ตัวแบบทางสถิติเพื่อการวิเคราะห์
ข้อความการแสดงความเห็นบนเว็บไซต์ต่อมหาวิทยาลัย

STATISTICAL MODEL FOR
WEB-BASED UNIVERSITY REVIEWS ANALYSIS

วุฒิพงษ์ ตุ้นยุทธ์
WUTTIPONG TUNYUT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2568
KMITL-2025-SC-M-050-007

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

STATISTICAL MODEL FOR WEB-BASED UNIVERSITY REVIEWS ANALYSIS

WUTTIPONG TUNYUT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS

DEPARTMENT OF STATISTICS SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2025

KMITL-2025-SC-M-050-007

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2025
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABAN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์ ตัวแบบทางสถิติเพื่อการวิเคราะห์ข้อความการแสดงความเห็นบนเว็บไซต์ต่อมหาวิทยาลัย
ชื่อนักศึกษา วุฒิพงษ์ ตุ้นยุทธ์
รหัสประจำตัว 64605105
ปริญญา วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ภาควิชา สถิติ
พ.ศ. 2568
อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.พรรณทิพา วาณิชย์จิรัฐติกาล

บทคัดย่อ

การศึกษานี้มุ่งวิเคราะห์ประเด็นที่ปรากฏในข้อความแสดงความคิดเห็นเกี่ยวกับมหาวิทยาลัย และประเมินความสัมพันธ์ระหว่างประเด็นดังกล่าวกับความรู้สึกที่ผู้คนมีต่อมหาวิทยาลัย งานวิจัยก่อนหน้ามักมุ่งเน้นการวิเคราะห์ความคิดเห็นในระบบการเรียนการสอนออนไลน์และหลักสูตรการศึกษา ขณะที่ความคิดเห็นเกี่ยวกับมหาวิทยาลัยยังได้รับการศึกษาอย่างจำกัด

การศึกษานี้ประกอบด้วยสามขั้นตอนหลัก ขั้นตอนแรก คือ การรวบรวมข้อมูลจากเว็บไซต์ด้านการศึกษาที่เปิดให้แสดงความคิดเห็น ขั้นตอนที่สอง คือ การวิเคราะห์ข้อความโดยใช้เทคนิค Google NLP เพื่อประเมินคะแนนความรู้สึก และการวิเคราะห์หัวข้อโดยอาศัยการจัดสรรของดีรีเคลแฝงร่วมกับ Word2Vec และการวิเคราะห์วลีแบบคู่คำ จากการวิเคราะห์สามารถจัดกลุ่มหัวข้อหลักได้ 9 กลุ่ม ได้แก่ (1) ประสบการณ์โดยรวมและการเรียน (2) สถานที่และชีวิตในมหาวิทยาลัย (3) ความท้าทายทางวิชาการและภาระงาน (4) การใช้ชีวิตและกิจกรรม (5) รูปแบบการสอนและบุคลากร (6) ค่าใช้จ่ายและต้นทุนการใช้ชีวิต (7) ทรัพยากรของมหาวิทยาลัย (8) ชื่อเสียงและความเป็นนานาชาติ และ (9) ประสบการณ์ด้านภาษาและวัฒนธรรม

ขั้นตอนสุดท้าย คือ การวิเคราะห์ด้วยการถดถอยบีตาแสดงให้เห็นว่า ประเด็นหลักที่ส่งผลเชิงบวกต่อความรู้สึกคือ ประสบการณ์โดยรวมและการเรียน การสอนและบุคลากร และการใช้ชีวิตและกิจกรรม ในทางกลับกัน หากมหาวิทยาลัยสามารถจัดการด้านกิจกรรมนักศึกษา การสอนและบุคลากร ความเป็นนานาชาติ และประสบการณ์ด้านภาษาและวัฒนธรรมได้ดี ความรู้สึกเชิงลบในข้อความก็จะไม่รุนแรงนัก นอกจากนี้ยังพบว่าประเด็นด้านสถานที่และชีวิตในมหาวิทยาลัย ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ไม่มีนัยสำคัญต่อคะแนนความรู้สึกหลังปี ค.ศ. 2019 ซึ่งอาจเป็นผลจากสถานการณ์โรคระบาด

ผลการเปรียบเทียบความคิดเห็นต่อมหาวิทยาลัยในสหราชอาณาจักร สหรัฐอเมริกา ออสเตรเลีย แคนาดา จีน และญี่ปุ่น แสดงให้เห็นว่าประเด็นที่ส่งผลต่อความรู้สึกของผู้ใช้แตกต่างกันในแต่ละประเทศ สะท้อนถึงลักษณะและจุดเน้นของระบบการศึกษาที่แตกต่างกันในแต่ละประเทศ

คำสำคัญ : **การประมวลภาษาธรรมชาติ การวิเคราะห์ข้อความ ข้อความแสดงความคิดเห็น ตัวแบบหัวข้อ การถดถอยบีตา อุดมศึกษา**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Statistical Model for Web-Based University Reviews Analysis
Student Name	Wuttipong Tunyut
Student ID	64605105
Degree	Master of Science (Applied Statistics)
Department	Statistics
Year	2024
Thesis Advisor	Asst. Prof. Dr. Puntipa Wanitjirattikal

Abstract

This study aims to analyze key themes appearing in university reviews and to assess their relationships with users' sentiments toward universities. While previous research has predominantly focused on reviews related to online learning systems and educational curricula, general university reviews have received limited scholarly attention.

The study consists of three main stages. The first stage involves collecting review data from educational websites that allow users to share their opinions about universities. The second stage includes sentiment analysis using Google Natural Language Processing (NLP), followed by topic modeling based on Latent Dirichlet Allocation (LDA), enhanced with Word2Vec and bigram phrase analysis. From this process, nine major topic groups are identified: (1) overall experience and learning, (2) location and campus life, (3) academic challenges and workload, (4) student life and activities, (5) teaching style and staff, (6) expenses and cost of living, (7) university resources, (8) reputation and international outlook, and (9) language and cultural experience.

The final stage applies beta regression analysis, which reveals that the most positively associated themes with user sentiment are overall experience and learning, teaching and staff, and student life and activities. Conversely, when universities effectively manage student engagement, teaching quality, internationalization, and cultural experiences, negative sentiments tend to be less intense. Additionally, themes related to location and campus life as well as cost of living show no significant relationship with sentiment scores after 2019, which may be attributed to the effects of the COVID-19 pandemic.

A cross-country comparison of university reviews from the United Kingdom, United States, Australia, Canada, China, and Japan indicates that the factors influencing

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

user sentiment differ across countries, reflecting the diverse characteristics and focal points of each nation's education system.

Keywords : **Natural Language Processing, Text Analysis, Reviews, Topic Modeling, Beta Regression Equation, Higher Education**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยความกรุณาช่วยเหลือและคุณความรู้ของท่านอาจารย์ ผศ.ดร.พรรณทิพา วาณิชย์จิรัฐติกาล ผู้เป็นที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอขอบพระคุณท่านอาจารย์เป็นอย่างสูง ผศ.ดร.สมศรี บัณฑิตวิไล รศ.ดร.อัชฌา อระวีพร และ รศ.ดร.วราฤทธิ์ พานิชกิจโกศลกุล ผู้ที่เป็นทั้งกรรมการวิทยานิพนธ์ เป็นผู้ให้ความรู้ทั้งจากการสอนและผลงานวิชาการของท่านทั้งสอง ซึ่งผู้วิจัยติดตามและให้ความเคารพยิ่ง

ผู้วิจัยขอขอบพระคุณ คณาจารย์จากสาขาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง คณาจารย์จากคณะเศรษฐศาสตร์ มหาวิทยาลัยสงขลานครินทร์ คณาจารย์จากคณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ท่านอาจารย์ ดร. วิโรจน์ ณ ระนอง และ คณะเพื่อนร่วมงานจากทั้งจาก บริษัท ทีโอที จำกัด (มหาชน) และสถาบันวิจัยเพื่อการพัฒนาประเทศไทย ซึ่งมีส่วนร่วมสำคัญในการเติบโตทางความคิดของผู้วิจัย

วุฒิพงษ์ ตุ้นยุทธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อ	ก
Abstract	ข
กิตติกรรมประกาศ	ง
สารบัญ	จ
สารบัญตาราง	ช
สารบัญรูป	ซ
คำย่อ/สัญลักษณ์	ฌ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	3
1.3 ขอบเขตของงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 การวิเคราะห์เชิงปริมาณข้อมูลข้อความ	4
2.1.1 ภาพรวมการวิเคราะห์ข้อความ และ การประมวลภาษาธรรมชาติ	4
2.1.2 ขั้นตอนการประมวลข้อมูลข้อความ โดยการแปลงประโยคเป็นถุงคำ	7
2.1.3 ขั้นตอนการประมวลข้อมูลข้อความ โดยการแปลงคำเป็นชุดรหัสตัวเลข	10
2.2 ตัวแบบทางสถิติเพื่อการวิเคราะห์ข้อมูลข้อความ	12
2.2.1 การจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA)	12
2.2.2 การวิเคราะห์แบ่งกลุ่ม (Cluster Analysis)	17
2.2.3 ค่าความคล้ายคลึงโคไซน์ และระยะห่างโคไซน์	18
2.3 ตัวแบบทางสถิติเพื่อการวิเคราะห์ข้อมูลร้อยละ	19
2.3.1 การถดถอยบีตา (Beta Regression)	19
2.4 งานวิจัยที่เกี่ยวข้อง	21
บทที่ 3 วิธีการดำเนินงานวิจัย	28
3.1 ข้อมูลที่ใช้ในการศึกษา	28
3.1.1 ข้อมูลข้อความการแสดงความเห็นต่อมหาวิทยาลัย	28
3.1.2 ข้อมูลตัวแปรคุณลักษณะของมหาวิทยาลัย	31
3.2 ขั้นตอนการศึกษา	33
3.2.1 การรวบรวมข้อมูลและภาพรวม	33
3.2.2 ขั้นตอนการวิเคราะห์ข้อมูลข้อความ	34
3.2.3 ขั้นตอนการวิเคราะห์เชิงสถิติ	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

บทที่ 4 ผลการวิจัยและการอภิปรายผล 38
4.1 การวิเคราะห์ความรู้สึกที่มีต่อมหาวิทยาลัย 38
4.2 การวิเคราะห์ประเด็นที่มหาวิทยาลัยถูกกล่าวถึง 42
4.3 การแจกแจงของประเด็นที่มหาวิทยาลัยถูกกล่าวถึง 46
4.4 การวิเคราะห์ความสัมพันธ์ระหว่างความรู้สึกและประเด็นที่มหาวิทยาลัยถูกกล่าวถึง 52
4.4 การอภิปรายผลการศึกษา 76
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 78
5.1 สรุปผลการวิจัย 78
5.2 ข้อเสนอแนะ 81
เอกสารอ้างอิง 82
ภาคผนวก 88
ภาคผนวก ก การแจกแจงของตัวแปรคะแนนความรู้สึกของข้อความ 89
ประวัติผู้วิจัย 91

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
2.1	ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นตัวแปรหุ่น	8
2.2	ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นค่าความถี่ของคำ	9
2.3	ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นค่าความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร	9
2.4	ตัวอย่างสรุปการทบทวนวรรณกรรมที่มีความน่าสนใจ	23
3.1	สรุปแหล่งข้อมูลข้อความเห็นต่อมหาวิทยาลัย	28
3.2	รูปแบบการแปลงข้อมูล HTML CODE เป็นตาราง	31
3.3	คำอธิบายตัวแปร	31
4.1	ข้อมูลคะแนนที่ผู้เขียนข้อความประเมินและความยาวของข้อความ	38
4.2	สถิติเชิงพรรณนาของคะแนนความรู้สึกต่อมหาวิทยาลัย	39
4.3	ค่าสัมประสิทธิสหสัมพันธ์ตัวแปรคะแนนความรู้สึกและคะแนนที่ผู้เขียนประเมิน	40
4.3	ตารางคำหลักของแต่ละหัวข้อจากการทำการจัดสรรของดีรีเคลแฝง	43
4.4	การแจกแจงของหัวข้อของข้อความเห็น จำแนกตามความรู้สึกเชิงลบเชิงบวก	47
4.5	การแจกแจงของหัวข้อของข้อความเห็น จำแนกตามความรู้สึกเชิงลบเชิงบวกและช่วงเวลา	48
4.6	การแจกแจงของหัวข้อของข้อความเห็น (เฉพาะเชิงบวก) จำแนกตามประเทศ	49
4.7	การแจกแจงของหัวข้อของข้อความเห็น (เฉพาะเชิงลบ) จำแนกตามประเทศ	50
4.8	ผลการทดสอบภาวะสารูปดี (GOODNESS OF FIT TEST) ของตัวแปรตาม	53
4.9	สถิติเชิงพรรณนาของข้อมูลที่ใช้ในการวิเคราะห์การถดถอยบีตา	54
4.10	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7	56
4.11	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7	57
4.12	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงลบ	58
4.13	ผลประมาณการการถดถอย กรณีก่อนและหลัง ค.ศ. 2019	62
4.14	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7 แยกรายประเทศ	65
4.15	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7 แยกรายประเทศ	66
4.16	ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงลบ แยกรายประเทศ	67
4.17	สรุปผลตัวแปรหัวข้อที่มีความสำคัญในการกำหนดความรู้สึกที่มีต่อมหาวิทยาลัย	68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	ขั้นตอนการแปลงข้อมูลข้อความเป็นข้อมูลเชิงปริมาณ	7
2.2	ความสัมพันธ์ของพารามิเตอร์ในตัวแบบ LDA	15
2.2	ค่าโคไซน์กับเวกเตอร์แสดงตำแหน่งของคำ	19
3.1	ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ WHAIUNI.COM	29
3.2	ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ UNIREVIEW.COM.AU	29
3.3	ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ IAGORA.COM	30
3.4	ตัวอย่างรูปแบบข้อมูล HTML CODE	30
3.5	ขั้นตอนการศึกษา	33
4.1	ค่าเฉลี่ยคะแนนความรู้สึกในแต่ละช่วงคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์	40
4.2	ค่าเฉลี่ยคะแนนความรู้สึกในแต่ละปี	41
4.3	ค่าคะแนนความเชื่อมโยงจากการทำการจัดสรรของดีรีเคลแฝง	43
4.4	แผนภาพเดนโดแกรมผลการแบ่งกลุ่มแบบมีขั้นตอน	45
4.5	การแจกแจงของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัย	53
ก.1	รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก	89
ก.2	รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก มีค่าน้อยกว่า 0.7	89
ก.3	รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก มีค่ามากกว่าหรือเท่ากับ 0.7	90
ก.4	รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงลบ	90

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำย่อ/สัญลักษณ์

คำย่อ	ความหมาย
BML	การถดถอยบีตาที่ประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด (Beta Regression with Maximum Likelihood Estimator)
CNN	โครงข่ายประสาทเทียมแบบสังวัตนาการ (Convolutional Neural Network)
CTM	ตัวแบบหัวข้อที่เกี่ยวข้อง Correlated Topic Model
LDA	การจัดสรรดีรีเคลแฝง (Latent Dirichlet Allocation)
ML	ภาวะน่าจะเป็นสูงสุด (Maximum Likelihood)
MSE	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error)
NLG	การสร้างภาษาธรรมชาติ (Natural Language Generating)
NLTK	ชุดคำสั่งเพื่อประมวลภาษาธรรมชาติ (Natural Language Toolkit)
NLP	การประมวลภาษาธรรมชาติ (Natural Language Processing)
NLU	การเข้าใจภาษาธรรมชาติ (Natural Language Understanding:)
OLS	กำลังสองน้อยที่สุด (Ordinary Least Squares)
PDF	ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function)
TF-IDF	ความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร (Term Frequency - Inverse Document Frequency)
TPBR	วิธีประมาณพารามิเตอร์การถดถอยบีตาแบบพารามิเตอร์ 2 ตัว (Two-Parameter Estimator for Beta Regression)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยบนสื่อออนไลน์เป็นข้อมูลขนาดใหญ่ที่มีค่ามาก เพราะเป็นข้อมูลที่บอกประสบการณ์ ความเห็น รวมถึงประเด็นที่ผู้ให้ข้อมูลคิดว่าแต่ละมหาวิทยาลัยต้องพัฒนา ข้อมูลประเภทนี้สามารถนำมาใช้สร้างประโยชน์ได้ทั้งในระดับบุคคลเช่น ผู้เรียน ผู้สอน เจ้าหน้าที่ ในระดับองค์กรที่มหาวิทยาลัยจะต้องพัฒนาตนเองเพื่อตอบสนองต่อภาวะการแข่งขันที่รุนแรงเพิ่มมากขึ้นตลอดเวลาจากภาวะประชากรวัยเรียนที่มีแนวโน้มลดต่ำลงเรื่อย ๆ (มหาวิทยาลัยธรรมศาสตร์, 2565) และในระดับประเทศที่มหาวิทยาลัยถือเป็นหน่วยผลิตทุนมนุษย์ที่สำคัญและบ่อยครั้งก็เป็นหน่วยที่ได้รับการสนับสนุนจากเงินภาษีของประเทศ

อย่างไรก็ตาม การใช้ประโยชน์จากข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยบนสื่อออนไลน์มีความท้าทายกว่าข้อมูลอื่น ๆ ด้วยเหตุผลหลัก 3 ประการ ประการแรก คือ ข้อมูลข้อความเป็นข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) การประมวลผลข้อมูลจึงจำเป็นต้องใช้เทคนิคทางวิทยาการคอมพิวเตอร์เข้าร่วม ประการที่สอง คือ ข้อมูลมักจะกระจายอยู่ในแหล่งต่าง ๆ ที่หน่วยงานมหาวิทยาลัยไม่ได้เป็นเจ้าของ ประการสุดท้าย คือ ข้อมูลข้อความมักจะไม่มีป้ายกำกับ (No Label) ที่จะใช้ในการแปลงข้อมูลเชิงคุณภาพเป็นข้อมูลเชิงปริมาณ เช่น ข้อความในเว็บบอร์ดสาธารณะ ซึ่งส่งผลต่อความยากในการวิเคราะห์ข้อมูล

โดยปกติแล้ว มหาวิทยาลัยจะมีข้อมูลการประเมินที่อยู่ในรูปแบบที่พร้อมใช้อยู่แล้ว แต่ข้อมูลที่มีก็ไม่อาจจะนำมาใช้ทดแทนข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยได้ เช่น ข้อมูลการจัดอันดับมหาวิทยาลัยจากหน่วยงานที่ได้รับการยอมรับอย่าง Times Higher Education (THE) และ QS World University ซึ่งจะเผยแพร่ข้อมูลการประเมินในรูปคะแนนของดัชนีประกอบ (Composite Index) (Elsevier, 2021) ที่สามารถบอกได้ว่าแต่ละมหาวิทยาลัยมีความโดดเด่นทางด้านไหน เช่น ด้านการวิจัย ด้านความเป็นนานาชาติ แต่เมื่อพิจารณาเกณฑ์การประเมินจะพบว่าเป็นการประเมินที่ไม่ได้ให้รายละเอียดมากนัก เน้นการสำรวจความเห็นกับกลุ่มบุคคล และไม่ได้มีมุมมองเรื่องความพึงพอใจของผู้เรียนซึ่งเป็นผู้ที่ได้รับผลกระทบโดยตรงจากการเข้าศึกษา

ขณะที่ข้อมูลรวบรวมความเห็นที่มหาวิทยาลัยจัดทำเอง มักจะอาศัยการใช้แบบสอบถามที่ต่างองค์กรต่างทำ ไม่ได้มีการแบ่งปันข้อมูลเพื่อให้เกิดการเรียนรู้จากข้อมูลขนาดใหญ่ และมีต้นทุนในการดำเนินการสร้างข้อมูลใหม่ ด้วยข้อจำกัดของข้อมูลที่มีทำให้การแสวงหาแนวทางเพื่อใช้ประโยชน์จากข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยบนสื่อออนไลน์เป็นสิ่งที่มีความสำคัญ

ผู้วิจัยได้จัดทำการศึกษาครั้งนี้โดยมีวัตถุประสงค์หลัก คือ การพัฒนาและนำเสนอตัวแบบเพื่อประมวลข้อความการแสดงความคิดเห็นต่อมหาวิทยาลัยให้กลายเป็นข้อมูลที่พร้อมนำไปใช้ในการวิเคราะห์เชิงปริมาณได้ ศึกษาปัจจัยที่ส่งผลต่อความรู้สึกด้านบวกและด้านลบของผู้คนที่มีต่อมหาวิทยาลัย และเปรียบเทียบการให้คุณค่าต่อลักษณะของมหาวิทยาลัยในกลุ่มมหาวิทยาลัยที่แตกต่างกันได้ เช่น มหาวิทยาลัยที่อยู่ต่างประเทศกันที่มีรูปแบบการจัดการเรียนการสอนที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการทบทวนการศึกษาในอดีตพบงานจำนวนมากที่มีการวิเคราะห์ข้อความแสดงความคิดเห็น โดยส่วนมากจะเป็นงานที่ใช้ข้อความแสดงความคิดเห็นเกี่ยวกับสินค้าและบริการ เช่น ข้อความแสดงความคิดเห็นบนแพลตฟอร์มอีคอมเมิร์ซ (Biswas *et al.*, 2022; Ghose and Ipeirotis, 2011) อย่างไรก็ตาม ในด้านการวิเคราะห์ข้อความเกี่ยวกับการศึกษา ส่วนมากจะยังคงเป็นการพัฒนาตัวแบบเพื่อการจัดกลุ่มข้อความ หรือการจำแนกข้อความ โดยการใช้ข้อมูลแสดงความคิดเห็นที่ปรากฏบนแพลตฟอร์มการเรียนการสอนแบบออนไลน์ (Massive Open Online Course: MOOCs) (Loyalka *et al.*, 2019; Barrón Estrada *et al.*, 2020; ONAN, 2021) หรือข้อมูลจากแบบสอบถาม (Ali Kandhro *et al.*, 2019; Giang, Dien and Khoa, 2020)

ในกรณีการวิเคราะห์ข้อความที่นอกเหนือการจัดกลุ่มหรือจำแนกข้อความ พบการศึกษาจำนวนหนึ่งที่วิเคราะห์ประเด็นที่ปรากฏในประกาศรับสมัครงานว่ามีการกล่าวถึงหัวข้อทักษะใดบ้าง (Mezzanzanica, 2020) และการวิเคราะห์ความสอดคล้องกับหลักสูตรการศึกษา (Almaleh *et al.*, 2019; Spada *et al.*, 2022) การทบทวนเอกสารไม่พบการศึกษาที่ศึกษาความสัมพันธ์ระหว่างความรู้สึกของผู้เรียนกับคุณลักษณะของสถานศึกษา แต่พบการศึกษาที่ศึกษาความสัมพันธ์ระหว่างความรู้สึกของผู้เรียนกับคุณลักษณะของวิชาทีเรียน (Li *et al.*, 2022)

ในการศึกษานี้ผู้วิจัยได้ประยุกต์ใช้กรอบวิธีการศึกษาของ Huai and Van de Voorde (2022) เพื่อศึกษาหาคุณลักษณะที่ส่งผลต่อความรู้สึกต่อมหาวิทยาลัยทั้งในแง่บวกและแง่ลบ โดยปรับกระบวนการวิเคราะห์ข้อความจากการวิเคราะห์แบ่งกลุ่ม (Cluster Analysis) เป็นการสร้างตัวแบบหัวข้อ (Topic Modeling) เพื่อให้สามารถดำเนินการศึกษาได้ตามวัตถุประสงค์ที่กำหนดไว้ ผู้วิจัยเลือกใช้ข้อมูลข้อความเห็นต่อมหาวิทยาลัยในสหราชอาณาจักรจาก WhatUNI.com ข้อความเห็นต่อมหาวิทยาลัยในออสเตรเลียจาก UniReview.com.au ข้อความเห็นต่อมหาวิทยาลัยในหลายประเทศจาก Igora.com และข้อมูลคุณลักษณะของมหาวิทยาลัยจากเว็บไซต์ Timeshighereducation.com

การศึกษานี้จะแบ่งขั้นตอนการศึกษาออกเป็น 3 ส่วน คือ ขั้นตอนการรวบรวมข้อมูล ขั้นตอนการวิเคราะห์ข้อมูลข้อความและขั้นตอนการวิเคราะห์เชิงสถิติ โดยการรวบรวมข้อมูลใช้วิธีการขูดข้อมูลจากเว็บไซต์และจัดรูปข้อมูล HTML จากนั้นจึงเป็นขั้นตอนการวิเคราะห์ข้อมูลข้อความ เพื่อแปลงข้อมูลข้อความให้เป็นข้อมูลตัวเลข โดยเริ่มจากการใช้ตัวแบบ Google NLP ประเมินความรู้สึกของข้อความ ใช้ Natural Language Toolkit (NLTK) และอื่น ๆ เพื่อทำความสะอาดข้อมูลข้อความ ใช้ตัวแบบ Word2Vec เพื่อแปลงข้อความเป็นเวกเตอร์โดยคำนึงถึงหน้าที่และความหมายของคำร่วมกับตัวแบบการวิเคราะห์คำที่อยู่ติดกันเป็นคู่ (Bigram Language Model: Bigram) จากนั้นจึงใช้ตัวแบบการจัดสรรดีรีเคลแฝง (Latent Dirichlet Allocation: LDA) เพื่อระบุหัวข้อที่ข้อความกำลังกล่าวถึง จากนั้นจึงสร้างตัวแปรความถี่ของหัวข้อที่ถูกกล่าวถึง ขั้นตอนนี้ทำให้สามารถเก็บค่าสถิติได้ว่ามีปัจจัยอะไรบ้างที่ถูกกล่าวถึงบ่อยเมื่อผู้คนแสดงความเห็นต่อมหาวิทยาลัย

ในขั้นตอนต่อไปคือการวิเคราะห์เชิงสถิติ โดยจะเป็นการศึกษาว่าปัจจัยที่ศึกษาในขั้นตอนก่อนหน้าว่ามีปัจจัยใดบ้างที่ส่งผลอย่างมีนัยสำคัญทางสถิติต่อความรู้สึกในแง่บวกและแง่ลบที่ผู้คนมีต่อมหาวิทยาลัย ดำเนินการด้วยการประมาณการถดถอยบีตา เนื่องจากตัวแปรตาม คือ ค่าเฉลี่ยคะแนนความรู้สึกในแง่บวกและแง่ลบมีค่าในช่วง 0 ถึง 1 โดยมีตัวแปรต้น คือ ตัวแปรความถี่ของหัวข้อต่าง ๆ ที่ถูกกล่าวถึง และมีตัวแปรคุณลักษณะของมหาวิทยาลัยเป็นตัวแปรควบคุม และมีการวิเคราะห์การกระจายและพลวัตของข้อมูลเพื่อศึกษาการเปลี่ยนแปลงของประเด็นที่มหาวิทยาลัยถูกพูดถึงข้ามช่วงเวลา และความแตกต่างของประเด็นที่มีความสำคัญในแต่ละพื้นที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การศึกษานี้แบ่งเนื้อหาเป็น 5 ส่วน ส่วนแรกนี้เป็นการอธิบายที่มาและความสำคัญ วัตถุประสงค์การศึกษาและภาพรวมการศึกษา ส่วนที่ 2 เป็นการทบทวนวรรณกรรมที่เกี่ยวข้อง โดยเฉพาะการศึกษาที่ดำเนินการใกล้เคียงกันและทฤษฎีต่าง ๆ ที่ใช้ในการวิเคราะห์ ส่วนที่ 3 แสดงการออกแบบกระบวนการศึกษา ส่วนที่ 4 จะแสดงผลการศึกษาตั้งแต่ขั้นตอนของข้อมูล การสกัดสารสนเทศน์จากข้อมูล และการประมวลตัวแบบ และส่วนที่ 5 จะเป็นการสรุปผล อภิปรายผลและข้อเสนอในการพัฒนาการศึกษาต่อในอนาคต

1.2 วัตถุประสงค์ของงานวิจัย

1) พัฒนาและนำเสนอตัวแบบเพื่อประมวลข้อความการแสดงความคิดเห็นต่อมหาวิทยาลัยให้กลายเป็นข้อมูลที่พร้อมนำไปใช้ในการวิเคราะห์เชิงปริมาณได้
2) ระบุลักษณะของมหาวิทยาลัยที่มีความสำคัญต่อทัศนคติด้านบวกและด้านลบต่อมหาวิทยาลัย
3) เปรียบเทียบการให้คุณค่าต่อลักษณะของมหาวิทยาลัยในกลุ่มมหาวิทยาลัยที่แตกต่างกัน

1.3 ขอบเขตของงานวิจัย

1) กลุ่มตัวอย่าง: มหาวิทยาลัยในสหราชอาณาจักร สหรัฐอเมริกา ออสเตรเลีย แคนาดา จีน ญี่ปุ่นและไทย
2) ระยะเวลา: พ.ศ. 2558-2565 (ค.ศ. 2015-2022)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้ตัวแบบรวมถึงกรอบการดำเนินการในการประมวลข้อความการแสดงความคิดเห็นต่อมหาวิทยาลัยให้กลายเป็นข้อมูลที่พร้อมนำไปใช้ในการวิเคราะห์เชิงปริมาณได้
2) สามารถระบุลักษณะของมหาวิทยาลัยที่มีความสำคัญต่อทัศนคติด้านบวกและด้านลบต่อมหาวิทยาลัย เช่น ตำแหน่งที่ตั้ง สถานที่ ชื่อเสียงของสถาบัน อาจารย์ผู้สอน
3) สามารถเปรียบเทียบการให้คุณค่าต่อลักษณะของมหาวิทยาลัยในกลุ่มมหาวิทยาลัยที่แตกต่างกัน เช่น การให้ความสำคัญกับสถานที่ในช่วงก่อนและหลังการมีการระบาดของโควิด-19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การวิเคราะห์เชิงปริมาณข้อมูลข้อความ

2.1.1 ภาพรวมการวิเคราะห์ข้อความ และ การประมวลภาษาธรรมชาติ

การวิเคราะห์ข้อความ (Text Analysis) และการประมวลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นวิธีการในการประมวลข้อมูลภาษาที่มนุษย์ใช้สื่อสารด้วยคอมพิวเตอร์ ทำให้การประมวลข้อมูลภาษาสามารถทำได้อย่างมีประสิทธิภาพมากขึ้น โดยทั้ง 2 คำอาจจะถูกใช้แทนการได้ ขึ้นอยู่กับการกำหนดขอบเขตนิยามซึ่งแตกต่างกันในแต่ละการศึกษา แต่โดยทั่วไปแล้วการวิเคราะห์ข้อความจะมุ่งเน้นการค้นหาความรู้หรือสารสนเทศน์จากข้อมูลข้อความ ขณะที่การประมวลภาษาธรรมชาติจะมุ่งเน้นการทำให้คอมพิวเตอร์สามารถเข้าในภาษาที่มนุษย์ใช้สื่อสาร (Kao and Poteet, 2007)

แม้ว่าการศึกษาฉบับนี้จะมุ่งเน้นการวิเคราะห์ข้อความ แต่กระบวนการวิเคราะห์ส่วนหนึ่งจำเป็นที่จะต้องมีความเข้าใจกระบวนการประมวลภาษาธรรมชาติ เพราะข้อมูลข้อความมีลักษณะเฉพาะที่แตกต่างจากข้อมูลตัวเลข โดยเฉพาะการที่ข้อมูลข้อความต่าง ๆ อาจจะมีการพูดถึงประเด็นเดียวกัน แต่มีการใช้คำที่แตกต่างกัน เช่น การนับจำนวนข้อความที่มีการกล่าวถึงมหาวิทยาลัยเกี่ยวกับสถานที่ ข้อความอาจจะมีการใช้คำต่าง ๆ เช่น วิทยาเขต ศูนย์ อาคาร พื้นที่ ซึ่งในกรณีนี้จำเป็นต้องมีการใช้เทคนิคกระบวนการประมวลภาษาธรรมชาติเข้ามาช่วยในการจัดกลุ่มคำเพื่อให้สามารถวิเคราะห์ได้ว่าข้อความแต่ละประโยคบรรจุประเด็นใดไว้บ้าง

การประมวลภาษาธรรมชาติ คือ วิธีการประมวลข้อมูลภาษาที่มนุษย์ใช้สื่อสารให้สามารถดำเนินการผ่านระบบคอมพิวเตอร์ โดยมุ่งเน้นการสร้างให้เกิดระบบชุดข้อมูลที่เป็นตัวแทนภาษาที่มนุษย์ใช้สื่อสาร ซึ่งการประมวลข้างต้นสามารถเป็นทั้งการประมวลจากภาษาธรรมชาติเป็นภาษาคอมพิวเตอร์โดยการใส่รหัส (Encoding) และการประมวลจากภาษาคอมพิวเตอร์เป็นภาษาธรรมชาติโดยการถอดรหัส (Decoding) โดยการดำเนินการเหล่านี้จำเป็นที่จะต้องมีความรู้ด้านโครงสร้างภาษาธรรมชาติ เช่น การแบ่งหน้าที่ของคำ (Part of Speech) และ ไวยกรณ์ (Grammar) ซึ่งทำให้การประมวลภาษาธรรมชาติเป็นสหวิทยาการที่เป็นส่วนผสมระหว่าง ภาษาศาสตร์ (Linguistics) และ วิทยาศาสตร์คอมพิวเตอร์ (Computer science) (Khurana et al., 2022)

การประมวลภาษาธรรมชาติ สามารถแบ่งเป็น 2 ส่วน คือ

1) การเข้าใจภาษาธรรมชาติ (Natural Language Understanding: NLU) คือ กระบวนการทำให้คอมพิวเตอร์เข้าใจในภาษามนุษย์ ซึ่งจะไล่ระดับตามความลึกซึ้งของการสร้างความเข้าใจดังนี้
    - สัทวิทยา (Phonology) คือ การศึกษาระบบเสียงในภาษาพูด
    - สัณฐานวิทยา (Morphology) คือ การศึกษาเกี่ยวกับโครงสร้างของคำศัพท์เพื่อแยกรากศัพท์ ส่วนต่อเติมเพื่อแปลงความหมายหรือหน้าที่ของคำ เช่น ส่วนข้างหน้า (Prefix) และส่วนข้างหลัง (Suffix)
    - ระบบคำ (Lexical) คือ การศึกษาเกี่ยวกับระบบการให้ความหมายของคำและหน้าที่ของคำตามบริบท (Part of Speech) ซึ่งทำให้สามารถใช้คำอื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แทนคำเดิมและยังคงความหมายเดิมไว้ได้ รวมถึงการสกัดคำที่ไม่มีส่วนสำคัญในการให้ความหมายในประโยค เช่น คำหยุด (Stop Words) เช่น คำว่า the คำว่า and การตัดส่วนท้ายของคำ (Stemming) และการแปลงคำด้วยรายการคำในพจนานุกรม (Lemmatization) เพื่อลดจำนวนคำที่จะถูกนำมาเป็นตัวแปรในขั้นตอนการวิเคราะห์ต่อไป

- วากยสัมพันธ์ (Syntactic) เป็นการวิเคราะห์การเชื่อมโยงกันของคำตามหน้าที่ ซึ่งแตกต่างจากระบบคำที่ตัดคำหยุด ส่วนขยายคำและหน้าที่ของคำตามบริบทออกไป
- ความหมาย (Semantic) เป็นการศึกษาเกี่ยวกับการแปลงความหมายที่เหมาะสมของประโยค
- วาทกรรม (Discourse) คือการวิเคราะห์ความหมายของประโยคพร้อม ๆ กันหลายประโยค ซึ่งทำให้เข้าใจความหมายของสรรพนามที่ถูกกล่าวถึงในแต่ละประโยคว่ากำลังหมายถึงใครหรืออะไร
- วัจนปฏิบัติ (Pragmatics) คือการเข้าใจความหมายของประโยคที่แท้จริงตามความหมายของผู้พูด คือ กระบวนการนำความรู้นอกเหนือจากที่ปรากฏในข้อความ เช่น บริบทการพูด (Context) มาใช้ตีความร่วม เช่น กรณีการประชดที่ต้องพิจารณาจากบริบทการพูดถึงจะทราบได้ว่าความหมายที่ผู้พูดกำลังสื่อถึงคืออะไร

2) การสร้างภาษาธรรมชาติ (Natural Language Generating: NLG) คือ กระบวนการผลิตข้อมูลภาษาที่ใช้แสดงข้อมูลในระบบ

การประมวลภาษาธรรมชาติมีการพัฒนากระบวนการวิเคราะห์ต่าง ๆ มาอย่างยาวนาน โดยการศึกษาของ Anandarajan, Hill and Nolan (2019) และ Khurana *et al* (2022) ได้สรุปพัฒนาการเทคนิคต่าง ๆ เริ่มตั้งแต่ ยุค 1940s เริ่มมีการใช้การแปลงภาษาโดยเครื่อง โดยส่วนมากจะเป็นการแปลงภาษาอังกฤษกับภาษารัสเซีย ก่อนที่ช่วงต้นยุค 1950s-80s จะเริ่มมีการใช้ปัญญาประดิษฐ์ในการวิเคราะห์ข้อความ เช่น โครงการ BASEBALL Q-A Systems มีการพัฒนาภาษาศาสตร์โดยใช้คอมพิวเตอร์ (Computational Linguistics) และการพัฒนาระบบการวิเคราะห์ฐานข้อมูลขนาดใหญ่

