะห์ความรู้สึกได้ 3 ความรู้สึก ได้แก่ ความรู้สึกเชิงบวก (Positive) จำนวน 232 ความคิดเห็น ความรู้สึกเชิงลบ (Negative) จำนวน 22 ความคิดเห็น และความรู้สึกเป็นกลาง (Neutral) จำนวน 39 ความคิดเห็น และสร้างแบบจำลองหัวข้อเกี่ยวกับความคิดเห็นพบว่ามีคำที่ถูกลำถึงมากที่สุด 10 อันดับแรก ได้แก่ อร่อย รสชาติดี กาแฟ coffee café น่ารัก มุมถ่ายรูป เยอะ บรรยากาศดี อ่างแก้ว ไต๋ถุนบ้าน ซึ่งสามารถนำเสนอข้อมูลการวิเคราะห์ความรู้สึกทางอารมณ์และจำแนกประเด็นข้อมูลเพื่อพัฒนาผลิตภัณฑ์และการให้บริการของธุรกิจอาหารและเครื่องดื่มในจังหวัดเชียงใหม่ต่อไป

คำสำคัญ : การวิเคราะห์ความคิดเห็น, แบบจำลองหัวข้อ, ของกินเชียงใหม่, รีวิวเชียงใหม่

Title	SENTIMENT ANALYSIS ON TWITTER FOR CHIANG MAI FOOD REVIEWS HASHTAG AND CHIANG MAI REVIEWS HASHTAG
Author	Miss Gittiga Tiyupsain
Degree	Master of Business Administration in Business Administration
Advisory Committee Chairperson	Dr. Jakkapong Sukphan

ABSTRACT

The objective of this study is to conduct sentiment analysis on Twitter comment related to the Chiang Mai food hashtag and Chiang Mai review hashtag. The goal is to generate a Topic Modeling based on Twitter comments related to the Chiang Mai food hashtag and the Chiang Mai review hashtag derived from posts on the social media platform Twitter. The article discusses the use of a Chiang Mai food hashtag and provides reviews of Chiang Mai. Therefore, when evaluating the accuracy of the opinion classification model, it is necessary to divide the data into a training dataset. A total of 293 Twitter posts were collected using "Chiang Mai food" and "Chiang Mai review" hashtags to establish a model for the experimental approach with this process: 1) Gathering of information 2) Data preprocessing 3) Sentiment Analysis 4) Topic Modeling. The study revealed that emotions can be categorized into three levels: Positive, Negative or Neutral. Additionally, a topic modeling created based on the comments revealed the top 10 themes based on the most frequently cited phrases. These subjects include descriptors such as delicious, nice flavor, coffee, coffee café, cute, many good spots to take photos, pleasant atmosphere, Ang Kaew, and Taithoonbaan. The study findings may be used to assess emotions and categorize data points, which can then be used to enhance the development of goods and services in the food and beverage industry in Chiang Mai province.

Keywords : Sentiment Analysis, Topic Modeling, Chiang Mai food, Chiang Mai reviews

กิตติกรรมประกาศ

งานวิจัยฉบับนี้ สำเร็จได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างสูงจาก อาจารย์ ดร.จักรพงษ์ สุขพันธ์ อาจารย์ที่ปรึกษาหลัก ที่กรุณาให้คำแนะนำปรึกษาตลอดจนตรวจสอบ ปรับปรุงแก้ไข ข้อบกพร่องต่าง ๆ ที่เกิดขึ้นระหว่างดำเนินการวิจัย ให้งานวิจัยนี้ถูกต้องสมบูรณ์ยิ่งขึ้น และขอกราบ ขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ อาจารย์ ดร.ประภาพร กิจดำรงธรรม และผู้ช่วยศาสตราจารย์ ดร.ดลยา ไชยวงศ์ และคณาจารย์ คณะบริหารธุรกิจ สาขาวิชาบริหารธุรกิจ ทุกท่านที่ได้อบรมและสั่งสอนให้วิชา ความรู้เพื่อนำมาต่อยอดและให้คำแนะนำ

ขอขอบพระคุณสำนักงานบริหารและพัฒนาวิชาการ มหาวิทยาลัยแม่โจ้ เจ้าหน้าที่ที่เกี่ยวข้อง ที่ประสานงานให้สำเร็จด้วยดี

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณบิดามารดา ที่สนับสนุนเรื่องการศึกษาในทุก ๆ ด้านตลอดจน กระทั่งทุกอย่างประสบความสำเร็จ

กฤติกา ตีพิมพ์เสน

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญรูปภาพ.....	ฌ
บทที่ 1 บทนำ	1
1. ความสำคัญของปัญหา.....	1
2. วัตถุประสงค์ของการวิจัย.....	3
3. ขอบเขตของการวิจัย.....	3
4. ประโยชน์ที่คาดว่าจะได้รับ.....	4
5. นิยามศัพท์.....	4
บทที่ 2 การตรวจเอกสาร	5
1. แนวคิดและทฤษฎี.....	5
2. งานวิจัยที่เกี่ยวข้อง	17
3. กระบวนการในการทำวิจัย.....	26
บทที่ 3 วิธีการวิจัย	28
1. ประชากรและกลุ่มตัวอย่าง.....	28
2. เครื่องมือในการวิจัย.....	28
3. วิธีการเก็บรวบรวมข้อมูล	30
4. การวิเคราะห์ข้อมูล.....	33

5. การประเมินแบบจำลอง..... 37

บทที่ 4 ผลการวิจัย..... 38

 1. ผลการศึกษาการวิเคราะห์ความรู้สึก (Sentiment Analysis)..... 38

 2. ผลการศึกษาการสร้างแบบจำลองหัวข้อ (Topic Modeling)..... 44

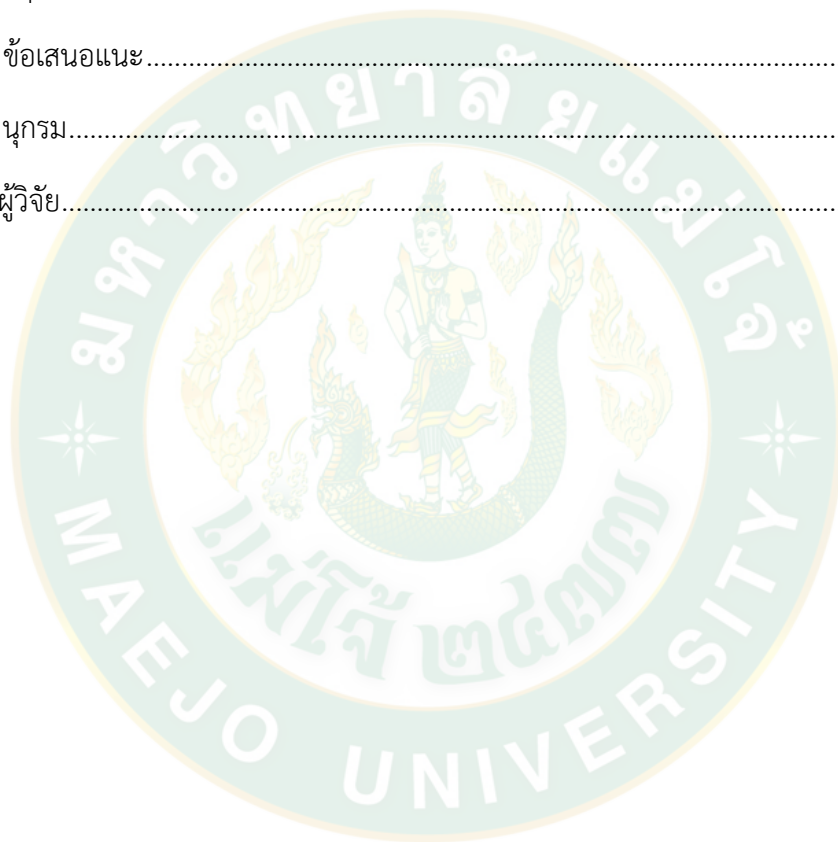
บทที่ 5 สรุปและอภิปรายผล..... 46

 1. สรุปและอภิปรายผล..... 46

 2. ข้อเสนอแนะ..... 49

บรรณานุกรม..... 50

ประวัติผู้วิจัย..... 54



สารบัญตาราง

	หน้า
ตารางที่ 1 ตัวอย่างความคิดเห็น.....	32
ตารางที่ 2 ตัวอย่างผลลัพธ์การจำแนกความรู้สึกเชิงบวก (Positive)	44
ตารางที่ 3 ตัวอย่างผลลัพธ์การจำแนกความรู้สึกเชิงลบ (Negative).....	44
ตารางที่ 4 ตัวอย่างผลลัพธ์การจำแนกความรู้สึกเป็นกลาง (Neutral)	45



สารบัญรูปภาพ

	หน้า
ภาพที่ 1 ความคิดเห็นของผู้บริโภค	30
ภาพที่ 2 ความคิดเห็นของผู้บริโภค	31
ภาพที่ 3 เก็บข้อมูลผ่านโปรแกรม Microsoft Excel.....	31
ภาพที่ 4 แสดงขั้นตอนการดำเนินงานวิจัยและการออกแบบจำลอง.....	32
ภาพที่ 5 ตัวอย่างข้อความความคิดเห็น	33
ภาพที่ 6 ตัวอย่างข้อความที่ถูกตัดคำในรูปแบบของไฟล์ XLSX.....	34
ภาพที่ 7 ตัวอย่างข้อความที่ถูกตัดคำ.....	34
ภาพที่ 8 การสกัดคุณลักษณะของข้อความโดยโปรแกรม Orange Data mining	35
ภาพที่ 9 ตัวอย่างการใช้โปรแกรม S-Sense วิเคราะห์ความรู้สึก (Sentiment Analysis).....	35
ภาพที่ 10 การสร้างแบบจำลองโมเดลโดยใช้อัลกอริทึมการจำแนก 3 อัลกอริทึม ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree.....	36
ภาพที่ 11 การคำนวณหาค่าน้ำหนักของคำ (IDF) โดยโปรแกรม Orange Data mining	36
ภาพที่ 12 กลุ่มรวมของคำ (Word Cloud)	37
ภาพที่ 13 แบบจำลองความรู้สึกเชิงบวก (Positive).....	39
ภาพที่ 14 แบบจำลองความรู้สึกเชิงลบ (Negative).....	40
ภาพที่ 15 แบบจำลองความรู้สึกเป็นกลาง (Neutral).....	41
ภาพที่ 16 แบบจำลองค่าเฉลี่ยความรู้สึกรวม (Total).....	42
ภาพที่ 17 กลุ่มรวมของคำ (Word Cloud)	45

บทที่ 1

บทนำ

1. ความสำคัญของปัญหา

หลังสถานการณ์การแพร่ระบาดของโควิด-19 (COVID-19) ในปี 2565 ประเทศไทยและอีกหลาย ๆ ประเทศ สามารถควบคุมการแพร่ระบาดของเชื้อไวรัส COVID-19 ได้ดี ทำให้การติดเชื้อทุเลาลง มีแนวโน้มปรับตัวดีขึ้น ส่งผลให้รัฐบาลแต่ละประเทศ รวมถึงประเทศไทยได้ผ่อนคลายมาตรการควบคุม ส่งผลให้ภาพรวมการท่องเที่ยวในประเทศไทยปรับตัวดีขึ้นจากการผ่อนคลายมาตรการควบคุมการเดินทาง และการอนุญาตให้กิจการต่าง ๆ กลับมาดำเนินการได้ ประชาชนได้รับวัคซีนเพื่อสร้างภูมิคุ้มกันหมู่ อีกทั้งมีความพร้อมในการรับมือที่จะอยู่ร่วมกับโควิด-19 (COVID-19) ทำให้ความวิตกกังวลของคนไทยคลี่คลายลง กล้าที่จะออกเดินทางท่องเที่ยว แม้ยังคงอยู่ภายใต้มาตรการ Social Distancing ก็ตามและล่าสุดประเทศไทยประกาศให้โควิด-19 (COVID-19) กลายเป็นโรคประจำถิ่น ทำให้นานาประเทศรวมทั้งประเทศไทยประกาศปลดล็อก และผ่อนคลายมาตรการต่าง ๆ ตามเงื่อนไขของแต่ละประเทศ พร้อมกับเปิดรับนักท่องเที่ยวเข้าประเทศแบบไม่มีเงื่อนไข ทำให้ผู้คนที่ต่างประเทศ รวมถึงประเทศไทยต่างวางแผนออกเดินทางท่องเที่ยวเป็นจำนวนมาก บรรยากาศการเดินทางท่องเที่ยวเริ่มกลับมาคึกคักอีกครั้ง ถึงแม้จะอยู่ในระดับต่ำกว่าช่วงก่อนสถานการณ์การแพร่ระบาดของโควิด-19 (COVID-19)

ความพร้อมของประเทศไทยทางด้านแหล่งท่องเที่ยวที่มีความหลากหลาย สามารถตอบสนองความต้องการของนักท่องเที่ยวในทุกกลุ่ม โดยประเทศไทยมีชื่อเสียงทั่วระดับโลกในด้านการท่องเที่ยวทางธรรมชาติที่อุดมสมบูรณ์ เช่น ทะเล และชายหาดที่สวยงามในจังหวัดสุราษฎร์ธานี ชลบุรี ระยอง ตราด พังงา ภูเก็ต และกระบี่ ภูเขา อุทยานแห่งชาติต่าง ๆ ในจังหวัดเชียงใหม่ น่าน เพชรบูรณ์ รวมถึงการท่องเที่ยวทางการแพทย์ และการท่องเที่ยวเพื่อความบันเทิงที่มีหลายรูปแบบ อาทิ โรงละคร แหล่งช้อปปิ้ง สวนสนุก และสถานบันเทิงต่าง ๆ ที่สามารถรองรับนักท่องเที่ยวได้ทุกระดับ ตั้งแต่นักท่องเที่ยวที่มีรายได้สูงที่สามารถใช้จ่ายในระดับสูงไปจนถึงนักท่องเที่ยวที่มีงบประมาณจำกัด จึงกล่าวได้ว่าประเทศไทยเป็นศูนย์รวมแหล่งท่องเที่ยวที่ครบวงจรที่สามารถตอบโจทย์ทุกกลุ่มนักท่องเที่ยว (กาญจนา จันทร์ชิต และคณะ, 2565)

หลังสถานการณ์การแพร่ระบาดของโควิด-19 (COVID-19) การท่องเที่ยวในประเทศไทยฟื้นตัวดีขึ้นอย่างต่อเนื่อง ส่งผลในเชิงบวกต่อธุรกิจร้านอาหาร ผู้บริโภคส่วนใหญ่กลับมาดำเนินชีวิตนอกบ้านตามปกติ ทำให้มีการใช้บริการร้านอาหาร คาเฟ่เครื่องดื่ม รวมถึงจัดกิจกรรมสร้างสรรค์เพิ่มมากขึ้น ซึ่งส่งผลดีต่อร้านค้าร้านอาหารอย่างมาก

ระยะเวลาในช่วงสถานการณ์การแพร่ระบาดของโควิด-19 (COVID-19) ทุกคนล้วนต้องปรับตัวเพื่ออยู่รอด มีแพลตฟอร์มเดลิเวอรีเข้ามาเป็นข้อต่อสำคัญที่ช่วยให้กิจการร้านอาหารยังสามารถไปต่อได้ท่ามกลางวิกฤต และปรับเปลี่ยนพฤติกรรมกรรมการบริโภคของผู้คนให้ต่างไปจากเดิม ส่งผลต่อการฟื้นตัวของธุรกิจร้านอาหารที่มีรูปแบบที่ต่างกันไป โดยกลุ่มผู้ประกอบการที่ประสบความสำเร็จในการทำตลาดผ่านแพลตฟอร์มเดลิเวอรีจะมีการฟื้นตัวที่ดีและสามารถขยายฐานลูกค้าได้เพิ่มขึ้น ดังนั้น การสั่งอาหารผ่านแพลตฟอร์มต่าง ๆ จึงกลายเป็นพฤติกรรมปกติของผู้คนในปัจจุบัน จำนวนผู้ใช้บริการจึงไม่เปลี่ยนแปลงอย่างมีนัย ดังนั้น ทิศทางการเติบโตจึงขึ้นอยู่กับความสามารถในการบริหารจัดการด้านการตลาดให้ตอบโจทย์ผู้บริโภคมากที่สุด ท่ามกลางภาวะต้นทุนและค่าครองชีพที่ปรับตัวสูง (ศุภย์วิเคราะห์เศรษฐกิจ, 2565)

ในยุคนี้เทคโนโลยีเข้ามามีบทบาท ส่งผลต่อชีวิตประจำวันที่จะช่วยให้การดำเนินชีวิตสะดวกสบายมากขึ้น ในการทำธุรกิจก็เช่นกัน เทคโนโลยีเข้ามามีส่วนช่วยให้ทุกอย่างง่ายขึ้นทั้งด้านผู้ขาย และด้านผู้ซื้อ เมื่อเรามีอินเทอร์เน็ตที่เป็นเทคโนโลยีหนึ่งที่สามารถเชื่อมต่อผ่านช่องทางออนไลน์ ทุกคนสามารถเข้าถึงได้จากการสืบค้นด้วยคีย์เวิร์ด (Keyword) เช่น ต้องการทำอะไร ต้องการทานอะไร ต้องการทราบอะไร เป็นต้น ไม่ว่าจะต้องการข้อมูลด้านใด ข้อมูลในอินเทอร์เน็ตนั้นมีเยอะมาก และทุกคนสามารถเข้าถึงได้

จังหวัดเชียงใหม่เป็นหนึ่งในจังหวัดที่น่าเดินทางไปท่องเที่ยว และเป็นจังหวัดที่มีการค้นหาในอินเทอร์เน็ตมากที่สุด ด้วยปัจจุบันมีอินเทอร์เน็ตที่ง่ายต่อการสืบค้น ผู้คนส่วนมากจะวางแผน และหาข้อมูลก่อนเสมอ ไม่ว่าจะต้องการซื้อของ ต้องการเดินทางไปสถานที่ไหน หรือต้องการสิ่งใด ก็มักจะอ่านจากการแสดงความคิดเห็น (Comment) คุปพิจารณา (Reviews) เนื่องจากจะให้เห็นภาพ วิดีโอ และคำแนะนำจากผู้ที่เคยซื้อไปก่อนหน้า ผู้ใช้งานจะใส่คำอธิบายไว้พร้อมใส่คีย์เวิร์ด (Keyword) ของสิ่งนั้น ๆ ข้อมูลเหล่านี้เป็นประโยชน์ และเป็นส่วนสำคัญที่ใช้เป็นเหตุผลในการตัดสินใจซื้ออีกด้วย

ปัจจุบันมีผู้ใช้งานอินเทอร์เน็ตจำนวนมาก ซึ่งมีแนวโน้มขยายตัวอย่างต่อเนื่อง ทำให้ผู้ประกอบการร้านอาหารและเครื่องดื่มในจังหวัดเชียงใหม่สร้างเพจเพื่อนำเสนอร้านของตนผ่านช่องทางออนไลน์เพิ่มขึ้น เพื่อเป็นการโปรโมทให้ลูกค้าเข้าถึงได้ง่าย จากเมื่อก่อนที่ขายเฉพาะออฟไลน์ (หน้าร้าน) แต่ในขณะเดียวกันแอปพลิเคชันไม่ว่าจะ Facebook, Twitter, Tiktok, Instagram, Pantip, Wongnai, Youtube และแพลตฟอร์มต่าง ๆ มีการปิดกั้นการมองเห็น ทำให้ไม่สามารถเจอเพจได้ในอันดับต้นบนฟีด เพราะไม่ได้ซื้อโฆษณาเพื่อยิงไปสู่กลุ่มเป้าหมาย จึงต้องมีค่าใช้จ่ายในส่วนนี้เพื่อทำให้ข้อมูลของผู้ประกอบการถูกส่งไปยังกลุ่มเป้าหมาย ช่วยให้กลุ่มเป้าหมายสามารถเจอเพจได้ในฟีดอันดับแรก ๆ แต่ก็ยังไม่สามารถการันตีว่าจะทำให้ยอดขายเพิ่มขึ้น User Generated Content : UGC เลยเป็นทางออกในยุคปัจจุบันนี้ จะเรียกง่าย ๆ ว่าการบอกต่อ ปากต่อปาก (Solomon, 2011) ก็ว่าได้ แต่ต่างกันที่ UGC บอกต่อผ่านช่องทางออนไลน์หรือโซเชียลมีเดีย ซึ่งไม่ว่าใครก็

สามารถเข้าถึงข้อมูลได้ ถึงอย่างไรก็ตามด้วยข้อมูลที่มีมากมายทำให้เราอยากอ่านข้อมูลได้ครบทั้งหมด ถึงแม้สามารถทำได้เราก็ต้องแลกด้วยเวลา และค่าใช้จ่ายที่นับไม่ถ้วน

การรวบรวมข้อมูลมหาศาลจากช่องทางออนไลน์ใช้เวลาและค่าใช้จ่ายที่สูงตามจำนวนข้อมูล การใช้แนวคิดเหมืองข้อมูล (Data Mining) ก็เป็นทางเลือกที่ดีในการใช้เพื่อเป็นการเก็บรวบรวมข้อมูลที่เยอะด้วยเทคนิคการประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP ในการวิเคราะห์ความรู้สึก (Sentiment Analysis) ในการสกัดข้อความที่เป็นความคิดเห็น และแบ่งกลุ่มข้อความโดยการวิเคราะห์แบบจำลองหัวข้อ (Topic Modeling) ที่เป็นเทคนิคการเรียนรู้ Machine Learning ในการวิเคราะห์ข้อมูลตัวอักษรเพื่อกำหนดกลุ่มคำที่มีความคล้ายคลึงมารวมเป็นกลุ่มเดียวกัน ทำให้เราสามารถนำกลุ่มคำที่ได้จากการวิเคราะห์ไปต่อยอดในกิจการได้

ในงานวิจัยนี้จึงได้ศึกษาการทำเหมืองข้อมูล (Data Mining) เพื่อทำการวิเคราะห์ความรู้สึก (Sentiment Analysis) ที่สามารถวิเคราะห์ความคิดเห็นนั้นว่าบ่งบอกความรู้สึกในเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral) และเพื่อสร้างแบบจำลองหัวข้อ (Topic Modeling) จากการสกัดข้อมูล ตัดคำที่ไม่มีความหมาย หรือคัดแยกข้อมูล กำหนดกลุ่มคำที่มีความคล้ายคลึงมารวมเป็นกลุ่มเดียวกัน เพื่อให้ผู้ประกอบการที่เกี่ยวข้องกับของกินในจังหวัดเชียงใหม่ได้นำไปพัฒนา ปรับปรุง แก้ไขให้ดีขึ้น สำหรับรองรับลูกค้าเดิม และเป็นทางเลือกที่ดีให้กับลูกค้าใหม่ในอนาคต

2. วัตถุประสงค์ของการวิจัย

2.1 เพื่อวิเคราะห์ความรู้สึก (Sentiment Analysis) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่

2.2 เพื่อสร้างแบบจำลองหัวข้อ (Topic Modeling) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่

3. ขอบเขตของการวิจัย

3.1 ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ข้อมูลจากความคิดเห็นเกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ ชุดข้อมูลในช่วงวันที่ 1 พฤษภาคม 2565 ถึง วันที่ 30 เมษายน 2566

4. ประโยชน์ที่คาดว่าจะได้รับ

4.1 เพื่อทราบผลวิเคราะห์แบบจำลองหัวข้อเกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์ สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กกรีวิวเชียงใหม่

4.2 เพื่อทราบผลวิเคราะห์ความรู้สึกเกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับ แฮชแท็กของกินเชียงใหม่ และแฮชแท็กกรีวิวเชียงใหม่ นั้นมีความรู้สึกของข้อความ เป็น ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral)

4.3 เพื่อนำผลลัพธ์ของการวิเคราะห์เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับ แฮชแท็กของกินเชียงใหม่ และแฮชแท็กกรีวิวเชียงใหม่ มาเป็นแนวทางแก้ไข ปรับปรุง พัฒนา ในเรื่อง การทำโฆษณา ประชาสัมพันธ์ หรือออกโปรโมชั่น สำหรับผู้ประกอบการร้านค้าที่เกี่ยวข้องกับของกิน เช่น ร้านอาหาร ร้านขนม และร้านเครื่องดื่ม ในจังหวัดเชียงใหม่

5. นิยามศัพท์

5.1 ทวิตเตอร์ หมายถึง หนึ่งในแพลตฟอร์มออนไลน์ที่สามารถให้ผู้ใช้งานนั้นเข้ามาแสดงความคิดเห็นได้ (ปัจจุบันได้เปลี่ยนชื่อเป็น X)