ในยุค 1990s เป็นต้นมา การประมวลภาษาธรรมชาติ ให้ความสำคัญกับการสกัดสารสนเทศน์จากข้อมูลภาษาซึ่งทำให้เกิดการพัฒนาตัวแบบทางภาษา โดยในปี ค.ศ. 2000 ได้มีการนำเสนอตัวแบบ Neural Language Model (โดย Bengio, Ducharme and Vincent, 2000) เพื่อประเมินความเป็นไปได้ของคำต่อไปโดยการใช้ข้อมูลคำที่ปรากฏก่อนหน้า ต่อมาในปี ค.ศ. 2008 ได้มีการเสนอการใช้ Multitask Learning ในการประมวลภาษาธรรมชาติ (โดย Collobert and Weston, 2008) ซึ่งคือ การสร้างตัวแบบประมวลภาษาที่ทำหน้าที่มากกว่าหนึ่งหน้าที่เข้าด้วยกัน ในปี ค.ศ. 2013 ได้มีการเสนอให้มีการแปลงคำเป็นชุดรหัสตัวเลข (Word Embedding) (โดย Mikolov *et al.*, 2013) ซึ่งเป็นการแปลงคำเป็นเวกเตอร์ นำมาใช้แทนการแปลงคำเป็นตัวแปรหุ่น ซึ่งมีส่วนช่วยในการพัฒนาการศึกษาความสัมพันธ์ของคำ เช่น คำที่มีความหมายใกล้เคียงกันหรือมีหน้าที่ในประโยคเดียวกันจะมีเวกเตอร์ที่ใกล้กัน เพื่อประโยชน์ในการวิเคราะห์ข้อความในระดับ วากยสัมพันธ์ (Syntactic) ขึ้นไป ในปี ค.ศ. 2013 มีการนำโครงข่ายประสาทเทียม (Neural Network: NN) เข้ามา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใช้ในกระบวนการวิเคราะห์ ในปี ค.ศ. 2014 มีการนำเสนอ Sequence-to-Sequence Model (โดย Sutskever, Vinyals and Le, 2014) หรือตัวแบบที่แปลงข้อความเป็นเวกเตอร์ และแปลงเวกเตอร์เป็นข้อความอีกภาษา โดยการใช้ความน่าจะเป็นมาช่วย ใช้ประโยชน์ได้มากในการแปลภาษาหนึ่งไปอีกภาษาหนึ่งให้แม่นยำมากขึ้น และในปี ค.ศ. 2015 ได้มีการเสนอ Attention Mechanism (โดย Bahdanau, Cho and Bengio, 2015 และถูกอ้างอิงโดย Vaswani *et al.*, 2017)) ซึ่งช่วยเพิ่มความแม่นยำในการแปลภาษาด้วยการมีกลไกการกำหนดว่าคำในภาษาปลายทางควรถูกกำหนดจากคำใดบ้างในภาษาต้นทางและในค่าน้ำหนักเท่าไหร่

ขณะที่การวิเคราะห์ข้อความ (Text Analysis) ที่มุ่งเน้นการหาความรู้จากข้อมูลข้อความเองก็ถูกพัฒนาใช้ในงานต่าง ๆ (Anandarajan, Hill and Nolan, 2019) เช่น

- การวิเคราะห์ความหมายแฝง (Latent Semantic Analysis: LSA) คือ การวิเคราะห์ความสัมพันธ์ระหว่างชุดของเอกสารและคำศัพท์ เมื่อเราแปลงชุดเอกสารเป็นข้อมูลเมทริกซ์ความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร (Term Frequency - Inverse Document Frequency) สามารถใช้เทคนิค Singular Value Decomposition (SVD) เพื่อแปลงข้อมูลเป็นเวกเตอร์ และใช้การวัดระยะเวกเตอร์ เช่น ค่าความคล้ายคลึงโคไซน์ (Cosine similarity) เพื่อพิจารณาความเหมือนของเอกสาร
- การวิเคราะห์แบ่งกลุ่ม (Cluster Analysis) คือ เทคนิคการจำแนกกลุ่มชุดข้อมูลเมื่อข้อมูลที่นำมาใช้สร้างตัวแบบไม่มีข้อมูลระบุกลุ่ม ซึ่งมีเทคนิคสำคัญ คือ การแบ่งกลุ่มข้อมูลแบบเคมีน (K-means Clustering)
- ตัวแบบความน่าจะเป็นของหัวข้อ (Probabilistic Topic Models) ใช้สำหรับการประเมินคำสำคัญที่จะมากำหนดเป็นหัวข้อของแต่ละชุดข้อมูล ว่าข้อมูลแต่ละชุดควรใช้คำสำคัญกลุ่มไหนในความน่าจะเป็นเท่าใด ใช้เพื่อจัดกลุ่มข้อความที่คล้ายกันให้อยู่ในกลุ่มเดียวกัน
- การวิเคราะห์จำแนกกลุ่ม (Classification Analysis) เทคนิคการจำแนกกลุ่มชุดข้อมูลเมื่อข้อมูลที่นำมาใช้สร้างตัวแบบมีข้อมูลระบุกลุ่มซึ่งมีเทคนิคสำคัญ คือ โครงข่ายประสาทเทียม (Neural Network)
- การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นวิธีเพื่อระบุว่าข้อความมีโอกาสให้ความรู้สึกแง่บวกและแง่ลบในความน่าจะเป็นที่เท่าใด อาจจะใช้วิธีการประเมินจากระบบคำ (Lexical) หรือ การสร้างตัวแบบขึ้นมาใหม่จากการวิเคราะห์แบ่งกลุ่ม หรือ การวิเคราะห์จำแนกกลุ่ม

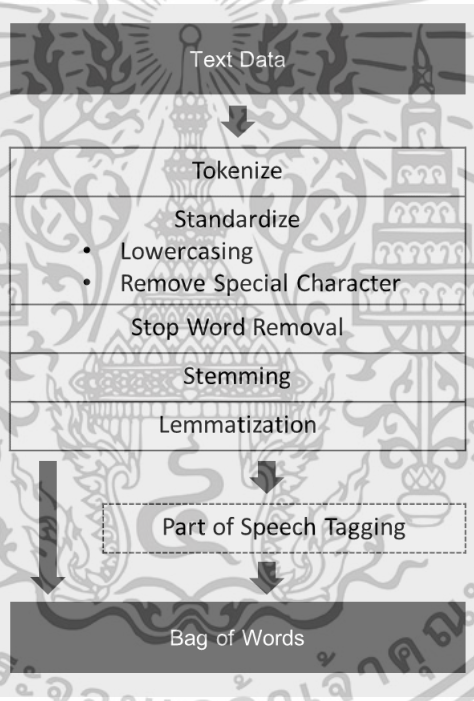
การใช้ประโยชน์จาก การวิเคราะห์ข้อความ (Text Analysis) และการประมวลภาษาธรรมชาติ (Natural Language Processing: NLP) ในการวิจัยเกิดขึ้นในหลากหลายรูปแบบ เช่น การสร้างระบบคำอ้างอิง (Co-Reference Resolution) การวิเคราะห์วาทกรรม (Discourse Analysis) การแปลข้อความ (Machine Translation) การวิเคราะห์ประเด็นสำคัญของเอกสาร และในกรณีของการศึกษานี้ คือ การวิเคราะห์ข้อความเห็น (Reviews Analysis) ซึ่งจะมีองค์ประกอบทั้ง การวิเคราะห์ความรู้สึก (Sentiment Analysis) การวิเคราะห์แบ่งกลุ่ม (Cluster Analysis) และตัวแบบระบบคำ (Lexical Model)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 ขั้นตอนการประมวลข้อมูลข้อความ โดยการแปลงประโยคเป็นถุงคำ (Bag of Words)

ข้อมูลข้อความเป็นข้อมูลประเภทไม่มีโครงสร้าง (Unstructured Data) ซึ่งเป็นข้อมูลที่ไม่ได้มีการนิยามโครงสร้างของข้อมูลไว้เหมือนข้อมูลตารางที่ระบุความหมายของแถว (มักจะนิยามแทนชุดตัวอย่าง) และคอลัมน์ (มักจะนิยามแทนตัวแปร) เพื่อให้สามารถทำการวิเคราะห์ต่อได้จึงจำเป็นที่จะต้องกำหนดโครงสร้างให้กับข้อมูลให้ข้อมูลกลายเป็นข้อมูลที่มีโครงสร้าง (Structured Data) โดยการใช้วิธีการแปลงข้อมูลเชิงคุณภาพให้เป็นข้อมูลเชิงปริมาณ

ในหัวข้อนี้จะเป็นการอธิบายถึงการแปลงข้อมูลข้อความให้เป็นข้อมูลที่มีโครงสร้างด้วยวิธีการพื้นฐานที่สุด คือ การแปลงเป็นถุงคำ (Bag of Words) และแปลงค่าเป็นเมทริกซ์ตัวแปรคำ ซึ่งเป็นวิธีที่ยังไม่คำนึงถึงหน้าที่ของคำ (Part of Speech) และ ความหมาย (Semantic) ของคำ โดยในเบื้องต้น ก่อนการแปลงเป็นถุงคำ (Bag of Words) เพื่อไม่ให้มีจำนวนตัวแปรมากเกินไป ผู้วิจัยจะต้องประมวลข้อมูลคำตามขั้นตอนดังรูปต่อไปนี้



รูปที่ 2.1 ขั้นตอนการแปลงข้อมูลข้อความเป็นข้อมูลเชิงปริมาณ

ที่มา: ผู้วิจัยสรุปจากขั้นตอนที่ระบุใน Grimmer, Roberts and Stewart (2022)

โดยเมื่อผู้วิจัยมีข้อมูลข้อความจะต้องนำข้อมูลมาแบ่งคำ (Tokenize) ซึ่งขั้นตอนนี้อาจจะมีความยากลำบากกับภาษาที่ไม่ได้เป็นอักษรภาพแบบภาษาจีนหรือภาษาที่ไม่ได้มีการวรรคคำแบบภาษาอังกฤษ เช่น ภาษาไทย ที่การแบ่งคำอาจจะต้องใช้ตัวแบบภาษาเข้ามาช่วย

จากนั้นจึงทำให้คำเป็นมาตรฐาน (Standardize) โดยการแปลงให้เป็นตัวพิมพ์เล็ก เพื่อตัดผลของการขึ้นต้นประโยคออกไป เช่น คำว่า Student กับ คำว่า student ให้เป็นคำเดียวกัน และกำจัดอักษรพิเศษต่าง ๆ (Remove Special Character) เช่น ? / & รวมถึงตัวเลขและคำเชื่อม ซึ่งจะไม่เอามาใช้วิเคราะห์ต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นจึงทำการตัดส่วนท้ายของคำ (Stemming) และการแปลงคำด้วยรายการคำในพจนาณุกรม (Lemmatization) เช่น คำว่า am, are, is เป็น be คำว่า car, cars, car's, cars' เป็น car คำว่า independently เป็น independent เพื่อลดจำนวนคำที่จะถูกนำมาเป็นตัวแปรในขั้นตอนการวิเคราะห์ต่อไป นอกจากนี้ในบางกรณีอาจจะมีการระบุหน้าที่ของคำ (Part of Speech Tagging) ในกรณีที่ไม่ต้องการแปลงทุกคำที่มีรากเดียวกันเป็นค่าเดียวกัน หรือกรณีที่ต้องการประมวลภาษาธรรมชาติในระดับสูง เช่น การระบุความหมาย (Semantic) ของประโยค

เมื่อถึงขั้นตอนนี้ ข้อมูลข้อความจะกลายเป็นข้อมูลถุงคำซึ่งมีลักษณะเป็นรายการคำ (Word List) ของแต่ละตัวอย่าง สมมติว่าข้อมูลมี 3 ตัวอย่าง คือ

ข้อความ 1 Great University and Great Class

ข้อความ 2 They can do better to provide better class!!

ข้อความ 3 Only two score, worst class ever.

เมื่อผ่านกระบวนการข้างต้น จะเหลือรายการคำเพียง great, university, class, they, better, provide, worst

ในการแปลงรายการคำเป็นเมทริกซ์คำทำโดยกำหนดแถว คือ ข้อความ คอลัมน์ คือ รายการคำ ในที่นี้จะทำได้ 3 วิธี คือ

วิธีที่ 1 คือ การแปลงข้อความเป็นเมทริกซ์ตัวแปรหุ่น (One Hot Encoding) ซึ่งผู้วิจัยให้ค่าตัวแปรหุ่นของแต่ละประโยค เท่ากับ 1 เมื่อประโยคปรากฏคำตามรายการตัวแปรนั้น และเท่ากับ 0 ถ้าไม่

ตารางที่ 2.1 ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นตัวแปรหุ่น

ประโยค	great	university	class	they	better	provide	worst
Great University and Great Class	1	1	1	0	0	0	0
They can do better to provide better class!!	0	0	1	1	1	1	0
Only two score, worst class ever.	0	0	1	0	0	0	1

ที่มา: ผู้วิจัย

วิธีที่ 2 คือ การแปลงข้อความเป็นเมทริกซ์ค่าความถี่ของคำ (Frequency Encoding) ซึ่งผู้วิจัยให้ค่าตัวแปรของแต่ละประโยค เท่ากับ ค่าความถี่ของคำที่ปรากฏ วิธีนี้คำที่ปรากฏบ่อยในประโยคจะมีค่าน้ำหนักมากเมื่อนำไปคำนวณในขั้นต่อไป อย่างไรก็ตามหากขั้นตอนการเตรียมถุงคำ เช่น การทำคำให้เป็นมาตรฐาน การตัดส่วนท้ายของคำ (Stemming) และการแปลงคำด้วยรายการคำในพจนาณุกรม (Lemmatization) จนเป็นรายการคำ (Word List) ไม่ดีจะทำให้คำบางคำ เช่น คำสรรพนาม คำเชื่อม I, They, and, the, this คำพวกนี้จะมีค่าน้ำหนักมากทั้ง ๆ ที่ไม่ได้มีความหมายมากในประโยค ทำให้ไม่มีประโยชน์การวิเคราะห์ในขั้นตอนต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.2 ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นค่าความถี่ของคำ

ประโยค	great	university	class	they	better	provide	worst
Great University and Great Class	2	1	1	0	0	0	0
They can do better to provide better class!!	0	0	1	1	2	1	0
Only two score, worst class ever.	0	0	1	0	0	0	1

ที่มา: ผู้วิจัย

วิธีที่ 3 คือ การแปลงข้อความเป็นเมทริกซ์ความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร (Term Frequency - Inverse Document Frequency: TFIDF) ซึ่งผู้วิจัยให้ค่าตัวแปรของแต่ละประโยค เท่ากับ ค่า TFIDF โดย คำนวณตามสูตรต่อไปนี้

$$TFIDF_{t,i} = TF_{t,i} \times IDF_t$$
$$TF_{t,i} = \frac{f(t_{t,i})}{\sum_{t' \in i} f(t_{t',i})} \quad (2.1)$$
$$IDF_t = log\left(\frac{N}{n_t}\right)$$

เมื่อ t คือ คำ
i คือ ข้อความ หรือ เอกสาร
f คือ การนับจำนวนคำ
N คือ จำนวนชุดข้อความ
n_t คือ จำนวนชุดข้อความที่ปรากฏคำ t

ตารางที่ 2.3 ตัวอย่างเมทริกซ์คำเมื่อแปลงเป็นค่าความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร

ประโยค	great	university	class	they	better	provide	worst
Great University and Great Class	0.24	0.12	0.00	0.00	0.00	0.00	0.00
They can do better to provide better class!!	0.00	0.00	0.00	0.10	0.19	0.10	0.00
Only two score, worst class ever.	0.00	0.00	0.00	0.00	0.00	0.00	0.24

ที่มา: ผู้วิจัย

วิธีการแปลงข้อความเป็นเมทริกซ์ความถี่ของคำ-ส่วนกลับความถี่ของเอกสาร (Term Frequency - Inverse Document Frequency) ช่วยให้คำที่ปรากฏมากในบางเอกสารมีค่าน้ำหนักมาก คำที่ปรากฏมากในหลายเอกสารมีค่าน้ำหนักน้อย ซึ่งวิธีการนี้ทำให้คำเชื่อม คำสรรพนาม ที่ปรากฏบ่อยแต่ไม่มีความหมายในการวิเคราะห์มีค่าน้ำหนักน้อยมาก เช่นกรณีข้างบนที่คำว่า Class ปรากฏในทุกประโยค ทำให้คำนี้ไม่มีน้ำหนักเพราะไม่สามารถนำมาใช้วิเคราะห์อะไรต่อได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแปลงข้อมูลข้อความเป็นเมทริกซ์ ทำให้ข้อมูลสามารถนำไปใช้ในการวิเคราะห์ต่อได้เช่น การจำแนกข้อความชมกับข้อความต่อว่าด้วยตัวแบบทางสถิติต่าง ๆ และเนื่องจากแต่ละประโยคเปรียบเสมือน 1 เวกเตอร์ (ตารางที่ 2.3 แนวนอน) ทำให้สามารถใช้ตัวแบบทางสถิติเช่นการวัดค่าความคล้ายคลึงโคไซน์ (Cosine similarity) มาใช้ในการวิเคราะห์ว่าข้อความแต่ละข้อความมีข้อความใดบ้างที่มีความคล้ายกันโดยอาศัยเมทริกซ์คำ

หรือในอีกทางหนึ่งคือ การวัดความคล้ายกันของคำโดยมองแต่ละคำเปรียบเสมือน 1 เวกเตอร์ (ตารางที่ 2.3 แนวตั้ง) ซึ่งในการศึกษานี้ต้องการหาคุณลักษณะของมหาวิทยาลัยที่ส่งผลต่อมุมมองของผู้วิจารณ์ ทำให้ไม่ใช่เพียงต้องการประเมินความใกล้เคียงกันของประโยคเพื่อแยกข้อความชมกับข้อความต่อว่า แต่ยังต้องการศึกษาไปถึงความใกล้เคียงกันของคำเพื่อแยกกลุ่มคำที่มีผลต่อการแยกข้อความชมกับข้อความต่อว่า

อย่างไรก็ตาม เมทริกซ์คำที่เกิดขึ้นนี้ถูกสร้างได้จากแค่การปรากฏของคำ ซึ่งในกระบวนการวิเคราะห์จริงผู้วิเคราะห์อาจจะมีปัญหา เช่น ค่าที่ได้ของแต่ละคำยังขาดข้อมูลที่แสดงถึงความสัมพันธ์กับคำอื่น ๆ ในประโยค ซึ่งส่งผลต่อประสิทธิภาพในการนำไปใช้วิเคราะห์ต่อไป

2.1.3 ขั้นตอนการประมวลข้อมูลข้อความ โดยการแปลงคำเป็นชุดรหัสตัวเลข (Word Embedding)

ในหัวข้อก่อนหน้าคำในประโยคจะถูกทำให้เป็นเวกเตอร์โดยการใช้ข้อมูลการปรากฏหรือความถี่ของคำ ซึ่งวิธีการข้างต้นจะก่อปัญหาสำคัญ 2 ประการ คือ

1) เป็นการแปลงข้อมูลที่ไม่คำนึงถึงหน้าที่ของคำ (Part of Speech) และความหมาย (Semantic) ของคำ
2) เป็นการแปลงข้อมูลที่ทำให้การเกิดเมทริกซ์คำที่มีเลข 0 จำนวนมาก (Sparse Matrix) ซึ่งส่งผลต่อการวิเคราะห์อื่น ๆ เช่น ทำให้ตัวแบบทางสถิติมีตัวแปรมากจนอาจจะทำให้เกิดปัญหา Overfitting (Jurafsky and Martin, 2009)

การแปลงคำเป็นเวกเตอร์ชุดรหัสตัวเลขจะสามารถทำได้ดีมากขึ้นโดยการประกอบตัวแบบภาษาและการใช้ข้อมูลนอกตัวแบบเข้ามาช่วย ในการศึกษานี้ผู้วิจัยใช้ตัวแบบ Word2Vec ซึ่งเป็นตัวแบบที่เป็นที่นิยมในการแปลงคำเป็นชุดรหัสตัวเลข โดยตัวแบบนี้ถูกพัฒนาโดย Tomas Mikolov และคณะ ซึ่งเป็นทีมนักวิจัยของ Google ในปี พ.ศ. 2556 (Mikolov, Sutskever, Chen, G. Corrado, *et al.*, 2013)

การแปลงคำเป็นชุดรหัสตัวเลข (Word Embedding) ด้วยวิธี Word2Vec ทำให้สามารถแปลงข้อมูลคำและวลีต่าง ๆ เป็นค่าเวกเตอร์ โดยเวกเตอร์ของคำและวลีใด ๆ ที่มีหน้าที่เดียวกันและมีความหมายคล้ายกันจะมีระยะห่างใกล้กัน ซึ่งช่วยให้ขั้นตอนการวิเคราะห์มีความแม่นยำมากขึ้น และเนื่องจากว่าเวกเตอร์ไม่ได้ขึ้นกับการปรากฏของคำในชุดข้อมูลทำให้สามารถที่จะได้เมทริกซ์คำที่มีขนาดเล็กกว่าและไม่มีเลข 0 มาก (Dense Matrix)

วิธี Word2Vec เป็นวิธีที่สร้างตัวแบบเพื่อการจำแนก (Classifier) เพื่อทำนายความน่าจะเป็นของการปรากฏร่วมกันของคำที่เป็นเป้าหมาย (Target Words) และคำบริบท (Context Words) หรือคำที่อยู่รอบข้างคำเป้าหมาย ซึ่งทำให้ต้องมีการกำหนดว่าจะใช้กลุ่มคำใดเป็นตัวแปรต้นและตัวแปรตาม และวิธีการคำนวณค่าเหมาะสมของสมการเป้าหมาย ทำให้ตัวแบบสามารถกำหนดกระบวนการ (Algorithm) ได้ 2 วิธี (Chaubard *et al.*, 2019) คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) วิธี Skip-gram คือ การทำนายคำบริบท (Context Words) จากคำเป้าหมาย (Target Word)
2) วิธี CBOW (Continuous Bag of Words) คือ การทำนายคำเป้าหมาย (Target Word) จากคำบริบท (Context Words)

และกำหนดวิธีการคำนวณค่าเหมาะสมของสมการเป้าหมาย (Training Method) ได้ 2 วิธี

1) วิธี Negative Sampling สุ่มคำที่ไม่อยู่ในบริบท (Negative Sample) ในฐานข้อมูลระบบคำ (Lexicon) มาใช้ในการคำนวณ
2) วิธี Hierarchical Softmax เป็นวิธีการสร้างโครงข่ายต้นไม้ของคำเพื่อนำมาใช้ในการคำนวณ

โดยในการวิจัยนี้ผู้วิจัยได้ใช้วิธี Skip-gram with Negative Sampling ในการแปลงข้อมูลคำเป็นเวกเตอร์ เนื่องจากเป็นวิธีที่ถูกทดสอบในบทความของ Mikolov, Sutskever, Chen, G. Corrado and Dean (2013) ว่าเป็นวิธีที่มีความแม่นยำที่สุดโดยมีขั้นตอนคือ

1) กำหนดคำเป้าหมาย (Target Word) และกำหนดคำบริบท (Context Words) ใกล้เคียงของคำเป้าหมาย ให้เป็นกลุ่มข้อมูลที่ให้ค่าถูกต้อง (Positive Sample) หรือเป็นคำบริบทจริง
2) สุ่มคำอื่นที่เป็นกลุ่มข้อมูลที่ให้ค่าไม่ถูกต้อง (Negative Sample) หรือเป็นคำบริบทไม่จริงของคำเป้าหมาย
3) ใช้การถดถอยโลจิสติก (Logistic Regression) มาสร้างตัวแบบเพื่อแยกกลุ่มข้อมูลที่ให้ค่าถูกต้องกับกลุ่มข้อมูลที่ให้ค่าไม่ถูกต้อง
4) ค่าพารามิเตอร์ที่ได้จากการคำนวณจะเป็นชุดรหัสตัวเลขของแต่ละคำ

ในทางคณิตศาสตร์ กำหนดคำเป้าหมาย คือ w และคำบริบท คือ c ซึ่งมีทั้งที่เป็นคำบริบทจริง (การเป็น Positive Sample: D=1) และคำบริบทไม่จริง (การเป็น Negative Sample: D=0)

ฉะนั้น โอกาสที่ c จะเป็นคำบริบทของ w คือ

$$P(D = 1|w, c)$$

กำหนดให้ c·w คือ ระยะห่างระหว่างคำทั้ง 2 คำนวณจากค่าความคล้ายคลึงโคไซน์

ฉะนั้น ตัวแบบที่ใช้ทำนายว่า c จะเป็นคำบริบทของคำเป้าหมาย w จะเป็นดังการถดถอยโลจิสติกต่อไปนี้

$$P(D = 1|w, c, \theta) = \sigma(c \cdot w) = \frac{1}{1 + exp(-c \cdot w)} \quad (2.2)$$

เนื่องจากคำบริบท c มีหลายคำ ฉะนั้น ในแต่ละคำเป้าหมายจะมีคำบริบทจริง D คำ สมมติว่าแต่ละคำเป็นอิสระต่อกัน ฉะนั้น 1 คำเป้าหมายจะคำนวณโอกาสของแต่ละคำบริบทจริง ดังสมการต่อไปนี้

$$logP(D = 1|w, c, \theta) = \sum_D log\left(\frac{1}{1 + exp(-c \cdot w)}\right) \quad (2.3)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในทางตรงข้าม ในแต่ละคำเป้าหมายจะมีคำบริบทไม่จริง D’ คำ สมมติว่าแต่ละคำเป็นอิสระต่อกัน ฉะนั้น 1 คำเป้าหมายจะคำนวณโอกาสของแต่ละคำบริบทไม่จริง ดังสมการต่อไปนี้

$$logP(D = 0|w, c, \theta) = \sum_{D'} log\left(\frac{1}{1 + exp(c \cdot w)}\right) \quad (2.4)$$

ในแต่ละคำเป้าหมายจำมีพารามิเตอร์ θ ทั้งหมด D+D’ ตัว ซึ่งคือขนาดของเวกเตอร์ที่จะแทนแต่ละคำเป้าหมาย โดยที่ θ

$$\theta = argmax_{\theta} \sum_{D} log\left(\frac{1}{1 + exp(-c \cdot w)}\right) + \sum_{D'} log\left(\frac{1}{1 + exp(c \cdot w)}\right) \quad (2.5)$$

ซึ่งลักษณะของฟังก์ชันแบบนี้ จะสังเกตว่าคำที่มีคำบริบทใกล้เคียงกัน จะมีค่า θ ที่ใกล้เคียงกัน กล่าวคือ คำที่มีหน้าที่หรือความหมายใกล้เคียงกันจะถูกแปลงเป็นค่าเวกเตอร์ที่มีระยะห่างใกล้กันนั้นเอง

2.2 ตัวแบบทางสถิติเพื่อการวิเคราะห์ข้อมูลข้อความ

ในบทก่อนหน้าผู้วิจัยได้ทบทวนวรรณกรรมเกี่ยวกับการแปลงข้อมูลข้อความสู่ข้อมูลตัวเลขที่มีโครงสร้างทั้งในลักษณะของเมทริกซ์คำ เวกเตอร์ประโยคและเวกเตอร์คำ เนื้อหาในส่วนต่อไปนี้จะเป็นการประมวลทางสถิติที่สำคัญในการประมวลชุดข้อมูลตัวเลขที่มาจากข้อความ

2.2.1 การจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA)

การจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA) เป็นหนึ่งในเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) แบบการเรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) หรือ เป็นตัวแบบที่ไม่มีการระบุตัวแปรตามในชุดข้อมูล ใช้เพื่อเป็นตัวแบบหัวข้อ (Topic Modeling) ของชุดข้อมูล (ประโยค หรือ เอกสาร) ซึ่งผู้ใช้งานจะต้องกำหนดจำนวนของหัวข้อที่เป็นไปได้ (K) จากนั้นตัวแบบจะทำนายความน่าจะเป็น (Generative Probabilistic Model) ที่แต่ละชุดข้อมูลจะนำเสนอหัวข้อ (Topics) ใด ๆ โดยการพิจารณาจากตัวแปร (คำ หรือ วลี) ที่ปรากฏในชุดข้อมูลนั้น

การจัดสรรของดีรีเคลแฝงถูกคิดค้นโดย Blei, Ng and Jordan (2003) ซึ่งตัวแบบนี้จะใช้สำหรับการจัดกลุ่มชุดข้อมูล แต่จะไม่เหมือนกับการวิเคราะห์แบ่งกลุ่ม (Cluster Analysis) เช่น วิธีการแบ่งกลุ่มข้อมูลแบบเคมีน (K-means Clustering) เพราะ การจัดสรรของดีรีเคลแฝงจะไม่ได้แยกให้ชุดข้อมูลถูกกำหนดเป็นกลุ่มใดกลุ่มหนึ่ง แต่จะเป็นการทำนายโอกาสที่ชุดข้อมูลนั้น ๆ จะอยู่ในกลุ่มต่าง ๆ และแตกต่างจากตัวแบบทำนายหัวข้อแบบ ดัชนีความหมายแฝง (Latent Semantic Indexing: LSI) ที่ใช้เทคนิค Singular Values Decomposition (SVD) ในการลดจำนวนมิติของชุดข้อมูลทั้งหมด (เมทริกซ์เอกสารและคำ) และใช้ ค่าความคล้ายคลึงโคไซน์ (Cosine Similarity) เพื่อจัดกลุ่มชุดข้อมูล หรืออาจจะกล่าวได้ว่า การจัดสรรของดีรีเคลแฝงเป็นตัวแบบที่สร้างตัวแปรหัวข้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Generative Model) ขณะที่ดัชนีความหมายแฝงเป็นตัวแบบที่จัดกลุ่มตัวแปรหัวข้อ (Discriminative Model)

การจัดสรรของดีรีเคลแฝงใช้การแจกแจงดีรีเคล (Dirichlet distribution) ในองค์ประกอบของตัวแบบ โดยที่การแจกแจงดีรีเคลเป็นการแจกแจงความน่าจะเป็นที่ใช้ในการจำลองความสัมพันธ์ระหว่างตัวแปรที่มีค่าเป็นจำนวนจริงบวกและผลรวมของค่าทุก ๆ ตัวแปรมีค่าเท่ากับ 1 เช่น ความน่าจะเป็นของหัวข้อ (Topics) จำนวน K หัวข้อ ที่ผลรวมจะเท่ากับ 1

ในทางคณิตศาสตร์ การแจกแจงดีรีเคล คือ ความน่าจะเป็นของค่าตัวแปรสุ่ม X ที่อยู่ในช่วง [0, 1] และมีผลรวมเท่ากับ 1 มีพารามิเตอร์ คือ $\boldsymbol{\alpha}$ มีฟังก์ชันการแจกแจงดังนี้ ดังนี้

$$Dir(x_1,\ x_2,\ \ldots,\ x_K\ /\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_K) = \frac{1}{B(\alpha)}\prod_{i=1}^{K} x_i^{\alpha_i - 1},$$

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{K}\alpha_i)}, \tag{2.6}$$

$$\Gamma(\boldsymbol{\alpha}) = \int_0^{\infty} d\,t\, t^{\boldsymbol{\alpha}-1}e^{t}.$$

โดยที่

$x_1,\ x_2,\ \ldots,\ x_K$ คือ ค่าตัวแปรสุ่มที่อยู่ในช่วง [0, 1] และมีผลรวมเท่ากับ 1

$\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_K$ คือ พารามิเตอร์ที่เรียกว่าพารามิเตอร์ความเชื่อมั่น (Concentration Parameters)

K คือ จำนวนหัวข้อ

B คือ ฟังก์ชันการแจกแจงบีตา (Beta Distribution Function)

Γ คือ ฟังก์ชันการแจกแจงแกมมา (Gamma Distribution Function)

นอกจากนี้ การจัดสรรของดีรีเคลแฝงยังใช้การแจกแจงแบบพหุนาม (Multinomial Distribution) ซึ่งเป็นรูปทั่วไปของการแจกแจงทวินาม โดยผลลัพธ์ของการสุ่มแต่ละครั้งมีมากกว่า 2 อย่าง เป็นรูปแบบของการหาค่าความน่าจะเป็นของการเกิดเหตุการณ์จากการทดลอง n ครั้ง ซึ่งทำให้เป็นรูปแบบการแจกแจงที่มักจะนำมาใช้ในการจำแนกกลุ่ม มีฟังก์ชันการแจกแจงดังนี้

$$Multinomial(x|\boldsymbol{\theta}) = \frac{n!}{\prod_{i=1}^{V} x_i}\prod_{i=1}^{V}\theta_1^{x_1},$$

$$n = \sum_{i=1}^{V} x_i, \tag{2.7}$$

$$\sum_{i=1}^{V}\theta_{i=1}$$

โดยที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$x_1, x_2, \ldots, x_V$ คือ ค่าตัวแปรสุ่มของคำ
(x = 0 เมื่อไม่ปรากฏคำนี้ และ x = 1 เมื่อปรากฏคำนี้)

$\theta_1, \theta_2, \ldots, \theta_V$ คือ พารามิเตอร์
$\theta_i = P(w_i)$ คือ โอกาสที่คำที่ i จะปรากฏ

V คือ จำนวนคำ

การจัดสรรของดีรีเคลแฝง กำหนดให้

1) หัวข้อแต่ละหัวข้อ (Topics) แทนด้วย k = 1, 2, ..., K คือ การแจกแจงของคำ (Words) แทนด้วย v = 1,2, ..., V
2) K ถูกกำหนดจากภายนอก
3) ชุดข้อมูล (ประโยค หรือ เอกสาร) แต่ละชุด แทนด้วย d = 1, 2, ..., D ประกอบไปด้วยหัวข้อที่หลากหลาย (Mixture of Topics)
4) คำแต่ละคำ ถูกสุ่มจากการกระจายที่หลากหลาย
5) ชุดข้อมูล (ประโยค หรือ เอกสาร) แต่ละชุด คือ การเรียงกันของคำจำนวน N คำ แทนด้วย w = (w_1, w_2, ..., w_N)
6) คำในประโยค และ เอกสาร เป็นสิ่งที่ถูกสังเกตได้
แต่ หัวข้อของแต่ละคำ การแจกแจงของหัวข้อในแต่ละเอกสาร การแจกแจงของคำในแต่ละหัวข้อเป็นตัวแปรแฝง (Latent Variable) ซึ่งจะต้องถูกสร้างขึ้นมา

โดยที่

θ_d คือ พารามิเตอร์แสดงสัดส่วนของหัวข้อต่าง ๆ ของเอกสาร d
เป็นเวกเตอร์ขนาด 1xK แต่ละสมาชิกมีค่า [0,1] โดยที่ผลรวมเท่ากับ 1

β_k คือ พารามิเตอร์แสดงการแจกแจงของคำในหัวข้อ k
เป็นเวกเตอร์ขนาด 1xV แต่ละสมาชิกมีค่า [0,1] โดยที่ผลรวมเท่ากับ 1

$z_{d,n}$ คือ หัวข้อของคำลำดับที่ n ของเอกสาร d
เป็นจำนวนเต็มมีค่า [1,2, ...,K]

$w_{d,n}$ คือ คำลำดับที่ n ของเอกสาร d
เป็นจำนวนเต็มมีค่า [1,2, ...,V]

α คือ การแจกแจงก่อนที่เป็น (Prior Distribution) ของ θ_d
เป็นเวกเตอร์ขนาด 1xK แต่ละสมาชิกมีค่า [0,1] โดยที่ผลรวมเท่ากับ 1

η คือ การแจกแจงก่อนที่เป็น (Prior Distribution) ของ β_k
เป็นเวกเตอร์ขนาด 1xV แต่ละสมาชิกมีค่า [0,1] โดยที่ผลรวมเท่ากับ 1

การจัดสรรของดีรีเคลแฝงกำหนดให้

$\theta_d \sim Dir(\alpha)$, $\beta_k \sim Dir(\eta)$,

$z_{d,n} \sim Multinomial(\theta_d)$ และ $w_{d,n} \sim Multinomial(\beta_{z_{d,n}})$

จำนวนตัวแปรที่สังเกตุได้จะมีค่าเท่ากับ NxD คือ จำนวนคำที่สังเกตุได้ของทุกเอกสาร
ความน่าจะเป็นของหัวข้อ k ของเอกสาร d คือ

$$\theta_{d,k} = p(z_{d,n} = k | \theta_{d,k}) \quad (2.8)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

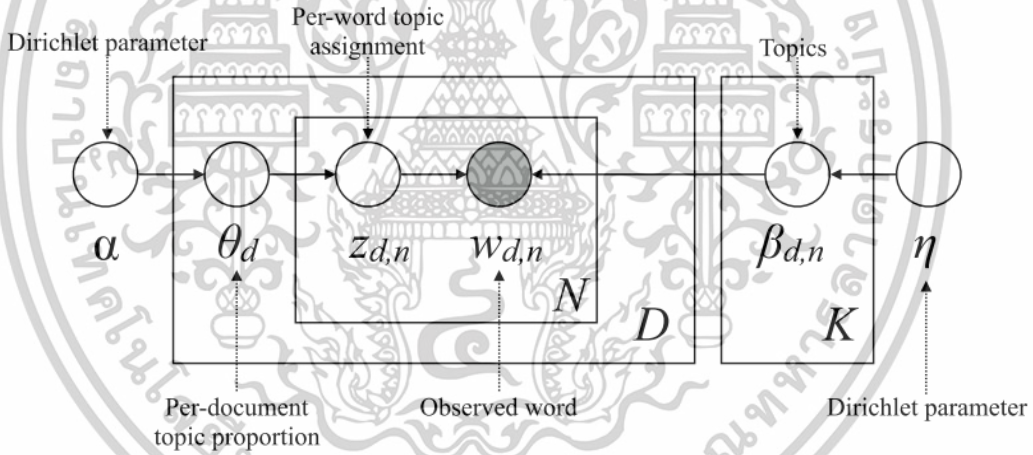
ความน่าจะเป็นของคำลำดับที่ n ของเอกสาร d ที่จะแสดงหัวข้อ k คือ

$$\beta_{z_{d,n},v} = p(w_{d,n} = v|z_{d,n}, \beta_1, \beta_2, \ldots, \beta_k) \quad (2.9)$$

การแจกแจงร่วม (Joint Distribution) ของทุกตัวแปร ทั้งตัวแปรคำและหัวข้อ คือ

$$p(w_{1:D}, z_{1:D}, \beta_{1:K}, w_{1:D}|\alpha, \eta) =$$
$$\prod_{k=1}^{K} p(\beta_k|\eta) \prod_{d=1}^{D} \left[p(\theta_d|\alpha) \prod_{n=1}^{N} \left[p(z_{d,n}|\alpha) p(w_{d,n}|z_{d,n}, \beta_{1:K}) \right] \right] \quad (2.10)$$

จากจำนวนเอกสาร D จำนวนหัวข้อ K และ การแจกแจงก่อนที่เป็น (Prior Distribution) เราจะหาพารามิเตอร์ที่ทำให้ค่าการแจกแจงร่วมนี้มีค่าสูงที่สุด โดยความสัมพันธ์ของทุกค่าพารามิเตอร์สามารถสรุปได้ดังรูปต่อไปนี้



รูปที่ 2.2 ความสัมพันธ์ของพารามิเตอร์ในตัวแบบ LDA
ที่มา: รูปที่ 1 ใน Syed and Spruit (2017)

วิธีการแก้ปัญหาการหาค่าสูงสุดสามารถทำได้หลายรูปแบบ เช่น ในบทความของ Blei, Ng and Jordan (2003) เสนอการใช้วิธี Variational Bayesian inference (VB) ร่วมกับ Expectation-Maximization Algorithm (EM Method) ซึ่งเป็นวิธีจากการศึกษาของ Dempster, Laird and Rubin (1977) นอกจากนี้จะมีวิธี Gibbs Sampling LDA (Griffiths and Steyvers, 2004) และวิธี Online Variational Bayes (VB) algorithm LDA (Hoffman, Bach and Blei, 2010) ซึ่งการศึกษานี้ใช้วิธี Online Variational Bayes (VB) algorithm LDA เนื่องจากเป็นวิธีที่ใช้เวลาในการประมวลผลต่ำกว่าอีก 2 วิธีก่อนหน้าและให้ผลค่อนข้างดี (Špeh, Muhič and Rupnik, 2013) ซึ่งสามารถดำเนินการผ่านชุดคำสั่งของ Gensim และประมวลผลโดยโปรแกรมภาษา Python

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้ เนื่องจากว่าตัวแบบการจัดสรรของดีรีเคลแผง ผู้วิเคราะห์จะต้องกำหนดค่าจำนวนหัวข้อ (K) ซึ่งการเลือกค่าจำนวนหัวข้อที่เหมาะสมนั้นจะสามารถพิจารณาได้ 2 วิธี คือ

1) ความสามารถในการตีความ (Interpretability) โดยพิจารณาจากความสามารถในการอธิบายหัวข้อจากกลุ่มคำที่ถูกกำหนดในแต่ละหัวข้อ ว่าหัวข้อดังกล่าวหมายถึงอะไร
2) ค่าคะแนนความเชื่อมโยง (Coherence Score)
   คือ การวัดความใกล้เคียงกันของคำหลักในแต่ละหัวข้อ โดยถ้าคำหลักในแต่ละหัวข้อมีความเชื่อมโยงกันสูง แปลว่า การแบ่งหัวข้อมีความชัดเจน (Less diffuse or More focused) (Röder, Both and Hinneburg, 2015) ซึ่งสามารถคำนวณได้หลายวิธี ตัวอย่างเช่น
   2.1. ค่าคะแนนความเชื่อมโยงแบบ Pointwise Mutual Information (PMI)

$$CV_{PMI} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} log\left(\frac{P(w_i, w_j) + \varepsilon}{P(w_i) \cdot P(w_j)}\right) \quad (2.11)$$

โดยที่

N คือ จำนวนคำที่เป็นคำหลักของแต่ละหัวข้อ
w_i คือ คำหลัก i ใด ๆ
$P(w_i)$ คือ ความน่าจะเป็นที่จะเจอคำหลัก i
$P(w_i, w_j)$ คือ ความน่าจะเป็นที่จะเจอคำหลัก i และ คำหลัก j ในเอกสารเดียวกัน

ค่า CV_{PMI} มีค่า [0,1] โดยที่คะแนนยิ่งมีค่ามาก นั้นหมายถึง การจัดสรรของดีรีเคลแผงทำให้เกิดการแบ่งหัวข้อได้ชัดเจนดี (Duraivel, Lavanya and Augustine, 2022) คำหลักในแต่ละหัวข้อมีความหมายใกล้เคียงกันเนื่องจากมักจะปรากฏพร้อมกัน

2.2. ค่าคะแนนความเชื่อมโยงแบบ University of Massachusetts (UMass)

$$CV_{UMass} = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} log\left(\frac{P(w_i, w_j) + \varepsilon}{P(w_j)}\right) \quad (2.12)$$

โดยที่

N คือ จำนวนคำที่เป็นคำหลักของแต่ละหัวข้อ
w_i คือ คำหลัก i ใด ๆ
$P(w_i)$ คือ ความน่าจะเป็นที่จะเจอคำหลัก i
$P(w_i, w_j)$ คือ ความน่าจะเป็นที่จะเจอคำหลัก i และ คำหลัก j ในเอกสารเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า CV_{UMass} โดยทั่วไปจะมีค่า [-14,14] โดยที่คะแนนยิ่งเข้าใกล้ 0 นั้น หมายถึง การจัดสรรของดีรีเคลแผงทำให้เกิดการแบ่งหัวข้อได้ชัดเจนดี คำหลักในแต่ละหัวข้อมีความหมายใกล้เคียงกันเนื่องจากมักจะปรากฏพร้อมกัน

ทั้งนี้ ไม่มีเกณฑ์ตายตัวว่าช่วงค่าคะแนนความเชื่อมโยงเท่าไหร่จะถือว่าอยู่ในเกณฑ์ที่เหมาะสมเพราะขึ้นกับชุดข้อมูลที่นำมาวิเคราะห์โดยทั่วไป ในหลายกรณีการเพิ่มจำนวนหัวข้อทำให้ค่า CV_{PMI} เข้าใกล้ 1 หรือ ค่า CV_{UMass} เข้าใกล้ 0 ซึ่งเป็นค่าเหมาะสมสูงสุดได้ แต่อาจจะสูญเสียความสามารถในการตีความไปได้ และเนื่องจากค่าคะแนนความเชื่อมโยงคำนวณจากการปรากฏคู่กันของคำ ทำให้อาจจะเกิดปัญหากับกรณีคำที่มีความเหมาะเหมือนกัน ซึ่งเมื่อเอกสารใช้คำหนึ่งก็มักจะไม่ใช้อีกคำหนึ่ง ส่งผลต่อค่าคะแนนความเชื่อมโยงได้ ฉะนั้น จึงควรพิจารณาค่าคะแนนความเชื่อมโยงกับความสามารถในการตีความควบคู่กัน

2.2.2 การวิเคราะห์แบ่งกลุ่ม (Cluster Analysis)

การวิเคราะห์แบ่งกลุ่ม (Cluster Analysis) คือ การวิเคราะห์ลักษณะของข้อมูลเพื่อแบ่งกลุ่มเมื่อข้อมูลที่นำมาใช้สร้างตัวแบบไม่มีข้อมูลระบุกลุ่ม หรือไม่มีตัวแปรเป้าหมาย เพื่อให้ภายในกลุ่มมีลักษณะคล้ายกัน และระหว่างกลุ่มมีลักษณะแตกต่างกัน (สายชล สินสมบูรณ์ทอง, 2559) โดยการแบ่งกลุ่มจะมี 2 ประเภท คือ

1) การแบ่งกลุ่มแบบมีขั้นตอน (Hierarchical Cluster Analysis)
คือ วิธีการแบ่งกลุ่มโดยค่อย ๆ จับกลุ่มข้อมูลให้มีจำนวนกลุ่มลดลงเรื่อย ๆ เมื่อมีการจับกลุ่มของชุดข้อมูลแล้วจะไม่มีการย้ายกลุ่ม แต่จะมีการควบรวมกลุ่มย่อยเข้าไว้ด้วยกันไปเรื่อย ๆ เช่น วิธี Agglomeration
2) การแบ่งกลุ่มแบบไม่มีขั้นตอน (Nonhierarchical Cluster Analysis)
คือ วิธีการแบ่งกลุ่มโดยมีการย้ายกลุ่มเรื่อย ๆ จนกว่าจะได้กลุ่มที่ระยะห่างภายในกลุ่มน้อยที่สุดและระยะระหว่างกลุ่มสูงที่สุด เช่น วิธีการแบ่งกลุ่มข้อมูลแบบเคมีน (k-means Clustering)

ในการศึกษานี้จะใช้การแบ่งกลุ่มข้อมูลแบบมีขั้นตอน

ขั้นตอนที่ 1 คำนวณระยะห่างของแต่ละจุดข้อมูล โดยใช้สูตรตามความเหมาะสม เช่น ค่าระยะห่างโคไซน์ (Cosine Distance) หรือ ระยะห่างยุคลิด (Euclidean Distance)

ขั้นตอนที่ 2 เลือกค่าระยะห่างที่น้อยที่สุด รวมกลุ่มชุดข้อมูลที่มีระยะห่างน้อยที่สุดเป็นกลุ่มเดียวกัน

ขั้นตอนที่ 3 ทำขั้นตอนที่ 1 และ 2 ซ้ำ จนกว่าชุดข้อมูลทั้งหมดจะเป็นกลุ่มเดียวกัน K =1

ขั้นตอนที่ 4 พิจารณาจำนวน K ที่เหมาะสม

การพิจารณาจำนวนกลุ่ม (K) ที่เหมาะสมอาจจะพิจารณาจากการทดลองทำไปเรื่อย ๆ และพิจารณาจาก

1) ความสามารถในการตีความ (Interpretability) โดยพิจารณาจากความสามารถในการอธิบายความแตกต่างระหว่างกลุ่มและความเหมือนภายในกลุ่มของการแบ่งกลุ่มที่เกิดขึ้น ซึ่งขึ้นกับการวิเคราะห์ของผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) พิจารณาจากค่าสถิติ เช่น ค่าสัมประสิทธิ์เงา (Silhouette Coefficient)

ค่าสัมประสิทธิ์เงา (Silhouette Coefficient)

ค่าสัมประสิทธิ์เงา (Silhouette Coefficient) (ถูกนำเสนอโดย Rousseeuw, 1987) ซึ่งเป็นค่าสัมประสิทธิ์ที่ใช้วัดประสิทธิภาพการแบ่งกลุ่มจากระยะห่างของหน่วยข้อมูลภายในกลุ่ม และระหว่างกลุ่ม โดยมีสูตรการคำนวณกรณีกำหนดจำนวนกลุ่มเท่ากับ K คือ

$$s(K) = \frac{b(K) - a(K)}{MAX\big(b(K), a(K)\big)}$$

$$a(K) = \frac{1}{|Ck| - 1} \sum_{i \in k, i' \in k} D(x_{jk} - x_{i'k}) \qquad (2.13)$$

$$b(K) = \min_{k' \neq k} \frac{1}{|Ck'|} \sum_{i \in k, i' \in k'} D(x_{\mathrm{ik}} - x_{i'k'})$$

เมื่อ $|Ck|$ คือ จำนวนสมาชิกในกลุ่ม k

D คือ ระยะทางแบบยุคลิด (Euclidean Distance)

$$|x_i, x_j| = \sqrt{\sum_p \left(x_i - x_j\right)^2} \qquad (2.14)$$

จากสูตรข้างต้น จะพบว่า a(K) คือ ระยะห่างโดยเฉลี่ยของข้อมูลในกลุ่ม k ขณะที่ b(K) คือ ระยะห่างที่สั้นที่สุดระหว่างของ $x_{"I}$" ในกลุ่ม k และ x'_{i}" ในกลุ่ม k' หรือคือ ระยะห่างระหว่างชุดข้อมูลที่เป็นเพื่อนบ้านที่ใกล้ที่สุดอยู่ต่างกลุ่มกัน

ค่า s(K) จะมีค่าที่ -1 ถึง 1 ถ้าค่า s(K) เข้าใกล้ 1 สะท้อนถึงการจัดกลุ่มที่ดีเพราะ a(K) คือ ระยะห่างโดยเฉลี่ยของข้อมูลในกลุ่มมีค่าน้อย และ b(K) คือ ค่าระยะห่างระหว่างชุดข้อมูลต่างกลุ่มมีค่ามาก ในทางตรงข้าม ถ้าค่า s(K) เข้าใกล้ -1 สะท้อนถึงการจัดกลุ่มที่ไม่ดีเพราะ a(K) คือ ระยะห่างโดยเฉลี่ยของข้อมูลในกลุ่มมีค่ามาก และ b(K) คือ ค่าระยะห่างระหว่างชุดข้อมูลต่างกลุ่มมีค่าน้อย (Tibshirani, Walther and Hastie, 2001; Han, Kamber and Pei, 2012a; Belyadi and Haghighat, 2021)

2.2.3 ค่าความคล้ายคลึงโคไซน์ (Cosine Similarity) และระยะห่างโคไซน์ (Cosine Distance)

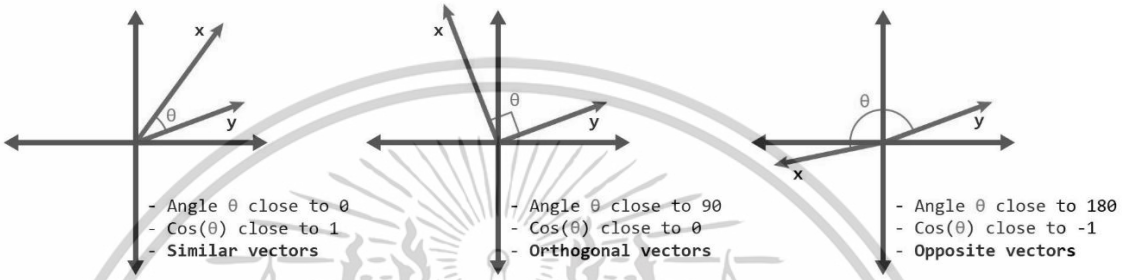
ค่าความคล้ายคลึงโคไซน์ (Cosine Similarity) เป็นวิธีการวัดระยะห่างของเวกเตอร์ ซึ่งใช้ในการประเมินความคล้ายคลึงของคำต่าง ๆ ที่ปรากฏในชุดข้อมูล

โดยหากกำหนดให้มีชุดข้อมูล $x_i = x_1, x_2, \ldots$ เป็นเวกเตอร์ที่แต่ละ x_i มีความยาว P มิติ (p = 1, 2, ..., P) ค่าความคล้ายคลึงโคไซน์ (Cosine Similarity: COS) ของข้อมูลคำนวณได้ตามสูตร (Han, Kamber and Pei, 2012b; Kotu and Deshpande, 2019) ต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Cosine\ Similarity(x_i, x_j) = \text{COS}(\theta) = \frac{\sum_p x_i \times x_j}{\sqrt{\sum_p x_i^2} \times \sqrt{\sum_p x_j^2}} \quad (2.15)$$

ค่าความคล้ายคลึงโคไซน์มีขนาด -1 ถึง 1 โดยหากค่าความคล้ายคลึงโคไซน์เข้าใกล้ 1 แสดงว่าข้อมูลมีความใกล้เคียงกันมาก และให้ค่า θ น้อย หากค่าความคล้ายคลึงโคไซน์เข้าใกล้ -1 แสดงว่าข้อมูลมีความใกล้เคียงกันน้อย และให้ค่า θ มาก โดยค่า θ คือ มุมที่เกิดจากเวกเตอร์ 2 ตัว ที่ลากจากจุดกำเนิดไปยังจุดของข้อมูลตามรูปต่อไปนี้



รูปที่ 2.2 ค่าโคไซน์กับเวกเตอร์แสดงตำแหน่งของคำ

ที่มา: Karabiber (2023)

นอกจากนี้ยังสามารถคำนวณระยะห่างโคไซน์ (Cosine Distance) ซึ่งเป็นค่าบอกความต่างของเวกเตอร์ได้มีขนาด 0 ถึง 2 โดยหากค่าระยะห่างโคไซน์เข้าใกล้ 0 แสดงว่าข้อมูลมีความใกล้เคียงกันมาก

$$Cosine\ Distance(x_i, x_j) = 1 - Cosine\ Similarity(x_i, x_j) \quad (2.16)$$

2.3 ตัวแบบทางสถิติเพื่อการวิเคราะห์ข้อมูลร้อยละ

2.3.1 การถดถอยบีตา (Beta Regression)

ในการศึกษานี้ขั้นตอนการวิเคราะห์ทางสถิติ ตัวแปรตาม คือ ค่าเฉลี่ยคะแนนความรู้สึกต่อมหาวิทยาลัยในแง่บวกและแง่ลบที่มหาวิทยาลัยได้รับ ซึ่งจะเป็นตัวแปรที่มีค่าอยู่ในช่วงจำกัด 0 ถึง 1 ทำให้การใช้การถดถอยเชิงเส้นโดยทั่วไปอาจจะไม่เหมาะนัก รวมทั้งเนื่องจากตัวแปรอาจไม่ได้มีค่าที่เน้นไปที่ค่า 0 และ ค่า 1 ทำให้ไม่เหมาะกับการใช้การถดถอยโลจีสติก (Logistic Regression)

การถดถอยบีตา (Beta Regression) เป็นรูปแบบการถดถอยที่ถูกนำเสนอโดย Ferrari and Cribari-Neto (2004) เป็นรูปแบบการถดถอยที่เป็นที่นิยมในการใช้วิเคราะห์ข้อมูลที่มีค่าอยู่ในช่วงจำกัด 0 ถึง 1 เช่น ข้อมูลสัดส่วน ข้อมูลความน่าจะเป็น (Swearingen *et al.*, 2011; Abonazel *et al.*, 2022)

การถดถอยบีตาไม่ได้มีข้อสมมติว่าข้อมูลต้องแจกแจงแบบปกติ โดยการถดถอยบีตาเป็นหนึ่งในตัวแบบเชิงเส้นทั่วไป (General Linear Model: GLM) และสามารถที่จะประมาณการค่าพารามิเตอร์ได้ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML) โดยสมมติให้ค่า y เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวแปรสุ่มต่อเนื่องที่มีแจกแจงแบบบีตา ทำให้มี ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function: PDF) ดังนี้

$$f(y;\ \mu,\emptyset)=\frac{\Gamma(\emptyset)}{\Gamma(\mu\emptyset)\Gamma\big((1-\mu)\emptyset\big)}y^{(\mu\emptyset)-1}(1-y)^{(1-\mu)\emptyset-1} \tag{2.17}$$

โดยที่

$$0<y<1\ ,\ 0<\mu<1, \emptyset>0$$
$$\emptyset=\frac{1-\sigma^2}{\sigma^2}\ ;\ E(y)=\mu\ ;\ Var(y)=\mu(1-\mu)\sigma^2$$

ค่า μ_i ขึ้นกับ ตัวแปรภายนอก X ผ่านฟังก์ชันเชื่อมโลจิสติก จะได้

$$g(\mu_i)=\log\left(\frac{\mu_i}{1-\mu_i}\right)=\mathrm{X}^T\beta=\eta$$

เมื่อ β คือ พารามิเตอร์ไม่ทราบค่า

ประมาณการค่าพารามิเตอร์ได้ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Beta Maximum Likelihood: BML) โดยจะได้ฟังก์ชันล็อกภาวะน่าจะเป็น (Loglikelihood Function) คือ

$$\begin{aligned}L(\mu_i,\emptyset;y_i)=\sum_{i=1}^{n}\{&log\Gamma(\emptyset)-log\Gamma\big(\mu_i(\emptyset)\big)\\&-log\Gamma\big((1-\mu_i)(\emptyset)\big)+(\mu_i(\emptyset)-1)\,log(y_i)\\&+\big((1-\mu_i)(\emptyset)-1\big)\,log(1-y_i)\}\end{aligned} \tag{2.18}$$

สมการข้างต้นหากหาค่าอนุพันธ์เทียบกับ β จะได้ฟังก์ชันคะแนน (Score Function)

$$\mathrm{S}(\beta)=\emptyset\mathrm{X}^{\mathrm{T}}\mathrm{A}(\mathrm{y}^*-\mu^*) \tag{2.19}$$

โดยที่

$$A=diag\left(\frac{1}{g'(\mu_1)},\frac{1}{g'(\mu_2)},\dots,\frac{1}{g'(\mu_n)}\right)$$
$$y^*=(y_1^*,y_2^*,\dots,y_n^*)^T$$
$$y_i^*=log\left(\frac{y_i}{1-y_i}\right)$$
$$\mu^*=(\mu_1^*,\mu_2^*,\dots,\mu_n^*)^T$$
$$\mu_i^*=\varphi\big((1-\mu_i)\emptyset\big)$$

เมื่อ φ คือ ฟังก์ชันไดแกมมา (Digamma Function)

ค่าประมาณของพารามิเตอร์จะมีค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\hat{\beta}_{ML} = \left(X^T \widehat{W} X\right)^{-1} X^T \widehat{W} \hat{z} \tag{2.20}$$

โดยที่

$$\hat{z} = \hat{\eta} + \widehat{W}^{-1} \hat{A}(y^* - \mu^*)$$

$$\widehat{W} = diag(\widehat{w}_1, \widehat{w}_2, \dots, \widehat{w}_n)$$

$$\widehat{w}_i = \frac{1-\hat{\sigma}^2}{\hat{\sigma}^2}\left\{\varphi'\left(\frac{\hat{\mu}_i(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right) + \varphi'\left(\frac{(1-\hat{\mu}_i)(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right)\right\}\frac{1}{\left(g'(\hat{\mu}_i)\right)^2}$$

โดย Abonazel and Taha (2021) และ Qasim, Månsson and Golam Kibria, (2021) ได้เสนอการถดถอยบีตาแบบริดจ์ (Ridge Beta Regression: RBR) Karlsson, Månsson and Kibria (2020) ได้เสนอการถดถอยบีตาแบบ Liu (Liu Beta Regression: LBR) และ Abonazel et al., (2022) ได้เสนอวิธีประมาณพารามิเตอร์การถดถอยบีตาแบบพารามิเตอร์ 2 ตัว (Two-Parameter Estimator for Beta Regression: TPBR) ซึ่งทั้งหมดเป็นตัวแบบที่ช่วยแก้ไขปัญหาความสัมพันธ์เชิงเส้นแบบพหุ (Multicollinearity)

การถดถอยบีตาแบบริดจ์ (Ridge Beta Regression: RBR) (Abonazel *et al.*, 2022) สามารถปรับด้วยค่าการลงโทษ k ค่าประมาณของพารามิเตอร์จะมีค่า

$$\hat{\beta}_{RBR} = \left(X^T \widehat{W} X + kI\right)^{-1} X^T \widehat{W} \hat{z} \tag{2.21}$$

เมื่อ I คือ เมทริกซ์เอกลักษณ์ (Identity Matrix)

โดยที่

$$k \geq 0, ||\hat{\beta}_{ML}|| > ||\hat{\beta}_{RBR}||$$

จะสังเกตว่าค่า ถ้า k = 0 แล้ว $\hat{\beta}_{ML} = \hat{\beta}_{RBR}$

โดยค่าประมาณการของพารามิเตอร์จากวิธีภาวะน่าจะเป็นสูงสุด และวิธีแบบริดจ์มีความสัมพันธ์กันตามสมการต่อไปนี้ (Akram *et al.*, 2021; Qasim, Månsson and Golam Kibria, 2021)

$$\hat{\beta}_{RBR} = \left(X^T \widehat{W} X + kI\right)^{-1} (X^T \widehat{W} X)\hat{\beta}_{ML} \tag{2.22}$$

2.4 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนการศึกษาในอดีตพบงานจำนวนมากที่มีการวิเคราะห์ข้อความแสดงความคิดเห็น โดยส่วนมากจะเป็นงานที่ใช้ข้อความแสดงความคิดเห็นเกี่ยวกับสินค้าและบริการ เช่น ข้อความแสดงความคิดเห็นบนแพลตฟอร์มอีคอมเมิร์ซ เช่น Amazon.com หรือ บนแพลตฟอร์มบริหารที่พัก เช่น Agoda.com Booking.com ซึ่งส่วนมากจะเป็นการศึกษาในด้านการเข้าใจภาษาธรรมชาติ (Natural Language Understanding: NLU) คือ การศึกษาคุณสมบัติของข้อความที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มีลักษณะที่แตกต่างกัน เช่น ลักษณะของข้อความแบบใดที่จะทำให้ข้อความได้รับการยอมรับว่ามีประโยชน์หรือเป็นที่นิยม

ตัวอย่าง การศึกษาของ Biswas *et al.* (2022) ที่วิเคราะห์ลักษณะของข้อความความคิดเห็นและตัวแปรที่เกี่ยวกับผู้ให้ความคิดเห็นส่งผลต่อการรู้สึกว่าข้อความเป็นประโยชน์ โดยใช้ Correlated Topic Models (CTM) ในการจำแนกประเด็นสำคัญที่ถูกพูดถึงในข้อความ นำไปสร้างตัวแปรต้นเกี่ยวกับข้อความซึ่งกำหนดจากทฤษฎีสารสนเทศน์ และใช้การถดถอยปัวซง (Poisson Regression) และการถดถอยทวินามเชิงลบ (Negative Binomial Regression) ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้นกับการโหวตว่าข้อความเป็นประโยชน์ ซึ่งถือเป็นหนึ่งในการศึกษาที่มีการสกัดปัจจัยจากข้อความโดยการใช้ตัวแบบทางภาษาศาสตร์ โดยมีการศึกษาในลักษณะคล้ายกันนี้จำนวนมากที่เลือกใช้วิธีที่แตกต่างกันและมีข้อสรุปที่แตกต่าง (ดูใน Ghose and Ipeirotis, 2011; Chua and Banerjee, 2015; Kuan *et al.*, 2015; Filieri *et al.*, 2018; Du *et al.*, 2021; Meek, Wilk and Lambert, 2021; Wu, Mai and Li, 2021; Yang, Zhou and Chen, 2021) ทำให้การวิจัยด้านการประมวลภาษาธรรมชาติด้วยข้อความแสดงความคิดเห็นต่อสินค้ามีความก้าวหน้ามาก

อย่างไรก็ตาม ในด้านการวิเคราะห์ข้อความเกี่ยวกับการศึกษา การวิเคราะห์ข้อความเกี่ยวกับการศึกษาส่วนมากจะยังคงเป็นการพัฒนาตัวแบบเพื่อการจัดกลุ่มข้อความ หรือการจำแนกข้อความโดยการสร้างตัวแบบที่หลากหลายและแสดงผลเปรียบเทียบประสิทธิภาพของแต่ละวิธี พบว่าการวิจัยส่วนมากจะใช้ข้อมูลแสดงความคิดเห็นที่ปรากฏบนแพลตฟอร์มการเรียนการสอนแบบออนไลน์ (Massive Open Online Course: MOOCs) เช่น การศึกษาของ Li et al. (2019) Barrón Estrada et al. (2020) และ Onan (2021) จะนำข้อความแสดงความคิดเห็นต่อ MOOCs ไปใช้สร้างตัวแบบโดยใช้วิธี Lexicon Based Methods (ใช้ฐานข้อมูลพจนานุกรมที่แบ่งกลุ่มคำเชิงบวกเชิงลบ) ในการสร้างตัวแปรลักษณะของข้อความเชิงบวกเชิงลบ

นอกจากข้อมูลข้อความแสดงความคิดเห็นบนเว็บไซต์ ยังปรากฏการศึกษาที่ทำการวิเคราะห์ข้อความโดยใช้วิธีการรวบรวมข้อมูลโดยวิธีอื่น ซึ่งส่วนมากจะใช้กับกลุ่มประชากรที่ไม่ได้ใช้ภาษาอังกฤษ เช่น การศึกษาของ Giang et al. (2020) ใช้วิธีให้ผู้เชี่ยวชาญมาจัดกลุ่มข้อความ และพบบางการศึกษาที่ใช้ข้อมูลคะแนนความพึงพอใจในการแบ่งกลุ่มข้อความ เช่น การศึกษาของ Kandhro et al. (2019) ซึ่งใช้ข้อมูลที่ได้จากแบบสอบถาม

ในกรณีการวิเคราะห์ข้อความ (Text Analysis) ที่นอกเหนือการจัดกลุ่มหรือจำแนกข้อความ พบการศึกษาจำนวนหนึ่งที่วิเคราะห์ประเด็นที่ปรากฏในประกาศรับสมัครงานว่ามีการกล่าวถึงหัวข้อทักษะใดบ้าง เช่น ใน Mezzanzanica (2020) ที่ดูว่าตำแหน่งงานแบบไหนที่มีมากขึ้นในระยะหลัง หรือ Pejic-Bach et al. (2020) ที่พิจารณาว่ากลุ่มอาชีพและทักษะแบบไหนที่ในประกาศรับสมัครงานเสนอตำแหน่งงานที่สูง นอกจากนี้ยังมีงานที่ต่อยอดด้วยการใช้ข้อมูลคำอธิบายทักษะมาวัดความสอดคล้องกับคำอธิบายรายวิชา เพื่อพิจารณาว่าการออกแบบหลักสูตรการศึกษามีการระบุเนื้อหาทักษะสอดคล้องกับความต้องการของตลาดแรงงานหรือไม่ (Almaleh *et al.*, 2019; Spada *et al.*, 2022)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.4 ตัวอย่างสรุปการทบทวนวรรณกรรมที่มีความน่าสนใจ

ผู้เขียนและปี	ข้อมูลที่ใช้	ตัวแบบที่ใช้	ข้อสรุปสำคัญ
Biswas *et al.* (2022)	- ความเห็นต่อสินค้าบน Amazon.com	- Cluster - Correlated Topic Models (CTM) - Regression	- ใช้ CTM เพื่อวิเคราะห์ประเด็นที่แต่ละข้อความกำลังแสดงถึง - สร้างตัวแปรลักษณะของข้อความตามทฤษฎี - ใช้การถดถอยเพื่อหาลักษณะของข้อความที่มีความสำคัญในการกำหนดประโยชน์ของข้อความ
Li et al. (2019)	- ความเห็นต่อ MOOCs	- Classification - BERT-CNN model	- ข้อมูลถูกจัดกลุ่มโดยใช้ Lexicon Based Methods - ผลการทดสอบตัวแบบให้ค่า Accuracy = 81.3% และค่า F1 Score = 92.8% - ตัวแบบมีค่าพารามิเตอร์กว่า 60 ล้านตัว
Ali Kandhro et al. (2019)	- ข้อมูลสำรวจเอง โดยเก็บตัวอย่างจากนักศึกษาในอินเดีย	- Classification - Naive Bayes - Stochastic Gradient Decent - Support Vector Machines - Random Forest - Multilayer Perceptron Classifier	- ข้อมูลถูกจัดกลุ่มโดยใช้ Lexicon Based Methods - แบ่งกลุ่มข้อความเป็น 3 กลุ่ม คือ Positive Neutral และ Negative - วิธี Naive Bayes และ Multilayer Perceptron Classifier ให้ค่า Accuracy และ F1-score สูงที่สุด
Giang et al. (2020)	- ข้อมูลสำรวจเอง โดยเก็บตัวอย่างจากนักศึกษาในเวียดนาม 5,000 ชุด	- Classification - Naive Bayes - Maximum Entropy - Support Vector Machines	- ให้ผู้ที่มีความรู้ด้านภาษาเวียดนามมาแบ่งกลุ่มข้อความเป็น 3 กลุ่ม คือ Positive Neutral และ Negative

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้เขียนและปี	ข้อมูลที่ใช้	ตัวแบบที่ใช้	ข้อสรุปสำคัญ
			- Support Vector Machines ให้ค่า Accuracy สูงที่สุด
Barrón Estrada et al. (2020)	- ข้อมูลจาก MOOCs และ Youtube	- Classification - Machine Learning - Deep Learning - Evolutionary Approach Called EvoMSA	- ข้อมูลถูกจัดกลุ่มโดยใช้ SentiDict - แบ่งข้อมูล 2 Class คือ Positive และ Negative - พิจารณาแยกว่ากำลังแสดงความเห็นถึงเรื่องอะไร Teachers, Exams, Homework, Academic Projects - EvoMSA ให้ค่า Accuracy สูงที่สุด
Onan (2021)	- ความเห็นต่อ MOOCs 66,000 ข้อความ	- Classification - K-Nearest Neighbor - Logistic Regression - Naïve Bayes - Random Forest - Support Vector Machines - Deep Learning Algorithms - อื่น ๆ	- ข้อมูลถูกจัดกลุ่มโดยใช้ Lexicon Based Methods - แบ่งข้อมูล 2 Class คือ Positive และ Negative - บทความใช้ตัวแบบหลากหลาย - ในแต่ละตัวแบบ จะมีการปรับตัวแบบทั้งแบบปกติ แบบใช้เทคนิค Bagging, Boost Random, Subspace Method Voting, and Stacking - ผลการทดสอบพบว่าตัวแบบในกลุ่ม Deep Learning ให้ค่า Accuracy สูงที่สุด
Mezzanzanica (2020)	- ประกาศรับสมัครงาน 1,098,589 ฉบับ	- Classification - Probabilistic Topic Models	- วิเคราะห์ข้อความเพื่อหาว่าในกลุ่มอุตสาหกรรมต่าง ๆ ประกาศรับสมัครงานต้องการให้ทำหน้าที่แบบไหน และต้องการทักษะแบบไหน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้เขียนและปี	ข้อมูลที่ใช้	ตัวแบบที่ใช้	ข้อสรุปสำคัญ
			- และดูการเปลี่ยนแปลงข้ามช่วงเวลา พบว่าหน้าที่การตลาดสื่อสังคมออนไลน์ ผู้จัดการส่งออก และผู้จัดการแบรนด์เป็นที่ต้องการของตลาดเพิ่มมากขึ้น
Pejic-Bach *et al. (*2020)	- ประกาศรับสมัครงานที่เกี่ยวกับอีตสาหกรรม 4.0	- TFIDF - Probabilistic Topic Models - Clustering	- จัดกลุ่มของอาชีพตามทักษะที่ระบุในประกาศรับสมัครงาน และเทียบว่าในแต่ละกลุ่มมีสัดส่วนของพนักงานระดับไหนบ้าง ตั้งแต่ฝึกงานจนถึงผู้บริหารระดับสูง - พบว่ากลุ่มผู้เชี่ยวชาญด้านซอฟต์แวร์มีสัดส่วนของตำแหน่งระดับอาวุโสสูงมากถ้าเทียบกับกลุ่มอาชีพอื่น ๆ
Spada *et al.* (2022)	- ประกาศรับสมัครงาน 952 ฉบับ - คำอธิบายหลักสูตร 300 ตัวอย่าง	- Bag- of-Words (BoW) - Bidirectional Encoder Representations from Transformers (BERT) - Cosine Similarity	- แปลงชุดประกาศรับสมัครงาน คำอธิบายหลักสูตร และรายการทักษะเป็นเวกเตอร์ - วัดความคล้ายระหว่างประกาศรับสมัครงานและรายการทักษะ และความคล้ายระหว่างคำอธิบายหลักสูตรและรายการทักษะ - พิจารณาว่าทักษะใดที่ระดับความคล้ายระหว่างคำอธิบายหลักสูตรแตกต่างไปจากประกาศรับสมัครงาน
Almaleh *et al.* (2019)	- หลักสูตร IT วิทยาการคอมพิวเตอร์	- Classification - Naive Bayes - Cosine Similarity	- เทียบความสอดคล้องกันของประกาศรับสมัครงานและคำอธิบายเนื้อหาวิชาในหลักสูตรที่เรียน แยก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้เขียนและปี	ข้อมูลที่ใช้	ตัวแบบที่ใช้	ข้อสรุปสำคัญ
	- ประกาศรับสมัครงาน IT วิทยาการคอมพิวเตอร์		กลุ่มวิชากับประกาศงานออกเป็นหน้าที่ต่าง ๆ - พิจารณาความสอดคล้องจากเวกเตอร์คำในแต่ละกลุ่ม - พบว่ากลุ่มวิชาเพื่อทำงานเป็นฝ่ายสนับสนุนมีเนื้อหาสอดคล้องกับประกาศหางานมาก - กลุ่มวิชานักวิเคราะห์และนักพัฒนาระบบมีเนื้อหาสอดคล้องกับประกาศหางานน้อย
Li *et al.* (2022)	- ความเห็นต่อ MOOCs 187,574 ข้อความ	- Sentiment Analysis - Probabilistic Topic Models - Correlation Coefficient	- แยกประเภทหลักสูตรเป็นหลักสูตรทักษะและหลักสูตรความรู้ เนื่องจากเชื่อว่ามีลักษณะที่แตกต่างกัน - ใช้ Sentiment Analysis ประเมินคะแนนความรู้สึก และ Probabilistic Topic Models ระบุหัวข้อที่ข้อความกล่าวถึง - ใช้ Correlation Coefficient เพื่อบอกว่าประเด็นใดบ้างที่มีความสัมพันธ์กัน

ที่มา: ประมวลโดยผู้วิจัย

หมายเหตุ

1) MOOCs: Massive Open Online Courses
2) SPOCs: Small Private Online Courses

การทบทวนเอกสารไม่พบการศึกษาที่ศึกษาความสัมพันธ์ระหว่างความรู้สึกของผู้เรียนกับคุณลักษณะของสถานศึกษา แต่พบการศึกษาที่ศึกษาความสัมพันธ์ระหว่างความรู้สึกของผู้เรียนกับวิชาที่เรียน คือ การศึกษาของ Li et al. (2022) ซึ่งใช้ การวิเคราะห์ความรู้สึก (Sentiment Analysis) ประเมินคะแนนความรู้สึก และ ตัวแบบความน่าจะเป็นของหัวข้อ (Probabilistic Topic Models)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบุหัวข้อที่ข้อความกล่าวถึง เพื่อดูว่าหลักสูตรในกลุ่มต่าง ๆ ได้คะแนนความรู้สึกในหัวข้อต่าง ๆ เป็นเท่าใด มีหัวข้อใดบ้างที่ควรพัฒนา และใช้สัมประสิทธิ์สหสัมพันธ์วิเคราะห์ว่าในแต่ละกลุ่มหลักสูตรทักษะด้านไหนบ้างที่มีความสัมพันธ์กัน ทั้งนี้ ในการศึกษาของ Li et al. (2022) ได้แสดงความสัมพันธ์ระหว่างความรู้สึกของผู้เรียนกับหัวข้อที่ข้อความกล่าวถึงในแต่ละวิชาจากการเปรียบเทียบค่าเฉลี่ยระหว่างวิชาเท่านั้น ไม่ได้มีการวิเคราะห์ว่าหัวข้อใดมีความสำคัญทางสถิติในการกำหนดความพึงพอใจของผู้เรียนในแต่ละวิชา