5.2 แฮชแท็ก (Hashtag) คือ เครื่องมือที่ช่วยแยกข้อมูลประเภทเดียวกันให้อยู่ด้วยกัน เมื่อต้องการที่จะค้นหาข้อมูลอะไร เพียงแค่ค้นหาด้วยการพิมพ์ # ต่อด้วยคำที่ต้องการค้นหา ก็จะพบโพสต์และความคิดเห็นที่เกี่ยวข้องกับสิ่งที่ค้นหา

5.3 ความคิดเห็น (Comment) หมายถึง ข้อความแสดงความคิดเห็นของผู้ใช้งานแพลตฟอร์มในทวิตเตอร์

5.4 ข้อความวิจารณ์ (Reviews) หมายถึง การวิจารณ์ของผู้ใช้งานแพลตฟอร์มในทวิตเตอร์

5.5 เนื้อหาที่ถูกสร้างขึ้นจากผู้บริโภค (User Generated Content) หมายถึง ผู้ที่ใช้งานแพลตฟอร์มในทวิตเตอร์

5.6 แบบจำลองหัวข้อ (Topic Modeling) หมายถึง ใจความสำคัญของข้อมูลทั้งหมด ที่สรุปว่าข้อมูลนั้นกล่าวถึงเรื่องอะไร

5.7 การวิเคราะห์ความรู้สึก (Sentiment Analysis) หมายถึง การวิเคราะห์ความรู้สึกหรืออารมณ์ ได้แก่ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral)

บทที่ 2

การตรวจเอกสาร

การวิจัยเรื่อง การวิเคราะห์ความคิดเห็นจากทวีตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง โดยมีขอบข่ายการศึกษาตามลำดับ ดังนี้

1. แนวคิดและทฤษฎี
2. งานวิจัยที่เกี่ยวข้อง

1. แนวคิดและทฤษฎี

ในปัจจุบันนี้กลยุทธ์ทางการตลาดผ่านช่องทางออนไลน์ (Social Media Marketing) มีผลต่อการตัดสินใจซื้อค่อนข้างมาก ถ้าเข้าใจพฤติกรรมผู้บริโภค จะสามารถรู้ใจลูกค้าได้ ฉะนั้นเราจะเข้าใจความต้องการของลูกค้า ก็ต้องมีการเก็บข้อมูล ซึ่งข้อมูลก็เกิดการเปลี่ยนแปลงรายวัน และ/หรือเรื่อย ๆ เท่ากับข้อมูลนั้นมีจำนวนมาก บางกลุ่มนิยมเก็บข้อมูลแบบออนไลน์เพราะเป็นความเคยชิน และคิดว่าสะดวกกว่า เช่น การไปสังเกต (Observe) การทำแบบสำรวจ (Survey) การทำแบบสอบถาม (Questionnaire) การสัมภาษณ์กลุ่มเป้าหมาย (Focus Group) หรือมีการสัมภาษณ์เชิงลึก (Depth Interview)

1.1 พฤติกรรมผู้บริโภค

พฤติกรรมผู้บริโภค คือ พฤติกรรมของผู้ซื้อสินค้า หรือบริการ ในการตัดสินใจ เพื่อซื้อสินค้านั้นเอง โดยจะมีสิ่งจูงใจ พฤติกรรมต่าง ๆ ไม่ว่าจะเป็นพฤติกรรมการกิน พฤติกรรมการเดินทาง พฤติกรรมสุขภาพ และอื่น ๆ จะเห็นได้ว่ามีการเปลี่ยนแปลงของพฤติกรรมอยู่ตลอดเวลา และเปลี่ยนแปลงมากในปัจจุบัน

ในอดีตการสังเกตพฤติกรรม นักวิชาการมักศึกษาพฤติกรรมโดยเริ่มจากการไปสังเกต (Observe) (สุภางค์ จันทวานิช, 2547) การสังเกตคือการเฝ้าดูสิ่งที่เกิดขึ้นอย่างใส่ใจและมีระเบียบวิธีเพื่อวิเคราะห์หรือหาความสัมพันธ์ของสิ่งที่เกิดขึ้นนั้นกับสิ่งอื่น ประเภทของการสังเกต การสังเกตที่ใช้ในการวิจัยเชิงคุณภาพมี 2 แบบ

1) การสังเกตแบบมีส่วนร่วม (Participant Observation) คือการสังเกตที่ผู้สังเกตเข้าไปใช้ชีวิตร่วมกับกลุ่มคนที่ถูกศึกษา มีการทำกิจกรรมร่วมกัน จนผู้ถูกศึกษายอมรับว่าผู้สังเกตมีสถานภาพบทบาทเช่นเดียวกับตน ผู้สังเกตจะต้องปรับตัวให้เข้ากับกลุ่มคนที่ศึกษา โดยอาจเข้าไปฝังตัวอยู่ในเหตุการณ์ เข้าไปอาศัยอยู่ในชุมชนเป็นเวลานาน

2) การสังเกตแบบไม่มีส่วนร่วม (Non-Participant Observation) คือการสังเกตที่ผู้วิจัยเฝ้าสังเกตอยู่วงนอก ไม่เข้าไปร่วมในกิจกรรมที่ทำอยู่

กล่าวโดยสรุป ทั้งการสังเกตแบบมีส่วนร่วมและการสังเกตแบบไม่มีส่วนร่วมนั้น ต่างมีวัตถุประสงค์ เพื่อสังเกตพฤติกรรมและเหตุการณ์ เพื่อนำมาหาความสัมพันธ์และความหมายของปรากฏการณ์ที่เกิดขึ้น

การทำแบบสำรวจ (Survey) แบบสำรวจเป็นเครื่องมือที่ใช้สำหรับเก็บรวบรวมความคิดเห็นของประชาชน เพื่อให้ผู้ออกแบบและพัฒนานโยบายสาธารณะสามารถนำความคิดเห็นนั้นไปใช้เป็นการผลิตนโยบาย การสร้างแบบสำรวจเพื่อรวบรวมความคิดเห็นจึงเป็นส่วนสำคัญที่ทำให้ประชาชนสามารถมีส่วนร่วมในการออกแบบนโยบายสาธารณะที่เป็นประโยชน์ต่อทุกคนในประเทศ (กานต์ ศุภจารูทิศดี, 2563)

แบบสอบถาม (Questionnaire) คือ แบบฟอร์มที่รวมชุดคำถาม ที่ถูกรวบรวมไว้อย่างมีหลักเกณฑ์และเป็นระบบ เพื่อใช้วัดสิ่งที่ผู้วิจัยต้องการจะวัดจากกลุ่มตัวอย่างหรือประชากรเป้าหมาย เพื่อให้ได้ข้อเท็จจริงทั้งในอดีต ปัจจุบันและการคาดคะเนเหตุการณ์ในอนาคต แบบสอบถามสามารถรวบรวมข้อมูลต่าง ๆ ของกลุ่มตัวอย่าง เช่น ข้อมูลส่วนบุคคล พฤติกรรม ความพึงพอใจ ความต้องการ ความคิดเห็น ข้อเสนอแนะ เป็นต้น แบบสอบถามที่สร้างขึ้นนี้จะนำไปให้กลุ่มตัวอย่างตอบ เพื่อนำข้อมูลมาประมวลผลและวิเคราะห์ต่อไป มีการใช้กันอย่างกว้างขวางทั้งในทางธุรกิจ หน่วยงานภาครัฐ องค์กรที่ไม่แสวงหากำไร เพราะสามารถรวบรวมข้อมูลโดยตรงจากกลุ่มตัวอย่าง ทำให้นำมาใช้ในการใช้ประโยชน์ได้มาก

การสัมภาษณ์กลุ่มเป้าหมาย (Focus Group) คือการพูดคุยเป็นกลุ่มกับตัวแทนกลุ่มเป้าหมายที่ตรงกับประเด็นที่ต้องการศึกษา โดยผู้เข้าร่วมวิจัยสามารถพูดคุย แสดงความคิดเห็นได้อย่างอิสระเสรี ในขณะเดียวกันฝั่งผู้วิจัยจะเป็นผู้ดำเนินการสนทนากลุ่ม (Moderator) ด้วยการถามคำถามต่าง ๆ เชิญชวนให้ทุกคนได้แสดงออกทางความคิดผ่านการสนทนา ถกประเด็นต่าง ๆ ร่วมกัน หรือจะใช้เครื่องมือต่าง ๆ เป็นตัวช่วยเสริมกระตุ้นความคิด ให้เข้าใจ เห็นภาพที่ชัดเจน และควบคุมรักษาบรรยากาศของการพูดคุย เพื่อให้ได้ข้อมูลที่มีคุณภาพสามารถนำไปวิเคราะห์ต่อได้ ซึ่งการทำ Focus Group มี 2 รูปแบบ ดังนี้

1) การสัมภาษณ์กลุ่ม - Focus Group Interview Focus Group Interview เป็นการสัมภาษณ์กลุ่ม ซักถามเฉพาะเจาะจงซึ่งจะมีความคล้ายกับการสัมภาษณ์เชิงลึก In-depth Interview ตรงที่สัมภาษณ์ลงลึกถึงรายบุคคล เพียงแต่ทำการสนทนาเป็นกลุ่มและไล่เรียงถามลงลึกเป็นคน ๆ ไป โดยเหมาะกับการศึกษาพฤติกรรม ประสบการณ์ การใช้ชีวิต ไลฟ์สไตล์ รสนิยม ความชอบ การบริโภค หรือทดสอบค้นหาแนวทางที่ควรพัฒนาต่อไป

2) การสนทนากลุ่ม - Focus Group Discussion Focus Group Discussion เป็นการสนทนากลุ่มพูดคุยถกถามความคิดเห็นในประเด็นที่สนใจซึ่งจะเป็นวิธีการที่เปิดโอกาสให้ผู้เข้าร่วมสามารถพูดคุยแลกเปลี่ยนความคิดเห็น ประสบการณ์ ทั้งมุมมองที่ตรงกันและต่างกัน ปรัชญาหรือมีปฏิสัมพันธ์โต้ตอบกันเพื่อหาข้อสรุป หรือแนวทางในประเด็น วัตถุประสงค์ที่ตั้งไว้ โดยชุดคำถามส่วนใหญ่จะทำในรูปแบบ Semi-Structure กึ่งมีโครงสร้าง คือ ทำชุดถามเป็นหัวข้อคลุมประเด็นที่ต้องการได้คำตอบ และในระหว่างการสนทนาสามารถเพิ่มหรือลดคำถามได้ตามแนวบริบทเนื้อหาที่ถกหารือกันในกลุ่ม ด้วยโครงสร้างที่ยืดหยุ่นเพื่อให้ได้ผลลัพธ์ตามที่คาดหวัง

การสัมภาษณ์เชิงลึก (In-depth Interview) คือ การสัมภาษณ์เชิงลึก เจาะลึกลงรายละเอียดรายบุคคล ทำให้ได้ข้อมูลที่ลงลึกถึงพฤติกรรม ทักษะคิด มุมมองความคิดเห็น ความเชื่อ ความคิด ความต้องการ ความรู้สึก ความรู้ความเชี่ยวชาญในด้านที่ถนัด และที่สำคัญทำให้ได้ข้อมูลเชิงลึก Insight เข้าใจลูกค้า เข้าใจผู้ใช้ เข้าใจเหตุผลที่อยู่เบื้องหลังของการตัดสินใจ การกระทำต่าง ๆ ลูกค้าคิดอะไร ทำไมลูกค้าถึงซื้อ ไม่ซื้อ เพราะอะไรถึงทำให้ตัดสินใจ หรือทำเช่นนั้น

In-depth Interview with Empathy and Observation เป็นวิธีการสัมภาษณ์ที่ใช้การเข้าใจ มองในมุมมองของผู้ให้ข้อมูล (ผู้ให้สัมภาษณ์หรือผู้ร่วมวิจัย) หรือเรียกว่า Empathy ประกอบกับทักษะการช่างสังเกต Observation จากสีหน้า ท่าทาง การเคลื่อนไหวร่างกาย การตอบสนองต่อคำถามต่าง ๆ จะเป็นอีกหนึ่งตัวช่วยให้พิสูจน์ ตรวจสอบความน่าเชื่อถือของข้อมูล ไปจนถึงเข้าใจในเหตุการณ์ได้ดียิ่งขึ้น ซึ่งการสัมภาษณ์เชิงลึกจะช่วยทำให้รู้ว่า ลูกค้า-ผู้ใช้ต้องการอะไร (Customer Needs) รับฟังเสียงของกลุ่มเป้าหมาย (Customer Voice) นำไปยกระดับธุรกิจ พัฒนาผลิตภัณฑ์ให้ตรงใจผู้ใช้ ออกแบบบริการ สร้างประสบการณ์ นวัตกรรมใหม่ให้ตรงกับพฤติกรรมที่เปลี่ยนแปลงไปของลูกค้า เพิ่มมูลค่า โอกาสความเป็นไปได้ใหม่ ๆ ในอนาคต

ในขณะที่ในยุคดิจิทัลแบบนี้ ด้วยสังคมโลกที่มีการเติบโตตลอดเวลาตามเทคโนโลยีที่มีการพัฒนาอย่างต่อเนื่อง ผู้คนส่วนมากแน่นอนว่าเคยใช้สื่อออนไลน์ในการติดต่อสื่อสาร หรือติดตามสิ่งที่สนใจจากสื่อโซเชียลมีเดีย ไม่เพียงเท่านั้นทุกคนอาจจะเคยเห็นว่าสื่อสังคมออนไลน์ต่าง ๆ จะมีการโพสต์เกี่ยวกับคอมเมนต์ ความรู้สึก คำแนะนำที่เกี่ยวกับสินค้านั้น ๆ หรือบริการต่าง ๆ มากมาย ซึ่งข้อความเหล่านั้นมีเป็นจำนวนมาก ซึ่งสังเกตเห็นผู้คนในปัจจุบันมักตัดสินใจซื้อสินค้าและบริการจากรีวิวหรือเรียกอีกอย่างหนึ่งว่า User Generated Content เราสามารถนำข้อความไปใช้ประโยชน์ได้หลายอย่าง สำหรับการใช้ในการดำเนินธุรกิจ เช่น การรับฟังเสียงสะท้อนของลูกค้า ทุกวันนี้ตามหน้าฟีดของเว็บบอร์ด หรือโซเชียลมีเดียต่าง ๆ มีคนจำนวนมากเข้ามาแสดงความคิดเห็น หรือคอมเมนต์ หรือการบอกกล่าวไปยังเจ้าของผลิตภัณฑ์ เพื่อให้เข้ามาแก้ไขปัญหาให้เขาอย่างเร่งด่วน ข้อความเหล่านั้นมีจำนวนมาก หากปล่อยผ่านจะไม่มีประโยชน์ และบางครั้งอาจจะเป็นโทษเสียหายนต่อธุรกิจ ด้วยข้อความที่มีมากมายมหาศาลเราสามารถจัดการข้อความเหล่านั้นด้วยการรวบรวม

ข้อความเหล่านั้นมาวิเคราะห์เพื่อหา Topic Modeling และการที่เราทำ Sentiment Analysis โดยเทคนิคการทำเหมืองข้อมูล (Data Mining) ได้ จะทำให้เราสามารถคัดกรองข้อความที่มีความรู้สึกที่ไม่ดีมาได้ และยังสามารถเข้าไปแก้ไขปัญหาได้ตรงจุดอีกด้วย เพื่อลดแรงงานคนในการเข้าไปอ่านทุกข้อความ เทคนิคนี้ใช้เวลาและค่าใช้จ่ายน้อยกว่า ทำให้เราประหยัดทั้งเวลาและค่าใช้จ่าย

1.2 เนื้อหาที่ถูกสร้างขึ้นจากผู้บริโภค (User Generated Content)

User Generated Content : UGC คือ การที่ผู้บริโภคหรือลูกค้ากลุ่มเป้าหมาย ลงมือผลิตคอนเทนต์ด้วยตัวเอง ซึ่งจะมีการกล่าวถึงแบรนด์ ที่ผู้บริโภคหรือลูกค้ากลุ่มเป้าหมายมีความสนใจ โดยที่แบรนด์ไม่จำเป็นต้องเสียเงินจ้างผู้บริโภคกลุ่มนี้เลย ซึ่งมีส่วนสำคัญที่ก่อให้เกิดความตั้งใจซื้อ (Bai et al, 2015) โลกอินเทอร์เน็ตมีการพัฒนาอย่างต่อเนื่อง ซึ่งเอื้อต่อการทำไวรัล มาเก้ตติ้ง (Viral Marketing) มีความสามารถในการแพร่กระจาย ข้อมูลที่รวดเร็วและรุนแรงอย่างมาก เนื่องจากระบบอินเทอร์เน็ตมีการเปลี่ยนแปลงรูปแบบการสื่อสาร จากการสื่อสารทางเดียว (One Way) เป็นการสื่อสารแบบสองทาง (Two Way) หมายถึง การที่ผู้ใช้กลายเป็นหนึ่งในผู้สร้างสรรค์เนื้อหาต่าง ๆ ผ่านระบบอินเทอร์เน็ต ซึ่งเนื้อหาเหล่านี้ถูก User Generated Content และกลายเป็นสื่อที่กำลังได้รับความนิยมอย่างมาก ปัจจุบัน User Generated Content มีหลายรูปแบบ เช่น บอร์ดสนทนา : www.pantip.com บล็อก : www.blogger.com สารานุกรมเสรีวิกิพีเดีย : www.wiki.org เครือข่ายสังคมออนไลน์ต่าง ๆ : www.facebook.com, www.twitter.com เว็บไซต์ข่าวต่าง ๆ ที่มีช่องทางให้ผู้ใช้สามารถแสดงความคิดเห็น : www.cnn.com เป็นต้น นอกจากนี้ User Generated Content ยังสามารถพบเห็นได้ทั่วไปในการค้าบนอินเทอร์เน็ต ในรูปแบบของการที่ผู้บริโภคซึ่งเคยซื้อสินค้ามาเขียนคำวิจารณ์หรือให้คะแนน เพื่อเป็นตัวช่วยสำหรับผู้บริโภคคนอื่นในการตัดสินใจซื้อสินค้าในเว็บไซต์นั้น ๆ (วรวิวัฒน์ จันทรศิริวัฒนา, 2557)

ผู้บริโภคในปัจจุบันแตกต่างจากผู้บริโภคในอดีตอย่างมาก เพราะเมื่อจะทำการซื้อสินค้าอะไรในยุคนี้ผู้บริโภคไม่ได้เลือกซื้อเลยทันที หรือทำการตัดสินใจอยู่นานด้วยตัวเองคนเดียว แต่เป็นการที่ผู้บริโภคนั้นเข้าไปหาข้อมูล สอบถามข้อมูลด้วยตัวเองกับเพื่อนหรือคนที่ไม่ได้รู้จักเลยผ่านเว็บบอร์ดหรือหน้าแฟนเพจต่าง ๆ ยิ่งสินค้านั้นมีมูลค่าที่มากการใช้เวลาในการสอบถามและหาข้อมูลก็มากตาม ในตอนนี้ แม้แต่ตัวเราเองนั้นก็หาข้อมูลกับผู้บริโภคด้วยกันในการตัดสินใจที่จะจ่ายเงิน แม้กระทั่งการไปทานร้านอาหารในยุคนี้ ก่อนที่จะไปลองทานอาหารที่ร้าน ทุกคนล้วนมีการอ่านรีวิวจากผู้บริโภคด้วยกันเองทั้งนั้น ผู้บริโภคจึงมองหาสิ่งที่เรียกว่าความน่าเชื่อถือจากผู้บริโภคด้วยกันเอง สิ่งนั้นคือ User Generated Content : UGC ความคิดเห็นของผู้บริโภคในยุคนี้กลายเป็นสิ่งที่ทำให้ผู้บริโภคด้วยกันเองนั้นเกิดความเชื่อถือ เพราะเป็นเรื่องทางจิตวิทยาของผู้บริโภคที่คิดว่าผู้บริโภค

ด้วยกันเองนั้น ไม่ได้มีส่วนได้ส่วนเสียกับแบรนด์ และเสียเงินในการซื้อของหรือใช้บริการนั้นจริง ๆ ทำให้หน้าเชื่อได้ว่าความเห็นนั้นจะบริสุทธิ์ใจในการบอกต่อ (Park et al., 2007)

พฤติกรรมผู้บริโภคในยุคนี้คือการค้นหาหรือดูเนื้อหาจากออนไลน์เหล่านี้ก่อนเริ่มกระบวนการซื้อของทั้งนั้น เพราะผู้บริโภคนั้นไม่ได้อยากเสี่ยงที่จะเสียเงิน โดยรวมถึงเพื่อลดความเสี่ยงที่จะได้สินค้าและบริการที่ไม่ตรงตามความต้องการหรือไม่มีคุณภาพ (O'Connor, 2008)

จากการศึกษาของ Stackla (2021) ได้ทำการหาคำตอบของคำถามที่ว่า ลักษณะคอนเทนต์ประเภทไหนที่โน้มน้าวใจผู้บริโภค จนนำไปสู่การซื้อสินค้าได้มากที่สุด คือ User Generated Content คือคอนเทนต์ที่สร้างความน่าเชื่อถือให้แบรนด์ได้ดีที่สุด

User Generated Content : UGC คือเนื้อหาคอนเทนต์ทุกประเภทที่ถูกสร้างขึ้นมาโดยเหล่าผู้บริโภคตัวจริงและมีความเกี่ยวข้องกับสินค้า บริการหรือแบรนด์โดยตรง ไม่ว่าจะเป็นโพสต์รีวิวสินค้าบน Facebook ภาพถ่ายคู่กับสินค้าที่ถูกโพสต์ลงใน Instagram ทวิตต่าง ๆ บนเว็บไซต์อย่าง Twitter จนไปถึงบทความและกระดานเว็บบอร์ดทั่วไป ซึ่งต้องเป็นคอนเทนต์ที่ไม่ได้เกิดจากคนที่มีความเกี่ยวข้องกับแบรนด์ ไม่ใช่อินฟลูเอนเซอร์ และไม่ใช่คนที่ได้ผลประโยชน์จากแบรนด์นั่นเอง การได้อ่าน User Generated Content เช่น คอนเทนต์จำพวกรีวิวจากผู้ใช้งานจริง ก็จะเป็นการช่วยเพิ่มความน่าเชื่อถือให้สินค้า/แบรนด์ได้เป็นอย่างดี เนื่องจากลูกค้าจะได้เห็น ภาพ วิดีโอ และคำแนะนำของการใช้งาน จากผู้ที่เคยซื้อไปก่อนหน้านี้

User Generated Content : UGC คือเนื้อหาคอนเทนต์ที่สร้างขึ้นโดยกลุ่มลูกค้า เป็นสิ่งที่เจ้าของธุรกิจไม่ควรมองข้าม เนื่องจากผู้บริโภคเชื่อมั่นและวางใจในเนื้อหาประเภทนี้มากกว่าคอนเทนต์จากทางร้านหรือแบรนด์ และยังเป็นเนื้อหาที่ลูกค้าสร้างขึ้นเองด้วยความเต็มใจโดยที่แบรนด์ไม่ต้องจ่ายเงิน ช่วยลดค่าใช้จ่ายในการทำโฆษณาของแบรนด์อีกด้วย (เอ็ก ดิจิทัล, 2564)

นอกจากนี้ ยังมีรูปแบบของการที่ผู้บริโภคมาแบ่งปันประสบการณ์การใช้สินค้า และบริการ โดยการเขียนบทวิจารณ์หรือให้คะแนน Rating เพื่อเป็นตัวช่วยสำหรับผู้บริโภคที่กำลังจะตัดสินใจซื้อสินค้าและบริการต่าง ๆ เช่น Tripadvisor.com, Agoda.com และ Amazon.com เป็นต้น ซึ่ง Agoda.com ถือเป็นเว็บไซต์ที่ผู้บริโภคนิยมใช้บริการจองโรงแรมและที่พัก และยังมีพื้นที่บริการสำหรับให้ผู้บริโภคแสดงบทวิจารณ์เกี่ยวกับสถานที่พักที่ตนเคยไปใช้บริการมาก่อนหน้า เพื่อเป็นข้อมูลให้ผู้บริโภคอื่นใช้ในการตัดสินใจ โดย Dellarocas (2003) กล่าวว่า UGC เช่น e-WOM (Electronic Word-of-Mouth) นอกจากจะเป็นประโยชน์ต่อผู้บริโภคแล้ว ยังมีความสำคัญต่อเจ้าของธุรกิจออนไลน์อีกด้วย คือ สามารถช่วยสร้างชื่อเสียง สร้างความมั่นใจในคุณภาพของสินค้า จนกระทั่งสามารถนำบทวิจารณ์ไปพัฒนาสินค้าต่อได้ ซึ่ง Pasi (2011) ทำการศึกษาอิทธิพลของบทวิจารณ์ของผู้บริโภคในเว็บไซต์ Tripadvisor.com ที่ส่งผลต่อผลการดำเนินงานของโรงแรม พบว่า ผลการดำเนินงานมีความสัมพันธ์โดยตรงกับจำนวนบทวิจารณ์ซึ่งสอดคล้องกับผลการวิจัยของ Ghose &