ในศึกษานี้ผู้วิจัยได้ประยุกต์ใช้กรอบวิธีการศึกษาของ Huai and Van de Voorde (2022) ซึ่งเป็นการศึกษาหาคุณลักษณะที่ส่งผลต่อความรู้สึกต่อสวนในเขตเมืองทั้งในแง่บวกและแง่ลบ โดยใช้ข้อมูลการแสดงความคิดเห็นต่อสวนในพื้นที่เขตเมืองเซี่ยงไฮ้ และ ข้อความการให้ความเห็นต่อพื้นที่สวนในเมืองบรัสเซลส์ โดยใช้วิธีการประมวลผลภาษาธรรมชาติตั้งแต่กระบวนการแปลงข้อความเป็นเวกเตอร์ วิเคราะห์ความรู้สึกของข้อความเพื่อแยกข้อความแสดงความรู้สึกเชิงบวกและเชิงลบ และใช้การวิเคราะห์การแบ่งกลุ่มข้อความ พบปัจจัยด้านสภาพแวดล้อม 15 ปัจจัยที่มีผลต่อความรู้สึกต่อพื้นที่สวนในเขตเมือง จากนั้นจึงประมาณการถดถอยบีตา พบว่าการที่สวนมีพื้นที่แหล่งน้ำและมีสีสันหลากหลายส่งผลบวกต่อความรู้สึกที่มีต่อพื้นที่สวนในเขตเมืองเซี่ยงไฮ้ ในทางกลับกันการที่สวนมีความสงบส่งผลบวกต่อความรู้สึกที่มีต่อพื้นที่สวนในเขตเมืองบรัสเซลส์ ซึ่งการศึกษาของ Huai and Van de Voorde (2022) เป็นตัวอย่างที่ดีในการวิเคราะห์ข้อความแสดงความเห็นเพื่อหาปัจจัยที่ส่งผลต่อความรู้สึกต่อสิ่งที่ถูกแสดงความเห็น และการเทียบเคียงระหว่างกลุ่มประชากรต่าง ๆ ที่มีโครงสร้างต่างกัน ขั้นตอนที่การศึกษาฉบับนี้จะประยุกต์ใช้แสดงในบทที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

3.1 ข้อมูลที่ใช้ในการศึกษา

3.1.1 ข้อมูลข้อความการแสดงความเห็นต่อมหาวิทยาลัย

การศึกษาฉบับนี้รวบรวมข้อมูลข้อความเห็นต่อมหาวิทยาลัย (University Reviews) จากสื่อออนไลน์ โดยการทำการขูดข้อมูลจากเว็บไซต์ที่เปิดให้ผู้คนเข้ามาเขียนแสดงความเห็นซึ่งมีหลายแหล่ง โดยผู้วิจัยสรุปแหล่งข้อมูล ขอบเขตข้อมูลและข้อมูลที่ปรากฏในแต่ละแหล่งตามตารางที่ 3.1

ตารางที่ 3.1 สรุปแหล่งข้อมูลข้อความเห็นต่อมหาวิทยาลัย

ประเทศ	แหล่งข้อมูล	ขอบเขตข้อมูล	ข้อมูล
สหราช อาณาจักร	WhatUNI.com	2013-2022 178,505 ชุด ข้อความ	ชื่อมหาวิทยาลัย วันที่เขียน ความเห็นต่อมหาวิทยาลัย คะแนนต่อมหาวิทยาลัย ความเห็นต่อวิชาเรียน คะแนนด้านต่าง ๆ
ออสเตรเลีย	UniReview.com.au	2014-2022	ชื่อมหาวิทยาลัย วันที่เขียน ความเห็นต่อมหาวิทยาลัย คะแนนต่อมหาวิทยาลัย
หลาย ประเทศ	lagora.com	ไม่ระบุ	ชื่อมหาวิทยาลัย วันที่เขียน ความเห็นต่อมหาวิทยาลัย คะแนนต่อมหาวิทยาลัย คะแนนด้านต่าง ๆ

ที่มา: ผู้วิจัย

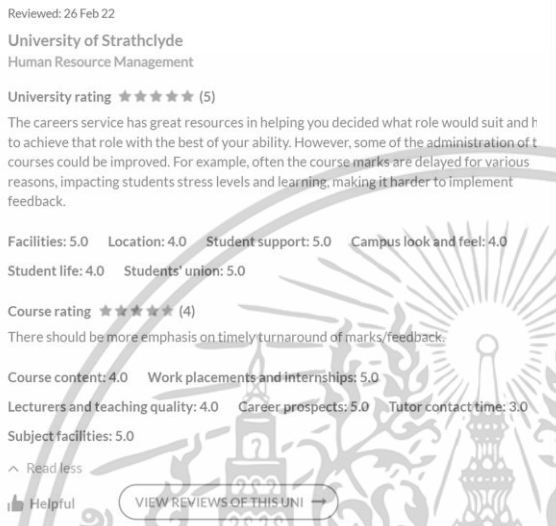
โดยการศึกษานี้ผู้วิจัยเลือกใช้ข้อมูลข้อความเห็นต่อมหาวิทยาลัยในสหราชอาณาจักรจาก WhatUNI.com ข้อความเห็นต่อมหาวิทยาลัยในออสเตรเลียจาก UniReview.com.au และข้อความเห็นต่อมหาวิทยาลัยในหลายประเทศจาก lagora.com โดยเลือกประเทศ สหราชอาณาจักร ออสเตรเลีย สหรัฐอเมริกา จีน ญี่ปุ่นและไทย เป็นประเทศที่เลือกนำมาใช้ในการศึกษา

ผู้วิจัยใช้โปรแกรมภาษา Python ผ่าน Google Colab ในการคัดลอกข้อมูล (Scrapping) ที่แสดงบนเว็บไซต์ทั้ง 3 ทุกชุดข้อมูล โดยการคัดลอก HTML Code ของแต่ละหน้าเว็บ เช่นกรณีของ Whaiuni.com ที่แสดงความเห็นหน้าละประมาณ 10 ชุดข้อความ (ขึ้นกับความยาวของข้อความ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

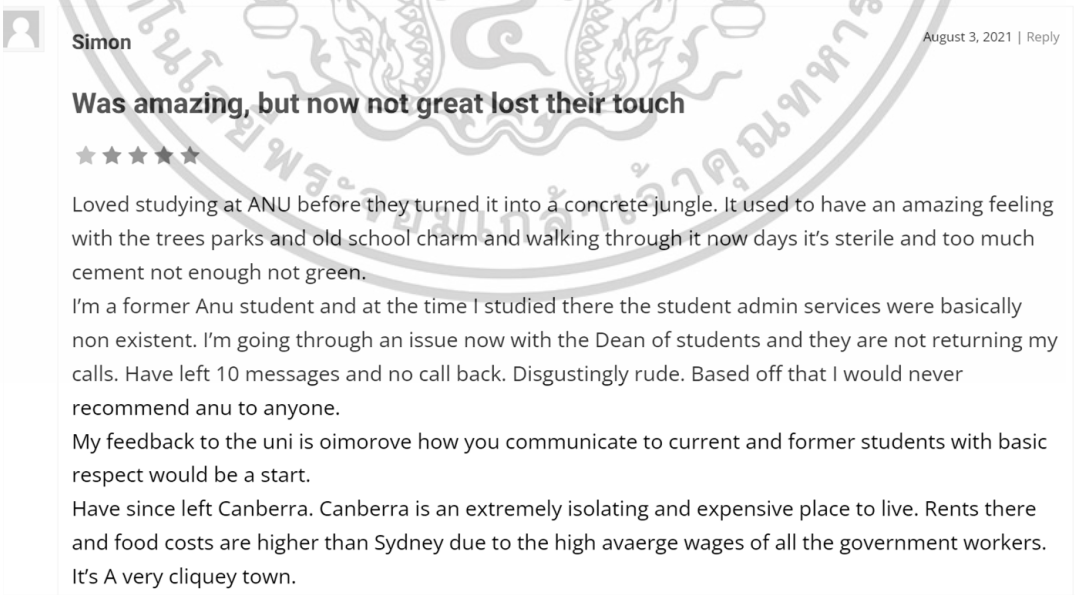
รวมทั้งสิ้น 14,716 หน้า ในวันที่ 20 ธันวาคม พ.ศ. 2565 และบันทึกข้อมูล HTML Code ในรูปแบบ .txt ซึ่งเป็นข้อมูลที่มีความยาวรวม 30.76 ล้านบรรทัด ทั้งนี้ ในแต่ละชุดข้อมูลจะมีองค์ประกอบไม่เหมือนกัน ขึ้นกับช่วงเวลาที่ผู้ให้ข้อมูลกรอกข้อมูลบนเว็บไซต์ (เว็บไซต์มีการปรับโครงสร้างช่วงปลายปี 2564) และหัวข้อประเด็นย่อยที่ผู้ให้ข้อมูลเลือกที่จะแสดงความคิดเห็น

ข้อความ 12 ธ.ค. 2564 เป็นต้นมา



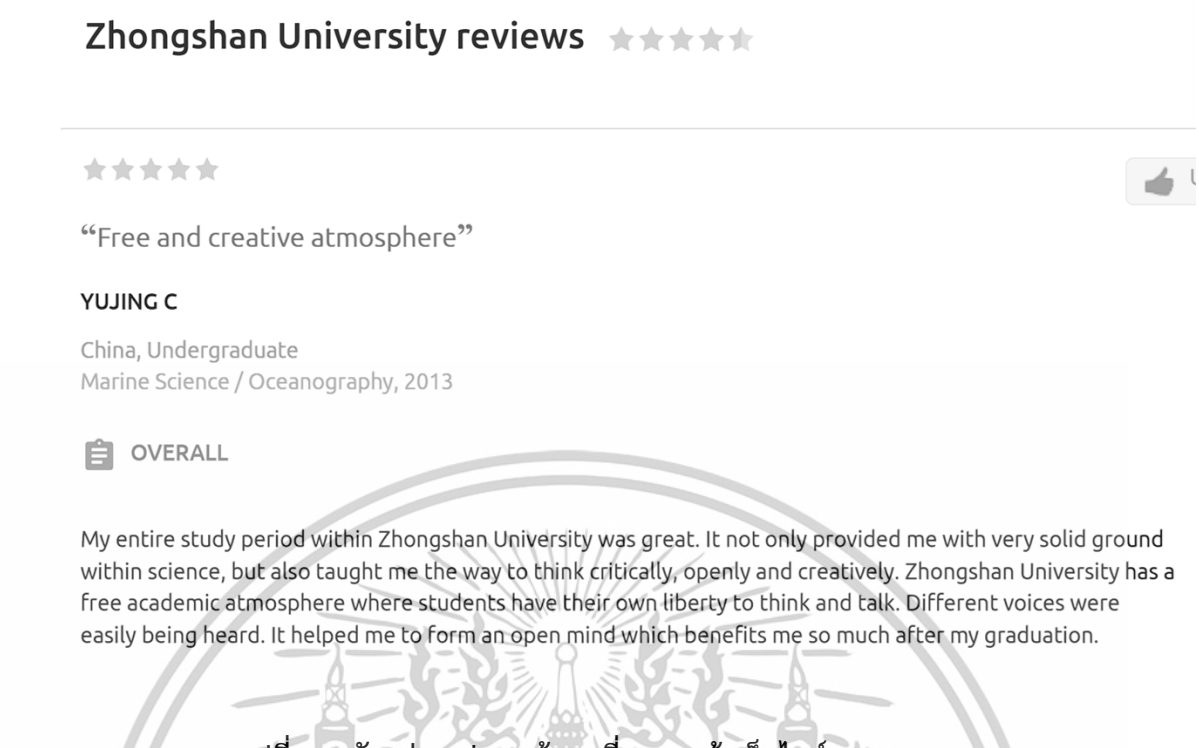
ข้อความก่อน 12 ธ.ค. 2564

รูปที่ 3.1 ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ Whaiuni.com
ที่มา: ผู้วิจัยคัดลอกจาก Whaiuni.com



รูปที่ 3.2 ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ UniReview.com.au
ที่มา: ผู้วิจัยคัดลอกจาก UniReview.com.au

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 ตัวอย่างรูปแบบข้อมูลที่แสดงหน้าเว็บไซต์ lagora.com

ที่มา: ผู้วิจัยคัดลอกจาก lagora.com

```
<p>Reviewed: 07 Apr 22</p>
</div>
</div>
<h2><a href="/university-profile/kingston-university/910/" onclick="viewProfileGaLogging('Kingston Universi
<h3><a href="/degrees/business-phd-doctor-of-philosophy/kingston-university/cd/58174562/910/" title="Busine
<div class="reviw_rating">
<div class="rate_new">
<span class="cat_rat rw_qus_des">University rating</span>
<span class="ml5 rat rat4"></span>
<span class="cat_rat rw_qus_des">(4)</span>
<p id="overallCommentId_1" class="rev_dec rm_height">The staff at the university are very helpful and have
</div>
<div id="viewMore_1" style="display:none;">
<div class="rate_new">
<div class="ext_fac_cntr">
<div class="ext_fac">Facilities: <span>4.0</span></div>
<div class="ext_fac">Location: <span>5.0</span></div>
<div class="ext_fac">Student support: <span>5.0</span></div>
<div class="ext_fac">Campus look and feel: <span>4.0</span></div>
<div class="ext_fac">Student life: <span>4.0</span></div>
<div class="ext_fac">Students' union: <span>4.0</span></div>
</div>
</div>
<div class="rate_new">
<span class="cat_rat rw_qus_des">Course rating</span>
<span class="ml5 rat rat4"></span>
<span class="cat_rat rw_qus_des">(4)</span>
<p class="rev_dec">Course is self taught as it’s a PhD however I have always received full support from my
<div class="ext_fac_cntr mt16">
<div class="ext_fac">Course content: <span>4.0</span></div>
<div class="ext_fac">Work placements and internships: <span>3.0</span></div>
```

รูปที่ 3.4 ตัวอย่างรูปแบบข้อมูล HTML Code

ที่มา: ผู้วิจัยคัดลอกจาก Whaiuni.com

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 รูปแบบการแปลงข้อมูล HTML Code เป็นตาราง

ประเทศ	ชื่อมหาวิทยาลัย	ข้อความเห็น	วันที่เขียน	ตัวแปรอื่น
ข้อความ (str)	ข้อความ (str)	ข้อความ (str)	วัตถุเวลา	

ที่มา: ผู้วิจัย

โดยข้อมูลที่ได้มาจะทำการสกัดวัตถุใส่ในตารางตามที่แสดงในตารางที่ 3.2 ประกอบด้วย ประเทศ ชื่อมหาวิทยาลัย ข้อความ วันที่เขียน และตัวแปรอื่น เพื่อเตรียมพร้อมสำหรับการวิเคราะห์ในขั้นตอนต่อไป จะสังเกตได้ว่าข้อมูลทั้งหมดยังไม่เป็นข้อมูลที่มีโครงสร้าง โดยข้อมูลสำคัญยังคงเป็นข้อความซึ่งจะต้องอาศัยการใส่รหัส (Encode) โดยจะดำเนินการในขั้นตอนของการวิเคราะห์ข้อความต่อไป

3.1.2 ข้อมูลตัวแปรคุณลักษณะของมหาวิทยาลัย

ในขั้นตอนของการวิเคราะห์ความสัมพันธ์ระหว่างประเด็นที่มหาวิทยาลัยถูกกล่าวถึงและความรู้สึกต่อมหาวิทยาลัย รวมถึงการอธิบายข้อมูลต่าง ๆ ที่ได้จากการสกัดข้อมูลจากข้อความ จำเป็นที่จะต้องมีข้อมูลตัวแปรคุณลักษณะของมหาวิทยาลัยมาใช้ประกอบการวิเคราะห์ โดยในการศึกษานี้ได้ใช้ข้อมูลที่ปรากฏในเว็บไซต์ที่มาของข้อมูลข้อความ ซึ่งจะมีการระบุชื่อมหาวิทยาลัย ประเทศ และเพิ่มเติมข้อมูลที่ปรากฏในเว็บไซต์ Timeshighereducation.com

โดยในเบื้องต้นข้อมูลที่ใช้เพื่อวิเคราะห์ในการศึกษาจะถูกแปลงเป็นตัวแปรตามรายการในตารางที่ 3.3

ตารางที่ 3.3 คำอธิบายตัวแปร

ชื่อตัวแปร	คำอธิบาย
ตัวแปรรายข้อความ	
Uname	ชื่อมหาวิทยาลัย
UReview	ข้อความเห็นต่อมหาวิทยาลัย
Uscore	คะแนนที่ถูกประเมินโดยผู้เขียนข้อความในเว็บไซต์ต้นทาง
Web	ชื่อเว็บไซต์ของข้อความ
Country	ประเทศที่มหาวิทยาลัยตั้งอยู่
Date	วันที่ข้อความถูกเขียน
ตัวแปรจากการวิเคราะห์ความรู้สึก	
ScoreGG_Doc	คะแนนความรู้สึกของข้อความ มีค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก)
ScoreGG_Sentence	คะแนนความรู้สึกของประโยค มีค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก)
Pos	ตัวแปรหุ่น = 1 เมื่อ ScoreGG_Sentence ≥ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	คำอธิบาย
Neg	ตัวแปรหุ่น = 1 เมื่อ ScoreGG_Sentence < 0
ตัวแปรรายมหาวิทยาลัย (จากข้อความเห็น)	
ScorePos	ค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงบวกที่มหาวิทยาลัยได้รับ มีค่า 0 ถึง 1 (ค่ายิ่งมากยิ่งบวกมาก)
ScorePos_1	ค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงบวกที่มหาวิทยาลัยได้รับ มีค่า 0 ถึง 0.7 เฉพาะกลุ่มที่มีค่าน้อยกว่า 0.7 (ค่ายิ่งมากยิ่งบวกมาก)
ScorePos_2	ค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงบวกที่มหาวิทยาลัยได้รับ มีค่า 0.7 ถึง 1 เฉพาะกลุ่มที่มีค่ามากกว่าหรือเท่ากับ 0.7 (ค่ายิ่งมากยิ่งบวกมาก)
ScoreNeg	ค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงลบที่มหาวิทยาลัยได้รับ มีค่า 0 ถึง 1 (ค่ายิ่งน้อยยิ่งลบมาก)
Px (จากแหล่งอื่น)	ร้อยละความถี่ของการกล่าวถึงปัจจัย x
Uranking_THE	อันดับของมหาวิทยาลัยใน Times Higher Education World University Rankings
Top_200	ตัวแปรหุ่น = 1 เมื่อ Uranking_THE ≤ 200
Top_500	ตัวแปรหุ่น = 1 เมื่อ 200 < Uranking_THE ≤ 500
Other	ตัวแปรหุ่น = 1 เมื่อ Top_200 = 0 และ Top_500 = 0
SS_ratio	จำนวนนักศึกษาต่อบุคลากร
Inter_PC	ร้อยละของนักศึกษาต่างชาติ
FM_PC	ร้อยละของนักศึกษาผู้หญิง
S_FTE	จำนวนนักศึกษาเต็มเวลา

ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

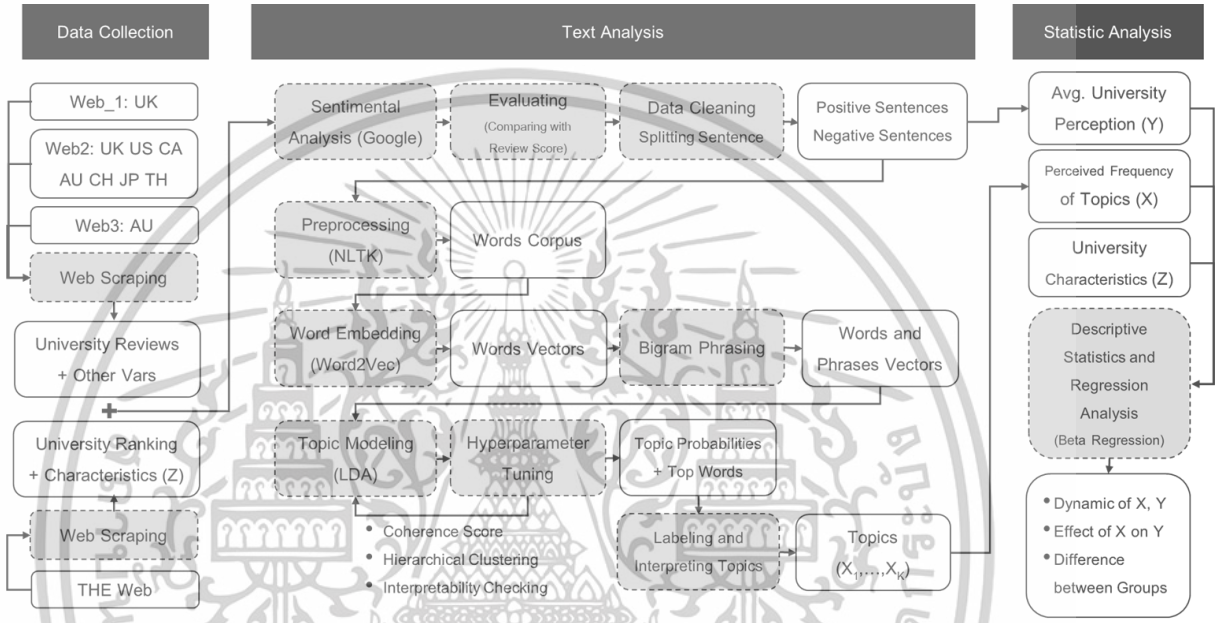
3.2 ขั้นตอนการศึกษา

3.2.1 การรวบรวมข้อมูลและภาพรวม

การศึกษานี้จะแบ่งขั้นตอนการศึกษาออกเป็น 3 ส่วน คือ

1) ขั้นตอนการรวบรวมข้อมูล (Data Collection)
2) ขั้นตอนการวิเคราะห์ข้อมูลข้อความ (Text Analysis)
3) ขั้นตอนการวิเคราะห์เชิงสถิติ (Statistical Analysis)

ขั้นตอนการดำเนินการทั้งหมดได้แสดงตามรูปต่อไปนี้



รูปที่ 3.5 ขั้นตอนการศึกษา

ที่มา: ผู้วิจัย

การศึกษานี้ใช้โปรแกรมภาษา Python คือ โปรแกรม Spyder IDE และ Google Colab ในขั้นตอนการรวบรวมข้อมูลและขั้นตอนการวิเคราะห์ข้อมูลข้อความ และใช้โปรแกรมภาษา R (Version 4.2.2) ผ่านโปรแกรม RStudio ในขั้นตอนการวิเคราะห์เชิงสถิติ

ขั้นตอนการรวบรวมข้อมูลได้ถูกอธิบายไว้แล้วในหัวข้อก่อนหน้า ซึ่งเมื่อผ่านขั้นตอนนี้จะได้ชุดตารางข้อมูลที่แต่ละหน่วยข้อมูลจะประกอบไปด้วย ข้อความเห็นต่อมหาวิทยาลัย (University Review: UReview) และตัวแปรอื่น ๆ เช่น ประเทศ ชื่อมหาวิทยาลัย วันที่เขียน

โดยในขั้นตอนนี้ ผู้วิจัยจะทำการทำความสะอาดข้อมูล โดยเฉพาะข้อมูลข้อความที่ควรจะเป็นข้อมูลขาดหาย เช่น n./a n.a n/.a n//a n/a nA na Ñ/A . .. zzz N/a Na N/,¥ n/A NA n/A Nil N/A xxx X - G H ajshfaksfgkas =hgchg n/a Xxx ให้ข้อความกลุ่มนี้กลายเป็นข้อมูลว่าง (Missing Value) และทำการแทนค่าอักขระพิเศษ เช่น : ; - + ! ด้วย . หรือ การเว้นวรรค ตามความเหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นจะเข้าสู่กระบวนการวิเคราะห์ข้อมูลข้อความเพื่อแปลงข้อมูลข้อความต่าง ๆ ให้กลายเป็นข้อมูลเชิงปริมาณ เพื่อให้พร้อมสู่ขั้นตอนการวิเคราะห์เชิงสถิติตามวัตถุประสงค์ของการศึกษา

3.2.2 ขั้นตอนการวิเคราะห์ข้อมูลข้อความ

การวิเคราะห์ข้อมูลข้อความมีวัตถุประสงค์หลัก คือ การประเมินความรู้สึกของข้อความและประเด็นที่ข้อความนั้นกำลังกล่าวถึงจากชุดข้อความความเห็นที่มี โดยธรรมชาติของข้อมูลข้อความจะเป็นข้อมูลไม่มีโครงสร้างที่มีมิติสูงมาก เพราะมนุษย์มีการใช้คำที่หลากหลาย การประมวลข้อมูลข้อความจึงจำเป็นที่จะต้องอาศัยความรู้ในการประมวลภาษาธรรมชาติ

โดยการวิเคราะห์ข้อมูลข้อความมีขั้นตอน ดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ความรู้สึก (Sentimental Analysis) รายประโยค

ในขั้นตอนนี้ผู้วิจัยจะใช้ตัวแบบที่ถูกพัฒนามาก่อนหน้า (Pre-trained Model) ของ Google คือ Google NLP ในการสร้างตัวแปรคะแนนความรู้สึก (Sentiment) โดยการนำข้อความเห็นต่อมหาวิทยาลัย (University Review: UReview) ไปสร้างเป็นคะแนนความรู้สึก Google NLP จะให้ค่าคะแนนความรู้สึกของข้อความ (Sentiment_GG) แต่ละข้อความในช่วงตั้งแต่ค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก) นอกจากนี้ Google NLP ยังให้ค่าคะแนนความรู้สึกของประโยค (ScoreGG_Sentence) แต่ละประโยคในแต่ละข้อความในช่วงตั้งแต่ค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก)

โดยผู้วิจัยจะประเมินคุณภาพของคะแนนความรู้สึก (Sentiment Score) ด้วยการวิเคราะห์ความสัมพันธ์กับ คะแนนที่ถูกประเมินโดยผู้เขียนข้อความ (Score) ผ่านการใช้รูปและการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งหาก Google NLP ไม่สามารถให้ค่าคะแนนความรู้สึกที่มีคุณภาพ การวิจัยอาจจะต้องหาวิธีการประเมินความรู้สึกอื่น เช่น การใช้ตัวแบบที่ถูกพัฒนามาก่อนหน้าอื่นนอกจาก Google NLP การใช้คะแนนที่ถูกประเมินโดยผู้เขียนข้อความมาเป็นคะแนนความรู้สึก อย่างไรก็ตามวิธีการหลังนี้อาจจะมีปัญหาในการนำการศึกษาไปใช้ประโยชน์กับข้อมูลชุดอื่นเพราะข้อมูลแต่ละแหล่งมีการออกแบบระบบการให้คะแนนที่แตกต่างกันและอาจจะมีปัญหาความเอนเอียงจากการกำหนดค่าเริ่มต้น (Default Bias) เช่น เว็บไซต์มักจะกำหนดค่าคะแนนตั้งต้นไว้ที่คะแนนเต็มทำให้ผู้ให้ข้อมูลที่ไม่วิเคราะห์อย่างถี่ถ้วนอาจจะข้ามขั้นตอนการประเมินและให้คะแนนเต็มโดยไม่สอดคล้องกับความเห็นของตนได้ ต่างจากกรณีของข้อมูลข้อความเห็นที่มักจะมีค่าเริ่มต้น คือ ค่าว่าง

จากนั้น ผู้วิจัยจะปรับตารางข้อมูลจากรายข้อความเป็นรายประโยค โดยแต่ละข้อความเห็นจะแยกเป็นประโยคเชิงบวก และประโยคเชิงลบ เนื่องจากข้อความเห็นแต่ละข้อความเห็นอาจจะกล่าวถึงประเด็นที่หลายหลาย ผู้ให้ความเห็นอาจจะกล่าวชมเชยมหาวิทยาลัยในบางเรื่องและกล่าวตำหนิมหาวิทยาลัยในเรื่องอื่น ๆ

ผู้วิจัยจะสร้างตัวแปรหุ่นข้อความเชิงบวก (Pos) มีค่า 1 เมื่อ ScoreGG_Sentence ≥ 0 และสร้างตัวแปรหุ่นข้อความเชิงลบ (Neg) มีค่า 1 เมื่อ ScoreGG_Sentence < 0 โดยแต่ละประโยคจะมีค่าคะแนนความรู้สึกเชิงบวก หรือเชิงลบ (ScorePos หรือ ScoreNeg) และเพื่อความสะดวกในขั้นตอนการวิเคราะห์ผู้วิจัยจะปรับค่าขนาดของคะแนนความรู้สึกเชิงลบ (Sentiment) จากช่วง (-1,0) เป็น (0,1) ด้วยการบวก 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 2 การเตรียมข้อมูลข้อความก่อนการประมวลผล (Preprocessing)

ในขั้นตอนนี้ผู้วิจัยจะดำเนินการนำข้อมูลมาแบ่งคำ (Tokenize) และ ทำให้คำเป็นมาตรฐาน (Standardize) โดยการแปลงให้เป็นตัวพิมพ์เล็ก กำจัดอักษรพิเศษต่าง ๆ (Remove Special Character) เช่น เครื่องหมายวรรคตอน (Punctuation) คำที่มีความยาวต่ำกว่า 3 ตัวอักษร การตัดส่วนท้ายของคำ (Stemming) การแปลงคำด้วยรายการคำในพจนานุกรม (Lemmatization) จากนั้นจึงทำคลังคำ (Word Corpus) ที่มีเฉพาะ คำนาม คำคุณศัพท์ และคำกริยา ซึ่งขั้นตอนนี้ทำได้โดยใช้เครื่องมือ Natural Language Toolkit (NLTK)

ขั้นตอนที่ 3 การแปลงคำเป็นชุดรหัสตัวเลข (Word Embedding)

ในขั้นตอนนี้ผู้วิจัยจะใช้ตัวแบบที่ถูกพัฒนามาก่อนหน้า (Pre-trained Model) ของ Google คือ Word2Vec มาแปลงคำให้กลายเป็นเวกเตอร์ เพื่อให้ได้เวกเตอร์ของคำที่มีหน้าที่เดียวกันและมีความหมายคล้ายกันมีค่าเวกเตอร์ที่มีระยะห่างใกล้กัน (รายละเอียดในหัวข้อ 2.1.3) โดยในการวิจัยนี้ผู้วิจัยได้ใช้วิธี Skip-gram with Negative Sampling ในการแปลงข้อมูลคำเป็นเวกเตอร์

นอกจากจะแปลงคำต่าง ๆ ในคลังคำ (Word Corpus) จากขั้นตอนก่อนหน้าให้กลายเป็นเวกเตอร์ ผู้วิจัยยังใช้ตัวแบบการวิเคราะห์คำที่อยู่ติดกันเป็นคู่ (Bigram) เพื่อดึงกลุ่มคำที่ปรากฏเป็นคู่เพิ่มเข้าไปในคลังคำ เพื่อไม่ให้ตกหล่นกลุ่มคำที่มีหน้าที่ในการอธิบายหัวข้อ เช่น คำว่า Overall_Experience, Study_Space, Home_Country ที่มักปรากฏเป็นคู่กัน ขั้นตอนนี้ทำได้โดยการใช้ชุดคำสั่ง Phrases ใน Gensim

ขั้นตอนที่ 4 การสร้างตัวแบบหัวข้อ (Topic Modeling)

ในขั้นตอนนี้ผู้วิจัยจะสร้างตัวแบบหัวข้อ (Topic Modeling) ด้วยการทำการจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA) โดยใช้คลังคำและวลีที่ถูกแปลงเป็นเวกเตอร์แล้ว ที่ได้จากขั้นตอนก่อนหน้า (รายละเอียดในหัวข้อ 2.2.1) ในขั้นตอนนี้ผู้วิจัยจะใช้คำสั่ง LdaMulticore ของ Gensim ในการวิเคราะห์ ซึ่งการวิเคราะห์นี้จะให้

1) การแจกแจงหัวข้อของแต่ละประโยค (Topic Distribution) ซึ่งแสดงความน่าจะเป็นที่ประโยคแต่ละประโยค กำลังหมายถึงหัวข้อใด ๆ
2) คำหลัก (Top Words) ของแต่ละหัวข้อ ที่ใช้อธิบายความน่าจะเป็นที่ประโยคนั้นกำลังหมายถึงหัวข้อใด ๆ

ในการจัดสรรของดีรีเคลแฝง ผู้วิจัยจะต้องปรับค่าไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning) สำคัญคือ จำนวนหัวข้อทั้งหมด (K) ซึ่งในกรณีนี้ผู้วิจัยได้ทำการทดสอบกรณี K = 7-20 โดยการ

1) ทำการจัดสรรของดีรีเคลแฝง ณ ค่า K ต่าง ๆ จากนั้นจึงเก็บผลลัพธ์ คือ การแจกแจงหัวข้อของแต่ละประโยค และคำหลัก ของแต่ละกรณี
2) คำนวณค่าคะแนนความเชื่อมโยงแบบ Pointwise Mutual Information (PMI) และ ค่าคะแนนความเชื่อมโยงแบบ University of Massachusetts (UMass) ของแต่ละ K และพิจารณาจำนวน K ที่เหมาะสมสูงสุด (K*) ของแต่ละค่าคะแนนความเชื่อมโยง
3) พิจารณาความสามารถในการตีความ (Interpretability) ของกรณี K*±2 โดยการพิจารณาจากคำหลักว่าสามารถตีความหัวข้อของแต่ละกลุ่มได้หรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) วิเคราะห์การแบ่งกลุ่มแบบมีขั้นตอน (Hierarchical Cluster Analysis) เพื่อพิจารณาว่าในแต่ละ K* สามารถยุบรวมหัวข้อลงได้หรือไม่ โดยการแปลงกลุ่มคำหลักให้เป็นเวกเตอร์เอกสารด้วยคำสั่ง Doc2Vec ใน Gensim และใช้การวัดค่าระยะห่างโคไซน์ในการคำนวณ จากนั้นจึงตรวจสอบความสามารถในการตีความของแต่ละกลุ่มหัวข้อที่มีระยะใกล้กันว่าสมควรจะถูกยุบรวมหรือไม่

การวิเคราะห์ข้างต้นจะทำจนกว่าจะได้จำนวน K ที่เหมาะสม สามารถตีความได้ว่าแต่ละหัวข้อหมายถึงอะไร และแต่ละหัวข้อมีเนื้อหาที่แยกจากกันได้

ขั้นตอนที่ 5 การระบุและตีความหัวข้อ (Labeling and Interpreting Topics)

หลังจากที่ทราบค่าจำนานหัวข้อที่เหมาะสมและสรุปได้แล้วว่าแต่ละกลุ่มคำหลักหมายถึงประเด็นอะไร จากนั้นผู้วิจัยจะพิจารณาว่าแต่ละประโยคกำลังหมายถึงประเด็นใดบ้าง โดยการพิจารณาจากการแจกแจงหัวข้อของแต่ละประโยค (Topic Distribution) ซึ่งในกรณีนี้จะเป็นการสร้างตัวแปรหุ่น = 1 เมื่อ ประโยคนั้นมีความน่าจะเป็นที่จะหมายถึงหัวข้อนั้นมากกว่า 1/K ซึ่งในแต่ละประโยคอาจจะหมายถึงมากกว่า 1 หัวข้อได้

3.2.3 ขั้นตอนการวิเคราะห์เชิงสถิติ

ในขั้นตอนก่อนหน้าเราทราบว่ามีปัจจัยอะไรบ้างที่ถูกกล่าวถึงบ่อยเมื่อผู้คนแสดงความเห็นต่อมหาวิทยาลัย ในขั้นตอนนี้จะเป็นการศึกษาว่าปัจจัยเหล่านั้นมีปัจจัยใดบ้างที่ส่งผลอย่างมีนัยสำคัญทางสถิติต่อการเปลี่ยนแปลงของคะแนนความรู้สึกในแง่บวกและแง่ลบที่ผู้คนมีต่อมหาวิทยาลัย

ขั้นตอนที่ 1 การแปลงข้อมูลรายข้อความเป็นรายมหาวิทยาลัย

ในขั้นตอนนี้ผู้วิจัยจะสร้างชุดตัวแปรเป็นรายมหาวิทยาลัย จากเดิมที่ชุดข้อมูลแทนค่าสังเกตุรายข้อความ โดยผู้วิจัยทำการหาค่าเฉลี่ยตัวแปรรายข้อความทั้งหมดจะได้ตัวแปรค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงบวกที่มหาวิทยาลัยได้รับ ค่าเฉลี่ยคะแนนความรู้สึกของข้อความเชิงลบที่มหาวิทยาลัยได้รับและร้อยละความถี่ของการกล่าวถึงหัวข้อต่าง ๆ

สาเหตุที่การวิเคราะห์เชิงสถิติใช้ข้อมูลรายมหาวิทยาลัยแทนรายข้อความ เพราะเป็นธรรมชาติของข้อมูลข้อความที่จะมีมิติไม่ครบในชุดข้อมูลเดียว เช่น ข้อความอาจจะชมเรื่องสถานที่การที่ข้อความไม่ได้กล่าวถึงปัจจัยอื่น ๆ เช่น ชั้นเรียน บุคลากร เราไม่สามารถสรุปได้ว่าหน่วยของข้อมูลให้ค่ากับมิติเหล่านี้อย่างไร

ขั้นตอนที่ 2 การวิเคราะห์การกระจายและพลวัตของข้อมูล

ในขั้นตอนนี้ผู้วิจัยจะนำตัวแปรที่มีทั้งในระดับข้อความ และระดับมหาวิทยาลัย โดยเริ่มจากการศึกษาว่าข้อมูลแต่ละชุดกลุ่มย่อย (แบ่งตาม เชิงบวกเชิงลบ ก่อนและหลัง ค.ศ. 2019 และประเทศต่าง ๆ) มีการกล่าวถึงประเด็นที่คล้ายกันหรือไม่ รวมถึงระดับความบ่อยในการพูดถึงประเด็นต่าง ๆ ที่แต่ละประเทศอาจจะมีไม่เท่ากัน เช่น มหาวิทยาลัยในสหราชอาณาจักรอาจจะมีคำชมด้านชั้นเรียนน้อยกว่ามหาวิทยาลัยในสหรัฐอเมริกาเพราะสหราชอาณาจักรไม่มีการจัดให้มีชั้นเรียน (Coursework) ในหลายหลักสูตร

นอกจากนี้ ยังสามารถพิจารณาความถี่ที่เปลี่ยนแปลงไปข้ามช่วงเวลาเพื่อศึกษาว่าในแต่ละช่วงเวลาผู้คนให้ความสำคัญกับประเด็นด้านไหนอย่างไรบ้าง รวมทั้งศึกษาทิศทางของประเด็นที่ถูกพูดถึงมากขึ้นหรือน้อยลงเมื่อเวลาผ่านไป เช่น กรณีหลังปี ค.ศ. 2019 การกล่าวชมในด้านสถานที่น่าจะมีลดลงเพราะการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 3 การประมาณการถดถอย

ในขั้นตอนนี้ผู้วิจัยจะประมาณการถดถอยบีตา โดยมีตัวแปรตาม คือ ค่าเฉลี่ยของคะแนนความรู้สึกต่อมหาวิทยาลัย ตัวแปรต้น คือ ตัวแปรร้อยละความถี่ของการกล่าวถึงปัจจัย x ของมหาวิทยาลัย และมีตัวแปรคุณลักษณะของมหาวิทยาลัยเป็นตัวแปรควบคุม สาเหตุที่ใช้การถดถอยบีตา เพราะ ตัวแปรตามเป็นตัวแปรที่มีค่าอยู่ในช่วงจำกัด 0 ถึง 1 (รายละเอียดในหัวข้อ 2.3)

ผู้วิจัยจะเริ่มต้นจากการตรวจสอบการแจกแจงของตัวแปรตาม คือ ค่าเฉลี่ยคะแนนความรู้สึกทั้งเชิงบวกและเชิงลบ เพื่อทดสอบความเหมาะสมของตัวแบบ โดยใช้แผนภาพการแจกแจงและการทดสอบค่าสถิติ Kolmogorov–Smirnov test ทั้งนี้ หากพบว่าข้อมูลไม่ได้มีการกระจายแบบบีตา จะมีการพิจารณาแปลงข้อมูล เลือกกลุ่มข้อมูลหรือปรับตัวแบบอื่นที่เหมาะสม

เมื่อประการการถดถอยทดสอบปัญหาความสัมพันธ์เชิงเส้นแบบพหุ โดยการคำนวณค่า Variance Inflation Factor (VIF) ซึ่งหากไม่พบปัญหาความสัมพันธ์เชิงเส้นแบบพหุจะใช้วิธีการประมาณการภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) แต่หากพบปัญหาความสัมพันธ์เชิงเส้นแบบพหุจะใช้วิธีประมาณพารามิเตอร์การถดถอยบีตาแบบพารามิเตอร์ 2 ตัว (Two-Parameter Estimator for Beta Regression: TPBR) ตามที่เสนอใน Abonazel *et al.* (2022) หรือการประมาณการด้วยการถดถอยบีตาแบบริดจ์ (Ridge Beta Regression: RBR)

โดยจะประมาณการสมการแยกชุดกลุ่มย่อยของข้อมูล ผลที่ได้จะทำให้ทราบได้ว่าปัจจัยใดบ้างที่ส่งผลอย่างมีนัยสำคัญทางสถิติต่อความรู้สึกในแง่บวกและแง่ลบที่ผู้คนมีต่อมหาวิทยาลัย รวมทั้ง เทียบผลการประมาณการของแต่ละสมการ เพื่อศึกษาความแตกต่างของสมการสหราชอาณาจักร (UK) สมการสหรัฐอเมริกา (US) สมการออสเตรเลีย (AUS) และประเทศอื่น ๆ ซึ่งจะสะท้อนการให้ค่าปัจจัยต่าง ๆ ที่แตกต่างกันในแต่ละพื้นที่

ขั้นตอนที่ 4 อภิปรายผลการศึกษา

ในขั้นตอนนี้ผู้วิจัยจะสรุปผลการศึกษาและตอบวัตถุประสงค์การวิจัยที่กำหนดไว้ รวมถึงอภิปรายผลการศึกษาโดยเทียบเคียงกับการศึกษาก่อนหน้า สรุปข้อค้นพบใหม่และแนวทางการพัฒนาและการใช้ประโยชน์จากการศึกษานี้ในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

4.1 การวิเคราะห์ความรู้สึกที่มีต่อมหาวิทยาลัย

ผู้วิจัยใช้ข้อมูลความเห็นต่อมหาวิทยาลัยในประเทศ สหราชอาณาจักร สหรัฐอเมริกา ออสเตรเลีย แคนาดา จีน ญี่ปุ่นและไทย จาก WhatUNI.com (ปี ค.ศ. 2013-2022), UniReview.com.au (ปี ค.ศ. 2014-2023) และ lagora.com (ปี ค.ศ. 2000-2023) ไม่รวมความเห็นที่ว่างเปล่า ทั้งที่ไม่มีข้อความ และพิมพ์ข้อความในลักษณะไม่ให้ความเห็น เช่น N/A, .., -, na, nan รวมทั้งสิ้น 126,140 ข้อความ และตัดชุดข้อมูลที่มีข้อความที่ซ้ำ ๆ กัน (Repetitious Message) ซึ่งมักจะเป็นคำหรือประโยคสั้น ๆ เช่น Good, Bad, Interesting, No complaints ออก คงเหลือชุดข้อมูล 119,806 ข้อความ ข้อมูลคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์ (Uscore) และความยาวของข้อความ (Length) ตามตารางที่ 4.1

ตารางที่ 4.1 ข้อมูลคะแนนที่ผู้เขียนข้อความประเมินและความยาวของข้อความ

ตัวแปร/ที่มา	ประเทศ [6]	จำนวน	ค่าเฉลี่ย	SD	ค่าต่ำสุด	P50	P75	ค่าสูงสุด
Uscore [1]		118,129	4.2	0.9	0.0	4.0	4.0	5.0
WhatUNI [3]	UK	100,386	4.2	0.9	1.0	4.0	4.0	5.0
lagora.com [4, 5]	AU	1,468	4.2	1.0	0.0	3.8	4.5	5.0
lagora.com	TH	300	4.0	1.1	1.0	3.8	3.8	5.0
lagora.com	UK	8,069	4.0	0.9	0.0	3.8	3.8	5.0
lagora.com	US	4,790	4.2	0.9	0.0	3.8	4.3	5.0
lagora.com	JP	489	4.0	1.1	0.0	3.8	3.9	5.0
lagora.com	CN	655	4.0	0.9	0.0	3.8	3.8	5.0
lagora.com	CA	1,972	4.3	0.8	0.0	3.8	4.3	5.0
Length [2]		119,806	33.53	32.63	1	12	31	43
WhatUNI	UK	100,386	33.61	28.23	1	12	31	44
UniReview	AU	1,677	116.80	134.53	2	33	71	149
lagora.com	AU	1,468	24.62	15.43	1	11	23	37
lagora.com	TH	300	26.96	15.48	2	12	28	41
lagora.com	UK	8,069	25.07	15.57	1	11	24	38
lagora.com	US	4,790	24.78	15.71	1	11	22	38
lagora.com	JP	489	31.51	16.47	1	17	34	47
lagora.com	CN	655	26.63	15.81	2	12	26	41
lagora.com	CA	1,972	25.06	15.35	1	11	24	38

ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายเหตุ:

1. Uscore คือ ข้อมูลคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์
2. Length คือ ความยาวของข้อความ
3. WhatUNI.com คะแนนในช่วง 1, 2, 3, 4, 5
4. UniReview.com.au ไม่มีการให้คะแนน
5. lagora.com คะแนนในช่วง [0,5]
6. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น TH คือ ไทย

จากข้อมูลข้างต้นจะพบว่าข้อความเห็นในเว็บไซต์ lagora.com ซึ่งมีความเห็นต่อมหาวิทยาลัยในหลายประเทศ ข้อความส่วนมากจะเป็นข้อความแสดงความเห็นต่อมหาวิทยาลัยในสหราชอาณาจักร สหรัฐอเมริกา แคนาดาและออสเตรเลีย ซึ่งเป็นประเทศที่มีมหาวิทยาลัยที่เป็นที่นิยมของนักศึกษานานาชาติ ขณะที่ข้อความเห็นต่อมหาวิทยาลัยในไทย จีนและญี่ปุ่นจะปรากฏอยู่น้อย

ข้อความเห็นจากเว็บไซต์ WhatUNI.com จะเป็นข้อความเห็นที่มีจำนวนมากกว่าเว็บอื่น ๆ ขณะที่ข้อความเห็นจาก UniReview.com.au จะมีความยาวกว่าเว็บอื่น ๆ มาก ซึ่งข้อความเห็นจากเว็บไซต์ WhatUNI.com และ lagora.com มีความยาวใกล้เคียงกัน จึงเป็นที่น่าสังเกตุว่าลักษณะการออกแบบเว็บไซต์น่าจะส่งผลต่อพฤติกรรมการให้ความเห็นต่อมหาวิทยาลัยด้วย

ผู้วิจัยได้นำข้อมูลข้อความไปประเมินเป็นคะแนนความรู้สึกต่อมหาวิทยาลัย (ScoreGG_Doc) โดยการใช้โปรแกรม Google NLP คะแนนที่ได้มีค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก) โปรแกรมให้ผลการประเมิน 111,640 ข้อความ มีข้อความ 8,166 ที่ไม่สามารถประเมินผลได้ คิดเป็นร้อยละ 6.8% โดยส่วนหนึ่งเป็นข้อความที่เป็นคำขนาดสั้น ๆ ไม่เป็นประโยค จึงละออกจากวิเคราะห์ต่อจากนี้ ได้ผลการประเมินสรุปตามตารางที่ 4.2

ตารางที่ 4.2 สถิติเชิงพรรณนาของคะแนนความรู้สึกต่อมหาวิทยาลัย (ScoreGG_Doc)

ตัวแปร/ที่มา	ประเทศ [2]	จำนวน	ค่าเฉลี่ย	SD	ค่าต่ำสุด	P50	P75	ค่าสูงสุด
ScoreGG_Doc [1]		111,640	0.49	0.49	-0.90	0.70	0.90	0.90
WhatUNI	UK	94,665	0.54	0.48	-0.90	0.80	0.90	0.90
UniReview	AU	1,487	-0.22	0.44	-0.90	-0.30	0.10	0.90
lagora.com	AU	1,272	0.29	0.46	-0.80	0.30	0.70	0.90
lagora.com	TH	264	0.11	0.46	-0.80	0.20	0.40	0.90
lagora.com	UK	7,118	0.26	0.47	-0.90	0.30	0.70	0.90
lagora.com	US	4,172	0.28	0.46	-0.90	0.30	0.60	0.90
lagora.com	JP	414	0.16	0.47	-0.90	0.20	0.50	0.90
lagora.com	CN	558	0.23	0.44	-0.80	0.30	0.60	0.90
lagora.com	CA	1,690	0.31	0.44	-0.90	0.40	0.70	0.90

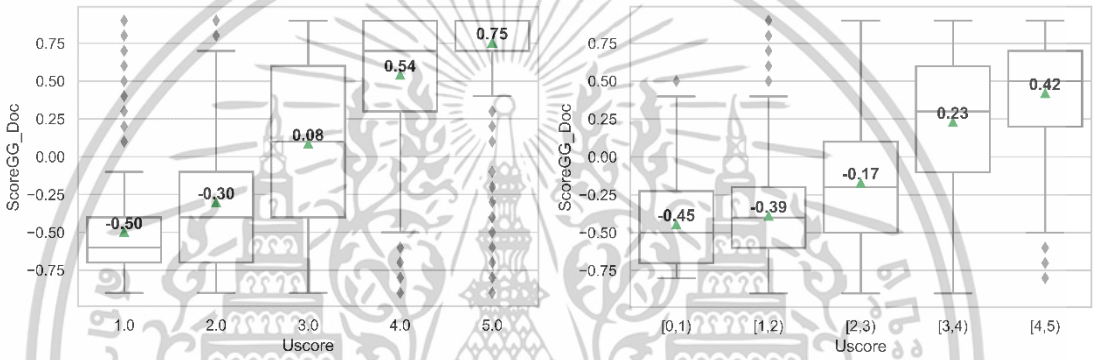
ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายเหตุ:

1. ScoreGG_Doc คือ คะแนนความรู้สึกของข้อความ มีค่า -1 (เชิงลบ) ถึง 1 (เชิงบวก)
2. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น TH คือ ไทย

ทั้งนี้ เพื่อประเมินคุณภาพของผลการวิเคราะห์คะแนนความรู้สึก ผู้วิจัยได้เทียบเคียงคะแนนความรู้สึกต่อมหาวิทยาลัย (ScoreGG_Doc) กับคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์ (Uscore) และทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ โดยหากตัวแบบ Google NLP สามารถประเมินค่าความรู้สึกได้ดี ผลคะแนนความรู้สึกควรมีความสัมพันธ์ในทิศทางเดียวกันกับคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์ ซึ่งพิจารณาได้จากรูปที่ 4.1 และตารางค่าสัมประสิทธิสหสัมพันธ์ในตารางที่ 4.3



ก. กรณี WhatUNI.com ข. กรณี lagora.com

รูปที่ 4.1 ค่าเฉลี่ยคะแนนความรู้สึกในแต่ละช่วงคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์

ที่มา: ผู้วิจัย

หมายเหตุ:

1. Uscore คือ คะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์
2. ScoreGG_Doc คือ คะแนนความรู้สึกต่อมหาวิทยาลัย
3. WhatUNI.com คะแนนในช่วง 1, 2, 3, 4, 5
4. lagora.com คะแนนในช่วง (0,5)

ตารางที่ 4.3 ค่าสัมประสิทธิสหสัมพันธ์ตัวแปรคะแนนความรู้สึกและคะแนนที่ผู้เขียนประเมิน

ตัวแปร 1	ตัวแปร 2	ที่มาของข้อมูล	ค่าสัมประสิทธิสหสัมพันธ์ Pearson (ρ)
Uscore	ScoreGG_Doc	WhatUNI	0.577 ***
Uscore	ScoreGG_Doc	lagora.com	0.430 ***

ที่มา: ผู้วิจัย

หมายเหตุ:

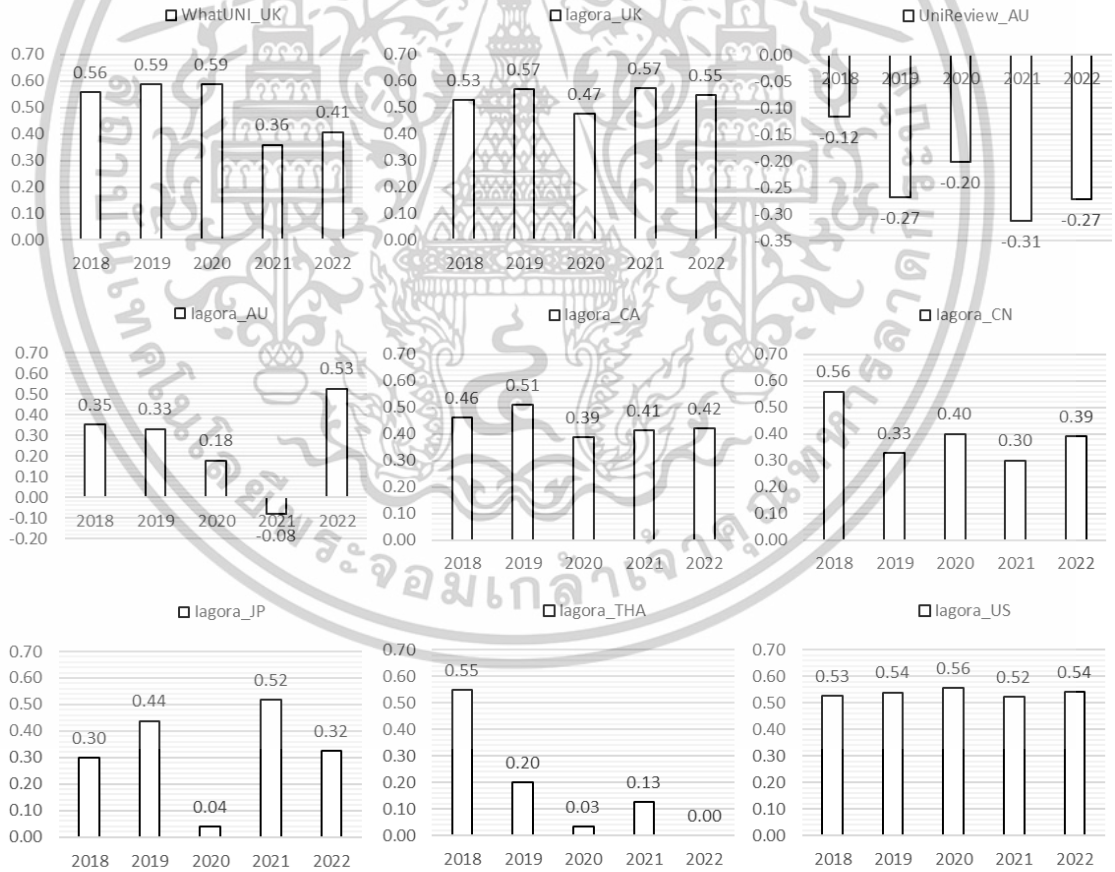
1. Uscore คือ ข้อมูลคะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ScoreGG_Doc คือ คะแนนความรู้สึกต่อมหาวิทยาลัย
3. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

จากรูปที่ 4.1 พบว่าผลการวิเคราะห์ค่าคะแนนความรู้สึกของข้อความเห็นต่อมหาวิทยาลัยโดย Google NLP ให้ผลดีพอสมควร เนื่องจากข้อความที่ผู้เขียนมีการให้คะแนนประเมินยิ่งสูง ค่าคะแนนความรู้สึกของข้อความเห็นต่อมหาวิทยาลัยจะยิ่งสูงตาม (ในการศึกษานี้ ผู้เขียนได้ทดลองการวิเคราะห์ความรู้สึก (Sentimental Analysis) ด้วยตัวแบบอื่นนอกเหนือจาก Google NLP โดยใช้ VADER Sentiment ซึ่งพบว่า ค่าคะแนนความรู้สึกของข้อความเห็นต่อมหาวิทยาลัยจาก Google NLP มีความสอดคล้องกันกับ คะแนนที่ผู้เขียนข้อความประเมินในเว็บไซต์ มากกว่าคะแนนจาก VADER Sentiment เมื่อพิจารณาจากค่าสัมประสิทธิสหสัมพันธ์)

ขณะที่ในตารางที่ 4.3 แสดงผลการทดสอบค่าสัมประสิทธิสหสัมพันธ์ (ρ) โดยมีสมมติฐานหลัก คือ H_0: $\rho = 0$ ผลการทดสอบพบว่าปฏิเสธสมมติฐานหลัก คะแนนความรู้สึกและคะแนนที่ผู้เขียนประเมินมีค่าสัมประสิทธิสหสัมพันธ์เป็นบวกและมีนัยสำคัญทางสถิติ ณ ระดับนัยสำคัญ 0.01



รูปที่ 4.2 ค่าเฉลี่ยคะแนนความรู้สึกในแต่ละปี

ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายเหตุ:

1. กราฟแท่งแสดงค่าเฉลี่ยคะแนนความรู้สึกของข้อความต่อมหาวิทยาลัย
2. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น TH คือ ไทย

จากรูปที่ 4.2 พบว่าค่าเฉลี่ยคะแนนความรู้สึกต่อมหาวิทยาลัยของมีความแตกต่างกันไปในแต่ละประเทศและแต่ละเว็บไซต์ เว็บไซต์ UniReview.com.au เป็นเว็บไซต์ที่มีข้อความเห็นเชิงลบต่อมหาวิทยาลัยในออสเตรเลียจำนวนมากเมื่อเทียบกับอีก 2 แหล่งที่มักจะมีข้อความเชิงบวกมากกว่า อย่างไรก็ตามเมื่อเทียบกับเว็บไซต์ lagora.com กรณีข้อความเห็นต่อมหาวิทยาลัยในประเทศออสเตรเลีย จะพบว่าค่าคะแนนความรู้สึกต่อมหาวิทยาลัยมีแนวโน้มขึ้นลงสอดคล้องกัน

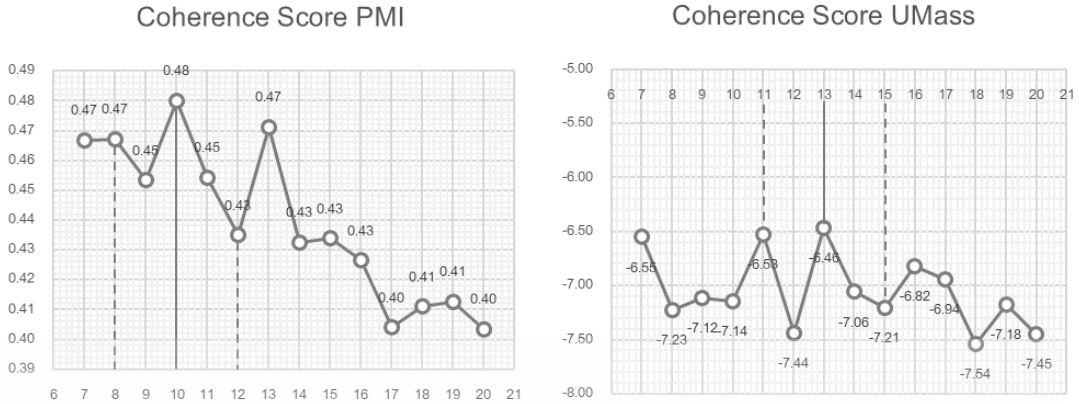
ทั้งนี้ โดยภาพรวมไม่พบว่าค่าเฉลี่ยคะแนนมีแนวโน้มเปลี่ยนแปลงไปในทิศทางใดทิศทางหนึ่งอย่างชัดเจน ยกเว้นกรณีประเทศไทยที่มีแนวโน้มลดลงในช่วงหลังปี ค.ศ. 2018 เป็นต้นมา จึงเป็นที่น่าสนใจว่ามหาวิทยาลัยในแต่ละประเทศถูกกล่าวถึงในประเด็นหัวข้อใดบ้าง และประเด็นหัวข้อใดบ้างใดบ้างที่ส่งผลต่อการเปลี่ยนแปลงของคะแนนความรู้สึกต่อมหาวิทยาลัย ซึ่งจำเป็นต้องใช้การทำตัวแบบหัวข้อ (Topic Modeling) ในหัวข้อต่อไปในการศึกษา

4.2 การวิเคราะห์ประเด็นที่มหาวิทยาลัยถูกกล่าวถึง

ในหัวข้อนี้ผู้วิจัยได้สกัดปัจจัยที่มหาวิทยาลัยถูกกล่าวถึงจากข้อความความเห็นที่มี โดยใช้ตัวแบบหัวข้อ (Topic Modeling) ด้วยการทำการจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA) ซึ่งใช้การเตรียมข้อมูลข้อความก่อนการประมวลผลด้วยชุดคำสั่ง Natural Language Toolkit (NLTK) จากนั้นจึงแปลงคำเป็นชุดรหัสตัวเลข (Word Embedding) ด้วย Word2Vec วิธี Skip-gram with Negative Sampling ร่วมกับการเพิ่มคำที่อยู่ติดกันเป็นคู่ (Bigram) ไปในคลังคำศัพท์ที่นำมาใช้วิเคราะห์ (รายละเอียดในส่วนก่อนหน้า)

ผลการจัดสรรของดีรีเคลแฝง ผู้วิจัยปรับค่าไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning) สำคัญ คือ จำนวนหัวข้อทั้งหมด (K) โดยการพิจารณาจากค่าคะแนนความเชื่อมโยงแบบ Pointwise Mutual Information (PMI) และค่าคะแนนความเชื่อมโยงแบบ University of Massachusetts (UMass) ของแต่ละค่า K ในช่วง 7 ถึง 20 ได้ค่าคะแนนความเชื่อมโยง ตามที่แสดงในรูปที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 ค่าคะแนนความเชื่อมโยงจากการทำการจัดสรรของดีรีเคลแฝง

ที่มา: ผู้วิจัย

โดยค่าคะแนนความเชื่อมโยงแบบ PMI และ UMass ระบุจำนวนหัวข้อที่เหมาะสม เท่ากับ 10 และ 13 หัวข้อตามลำดับ ในขั้นตอนต่อมาผู้วิจัยจึงพิจารณาความสามารถในการตีความ (Interpretability) ของกรณี K*±2 คือ K ในช่วง 8 ถึง 15 โดยการพิจารณาจากคำหลักว่าสามารถตีความหัวข้อของแต่ละกลุ่มได้หรือไม่ และ ทำการวิเคราะห์การแบ่งกลุ่มแบบมีขั้นตอน (Hierarchical Cluster Analysis) เพื่อดูความใกล้เคียงกันของแต่ละหัวข้อ ตรวจสอบความสามารถในการตีความของแต่ละกลุ่มหัวข้อที่มีระยะใกล้กันว่าสมควรจะถูกยุบรวมหรือไม่

พบว่า ค่า K ที่เหมาะสม คือ K = 9 ได้คำหลัก การตีความและผลการวิเคราะห์การแบ่งกลุ่มแบบมีขั้นตอน ตามตารางที่ 4.3 และรูปที่ 4.4

ตารางที่ 4.3 ตารางคำหลักของแต่ละหัวข้อจากการทำการจัดสรรของดีรีเคลแฝง

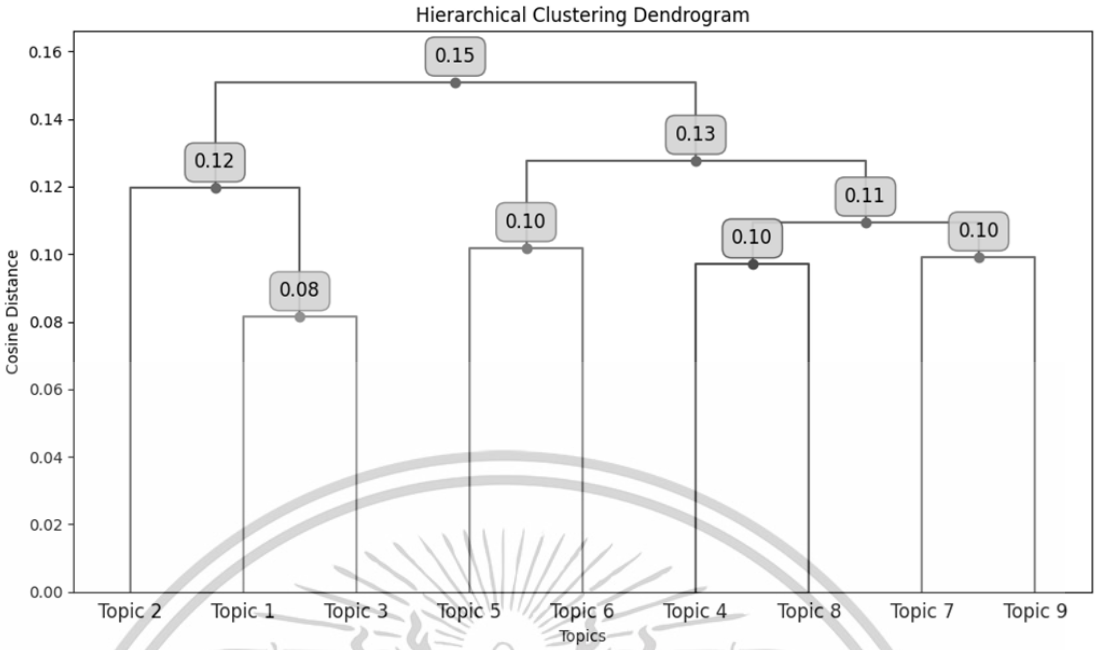
หัวข้อ (Topic) และการตีความ	คำหลัก (Top Words)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน (Overall Experience and Learning)	experience, course, life, study, enjoy, time, people, friend, year, work, opportunity, help, support, learn, overall_experience, thing, excellent, much, feel, overall
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย (Facilities and Campus Life)	facility, course, campus, people, city, union, accommodation, place, nice, society, everything, amaze, lot, many, location, lecture, support, club, thing, night
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน (Academic Challenges and Workload)	time, course, work, year, lecture, much, support, hard, difficult, feel, study, help, module, week, teach, thing, take, make, think, improve

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ (Topic) และการตีความ	คำหลัก (Top Words)
หัวข้อ 4 การใช้ชีวิตและกิจกรรม (Student Life and Social Activities)	year, first_year, second_year, travel, poor, hour, hall, miss, home, fine, live_campus, country, room, food, exchange, accomodation, house, park, world, party
หัวข้อ 5 รูปแบบการสอนและบุคลากร (Teaching System and Staff)	issue, staff, lecturer, expect, swansea, care, assignment, communication, better, program, help, feel, support, feedback, choose, others, service, situation, contact, department
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต (Expense and Cost of Living)	expensive, cheap, overall, nice_people, london, school, living, accommodation_expensive, cafe, convenient, place, everything_expensive, study_space, small_class, street, england, amazing_experience, home_country, size, residence
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย (Infrastructure and Resources)	library, decent, class, tutor, space, book, improvement, prospect, organisation, course_structure, ready, complain, satisfy, train, science, classroom, cardiff, library_facility, drink, excellent
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ (Reputation and Internationality)	problem, english, local, average, teacher, unit, none, reputation, speak_english, contact_hour, difficulty, international, call, business, disappointed, language, psychology, chinese, understanding, rubbish
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม (Language and Cultural Experience)	international, housing, language, professor, bad, visit, live, please, close, manchester, city, make_lot, freedom, distance, market, price, locate, rent, dorm, exchange_rate

ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 แผนภาพเดนโดแกรมผลการแบ่งกลุ่มแบบมีขั้นตอน

ที่มา: ผู้วิจัย

หมายเหตุ:

1. แปลงกลุ่มคำหลักเป็นเวกเตอร์ด้วย Doc2Vec ใน Gensim
2. คำนวณโดยการวัดระยะห่างโคไซน์ (Cosine Distance)

จากการวิเคราะห์ความสามารถในการตีความ พบว่าจำนวนหัวข้อเท่ากับ 9 เป็นจำนวนที่เหมาะสมเพราะแต่ละหัวข้อให้คำหลักที่สามารถตีความได้และแยกประเด็นออกจากกันได้ โดยในรูปที่ 4.4 แสดงให้เห็นว่าหัวข้อที่ 1 และ หัวข้อที่ 3 เป็นหัวข้อที่มีระยะของเวกเตอร์คำหลักใกล้เคียงกันมากสุด เนื่องจากเป็นคำหลักที่เกี่ยวกับการเรียนทั้งคู่ แต่ทั้ง 2 นำเสนอประเด็นที่แตกต่างกัน หัวข้อที่ 1 จะกล่าวถึงการเรียนในมุมมองประสบการณ์โดยภาพรวม ขณะที่ หัวข้อที่ 3 กล่าวถึง จะกล่าวถึงการเรียนในมุมมองความท้าทายทางวิชาการและภาระงาน ทำให้ไม่ควรควบรวมทั้ง 2 กลุ่มเข้าด้วยกัน

โดยสรุปตัวแบบดีรีเคลแฝง สามารถจำแนกหัวข้อของข้อความเห็นต่อมหาวิทยาลัยออกมาเป็น 9 หัวข้อ โดยในแต่ละหัวข้อกล่าวถึงประเด็นดังต่อไปนี้

- **หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน (Overall Experience and Learning)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับประสบการณ์โดยรวมของการเป็นนักศึกษา รวมถึงการเรียนรู้ การศึกษา และการเจริญเติบโตของตัวเอง
- **หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย (Facilities and Campus Life)** หัวข้อนี้แทนการสนทนาเกี่ยวกับสิ่งอำนวยความสะดวก สถานที่ วิทยาเขต และสภาพแวดล้อมโดยรวม รวมถึงชีวิตในมหาวิทยาลัย
- **หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน (Academic Challenges and Workload)** หัวข้อนี้แทนความยากลำบากและภาระงานที่เกี่ยวข้องกับหลักสูตรวิชาการ การบรรยาย และการมอบหมายงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **หัวข้อ 4 การใช้ชีวิตและกิจกรรม (Student Life and Social Activities)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับชีวิตสังคมของนักศึกษา กิจกรรมนอกเหนือจากวิชาการ และกิจกรรมที่เกิดขึ้นทั้งในและนอกมหาวิทยาลัย
- **หัวข้อ 5 รูปแบบการสอนและบุคลากร (Teaching System and Staff)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับอาจารย์ บุคลากร คุณภาพของการสอน ผลของวิธีการสอน และระบบสนับสนุนที่มีให้แก่นักศึกษาเพื่อเสริมสร้างประสบการณ์การเรียนรู้
- **หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต (Expense and Cost of Living)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับค่าใช้จ่ายที่นักศึกษาต้องรับผิดชอบขณะที่เข้ามหาวิทยาลัย รวมถึงค่าเช่าที่พักอาศัย อาหาร การเดินทาง และค่าใช้จ่ายอื่น ๆ
- **หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย (Infrastructure and Resources)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับความพร้อมใช้งานและคุณภาพของทรัพยากรวิชาการ เช่น ห้องสมุด ห้องปฏิบัติการ ห้องเรียน และโครงสร้างพื้นฐานเทคโนโลยี
- **หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ (Reputation and Internationality)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับชื่อเสียงของมหาวิทยาลัย การยอมรับระดับนานาชาติและการสื่อสารในการเรียนการสอน
- **หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม (Language and Cultural Experience)** หัวข้อนี้แทนการสนทนาที่เกี่ยวข้องกับประสบการณ์ของนักศึกษาชาวต่างชาติ เช่น ความท้าทายทางภาษา ความหลากหลายทางวัฒนธรรม และโอกาสในการแลกเปลี่ยนทางวัฒนธรรม โดยเฉพาะการมีประสบการณ์การเรียนรู้จากสถานการณ์ทางวัฒนธรรมที่แตกต่างกัน

ตัวแบบดีรีเคลแฝงได้คำนวณการแจกแจงความน่าจะเป็นของข้อความเห็นแต่ละข้อความเห็นในหัวข้อที่ 1-9 ซึ่งทำให้คำนวณได้ว่าแต่ละข้อความเห็นต่าง ๆ นั้นกำลังหมายถึงหัวข้อใดบ้าง

4.3 การแจกแจงของประเด็นที่มหาวิทยาลัยถูกกล่าวถึง

ผู้วิจัยจะแสดงการแจกแจงของประเด็นที่มหาวิทยาลัยถูกกล่าวถึง โดยจำแนกตามความรู้สึกเชิงบวกและเชิงลบ ช่วงเวลาของข้อความ และประเทศที่มหาวิทยาลัยตั้งอยู่ เพื่อพิจารณาประเด็นที่มหาวิทยาลัยมักจะถูกชมและตำหนิ ผลของช่วงเวลาก่อนและหลังเกิดการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 และความแตกต่างของวัฒนธรรมในแต่ละประเทศตามวัตถุประสงค์ของการวิจัยนี้ ในการวิเคราะห์ต่อจากนี้ ผู้วิจัยจะปรับค่าความน่าจะเป็นของหัวข้อในแต่ละข้อความ โดยหากข้อความนั้นมีความน่าจะเป็นมากกว่า 1/K จะแสดงว่าข้อความนั้นกำลังกล่าวถึงหัวข้อนี้

ตารางที่ 4.4 แสดงข้อมูลการแจกแจงของหัวข้อของข้อความเห็นที่จำแนกตามความรู้สึกเชิงบวกเชิงลบ จะพบว่าข้อความเห็นต่อมหาวิทยาลัยในเชิงบวกส่วนมากจะกล่าวถึงประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย ความท้าทายทางวิชาการและภาระงาน การใช้ชีวิตและกิจกรรม ตามลำดับ ขณะที่เมื่อเป็นข้อความเห็นต่อมหาวิทยาลัยในเชิงลบส่วนมากจะกล่าวถึง ความท้าทายทางวิชาการและภาระงาน ประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย และรูปแบบการสอนและบุคลากร ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สะท้อนว่านอกเหนือจากประสบการณ์โดยภาพรวม ปัจจัยที่จะจะทำให้คนกล่าวถึงมหาวิทยาลัยด้วยความชื่นชมมากที่สุด คือ เรื่องสภาพแวดล้อมโดยรวมและชีวิตในมหาวิทยาลัย ขณะที่เรื่องที่ผู้คนบ่นถึงมหาวิทยาลัยมากที่สุด คือ เรื่องภาระงาน ความยากในการเรียน รวมถึงคุณภาพของการเรียนการสอน ซึ่งปัจจัยที่กล่าวมานี้น่าจะเป็นปัจจัยสำคัญที่กำหนดความรู้สึกของผู้ให้ความเห็นต่อมหาวิทยาลัย

ตารางที่ 4.4 การแจกแจงของหัวข้อของข้อความเห็น จำแนกตามความรู้สึกเชิงลบเชิงบวก

หัวข้อ	ข้อความเชิงบวก	ข้อความเชิงลบ	ผลต่าง
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน (Overall Experience and Learning)	72.3%	44.0%	-28.3%
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย (Facilities and Campus Life)	43.9%	37.4%	-6.5%
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน (Academic Challenges and Workload)	26.9%	50.1%	23.1%
หัวข้อ 4 การใช้ชีวิตและกิจกรรม (Student Life and Social Activities)	14.3%	19.7%	5.4%
หัวข้อ 5 รูปแบบการสอนและบุคลากร (Teaching System and Staff)	14.2%	23.7%	9.5%
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต (Expense and Cost of Living)	8.2%	12.8%	4.7%
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย (Infrastructure and Resources)	9.2%	13.2%	4.0%
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ (Reputation and Internationality)	6.8%	11.7%	4.9%
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม (Language and Cultural Experience)	7.5%	10.4%	3.0%
จำนวนตัวอย่าง	101,323	38,096	