Ipeirotis (2006) และ Ye et al. (2011) ที่พบว่า เมื่อจำนวนบทวิจารณ์เกี่ยวกับโรงแรมมากขึ้น จะทำให้ยอดจองห้องพักมากขึ้นด้วยเช่นกัน

เนื้อหาของบทวิจารณ์ที่โพสต์ตามสื่อออนไลน์หรือเว็บไซต์ต่าง ๆ อาจมีหลายรูปแบบขึ้นอยู่กับคุณลักษณะของสินค้าและบริการที่ถูกระบุ โดยทั่วไปสามารถแบ่งออกเป็น 2 แบบ คือ ลักษณะเด่นสำคัญและประสบการณ์ด้านการบริการ โดยมักประกอบด้วยบทวิจารณ์เชิงบวกและเชิงลบต่อสินค้าและบริการ (Ye et al., 2009) ซึ่งบทวิจารณ์เหล่านี้คือข้อมูลที่มาจกคนแปลกหน้าที่ไม่รู้จักและไม่เคยพบหน้ากันมาก่อน ผู้บริโภคจึงต้องพิจารณาความน่าเชื่อถือของทั้งผู้เขียนและเว็บไซต์ที่นำเสนอข้อมูลก่อนที่จะเชื่อถือข้อมูลเหล่านั้น โดยที่ผู้บริโภคจะพิจารณาความน่าเชื่อถือของแหล่งข้อมูลจากเนื้อหาของบทวิจารณ์นั้น ๆ (Schlosser, 2005) หากผู้บริโภคเชื่อถือในข้อมูลบทวิจารณ์ที่ได้รับ จะเกิดความคาดหวังว่าจะได้รับสินค้าและบริการแบบเดียวกันกับที่ผู้แสดงบทวิจารณ์ได้รับมา เมื่อผู้บริโภคมีความคาดหวังต่อสินค้าและบริการสูงขึ้น ความเป็นไปได้ในการซื้อก็จะสูงขึ้นด้วย จึงกล่าวได้ว่า ความคาดหวังของผู้บริโภคมีความสัมพันธ์เชิงบวกกับความตั้งใจซื้อ (Wu, 2013)

พีรพล กิรดิธนากาญจน์ (2553) กล่าวว่า ผู้บริโภคส่วนใหญ่มักให้ความเชื่อถือข้อมูล ที่ได้รับจากกลุ่มผู้ทรงอิทธิพลเป็นอย่างมากเพราะเชื่อมั่นว่าข้อมูลที่ได้รับเป็นข้อมูลจากผู้บริโภคด้วยกันเอง (Peer to Peer) หรือจากผู้เชี่ยวชาญ (Expert to Peer) ในเรื่องใดเรื่องหนึ่งมากกว่าตน และคิดว่าข้อมูลที่ได้รับไม่ใช่การโฆษณาโดยเจ้าของสินค้าและบริการจึงยินดีที่จะเปิดรับโดย ปราศจากอคติ ซึ่งรูปแบบการตัดสินใจซื้อสินค้าของผู้บริโภคในยุคออนไลน์โดยทั่วไป มักเริ่มเมื่อผู้บริโภคได้รับข้อมูลจากโฆษณาในสื่อต่าง ๆ จากนั้นผู้บริโภคจะค้นหาข้อมูลผ่านทางเว็บไซต์ชุมชน (Community Web) หรือเครือข่ายสังคมออนไลน์ (Social Networks) เพื่อสอบถาม แลกเปลี่ยนข้อมูล และพูดคุยเรื่องที่ตนเองสนใจพร้อมทั้งรับความคิดเห็นจากผู้ทรงอิทธิพล เพื่อประมวลผลและตีความก่อนที่จะตัดสินใจเลือกซื้อสินค้าและบริการตามคำบอกกล่าวของผู้ทรงอิทธิพล ในอีกกรณีหนึ่งผู้บริโภคอาจมีความเชื่อถือผู้ทรงอิทธิพลคนใดคนหนึ่งอยู่ก่อนแล้วและเมื่อได้รับรู้ว่าผู้ทรงอิทธิพลที่ตนชื่นชอบนั้นใช้สินค้าหรือบริการใดตนเองก็ต้องการที่จะใช้สินค้าหรือบริการนั้น พร้อมทั้งยังยินดีที่จะบอกต่อไปยังคนใกล้ชิดด้วยในมุมมองของนักการตลาดหรือผู้ประกอบการส่วนใหญ่จะใช้วิธีการสร้างการบอกต่อผ่านสื่อออนไลน์โดยให้ความสำคัญในการคัดเลือกผู้ทรงอิทธิพลจากวงการต่าง ๆ เช่น นักแสดง นักการเมือง หรือบล็อกเกอร์ที่ได้รับความนิยม โดยให้ทดลองใช้ผลิตภัณฑ์ของตราสินค้าจากนั้นให้ทำหน้าที่เป็นผู้วิจารณ์ (Reviewer) ผลิตภัณฑ์หรือแสดงภาพว่า ตนใช้ผลิตภัณฑ์นั้น ๆ ผ่านกลุ่มสังคมออนไลน์ เพื่อให้ผู้บริโภครู้สึกเชื่อถือและเกิดความต้องการ ผลิตภัณฑ์ซึ่งเมื่อผู้บริโภคตัดสินใจทดลองใช้แล้วเกิดความพึงพอใจในผลิตภัณฑ์ก็จะกระจายข่าวสารต่อไปยังเพื่อนหรือบุคคลใกล้ชิดต่อไป

ข้อมูลที่ได้รับคามนิยมจากผู้บริโภคเป็นอย่างมากนั้นพบว่าจะอยู่ในรูปแบบของ Consumer Review : CR คือข้อมูลจากผู้บริโภคที่เคยมีประสบการณ์ในการใช้สินค้าหรือบริการนั้นแล้ว โดยคน

ส่วนใหญ่มักมองว่ามีความน่าเชื่อถือมากกว่าข้อมูลจากแบรนด์ที่มีเพียงคำโฆษณา ซึ่งได้ทำการสำรวจพฤติกรรมของผู้บริโภคร้อยละ 86 ของผู้บริโภคไม่เชื่อในคำโฆษณาของผู้ประกอบการ และกว่าร้อยละ 70 ที่เชื่อในความคิดเห็นของผู้บริโภคผ่านช่องทางออนไลน์ (นิวัฒน์ ชาติะวิทวิทยกุล, 2555)

ณภัคอร ปุณยภานุสสร (2553) ได้กล่าวถึง ทฤษฎีเกี่ยวกับการตัดสินใจซื้อ (AIDA Model) ดังนี้ 1) ความตั้งใจ (Attention) เริ่มต้นที่ผู้บริโภคจะต้องรู้จักสินค้าและบริการนั้นก่อน ซึ่งก็เกิดขึ้นเมื่อผู้บริโภคมีความตั้งใจในการรับสาร 2) ความสนใจ (Interest) นอกจากการรู้จักสินค้าแล้วก็ยังไม่เพียงพอ ผู้บริโภคต้องถูกเร้าความสนใจจนสามารถแยกแยะสินค้านั้นออกจากสินค้าอื่น ๆ ที่มีอยู่ในตลาด 3) ความปรารถนา (Desire) ความสนใจที่ถูกกระตุ้นจะกลายเป็นความอยาก ความปรารถนาที่จะได้ครอบครองสินค้านั้น ตามปกติความปรารถนาจะเกิดขึ้นจากการเห็นประโยชน์ ที่สินค้านั้นเสนอให้หรือเกิดจากการเห็นหนทางแก้ปัญหาที่มีอยู่ที่สินค้านั้นหยิบยื่นให้ 4) การกระทำ (Action) เมื่อความปรารถนาเกิดขึ้นแล้ว กระบวนการซื้อจะสมบูรณ์ก็ต่อเมื่อผู้บริโภคก้าวเข้าสู่ขั้นตอนสุดท้ายคือการซื้อสินค้า จากข้อมูลสรุปได้ว่ากระบวนการตัดสินใจซื้อของผู้บริโภค จะเกิดจากการที่ผู้บริโภคเกิดความสนใจหรือความต้องการซื้อ และเกิดการแสวงหาข้อมูลและเลือกทางเลือกที่ดีที่สุด และทำการตัดสินใจซื้อสินค้าหรือบริการในที่สุด

1.3 การตลาดผ่านสังคมออนไลน์ (Social Media Marketing)

Social Media Marketing : SMM คือ กระบวนการผลิตและสร้างคอนเทนต์ให้ตอบโต้กับช่องทาง Social Media ต่าง ๆ เพื่อสื่อสารกับกลุ่มลูกค้าและโปรโมทชื่อเสียงของแบรนด์

Kaur (2018) ได้ให้ความหมายว่า การตลาดผ่านสังคมออนไลน์ คือรูปแบบของการสร้างเครื่องมือทางการตลาดบนโลกอินเทอร์เน็ต ซึ่งมีการใช้สื่อออนไลน์ในการกระจายและสร้างมูลค่าทางการตลาดโดยการสร้างการติดต่อสื่อสารและสร้างตราสินค้าให้บรรลุตามความต้องการโดยพื้นฐาน การสร้างการตลาดสังคมออนไลน์นั้นเป็นการสร้างกิจกรรมทางการตลาด ทั้งเนื้อหา รูปภาพ ภาพเคลื่อนไหว เสียง เพื่อวัตถุประสงค์หลักของการตลาดเพื่อกระจายเนื้อหาไปยังผู้บริโภคซึ่งการตลาดผ่านสังคมออนไลน์ถือได้ ว่าเป็นความนิยมแนวใหม่ในทางธุรกิจที่จะเข้าถึงกลุ่มเป้าหมายได้อย่างง่าย นอกจากนี้การตลาดผ่านสังคมออนไลน์สามารถพัฒนาช่องทางที่หลากหลายในการเข้าถึงกลุ่มผู้บริโภคที่ต้องการได้หลากหลายรูปแบบและกิจกรรม เช่น อีเมลล์หรือการโฆษณาออนไลน์บนเว็บไซต์ เป็นต้น

ศุภวิชัยกสิกร (2561) การทำธุรกิจกับการตลาดเป็นของคู่กัน กลยุทธ์การตลาดแบบดั้งเดิมที่คุ้นเคยกันดี คือ การใช้สื่อโฆษณาทั้งสิ่งพิมพ์ นิตยสาร สื่อเคลื่อนไหว โบรชัวร์ หรือ บิลบอร์ด ซึ่งเป็นช่องทางที่ผู้ประกอบการใช้ในการเข้าถึงผู้บริโภคมาอย่างยาวนานจนกระทั่งมีสื่อกลาง

อินเทอร์เน็ตมาถึง และสื่อสังคมออนไลน์กำเนิดขึ้นการตลาดผ่านสื่อสังคมออนไลน์ หรือ Social Media Marketing : SMM กลายเป็นช่องทางที่เข้ามาแทนที่และกำลังได้รับความนิยมสูงสุดในตอนนี้

ธนิดา อัครโยธิน (2561) ได้กล่าวว่า การตลาดออนไลน์ เป็นการนำเสนอผลิตภัณฑ์หรือบริการที่ใช้อินเทอร์เน็ตเป็นสื่อกลางที่ช่วยในการโฆษณา ประชาสัมพันธ์และใช้เป็นช่องทางในการติดต่อสื่อสารกันระหว่างเจ้าของธุรกิจกับลูกค้าเป็นการลงทุนทำการตลาดที่มีราคาไม่สูงสามารถแก้ไขหรืออัปเดตข้อมูลได้เสมอไม่ว่าจะเป็นการโพสต์หรือแชร์ข้อมูลข่าวสารที่ธุรกิจต้องการให้ลูกค้าทราบก็สามารถทำได้ง่าย สะดวก และรวดเร็ว เกิดจากการเข้ามาของการพาณิชย์อิเล็กทรอนิกส์ (E-Commerce) ซึ่งเป็นการดำเนินธุรกิจทางการค้าผ่านสื่ออิเล็กทรอนิกส์ การใช้เทคโนโลยีเป็นสื่อกลางระหว่างผู้ที่เกี่ยวข้อง ได้แก่ ตัวบุคคล องค์กร หรือตัวบุคคลกับองค์กรเพื่อช่วยสนับสนุนและอำนวยความสะดวกในการดำเนินกิจกรรมต่าง ๆ

1.4 เหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการสกัดข้อมูลในลักษณะข้อความที่มีขนาดใหญ่ โดยทำการสกัดคำ ค้นหารูปแบบและความสัมพันธ์ที่อยู่ภายในข้อมูล เพื่อให้เกิดความหมายและสามารถใช้ประโยชน์จากข้อมูลได้อย่างครอบคลุม ด้วยวิธีการทำเหมืองข้อมูลนี้เป็นการสกัดสารสนเทศที่มีคุณค่าและค้นหาความสัมพันธ์ของข้อความจำนวนมาก (เกียรติพงษ์ อุดมธนะธีระ, 2551) โดยการวิเคราะห์ข้อมูลจากชุดข้อมูลที่มีโครงสร้างและไม่มีโครงสร้างอย่างมีเป้าหมายจึงทำให้สามารถนำมาใช้ในการทำนายและจัดกลุ่มข้อมูล การทำเหมืองข้อมูลจึงเหมาะกับหลากหลายสาขาวิชาที่สามารถนำไปประยุกต์ใช้ให้เกิดประโยชน์ การเริ่มต้นวิเคราะห์ข้อมูลจะได้ข้อมูลที่มีลักษณะไม่มีโครงสร้างจึงต้องปรับให้อยู่ในข้อมูลที่มีโครงสร้าง เพื่อสะดวกต่อการวิเคราะห์โดยผ่านวิธีการวิเคราะห์ การตีความ การให้ความหมาย โดยการทำเหมืองข้อมูลมักจะเป็นหน้าที่ของทีมนักวิทยาศาสตร์ข้อมูล (Data Science) ที่ทำหน้าที่แสวงหาสารสนเทศที่น่าสนใจจากข้อมูลขนาดใหญ่ที่ยุงยากและซับซ้อน นำไปสู่การสร้างแบบจำลองการทำนายการเรียนรู้ของเครื่อง (Machine Learning) เพื่อให้ได้คำแนะนำที่ลุ่มลึกและเป็นประโยชน์แก่องค์กรและหน่วยงานต่าง ๆ (Kwartler, 2017) และปัจจุบันมีการคิดค้นซอฟต์แวร์ในการวิเคราะห์อย่างหลากหลาย เช่น RapidMiner, Weka, KEEL, KNIME, Orange and SPSS ซึ่งติดลิขสิทธิ์ในการวิเคราะห์

แต่ถ้าหากสามารถเขียนโปรแกรมได้ การใช้ R หรือ Python จะทำให้สามารถวิเคราะห์ข้อมูลได้หลายรูปแบบตามการเขียนโปรแกรมและสามารถนำเสนอได้โดยไม่ต้องติดลิขสิทธิ์

การประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) หรือ Natural Language (Processing Linguistics+Programming+Computer Science) เป็นศาสตร์หนึ่งในสาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ที่เกี่ยวข้องกับปัญญาประดิษฐ์ (Artificial Intelligence) และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) เป็นศาสตร์ที่จะทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์และสามารถโต้ตอบด้วยภาษามนุษย์ โดยข้อความหรือเสียง หรือการเก็บข้อมูล โดยมีระบบฐานข้อมูลในการเก็บบันทึกการโต้ตอบ นำข้อมูลเหล่านั้นมาตรวจจับคำสำคัญจากข้อมูล เพื่อประมวลหาคำตอบ และสามารถเข้าถึงลูกค้าได้มากขึ้น เป็นเทคนิควิธีสำหรับการใช้เครื่องคอมพิวเตอร์ประมวลวิเคราะห์ข้อมูลภาษาจำนวนมาก หรือเทคนิควิธีสำหรับการ automate หน้าที่ทางภาษาต่าง ๆ

1.5 แบบจำลองหัวข้อ (Topic Modeling)

แบบจำลองหัวข้อ (Topic Modeling) เป็นเทคนิคการเรียนรู้ Machine Learning เป็นการเรียนรู้ของเครื่องแบบ Unsupervised Learning ใช้ในการวิเคราะห์ข้อมูลตัวอักษร เพื่อกำหนดกลุ่มคำสำหรับชุดเอกสาร (Document) เป็นการจำลองหัวข้อแบบการกระจายตัวของข้อมูล เพื่อนำไปใช้ในการจัดกลุ่มข้อมูล โดยเอกสาร แต่ละเอกสารจะประกอบด้วยการรวมกันหลาย ๆ หัวข้อ แต่ละหัวข้อมีการแจกแจงความน่าจะเป็นของคำนั้น ๆ ที่เกิดขึ้นในแต่ละหัวข้อ

การจัดสรรDirichletแฝง (Latent Dirichlet Allocation : LDA) เป็นแบบจำลองความน่าจะเป็นสำหรับการทำแบบจำลองหัวข้อ (Topic Modeling) เป็นการหาหัวข้อ (Topic) ที่ซ่อนอยู่ในเอกสาร โดยมีหัวข้อที่แฝงอยู่ในเอกสาร ถูกสร้างจากแนวคิดที่ว่า ในเอกสารนั้นจะประกอบไปด้วยหัวข้อที่ถูกรวมกันแบบไม่มีรูปแบบและกระจัดกระจายในกลุ่มของคำศัพท์ของเอกสารนั้น โดยการทำให้ LDA จะทำการหาค่าความน่าจะเป็นของความสัมพันธ์ของหัวข้อของแต่ละประโยค และค่าความน่าจะเป็นของคำในแต่ละหัวข้อ ลักษณะของเทคนิคนี้เป็นแบบการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และใช้หลักความน่าจะเป็นในการหาพารามิเตอร์ (Probabilistic Model) หลักการทำงานของ LDA ได้แก่ 1) เลือกพารามิเตอร์ โดยกำหนดจำนวนหัวข้อ ตัวอย่างเช่น Clustering algorithms เช่น K-mean ที่เลือกจำนวน Cluster หรือค่า k 2) LDA จะทำการอ่านทุกคำ ของแต่ละเอกสารและกำหนดหัวข้อ โดยผลลัพธ์ที่ได้คือ กลุ่มคำของหัวข้อ (Topic Representation) ที่อยู่ร่วมกันในถ้อยคำ (Bag of Words) และหัวข้อของแต่ละเอกสาร จำนวนหัวข้อที่ดีที่สุดมี 2 วิธี คือ การหา Perplexity และ Log-Likelihood function

วิธีแรก Perplexity หาจำนวน k ที่เหมาะสม และประเมินประสิทธิภาพของโมเดลในการอธิบายข้อมูลค่า Perplexity ต่ำมาก ก็หมายความว่าโมเดลนั้นมีประสิทธิภาพดีมาก วิธีการหาตอนนี้เมื่อมีการกำหนดค่า k จากการ LDA ได้กลุ่มของคำ นำค่าเหล่านั้นที่ได้กำหนดความเชื่อมโยงกับ

แต่ละหัวข้อเปรียบเทียบกับกระจายของคำ ในเอกสาร เพื่อหา cluster ของหัวข้อของเอกสารนั้น ๆ สามารถอธิบายเป็นสมการได้ดังนี้

$$\text{perplexity}(D_{\text{test}}) = \exp \left(-\frac{\sum_{d=1}^m \mathbf{1} \log[p(\mathbf{w}_d)]}{2a \sum_{d=1}^m \mathbf{1} N_d} \right)$$

วิธีที่สอง Log-Likelihood function การหาค่าความใกล้เคียง โดยผลลัพธ์ที่ได้หากมีค่าสูง แสดงว่ามีความใกล้เคียงกันมาก

$$L(\theta, m/D) = p_{\theta, m}(d) = P_{\theta, m}(D = d)$$

$$L(\theta, m/D) = p(D/\theta, m)$$

ขั้นตอนการเตรียมข้อมูลภาษาไทย (Thai text preprocessing)

ทำความสะอาดข้อความ (Clean text) นำข้อความมาทำการตัดเครื่องหมายพิเศษต่าง ๆ ที่ไม่จำเป็น

ตัดคำ (Tokenization) คือ กระบวนการตัดคำออกจากข้อความยาว ๆ ซึ่งแต่ละคำที่ถูกตัดออกมาจะถูกเรียกว่า Token โดยกระบวนการนี้มักจะถูกทำเป็นกระบวนการแรก ๆ

สำหรับภาษาไทยมี library ในการตัดคำภาษาไทย เช่น PyThaiNLP มีการรองรับหลายอัลกอริทึม ได้แก่ newmm, longest, multi_cut, pyicu, deepcut, attacut, Thai Character Clusters (TCCs) และ Enhanced Thai Character Cluster (ETCC)

การตัด Stop words หรือคำ ที่ไม่ได้สื่อความหมายและพบบ่อย เช่น การ, ความ, ที่, ซึ่ง และการตัดคำจะช่วยให้คำศัพท์ (Bag of Words) เก็บคำศัพท์ไม่เยอะเกินความจำเป็น โดยเฉพาะเมื่อเรามีชุดข้อมูลที่มีขนาดใหญ่ แต่การตัดคำบางส่วนก็อาจทำให้ความแม่นยำของแบบจำลองลดลง และการแปลงคำศัพท์ให้อยู่ในรากศัพท์ เพื่อลดจำนวนคำศัพท์ลง

การสร้างคลังคำศัพท์จากข้อความทั้งหมด (Bag of Words) เป็นการแปลงคำที่มีลักษณะเฉพาะเป็นรหัส มีการสกัดคุณลักษณะ 2 แบบ คือ การสกัดคุณลักษณะ (Feature) จาก การนับคำ โดยนับจำนวนครั้งที่แต่ละคำ จากคำทั้งหมดปรากฏในแต่ละข้อความ และการสกัดคุณลักษณะด้วย TF-IDF คือ การคำนวณค่า TF-IDF ของแต่ละคำ TF-IDF (term frequency-inverse document frequency) นั่นคือ การนับจำนวนครั้งที่แต่ละคำปรากฏในแต่ละข้อความ (Term frequency) หารด้วยคำทั้งหมดในข้อความนั้น และนำมาคูณกับจำนวนของข้อความทั้งหมด

หารด้วยจำนวนข้อความที่แต่ละคำปรากฏและเทค log โดยคำที่ปรากฏในข้อความมาก ๆ ค่า IDF จะมีค่ามากตาม แต่ถ้าคำนั้นปรากฏอยู่ในทุกข้อความก็จะให้ค่า IDF ที่ต่ำ

การวิเคราะห์การจัดกลุ่มข้อมูล (Clustering Analysis)

การจัดกลุ่มข้อมูล หรือ Clustering Analysis เป็น Machine Learning ประเภทการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) โดยเป็นวิธีที่เป็นการจำแนกข้อมูลที่มีความคล้ายกัน ให้อยู่ในกลุ่มเดียวกัน เทคนิคนี้ไม่ได้ใช้เพื่อหาความแม่นยำ แต่เป็นการหาความสัมพันธ์ของข้อมูลอีกรูปแบบหนึ่ง อัลกอริทึมที่ใช้ในการจัดกลุ่ม แบ่งเป็น 2 ประเภท ได้แก่

1) ประเภทการแบ่งกลุ่มชัดเจน (Hard Clustering) เทคนิคนี้มีการแบ่งกลุ่มข้อมูล แต่ละข้อมูลไปอยู่ในกลุ่มใดกลุ่มหนึ่งเท่านั้น เทคนิคประเภทนี้ ได้แก่ เทคนิคการจัดกลุ่ม ข้อมูลแบบเคมีน (K-means clustering algorithm)

2) ประเภทการแบ่งกลุ่มแบบไม่ชัดเจน (Soft Clustering) เทคนิคนี้มีการแบ่งข้อมูลแต่ละข้อมูลที่สามารถอยู่ได้ในหลาย ๆ กลุ่ม เทคนิคประเภทนี้ ได้แก่ เทคนิคการจัดกลุ่ม แบบข้อมูลแบบลำดับขั้น (Hierarchical Clustering methods)

โดยการจัดกลุ่มข้อมูลแบบเคมีน (K-means clustering Algorithm) เป็นวิธีการที่ถูกนำมาใช้บ่อยมากที่สุด เนื่องจากมีขั้นตอนในการทำงานไม่ซับซ้อน และเข้าใจง่าย โดยมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 กำหนดค่า K หรือกำหนดจำนวนกลุ่มข้อมูลที่ต้องการ

ขั้นตอนที่ 2 สุ่มวางตำแหน่งจุดศูนย์กลาง (Centroid) ของข้อมูลแต่ละกลุ่ม

ขั้นตอนที่ 3 มีการคำนวณระยะทางจากจุดศูนย์กลางกับข้อมูลในชุดข้อมูล จากนั้นแต่ละข้อมูลจะถูกจัดอยู่ในกลุ่มของศูนย์กลางที่มีระยะทางใกล้ที่สุดเท่านั้น