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ผลรวมของแต่ละคอลัมน์สามารถจะมากกว่า 100% เพราะข้อความเห็น 1 ข้อความสามารถแสดงถึงหัวข้อมากกว่า 1 หัวข้อ
2. ผลต่าง คือ เชิงบวก – เชิงลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 การแจกแจงของหัวข้อของข้อความเห็น จำแนกตามความรู้สึกเชิงลบเชิงบวก และช่วงเวลา

	ข้อความทั้งหมด			ข้อความเชิงบวก			ข้อความเชิงลบ		
หัวข้อ	จนถึง 2019	หลัง 2019	ผลต่าง	จนถึง 2019	หลัง 2019	ผลต่าง	จนถึง 2019	หลัง 2019	ผลต่าง
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	64.4%	64.9%	0.4%	72.2%	72.3%	0.1%	43.0%	47.4%	4.3%
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	41.9%	43.0%	1.1%	43.5%	45.7%	2.2%	37.5%	36.8%	-0.7%
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	32.4%	36.7%	4.3%	26.6%	28.1%	1.5%	48.2%	56.9%	8.7%
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	16.2%	14.1%	-2.2%	14.8%	12.4%	-2.4%	20.2%	18.1%	-2.1%
หัวข้อ 5 รูปแบบการสอนและบุคลากร	15.9%	20.5%	4.7%	13.6%	16.7%	3.0%	22.0%	29.6%	7.5%
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	10.0%	7.2%	-2.8%	8.5%	6.8%	-1.7%	14.1%	8.2%	-5.9%
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	10.2%	10.5%	0.3%	9.1%	9.3%	0.2%	13.2%	13.2%	-0.1%
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	8.3%	7.7%	-0.6%	7.0%	6.2%	-0.8%	11.9%	11.1%	-0.8%
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	8.7%	6.7%	-2.0%	7.9%	5.9%	-2.0%	10.9%	8.6%	-2.3%
จำนวนตัวอย่าง	111,713	27,706		81,896	19,427		29,817	8,279	

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ผลรวมของแต่ละคอลัมน์สามารถจะมากกว่า 100% เพราะข้อความเห็น 1 ข้อความสามารถแสดงถึงหัวข้อมากกว่า 1 หัวข้อ
2. ผลต่าง คือ หลัง 2019 – ก่อน 2019

ตารางที่ 4.6 การแจกแจงของหัวข้อของข้อความเห็น (เฉพาะเชิงบวก) จำแนกตามประเทศ

หัวข้อ	UK	US	AU	CA	CN	JP	THA
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	74.3%	48.8%	55.7%	47.7%	44.2%	52.4%	45.4%
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	44.5%	37.1%	33.1%	38.7%	38.5%	45.9%	34.3%
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	26.5%	27.7%	40.3%	27.3%	31.2%	36.8%	31.0%
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	13.7%	24.1%	17.4%	22.7%	21.7%	21.1%	22.7%
หัวข้อ 5 รูปแบบการสอนและบุคลากร	13.8%	16.3%	27.5%	14.6%	24.3%	15.9%	16.7%
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	7.3%	19.9%	16.0%	17.3%	17.9%	11.4%	18.5%
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	9.1%	9.7%	10.4%	9.2%	7.3%	15.1%	7.9%
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	5.7%	18.4%	17.4%	24.5%	23.5%	17.0%	16.7%
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	6.5%	19.6%	13.6%	20.4%	20.5%	19.2%	19.4%
จำนวนตัวอย่าง	92,853	3,706	2,127	1,544	507	370	216

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ผลรวมของแต่ละคอลัมน์สามารถจะมากกว่า 100% เพราะข้อความเห็น 1 ข้อความสามารถแสดงถึงหัวข้อมากกว่า 1 หัวข้อ
2. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น TH คือ ไทย

ตารางที่ 4.7 การแจกแจงของหัวข้อของข้อความเห็น (เฉพาะเชิงลบ) จำแนกตามประเทศ

หัวข้อ	UK	US	AU	CA	CN	JP	THA
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	44%	36%	52.1%	34%	30%	38%	38%
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	39%	30%	23.1%	32%	32%	44%	30%
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	50%	40%	64.8%	36%	43%	47%	51%
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	20%	22%	14.4%	24%	20%	19%	22%
หัวข้อ 5 รูปแบบการสอนและบุคลากร	22%	19%	52.1%	21%	22%	20%	25%
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	12%	22%	11.3%	18%	21%	14%	13%
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	13%	12%	11.9%	10%	12%	15%	10%
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	10%	16%	22.2%	23%	26%	21%	29%
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	9%	19%	12.5%	19%	22%	21%	20%
จำนวนตัวอย่าง	32,341	2,098	2,053	831	321	289	163

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ผลรวมของแต่ละคอลัมน์สามารถจะมากกว่า 100% เพราะข้อความเห็น 1 ข้อความสามารถแสดงถึงหัวข้อมากกว่า 1 หัวข้อ
2. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น TH คือ ไทย

ตารางที่ 4.5 แสดงข้อมูลการแจกแจงของหัวข้อของข้อความเห็นที่จำแนกตามความรู้สึกเชิงลบเชิงบวก และช่วงเวลา จะพบว่าในช่วงก่อนและหลังปี ค.ศ. 2019 ซึ่งเกิดเหตการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ทำให้รูปแบบการเรียนการสอนเปลี่ยนไป โดยภาพรวมข้อความเห็นเชิงบวกก่อนและหลังโควิดจะพูดถึงประเด็นไม่ต่างกันมาก โดยมีการกล่าวถึง รูปแบบการสอนและบุคลากรมากขึ้น ขณะที่กล่าวถึงการใช้ชีวิตและกิจกรรม ประสบการณ์ด้านภาษาและวัฒนธรรม ลดลง

ขณะที่กรณีของข้อความเห็นเชิงลบจะพบว่ามีการกล่าวถึง ความท้าทายทางวิชาการและภาระงาน รูปแบบการสอนและบุคลากร เพิ่มขึ้นมาก และมีการกล่าวถึง ค่าใช้จ่ายและต้นทุนการใช้ชีวิต และ ประสบการณ์ด้านภาษาและวัฒนธรรม ลดลง ซึ่งสอดคล้องกับเหตุการณ์การระบาดของโรคที่ทำให้รูปแบบการสอนเปลี่ยนเป็นแบบออนไลน์มากขึ้น ทำให้รูปแบบการสอนและบุคลากรมีการเปลี่ยนแปลงมาก ขณะที่การเรียนออนไลน์อาจจะทำให้ผู้เรียนมีค่าใช้จ่ายบางอย่างที่ลดลง รวมถึงมีประสบการณ์ด้านภาษาและวัฒนธรรมที่ลดลงลดลงจากการไม่ได้พบเจอผู้คนมากเท่าเดิม

ตารางที่ 4.6 แสดงผลการพิจารณารายประเทศในข้อความที่มีความรู้สึกเชิงบวก พบว่าทุกประเทศจะเน้นกล่าวถึง ประสบการณ์โดยภาพรวมและการเรียน เป็นหลัก รองลงมาคือ สถานที่และชีวิตในมหาวิทยาลัย ความท้าทายทางวิชาการและภาระงาน กรณีออสเตรเลียเป็นประเทศเดียวที่มีการกล่าวถึง ความท้าทายทางวิชาการและภาระงาน มากกว่า สถานที่และชีวิตในมหาวิทยาลัย

หากไม่นับ 3 หัวข้อข้างต้น แต่ละประเทศก็จะมีการกระจายของหัวข้อรองลงมาที่แตกต่างกัน กรณีของสหราชอาณาจักรจะเป็นที่สังเกตว่ามีการกล่าวถึง ชื่อเสียงและความเป็นนานาชาติ ประสบการณ์ด้านภาษาและวัฒนธรรม น้อยกว่าประเทศอื่น ๆ ส่วนหนึ่งน่าจะเป็นผลมากจากการที่ที่มาของข้อมูลสหราชอาณาจักมีกรณีของมหาวิทยาลัยที่เป็นมหาวิทยาลัยท้องถิ่นจำนวนมาก ประเด็นหลักรองลงมาที่ผู้ให้ความเห็นเขียนชื่นชม กรณีสหราชอาณาจักร และกรณีสหรัฐอเมริกา กรณีญี่ปุ่น กรณีไทย คือ การใช้ชีวิตและกิจกรรม กรณีออสเตรเลีย กรณีจีน คือ รูปแบบการสอนและบุคลากร กรณีแคนาดา คือ ชื่อเสียงและความเป็นนานาชาติ และเป็นที่น่าสังเกตว่ากรณีของญี่ปุ่นที่มีการกล่าวชมถึง ทรัพยากรของมหาวิทยาลัย มากกว่าประเทศอื่น ๆ ตรงข้ามกับกรณีของไทย

ตารางที่ 4.7 แสดงผลการพิจารณารายประเทศในข้อความที่มีความรู้สึกเชิงลบ พบว่าแทบทุกประเทศจะเน้นกล่าวถึง ความท้าทายทางวิชาการและภาระงาน เป็นหลัก รองลงมาคือ ประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย ยกเว้น กรณีออสเตรเลียที่มีการบ่นถึงรูปแบบการสอนและบุคลากร มากกว่าประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย

ประเด็นหลักรองลงมาที่ผู้ให้ความเห็นเขียนตำหนิ กรณีสหราชอาณาจักร คือ รูปแบบการสอนและบุคลากร กรณีสหรัฐอเมริกา คือ การใช้ชีวิตและกิจกรรม รวมถึง ค่าใช้จ่ายและต้นทุนการใช้ชีวิต กรณีแคนาดา คือ การใช้ชีวิตและกิจกรรม ขณะที่กรณีของจีน ญี่ปุ่น คือ ชื่อเสียงและความเป็นนานาชาติ ขณะที่กรณีของไทย คือ ชื่อเสียงและความเป็นนานาชาติ รูปแบบการสอนและบุคลากร เป็นที่น่าสังเกตว่ากรณีของไทยจะมีการบ่นถึง ชื่อเสียงและความเป็นนานาชาติ ในสัดส่วนที่สูงกกว่าประเทศอื่น ๆ

ผลการศึกษาโดยภาพรวมพบว่า ข้อความคิดเห็นต่อมหาวิทยาลัย โดยมากจะกล่าวถึงประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย ความท้าทายทางวิชาการและภาระงาน เป็นหลักในเกือบทุกประเทศ ขณะที่หัวข้อรองลงมาจะแตกต่างกันไปในแต่ละประเทศ รวมถึงหัวข้อที่ถูกกล่าวถึงบ่อยในข้อความชื่นชมก็อาจจะถูกกล่าวถึงบ่อยในข้อความตำหนิเช่นกัน เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวสะท้อนว่าผู้ที่ให้ความเห็นในกรณีประเทศนั้นให้ความสำคัญกับหัวข้อนี้มาก และทำให้การวิเคราะห์ความสัมพันธ์ระหว่างความรู้สึกและประเด็นที่มหาวิทยาลัยถูกกล่าวถึง จึงควรแยกประมวลผลกรณีข้อความเห็นที่มีความรู้สึกเชิงบวกกับกรณีข้อความเห็นที่มีความรู้สึกเชิงลบ

4.4 การวิเคราะห์ความสัมพันธ์ระหว่างความรู้สึกและประเด็นที่มหาวิทยาลัยถูกกล่าวถึง

ในขั้นตอนนี้ผู้วิจัยจะวิเคราะห์ความสัมพันธ์ว่าประเด็นที่มหาวิทยาลัยถูกกล่าวถึงแต่ละประเด็นจะส่งผลต่อค่าคะแนนความรู้สึกที่ผู้เขียนความเห็นมีต่อมหาวิทยาลัยอย่างไร โดยจะแยกกรณีข้อความเห็นเชิงบวก และข้อความเห็นเชิงลบ เพราะทั้ง 2 ส่วนจะมีปัจจัยที่ถูกกล่าวถึงเป็นหลักที่แตกต่างกัน ดังที่แสดงในหัวข้อก่อนหน้า

การวิเคราะห์นี้ผู้วิจัยจะแปลงข้อมูลจากรายข้อความเห็น เป็นรายมหาวิทยาลัย เนื่องจากเป็นธรรมชาติของข้อมูลข้อความเห็นที่ผู้เขียนแต่ละข้อความจะเขียนแค่เฉพาะในบางประเด็นเท่านั้น ไม่ได้ให้ข้อมูลครบทุกประเด็น ทำให้เมทริกซ์ข้อมูลมีลักษณะเป็นเมทริกซ์ที่มีค่า 0 จำนวนมาก (Sparse Matrix) จึงมีความจำเป็นที่จะต้องแปลงข้อมูลให้เป็นเมทริกซ์ที่มีความหนาแน่นมากขึ้น (Dense Matrix) เพื่อความสะดวกในการวิเคราะห์ นอกจากนี้ผู้วิจัยจะละทิ้งชุดข้อมูลมหาวิทยาลัยที่มีจำนวนข้อความเห็นต่ำกว่า 5 ออก เพื่อลดผลของกรณีค่าสุดโต่ง

ผู้วิจัยวิเคราะห์ผลของประเด็นที่มหาวิทยาลัยถูกกล่าวถึงต่อค่าคะแนนความรู้สึกด้วยการใช้ตัวแบบการถดถอยบีตา โดยมีตัวแปรตาม คือ ค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) และ ค่าเฉลี่ยคะแนนความรู้สึกค่าที่มีต่อมหาวิทยาลัยเชิงลบ (ScoreNeg) ขณะที่ตัวแปรต้น คือ ร้อยละความถี่ของการกล่าวถึงหัวข้อต่าง ๆ และตัวแปรคุณลักษณะของมหาวิทยาลัย

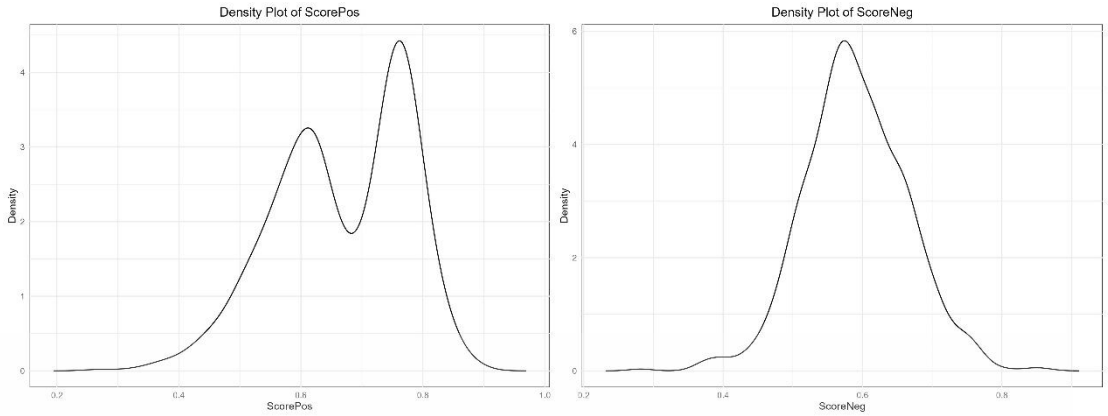
เนื่องจากตัวแปรตาม คือ ค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) มีค่า 0 ถึง 1 ค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงลบ (ScoreNeg) มีค่า -1 ถึง 0 ผู้วิจัยแปลงค่า ScoreNeg ด้วยการบวก 1 ทำให้ทั้ง ScorePos และ ScoreNeg มีค่า 0 ถึง 1 โดยค่ายิ่งมากยิ่งหมายถึงข้อความแสดงความรู้สึกที่ดีกว่า คือ รู้สึกบวกมากขึ้น หรือ รู้สึกลบลดลง

ตัวแบบการถดถอยบีตา มีสมมติฐานหลัก คือ ตัวแปรตามมีการแจกแจงแบบบีตา

$$H_0: y \sim Beta(\mu, \theta)$$

ฉะนั้น ผู้วิจัยจึงทดสอบการแจกแจงของข้อมูลตัวแปรตาม ด้วยการใช้รูปแสดงการแจกแจงและการทดสอบภาวะสารูปดี (Goodness of Fit Test) โดยจากรูปที่ 4.5 แสดงการแจกแจง จะพบว่าค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) ไม่ได้มีการกระจายที่เป็นรูปทรงคล้ายแบบการแจกแจงบีตา เพราะมีลักษณะเป็นจุดพีคสองจุด (Double Peaks) ซึ่งน่าจะเกิดจากลักษณะของมหาวิทยาลัยที่แบ่งเป็น 2 กลุ่ม คือมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก และมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก เพื่อให้สามารถจัดทำการวิเคราะห์ได้สอดคล้องกับลักษณะของข้อมูลจึงควรแยกคำชม 2 กลุ่มนี้ออกจากกัน โดยมีจุดแบ่งที่ ScorePos = 0.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก. ค่าคะแนนความรู้สึกเชิงบวก ข. ค่าคะแนนความรู้สึกเชิงลบ

รูปที่ 4.5 การแจกแจงของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัย

ที่มา: ผู้วิจัย

ตารางที่ 4.8 ผลการทดสอบภาวะสารูปดี (Goodness of Fit Test) ของตัวแปรตาม

ตัวแปร	$\hat{\alpha}_1$	$\hat{\alpha}_2$	D	P-Value
ScorePos	12.36	6.17	0.107	0.000
ScorePos < 0.7	28.13	20.24	0.072	0.016
ScorePos ≥ 0.7	104.91	31.74	0.066	0.059
ScoreNeg	25.28	17.55	0.041	0.165

ที่มา: ผู้วิจัย

หมายเหตุ:

1. สมมติฐานหลัก คือ $H_0: y \sim Beta(\mu, \theta)$
2. $\hat{\alpha}_1, \hat{\alpha}_2$ คือ ค่าพารามิเตอร์แสดงรูปทรง (Shape) การแจกแจงบีตา โดยที่ $\hat{\alpha}_1 = \hat{\mu} * \hat{\theta}$ และ $\hat{\alpha}_2 = (1 - \hat{\mu}) * \hat{\theta}$
3. ทดสอบโดยใช้ Kolmogorov–Smirnov test
4. D คือ ค่าสถิติ Kolmogorov–Smirnov

ตารางที่ 4.8 แสดงผลการทดสอบภาวะสารูปดี (Goodness of Fit Test) ของชุดข้อมูลกับการแจกแจงบีตา กำหนดสมมติฐานหลัก คือ ตัวแปรมีการแจกแจงแบบบีตา ผลการทดสอบ ณ ระดับนัยสำคัญ 0.01 พบว่า ค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) ไม่ได้มีการแจกแจงแบบบีตา แต่เมื่อแบ่งกลุ่ม มหาวิทยาลัยที่มีค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) น้อยกว่า 0.7 และมหาวิทยาลัยที่มีค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก (ScorePos) มากกว่าหรือเท่ากับ 0.7 พบว่าทั้ง 2 กลุ่มมีค่าเฉลี่ยของค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวกที่มีการแจกแจงแบบบีตา ขณะที่ค่าเฉลี่ย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คะแนนความรู้สึกค่าที่มีต่อมหาวิทยาลัยเชิงลบ (ScoreNeg) มีการแจกแจงแบบบีตาเช่นกัน (รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการในภาคผนวก ก.)

ฉะนั้น ในขั้นตอนของการวิเคราะห์การถดถอยจะแบ่งข้อมูลเป็น 3 กลุ่มใหญ่ คือ

1) กลุ่มค่าเฉลี่ยคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก น้อยกว่า 0.7 (ScorePos_1)
2) กลุ่มค่าเฉลี่ยคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก มากกว่าหรือเท่ากับ 0.7 (ScorePos_2)
3) กลุ่มค่าเฉลี่ยค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงลบ (ScoreNeg)

ข้อมูลที่ใช้ในการวิเคราะห์การถดถอยบีตามีค่าสถิติเชิงพรรณนาตามตารางที่ 4.9

ตารางที่ 4.9 สถิติเชิงพรรณนาของข้อมูลที่ใช้ในการวิเคราะห์การถดถอยบีตา

ตัวแปร	ตัวอย่าง	ค่าเฉลี่ย	SD	ต่ำสุด	สูงสุด
กรณีคะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7					
คะแนนความรู้สึกเชิงบวก	468	0.582	0.071	0.27	0.7
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	468	0.415	0.101	0	0.693
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	468	0.312	0.104	0	0.693
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	468	0.271	0.104	0	0.588
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	468	0.185	0.098	0	0.511
หัวข้อ 5 รูปแบบการสอนและบุคลากร	468	0.175	0.112	0	0.588
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	468	0.144	0.089	0	0.405
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	468	0.096	0.076	0	0.405
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	468	0.158	0.091	0	0.46
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	468	0.152	0.088	0	0.47
ช่วงเวลา	468			0	1
จนถึง ค.ศ. 2019	74			0	1
หลัง ค.ศ. 2020	394			0	1
กลุ่มอันดับมหาวิทยาลัย	468			0	1
Top 200	134			0	1
Top 500	177			0	1
อื่น ๆ	157			0	1
จำนวนนักศึกษาต่อบุคลากร	390	19.752	9.193	3.8	69
ร้อยละของนักศึกษาต่างชาติ	390	0.246	0.144	0	0.73
ร้อยละของนักศึกษาผู้หญิง	385	0.547	0.072	0.23	0.76
จำนวนนักศึกษาเต็มเวลา	390	23,713	13,078	1,015	77,468
กรณีคะแนนความรู้สึกเชิงบวก มากกว่า 0.7					
คะแนนความรู้สึกเชิงบวก	399	0.768	0.035	0.701	0.893
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	399	0.547	0.082	0.118	0.693
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	399	0.344	0.103	0	0.693
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	399	0.235	0.082	0	0.588
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	399	0.116	0.068	0	0.588
หัวข้อ 5 รูปแบบการสอนและบุคลากร	399	0.151	0.083	0	0.511

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวแปร	ตัวอย่าง	ค่าเฉลี่ย	SD	ต่ำสุด	สูงสุด
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	399	0.075	0.067	0	0.588
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	399	0.09	0.057	0	0.357
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	399	0.066	0.07	0	0.496
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	399	0.063	0.061	0	0.588
ช่วงเวลา	399			0	1
จนถึง ค.ศ. 2019	181			0	1
หลัง ค.ศ. 2020	218			0	1
กลุ่มอันดับมหาวิทยาลัย	399			0	1
Top 200	82			0	1
Top 500	98			0	1
อื่น ๆ	219			0	1
จำนวนนักศึกษาต่อบุคลากร	331	16.902	7.498	4	130.9
ร้อยละของนักศึกษาต่างชาติ	331	0.286	0.142	0.02	0.91
ร้อยละของนักศึกษาผู้หญิง	330	0.574	0.081	0.3	0.8
จำนวนนักศึกษาเต็มเวลา	331	15,516	11,191	615	77,468
กรณีคะแนนความรู้สึกเชิงลบ					
คะแนนความรู้สึกเชิงลบ	737	0.59	0.074	0.282	0.86
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	737	0.35	0.102	0	0.693
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	737	0.285	0.104	0	0.619
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	737	0.399	0.112	0	0.693
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	737	0.172	0.086	0	0.588
หัวข้อ 5 รูปแบบการสอนและบุคลากร	737	0.227	0.126	0	0.693
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	737	0.131	0.094	0	0.511
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	737	0.115	0.073	0	0.47
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	737	0.129	0.089	0	0.606
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	737	0.118	0.092	0	0.606
ช่วงเวลา	737			0	1
จนถึง ค.ศ. 2019	198			0	1
หลัง ค.ศ. 2020	539			0	1
กลุ่มอันดับมหาวิทยาลัย	737			0	1
Top 200	184			0	1
Top 500	240			0	1
อื่น ๆ	313			0	1
จำนวนนักศึกษาต่อบุคลากร	628	18.713	7.791	3.8	69
ร้อยละของนักศึกษาต่างชาติ	628	0.263	0.141	0	0.73
ร้อยละของนักศึกษาผู้หญิง	622	0.561	0.078	0.23	0.8
จำนวนนักศึกษาเต็มเวลา	628	19,789	12,676	615	77,468

ที่มา: ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงบวก

วิธีประมาณการ ตัวแปรต้น	OLS (1)	OLS (2)	OLS (3)	Beta (4)	Beta (5)	Beta (6)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.13***	0.12***	0.09**	0.54***	0.51***	0.37**
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.02	0.02	0.03	0.06	0.08	0.1
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.16***	-0.16***	-0.16***	-0.68***	-0.68***	-0.68***
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	0.06*	0.06*	0.06*	0.24*	0.27**	0.26*
หัวข้อ 5 รูปแบบการสอนและบุคลากร	-0.15***	-0.16***	-0.18***	-0.62***	-0.66***	-0.75***
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.02	0.02	0.01	0.07	0.09	0.05
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.06	-0.07*	-0.10**	-0.24	-0.30*	-0.42**
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	-0.01	-0.005	-0.01	-0.05	-0.03	-0.06
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	-0.12***	-0.11***	-0.12***	-0.49***	-0.45***	-0.51***
ตัวแปรหุ่น						
ข้อความหลัง 2019	-	0.01	0.02*	-	0.05	0.08*
THE Rank < Top 200	-	0.01	-0.001	-	0.04	-0.004
THE Rank Top 200-501	-	0.01	0.003	-	0.03	0.01
จำนวนนักศึกษาต่อบุคลากร	-	-	-0.001***	-	-	-0.005***
ร้อยละของนักศึกษาต่างชาติ	-	-	0.04*	-	-	0.17*
ร้อยละของนักศึกษาผู้หญิง	-	-	0.05	-	-	0.19
จำนวนนักศึกษาเต็มเวลา	-	-	0.0000*	-	-	0.0000*
ค่าคงที่	0.61***	0.60***	0.60***	0.43***	0.40***	0.42***
ค่า phi	-	-	-	60.82	61.29	74.91
จำนวนตัวอย่าง	468	468	385	468	468	385
R^2 / Pseudo R2	0.20	0.21	0.30	0.20	0.21	0.30
Adjusted R^2	0.19	0.19	0.27	-	-	-
AIC	-1236	-1234	-1078	-1244	-1242	-1085

ที่มา: ผู้วิจัย

หมายเหตุ:

1. OLS คือ วิธีการถดถอยกำลังสองน้อยสุด Beta คือ วิธีการถดถอยบีตา
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงบวก

วิธีประมาณการ ตัวแปรต้น	OLS (1)	OLS (2)	OLS (3)	Beta (4)	Beta (5)	Beta (6)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.20***	0.19***	0.16***	1.09***	1.04***	0.87***
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.004	0.01	0.0001	0.03	0.08	0.01
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.03*	-0.04*	-0.03	-0.18	-0.18*	-0.17
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	-0.01	-0.01	-0.03	-0.08	-0.05	-0.17
หัวข้อ 5 รูปแบบการสอนและบุคลากร	0.05**	0.05**	0.02	0.30***	0.33***	0.17
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.04	0.04*	0.02	0.23	0.26*	0.15
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.06**	-0.06**	-0.06**	-0.38**	-0.33**	-0.35**
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	0.004	0.004	0.002	0.02	0.02	0.01
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	-0.04	-0.03	-0.03	-0.21	-0.16	-0.13
ตัวแปรหุ่น						
ข้อความหลัง 2019	-	-0.01*	-0.01*	-	-0.04**	-0.03*
THE Rank < Top 200	-	-0.01***	0.002	-	-0.07***	0.01
THE Rank Top 200-501	-	-0.004	0.004	-	-0.02	0.02
จำนวนนักศึกษาต่อบุคลากร	-	-	-0.0001	-	-	-0.0005
ร้อยละของนักศึกษาต่างชาติ	-	-	-0.03***	-	-	-0.19***
ร้อยละของนักศึกษาผู้หญิง	-	-	-0.02	-	-	-0.13
จำนวนนักศึกษาเต็มเวลา	-	-	-0.0000*	-	-	-0.0000*
ค่าคงที่	0.67***	0.67***	0.72***	0.62***	0.66***	0.93***
ค่า phi	-	-	-	178.79	184.74	219.42
จำนวนตัวอย่าง	399	399	330	399	399	330
R^2 / Pseudo R2	0.24	0.26	0.22	0.24	0.26	0.21
Adjusted R^2	0.22	0.24	0.18	-	-	-
AIC	-1632	-1639	-1394	-1609	-1616	-1374

ที่มา: ผู้วิจัย

หมายเหตุ:

1. OLS คือ วิธีการถดถอยกำลังสองน้อยสุด Beta คือ วิธีการถดถอยบีตา
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงลบ

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงลบ

วิธีประมาณการ ตัวแปรต้น	OLS (1)	OLS (2)	OLS (3)	Beta (4)	Beta (5)	Beta (6)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.01	0.03	0.05	0.03	0.13	0.2
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.02	0.03	0.04	0.08	0.13	0.18
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.04	-0.01	0.03	-0.16	-0.03	0.13
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	0.08***	0.08**	0.04	0.36***	0.33***	0.19
หัวข้อ 5 รูปแบบการสอนและบุคลากร	0.12***	0.13***	0.11***	0.48***	0.55***	0.47***
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.04	0.003	0.05	0.14	-0.01	0.19
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.003	0.003	0.04	-0.03	-0.01	0.13
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	0.22***	0.19***	0.20***	0.90***	0.80***	0.84***
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	0.09***	0.06**	0.11***	0.37***	0.27**	0.46***
ตัวแปรหุ่น						
ข้อความหลัง 2019	-	-0.04***	-0.04***	-	-0.18***	-0.18***
THE Rank < Top 200	-	0.004	0.01	-	0.01	0.05
THE Rank Top 200-501	-	0.01	0.004	-	0.02	0.02
จำนวนนักศึกษาต่อบุคลากร	-	-	0.001***	-	-	0.005***
ร้อยละของนักศึกษาต่างชาติ	-	-	-0.09***	-	-	-0.38***
ร้อยละของนักศึกษาผู้หญิง	-	-	-0.04	-	-	-0.16
จำนวนนักศึกษาเต็มเวลา	-	-	0	-	-	0
ค่าคงที่	0.51***	0.50***	0.50***	0.05	0.02	-0.02
ค่า phi	-	-	-	49.3	52.53	62.87
จำนวนตัวอย่าง	737	737	622	737	737	622
R^2 / Pseudo R2	0.13	0.19	0.29	0.13	0.18	0.28
Adjusted R^2	0.12	0.17	0.27	-	-	-
AIC	-1836	-1878	-1688	-1824	-1865	-1672

ที่มา: ผู้วิจัย

หมายเหตุ:

1. OLS คือ วิธีการถดถอยกำลังสองน้อยสุด Beta คือ วิธีการถดถอยบีตา
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลสถิติเชิงพรรณนาจะพบว่าจำนวนตัวอย่างของตัวแปรคุณลักษณะของมหาวิทยาลัยมีจำนวนน้อยกว่าจำนวนมหาวิทยาลัย เพราะมหาวิทยาลัยจำนวนหนึ่งไม่ปรากฏข้อมูลบนเว็บไซต์ timeshighereducation.com ทำให้ไม่มีข้อมูลส่วนนี้

ในตารางที่ 4.10-4.12 ผู้วิจัยทดลองวิธีการถดถอยกำลังสองน้อยสุด (OLS Regression) และวิธีการถดถอยบีตา (Beta Regression) และทดลองการใส่ตัวแปรหุ่นช่วงเวลา ตัวแปรหุ่นกลุ่มอันดับมหาวิทยาลัย และตัวแปรรคุณลักษณะของมหาวิทยาลัย เพื่อทดสอบกรณีการเพิ่มลดตัวแปร เพื่อหาว่าตัวแบบไหนจะให้ค่าประมาณดีที่สุด

โดยภาพรวมผลการประมาณการด้วยวิธีการถดถอยกำลังสองน้อยสุด และ วิธีการถดถอยบีตา ให้ข้อสรุปสอดคล้องกันในเกือบทุกชุดข้อมูลและชุดตัวแปร โดยให้ค่าตัวประมาณสัมประสิทธิ์การถดถอยที่สอดคล้องในทิศทางเดียวกัน และให้ค่า AIC ที่ใกล้เคียงกันมาก ทั้งนี้ การถดถอยบีตาไม่ต้องใช้ข้อสมมติมากเท่าการถดถอยกำลังสองน้อยสุด และสอดคล้องกับธรรมชาติของข้อมูลที่เป็นช่วงระหว่าง 0 ถึง 1 มากกว่า

กรณีการถดถอยบีตา ในตารางที่ 4.10-4.12 ตัวแบบที่ 4 ให้ค่า AIC ต่ำที่สุด 1 กรณี คือ กรณีคะแนนความรู้สึกเชิงบวกน้อยกว่า 0.7 ขณะที่ตัวแบบที่ 5 ให้ค่า AIC ต่ำที่สุด 2 กรณี คือ กรณีคะแนนความรู้สึกเชิงบวกมากกว่าหรือเท่ากับ 0.7 และกรณีคะแนนความรู้สึกเชิงลบ ฉะนั้น ตัวแบบการถดถอยบีตาที่เหมาะสมที่สุด คือ ตัวแบบที่ 5 คือ ตัวแบบที่มีการใส่ตัวแปรหุ่นช่วงเวลา ตัวแปรหุ่นกลุ่มอันดับมหาวิทยาลัย และไม่ใส่ตัวแปรคุณลักษณะของมหาวิทยาลัย

โดยเมื่อพิจารณาผลของตัวแปรตาม ในเบื้องต้นหัวข้อทั้ง 9 ที่นำมาใช้ในการวิเคราะห์ล้วนแต่เป็นประเด็นที่สำคัญที่มีการเขียนถึงบ่อยในการกล่าวถึงมหาวิทยาลัย การทดสอบระดับนัยสำคัญทางสถิติในการถดถอยบีตาจึงเป็นการบอกปัจจัยที่กำหนดความแปรปรวนของค่าคะแนนความรู้สึกกลุ่มต่าง ๆ ซึ่งหมายถึงหากมหาวิทยาลัยจะพัฒนาคะแนนความรู้สึกที่มีต่อมหาวิทยาลัย มหาวิทยาลัยกลุ่มนั้น ๆ ควรจะเน้นในด้านไหน

ผลการประมาณการถดถอยบีตา (สมการที่ 2.17-2.20) จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}$) กรณีกำหนดตัวแปรต้นครบทุกตัวแปร ตามสมการต่อไปนี้

$$\hat{\mu} = \frac{exp(x^T\hat{\beta})}{1 + exp(x^T\hat{\beta})}, \tag{4.1}$$

โดยที่

$$\begin{aligned} x^T\hat{\beta} = {} & \hat{\beta}_0 + \hat{\beta}_1 P_1 + \hat{\beta}_2 P_2 + \hat{\beta}_3 P_3 + \hat{\beta}_4 P_4 + \hat{\beta}_5 P_5 + \hat{\beta}_6 P_6 \\ & + \hat{\beta}_7 P_7 + \hat{\beta}_8 P_8 + \hat{\beta}_9 P_9 + \hat{\beta}_{10} Period \\ & + \hat{\beta}_{11} TOP200 + \hat{\beta}_{12} TOP500 \end{aligned} \tag{4.2}$$

เมื่อ

μ คือ ค่าเฉลี่ยของค่าคะแนนความรู้สึกจากข้อความที่เขียนถึงมหาวิทยาลัย

$\hat{\beta}_0$ คือ ค่าคงที่

P_1 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน

P_2 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

P_3 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน

P_4 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 4 การใช้ชีวิตและกิจกรรม

P_5 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 5 รูปแบบการสอนและบุคลากร

P_6 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต

P_7 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย

P_8 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ

P_9 คือ ร้อยละของข้อความที่กล่าวถึง
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม

$Period$ คือ ตัวแปรหุ่นช่วงเวลาของข้อความ
= 1 เมื่อ ข้อความหลังปี ค.ศ. 2019
= 0 เมื่อ ข้อความก่อนและจนถึงปี ค.ศ. 2019

$TOP200$ คือ ตัวแปรหุ่นช่วงเวลาของข้อความ
= 1 เมื่อ อันดับของมหาวิทยาลัยน้อยกว่า 200
= 0 เมื่อ อันดับของมหาวิทยาลัยมากกว่าหรือเท่ากับ 200
หรือ ไม่มีในฐาน THE

$TOP500$ คือ ตัวแปรหุ่นช่วงเวลาของข้อความ
= 1 เมื่อ อันดับของมหาวิทยาลัย เท่ากับ 200-501
= 0 เมื่อ อันดับของมหาวิทยาลัยมากกว่า 501 หรือ ไม่มีในฐาน THE

จากตารางที่ 4.10 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}_i(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7) ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.4 + 0.51P_1 + 0.08P_2 - 0.68P_3 + 0.27P_4 - 0.66P_5 + 0.09P_6 - 0.30P_7 - 0.03P_8 - 0.45P_9 + 0.05Period + 0.04TOP200 + 0.03TOP500 \quad (4.3)$$

ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7) ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน และการใช้ชีวิตและกิจกรรม ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และภาระงาน รูปแบบการสอนและบุคลากร ทรัพยากรของมหาวิทยาลัย และประสบการณ์ด้านภาษาและวัฒนธรรม

จากตารางที่ 4.11 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูง (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7) ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.66 + 1.04P_1 + 0.08P_2 - 0.18P_3 - 0.05P_4 + 0.33P_5 + 0.26P_6 - 0.33P_7 + 0.02P_8 - 0.16P_9 - 0.04Period - 0.07TOP200 - 0.02TOP500 \quad (4.4)$$

ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7) ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน รูปแบบการสอนและบุคลากร และค่าใช้จ่ายและต้นทุนการใช้ชีวิต ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน และทรัพยากรของมหาวิทยาลัย

พบว่า กรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก จะต่างจาก กรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก ในตัวแปรด้านรูปแบบการสอนและบุคลากร ที่จะส่งผลลบต่อคะแนนความรู้สึกกรณีคำชมที่แสดงความชื่นชมในระดับไม่สูงมาก แต่จะส่งผลบวกต่อคะแนนความรู้สึกกรณีคำชมที่แสดงความชื่นชมในระดับสูงมาก ซึ่งเป็นตัวสะท้อนว่าหากมหาวิทยาลัยอยากจะได้รับความชื่นชมนอกจากจะต้องพัฒนาในทุกหัวข้อ ยังคงต้องเน้นไปที่ รูปแบบการสอนและบุคลากร เพื่อให้ได้รับคะแนนความรู้สึกเชิงบวกในระดับที่สูงมาก

อีกตัวแปรตามที่ให้ผลต่อคะแนนความรู้สึกเชิงบวกต่างกันในกลุ่มมหาวิทยาลัยที่ได้รับคำชื่นชมไม่สูงมากและสูงมาก คือ การใช้ชีวิตและกิจกรรม และประสบการณ์ด้านภาษาและวัฒนธรรม ที่จะส่งผลต่อคะแนนความรู้สึกกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก แต่จะไม่ส่งผลต่อการเปลี่ยนแปลงของคะแนนความรู้สึกกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก นั้นหมายถึงเมื่อมหาวิทยาลัยสามารถยกระดับให้ตนเองได้ค่าเฉลี่ยคะแนนความรู้สึกเชิงบวกในระดับที่สูงมากแล้ว อาจจะไม่ต้องเน้นเรื่องการใช้ชีวิตและกิจกรรม และประสบการณ์ด้านภาษาและวัฒนธรรม เพิ่มขึ้นมากอีกต่อไป แต่ควรไปเน้นเรื่องอื่น ๆ ที่จะส่งผลต่อคะแนนความรู้สึกเชิงบวกในระดับสูงแทน

จากตารางที่ 4.12 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงลบของมหาวิทยาลัย ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.02 + 0.13P_1 + 0.13P_2 - 0.03P_3 + 0.33P_4 + 0.55P_5 - 0.01P_6 - 0.01P_7 + 0.8P_8 + 0.27P_9 - 0.18Period + 0.01TOP200 + 0.02TOP500 \quad (4.5)$$

ในกรณีข้อความตำหนิที่มีค่าคะแนนความรู้สึกเชิงลบ จะพบว่าการกล่าวถึงปัจจัยต่าง ๆ โดยภาพรวมจะส่งผลบวกต่อค่าคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) โดยปัจจัยที่ช่วยทำให้ค่าคะแนน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความรู้สึกเชิงลบไม่ต่ำมาก คือ การใช้ชีวิตและกิจกรรม รูปแบบการสอนและบุคลากร ชื่อเสียงและความเป็นนานาชาติ และประสบการณ์ด้านภาษาและวัฒนธรรม ถ้ามหาวิทยาลัยจัดการประเด็นเหล่านี้ได้ดี แม้จะเป็นข้อความตำหนิก็จะไม่เป็นข้อความที่มีความรู้สึกเป็นลบมากนัก

ตารางที่ 4.13 ผลประมาณการการถดถอย กรณีก่อนและหลัง ค.ศ. 2019

ตัวแปรตาม ตัวแปรต้น / ก่อน หรือ หลัง 2019	ScorePos_1 ก่อน	 หลัง	ScorePos_2 ก่อน	 หลัง	ScoreNeg ก่อน	 หลัง
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.49***	0.53*	1.12***	0.99***	-0.08	0.72***
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.06	-0.03	0.11	0.04	0.23*	0.1
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.59***	-0.94***	-0.16	-0.18	-0.09	0.16
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	0.2	0.80**	-0.31	0.18	0.33**	0.01
หัวข้อ 5 รูปแบบการสอนและบุคลากร	-0.60***	-0.57**	0.32*	0.24	0.52***	0.49***
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.21	-0.29	0.45**	0.13	-0.07	-0.14
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.3	0.08	-0.45*	0.03	-0.02	-0.01
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	0.05	-0.21	0.11	0.0004	0.54***	1.38***
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	-0.27*	-1.55***	-0.21	-0.36	0.38***	-0.35
ตัวแปรหุ่น						
THE Rank < Top 200	-0.04	-0.05	-0.06*	-0.08***	0.03	-0.06
THE Rank Top 200-501	-0.05	0.07	0.02	-0.08***	0.04	-0.06
ค่าคงที่	0.33**	0.62**	0.61***	0.66***	0.11	-0.37**
ค่า phi	63.54	65.63	180.24	206.67	49.74	77.95
จำนวนตัวอย่าง	394	74	218	181	539	198
Pseudo R2	0.15	0.53	0.28	0.26	0.09	0.4
AIC	-1,057	-180	-870	-737	-1,334	-556

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ScorePos_1 คือ คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7
   ScorePos_2 คือ คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7
   ScoreNeg คือ คะแนนความรู้สึกเชิงลบ
2. ก่อน คือ จนถึง ค.ศ. 2019 หลัง คือ หลัง ค.ศ. 2019
3. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.13 คอลัมน์ ScorePos_1 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7)

กรณีก่อน ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.33 + 0.49P_1 + 0.06P_2 - 0.59P_3 + 0.2P_4 - 0.6P_5 + 0.21P_6 - 0.3P_7 + 0.05P_8 - 0.27P_9 - 0.04TOP200 - 0.05TOP500 \quad (4.6)$$

กรณีหลัง ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.62 + 0.53P_1 - 0.03P_2 - 0.94P_3 + 0.8P_4 - 0.57P_5 - 0.29P_6 + 0.08P_7 - 0.21P_8 - 1.55P_9 - 0.05TOP200 + 0.07TOP500 \quad (4.7)$$

จากตารางที่ 4.13 คอลัมน์ ScorePos_2 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูง (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7)

กรณีก่อน ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.61 + 1.12P_1 + 0.11P_2 - 0.16P_3 - 0.31P_4 + 0.32P_5 + 0.45P_6 - 0.45P_7 + 0.11P_8 - 0.21P_9 - 0.06TOP200 + 0.02TOP500 \quad (4.8)$$

กรณีหลัง ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.66 + 0.99P_1 + 0.04P_2 - 0.18P_3 + 0.18P_4 + 0.24P_5 + 0.13P_6 + 0.03P_7 + 0.0004P_8 - 0.36P_9 - 0.08TOP200 - 0.08TOP500 \quad (4.9)$$

ช่วงก่อนและหลัง ค.ศ. 2019 ที่เกิดเหตุกรณีการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ซึ่งทำให้เกิดการเปลี่ยนแปลงต่อรูปแบบการเรียนการสอนทั่วโลก

จากตารางที่ 4.13 คอลัมน์ ScorePos_1 กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยที่ได้รับคำชมในระดับไม่สูงมากจะพบว่าก่อนและหลัง ค.ศ. 2019 ปัจจัยที่ส่งผลต่อคะแนนความรู้สึกเชิงบวกจะไม่แตกต่างกัน ยกเว้นหัวข้อเรื่องการใช้ชีวิตและกิจกรรมที่ส่งผลต่อคะแนนความรู้สึกต่อมหาวิทยาลัยอย่างมีนัยสำคัญทางสถิติหลังช่วง ค.ศ. 2019

จากตารางที่ 4.13 คอลัมน์ ScorePos_2 กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยที่ได้รับคำชมในระดับสูง หัวข้อเรื่องรูปแบบการสอนและบุคลากร ค่าใช้จ่ายและต้นทุนการใช้ชีวิต และทรัพยากรของมหาวิทยาลัย ช่วงก่อน ค.ศ. 2019 จะเป็นปัจจัยที่ส่งผลต่อคะแนนความรู้สึกต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มหาวิทยาลัยอย่างมีนัยสำคัญทางสถิติ แต่หลัง ค.ศ. 2019 กลับไม่เป็นปัจจัยที่ส่งผลต่อคะแนนความรู้สึกต่อมหาวิทยาลัยอย่างมีนัยสำคัญทางสถิติ

จากตารางที่ 4.13 คอลัมน์ ScoreNeg จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงลบของมหาวิทยาลัย

กรณีก่อน ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.11 - 0.08P_1 + 0.23P_2 - 0.09P_3 + 0.33P_4 + 0.52P_5 - 0.07P_6 - 0.02P_7 + 0.54P_8 + 0.38P_9 + 0.03TOP200 + 0.04TOP500 \quad (4.10)$$

กรณีหลัง ค.ศ. 2019 ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = -0.37 + 0.72P_1 + 0.1P_2 + 0.16P_3 + 0.01P_4 + 0.49P_5 - 0.14P_6 - 0.01P_7 + 1.38P_8 - 0.35P_9 - 0.06TOP200 - 0.06TOP500 \quad (4.11)$$

จากตารางที่ 4.13 คอลัมน์ ScoreNeg กรณีข้อความเห็นเชิงลบ หลัง ค.ศ. 2019 มีการกล่าวถึง ประสบการณ์โดยภาพรวมและการเรียน จะทำให้ค่าคะแนนความรู้สึกมีค่าเพิ่มขึ้น (รู้สึกลบน้อยลง) ทั้ง ๆ ที่ก่อนหน้า ค.ศ. 2019 การกล่าวถึงปัจจัยด้านนี้ไม่ส่งผลต่อคะแนนความรู้สึก อีกด้านหนึ่งคือ สถานที่และชีวิตในมหาวิทยาลัย การใช้ชีวิตและกิจกรรม ประสบการณ์ด้านภาษาและวัฒนธรรม ที่เคยมีความสำคัญช่วยให้คะแนนความรู้สึกเชิงลบไม่ต่ำมากในช่วงก่อน ค.ศ. 2019 พอหลัง ค.ศ. 2019 ปัจจัยเหล่านี้ไม่ส่งผลสำคัญอีกต่อไป ส่วนหนึ่งน่าจะเกิดจากการปรับรูปแบบการเรียนการสอนทำให้นักเรียนพบปะกันน้อยลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.14 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7 แยกรายประเทศ

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงบวก

ประเทศ ตัวแปรต้น	UK (1)	US (2)	AU (3)	CA (4)	CN (5)	JP (6)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.88***	0.37	-0.21	0.2	1.01*	0.59
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	-0.11	-0.05	-0.34	0.27	0.75*	0.25
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.1	-0.68***	-1.24***	0.27	-1.20**	-0.16
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	0.26	0.03	0.72**	-0.74*	0.63	-0.16
หัวข้อ 5 รูปแบบการสอนและบุคลากร	-0.34**	-0.33	-0.43	-0.04	0.78	-1.66***
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.04	0.08	0.01	0.39	-1.24**	0.33
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	0.04	-0.44	-0.90**	-0.35	0.71	-2.11***
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	-0.0001	0.16	0.01	-0.72*	-0.75*	-0.01
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	-0.99***	-0.16	-0.77*	-0.52	-1.01**	0.96*
ตัวแปรหุ่น						
THE Rank < Top 200	0.07*	0.13**	-0.15	0.15*	-0.13	-0.07
THE Rank Top 200-501	-0.04	0.17***	-0.07	0.16**	-0.04	0.22
ค่าคงที่	0.27*	0.34*	1.08***	0.44	0.25	0.29
ค่า phi	161.6	53.35	54.28	175.35	261.26	479.65
จำนวนตัวอย่าง	143	155	94	34	18	14
Pseudo R2	0.4	0.16	0.29	0.35	0.61	0.77
AIC	-503	-374	-216	-102	-49	-40

ที่มา: ผู้วิจัย

หมายเหตุ:

1. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น (กรณี ไทย มีจำนวนตัวอย่างไม่พอ)
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7 แยกรายประเทศ

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงบวก

ประเทศ ตัวแปรต้น	UK (1)	US (2)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	1.29***	0.09
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.15	-0.45
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	-0.1	0.09
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	-0.05	-0.31
หัวข้อ 5 รูปแบบการสอนและบุคลากร	0.33**	0.51*
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.27*	0.36
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.36**	0.72*
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	-0.13	0.95**
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	-0.39**	0.80***
ตัวแปรหุ่น		
THE Rank < Top 200	-0.06**	0.16
THE Rank Top 200-501	-0.02	0.11
ค่าคงที่	0.48***	0.72***
ค่า phi	191.12	534.54
จำนวนตัวอย่าง	373	20
Pseudo R2	0.28	0.75
AIC	-1525	-77

ที่มา: ผู้วิจัย

หมายเหตุ:

1. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา (กรณี ประเทศอื่น มีจำนวนตัวอย่างไม่พอ)
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 ผลประมาณการการถดถอย กรณีคะแนนความรู้สึกเชิงลบ แยกรายประเทศ

ตัวแปรตาม คือ คะแนนความรู้สึกเชิงลบ

ประเทศ ตัวแปรต้น	UK (1)	US (2)	AU (3)	CA (4)	CN (5)	JP (6)
หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	0.33*	-0.1	-0.16	-0.34	6.09***	-0.73
หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	0.19	0.60**	0.16	0.03	0.48	1.91***
หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน	0.14	-0.75***	-0.03	-0.43	-1.48**	1.30***
หัวข้อ 4 การใช้ชีวิตและกิจกรรม	0.59***	0.41	0.13	-0.87**	-4.34***	0.06
หัวข้อ 5 รูปแบบการสอนและบุคลากร	-0.22	0.80***	0.36	-0.49	0.63	-1.04*
หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	0.45**	-0.44	0.17	0.94	-1.11*	1.94**
หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย	-0.29	0.35	-0.09	-0.02	-1.50***	-0.87
หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	0.69***	0.3	0.77***	0.18	4.00***	-0.08
หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	0.29	0.03	-0.08	0.19	-0.84	2.28***
ตัวแปรหุ่น						
THE Rank < Top 200	-0.04	0.02	-0.11	0.08	-0.64***	0.22
THE Rank Top 200-501	-0.03	0.03	-0.16*	0.08	-0.73***	-0.03
ค่าคงที่	-0.12	0.33	0.44	0.70*	0.1	-0.81*
ค่า phi	70.75	38.04	57.63	115.01	371.92	181.17
จำนวนตัวอย่าง	433	133	102	28	16	17
Pseudo R2	0.11	0.19	0.12	0.36	0.87	0.77
AIC	-1205	-276	-250	-68	-47	-39

ที่มา: ผู้วิจัย

หมายเหตุ:

1. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น (กรณี ไทย มีจำนวนตัวอย่างไม่พอ)
2. ระดับนัยสำคัญทางสถิติ จากค่า P-Value ***, **, * คือ ค่านัยสำคัญทางสถิติที่ระดับ 0.01, 0.05, 0.1 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 สรุปผลตัวแปรหัวข้อที่มีความสำคัญในการกำหนดความรู้สึกที่มีต่อมหาวิทยาลัย

ประเภทข้อความ	กลุ่มตัวอย่าง	หัวข้อที่ทำให้รู้สึกบวกมากขึ้น (ลบน้อยลง)	หัวข้อที่ทำให้รู้สึกบวกน้อยลง (ลบมากขึ้น)
ข้อความที่มีความรู้สึกเชิงบวก < 0.7	ทั้งหมด	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 4 การใช้ชีวิตและกิจกรรม	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
ข้อความที่มีความรู้สึกเชิงบวก ≥ 0.7	ทั้งหมด	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย
ข้อความที่มีความรู้สึกเชิงลบ	ทั้งหมด	หัวข้อ 4 การใช้ชีวิตและกิจกรรม หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	
ข้อความที่มีความรู้สึกเชิงบวก < 0.7	จนถึง ค.ศ. 2019	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
	หลัง ค.ศ. 2019	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 4 การใช้ชีวิตและกิจกรรม	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
ข้อความที่มีความรู้สึกเชิงบวก ≥ 0.7	จนถึง ค.ศ. 2019	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทข้อความ	กลุ่มตัวอย่าง	หัวข้อที่ทำให้รู้สึกบวกมากขึ้น (ลบน้อยลง)	หัวข้อที่ทำให้รู้สึกบวกน้อยลง (ลบมากขึ้น)
	หลัง ค.ศ. 2019	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	
ข้อความที่มีความรู้สึกเชิงลบ	จนถึง ค.ศ. 2019	หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย หัวข้อ 4 การใช้ชีวิตและกิจกรรม หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	
	หลัง ค.ศ. 2019	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	
ข้อความที่มีความรู้สึกเชิงบวก < 0.7	UK	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน	หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
	US		หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน
	AU	หัวข้อ 4 การใช้ชีวิตและกิจกรรม	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
	CA		หัวข้อ 4 การใช้ชีวิตและกิจกรรม หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทข้อความ	กลุ่มตัวอย่าง	หัวข้อที่ทำให้รู้สึกบวกมากขึ้น (ลบน้อยลง)	หัวข้อที่ทำให้รู้สึกบวกน้อยลง (ลบมากขึ้น)
	CN	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
	JP	หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย
ข้อความที่มีความรู้สึกเชิงบวก >= 0.7	UK	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต	หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม
	US	หัวข้อ 5 รูปแบบการสอนและบุคลากร หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	
ข้อความที่มีความรู้สึกเชิงลบ	UK	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 4 การใช้ชีวิตและกิจกรรม หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	
	US	หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย หัวข้อ 5 รูปแบบการสอนและบุคลากร	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเภทข้อความ	กลุ่มตัวอย่าง	หัวข้อที่ทำให้รู้สึกบวกมากขึ้น (ลบน้อยลง)	หัวข้อที่ทำให้รู้สึกบวกน้อยลง (ลบมากขึ้น)
	AU	หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	
	CA		หัวข้อ 4 การใช้ชีวิตและกิจกรรม
	CN	หัวข้อ 1 ประสบการณ์โดยภาพรวมและการเรียน หัวข้อ 8 ชื่อเสียงและความเป็นนานาชาติ	หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 4 การใช้ชีวิตและกิจกรรม หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต หัวข้อ 7 ทรัพยากรของมหาวิทยาลัย
	JP	หัวข้อ 2 สถานที่และชีวิตในมหาวิทยาลัย หัวข้อ 3 ความท้าทายทางวิชาการและภาระงาน หัวข้อ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต หัวข้อ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม	หัวข้อ 5 รูปแบบการสอนและบุคลากร

ที่มา: ผู้วิจัย

หมายเหตุ:

1. UK คือ สหราชอาณาจักร US คือ สหรัฐอเมริกา AU คือ ออสเตรเลีย CA คือ แคนาดา CN คือ จีน JP คือ ญี่ปุ่น (กรณี ไทย มีจำนวนตัวอย่างไม่พอ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.14 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7)

กรณีประเทศสหราชอาณาจักร ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.27 + 0.88P_1 - 0.11P_2 - 0.1P_3 + 0.26P_4 - 0.34P_5 + 0.04P_6 + 0.04P_7 - 0.0001P_8 - 0.99P_9 + 0.07TOP200 - 0.04TOP500 \quad (4.12)$$

กรณีประเทศสหรัฐอเมริกา ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.34 + 0.37P_1 - 0.05P_2 - 0.68P_3 + 0.03P_4 - 0.33P_5 + 0.08P_6 - 0.44P_7 + 0.16P_8 - 0.16P_9 + 0.13TOP200 + 0.17TOP500 \quad (4.13)$$

กรณีประเทศออสเตรเลีย ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 1.08 - 0.21P_1 - 0.34P_2 - 1.24P_3 + 0.72P_4 - 0.43P_5 + 0.01P_6 - 0.9P_7 + 0.01P_8 - 0.77P_9 - 0.15TOP200 - 0.07TOP500 \quad (4.14)$$

กรณีประเทศแคนาดา ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.44 + 0.2P_1 + 0.27P_2 + 0.27P_3 - 0.74P_4 - 0.04P_5 + 0.39P_6 - 0.35P_7 - 0.72P_8 - 0.52P_9 + 0.15TOP200 + 0.16TOP500 \quad (4.15)$$

กรณีประเทศจีน ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.25 + 1.01P_1 + 0.75P_2 - 1.2P_3 + 0.63P_4 + 0.78P_5 - 1.24P_6 + 0.71P_7 - 0.75P_8 - 1.01P_9 - 0.13TOP200 - 0.04TOP500 \quad (4.16)$$

กรณีประเทศญี่ปุ่น ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.29 + 0.59P_1 + 0.25P_2 - 0.16P_3 - 0.16P_4 - 1.66P_5 + 0.33P_6 - 2.11P_7 - 0.01P_8 + 0.96P_9 - 0.07TOP200 + 0.22TOP500 \quad (4.17)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.15 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงบวกของมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูง (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7)

กรณีประเทศสหราชอาณาจักร ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.48 + 1.29P_1 + 0.15P_2 - 0.1P_3 - 0.05P_4 + 0.33P_5 + 0.27P_6 - 0.36P_7 - 0.13P_8 - 0.39P_9 - 0.06TOP200 - 0.02TOP500 \quad (4.18)$$

กรณีประเทศสหรัฐอเมริกา ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.72 + 0.09P_1 - 0.45P_2 + 0.09P_3 - 0.31P_4 + 0.51P_5 + 0.36P_6 + 0.72P_7 + 0.95P_8 + 0.8P_9 + 0.16TOP200 + 0.11TOP500 \quad (4.19)$$

จากตารางที่ 4.16 จะได้ตัวแบบสมการเพื่อประมาณการค่าเฉลี่ยของค่าคะแนนความรู้สึก ($\hat{\mu}(x^T\hat{\beta})$) ในกรณีข้อความเชิงลบของมหาวิทยาลัย

กรณีประเทศสหราชอาณาจักร ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = -0.12 + 0.33P_1 + 0.19P_2 + 0.14P_3 + 0.59P_4 - 0.22P_5 + 0.45P_6 - 0.29P_7 + 0.69P_8 + 0.29P_9 - 0.04TOP200 - 0.03TOP500 \quad (4.20)$$

กรณีประเทศสหรัฐอเมริกา ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.33 - 0.1P_1 + 0.6P_2 - 0.75P_3 + 0.41P_4 + 0.8P_5 - 0.44P_6 + 0.35P_7 + 0.3P_8 + 0.03P_9 + 0.02TOP200 + 0.03TOP500 \quad (4.21)$$

กรณีประเทศออสเตรเลีย ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.44 - 0.16P_1 + 0.16P_2 - 0.03P_3 + 0.13P_4 + 0.36P_5 + 0.17P_6 - 0.09P_7 + 0.77P_8 - 0.08P_9 - 0.11TOP200 - 0.16TOP500 \quad (4.22)$$

กรณีประเทศแคนาดา ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.7 - 0.34P_1 + 0.03P_2 - 0.43P_3 - 0.87P_4 - 0.49P_5 + 0.94P_6 - 0.02P_7 + 0.18P_8 + 0.19P_9 + 0.08TOP200 + 0.08TOP500 \quad (4.23)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีประเทศจีน ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = 0.1 + 6.09P_1 + 0.48P_2 - 1.48P_3 - 4.34P_4 + 0.63P_5 - 1.11P_6 - 1.5P_7 + 4P_8 - 0.84P_9 - 0.64TOP200 - 0.73TOP500 \quad 4.24$$

กรณีประเทศญี่ปุ่น ดังสมการต่อไปนี้ (คำอธิบายตัวแปรในท้ายสมการที่ 4.2)

$$x^T\hat{\beta} = -0.81 - 0.73P_1 + 1.91P_2 + 1.3P_3 + 0.06P_4 - 1.04P_5 + 1.94P_6 - 0.87P_7 - 0.08P_8 + 2.28P_9 + 0.22TOP200 - 0.03TOP500 \quad 4.25$$

จากตารางที่ 4.14 และ 4.15 คอลัมน์ UK เมื่อพิจารณาเป็นรายประเทศกรณีข้อความเห็นเชิงบวก กรณีสหราชอาณาจักร กลุ่มมหาวิทยาลัยที่ได้รับคำชมในระดับที่ไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7) ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ รูปแบบการสอนและบุคลากร ประสบการณ์ด้านภาษาและวัฒนธรรม ขณะที่กลุ่มมหาวิทยาลัยที่ได้รับคำชมในระดับที่สูงมาก (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7) ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน รูปแบบการสอนและบุคลากร ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ทรัพยากรของมหาวิทยาลัย ประสบการณ์ด้านภาษาและวัฒนธรรม ฉะนั้น โดยรวมปัจจัยที่ส่งผลอย่างมีนัยสำคัญ คือ ประสบการณ์โดยภาพรวมและการเรียน ปัจจัยที่ส่งผลลบอย่างมีนัยสำคัญ คือ ประสบการณ์ด้านภาษาและวัฒนธรรม ขณะที่คำชมจะมากน้อยจะขึ้นกับเป็นมหาวิทยาลัยที่จัดการด้าน รูปแบบการสอนและบุคลากร ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ได้ดีมากขึ้นหรือไม่

จากตารางที่ 4.14 และ 4.15 คอลัมน์ US กรณีข้อความเห็นเชิงบวกของสหรัฐอเมริกา กลุ่มมหาวิทยาลัยที่ได้รับคำชมในระดับที่ไม่สูงมาก (คะแนนความรู้สึกเชิงบวก น้อยกว่า 0.7) ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน ขณะที่ กลุ่มมหาวิทยาลัยที่ได้รับคำชมในระดับที่สูงมาก (คะแนนความรู้สึกเชิงบวก มากกว่าหรือเท่ากับ 0.7) ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ รูปแบบการสอนและบุคลากร ทรัพยากรของมหาวิทยาลัย ชื่อเสียงและความเป็นนานาชาติ ประสบการณ์ด้านภาษาและวัฒนธรรม ฉะนั้น โดยรวมปัจจัยที่ส่งผลอย่างมีนัยสำคัญ คือ รูปแบบการสอนและบุคลากร ทรัพยากรของมหาวิทยาลัย ชื่อเสียงและความเป็นนานาชาติ จะต่างจากสหราชอาณาจักรตรงที่ ประสบการณ์โดยภาพรวมและการเรียน ไม่ส่งต่อคะแนนความรู้สึกมากในกรณีของสหรัฐอเมริกา ขณะที่ ความท้าทายทางวิชาการและภาระงาน จะไม่ส่งต่อคะแนนความรู้สึกมากในกรณีของสหราชอาณาจักร อีกเรื่องคือ ชื่อเสียงและความเป็นนานาชาติ ประสบการณ์ด้านภาษาและวัฒนธรรม ที่จะมีการพูดถึงบ่อยในกรณีของสหรัฐอเมริกา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.14 คอลัมน์ AU กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยในออสเตรเลีย ออสเตรเลียเป็นประเทศเดียวที่การพูดถึง การใช้ชีวิตและกิจกรรม ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ ขณะที่ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน ทรัพยากรของมหาวิทยาลัย และประสบการณ์ด้านภาษาและวัฒนธรรม

จากตารางที่ 4.14 คอลัมน์ CA กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยในแคนาดา ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ การใช้ชีวิตและกิจกรรม และชื่อเสียงและความเป็นนานาชาติ

จากตารางที่ 4.14 คอลัมน์ CN กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยในจีน ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย ซึ่งจีนเป็นประเทศเดียวที่ สถานที่และชีวิตในมหาวิทยาลัยส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ชื่อเสียงและความเป็นนานาชาติ และประสบการณ์ด้านภาษาและวัฒนธรรม

จากตารางที่ 4.14 คอลัมน์ JP กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยในญี่ปุ่น ญี่ปุ่นเป็นประเทศเดียวที่การพูดถึง ประสบการณ์ด้านภาษาและวัฒนธรรม ส่งผลบวกต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงบวกอย่างมีนัยสำคัญทางสถิติ คือ รูปแบบการสอนและบุคลากร ทรัพยากรของมหาวิทยาลัย

จากตารางที่ 4.16 คอลัมน์ UK กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในสหราชอาณาจักร ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) อย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน การใช้ชีวิตและกิจกรรม ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ชื่อเสียงและความเป็นนานาชาติ หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่องชื่อเสียงและความเป็นนานาชาติช่วยให้รู้สึกลบน้อยลง แต่อาจจะไม่ได้ช่วยให้รู้สึกบวกมากขึ้น

จากตารางที่ 4.16 คอลัมน์ US กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในสหรัฐอเมริกา ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) อย่างมีนัยสำคัญทางสถิติ คือ รูปแบบการสอนและบุคลากร สถานที่และชีวิตในมหาวิทยาลัย ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบมากขึ้น) อย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน เช่นเดียวกับกรณีคะแนนความรู้สึกเชิงบวก หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่องสถานที่และชีวิตในมหาวิทยาลัย ช่วยให้รู้สึกลบน้อยลงแต่อาจจะไม่ได้ช่วยให้รู้สึกบวกมากขึ้น

จากตารางที่ 4.16 คอลัมน์ AU กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในออสเตรเลีย ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) อย่างมีนัยสำคัญทางสถิติ คือ ชื่อเสียงและความเป็นนานาชาติ หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่อง ชื่อเสียงและความเป็นนานาชาติ ช่วยให้รู้สึกลบน้อยลงแต่อาจจะไม่ได้ช่วยให้รู้สึกบวกมากขึ้น ขณะที่ การใช้ชีวิติและกิจกรรม ช่วยให้รู้สึกบวกมากขึ้นแต่อาจจะไม่ได้ช่วยให้รู้สึกลบลดลง

จากตารางที่ 4.16 คอลัมน์ CA กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในแคนาดา ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบมากขึ้น) อย่างมีนัยสำคัญทางสถิติ คือ การใช้ชีวิติและกิจกรรม หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่อง การใช้ชีวิติและกิจกรรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทั้งทำให้รู้สึกลบมากขึ้นและรู้สึกบวกลดลง ขณะที่ ชื่อเสียงและความเป็นนานาชาติ ไม่ทำให้รู้สึกลบลดลงแต่ทำให้รู้สึกบวกลดลง

จากตารางที่ 4.16 คอลัมน์ CN กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในจีน ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) อย่างมีนัยสำคัญทางสถิติ คือ ประสบการณ์โดยภาพรวมและการเรียน ชื่อเสียงและความเป็นนานาชาติ ตัวแปรที่ส่งผลลบต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบมากขึ้น) อย่างมีนัยสำคัญทางสถิติ คือ ความท้าทายทางวิชาการและภาระงาน การใช้ชีวิตและกิจกรรม ค่าใช้จ่ายและต้นทุนการใช้ชีวิตทรัพยากรของมหาวิทยาลัย หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่อง ชื่อเสียงและความเป็นนานาชาติ ช่วยให้รู้สึกลบน้อยลงแต่อาจจะไม่ได้ช่วยให้รู้สึกบวกมากขึ้น

จากตารางที่ 4.16 คอลัมน์ JP กรณีข้อความเห็นเชิงลบของมหาวิทยาลัยในญี่ปุ่น ตัวแปรที่ส่งผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง) อย่างมีนัยสำคัญทางสถิติ คือ สถานที่และชีวิตในมหาวิทยาลัย ความท้าทายทางวิชาการและภาระงาน ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ประสบการณ์ด้านภาษาและวัฒนธรรม หากพิจารณารวมกับกรณีคะแนนความรู้สึกเชิงบวก แสดงว่าเรื่อง สถานที่และชีวิตในมหาวิทยาลัย ความท้าทายทางวิชาการและภาระงาน ค่าใช้จ่ายและต้นทุนการใช้ชีวิต ช่วยให้รู้สึกลบน้อยลงแต่อาจจะไม่ได้ช่วยให้รู้สึกบวกมากขึ้น ขณะที่ ประสบการณ์ด้านภาษาและวัฒนธรรม ทั้งช่วยให้รู้สึกลบน้อยลงและช่วยให้รู้สึกบวกมากขึ้น

ฉะนั้น โดยภาพรวมแล้วหากแต่ละประเทศต้องการเพิ่มคะแนนความรู้สึกที่ผู้คนมีต่อมหาวิทยาลัย แต่ละประเทศควรมีจุดเน้นที่แตกต่างกัน กรณีของสหราชอาณาจักร คือการเน้นส่งเสริมด้านประสบการณ์โดยภาพรวมและการเรียน เน้นแก้ไขด้านประสบการณ์ด้านภาษาและวัฒนธรรม กรณีของสหรัฐ คือการเน้นส่งเสริมด้านรูปแบบการสอนและบุคลากร และด้านความเป็นนานาชาติ เน้นแก้ไขด้านประสบการณ์ความท้าทายทางวิชาการและภาระงาน กรณีของออสเตรเลีย คือการเน้นส่งเสริมด้านการใช้ชีวิตและกิจกรรม เน้นแก้ไขด้านความท้าทายทางวิชาการและภาระงาน กรณีของแคนาดา ไม่มีด้านไหนที่ต้องเน้นส่งเสริมเพิ่มมากเป็นพิเศษ เน้นแก้ไขด้านการใช้ชีวิตและกิจกรรม กรณีของจีน คือการเน้นส่งเสริมด้านประสบการณ์โดยภาพรวมและการเรียน เน้นแก้ไขด้านการใช้ชีวิตและกิจกรรม กรณีของญี่ปุ่น คือการเน้นส่งเสริมด้านประสบการณ์ด้านภาษาและวัฒนธรรม เน้นแก้ไขด้านรูปแบบการสอนและบุคลากร

4.4 การอภิปรายผลการศึกษา

การศึกษานี้วิเคราะห์ข้อความความเห็นของผู้คนต่อมหาวิทยาลัย และพิจารณาประเด็นที่มหาวิทยาลัยมักจะถูกกล่าวถึง และผลของประเด็นเหล่านี้ต่อความรู้สึกของผู้คนต่อมหาวิทยาลัย ซึ่งการดำเนินการในลักษณะคล้ายนี้มีการศึกษาอื่นที่เคยทำมาแล้ว แต่จะเป็นวิเคราะห์ข้อความความเห็นของผู้คนต่อสิ่งอื่นที่ไม่ใช่มหาวิทยาลัย เช่น ข้อความเห็นต่อสวนสาธารณะในเมือง (Huai and Van de Voorde, 2022) ข้อความเห็นต่อสินค้าในระบบอีคอมเมิร์ซ (Biswas *et al.*, 2022) ข้อความเห็นต่อการเรียนการสอนแบบออนไลน์ (Li et al., 2022)

การวิเคราะห์ข้อมูลข้อความในด้านการศึกษา ส่วนมากจะยังคงเป็นการพัฒนาตัวแบบเพื่อการจัดกลุ่มข้อความ หรือการจำแนกข้อความ โดยการสร้างตัวแบบสมการที่หลากหลายและแสดงผลเปรียบเทียบประสิทธิภาพของแต่ละวิธี พบว่าการวิจัยส่วนมากจะใช้ข้อมูลแสดงความคิดเห็นที่ปรากฏบนแพลตฟอร์มการเรียนการสอนแบบออนไลน์ (Massive Open Online Course: MOOCs)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่น การศึกษาของ Li et al. (2019) Barrón Estrada et al. (2020) และ Onan (2021) การวิเคราะห์ความสอดคล้องของหลักสูตรและความต้องการของตลาดแรงงาน (Almaleh *et al.*, 2019; Spada *et al.*, 2022)

ทั้งนี้ ข้อค้นพบในการศึกษาโดยเฉพาะในขั้นตอนของการแบ่งกลุ่มมหาวิทยาลัยในการวิเคราะห์จะให้ผลการศึกษาสอดคล้องกับการศึกษาที่ดำเนินการด้วยข้อมูล ตัวแปรและวิธีการอื่น โดยพบว่าการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ส่งผลอย่างมีนัยสำคัญต่อการเรียนการสอนในระดับมหาวิทยาลัย (เช่นใน Estrada Guillén, Monferrer Tirado and Rodríguez Sánchez, 2022; Tomej *et al.*, 2022; Ferrer *et al.*, 2023; Abbas, Ar and Hosseini, 2024) ขณะมหาวิทยาลัยในแต่ละประเทศมีรูปแบบการจัดบริการการศึกษาที่มีจุดเน้นที่แตกต่างกัน (เช่นใน Andrade, 2006; Araya, 2014)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยบนสื่อออนไลน์เป็นข้อมูลที่มีค่ามากเนื่องจากเป็นแหล่งข้อมูลที่มีขนาดใหญ่ที่สะท้อนประสบการณ์และความคิดเห็นจากบุคคลจำนวนมาก ทั้งนี้การใช้ประโยชน์จากข้อมูลการแสดงความคิดเห็นต่อมหาวิทยาลัยบนสื่อออนไลน์มีความท้าทายมากกว่าข้อมูลอื่น ๆ เนื่องจากข้อมูลนี้มักเป็นข้อความที่ไม่มีโครงสร้างและอยู่กระจัดกระจายในหลายแหล่ง

การวิเคราะห์ข้อความแสดงความคิดเห็นในอดีตมักมุ่งเน้นการจัดกลุ่มหรือการจำแนกข้อความ โดยมักเกี่ยวข้องกับสินค้าและบริการ แต่การใช้ข้อมูลนี้ในด้านการศึกษายังมีข้อจำกัดโดยมีการวิจัยซึ่งจะพบกรณีการวิเคราะห์ข้อความเห็นแพลตฟอร์มการเรียนการสอนออนไลน์ (MOOCs) ข้อความจากแบบสอบถามและข้อความจากตัวหลักสูตร

การศึกษาแบ่งขั้นตอนการศึกษาเป็น 3 ส่วนหลัก คือ การรวบรวมข้อมูล การวิเคราะห์ข้อมูลข้อความ และการวิเคราะห์เชิงสถิติ ในขั้นตอนการรวบรวมข้อมูล ใช้วิธีการขูดและจัดรูปข้อมูล HTML เพื่อดึงข้อมูลจากแหล่งต่าง ๆ และทำการวิเคราะห์ข้อมูลข้อความ เพื่อแปลงข้อความเป็นข้อมูลตัวเลขโดยใช้เทคโนโลยีและเครื่องมือต่าง ๆ เช่น Google NLP และ Natural Language Toolkit (NLTK) ผลการวิเคราะห์ค่าคะแนนความรู้สึก (Sentimental Analysis) ต่อมหาวิทยาลัยโดยใช้ Google NLP พบว่าได้ผลที่ดีเมื่อเทียบเคียงกับคะแนนที่ผู้ให้ความเห็นให้ไว้ในเว็บไซต์ของข้อความ

นอกจากนี้ ผู้วิจัยวิเคราะห์ “หัวข้อ” ที่แต่ละข้อความเห็นกล่าวถึงโดยใช้การจัดสรรของดีรีเคลแฝง (Latent Dirichlet Allocation: LDA) ร่วมกับการแปลงข้อมูลข้อความเป็นเวกเตอร์ด้วยวิธี Word2Vec และการวิเคราะห์คำที่อยู่ติดกันเป็นคู่ (Bigram) ปรับค่าไฮเปอร์พารามิเตอร์ด้วยการพิจารณาค่าคะแนนความเชื่อมโยง ความสามารถในการตีความและการวิเคราะห์การแบ่งกลุ่มแบบมีขั้นตอน (Hierarchical Cluster Analysis) พบว่าข้อมูลมีความเหมาะสมที่จะแบ่งออกเป็น 9 หัวข้อ คือ