ขั้นตอนที่ 4 หลังจากทำการจัดกลุ่มข้อมูลใหม่ และทำการคำนวณค่าเฉลี่ยของข้อมูลในกลุ่ม ทำการกำหนดศูนย์กลางของข้อมูลใหม่

ขั้นตอนที่ 5 ทำการทำซ้ำในขั้นขั้นตอน 3 และ 4 จนกระทั่งค่าจุดศูนย์กลางของกลุ่มข้อมูลใหม่ไม่เปลี่ยนแปลงหรือเปลี่ยนแปลงเพียงเล็กน้อย

1.6 การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นการนำเอาเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) ที่สามารถใช้วิเคราะห์ความรู้สึกทั้งในระดับเอกสาร ระดับประโยค และในระดับคุณลักษณะของวัตถุ รวมถึงการใช้วิเคราะห์ความคิดเห็นของผู้บริโภคที่มีความรู้สึกต่อสินค้าและบริการต่าง ๆ รวมถึงการทำเหมืองความคิดเห็น (Opinion Mining) โดยมีวัตถุประสงค์ในการสกัดข้อความที่เป็นความคิดเห็นที่มีต่อสิ่งใดสิ่งหนึ่งออกมาจาก

ข้อเท็จจริงเพื่อนำไปสกัดคุณลักษณะและส่วนประกอบของวัตถุที่แสดงความคิดเห็นออกมาเพื่อพิจารณาว่าความคิดเห็นนั้นแสดงความรู้สึกไปในเชิงบวก เชิงลบ หรือเป็นกลาง และเทคนิคการทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้เพื่อมุ่งเน้นการวิเคราะห์และตรวจสอบความรู้สึก (Opinion) ได้อย่างอัตโนมัติ รวมถึงการแทนคำในความคิดเห็นให้อยู่ในรูปแบบโครงสร้าง เช่น การแทนด้วยถุงคำ (Bag of Words) และค่าความถี่ของคำที่เกิดขึ้น (TF-IDF) และเทคนิคการสกัดคำ (Word Extraction) ซึ่งเป็นกระบวนการของการทำเหมืองข้อมูล เพื่อใช้การวิเคราะห์คำออกจากเอกสาร ข่าวสาร ข้อความ และสารสนเทศต่าง ๆ ที่เป็นตัวอักษร โดยสามารถนำไปทำการแบ่งกลุ่ม (Clustering) จำแนกกลุ่ม และหาความสัมพันธ์ขั้นตอนการแบ่งกลุ่มโดยใช้เทคนิคต่าง ๆ เช่น

1) การตัดคำ (Word Segmentation) เป็นการแยกแต่ละคำจากเอกสารออกจากกัน โดยยังคงมีความหมายที่ถูกต้องสมบูรณ์อยู่ โดยการตัดคำนั้นใช้ฐานข้อมูลพจนานุกรมคำศัพท์ในการแบ่งคำออกมา ในการตัดคำภาษาไทยจะนิยมใช้อัลกอริทึมจาก PyThaiNLP ในการตัดคำภาษาไทย โดยเฉพาะ เช่น อัลกอริทึม พัฒนาจากเทคนิคการแบ่งส่วนคำศัพท์ภาษาไทยด้วยพจนานุกรม (Dictionary-based) โดยใช้อัลกอริทึมการจับคู่สูงสุดและการจัดกลุ่มอักขระภาษาไทย อัลกอริทึม Longest Engine ซึ่งเป็นอัลกอริทึมพัฒนาจากเทคนิคการแยกคำภาษาไทยที่ยาวที่สุดที่จับคู่กับพจนานุกรม (Dictionary-based) และอัลกอริทึม AttaCut Engine ซึ่งเป็นอัลกอริทึมที่สามารถตัดคำภาษาไทยได้อย่างรวดเร็วและแม่นยำ โดยใช้อัลกอริทึมการจัดกลุ่ม (Clustering) อักขระภาษาไทย (Rules purposed)

2) การกำจัดคำหยุด (Stop word) เป็นการตัดคำที่ไม่มีความหมายออกจากเอกสารโดยกำจัดคำหยุดนั้นใช้ฐานข้อมูลคำศัพท์ที่เป็นคำที่ไม่มีความหมายในการนำมาประมวลผล เพื่อช่วยให้การจำแนกประเภท (Classification) มีความถูกต้องมากยิ่งขึ้น ในส่วนของการวิเคราะห์ความรู้สึกและการประมวลผลในภาษาไทยยังมีความยุ่งยากมากกว่าภาษาอังกฤษ เพราะโครงสร้างประโยคและไวยากรณ์มีความซับซ้อน ไม่มีการแบ่งส่วน แบ่งคำประโยคติดกัน อักขระที่ไม่มีขอบเขตคำ และไม่มีโครงสร้างที่ชัดเจน ทำให้เกิดความกำกวม นอกจากนี้ยังมีการพัฒนาเฟรมเวิร์ค (Framework) ที่เรียกว่า S-Sense (Social Media Sensing) สำหรับวิเคราะห์ความรู้สึกบนโซเชียลมีเดียสำหรับภาษาไทย ตัวแบบจะเปรียบเทียบประสิทธิภาพระหว่างชุด คำศัพท์ที่แตกต่างกัน 2 ชุด คือ คำศัพท์ทั่วไปและคำใบ้ ซึ่งการรวมคำใบ้เข้ากับเวกเตอร์คุณลักษณะสำหรับการสร้างแบบจำลองการจำแนกประเภททำให้มีการปรับปรุงอย่างมีนัยสำคัญในแง่ของความถูกต้อง และในปัจจุบันการใช้เทคนิคการเรียนรู้ด้วยเครื่อง (Machine Learning) จะเป็นที่นิยมมากกว่า โดยจะอาศัยการสร้างชุดข้อมูลฝึกฝน (Training Set) และนำชุดข้อมูลทดสอบ (Test Set) เข้ากระบวนการจัดกลุ่มหรือจำแนกประเภท (Classification) ในส่วนของการเลือกใช้ตัวจำแนกประเภทที่มีความนิยมในการพัฒนา Sentiment Analysis ภาษาไทย ได้แก่ Logistic Regression, Random Forest, Support Vector

Machine, Naive Bayes, K-Nearest Neighbors และ Decision Tree เป็นต้น และในการประเมินประสิทธิภาพของตัวแบบที่ได้โดยสรุปจาก Accuracy, Precision, Recall และ F-Measure (F1) นอกจากนี้ การวิเคราะห์ข้อความรู้สีกส่วนใหญ่จะเป็นการตัดสินใจเพียง 2 ระดับ คือ เชิงบวกและเชิงลบ อย่างไรก็ตาม จะมีความคิดเห็นที่เป็นกลาง ซึ่งถ้าหากถูกตัดสินเพียง 2 ระดับ อาจจะทำให้ข้อคิดเห็นโดยรวมเกิดอคติได้ ดังนั้นถ้าหากมีการวิเคราะห์เป็น 3 ระดับ คือ มีระดับเป็นกลางเพิ่มเข้ามาจะทำให้น้ำหนักของการตัดสินข้อความรู้สีกมีความถูกต้องมากขึ้น อีกทั้งการตัดคำ (Word Segmentation) จะต้องเลือกโมดูล (Module) ที่เหมาะสม ถูกพัฒนาขึ้นเพื่อใช้งานกับภาษาไทยโดยเฉพาะ ทั้งนี้เพื่อให้การตัดคำถูกต้องและพร้อมใช้งานในการฝึกฝนข้อมูล

2. งานวิจัยที่เกี่ยวข้อง

Sheikha (2006) ได้เสนอการทำเหมืองข้อมูลเกี่ยวกับ COVID-19 โดยมีการใช้อัลกอริทึม Latent Semantic Analysis : LSA และ Latent Dirichlet Allocation : LDA ในการลดมิติของข้อมูล ซึ่งจะถูกจัดกลุ่มโดยใช้อัลกอริทึม HDBSCAN และ K-Means โดยผลสรุปพบว่า จากรายงานไม่มีความสัมพันธ์ของข้อมูลโซเชียลมีเดียกับข้อมูล WHO โดยตรง แต่มีความสัมพันธ์กับการพาดหัวข่าว โดยการเปรียบเทียบคำที่ใช้บ่อยของกลุ่มคำในวันเดียวกัน โดย HDBSCAN และการจัดสรรตรีเคลแฝง (Latent Dirichlet Allocation : LDA) ให้ผลลัพธ์ที่ดีที่สุด

Kim & Gil (2008) ได้เสนอระบบการจำแนกเอกสารการวิจัยที่มีความคล้ายคลึงกัน โดยใช้เทคนิคการจัดสรรตรีเคลแฝง (Latent Dirichlet Allocation : LDA) ในการดึงคำที่มีนัยสำคัญจากบทคัดย่อ (Abstract) ของแต่ละบทความและตามหัวข้อ จากนั้นใช้อัลกอริทึมการจัดกลุ่ม โดยใช้ K-Means เพื่อใช้จำแนกบทความทั้งหมดที่เป็นรายงานการวิจัยที่มีหัวข้อคล้ายคลึงกัน โดยยึดตามค่า Term frequency-inverse document frequency (TF-IDF) ของในแต่ละบทความ โดยการประเมินผลจากค่า F-Score สูงสุดคือการทำ TF-IDF-LDA 30 โดยมี 15 คำสำคัญ และ 15 หัวข้อ

Lim (2009) ได้ทำการพัฒนาระบบหุ่นยนต์โต้ตอบการสนทนา (Chatbot) สุขภาพจิตชื่อว่า Woebot สำหรับคาดเดาผู้ใช้งานในเรื่องความรุนแรงของภาวะซึมเศร้า โดยป้อนข้อมูลไปยังตัวแยกประเภท (Classifier) 3 ประเภท ได้แก่ ข้อมูลการสำรวจพื้นฐานหรือเวกเตอร์ TF-IDF จากข้อความทั้งหมด หรือเวกเตอร์การหาความน่าจะเป็นจากแบบจำลอง Latent Dirichlet Allocation : LDA โดยงานวิจัยนี้พบว่าหัวข้อที่สร้างโดย LDA มีประสิทธิภาพที่ดีที่สุดสำหรับแสดงลักษณะอารมณ์จากข้อความ ซึ่งสามารถใช้เพื่อทำนายผลภาวะซึมเศร้า

Agarwal et al. (2011) การวิเคราะห์ความรู้สึกของข้อมูลใน Twitter ผลงานของบทความนี้คือ : 1) เราแนะนำคุณสมบัติชั่วคราวหน้าเฉพาะ POS 2) เราสำรวจการใช้ tree kernel เพื่อจัด

ความต้องการวิศวกรรมคุณลักษณะที่น่าเบื่อ พีเจอร์ใหม่ (ร่วมกับพีเจอร์ที่เสนอก่อนหน้านี้) และ ตรีเคอร์เนลทำงานโดยประมาณที่ระดับเดียวกัน ทั้งคู่มีประสิทธิภาพดีกว่าพื้นฐานที่ล้ำสมัย

Tian (2012) แหล่งชุมชนในการถามตอบคำถาม (Community Question Answering : CQA) กำลังกลายเป็นแหล่งข้อมูลที่สำคัญมาก ซึ่งผู้ใช้สามารถแบ่งปันความรู้ใน หัวข้อต่าง ๆ ระหว่างกันได้ แม้ว่าแพลตฟอร์มสามารถช่วยเหลือผู้ใช้งาน แต่ก็ยังก่อให้เกิดความท้าทายมากมายกับขนาดของชุมชนที่เติบโตขึ้นเรื่อย ๆ จำนวนคำถามที่โพสต์ทุกวันเป็นสาเหตุให้เกิดปัญหาในการค้นหาคำถามแก่ผู้ใช้ที่เหมาะสม ในบทความนี้ได้เสนอแนวทางในการทำนายคำตอบที่ดีที่สุดสำหรับคำถามใหม่บน CQA แนวทางของงานวิจัยนี้ได้พิจารณาทั้งความสนใจของผู้ใช้และความเชี่ยวชาญของผู้ใช้ที่เกี่ยวข้องกับหัวข้อของคำถามที่กำหนด ความสนใจของผู้ใช้ในหัวข้อต่าง ๆ สามารถเรียนรู้ได้จากการใช้การสร้างแบบจำลองหัวข้อ (Topic Modeling) กับคำถามก่อนหน้านี้ที่ตอบโดยผู้ใช้ ในขณะที่ความเชี่ยวชาญของผู้ใช้จะได้รับการเรียนรู้โดยใช้ประโยชน์จากกลไกการลงคะแนนร่วมกันใน CQA เราได้ใช้โมเดลของเรากับชุดข้อมูลที่ดึงมาจาก Stack Overflow ซึ่งเป็นหนึ่งใน CQA ที่ใหญ่ที่สุด ผลการศึกษาแสดงให้เห็นว่าแนวทางของเรา มีประสิทธิภาพดีกว่าแนวทางที่ใช้ TF-IDF โดยงานวิจัยนี้ประกอบด้วย 4 ส่วน ได้แก่ การจัดเตรียมข้อมูล (Data preprocessing) การสร้างโปรไฟล์ผู้ใช้งาน (User profile building) ความสนใจเฉพาะของผู้ใช้และการเรียนรู้แบบเชี่ยวชาญ (User topical interest and Expertise learning) และการสร้างแบบจำลองการจัดอันดับ (Ranking model building) โดยเริ่มต้นจากการสร้างโปรไฟล์ของผู้ใช้งานจากคำถามที่เคยถามก่อนหน้านี้ ประกอบด้วยข้อความและข้อมูลการลงคะแนน (Voting information) ซึ่งจะใช้ในการเรียนรู้หัวข้อตลอดจนความสนใจของผู้ใช้และความเชี่ยวชาญของผู้ใช้ในแต่ละหัวข้อที่เรียนรู้ งานวิจัยนี้ได้เสนอแบบจำลองการจัดอันดับเพื่อที่จะคำนวณหาความน่าจะเป็นที่ผู้ใช้จะเป็นผู้ที่ตอบคำถามได้ดีที่สุด ผู้ใช้มีความสนใจและเชี่ยวชาญในหัวข้อต่าง ๆ จะถูกบันทึกลงในแบบจำลองการจัดอันดับ โดยหลักของงานวิจัยนี้คือการจัดอันดับผู้ใช้ โดยใช้ค่าความน่าจะเป็นและค้นหาผู้ใช้ TOP-K ที่มีโอกาสสูงในการให้คำตอบที่ดีที่สุด

Touimi et al. (2013) กล่าวถึงระบบหุ่นยนต์โต้ตอบการสนทนา (Chatbot) ในเชิงความหมาย สำหรับการวิเคราะห์ แลกเปลี่ยน แบ่งปันความรู้และประสบการณ์ของผู้เรียน โดยงานวิจัยนี้ได้รับข้อความจากผู้ใช้งานด้วยระบบ หุ่นยนต์โต้ตอบการสนทนา (Chatbot) จำแนกคำที่มีนัยสำคัญด้วยแบบจำลอง Latent Dirichlet Allocation : LDA ทำการ mapping ไปยังโดเมน ontology ของ MOOC การประยุกต์ใช้แบบจำลองความน่าจะเป็นของ LDA ทำให้สามารถดึงความรู้ที่เกี่ยวข้องตามที่ผู้เรียนร้องขอได้

Twinandilla (2014) ได้เสนอการสรุปเอกสารหลายฉบับเพื่อให้ผู้อ่านสามารถเข้าใจเจตนาของเอกสารข่าวออนไลน์ได้อย่างง่ายดาย โดยการใช้เทคนิคการสร้างแบบจำลองหัวข้อ การจัดสรร

ดีรีเคลแฝง (Latent Dirichlet Allocation : LDA) ในการสรุปประโยคที่สำคัญในเอกสารหลาย ๆ ฉบับในภาพรวม โดยไม่ต้องจัดกลุ่มตามหัวข้อ และได้เสนอวิธีการสรุปแบบใหม่คือการรวม K-Means Clustering และ Latent Dirichlet Allocation : LDA - Significance Sentences เข้าด้วยกัน

Wexler et al. (2016) ได้เสนอการวิเคราะห์ฟอรัมการตั้งกระทู้ออนไลน์ เพื่อให้เข้าใจวิธีการหาข้อมูลของหญิงตั้งกระทู้ในชุมชนออนไลน์ โดยใช้การสร้างแบบจำลองหัวข้อ การจัดสรรดีรีเคลแฝง (Latent Dirichlet Allocation : LDA) โดยอัลกอริทึมนี้สามารถค้นพบถ้อยคำ (Bag of Words) ที่มีความน่าจะเป็นสูงที่จะปรากฏร่วมกัน โดยผลลัพธ์ในการจัดหมวดหมู่ได้หัวข้อที่ใหญ่ที่สุด ได้แก่ สุขภาพมารดา (45%) หัวข้อเกี่ยวกับทารก (29%) และ คน/ความสัมพันธ์ (10%) และในการจัดหมวดหมู่ตามไตรมาสการตั้งกระทู้ ผลลัพธ์ในไตรมาสแรกหญิงตั้งกระทู้มีความกังวลในเรื่องการแท้งบุตร ไตรมาสที่สองมีความกังวลในการคลอดบุตร และกิจกรรมนอนของทารกในช่วงหลังคลอด

Zhang (2017) บทความนี้ได้เสนอวิธี Supervised question-answer topic modeling approach สำหรับการดึงข้อมูลใน community question answering (CQA) โดยมีการตั้งสมมติฐานพื้นฐานคือ แม้ว่าคำถามและคำตอบจะมีความแตกต่างกันในหลายแง่มุม แต่พวกเขา มีการแบ่งปันปัจจัยแฝงบางอย่างที่แสดงให้เห็นถึงหัวข้อที่ผู้ถามและผู้ตอบต้องการ ยิ่งไปกว่านั้นยังมีตัวบ่งชี้คุณภาพของคำตอบจำลองวิธีที่ผู้ใช้งานคนอื่นประเมินคำตอบที่เกี่ยวข้องกับคำถามในคู่ของคำถาม-คำตอบ กำหนดให้ผู้ถาม และผู้ตอบตามหัวข้อทั่วไปในแบบจำลองได้ดีเพียงใด ด้วยคำแนะนำตัวอย่างของคุณภาพ คือ คำถามและคำตอบจะถูกจับคู่กับ latent space หรือ topic space และความคล้ายคลึงของคำถามสามารถวัดได้ด้วยความช่วยเหลือของข้อมูลในคำตอบที่ถูกย่อใน space ด้วย topic space และได้เพิ่มเติมการจับคู่คำถามใน Latent space ให้เป็น LMIR แบบดั้งเดิมและเสนอรูปแบบ Topic-based language model และยังสามารถใช้ข้อมูลจาก Metadata มากมายที่เกี่ยวข้องกับคำถาม และเรียนรู้คุณภาพของคำตอบโดยอัตโนมัติ โดยการใช้คำตอบที่ดีที่สุด เป็นมาตรฐานและดึงคุณลักษณะต่าง ๆ จากทั้งข้อความคำตอบและข้อมูล Metadata ที่เกี่ยวข้อง ด้วยข้อมูลการฝึกฝน และคุณลักษณะเรียนรู้ฟังก์ชันคะแนนเพื่อสร้างคำตอบที่มีคุณภาพ และปล่อยให้ควบคุมดูแลการเรียนรู้ของหัวข้อแบบจำลองนี้ นอกจากนี้จะนำข้อมูลขนาดใหญ่มาใช้ได้เป็นประโยชน์ได้เท่านั้น ยังสามารถช่วยประหยัดแรงงานของผู้สังเกตอีกด้วย บทความนี้ได้ทำการทดลองกับข้อมูลขนาดใหญ่จาก Yahoo! Answers และ Baidu Knows ผลการทดลองแสดงให้เห็นว่าแบบจำลองที่นำเสนอสามารถมีประสิทธิภาพเหนือกว่าโมเดลการดึงข้อมูลที่ทันสมัยใน CQA อย่างมีนัยสำคัญ

กานดา แผ้ววัฒนากุล และปราโมทย์ ลือนาม (2556) ความคิดเห็นและข้อเสนอแนะที่แพร่หลายจำนวนมากในเครือข่ายสังคมออนไลน์เป็นประโยชน์ต่อธุรกิจ ทำให้ทราบความรู้สึกของผู้บริโภคทั้งในแง่บวกและลบ ต่อคุณลักษณะของผลิตภัณฑ์ในแต่ละด้าน อันจะนำไปสู่การปรับปรุงคุณภาพของสินค้าและบริการ การวิเคราะห์ความคิดเห็นจำนวนมากสามารถทำได้

โดยอาศัยระบบคอมพิวเตอร์ที่ประมวลผลชุดขั้นตอนวิธีการทำเหมืองความคิดเห็น (Opinion mining) แบ่งเป็น 2 ขั้นตอนหลัก ได้แก่ 1) ขั้นตอนการแทนข้อความให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปประมวลผลได้ และ 2) ขั้นตอนการจำแนกประเภทความคิดเห็นออกเป็นข้อความความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง ทำให้ได้ข้อมูลที่จะช่วยเพิ่มความสามารถในการแข่งขัน ลดระยะเวลาและค่าใช้จ่ายในการสำรวจตลาด รวมถึงปรับกลยุทธ์ตอบสนองต่อความต้องการของผู้บริโภคที่เปลี่ยนแปลงอย่างรวดเร็ว บทความนี้กล่าวถึงแนวคิดพื้นฐาน และการประยุกต์ใช้เหมืองความคิดเห็น รวมทั้งทบทวนวรรณกรรมที่เกี่ยวข้อง พร้อมคำวิจารณ์ และข้อเสนอแนะในการศึกษาวิจัย

รวีสฤตา เทศเมือง และนิเวศ จิระวิชิตชัย (2560) ได้ทำการศึกษาเรื่องการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทซึน นำเสนอวิธีการจำแนกความคิดเห็นโดยการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการการรีวิวสินค้าออนไลน์ ด้านการบริการห้องพักโรงแรม รีสอร์ท จาก Agoda Thailand และ Twitter Thailand จำนวน 2,890 ข้อมูล จากการทดลองพบว่าคุณลักษณะที่ดีที่สุด คือ ซัพพอร์ตเวกเตอร์แมทซึน รองลงมาเป็น Naive Bayes, Decision Tree และ K nearest neighbor ตามลำดับ

อิสรภาพ ล้อรัตนไชยรงค์ (2560) งานวิจัยนี้มีจุดมุ่งหมายเพื่อสร้างคลังศัพท์บอกความรู้สึกจากบทวิจารณ์สินค้าและบริการออนไลน์ในภาษาไทยโดยใช้วิธีการประมวลภาษาธรรมชาติตามแนวทางการวิจัยด้านภาษาศาสตร์คอมพิวเตอร์ บทวิจารณ์ที่เลือกใช้มาจาก 3 แหล่งข้อมูล ได้แก่ บทวิจารณ์โรงแรมของ Agoda บทวิจารณ์ภาพยนตร์ของ Major Cineplex และบทวิจารณ์แอปพลิเคชันโทรศัพท์มือถือของ Microsoft ซึ่งมีการให้คะแนนร่วมกับการเขียนเนื้อหาบทวิจารณ์ การวิเคราะห์หาคำบอกความรู้สึกจากเนื้อหาบทวิจารณ์อาศัยสมมติฐานว่าคำบอกความรู้สึกจะเกิดร่วมกับคำบอกลักษณะสินค้าซึ่งเป็นคำในกลุ่มคำนามที่มีความถี่การปรากฏสูง การระบุข้อความรู้สึกรวมของคำบอกความรู้สึกจะดูจากค่า TF-IDF เชิงบวกและเชิงลบซึ่งคำนวณจากความถี่การปรากฏในกลุ่มข้อมูลบทวิจารณ์ที่มีการให้คะแนนเชิงบวกและเชิงลบตามลำดับ กระบวนการรวบรวมคำบอกความรู้สึกในงานวิจัยนี้ทดลองใช้วิธีการต่าง ๆ ในสามขั้นตอน คือ การกำหนดชนิดคำบอกความรู้สึก การกำหนดค่าขั้นต่ำของลำดับความถี่ของคำบอกลักษณะสินค้า และการกำหนดค่า TF-IDF ขั้นต่ำในการคัดเลือกคำบอกความรู้สึกเชิงบวกและเชิงลบ ผลที่ได้คือชุดคำบอกความรู้สึกที่แตกต่างกัน 112 ชุดจากแต่ละโดเมน จากนั้นชุดคำทั้งหมดจะนำไปทดสอบผลการวิเคราะห์ความรู้สึกเพื่อคัดเลือกชุดคำที่ดีที่สุดสำหรับการสร้างคลังศัพท์ ชุดคำบอกความรู้สึกที่ให้ผลการวิเคราะห์ดีที่สุดมาจากการใช้คำในกลุ่มคำกริยา คำคุณศัพท์ และคำวิเศษณ์เป็นคำบอกความรู้สึก เลือกใช้คำบอกลักษณะที่มีลำดับเปอร์เซ็นต์ไทล์ของความถี่ตั้งแต่ 90 ขึ้นไป และคัดเลือกคำบอกความรู้สึกที่มีผลรวมของค่า TF-IDF เชิงบวกและลบมากกว่าหรือเท่ากับ 0 หลังจากนั้นคำบอกความรู้สึกในคลังศัพท์ที่ได้จากแต่ละโดเมนจะนำมา

จำแนกประเภทเป็นคำบอกความรู้สึกแบบเจาะจงโดเมนและแบบไม่เจาะจงโดเมน รายการคำที่ได้จะนำมาวิเคราะห์โดยเปรียบเทียบการปรากฏในเนื้อหาบทวิจารณ์โดเมนต่าง ๆ เพื่อศึกษาความแตกต่างของการเลือกใช้คำบอกความรู้สึกในบทวิจารณ์สินค้าของแต่ละโดเมน ผลการวิเคราะห์พบว่าการใช้คำบอกความรู้สึกของผู้เขียนบทวิจารณ์จะเปลี่ยนไปตามความคาดหวังของผู้ใช้สินค้าหรือบริการและสไตล์การเขียนบทวิจารณ์ในโดเมนนั้น ๆ นอกจากนี้การวิเคราะห์ข้อมูลบทวิจารณ์ยังแสดงให้เห็นถึงปัจจัยสำคัญต่าง ๆ ที่ส่งผลต่อรายการคำบอกความรู้สึกในคลังศัพท์ที่ได้จากงานวิจัยนี้ ซึ่งอาจนำไปใช้ปรับปรุงวิธีการสร้างคลังศัพท์ให้ดีขึ้นได้ ประกอบด้วย การปรากฏของคำบอกลักษณะสินค้า ช่วงคะแนนของบทวิจารณ์เชิงบวกและเชิงลบ แรงจูงใจในการเขียนบทวิจารณ์ และการเลือกใช้สินค้าหรือบริการในโดเมนต่าง ๆ

สมัคร ชัยสงวน (2561) การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษานานาชาติโดยใช้ตัวจำแนกข้อมูล Naive Bayes สำหรับภาษาไทย และ 2) เพื่อประเมินมาตรการความถูกต้องของค่าความแม่นยำและค่าเรียกคืนของระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษานานาชาติโดยใช้ตัวจำแนกข้อมูล Naive Bayes สำหรับภาษาไทย งานวิจัยนี้รวบรวมข้อมูลจากเฟสบุ๊คเพจของมหาวิทยาลัยสุโขทัยธรรมมาธิราช ซึ่งเป็นข้อมูลการแสดงความรู้สึกของนักศึกษาที่ติดตามเฟสบุ๊คเพจของมหาวิทยาลัย เพื่อทำการวิเคราะห์ความรู้สึกของนักศึกษาแบบเรียลไทม์ด้วย Naive Bayes จากหลักการพื้นฐานของการประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) ด้วยวิธีการตัดคำภาษาไทย และประยุกต์ตัวจำแนกข้อมูล Naive Bayes โดยผลลัพธ์ทั้งหมดสามารถแสดงผลแบบเรียลไทม์กับระบบจัดการคลังคำศัพท์ สรุปผลการวิเคราะห์ความรู้สึก เครื่องมือทดสอบประโยค และระบบแจ้งเตือนข้อความผ่านทางไลน์ ผลการวิจัยนี้ได้ผลลัพธ์ ค่าความแม่นยำ ร้อยละ 80.03 ค่าเรียกคืน ร้อยละ 90.52 ค่าความถูกต้อง 89.63 และการวัดประสิทธิภาพโดยรวม ร้อยละ 84.81 ตามลำดับ

สันติ สุขเกษม (2561) การวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม โดยใช้เทคนิคการตัดคำแบบผสมผสาน มีวัตถุประสงค์ในการทำวิจัย ดังนี้ 1) เพื่อพัฒนาเทคนิคการตัดคำแบบผสมผสาน 2) เพื่อวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม 3) เพื่อประเมินประสิทธิภาพการวิเคราะห์ความคิดเห็น โดยการจำแนกแบบ Naive Bayes, Decision Tree และ K nearest neighbor และดำเนินการโดยรวบรวมข้อมูลจากเว็บไซต์ Agoda เป็นข้อมูลการแสดงความคิดเห็นของผู้ใช้บริการที่เข้าพักโรงแรมนำมาวิเคราะห์ความรู้สึกโดยหลักการวิธีการตัดคำแบบผสมผสาน ซึ่งผลของการวิจัยมีดังนี้ ผลการวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม โดยใช้เทคนิคการตัดคำแบบผสมผสาน การวิเคราะห์จำแนกแบบ Naive Bayes ผลวิเคราะห์ความคิดเห็นเชิงบวก (Positive) 99.36%, ผลวิเคราะห์ความคิดเห็นเชิงลบ (Negative) 98.89%, ผลวิเคราะห์ความคิดเห็นเป็นกลาง (Neutral) 99.43%, ความถูกต้อง (Accuracy) 99.66% การวิเคราะห์การจำแนกแบบ Decision

Tree ผลวิเคราะห์ความคิดเห็นเชิงบวก (Positive) 100%, ผลวิเคราะห์ความคิดเห็นเชิงลบ (Negative) 0.00%, ผลวิเคราะห์ความคิดเห็นเป็นกลาง (Neutral) 0.00%, ความถูกต้อง (Accuracy) 65.49% และการวิเคราะห์การจำแนกแบบ K nearest neighbor ผลวิเคราะห์ความคิดเห็นเชิงบวก (Positive) 83.68%, ผลวิเคราะห์ความคิดเห็นเชิงลบ (Negative) 26.16%, ผลวิเคราะห์ความคิดเห็นเป็นกลาง (Neutral) 45.00%, ความถูกต้อง (Accuracy) 70.21%

ณิชภา รัตนปริชาเวช (2562) ได้ทำการศึกษาเรื่องความสัมพันธ์ระหว่างประเด็นต่าง ๆ ที่ถูกพูดถึงในโซเชียลมีเดีย กรณีศึกษาเว็บบอร์ดกลุ่มแม่และเด็ก โดยใช้ข้อมูลความคิดเห็นทั้งหมดที่เกิดขึ้นในเว็บบอร์ด Pantip.com ห้องชานเรื่อน ผลการวิจัยพบว่า ผู้บริโภคกลุ่มแม่และเด็กส่วนใหญ่พูดคุยกันในประเด็นเรื่องค่าใช้จ่ายของโรงเรียน, ความกังวลต่อการตั้งครรภ์ และค่าใช้จ่ายของโรงพยาบาลที่ Jaccard Similarity 0.2, 0.19, 0.19 ตามลำดับทั้งยังมีการแสดงความเครียดและกังวลกับทุกเรื่อง Normalized Degree Centrality สูงสุด 0.648 โดยเฉพาะสถานการณ์ท้องไม่พร้อม

ชนกร ญาณกาย และวนิดา แก่นอากาศ (2562) การทำเหมืองข้อมูลเป็นวิธีการหนึ่งที่ใช้ในการค้นหาองค์ความรู้ในข้อมูล ข้อมูลประเภทข้อความ ข้อมูลประเภทที่สามารถค้นหาองค์ความรู้ได้หลากหลายรูปแบบ เช่น การสรุปข้อความ การหาความหมายแฝง การหาหัวข้อการจัดกลุ่มข้อความ Latent Dirichlet Allocation : LDA เป็นเทคนิคหนึ่งที่ใช้ในการค้นหาหัวข้อ (Topic) ของข้อมูล และสามารถทำการเพิ่มประสิทธิภาพได้ด้วยการปรับปรุงค่าโดยใช้ Optimization อัลกอริทึมซึ่งผู้วิจัยใช้ Ant colony optimization ในการปรับค่าตัวแปร ซึ่งการค้นหาหัวข้อจากเอกสาร มักจะใช้เวลาในการคำนวณค่านาน ผู้วิจัยจึงประยุกต์ใช้ map-reduce ซึ่งเป็นการทำงานภายใต้สภาพแวดล้อมของ Hadoop มาช่วยในการประมวลผลเพื่อให้สามารถทำงานได้เร็วขึ้นและทำการวัดค่าประสิทธิภาพของอัลกอริทึม LDA ผลการวิจัยพบว่า การประมวลผลชุดข้อมูลด้วยอัลกอริทึม LDA ที่ปรับปรุงค่าตัวแปรโดย ACO ที่ทำงานโดย Map-reduce มีความเร็วในการประมวลผลที่สูงขึ้น

พิศิษฐ์ บวรเลิศสุธี และวรัทธร ไพรีเกรง (2562) การตัดสินใจซื้อสินค้าและบริการออนไลน์ ผู้บริโภคส่วนใหญ่จะทำการสืบค้นข้อมูลเกี่ยวกับสินค้าและบริการรวมทั้งพิจารณาจากการแสดงความคิดเห็น (Comments) และบทวิจารณ์ (Reviews) จากผู้ที่เคยซื้อสินค้าไปแล้ว เพื่อเป็นแนวทางในการตัดสินใจเลือกซื้อ อย่างไรก็ตาม ข้อมูลการแสดงความคิดเห็นและบทวิจารณ์ในปัจจุบันนี้มีปริมาณมหาศาลและเกิดขึ้นใหม่อยู่ตลอดเวลา ซึ่งทำให้ผู้บริโภคต้องเสียเวลาในการวิเคราะห์ข้อดีข้อเสียของผลิตภัณฑ์เหล่านี้ด้วยตนเองเพื่อใช้ในการตัดสินใจเลือกซื้อสินค้าและบริการ ซึ่งถ้าหากมีการสร้างตัวแบบเพื่อแจ้งให้ผู้ใช้ทราบถึงการวิเคราะห์ความรู้สึกได้ จะช่วยให้ผู้ใช้ตัดสินใจได้รวดเร็วมากยิ่งขึ้น ดังนั้น งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์ สำหรับจำแนกประเภทบทความแนะนำสินค้าออนไลน์โดยใช้เทคนิคส่วนการเรียนรู้ของเครื่อง (Machine Learning) ร่วมกับเทคนิคการตัดคำ (Word Tokenization) และการสร้างคลังคำศัพท์

(Bag of Words) จากนั้นนำเข้ากระบวนการจำแนกประเภทผลการวิเคราะห์ 4 เทคนิค ได้แก่ LSTM, SGD, Logistic Regression และ Support Vector Machines ในการสร้างตัวแบบมีการทดลอง 5 ขั้นตอน ได้แก่ 1) การเตรียมข้อมูล 2) การตัดคำ 3) การฝึกอบรมข้อมูล 4) ขั้นตอนการแยกประเภท และ 5) การประเมินตัวแบบ ทำการทดลองโดยใช้ข้อมูลตัวอย่างการแสดงความคิดเห็นต่อสินค้าและบริการออนไลน์ภาษาไทย จำนวน 12,900 ข้อมูล ใช้เป็นข้อมูลในการสร้างตัวแบบ เพื่อช่วยให้ผู้บริโภคใช้ในการตัดสินใจซื้อสินค้าและบริการ และช่วยให้ผู้ประกอบการมีข้อมูลสำหรับการพัฒนาสินค้าและบริการต่อไปในอนาคต จากผลการทดลองสร้างตัวแบบการวิเคราะห์ระดับความรู้สึกทางอารมณ์ 3 ระดับ คือ ความรู้สึกเชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) โดยแต่ละตัวแบบให้ค่าความถูกต้องในการทำนายผล ดังนี้ LSTM 81.27%, Logistic Regression 69%, SGD 66% และ Support Vector Machines 65% สรุปได้ว่าการสร้างตัวแบบโดยใช้เทคนิคการจำแนกประเภทแบบ LSTM ให้คะแนนความถูกต้องสูงที่สุดในการทำนายผลด้วยการเรียนรู้เชิงลึก จึงเหมาะสมแก่การนำไปใช้สร้างตัวแบบในการวิเคราะห์ระดับความรู้สึกกับข้อความที่เป็นภาษาไทยโดยพิจารณาจากค่า F-Measure (F1)

ปิยวรรณ ทองพลอย (2563) ปัจจุบันมีผู้ใช้งานอินเทอร์เน็ตจำนวนมากเข้าปรึกษาปัญหาเกี่ยวกับแพทย์ผ่านชุมชนออนไลน์ (Community Question Answering : CQA) ซึ่งเป็นงานที่หนักสำหรับแพทย์ที่ต้องตอบคำถามให้ทัน ซึ่งปัจจุบันมีการนำระบบหุ่นยนต์โต้ตอบการสนทนา (Chatbot) มาประยุกต์ใช้ในการให้ข้อมูล ถามตอบปัญหา แต่การพัฒนาหุ่นยนต์โต้ตอบการสนทนา (Chatbot) นั้นมีข้อจำกัด เช่น มีราคาแพง และเป็นงานที่ทำหายในการที่จะทำให้คอมพิวเตอร์เข้าใจถึงภาษามนุษย์จากข้อความในเอกสารที่เป็นภาษาไทย งานวิจัยนี้จึงมีวัตถุประสงค์ ในการวิเคราะห์หาค่าที่มียุทธศาสตร์และการจำแนกหัวข้อในข้อความที่มีความคล้ายคลึงกัน เพื่อสร้างแบบจำลองที่สามารถนำไปพัฒนาระบบโต้ตอบอัตโนมัติ (Chatbot) ให้สามารถโต้ตอบกับผู้ใช้งานได้ตรงประเด็นมากขึ้น โดยใช้วิธีการสร้างแบบจำลองหัวข้อ (Topic Modeling) และการจำแนกกลุ่มข้อความ (Clustering) โดยงานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการจัดสรรดีริคเคลแฝง (Latent Dirichlet Allocation : LDA) ในการวิเคราะห์หาค่าที่มียุทธศาสตร์ และจำแนกหัวข้อในข้อความเกี่ยวกับการตั้งครรภ์ การมีเพศสัมพันธ์ และการคุมกำเนิด ในส่วนของการประเมินประสิทธิภาพได้ใช้การวัดผลแบบ extrinsic evaluation โดยใช้เทคนิค K-means ในการจัดกลุ่มหัวข้อและประเมินประสิทธิภาพการจำแนกกลุ่มด้วย Silhouette Coefficient

สุธิมา แสงจันทร์ (2563) ได้ทำการศึกษาเรื่องการวิเคราะห์ความรู้สึกแต่ละแง่มุมของผู้ใช้บริการส่งอาหารในประเทศไทย ผู้วิจัยได้สร้างตัวแบบสำหรับการวิเคราะห์ความรู้สึกแต่ละแง่มุมของผู้ใช้บริการส่งอาหารในประเทศไทย โดยสร้างตัวแบบหลัก 2 ตัวแบบ คือ ตัวแบบสำหรับการระบุ

แง่มุม ซึ่งตัวแบบซัพพอร์ตเวกเตอร์แมทซึนให้ประสิทธิภาพที่ดีที่สุด และตัวแบบสำหรับการระบุความรู้สึกซึ่งตัวแบบซัพพอร์ตเวกเตอร์แมทซึน และนาอิวเบย์ ให้ประสิทธิภาพต่างกันในแต่ละแง่มุม

พนิตา กาศกลางดอน (2563) การรับรู้ถึงปัญหาและการแก้ปัญหาที่ไม่ตรงจุดในการตอบสนองความต้องการของผู้ใช้บริการได้ทันที่ถือเป็นปัญหาสำคัญที่ส่งผลกระทบต่อคุณภาพการให้บริการของโรงพยาบาลในปัจจุบัน สาเหตุหลักของปัญหา เนื่องจากมีผู้ใช้บริการจำนวนมากที่ได้รับปัญหาจากการบริการในหลาย ๆ ด้านของโรงพยาบาล และด้วยวิธีการต่าง ๆ ในการรวบรวมข้อมูลของผู้ใช้บริการและการได้มาซึ่งข้อมูลล่าช้าซึ่งไม่ได้สะท้อนถึงปัญหาที่แท้จริง งานวิจัยนี้นำเสนอแนวทางแก้ไขปัญหาโดยรวบรวมความคิดเห็นของผู้มีประสบการณ์การใช้บริการจากเว็บบอร์ด ซึ่งรวดเร็วและได้รับข้อมูลที่แท้จริงจากการรีวิว ข้อมูลเหล่านี้จะถูกวิเคราะห์เพื่อหาความรู้เชิงลึกโดยการใช้โมเดลการเรียนรู้ของเครื่อง ด้วยข้อมูลข้อความจากประสบการณ์ผู้ป่วย และใช้ข้อมูลที่รวบรวมมาสร้างรายงานผลผ่านแดชบอร์ด ผลการเปรียบเทียบใช้การทดสอบการจำลองโดยใช้เทคนิคอัลกอริธึม 3 เทคนิค คือ นาร์อิวเบย์ (Naive Bayes), การสุ่มป่าไม้ (Random Forest) และโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้น แบบยาว (Long-Short-Term Memory) จากผลการทดสอบพบว่าอัลกอริธึม Random Forest มีประสิทธิภาพสูงสุดในการทำนายความคิดเห็นเชิงลบโดยมีค่าการเรียกคืน (Recall) ที่ 0.71 และผลวิเคราะห์จากการนำข้อมูลของโรงพยาบาลรัฐและเอกชนเพื่อหาข้อมูลเชิงลึก ในภาพรวมการวิเคราะห์ของระบบจัดการคุณภาพ โรงพยาบาลบนรายงานพบว่าในด้านบุคลากร ด้านการจัดการโครงสร้าง ด้านการดำเนินงาน ด้านการเงิน และด้านอื่น ๆ ของโรงพยาบาลรัฐถูกพูดถึงในเชิงลบมากกว่าโรงพยาบาลเอกชนทั้ง 5 ด้าน ด้านที่ชัดเจนที่สุดคือ โครงสร้างโรงพยาบาล

สหชัย งามชัยภูมิ (2563) บทวิจารณ์ความคิดเห็นของลูกค้าส่วนใหญ่มีรูปแบบเป็นข้อความสั้น ดังนั้นความยาวของข้อความที่มีอยู่จำกัด เป็นความท้าทายสำหรับการจำแนกบทวิจารณ์ของลูกค้า เพราะจำนวนคำที่แสดงในข้อความมีจำนวนน้อย ทำให้ไม่สามารถคัดเลือกคุณลักษณะที่เหมาะสมและมีความหมาย หรืออาจจะสกัดได้น้อยเกินไปจนยากต่อการสร้างตัวจำแนกความรู้สึกจากข้อความที่มีคุณภาพต่อการใช้งานที่ดีได้งานวิจัยนี้ได้นำเสนอกระบวนการจำแนกบทวิจารณ์ที่มีลักษณะข้อความสั้น ด้วยการสร้างโมเดลแบบผสมผสานด้วย 3 เทคนิค คือ Support Vector Machine, Naive Bayes และ K nearest neighbor เพื่อเพิ่มประสิทธิภาพ และความแม่นยำในการจำแนกความรู้สึกจากเอกสารของข้อความที่มีข้อความสั้น โดยการวัดประสิทธิภาพด้วยค่าความถูกต้อง = 0.97 ค่าความแม่นยำ = 0.98 ค่าความระลึก = 0.97 และ ค่า F-measure (F1) = 0.97

วสุวัตต์ อินทร์แปลง (2563) เหมืองข้อความ เป็นกระบวนการวิเคราะห์ข้อมูลตัวอักษรเพื่อสกัดข้อมูลที่เป็นประโยชน์จากแหล่งข้อมูล ปัจจุบันเทคนิคในการจำแนกเหมืองข้อความมีหลายวิธี งานวิจัยนี้มีวัตถุประสงค์เพื่อค้นหาเทคนิคการจำแนกจาก 4 เทคนิคที่มีประสิทธิภาพ คือ เทคนิค

Naive Bayes เทคนิค Support Vector Machine (SVM) เทคนิค K-Nearest Neighbor และเทคนิค C4.5 โดยเก็บรวบรวมข้อมูลความคิดเห็นต่อเกมมือถือฉบับจีจำนวน 3,798 ข้อความ ในกระบวนการคัดเลือกคำบ่งชี้ เพื่อใช้ในการแยกคุณลักษณะได้เลือกใช้คำวิเศษณ์ และคำสแลงบางคำที่ความหมายของคำเป็นคำวิเศษณ์เพื่อทำการแยกคุณลักษณะเชิงบวกและเชิงลบ ผลการศึกษาพบว่ามีความไม่สมดุลของคลาสในข้อมูล โดยมีจำนวนคลาสหนึ่งมากกว่าอีกคลาสหนึ่งเป็นจำนวนมาก ผู้วิจัยจึงได้ทำการแก้ปัญหา โดยการปรับความสมดุลของข้อมูลด้วยวิธี Synthetic Minority Over-sampling Technique : SMOTE และใช้หลักการ 10-fold cross validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพการจำแนกของแบบจำลองด้วยค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) เมื่อทำการทดสอบและวัดประสิทธิภาพของโมเดลพบว่า เทคนิค Support Vector Machine ให้ผลดีที่สุดในการวิเคราะห์ความคิดเห็น โดยให้ค่าความถูกต้อง 99.65% ค่าความไว 99.30% และค่าความจำเพาะ 100%

วันสวรรคต มีประเสริฐ และเอกรัฐ รัฐกาญจน์ (2564) การวิเคราะห์ความต้องการหรือการหาความรู้สึกของลูกค้า เป็นสิ่งที่บริษัท E-Commerce ควรรับฟังและนำไปปรับปรุงแก้ไขเพื่อสร้างความพึงพอใจให้แก่ลูกค้าและยังคงใช้บริการต่อไป ซึ่งงานวิจัยนี้มีวัตถุประสงค์เพื่อหาวิธีการวิเคราะห์ความคิดเห็นจากทวิตเตอร์ของลูกค้าที่มีต่อบริษัทช้อปปิ้งประเทศไทยโดยแยกแต่ละประเด็นสำคัญ ซึ่งแบ่งออกเป็น 5 ประเด็นคือ 1) บริษัท 2) ขนส่ง 3) โปรโมชั่น 4) ร้านค้า 5) ระบบผลของการวิเคราะห์ความคิดเห็นของลูกค้าโดยแยกหัวข้อประเด็นสำคัญพบว่าวิธีที่เหมาะสมที่สุดคือ Random Forest ส่วนการแยกความรู้สึกแต่ละประเด็นสำคัญนั้น วิธีที่ให้ประสิทธิภาพสูงที่สุดในการวิเคราะห์ข้อความภาษาไทย คือ WangchanBERTa นอกจากนี้ภาพรวมของการวิเคราะห์ความรู้สึกแต่ละประเด็นของลูกค้าบริษัทช้อปปิ้งในเชิงลึก พบว่าบริษัทช้อปปิ้งควรให้ความสำคัญกับการปรับปรุงและพัฒนากลยุทธ์ในเรื่องของขนส่งและระบบเป็นอันดับแรกเพื่อสร้างความพึงพอใจให้แก่ลูกค้าและเป็นผู้นำในด้านการให้บริการ E-Commerce

พัฒนพันธ์ กลิ่นหอม (2564) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษามุมมองของผู้บริโภคผ่านการแสดงความคิดเห็นของผู้บริโภคที่มีต่อสินค้าสมาร์ตโฟนบนสื่อสังคมออนไลน์ในประเทศไทย โดยผู้วิจัย ได้พัฒนาเครื่องมือ เพื่อใช้ในการวิเคราะห์ข้อมูลมหัตที่ไม่มีโครงสร้างในสื่อสังคมออนไลน์แพลตฟอร์มบริการเทคโนโลยีปัญญาประดิษฐ์ภายใต้แนวคิด “AI สัญชาติไทย” เพื่อตอบโจทย์ความต้องการใช้งานในประเทศไทย กรณีศึกษา : บริษัทผู้ผลิตสมาร์ตโฟนชั้นนำในด้านของยอดขายสามบริษัทในประเทศไทยที่มีการนำเสนอข้อมูลสมาร์ตโฟนผ่านสื่อสังคมออนไลน์ ผู้วิจัยเลือกสื่อสังคมออนไลน์ที่ได้รับความนิยมในประเทศไทย ได้แก่ เฟซบุ๊ก แพนเพจ และ ยูทูบ แชนแนล โดยการเก็บข้อมูลจากคอมเมนต์ ในสื่อสังคมออนไลน์ย้อนหลังระยะเวลา 5 ปีจำนวนมากว่าหนึ่งล้าน

ความคิดเห็นที่แตกต่างกัน ซึ่งเครื่องมือที่ผู้วิจัยพัฒนาขึ้นมาเสนอผ่านข้อมูลมหัตที่มีโครงสร้างสามารถจำแนกหมวดหมู่ของข้อความได้โดย แสดงในรูปแบบความถี่ของคำที่ผู้บริโภคพูดถึงเกี่ยวกับคุณลักษณะของสมาร์ทโฟน ความถี่ของการแสดงความคิดเห็นเชิงบวกและเชิงลบของผู้บริโภค และความถี่ของการพูดถึงเกี่ยวกับธุรกิจที่เกี่ยวข้องกับสมาร์ทโฟนของผู้บริโภค

ปิยนตร อรุณปรีดี (2565) “ภูเก็ต” เป็นจังหวัดขนาดเล็ก ๆ ในประเทศไทยแต่เป็นแหล่งรายได้หลักด้านการท่องเที่ยวอันดับต้น ๆ ของประเทศไทย เพราะจังหวัดภูเก็ตมีแหล่งท่องเที่ยว ทั้งทางธรรมชาติ อาหาร และกิจกรรมนันทนาการต่าง ๆ ที่สามารถดึงดูดนักท่องเที่ยวทั้งในและต่างประเทศได้อย่างมหาศาล ในโลกยุคปัจจุบัน อินเทอร์เน็ตเข้ามามีส่วนร่วมในการค้นหาข้อมูลหรือวางแผนการท่องเที่ยวเป็นอย่างมาก ยกตัวอย่างได้จากเว็บที่เกี่ยวข้องกับท่องเที่ยว สื่อการท่องเที่ยวต่าง ๆ ที่พัฒนารูปแบบการนำเสนอให้มีส่วนของบทวิจารณ์ออนไลน์เพื่อเป็นสื่อหลักในการแลกเปลี่ยนความคิดเห็นหรือแบ่งปันประสบการณ์ที่นักท่องเที่ยวที่ได้สัมผัสกับสถานที่เหล่านั้นจริง ๆ การเข้าถึงข้อมูลเหล่านี้กลายเป็นเครื่องมือในการตัดสินใจของนักท่องเที่ยวได้อย่างดีเยี่ยม แต่ปัญหาของบทวิจารณ์ออนไลน์นั้นคือ ข้อมูลจำนวนมาก ภาษาที่นักวิจารณ์ใช้ หรือสิ่งที่ผู้วิจารณ์ต้องการสื่อที่แท้จริง ดังนั้นนักวิจัยจึงทำการสกัดความรู้จากบทวิจารณ์ออนไลน์ของสถานที่ท่องเที่ยวในจังหวัดภูเก็ตจากเว็บไซต์ TripAdvisor ระหว่าง ปี พ.ศ. 2553 ถึง พ.ศ. 2563 มีสถานที่ท่องเที่ยวทั้งหมด 190 สถานที่ จำนวนบทวิจารณ์ทั้งหมด 76,183 บทวิจารณ์ โดยดำเนินการวิเคราะห์ในหลายรูปแบบ เช่น หากกฎความสัมพันธ์ (Association Rule) โดยใช้อัลกอริทึม FP-Growth ใน Rapid Miner เพื่อหาความเกี่ยวข้องของข้อมูล และการทำเหมืองข้อมูล (Data Mining) โดยใช้การวิเคราะห์ความถี่ของคำ (Text Frequency Analysis) ทั้งยังค้นหาประเภทของคำในประโยค (Part of Speech - POS) เพื่อสกัดหาสิ่งที่นักท่องเที่ยวต้องการ หรือกล่าวถึงมากที่สุดร่วมกับการใช้เทคนิค Term Frequency – Inverse Document Frequency (TF-IDF) ในการพิจารณาความสำคัญของคำประโยค ทั้งยังดำเนินการวิเคราะห์ความรู้สึก (Sentiment Analysis) เพื่อวิเคราะห์ความรู้สึกที่แท้จริงของนักท่องเที่ยว โดยท้ายสุดจะสรุปให้อยู่ในรูปแบบของ Dashboard เพื่อให้นักท่องเที่ยวหรือหน่วยงานองค์กรที่เกี่ยวข้อง สามารถเข้าถึงและใช้ประโยชน์ จากการวิเคราะห์ข้อมูลเหล่านี้ได้อย่างรวดเร็วและมีประสิทธิภาพ เพื่อเข้ามาเป็นส่วนหนึ่งในการพัฒนาการท่องเที่ยวได้อย่างยั่งยืน

3. กระบวนการในการทำวิจัย

งานวิจัยนี้ศึกษาการทำเหมืองข้อมูล (Data Mining) รวบรวมข้อมูลความคิดเห็น (Comment) ของผู้ใช้งานบนแพลตฟอร์มทวิตเตอร์ แชนแชตของกินเชียงใหม่ และแชตแท็กรีวิวเชียงใหม่ ด้วยเทคนิคการเรียนรู้ Machine Learning ใช้ข้อมูลในช่วงวันที่ 1 พฤษภาคม 2565 ถึงวันที่ 30 เมษายน 2566 โดยใช้ โปรแกรมเอสเซนส์ (S-Sense: Social Sensing) ที่ถูกคิดค้นและ

พัฒนาขึ้นมาด้วยเทคนิคการประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP ควบคู่กับใช้โปรแกรม Orange Data mining วิเคราะห์ข้อมูลรูปแบบการใช้เป็นลักษณะการเขียนโปรแกรมด้วยภาพ (Visual Programming) เพื่อให้สามารถวิเคราะห์ข้อความที่ใช้ภาษาพูดและไม่เป็นทางการได้อย่างมีประสิทธิภาพ และใช้เครื่องมือการวิเคราะห์ความรู้สึก (Sentiment Analysis) ที่สามารถวิเคราะห์ความคิดเห็นนั้นว่าบ่งบอกความรู้สึกในเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral) และใช้เครื่องมือการวิเคราะห์การสร้างแบบจำลองหัวข้อ (Topic Modeling) จากการสกัดข้อมูล ตัดคำที่ไม่มี ความหมาย หรือคัดแยกข้อมูล กำหนดกลุ่มคำที่มีความคล้ายคลึงมารวมเป็นกลุ่มเดียวกัน

โดยผู้วิจัยใช้การวิเคราะห์ความรู้สึก (Sentiment Analysis) จากการแสดงความคิดเห็น (Comment) ของผู้ใช้งานบนแพลตฟอร์มทวิตเตอร์ออกเป็น 3 ความรู้สึก ได้แก่

- 1) ความรู้สึกเชิงบวก (Positive)
- 2) ความรู้สึกเชิงลบ (Negative)
- 3) ความรู้สึกเป็นกลาง (Neutral)

และสร้างแบบจำลองหัวข้อ (Topic Modeling) กำหนดกลุ่มคำที่มีความคล้ายคลึงมารวมเป็นกลุ่มเดียวกัน และหาคำที่ผู้ใช้งานบนแพลตฟอร์มทวิตเตอร์มีการกล่าวถึงมากที่สุด

บทที่ 3

วิธีการวิจัย

การศึกษาวิจัยในหัวข้อการวิเคราะห์ความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่ ใช้การวิเคราะห์ข้อมูลผ่านคอมพิวเตอร์ โดยผู้วิจัยจะทำการรวบรวมความคิดเห็นของผู้ใช้งานบนทวิตเตอร์ ในบทนี้จะกล่าวถึงวิธีการวิจัยเพื่อให้ได้ตามวัตถุประสงค์ที่ตั้งไว้ มีขั้นตอนวิธีการวิจัยดังนี้

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือในการวิจัย
3. วิธีการเก็บรวบรวมข้อมูล
4. การวิเคราะห์ข้อมูล
5. การประเมินแบบจำลอง

1. ประชากรและกลุ่มตัวอย่าง

ข้อมูลจากความคิดเห็นเกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ ชุดข้อมูลในช่วงวันที่ 1 พฤษภาคม 2565 ถึง วันที่ 30 เมษายน 2566

2. เครื่องมือในการวิจัย

เครื่องมือที่ใช้ในงานวิจัยการวิเคราะห์ความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ โดยใช้โปรแกรมและเครื่องมือในการวิเคราะห์ดังต่อไปนี้

2.1 โปรแกรมเอสเซนส์ (S-Sense : Social Sensing)

เป็นโปรแกรมที่ใช้ในการวิเคราะห์ ประมวลผลข้อความบนเว็บไซต์เครือข่ายเชิงสังคมและเว็บบอร์ด เช่น Facebook, Twitter, YouTube, Pantip.com เป็นต้น เนื่องจากข้อความส่วนใหญ่บนอินเทอร์เน็ตนิยมใช้ภาษาที่ไม่เป็นทางการหรือภาษาพูด รวมทั้งมีโครงสร้างประโยคที่ไม่ถูกต้องตามหลักไวยากรณ์ภาษาไทย จึงทำให้ยากต่อการวิเคราะห์ ดังนั้นเทคโนโลยีของ S-Sense จึงถูกคิดค้นและพัฒนาขึ้นมาด้วยเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) และการทำเหมืองข้อมูล (Data Mining) เพื่อให้สามารถวิเคราะห์ข้อความที่ใช้ภาษาพูดและไม่เป็นทางการได้อย่างมีประสิทธิภาพ ระบบ S-Sense สามารถนำมาช่วยธุรกิจในการรับรู้กิจกรรม

รวมถึงการติดตามข้อมูลในโซเชียลมีเดียเพื่อการปรับปรุงแผนการตลาดและการคิดแคมเปญและติดตามทัศนคติของสาธารณะที่มีต่อสินค้าหรือบริการ เพื่อให้เข้าใจถึงความต้องการของลูกค้าได้ดี

การวิเคราะห์ความรู้สึกของผู้บริโภคในรูปแบบการแสดงความคิดเห็นบนโซเชียลมีเดียจากแพลตฟอร์มต่าง ๆ เพื่อประเมินและเปรียบเทียบการวิเคราะห์ค่าความเชื่อมั่นเพื่อทดสอบและเปรียบเทียบวิธีการใช้พจนานุกรมและการเรียนรู้ของเครื่องกับการวิเคราะห์ความรู้สึก การค้นหาความเคลื่อนไหวของราคาหุ้นที่มีความสัมพันธ์กับความคิดเห็น (ความรู้สึก) ของสาธารณชนที่มีต่อบริษัทนั้น อัลกอริทึมที่ใช้พิจารณาความรู้สึกความคิดเห็น ข่าว และราคาหุ้นในอดีตเพื่อคาดการณ์ราคาหุ้นในอนาคต ดำเนินการโดยใช้การเรียนรู้ของเครื่องด้วยโมเดล Deeplearning, Support Vector Machine, MNB, linear regression, Naive Bayes and LSTM (Long Short-Term Memory) ซึ่งให้ผลลัพธ์ที่สามารถยืนยันความสำเร็จได้ ในส่วนของการเลือกใช้ตัวจำแนกประเภทที่มีความนิยมในการพัฒนา Sentiment Analysis ภาษาไทย ได้แก่ Logistic Regression, Random Forest, Support Vector Machine, Naive Bayes, K-Nearest Neighbors และ Decision Tree เป็นต้น และในการประเมินประสิทธิภาพสรุปจาก Accuracy, Precision, Recall และ F-Measure (F1)

2.2 โปรแกรม Orange Data mining

Orange Data mining เป็นซอฟต์แวร์โอเพนซอร์สสำหรับการเรียนรู้ของเครื่องจักรและแสดงผลข้อมูล สำหรับการสร้างกระบวนการวิเคราะห์ข้อมูล (Data Analytics workflow) รูปแบบการใช้เป็นลักษณะการเขียนโปรแกรมด้วยภาพ (Visual Programming)

2.3 เครื่องมือการวิเคราะห์ การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นเทคนิคการประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP ในการสกัดข้อความที่เป็นความคิดเห็นมาประยุกต์ใช้เพื่อมุ่งเน้นการวิเคราะห์ และตรวจสอบความรู้สึกได้อย่างอัตโนมัติ โดยมีข้อมูลเป็นข้อความแสดงความคิดเห็นของผู้ใช้งาน ที่ได้แสดงความคิดเห็นไว้ในแพลตฟอร์มต่าง ๆ ทั้งในกระดานสนทนา (Web board) หรือในสื่อสังคมออนไลน์ เช่น Twitter, Facebook โดยนำข้อมูลมาวิเคราะห์และจำแนกประเภทของความรู้สึก เช่นความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral) (นันทกัศ สุทธิเลิศ, 2560)

2.4 เครื่องมือการวิเคราะห์ แบบจำลองหัวข้อ (Topic Modeling)

แบบจำลองหัวข้อ (Topic Modeling) คือ แบบจำลองทางสถิติประเภทหนึ่งสำหรับการค้นหา "หัวข้อ" ซึ่งนามธรรมที่เกิดขึ้นในชุดเอกสาร เมื่อพิจารณาว่าเอกสารเกี่ยวกับหัวข้อใดหัวข้อหนึ่ง โดยเทคนิคการสร้างแบบจำลองหัวข้อคือการวิเคราะห์ข้อมูลตัวอักษรเพื่อกำหนดกลุ่มคำที่มีความคล้ายคลึงมารวมเป็นกลุ่มเดียวกัน แบบจำลองหัวข้อรวบรวมสัญญาตญาณนี้ไว้ในกรอบทางคณิตศาสตร์ ซึ่งช่วยให้ตรวจสอบชุดของเอกสารและค้นพบตามสถิติของคำในแต่ละคำ เพื่อให้เราเข้าใจชุดข้อความที่ไม่มีโครงสร้างจำนวนมาก

3. วิธีการเก็บรวบรวมข้อมูล

ในการเก็บรวบรวมข้อมูล ผู้วิจัยจะทำการเก็บรวบรวมข้อมูลจากการแสดงความคิดเห็นจากผู้ใช้งานทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ ใช้ชุดข้อมูลในช่วงวันที่ 1 พฤษภาคม 2565 ถึง วันที่ 30 เมษายน 2566 ทั้งนี้ข้อความที่ได้ทำการเก็บรวบรวมนั้นจะเป็นข้อความที่มีลักษณะเป็นภาษาไทยที่ได้มีการแสดงความคิดเห็น (Comment) บนทวิตเตอร์แฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่เท่านั้น ซึ่งจำนวนของความคิดเห็นที่ได้เก็บรวบรวมนั้นมีจำนวนทั้งสิ้น 2,213 ความคิดเห็น โดยการเก็บรวบรวมข้อมูลนั้นจะเป็นการดึงข้อมูลความคิดเห็นของผู้บริโภคบันทึกเข้าสู่โปรแกรม Microsoft Excel ดังแสดงในภาพที่ 1-3

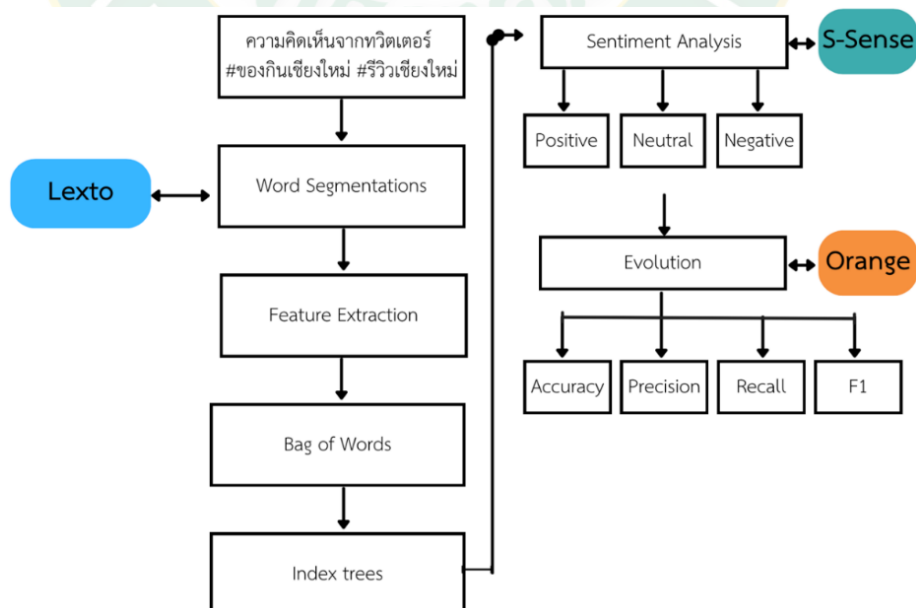


ภาพที่ 1 ความคิดเห็นของผู้บริโภค

ตารางที่ 1 ตัวอย่างความคิดเห็น

ลำดับที่	#ของกินเชียงใหม่ #รีวิวเชียงใหม่	วันที่แสดง ความคิดเห็น
1	วันก่อนไปสุกี้ข้างเผือกหลังมอ ✨ อร่อยหอมกระเพาะเหมือนกัน เพิ่มน้ำจิ้ม เพิ่มวุ้นเส้นเท่านั้น 😊 หนึบ ๆ ชอบมาก สาขานี้จะมีลวกจิ้มด้วย คนเยอะแต่รอคิวไม่นานนั่งสบาย ๆ 👍 🏠 : สาขาหลังมอ // #รีวิวเชียงใหม่ #ของกินเชียงใหม่	14/07/2565
2	วิถีลาบ เจริญเมือง คือสัมผัสครบรสมากแซบ ทุกอย่างอร่อยหมด ต้องมา 😊 #รีวิวเชียงใหม่ #เชียงใหม่ #อร่อยบอกต่อ #ของกินเชียงใหม่	11/11/2565
3	Sammai ราเมง ย่านนิมมาน ชูที่กลมกล่อมเข้มข้น เป็นราเมงอีกร้านที่ต้องกลับไปซ้ำ 😊 #รีวิวเชียงใหม่ #เชียงใหม่ #ราเมงเชียงใหม่ #ของกินเชียงใหม่	02/01/2566
4	ก๋วยเตี๋ยวหมูเขาไฟอุดมผล The best ใส่วก อร่อยที่สุดในชีวิตตตที่เคยกินได้มา ไม่ควาไม่เหนียว นุ่มแบบจะละลายแล้วอะ ไปกินก็ครั้งก็อร่อยเหมือนเดิม คนเยอะเป็นปกติ ก๋วยเตี๋ยวก็อร่อย แซบ เผ็ดถึงใจ รักร้านนี้ #reviewchiangmai #รีวิวเชียงใหม่ #อร่อยบอกต่อ #ของกินเชียงใหม่	18/03/2566
5	เป็นบัวลอยที่อร่อยมาก 😊👍 เจี๊ซัง อยู่ตรงข้างเผือก #รีวิวเชียงใหม่ ชอบแบบมาก ต้นปีมากิน รอบนี้ละติดใจกลับมาซ้ำ เป็นบัวลอยเม็ดเล็กแต่หนึบมากใส่เครื่องเพิ่มได้ตามใจเลยกินกับน้ำกะทิอุ่น ๆ คือฟินนอนหลับสบาย ใครยังไม่เคยลอง พลาดละนะ! #อร่อย #ของกินเชียงใหม่	31/03/2566

เนื่องจากข้อมูลที่ได้อยู่ในรูปแบบข้อความบนสื่อสังคมออนไลน์ และไม่ได้อยู่ในรูปแบบภาษาธรรมชาติ จึงต้องแปลงข้อความให้อยู่ในรูปแบบที่ภาษาคอมพิวเตอร์สามารถเรียนรู้ได้



ภาพที่ 4 แสดงขั้นตอนการดำเนินงานวิจัยและการออกแบบจำลอง

4. การวิเคราะห์ข้อมูล

4.1 ขั้นตอนที่หนึ่ง การเก็บรวบรวมข้อมูลการแสดงความคิดเห็นของผู้ใช้งาน

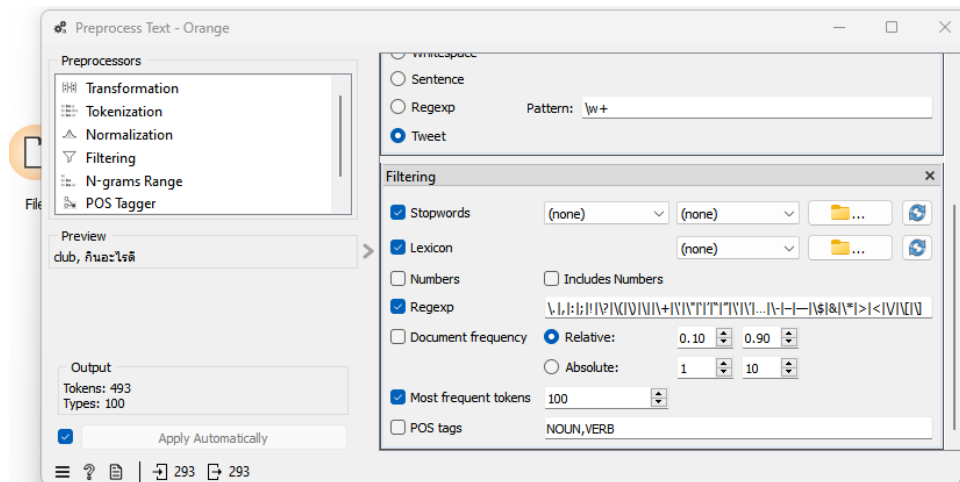
เก็บรวบรวมข้อมูลการแสดงความคิดเห็นของผู้ใช้งานจากทวีตเตอร์สำหรับแฮชแท็กของกิน เชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ บันทึกเข้าสู่โปรแกรม Microsoft Excel จะได้ไฟล์ข้อมูลในรูปแบบของไฟล์ XLSX (Microsoft Excel Open XML Spreadsheet) ในไฟล์มีจำนวนข้อความทั้งสิ้น 2,213 ความคิดเห็น



ภาพที่ 5 ตัวอย่างข้อความความคิดเห็น

4.2 ขั้นตอนที่สอง การจัดเตรียมข้อมูล (Text preprocessing)

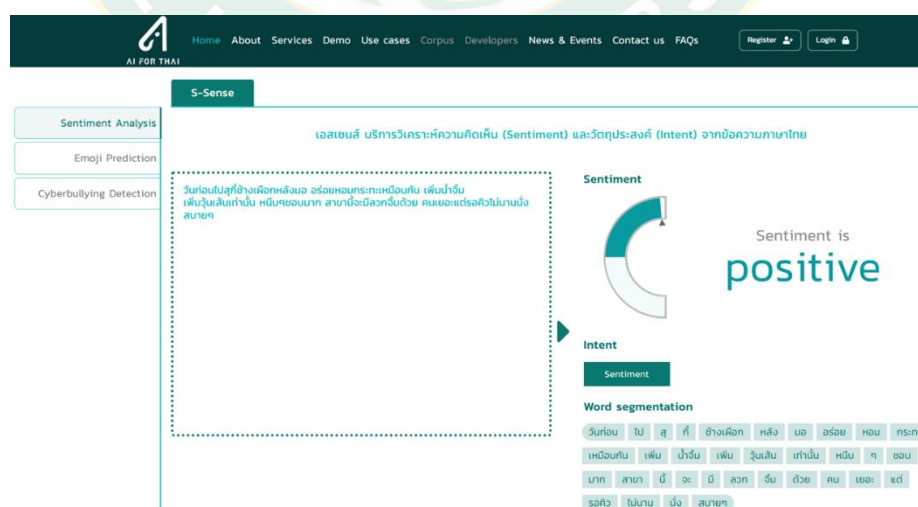
การจัดเตรียมข้อมูล ทำความสะอาดข้อมูล ได้ตัดประโยคที่ไม่เกี่ยวข้องกับข้อความออกจากข้อความที่เกิดจากการดึงข้อมูลและตัดตัวอักษรพิเศษ เช่น URL, Emoji, Hashtag เป็นต้น หลังจากผ่านกระบวนการนี้จะเหลือข้อความความคิดเห็นจำนวน 293 ความคิดเห็น จากนั้นทำการตัดคำ (Tokenization) โดยใช้โปรแกรมเล็กซ์โต (Lexto) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ เนื่องจากภาษาไทยนั้นไม่มีการใช้ตัวอักษรหรือสัญลักษณ์ที่นำมาใช้คั่นระหว่าง



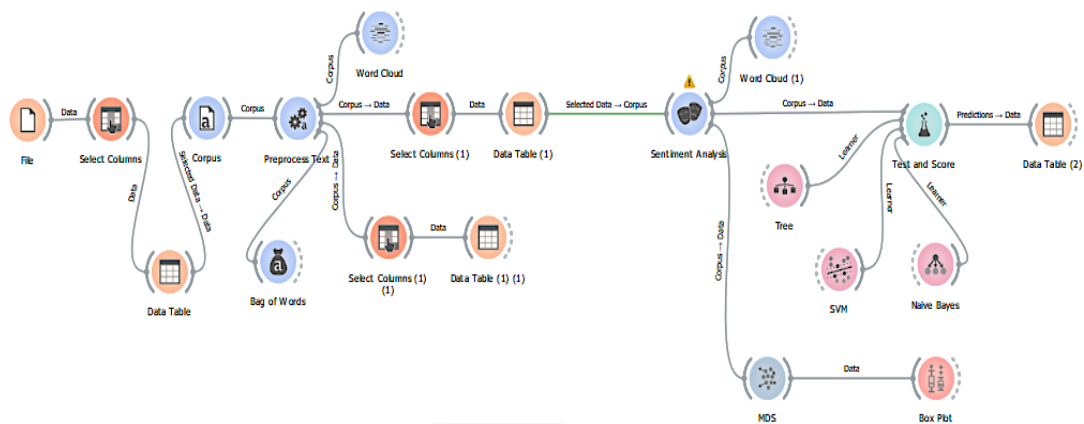
ภาพที่ 8 การสกัดคุณลักษณะของข้อความโดยโปรแกรม Orange Data mining

4.3 ขั้นตอนที่สาม การวิเคราะห์ความรู้สึก (Sentiment Analysis)

ใช้โปรแกรม S-Sense ทำการวิเคราะห์ความคิดเห็นบนสื่อโซเชียลมีเดียด้วยเทคนิคการประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP และการเรียนรู้ Machine Learning ดังภาพที่ 10 จากนั้นใช้การจำแนก 3 อัลกอริทึม ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree เพื่อเปรียบเทียบประสิทธิภาพการจำแนกหมวดหมู่วิเคราะห์คำเพื่อระบุผลความรู้สึกของความคิดเห็นได้ว่าคำนั้นอยู่ในความหมายความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral)



ภาพที่ 9 ตัวอย่างการใช้โปรแกรม S-Sense วิเคราะห์ความรู้สึก (Sentiment Analysis)

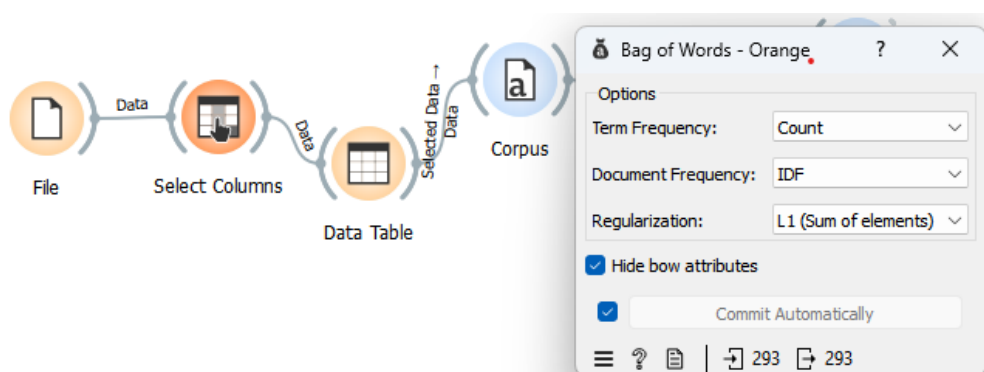


ภาพที่ 10 การสร้างแบบจำลองโมเดลโดยใช้อัลกอริทึมการจำแนก 3 อัลกอริทึม ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree โดยโปรแกรม Orange Data mining

4.4 ขั้นตอนทีสี่ การจำลองหัวข้อ (Topic Modeling)

การสกัดใจความสำคัญของข้อความ เพื่อหาความถี่ของคำที่มีลักษณะเฉพาะในแต่ละความคิดเห็น ช่วยในการเลือกคำที่มีนัยสำคัญในความคิดเห็น จากนั้นเมื่อเข้าสู่กระบวนการต่อ จะได้ถ่วงคำ (Bag of Word) เพื่ออธิบายกลุ่มรวมของคำโดยไม่ได้คำนึงถึงหลักไวยากรณ์ คำนามถี่ที่พบ และลำดับของคำมาใช้เป็น Feature ในการตัดจัดแบ่งข้อความข้อความ Classifier เมื่อได้คลังคำศัพท์ที่เกี่ยวข้องกับความคิดเห็นสำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ นำไปสร้างดัชนีคำสำคัญ (Indexing) ให้อยู่ในรูปแบบของเวกเตอร์ค่าน้ำหนักของคำ

ในขั้นตอนนี้โปรแกรม Orange Data mining มาช่วยในการสร้างคลังคำศัพท์ของข้อมูลในรูปแบบภาษาไทย



ภาพที่ 11 การคำนวณหาค่าน้ำหนักของคำ (IDF) โดยโปรแกรม Orange Data mining

บทที่ 4

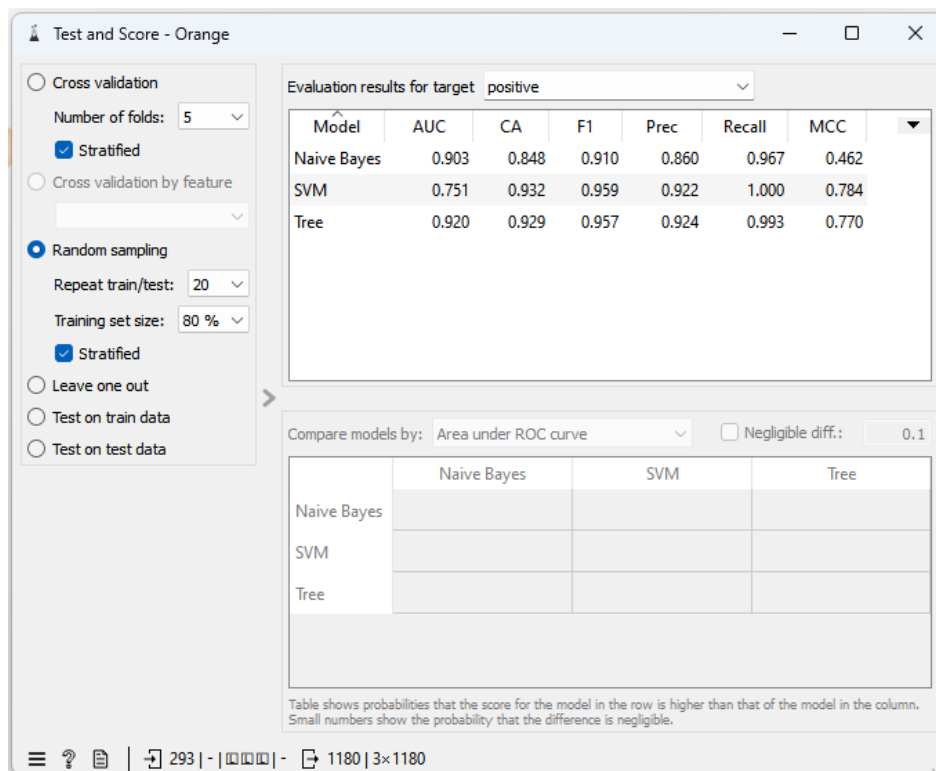
ผลการวิจัย

งานวิจัยนี้ได้ทำการวิเคราะห์ความคิดเห็นจากทวีตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และแฮชแท็กรีวิวเชียงใหม่ โดยใช้แนวคิดการทำเหมืองข้อมูล (Data Mining) การประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP มาใช้ในการวิเคราะห์ ใช้อัลกอริทึมการจำแนก 3 อัลกอริทึม ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree โดยโปรแกรม Orange Data mining โดยข้อมูลที่ใช้ในงานวิจัยนี้เก็บรวบรวมข้อมูลจากความคิดเห็นในโซเชียลมีเดียแพลตฟอร์มทวีตเตอร์ ซึ่งเป็นข้อมูลการแสดงความคิดเห็นจากผู้บริโภคในรูปแบบใส่แฮชแท็ก (Hashtag) ในช่วงวันที่ 1 พฤษภาคม 2565 ถึง วันที่ 30 เมษายน 2566 จำนวนข้อความ 2,213 ความคิดเห็น หลังจากผ่านกระบวนการต่าง ๆ แล้วจะเหลือข้อความความคิดเห็นที่เตรียมเข้าสู่โมเดลทั้ง 3 อัลกอริทึม จำนวน 293 ความคิดเห็น สามารถวิเคราะห์อารมณ์หรือความรู้สึก ได้แก่ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral)

1. ผลการศึกษาการวิเคราะห์ความรู้สึก (Sentiment Analysis)

การสร้างแบบจำลอง (Evaluation Model) เพื่อหาแบบจำลองที่เหมาะสมที่สุดในการวิเคราะห์ความคิดเห็นจากทวีตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่ที่เป็นภาษาไทย ด้วย 3 อัลกอริทึม ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree โดยใช้โปรแกรม Orange Data mining ในการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่า Accuracy, Precision, Recall และ F-measure (F1)

1.1 ความรู้สึกเชิงบวก (Positive)



ภาพที่ 13 แบบจำลองความรู้สึกเชิงบวก (Positive)

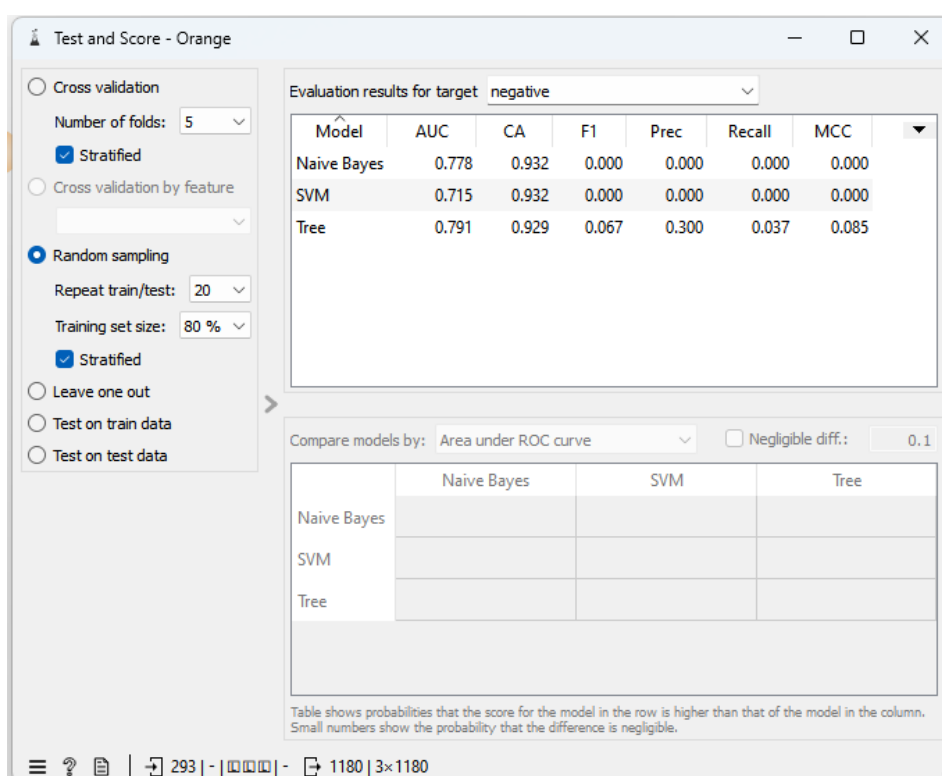
จากการวิเคราะห์ความคิดเห็นความรู้สึกเชิงบวก (Positive) โดยใช้อัลกอริทึม 3 แบบ ได้แก่ อัลกอริทึม Support Vector Machines (SVM) อัลกอริทึม Naive Bayes และอัลกอริทึม Decision Tree

อันดับที่ 1 การใช้อัลกอริทึมประเภท Decision Tree คะแนนความถูกต้อง (Accuracy) 0.920 ค่าการทำนาย (Precision) 0.924 ค่าความน่าจะเป็น (Recall) 0.993 และค่า F-measure 0.957 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความโดยใช้ อัลกอริทึมประเภท Decision Tree ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 2 การใช้อัลกอริทึมประเภท Naive Bayes คะแนนความถูกต้อง (Accuracy) 0.903 ค่าการทำนาย (Precision) 0.860 ค่าความน่าจะเป็น (Recall) 0.967 และค่า F-measure 0.910 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความโดยใช้อัลกอริทึมประเภท Naive Bayes ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 3 การใช้อัลกอริทึมประเภท Support Vector Machines (SVM) คะแนนความถูกต้อง (Accuracy) 0.751 ค่าการทำนาย (Precision) 0.922 ค่าความน่าจะเป็น (Recall) 1.000 และค่า F-measure 0.959 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความด้วยอัลกอริทึมประเภท SVM ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

1.2 ความรู้สึกเชิงลบ (Negative)



ภาพที่ 14 แบบจำลองความรู้สึกเชิงลบ (Negative)

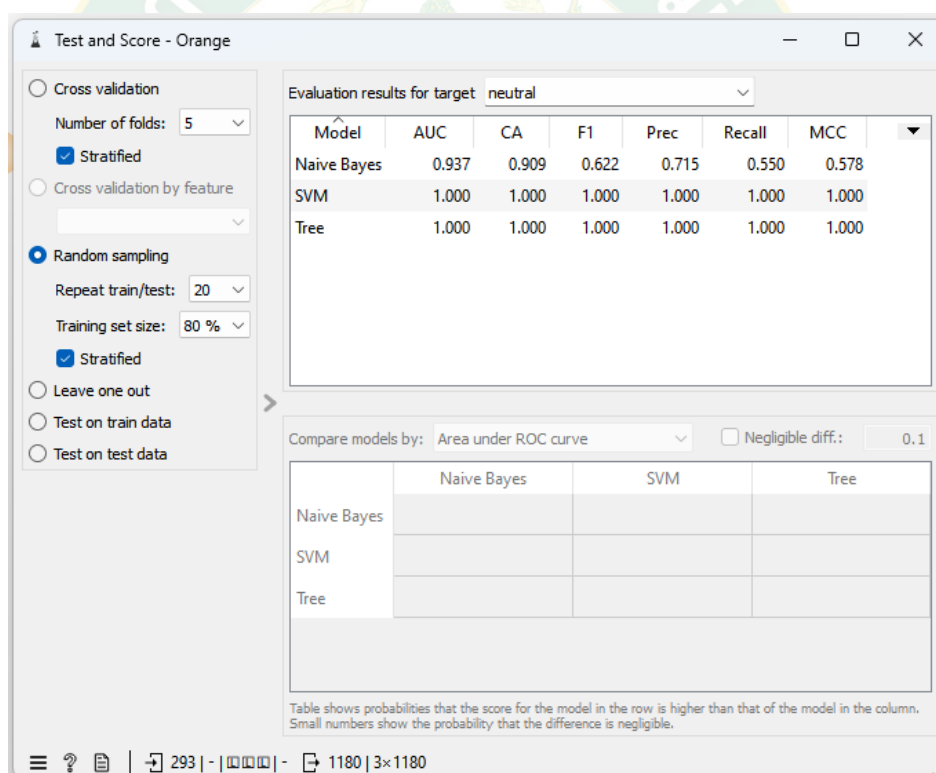
จากการวิเคราะห์ความคิดเห็นความรู้สึกเชิงลบ (Negative) โดยใช้อัลกอริทึม 3 แบบ ได้แก่ อัลกอริทึม Support Vector Machines (SVM) อัลกอริทึม Naive Bayes และอัลกอริทึม Decision Tree

อันดับที่ 1 การใช้อัลกอริทึมประเภท Decision Tree คะแนนความถูกต้อง (Accuracy) 0.791 ค่าการทำนาย (Precision) 0.300 ค่าความน่าจะเป็น (Recall) 0.037 และค่า F-measure 0.067 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความด้วยอัลกอริทึมประเภท Decision Tree ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 2 การใช้อัลกอริทึมประเภท Naive Bayes คะแนนความถูกต้อง (Accuracy) 0.778 ค่าการทำนาย (Precision) 0.000 ค่าความน่าจะเป็น (Recall) 0.000 และค่า F-measure 0.000 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความด้วยอัลกอริทึมประเภท Naive Bayes ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 3 การใช้อัลกอริทึมประเภท Support Vector Machines (SVM) คะแนนความถูกต้อง (Accuracy) 0.715 ค่าการทำนาย (Precision) 0.000 ค่าความน่าจะเป็น (Recall) 0.000 และค่า F-measure 0.000 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความด้วยอัลกอริทึมประเภท SVM ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

1.3 ความรู้สึกเป็นกลาง (Neutral)



ภาพที่ 15 แบบจำลองความรู้สึกเป็นกลาง (Neutral)

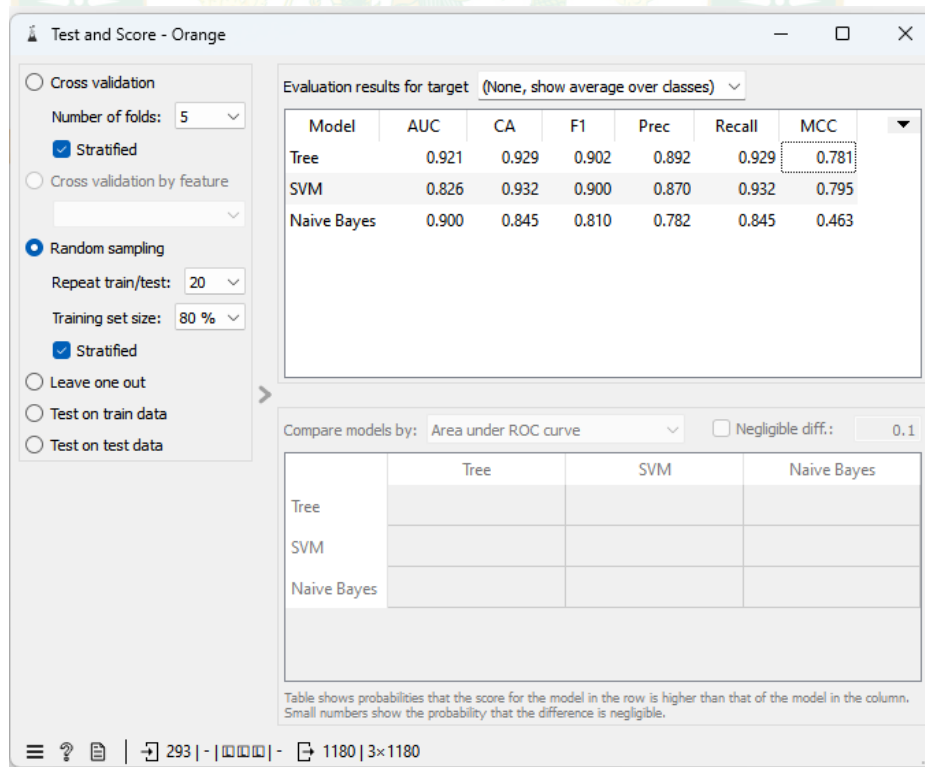
จากการวิเคราะห์ความคิดเห็นความรู้สึกเป็นกลาง (Neutral) โดยใช้อัลกอริทึม 3 แบบ ได้แก่ อัลกอริทึม Support Vector Machines (SVM) อัลกอริทึม Naive Bayes และอัลกอริทึม Decision Tree

อันดับที่ 1 การใช้อัลกอริทึมประเภท Decision Tree คะแนนความถูกต้อง (Accuracy) 1.000 ค่าการทำนาย (Precision) 1.000 ค่าความน่าจะเป็น (Recall) 1.000 และค่า F-measure 1.000 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อมูลโดยใช้อัลกอริทึมประเภท Decision Tree ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 2 การใช้อัลกอริทึมประเภท Support Vector Machines (SVM) คะแนนความถูกต้อง (Accuracy) 1.000 ค่าการทำนาย (Precision) 1.000 ค่าความน่าจะเป็น (Recall) 1.000 และค่า F-measure 1.000 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อมูลโดยใช้อัลกอริทึมประเภท SVM ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 3 การใช้อัลกอริทึมประเภท Naive Bayes คะแนนความถูกต้อง (Accuracy) 0.937 ค่าการทำนาย (Precision) 0.715 ค่าความน่าจะเป็น (Recall) 0.550 และค่า F-measure 0.622 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อมูลโดยใช้อัลกอริทึมประเภท Naive Bayes ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

1.4 ค่าเฉลี่ยความรู้สึกรวม



ภาพที่ 16 แบบจำลองค่าเฉลี่ยความรู้สึกรวม (Total)

จากการวิเคราะห์ความคิดเห็นค่าเฉลี่ยความรู้สึกรวม (Total) โดยใช้อัลกอริทึม 3 แบบ ได้แก่ อัลกอริทึม Support Vector Machines (SVM) อัลกอริทึม Naive Bayes และอัลกอริทึม Decision Tree

อันดับที่ 1 การใช้อัลกอริทึมประเภท Decision Tree คะแนนความถูกต้อง (Accuracy) 0.921 ค่าการทำนาย (Precision) 0.892 ค่าความน่าจะเป็น (Recall) 0.929 และค่า F-measure 0.902 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความโดยใช้อัลกอริทึมประเภท Decision Tree ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 2 การใช้อัลกอริทึมประเภท Naive Bayes คะแนนความถูกต้อง (Accuracy) 0.900 ค่าการทำนาย (Precision) 0.782 ค่าความน่าจะเป็น (Recall) 0.845 และค่า F-measure 0.810 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความโดยใช้อัลกอริทึมประเภท Naive Bayes ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

อันดับที่ 3 การใช้อัลกอริทึมประเภท Support Vector Machines (SVM) คะแนนความถูกต้อง (Accuracy) 0.826 ค่าการทำนาย (Precision) 0.870 ค่าความน่าจะเป็น (Recall) 0.932 และค่า F-measure 0.900 ในการทดลองนี้ได้จำลองการทำงานของอัลกอริทึมแบบจำลองวิเคราะห์ข้อความโดยใช้อัลกอริทึมประเภท SVM ซึ่งมีการใช้เทคนิคนี้ในงานวิจัยตามที่ระบุไว้

เมื่อพิจารณาการเปรียบเทียบค่าความถูกต้อง (Accuracy) ค่าการทำนาย (Precision) ค่าความน่าจะเป็น (Recall) และค่า F-measure (F1) ในการจำลองการทำงานของอัลกอริทึมทั้ง 3 อัลกอริทึม พบว่าอัลกอริทึมประเภท Decision Tree จำแนกข้อมูลโดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูล ทั้งความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) หรือความรู้สึกเป็นกลาง (Neutral) มีค่าความถูกต้อง (Accuracy), ค่าการทำนาย (Precision), ค่าความน่าจะเป็น (Recall) และค่า F-measure (F1) โดยรวมของทุกอัลกอริทึมแม่นยำกว่าตัวต้นแบบจำลองแบบอื่นที่ใช้ในการทดลองครั้งนี้ ดังนั้น ผลลัพธ์แบบจำลองการวิเคราะห์ความคิดเห็นจากทวีตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่ ด้วยอัลกอริทึมประเภท Decision Tree จึงเป็นขั้นตอนวิธีที่เหมาะสมที่สามารถนำมาประยุกต์ใช้งานได้อย่างมีประสิทธิภาพมากที่สุด ใน 3 อัลกอริทึมที่วิเคราะห์ หรืออีกนัยหนึ่งอาจกล่าวได้ว่าวิธีนี้สามารถแยกความรู้สึกในชุดข้อมูลภาษาไทยได้แม่นยำที่สุด

2. ผลการศึกษาการสร้างแบบจำลองหัวข้อ (Topic Modeling)

ตารางที่ 2 ตัวอย่างผลลัพธ์การจำแนกความรู้สึกเชิงบวก (Positive)

Topics	Text	Sentiment	คะแนน	ลำดับ
ของกิน	ซาบูหมู่อืด สาขาดถนนรอบเมืองเชียงใหม่ ภาพรวมคือดีเลยบริการดีนั่งเสิร์ฟไวมาก ดูแลดั่งมิตรสหาย รสชาติโอเค ราคาดีมากคนละ 219 บาทราคานี้รวมเครื่องดื่มและของหวานแล้วน้ำ เครื่องเคียงก็อร่อยโดยเฉพาะเห็ดเข็มทองทอดนี้อร่อยมากเติมไป 4 รอบ สายซาบูไปเถอะคุ้มค่า	Positive	99.99	1
ของกิน	สุกี้ข้างเผือก ของดีตลาดข้างเผือก เด็ด ๆ ยิ่งเป็นสุกี้แห้ง บอกเลยว่ามันหอมกระเทียมร้อน ๆ หมูนุ่ม ๆ กับผักกรอบ ๆ กำลังดี ยังไม่ทันจะจิ้มน้ำจิ้มนี้ว่าอร่อยแล้ว น้ำจิ้ม คือนัว กลมกล่อมกำลังดี	Positive	98.84	2
ของกิน	ก๋วยเตี๋ยวหมูเขาไฟอุตมผล ก๋วยเตี๋ยวควายวที่รสชาติเข้มข้นสุด ๆ ในเชียงใหม่ ต้มยำ น้ำตกคือดีมาก ใครที่เป็นสายแข่งอยากให้ลองก๋วยเตี๋ยวกูเขาไฟ มาแบบรสจัด หลากรสในซามเดียว ที่สุดคือใส่อ่อนลวกต้องสั่งกันทุกโต๊ะ นิ่มนวลหอมละมุนกำลังดี	Positive	97.56	3
ของกิน	ได้ถุนบ้าน เป็นร้านอาหารเช้าที่น่ารักมาก อร่อยด้วย ฮือ ดีไปหมด	Positive	96.00	4
สถานที่	อ่างแก้วในเวลาที่ฝนหยุดตก มองเห็นหมอกลอยตรงภูเขา ภูญาสี่เขียว ๆ อากาศเย็นแบบสดชื่นมาก ดีต่อใจที่สุด เป็นอีกที่ในเชียงใหม่ที่ชอบมาก	Positive	85.71	5

ตารางที่ 3 ตัวอย่างผลลัพธ์การจำแนกความรู้สึกเชิงลบ (Negative)

Topics	Text	Sentiment	คะแนน	ลำดับ
ของกิน	ร้านนี้กินมาตั้งแต่สมัยเรียนมัธยม ตั้งแต่ร้านเค้าอยู่ในบ้าน ในซอยเล็ก ๆ จนตอนนี้ย้ายร้านออกมาริมถนน คนเยอะนะคะ ที่จอดรถน้อยต้องจอดริมถนน และข้างกำแพงวัด ราคาแพงหน่อยแต่อร่อยนะ ปล.มีทั้งของคาวของหวานเลยจ้า	Negative	66.67	1
ของกิน	ได้ถุนบ้าน อาหารเช้า รถใหญ่หาที่จอดยาก โต๊ะน้อย ถ้าไปสายรอคิวนั่งนาน ไข่กะทะรอไม่นาน รอใจก็นานกว่า สั่งชยาหอม ๆ มีชาร้อนเสิร์ฟ	Negative	75.00	2
สถานที่	ริวิว ดอยอินทนนท์ หน้าฝน สวยงามไปอีกแบบ หมอกเยอะมาก ถ่ายรูปดี๊ดี แต่ต้องระวังทางขึ้น จะมองไม่ค่อยเห็นและค่อนข้างอันตรายสำหรับคนไม่ชินทาง	Negative	75.00	3
ของกิน	ใครว่าเชียงใหม่ทุกอย่างถูก มาดู 555555 ทดลองสั่งหมูปิ้ง 2 ร้าน จากแกริบ ช่าย 8 บาท ขวา 13 บาท ความต่างอะมึงดูละกัน ร้าน 8 บาท หมูพรีเมียมหรวองงู บางไม้แหม่งเล็กกว่าที่ถ่ายอีก	Negative	80.00	4
ของกิน	Building A Cafe คาเฟ่เปิดใหม่ในตัวเมืองเชียงใหม่ แต่หาไม่เจอ เพราะร้านอยู่ลึกลับมาก ทางเข้าที่จอดรถก็คือ นี่เข้าไปได้จริงหรือ แต่พอเข้าไปถึงคือข้างในลานจอดรถมีพื้นที่กว้างมากแหม่จอดรถเสร็จก็หาทางเข้าร้านไม่เจออีก จะลึกลับแอบซ่อนไปไหน	Negative	88.89	5

บทที่ 5

สรุปและอภิปรายผล

งานวิจัยนี้ศึกษาเพื่อวิเคราะห์ความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และรีวิวเชียงใหม่ ในชุดข้อมูลภาษาไทย โดยมีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึก (Sentiment Analysis) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ และเพื่อสร้างแบบจำลองหัวข้อ (Topic Modeling) เกี่ยวกับการแสดงความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ ใช้ชุดข้อมูลในช่วงวันที่ 1 พฤษภาคม 2565 ถึง วันที่ 30 เมษายน 2566 ซึ่งจำนวนของความคิดเห็นที่ได้เก็บรวบรวมนั้นมีจำนวนทั้งสิ้น 2,213 ความคิดเห็น หลังจากผ่านกระบวนการต่าง ๆ จะเหลือข้อความความคิดเห็นจำนวน 293 ความคิดเห็น โดยใช้แนวคิดการทำเหมืองข้อมูล (Data Mining) การประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP มาใช้ในการวิเคราะห์ ได้แก่ Support Vector Machines (SVM), Naive Bayes และ Decision Tree และทำการวัดประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) ค่าการทำนาย (Precision) ค่าความน่าจะเป็น (Recall) และค่า F-measure (F1) โดยแบ่งหัวข้อในการสรุปผลดังนี้

1. สรุปและอภิปรายผล
2. ข้อเสนอแนะ

1. สรุปและอภิปรายผล

จากการวิเคราะห์ข้อมูล

จากผลการวิจัย จะเห็นได้ว่าสามารถวิเคราะห์ระดับความรู้สึกทางอารมณ์ได้ 3 ระดับ ได้แก่ ความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) และความรู้สึกเป็นกลาง (Neutral) ที่นำแบบจำแนกประเภท Decision Tree มาเป็นส่วนหนึ่งในการวิเคราะห์ การวิเคราะห์ความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่และรีวิวเชียงใหม่ ซึ่งสอดคล้องกับการศึกษาของ สันติสุขเกษม (2561) ที่ศึกษาความคิดเห็นของข้อมูลโรงแรม โดยใช้เทคนิคการตัดคำแบบผสมผสาน โดยใช้ Decision Tree ในการประเมินประสิทธิภาพ ตัวแบบจำแนกประเภทมีค่าความถูกต้อง (Accuracy) สูงที่สุด ร้อยละ 92.10 ค่าการทำนาย (Precision) ร้อยละ 89.20 ค่าความน่าจะเป็น (Recall) ร้อยละ 92.90 และ Naive Bayes คะแนนความถูกต้อง (Accuracy) ร้อยละ 90.00 ค่าการทำนาย (Precision) ร้อยละ 78.20 ค่าความน่าจะเป็น (Recall) ร้อยละ 84.50 และ Support Vector Machines (SVM) คะแนนความถูกต้อง (Accuracy) ร้อยละ 82.60 ค่าการทำนาย

(Precision) ร้อยละ 87.00 ค่าความน่าจะเป็น (Recall) ร้อยละ 93.20 ซึ่งการวัดประสิทธิภาพของข้อมูลแต่ละครั้งต้องเลือกค่าสูงที่สุด ดังนั้นถ้าเลือกค่าน้อยกว่าจนไม่เหมาะสมกับข้อมูล จะทำให้ได้ผลลัพธ์ที่ไม่สามารถวิเคราะห์ข้อมูลหรือนำไปใช้ประโยชน์ได้ นอกจากนี้ในแบบจำลองจะมีข้อดีข้อเสีย และข้อจำกัด อัลกอริทึมที่แตกต่างกัน งานวิจัยนี้ทำการวิเคราะห์ความรู้สึกทางอารมณ์และจำแนกประเด็นข้อมูลเพื่อพัฒนาผลิตภัณฑ์และการให้บริการของธุรกิจอาหารและเครื่องดื่มในจังหวัดเชียงใหม่ต่อไป

งานวิจัยนี้เป็นการศึกษาวิธีการวิเคราะห์หาค่าที่มีนัยสำคัญและหัวข้อในข้อความ และทำการจัดกลุ่มข้อมูลที่มีลักษณะคล้ายคลึงกันอยู่ในกลุ่มเดียวกันจากผู้ใช้งานโซเชียลมีเดียแพลตฟอร์มทวิตเตอร์ในรูปแบบแฮชแท็ก (Hashtag) ชื่อว่า “ของกินเชียงใหม่” และ “รีวิวเชียงใหม่” โดยใช้โปรแกรมเล็กซ์โต (Lexto) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ซึ่งเหมาะสมกับการใช้ตัดคำในข้อความที่เป็นภาษาไทยจากความคิดเห็นของผู้ใช้งานในโซเชียลมีเดีย จากนั้นใช้โปรแกรมเอสเซนส์ (S-Sense : Social Sensing) เพื่อวิเคราะห์ความคิดเห็นบนโซเชียลมีเดียด้วยเทคนิคการประมวลผลภาษาธรรมชาติ Natural Language Processing : NLP ในการวิเคราะห์ข้อมูลภาษาไทยจำนวนมาก ด้วยโปรแกรม S-Sense เป็นที่นิยมใช้ในการวิเคราะห์ Sentiment Analysis สอดคล้องกับการศึกษาของศุภภาวีร์ มากดี (2563) กับ Haruechaiyasak et al. (2018) ได้ศึกษาการใช้โปรแกรม S-Sense ในการวิเคราะห์อารมณ์และความรู้สึกจากข้อความโดยใช้ข้อมูลจากเว็บเครือข่ายสังคมออนไลน์ สามารถจำแนกรู้สึก 2 ระดับ คือความคิดเห็นเชิงบวก (Positive) และความคิดเห็นเชิงลบ (Negative)

หลังจากใช้โปรแกรมเอสเซนส์ (S-Sense : Social Sensing) จะใช้โปรแกรม Orange Data mining มาช่วยในการสกัดคุณลักษณะของความคิดเห็นออกมา เพื่อสร้างคลังคำศัพท์ จะได้ถ่วงคำ (Bag of Word) เพื่ออธิบายกลุ่มรวมของคำโดยไม่ได้คำนึงถึงหลักไวยากรณ์ คำนามวลีที่พบ และลำดับของคำมาใช้เป็น Feature ในการตัดจัดแบ่งข้อความ Classifier นำไปสร้างดัชนีคำสำคัญ (Indexing) ให้อยู่ในรูปแบบของเวกเตอร์ค่าน้ำหนักของคำ การสร้างแบบจำลองงานวิจัยนี้มี 3 อัลกอริทึมที่นำมาวิเคราะห์ ผลลัพธ์อัลกอริทึมประเภท Decision Tree เป็นขั้นตอนวิธีที่เหมาะสมที่สามารถนำมาประยุกต์ใช้งานได้อย่างมีประสิทธิภาพมากที่สุด เนื่องจากค่าความถูกต้อง (Accuracy) ค่าการทำนาย (Precision) ค่าความน่าจะเป็น (Recall) และค่า F-measure มีค่ามากที่สุดของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูล ทั้งความรู้สึกเชิงบวก (Positive) ความรู้สึกเชิงลบ (Negative) และความรู้สึกเป็นกลาง (Neutral)

สรุปผลการจำแนกรู้สึกความคิดเห็นจากทวิตเตอร์สำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่จำนวน 293 ความคิดเห็น ได้แก่

ความรู้สึกเชิงบวก (Positive) จำนวน 232 ความคิดเห็น ความรู้สึกเชิงลบ (Negative) จำนวน 22 ความคิดเห็น และความรู้สึกเป็นกลาง (Neutral) จำนวน 39 ความคิดเห็น

ผลลัพธ์ของงานวิจัยครั้งนี้แสดงให้เห็นถึงการแสดงความคิดเห็นไปในเชิงบวกมากกว่าความคิดเห็นเชิงลบ สอดคล้องกับการศึกษาของพิศิษฐ์ บวรเลิศสุธี และวรภัทร ไพริเกรง (2562) การตัดสินใจซื้อสินค้าและบริการออนไลน์ ผู้บริโภคส่วนใหญ่จะทำการสืบค้นข้อมูลเกี่ยวกับสินค้าและบริการรวมทั้งพิจารณาจากการแสดงความคิดเห็น และบทวิจารณ์ จากผู้ที่เคยซื้อสินค้าไปแล้ว

เพื่อเป็นแนวทางในการตัดสินใจเลือกซื้อ ซึ่งถ้าหากมีการสร้างตัวแบบเพื่อแจ้งให้ผู้ใช้งานทราบถึงการวิเคราะห์ความรู้สึกได้จะช่วยให้ผู้ซื้อตัดสินใจได้รวดเร็วมากยิ่งขึ้น โดยใช้เป็นข้อมูลในการสร้างตัวแบบ เพื่อช่วยให้ผู้บริโภคใช้ในการตัดสินใจซื้อสินค้าและบริการ และช่วยให้ผู้ประกอบการมีข้อมูลสำหรับการพัฒนาสินค้าและบริการต่อไปในอนาคต

นอกจากนี้การแสดงความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบสอดคล้องกับการศึกษาของ อิศรภาพ ล้อรัตน์ไชยงค์ (2560) ทำการศึกษาเรื่องการสร้างคลังศัพท์บอกความรู้สึกในภาษาไทยจากบทวิจารณ์ออนไลน์ อีกทั้งสอดคล้องกับการศึกษา วิสุตา เทศเมือง และนิเวศ จิระวิจิตชัย (2560) ทำการศึกษาเรื่องการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ และนำเสนอวิธีการจำแนกความคิดเห็นโดยการวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์

จากข้อมูลความคิดเห็นบนแพลตฟอร์มวิดีโอสำหรับแฮชแท็กของกินเชียงใหม่ และแฮชแท็กรีวิวเชียงใหม่ มีการนับจำนวนคำ ได้ผลลัพธ์ 10 อันดับแรกของคำที่มีการกล่าวถึงมากที่สุด ได้แก่ อร่อย รสชาติดี กาแฟ coffee café น่ารัก มุมถ่ายรูปเยอะ บรรยากาศดี อ่างแก้ว ใต้ถุนบ้าน เป็นต้น

สอดคล้องกับงานวิจัยของ Kasikornbank (2561) ที่พบว่าสื่อสังคมออนไลน์กำเนิดขึ้น การตลาดผ่านสื่อสังคมออนไลน์ หรือ Social Media Marketing กลายเป็นช่องทางที่เข้ามาแทนที่ และกำลังได้รับความนิยมสูงสุดในตอนนี้ ผู้ประกอบการสามารถใช้ในการเข้าถึงผู้บริโภคได้ ดังนั้นผู้ประกอบการสามารถเลือกคำจากการวิเคราะห์ที่มีการกล่าวถึงมากที่สุด เพื่อทำการตลาดแบบ Social Media Marketing สอดคล้องกับการศึกษาของ ธนิตา อัครวโยธิน (2561) ได้กล่าวว่า การตลาดออนไลน์ เป็นการนำเสนอผลิตภัณฑ์หรือบริการที่ใช้อินเทอร์เน็ตเป็นสื่อกลางที่ช่วยในการโฆษณา ประชาสัมพันธ์และใช้เป็นช่องทางในการติดต่อสื่อสารกันระหว่างเจ้าของธุรกิจกับลูกค้าเป็นการลงทุนทำการตลาดที่มีราคาไม่สูง สามารถแก้ไขหรืออัปเดตข้อมูลได้เสมอไม่ว่าจะเป็นการโพสต์หรือแชร์ข้อมูลข่าวสารที่ธุรกิจต้องการให้ลูกค้าทราบก็สามารถทำได้ง่าย สะดวก และรวดเร็ว

2. ข้อเสนอแนะ

2.1 ข้อเสนอแนะจากผลการวิจัย (ข้อเสนอแนะเชิงปฏิบัติ)

งานวิจัยนี้ ศึกษาเพียงหากระบวนการวิเคราะห์ความคิดเห็นของผู้ใช้งานแพลตฟอร์มทวิตเตอร์ในฐานะผู้บริโภคเท่านั้น เพื่อให้ผู้ประกอบการ ร้านค้า กิจกรรมต่าง ๆ นำไปปรับใช้กับธุรกิจได้ ซึ่งปัจจุบัน Social Media ทุกคนสามารถเข้าถึงได้ ทำให้สังเกตเห็นว่าการมีคำหรือวลีที่คนใช้ค้นหาสิ่งที่ต้องการ (keyword) จะมีส่วนช่วยให้ผู้บริโภคค้นหาได้ง่าย ร้านค้าสามารถใช้สำหรับโปรโมททางการตลาด และสามารถสร้างฐานลูกค้าใหม่ได้มากขึ้น

2.2 ข้อเสนอแนะสำหรับการวิจัยในครั้งต่อไป

1) งานวิจัยนี้ ในขั้นตอนการเตรียมข้อมูลไม่ได้แก้คำผิด เมื่อเข้ากระบวนการ การตัดคำ คำเหล่านั้นอาจถูกตัดไป ดังนั้นหากแก้ไขคำผิด จะสามารถช่วยเพิ่มประสิทธิภาพในการวิเคราะห์ และอาจช่วยหาค่าที่มีนัยสำคัญได้มากขึ้น

2) งานวิจัยนี้ เป็นการวิเคราะห์จาก Social Media แพลตฟอร์มทวิตเตอร์ เพียงแหล่งเดียว เนื่องจากประเทศไทยยังมีแพลตฟอร์มอีกมากมาย ที่สามารถนำมาวิเคราะห์ความคิดเห็นของผู้บริโภคต่อธุรกิจที่มีความเกี่ยวข้องทางการตลาด ซึ่งอาจจะได้ผลการวิจัยที่แตกต่างจากการวิจัยครั้งนี้

3) งานวิจัยนี้ มีข้อจำกัด ในกระบวนการการวิเคราะห์ความคิดเห็น เนื่องด้วยโปรแกรมที่ใช้ในการวิจัยครั้งนี้เป็นโปรแกรมที่ใช้งานได้อย่างจำกัด จึงไม่สามารถแปลผลได้ครบสมบูรณ์ทั้งหมดจากข้อมูลที่ทำกรรวบรวมได้ ดังนั้นถ้าใช้โปรแกรมที่สามารถใช้งานได้โดยไม่จำกัด ก็อาจจะทำให้ผลที่ได้มีความแม่นยำมากยิ่งขึ้น

บรรณานุกรม

- กานดา แผ้ววัฒนากุล และปราโมทย์ ลีอนาม. (2556). การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์. *วารสารการจัดการสมัยใหม่*, 11(2), 11-20.
- กานต์ ศุภจารุกิตต์. (2563). *Online Survey: คนไทยพร้อมไหมกับการทำแบบสำรวจ*. สถาบันนโยบายสาธารณะและการพัฒนา. <https://ippd.or.th/online-survey/>
- ณภัคอร ปุณยภาภัสสร. (2553). *AIDA Model*. <https://punyapapassorn.blogspot.com/2010/08/aida-model.html>
- ธนกร ญาณกาย และวนิดา แก่นอากาศ. (2562). การเพิ่มประสิทธิภาพแบบจำลองหัวข้อด้วยสภาพแวดล้อมแบบข้อมูลขนาดใหญ่. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 21(3), 166-174.
- นันทศักดิ์ สุทธิเลิศ. (2560). *การวิเคราะห์ความรู้สึกผู้ใช้บริการสายการบินของบริษัทในประเทศไทย สหรัฐอเมริกา* [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยบูรพา]. ชลบุรี.
- ปิยนตร อรุณปรีดี. (2565). *การวิเคราะห์ผลสะท้อนกลับด้านการท่องเที่ยวจากบทวิจารณ์ออนไลน์ กรณีศึกษาจังหวัดภูเก็ต* [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยสงขลานครินทร์]. สงขลา.
- ปิยวรรณ ทองพลอย. (2563). *การวิเคราะห์ข้อความภาษาไทยเกี่ยวกับการตั้งครรภ์ด้วยวิธีการสร้างแบบจำลองหัวข้อ (Topic Modelling)* [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ]. กรุงเทพฯ.
- พนิดา กาศกลางดอน. (2563). *การวิเคราะห์ประสบการณ์จากการใช้บริการโรงพยาบาลในประเทศไทย จากความคิดเห็นของผู้ใช้บริการ* [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ]. กรุงเทพฯ.
- พัฒนพันธ์ กลิ่นหอม. (2564). *การวิเคราะห์มุมมองของผู้บริโภคผ่านการแสดงความคิดเห็น (ข้อความเมนต์) ของผู้บริโภคที่มีต่อสินค้าสมาร์ตโฟนบนโซเชียลมีเดีย* [วิทยานิพนธ์บริหารธุรกิจมหาบัณฑิต, มหาวิทยาลัยนเรศวร]. พิษณุโลก.
- พิศิษฐ์ บวรเลิศสุธี และวรภัทร ไพรีเกรง. (2562). การสร้างตัวแบบการวิเคราะห์ความรู้สึกทางอารมณ์สำหรับบทความแนะนำสินค้าออนไลน์ โดยการตรวจสอบความคิดเห็นเชิงลบ งานประชุมวิชาการระดับชาติ ครั้งที่ 14 *Digital Technology for Sustainable Wellbeing and Smart Society*, กรุงเทพฯ.

- วิสุตา เทศเมือง และนิเวศ จิระวิชิตชัย. (2560). การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์ โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทซิง. *วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยสยาม*, 18(1), 1-11.
- วันสุวรรณ มีประเสริฐ และเอกรัฐ รัชฎาญจน์. (2564). การวิเคราะห์ความคิดเห็นจากทวิตเตอร์ของลูกค้าบริษัทข้อป้ประเทศไทย. *วารสารระบบสารสนเทศด้านธุรกิจ (JISB)*, 7(3), 6-18.
- วรวิวัฒน์ จันทร์ศิริวัฒนา. (2557). อิทธิพลของบทวิจารณ์ออนไลน์ที่มีต่อความตั้งใจในการเลือกใช้บริการโรงแรมและที่พักของผู้บริโภค [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์]. กรุงเทพฯ.
- วสวัสดิ์ อินทร์แปลง. (2563). การวิเคราะห์ความคิดเห็นต่อเกมมือถือผับจีด้วยเหมืองข้อความ [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยมหาสารคาม]. มหาสารคาม.
- ศุภภาวีร์ มากดี. (2563). การวิเคราะห์ข้อมูลขนาดใหญ่สำหรับการท่องเที่ยวกรณีศึกษาการท่องเที่ยวในจังหวัดอุบลราชธานี. *วารสารเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏอุบลราชธานี*, 10(2), 139 - 152.
- สมัคร ชัยสงวน. (2561). การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษานนทบุรี โดยใช้ตัวจำแนกข้อมูล นาอ็ฟ เบย์ สำหรับภาษาไทย [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยสุโขทัยธรรมาธิราช]. นนทบุรี.
- สหชัย งามชัยภูมิ. (2563). กระบวนการแบบผสมผสานเพื่อการจำแนกความรู้สึกแบบข้อความสั้น [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยมหาสารคาม]. มหาสารคาม.
- สันติ สุขเกษม. (2561). การวิเคราะห์รูปแบบความคิดเห็นของข้อมูลโรงแรม โดยใช้เทคนิคการตัดคำแบบผสมผสาน [วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สถาบันเทคโนโลยีไทย-ญี่ปุ่น]. กรุงเทพฯ.
- สุภางค์ จันทวานิช. (2547). *วิธีการวิจัยเชิงคุณภาพ*. (พิมพ์ครั้งที่ 12). สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อิสรภาพ ล้อรัตนไชยยงค์. (2560). การสร้างคลังศัพท์บอกความรู้สึกในภาษาไทยจากบทวิจารณ์ออนไลน์ [วิทยานิพนธ์อักษรศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย]. กรุงเทพฯ.
- เอ็ก ดิจิทัล. (2564). *User-generated Content คืออะไร ทำไมจึงช่วยแบรนด์ทำการตลาดได้แบบสุดปัง?* egg digital. <https://www.eggdigital.com/what-is-user-generated-content/>
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). *Sentiment Analysis of Twitter Data Language in Social Media (LSM 2011)*, Oregon, Portland.

- Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2018). S-Sense: A sentiment analysis framework for social media monitoring applications. *Information Technology Journal*, 14(1), 11-22.
- Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., & Tang, J. Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10), e10894. <https://doi.org/10.1016/j.heliyon.2022.e10894>
- Munsaen, S., Chuchuen, C., & Sukphan, J. (2023). *Sentiment Analysis on Social Networking: A Case Study of Chiang Mai Tourism* the 7th TICC International Conference 2023 Toward Sustainable Development Goals: Digital Transformation and Beyond, Chiang Mai.
- Satauri, I., Satouri, B., & El Beqqali, O. (2023). Big Data Analysis in Commercial Social Networks: Analysis of Twitter Reviews for Marketing Decision Making. *European Journal of Information Technologies and Computer Science*, 3(2), 1-6. <https://doi.org/10.24018/compute.2023.3.2.94>
- Saura, J. R., Palos-Sanchez, P., & Grilo, A. (2019). Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining. *Sustainability*, 11(3), 917.
- Tesmuang, R., & Chirawichitchai, N. (2020). Sentiment Analysis of Thai Online Product Reviews using Genetic Algorithms with Support Vector Machine. *Progress in Applied Science and Technology*, 10(2), 7-13.

ประวัติผู้วิจัย

ชื่อ-สกุล

กฤติกา ตี๋พเสน

เกิดเมื่อ

22 มีนาคม 2541

ประวัติการศึกษา

พ.ศ.2563

คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ

สาขาวิชาการจัดการ-การจัดการทั่วไป

มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น