หัวข้อที่ 1 ประสบการณ์โดยภาพรวมและการเรียน (Overall Experience and Learning)
หัวข้อที่ 2 สถานที่และชีวิตในมหาวิทยาลัย (Facilities and Campus Life)
หัวข้อที่ 3 ความท้าทายทางวิชาการและภาระงาน (Academic Challenges and Workload)
หัวข้อที่ 4 การใช้ชีวิตและกิจกรรม (Student Life and Social Activities)
หัวข้อที่ 5 แบบการสอนและบุคลากร (Teaching System and Staff)
หัวข้อที่ 6 ค่าใช้จ่ายและต้นทุนการใช้ชีวิต (Expense and Cost of Living)
หัวข้อที่ 7 ทรัพยากรของมหาวิทยาลัย (Infrastructure and Resources)
หัวข้อที่ 8 ชื่อเสียงและความเป็นนานาชาติ (Reputation and Internationality)
หัวข้อที่ 9 ประสบการณ์ด้านภาษาและวัฒนธรรม (Language and Cultural Experience)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อพิจารณาการกระจายของหัวข้อ พบว่าข้อความเห็นต่อมหาวิทยาลัยเชิงบวกมักจะเน้นที่ประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย และความท้าทายทางวิชาการและภาระงาน ในขณะที่ข้อความเห็นเชิงลบมักมุ่งไปที่ความท้าทายทางวิชาการและภาระงาน ประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย และรูปแบบการสอนและบุคลากร

เมื่อพิจารณาการกระจายของหัวข้อจำแนกตามช่วงเวลา ในส่วนของข้อความเชิงบวกพบว่ามีความแตกต่างที่น่าสนใจ ระหว่างช่วงก่อนและหลังปี ค.ศ. 2019 (เหตุการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019) โดยหลังปี ค.ศ. 2019 มีการพูดถึงรูปแบบการสอนและบุคลากรมากขึ้น ในขณะที่การใช้ชีวิตและกิจกรรม และประสบการณ์ด้านภาษาและวัฒนธรรม มีการกล่าวถึงลดลง

ในกรณีของข้อความเชิงลบ มีการพูดถึง ความท้าทายทางวิชาการและภาระงาน รูปแบบการสอนและบุคลากร เพิ่มขึ้นมาก และมีการกล่าวถึงค่าใช้จ่ายและต้นทุนการใช้ชีวิต และประสบการณ์ด้านภาษาและวัฒนธรรม ลดลง ซึ่งน่าจะเป็นผลจากการที่การระบาดของโรคทำให้รูปแบบการสอนเปลี่ยนเป็นแบบออนไลน์มากขึ้น ส่งผลต่อการเปลี่ยนแปลงการพูดถึงมหาวิทยาลัย

เมื่อพิจารณาการกระจายของหัวข้อจำแนกตามกลุ่มประเทศ กรณีข้อความที่มีความรู้สึกเชิงบวกทุกประเทศมักเน้นการกล่าวถึง ประสบการณ์โดยภาพรวมและการเรียน เป็นหลัก ขณะเดียวกันในข้อความที่มีความรู้สึกเชิงลบแทบทุกประเทศมักเน้นการกล่าวถึง ความท้าทายทางวิชาการและภาระงานเป็นหลัก รองลงมาคือประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย โดยยกเว้นกรณีของออสเตรเลียที่มีการบ่นถึง รูปแบบการสอนและบุคลากร มากกว่าประสบการณ์โดยภาพรวมและการเรียน สถานที่และชีวิตในมหาวิทยาลัย ขณะที่กรณีของไทย การกระจายของหัวข้อใกล้เคียงกับประเทศอื่น ๆ แต่เป็นที่น่าสังเกตว่าไทยมีสัดส่วนการกล่าวชมเรื่องทรัพยากรของมหาวิทยาลัย ชื่อเสียงและความเป็นนานาชาติ ค่อนข้างต่ำ ขณะที่มีสัดส่วนการบ่นเรื่องชื่อเสียงและความเป็นนานาชาติ รูปแบบการสอนและบุคลากร ค่อนข้างสูงเมื่อเทียบกับประเทศอื่น ๆ

การวิเคราะห์ความสัมพันธ์ระหว่างประเด็นที่มหาวิทยาลัยถูกกล่าวถึงและความรู้สึกที่ผู้เขียนข้อความเห็นมีต่อมหาวิทยาลัย โดยใช้การถดถอยบีตา (Beta Regression) เริ่มจากการแปลงข้อมูลเป็นรายมหาวิทยาลัย และพิจารณาการกระจายของข้อมูลด้วยการใช้รูปแสดงการแจกแจงและการทดสอบภาวะสารูปดี (Goodness of Fit Test) โดยใช้ Kolmogorov–Smirnov test พบว่าข้อมูลข้อความเห็นเชิงบวกควรแบ่งมหาวิทยาลัยออกเป็น 2 กลุ่ม คือ กลุ่มที่ถูกกล่าวชื่นชมในระดับที่ไม่สูงมาก และกลุ่มที่ถูกกล่าวชื่นชมในระดับที่สูงมาก (ใช้เกณฑ์แบ่งที่ค่าเฉลี่ยคะแนนความรู้สึกในเชิงบวกเท่ากับ 0.7)

ตัวแปรต้น คือ ร้อยละความถี่ที่หัวข้อทั้ง 9 ถูกกล่าวถึง ซึ่งทั้งหมดล้วนแต่เป็นประเด็นที่สำคัญที่มีการเขียนถึงบ่อยในการกล่าวถึงมหาวิทยาลัย การทดสอบระดับนัยสำคัญทางสถิติในการถดถอยบีตาจึงเป็นการบอกปัจจัยที่กำหนดความแปรปรวนของค่าคะแนนความรู้สึกในกลุ่มต่าง ๆ ซึ่งหมายถึงหากมหาวิทยาลัยจะพัฒนาคะแนนความรู้สึกที่มีต่อมหาวิทยาลัย มหาวิทยาลัยกลุ่มนั้น ๆ ควรจะเน้นในหัวข้อไหน

เมื่อทำการวิเคราะห์การกระทบของแต่ละประเด็นต่อคะแนนความรู้สึก พบว่า

- ประสบการณ์โดยภาพรวมและการเรียน: มีผลบวกต่อคะแนนความรู้สึกเชิงบวกทั้งในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก และกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ความท้าทายทางวิชาการและภาระงาน: มีผลลบต่อคะแนนความรู้สึกเชิงบวกทั้งในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก และกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก
- การใช้ชีวิตและกิจกรรม: มีผลลบต่อคะแนนความรู้สึกเชิงบวกเฉพาะในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก มีผลบวกต่อคะแนนความรู้สึกเชิงลบ (รู้สึกลบน้อยลง)
- รูปแบบการสอนและบุคลากร: มีผลลบต่อคะแนนความรู้สึกเชิงบวกในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก มีผลบวกต่อคะแนนความรู้สึกเชิงบวกในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก และกรณีข้อความเชิงลบ
- ทรัพยากรของมหาวิทยาลัย: มีผลลบต่อคะแนนความรู้สึกเชิงบวกทั้งในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก และกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก
- ค่าใช้จ่ายและต้นทุนการใช้ชีวิต: มีผลบวกต่อคะแนนความรู้สึกเชิงบวกในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับสูงมาก
- ทรัพยากรของมหาวิทยาลัย: มีผลลบต่อคะแนนความรู้สึกเชิงบวกในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมทั้งในระดับไม่สูงมากและสูงมาก
- ชื่อเสียงและความเป็นนานาชาติ: มีผลบวกต่อคะแนนความรู้สึกเชิงลบ
- ประสบการณ์ด้านภาษาและวัฒนธรรม: มีผลลบต่อคะแนนความรู้สึกเชิงบวกเฉพาะในกรณีมหาวิทยาลัยที่ได้รับความชื่นชมในระดับไม่สูงมาก มีผลบวกต่อคะแนนความรู้สึกเชิงลบ

ช่วงก่อนและหลัง ค.ศ. 2019 ที่เกิดเหตุกรณีการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ซึ่งทำให้เกิดการเปลี่ยนแปลงต่อรูปแบบการเรียนการสอนทั่วโลก กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยที่ได้รับคำชมในระดับไม่สูงมากจะพบว่าก่อนและหลัง ค.ศ. 2019 ปัจจัยที่ส่งผลต่อคะแนนความรู้สึกเชิงบวกจะไม่แตกต่างกันมาก ขณะที่กรณีข้อความเห็นเชิงบวกของมหาวิทยาลัยที่ได้รับคำชมในระดับสูง หัวข้อเรื่องรูปแบบการสอนและบุคลากร ค่าใช้จ่ายและต้นทุนการใช้ชีวิต และทรัพยากรของมหาวิทยาลัย ช่วงก่อน ค.ศ. 2019 จะเป็นปัจจัยที่ส่งผลต่อคะแนนความรู้สึกต่อมหาวิทยาลัยอย่างมีนัยสำคัญทางสถิติ แต่หลัง ค.ศ. 2019 กลับไม่เป็นปัจจัยที่ส่งผลต่อคะแนนความรู้สึกต่อมหาวิทยาลัยอย่างมีนัยสำคัญทางสถิติ

ขณะที่กรณีข้อความเห็นเชิงลบ หลัง ค.ศ. 2019 มีการกล่าวถึง ประสบการณ์โดยภาพรวมและการเรียน จะทำให้ค่าคะแนนความรู้สึกมีค่าเพิ่มขึ้น (รู้สึกลบน้อยลง) ทั้ง ๆ ที่ก่อนหน้า ค.ศ. 2019 การกล่าวถึงปัจจัยด้านนี้ไม่ส่งผลต่อคะแนนความรู้สึก อีกด้านหนึ่งคือ สถานที่และชีวิตในมหาวิทยาลัย การใช้ชีวิตและกิจกรรม ประสบการณ์ด้านภาษาและวัฒนธรรม ที่เคยมีความสำคัญช่วยให้คะแนนความรู้สึกเชิงลบไม่ต่ำมากในช่วงก่อน ค.ศ. 2019 พอหลัง ค.ศ. 2019 ปัจจัยเหล่านี้ไม่ส่งผลสำคัญอีกต่อไป ส่วนหนึ่งน่าจะเกิดจากการปรับรูปแบบการเรียนการสอนทำให้นักเรียนพบปะกันน้อยลง

เมื่อพิจารณาผลการวิเคราะห์ตามกลุ่มประเทศ ถ้าประเทศต้องการเพิ่มคะแนนความรู้สึกที่ผู้คนมีต่อมหาวิทยาลัย แต่ละประเทศจะมีจุดเน้นที่แตกต่างกัน กรณีของสหราชอาณาจักร คือการเน้นส่งเสริมด้านประสบการณ์โดยภาพรวมและการเรียน เน้นแก้ไขด้านประสบการณ์ด้านภาษาและ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วัฒนธรรม กรณีของสหรัฐ คือการเน้นส่งเสริมด้านรูปแบบการสอบและบุคลากร เน้นแก้ไขด้านประสบการณ์ความท้าทายทางวิชาการและภาระงาน กรณีของออสเตรเลีย คือการเน้นส่งเสริมด้านการใช้ชีวิตและกิจกรรม เน้นแก้ไขด้านความท้าทายทางวิชาการและภาระงาน กรณีของแคนาดา ไม่มีด้านไหนที่ต้องเน้นเป็นพิเศษ เน้นแก้ไขด้านการใช้ชีวิตและกิจกรรม กรณีของจีน คือการเน้นส่งเสริมด้านประสบการณ์โดยภาพรวมและการเรียน เน้นแก้ไขด้านการใช้ชีวิตและกิจกรรม กรณีของญี่ปุ่น คือการเน้นส่งเสริมด้านประสบการณ์ด้านภาษาและวัฒนธรรม เน้นแก้ไขด้านรูปแบบการสอนและบุคลากร

การศึกษานี้แตกต่างจากการศึกษาในอดีตอื่น ที่มักจะเน้นวิเคราะห์ข้อความเห็นต่อสิ่งอื่นที่ไม่ใช่มหาวิทยาลัย เช่น ข้อความเห็นต่อสินค้าในระบบอีคอมเมิร์ซ ข้อความเห็นต่อการเรียนการสอนแบบออนไลน์ ทั้งนี้ หากเปรียบเทียบกับการศึกษาที่เน้นศึกษามหาวิทยาลัยจากชุดข้อมูลและวิธีการอื่น ๆ ข้อค้นพบในการศึกษานี้โดยภาพรวมจะให้ผลการศึกษาสอดคล้องกัน โดยพบว่าการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 ส่งผลอย่างมีนัยสำคัญต่อการเรียนการสอนในระดับมหาวิทยาลัย และมหาวิทยาลัยในแต่ละประเทศมีรูปแบบการจัดบริการการศึกษาที่มีจุดเน้นที่แตกต่างกัน

5.2 ข้อเสนอแนะ

การศึกษานี้วิเคราะห์ข้อความความเห็นของผู้คนต่อมหาวิทยาลัยและพิจารณาประเด็นที่มหาวิทยาลัยมักจะถูกกล่าวถึง ผ่านการทำตัวแบบหัวข้อ ซึ่งสามารถกำหนดหัวข้อที่มหาวิทยาลัยกล่าวถึงได้ 9 หัวข้อ อย่างไรก็ตาม ในโลกแห่งความเป็นจริงมหาวิทยาลัยอาจจะถูกกล่าวถึงในประเด็นมากกว่านี้ และในแต่ละหัวข้อก็อาจจะยังสามารถแบ่งเป็นหัวข้อย่อย ๆ ได้อีก ซึ่งจะให้ความรู้ที่ละเอียดมากกว่าขอบเขตของการศึกษานี้ได้

การวิเคราะห์ข้อมูลเกี่ยวกับการศึกษาที่มองในระดับมหาวิทยาลัยของแต่ละประเทศ ก็อาจจะเป็นมุมมองที่ยังคงค่อนข้างกว้าง หากผู้ที่สนใจศึกษาต้องการข้อมูลในรายละเอียดมากกว่านี้ อาจจะใช้วิธีการเทียบเคียงรายหลักสูตรเดียวกันของแต่ละประเทศ ทั้งนี้ เนื่องจากข้อจำกัดของข้อมูลที่มีจึงยังไม่สามารถดำเนินการเช่นนี้ได้ เพราะไม่ใช่ทุกแหล่งข้อมูลที่จะระบุหลักสูตรที่ผู้เขียนข้อความเห็นกำลังศึกษาหรือสำเร็จการศึกษามา

จำนวนตัวอย่างของกรณีประเทศไทยยังมีน้อยมากเมื่อเทียบกับกรณีของประเทศอื่น ๆ ส่วนหนึ่งเพราะแหล่งที่มาของข้อมูล เป็นแหล่งข้อมูลที่เป็นภาษาอังกฤษ ทำให้มีข้อมูลจากประเทศที่ไม่ได้ใช้ภาษาอังกฤษน้อย โดยในขั้นตอนการศึกษาผู้วิจัยพยายามรวบรวมข้อมูลจากแหล่งอื่น ๆ แล้ว เช่น บนสื่อสังคมออนไลน์ต่าง ๆ แต่ก็ไม่พบแหล่งข้อมูลใดที่ให้ข้อมูลที่จะสามารถนำมาใช้ได้ตามวัตถุประสงค์การศึกษา โดยหากจะวิเคราะห์กรณีของไทยเพิ่มเติมอาจจะทำได้โดยการเก็บข้อมูลปฐมภูมิเพิ่ม เช่น การเก็บแบบสอบถาม หรือ การสำรวจเพิ่มเติมจากแหล่งข้อมูลที่มีอยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

มหาวิทยาลัยธรรมศาสตร์ (2565) อาจารย์ธรรมศาสตร์ ชำแหละผลกระทบ ‘วิกฤตเด็กไทยเกิดน้อย’ เขย่าสังคมไทย. Available at: https://tu.ac.th/thammasat-310165-crisis-thai-children-born-less (Accessed: 22 May 2022).

สายชล สินสมบูรณ์ทอง (2559) การวิเคราะห์ตัวแปรหลายตัว. กรุงเทพฯ: จามจุรีโปรดักส์.

Abbas, A., Ar, A.Y. and Hosseini, S. (2024) ‘A global perspective of government initiatives to support higher education during the COVID-19 pandemic: A systematic review of literature’, *Research in Globalization*, 8, p. 100202. doi:10.1016/J.RESGLO.2024.100202.

Abonazel, M.R. *et al.* (2022) ‘A New Two-Parameter Estimator for Beta Regression Model: Method, Simulation, and Application’, *Frontiers in Applied Mathematics and Statistics*, 7(January). doi:10.3389/fams.2021.780322.

Abonazel, M.R. and Taha, I.M. (2023) ‘Beta ridge regression estimators: simulation and application’, *Communications in Statistics - Simulation and Computation*, 52(9), pp. 4280–4292. doi:10.1080/03610918.2021.1960373.

Akram, M.N. *et al.* (2021) ‘A new modified ridge-type estimator for the beta regression model: simulation and application’, *AIMS Mathematics*, 7(1), pp. 1035–1057. doi:10.3934/math.2022062.

Ali Kandhro, I. *et al.* (2019) ‘Student Feedback Sentiment Analysis Model Using Various Machine Learning Schemes A Review’, *Indian Journal of Science and Technology*, 14(12), pp. 1–9. doi:10.17485/ijst/2019/v12i14/143243.

Almaleh, A. *et al.* (2019) ‘Align my curriculum: A framework to bridge the gap between acquired university curriculum and required market skills’, *Sustainability (Switzerland)*, 11(9). doi:10.3390/su11092607.

Anandarajan, M., Hill, C. and Nolan, T. (2019) *Practical Text Analytics Maximizing the Value of Text Data*. Springer International Publishing (Advances in Analytics and Data Science). doi:10.1007/978-3-319-95663-3.

Andrade, M.S. (2006) ‘International students in English-speaking universities’, *Journal of Research in International Education*, 5(2), pp. 131–154. doi:10.1177/1475240906065589.

Araya, D. (2014) *Higher education in the global age : policy, practice and promise in emerging societies*. Routledge New York. doi:https://worldcat.org/title/844372735.

Bahdanau, D., Cho, K. and Bengio, Y. (2015) ‘Neural Machine Translation by Jointly Learning to Align and Translate’, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15. Available at: http://arxiv.org/abs/1409.0473.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Barrón Estrada, M.L. *et al.* (2020) 'Opinion mining and emotion recognition applied to learning environments', *Expert Systems with Applications*, 150, p. 113265. doi:10.1016/j.eswa.2020.113265.

Belyadi, H. and Haghighat, A. (2021) 'Unsupervised machine learning: clustering algorithms', in *Machine Learning Guide for Oil and Gas Using Python*. Elsevier, pp. 125–168. doi:10.1016/B978-0-12-821929-4.00002-0.

Bengio, Y., Ducharme, R. and Vincent, P. (2000) 'A Neural Probabilistic Language Model', in Leen, T., Dietterich, T., and Tresp, V. (eds) *Advances in Neural Information Processing Systems*. MIT Press. Available at: https://proceedings.neurips.cc/paper/2000/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf.

Biswas, B. *et al.* (2022) 'A critical assessment of consumer reviews: A hybrid NLP-based methodology', *Decision Support Systems*, 159(May 2021), p. 113799. doi:10.1016/j.dss.2022.113799.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) 'Latent dirichlet allocation', *Journal of Machine Learning Research*, 3, pp. 993–1022.

Chaubard, F. *et al.* (2019) 'CS224n: Natural Language Processing with Deep Learning Lecture Notes: Part I Word Vectors I: Introduction, SVD and Word2Vec'. Available at: https://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes01-wordvecs1.pdf.

Chua, A.Y.K. and Banerjee, S. (2015) 'Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth', *Journal of the Association for Information Science and Technology*, 66(2), pp. 354–362. doi:10.1002/asi.23180.

Collobert, R. and Weston, J. (2008) 'A unified architecture for natural language processing', in *Proceedings of the 25th international conference on Machine learning - ICML '08*. New York, New York, USA: ACM Press, pp. 160–167. doi:10.1145/1390156.1390177.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) 'Maximum Likelihood from Incomplete Data Via the EM Algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), pp. 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x.

Du, J. *et al.* (2021) 'Neighbor-aware review helpfulness prediction', *Decision Support Systems*, 148, p. 113581. doi:10.1016/j.dss.2021.113581.

Duraivel, S., Lavanya, L. and Augustine, A. (2022) 'Understanding Vaccine Hesitancy with Application of Latent Dirichlet Allocation to Reddit Corpora', *Indian Journal Of Science And Technology*, 15(37), pp. 1868–1875. doi:10.17485/IJST/v15i37.687.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Elsevier (2021) *University rankings guide: A closer look for research leaders* . Available at: https://www.elsevier.com/__data/assets/pdf_file/0017/1113920/university-rankings-guide-aug-2021.pdf (Accessed: 22 May 2022).

Estrada Guillén, M., Monferrer Tirado, D. and Rodríguez Sánchez, A. (2022) 'The impact of COVID-19 on university students and competences in education for sustainable development: Emotional intelligence, resilience and engagement', *Journal of Cleaner Production*, 380, p. 135057. doi:10.1016/J.JCLEPRO.2022.135057.

Ferrari, S. and Cribari-Neto, F. (2004) 'Beta Regression for Modelling Rates and Proportions', *Journal of Applied Statistics*, 31(7), pp. 799–815. doi:10.1080/0266476042000214501.

Ferrer, J. *et al.* (2023) 'Analyzing the impact of COVID-19 on the grades of university education: A case study with economics students', *Social Sciences & Humanities Open*, 7(1), p. 100428. doi:10.1016/J.SSAHO.2023.100428.

Filieri, R. *et al.* (2018) 'Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services', *Information & Management*, 55(8), pp. 956–970. doi:10.1016/j.im.2018.04.010.

Ghose, A. and Ipeirotis, P.G. (2011) 'Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics', *IEEE Transactions on Knowledge and Data Engineering*, 23(10), pp. 1498–1512. doi:10.1109/TKDE.2010.188.

Giang, N.T.P., Dien, T.T. and Khoa, T.T.M. (2020) 'Sentiment Analysis for University Students' Feedback', *Advances in Intelligent Systems and Computing*, 1130 AISC(October), pp. 55–66. doi:10.1007/978-3-030-39442-4_5.

Griffiths, T.L. and Steyvers, M. (2004) 'Finding scientific topics', *Proceedings of the National Academy of Sciences*, 101(suppl_1), pp. 5228–5235. doi:10.1073/pnas.0307752101.

Grimmer, J., Roberts, M.E. and Stewart, B.M. (2022) 'Text as Data: A New Framework for Machine Learning and the Social Sciences'. doi:https://www.amazon.com/Text-Data-Framework-Learning-Sciences/dp/0691207550.

Han, J., Kamber, M. and Pei, J. (2012a) 'Cluster Analysis', in *Data Mining*. Elsevier, pp. 443–495. doi:10.1016/B978-0-12-381479-1.00010-1.

Han, J., Kamber, M. and Pei, J. (2012b) 'Getting to Know Your Data', in *Data Mining*. Elsevier, pp. 39–82. doi:10.1016/B978-0-12-381479-1.00002-2.

Hoffman, M., Bach, F. and Blei, D. (2010) 'Online Learning for Latent Dirichlet Allocation', in Lafferty, J. et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

https://proceedings.neurips.cc/paper_files/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf.

Huai, S. and Van de Voorde, T. (2022) 'Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods', *Landscape and Urban Planning*, 218(November 2021), p. 104307. doi:10.1016/j.landurbplan.2021.104307.

Jurafsky, D. and Martin, J.H. (2009) *Speech and Language Processing*. 2nd edn. USA: Prentice-Hall.

Kao, A. and Poteet, S.R. (2007) *Natural Language Processing and Text Mining*. London: Springer. doi:10.1007/978-1-84628-754-1.

Karabiber, F. (2023) *Cosine Similarity*. Available at: https://www.learndatasci.com/glossary/cosine-similarity/ (Accessed: 2 January 2022).

Karlsson, P., Månsson, K. and Kibria, B.M.G. (2020) 'A Liu estimator for the beta regression model and its application to chemical data', *Journal of Chemometrics*, 34(10). doi:10.1002/cem.3300.

Khurana, D. *et al.* (2022) 'Natural language processing: state of the art, current trends and challenges', *Multimedia Tools and Applications* [Preprint]. doi:10.1007/s11042-022-13428-4.

Kotu, V. and Deshpande, B. (2019) 'Classification', in *Data Science*. Elsevier, pp. 65–163. doi:10.1016/B978-0-12-814761-0.00004-6.

Kuan, K. *et al.* (2015) 'What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems', *Journal of the Association for Information Systems*, 16(1), pp. 48–71. doi:10.17705/1jais.00386.

Li, L. *et al.* (2022) 'Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit', *Computers & Education*, 176(October 2021), p. 104354. doi:10.1016/j.compedu.2021.104354.

Li, X. *et al.* (2019) 'A Shallow BERT-CNN Model for Sentiment Analysis on MOOCs Comments', *TALE 2019 - 2019 IEEE International Conference on Engineering, Technology and Education* [Preprint]. doi:10.1109/TALE48000.2019.9225993.

Loyalka, P. *et al.* (2019) 'Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program', *American Economic Journal: Applied Economics*, 11(3), pp. 128–154. doi:10.1257/app.20170226.

Meek, S., Wilk, V. and Lambert, C. (2021) 'A big data exploration of the informational and normative influences on the helpfulness of online restaurant reviews', *Journal of Business Research*, 125, pp. 354–367.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

doi:10.1016/j.jbusres.2020.12.001.

Mezzanzanica, M. (2020) 'Italian Web Job Vacancies for Marketing-Related Professions', *Symphonya. Emerging Issues in Management*, pp. 110–124. doi:10.4468/2015.3.14mezzanzanica.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., *et al.* (2013) 'Distributed Representations of Words and Phrases and their Compositionality', in Burges, C.J. et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., *et al.* (2013) 'Distributed Representations of Words and Phrases and their Compositionality', *Advances in Neural Information Processing Systems*, pp. 1–9. Available at: http://arxiv.org/abs/1310.4546.

ONAN, A. (2021) 'Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach', *Computer Applications in Engineering Education*, 29(3), pp. 572–589. doi:10.1002/cae.22253.

Pejic-Bach, M. *et al.* (2020) 'Text mining of industry 4.0 job advertisements', *International Journal of Information Management*, 50(July), pp. 416–431. doi:10.1016/j.ijinfomgt.2019.07.014.

Qasim, M., Månsson, K. and Golam Kibria, B.M. (2021) 'On some beta ridge regression estimators: method, simulation and application', *Journal of Statistical Computation and Simulation*, 91(9), pp. 1699–1712. doi:10.1080/00949655.2020.1867549.

Röder, M., Both, A. and Hinneburg, A. (2015) 'Exploring the Space of Topic Coherence Measures', in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. New York, NY, USA: ACM (WSDM '15), pp. 399–408. doi:10.1145/2684822.2685324.

Rousseeuw, P.J. (1987) 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics*, 20(C), pp. 53–65. doi:10.1016/0377-0427(87)90125-7.

Spada, I. *et al.* (2022) 'Are universities ready to deliver digital skills and competences? A text mining-based case study of marketing courses in Italy', *Technological Forecasting and Social Change*, 182, p. 121869. doi:10.1016/j.techfore.2022.121869.

Špeh, J., Muhič, A. and Rupnik, J. (2013) 'Algorithms of the LDA model', pp. 1–5. Available at: http://arxiv.org/abs/1307.0317.

Sutskever, I., Vinyals, O. and Le, Q. V (2014) 'Sequence to Sequence Learning with

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Neural Networks', in Ghahramani, Z. et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf.

Swearingen, C.J. *et al.* (2011) 'Application of Beta Regression to Analyze Ischemic Stroke Volume in NINDS rt-PA Clinical Trials', *Neuroepidemiology*, 37(2), pp. 73–82. doi:10.1159/000330375.

Syed, S. and Spruit, M. (2017) 'Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation', in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, pp. 165–174. doi:10.1109/DSAA.2017.61.

Tibshirani, R., Walther, G. and Hastie, T. (2001) 'Estimating the number of clusters in a data set via the gap statistic', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), pp. 411–423. doi:10.1111/1467-9868.00293.

Tomej, K. *et al.* (2022) 'Blended and (not so) splendid teaching and learning: Higher education insights from university teachers during the Covid-19 pandemic', *International Journal of Educational Research Open*, 3, p. 100144. doi:10.1016/J.IJEDRO.2022.100144.

Vaswani, A. *et al.* (2017) 'Attention is all you need', *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), pp. 5999–6009.

Wu, C., Mai, F. and Li, X. (2021) 'The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model', *Information & Management*, 58(2), p. 103408. doi:10.1016/j.im.2020.103408.

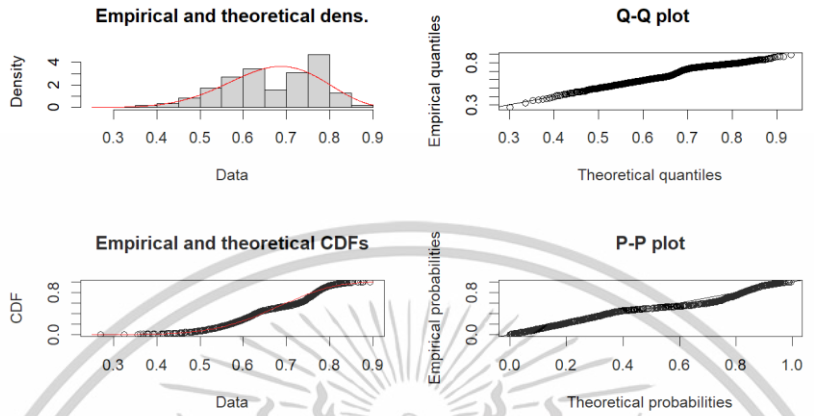
Yang, S., Zhou, C. and Chen, Y. (2021) 'Do topic consistency and linguistic style similarity affect online review helpfulness? An elaboration likelihood model perspective', *Information Processing & Management*, 58(3), p. 102521. doi:10.1016/j.ipm.2021.102521.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
การแจกแจงของตัวแปรคะแนนความรู้สึกของข้อความ

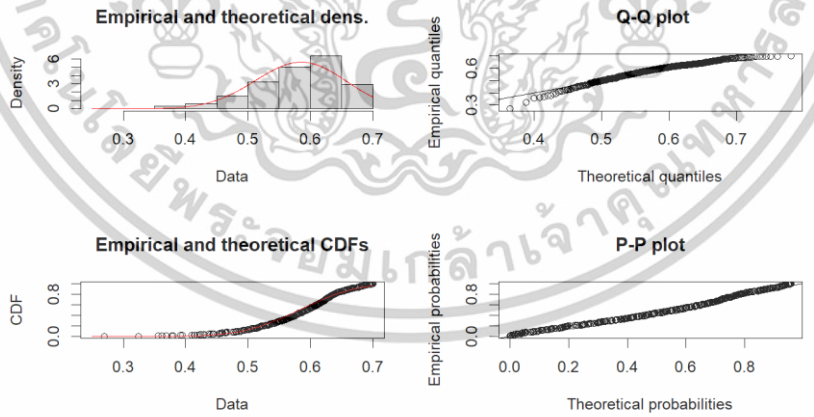


รูปที่ ก.1 รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ภาพซ้ายบน
   กราฟเส้น คือ ค่าความถี่จากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด
   กราฟแท่ง คือ ความถี่ของข้อมูลจริง



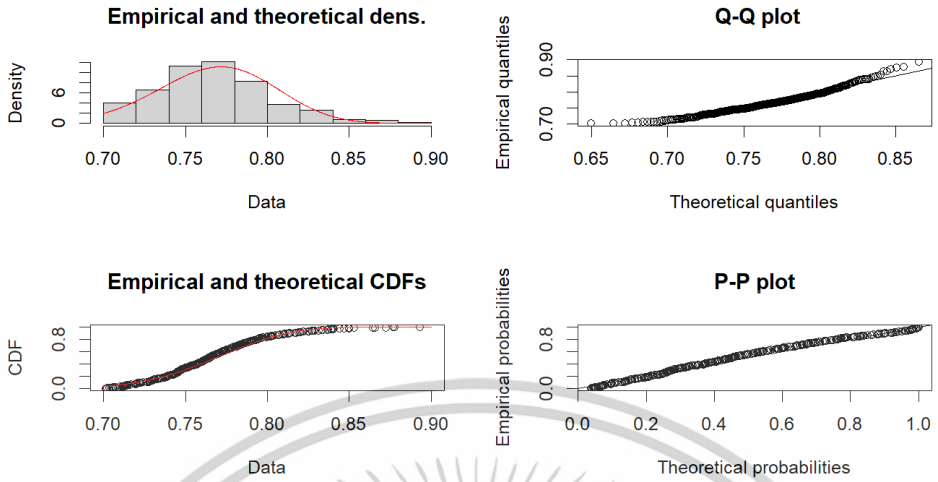
รูปที่ ก.2 รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก มีค่าน้อยกว่า 0.7

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ภาพซ้ายบน
   กราฟเส้น คือ ค่าความถี่จากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด
   กราฟแท่ง คือ ความถี่ของข้อมูลจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

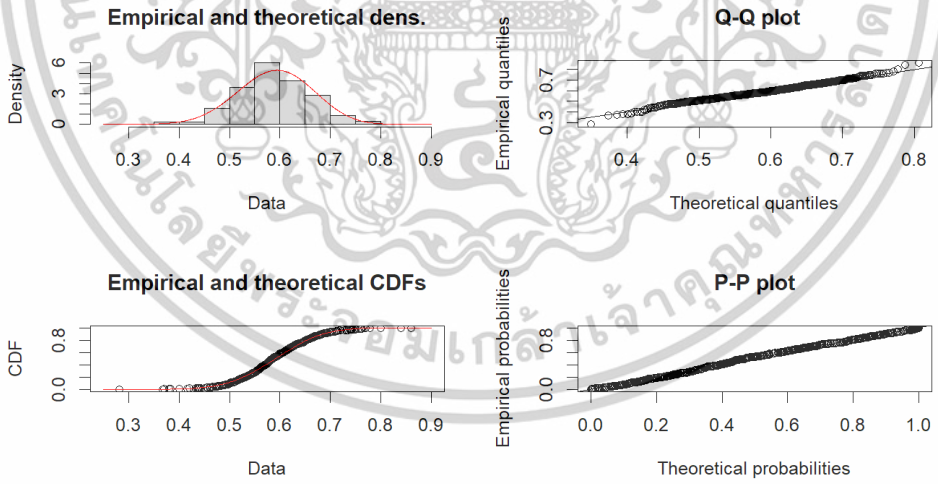


รูปที่ ก.3 รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงบวก มีค่ามากกว่าหรือเท่ากับ 0.7

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ภาพซ้ายบน
   กราฟเส้น คือ ค่าความถี่จากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด
   กราฟแท่ง คือ ความถี่ของข้อมูลจริง



รูปที่ ก.4 รูปแสดงการแจกแจงของข้อมูลจริงและการแจกแจงที่ประมาณการ กรณีค่าคะแนนความรู้สึกที่มีต่อมหาวิทยาลัยเชิงลบ

ที่มา: ผู้วิจัย

หมายเหตุ:

1. ภาพซ้ายบน
   กราฟเส้น คือ ค่าความถี่จากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด
   กราฟแท่ง คือ ความถี่ของข้อมูลจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้วิจัย

ชื่อ	นายวุฒิพงษ์ ตุ้นยุทธ์
วัน เดือน ปีเกิด	23 เมษายน พ.ศ. 2534
ที่อยู่ปัจจุบัน	กรุงเทพมหานคร ประเทศไทย
ประวัติการศึกษา	(2555) เศรษฐศาสตรบัณฑิต สาขาเศรษฐศาสตร์ มหาวิทยาลัยสงขลานครินทร์
	(2564) เศรษฐศาสตรมหาบัณฑิต สาขาเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์
	(2568) วิทยาศาสตรมหาบัณฑิต สาขาสถิติประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ทุนการศึกษาที่ได้รับ	1. ทุนการศึกษาระดับปริญญาตรีสาขาเศรษศาสตร์ ปีการศึกษา 2555 บริษัท ทีโอที จำกัด (มหาชน)
	2. ทุนการศึกษาระดับบัณฑิตศึกษาคณะวิทยาศาสตร์ ปีการศึกษา 2564-65 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ผลงานทางวิชาการ	1. Tunyut, W., & Wanitjirattikal, P. (2024). Simplified Approach to Constructing Coherent Topics and Subtopics from Text Data: A Case Study Using University Reviews. 2024 10th International Conference on Big Data and Information Analytics (BigDIA), BigDIA, 848–853. https://doi.org/10.1109/BigDIA63733.2024.10808587
	2. Tunyut, W. (2022). Intrahousehold Resources Allocation in Thailand: Sharing Rules and Determinants. Journal of Economics, Business and Management, 10(1), 29–33. https://doi.org/10.18178/joebm.2022.10.1.669
	3. NaRanong, V., & Tunyut, W. (2019). The Human Rights of Migrant Workers in the Thai Poultry Industry: The Regulatory Gaps and the Guidelines for the Promotion and Protection. TDRI QUARTERLY REVIEW, 34(3). https://tdri.or.th/2020/01/tdri-quarterly-review-september-2019/
	4. วุฒิพงษ์ ตุ้นยุทธ์. (31 ต.ค. 2566). การวัดค่าใช้จ่ายรายบุคคลและความเหลื่อมล้ำที่เกิดภายในครัวเรือนไทย. เศรษฐสาร. คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์. https://setthasarn.econ.tu.ac.th/blog/detail/663
	5. วิโรจน์ ณ ระนอง, วุฒิพงษ์ ตุ้นยุทธ์, อรุณ สถิตพงศ์สถาพร, บุญวรา สุมะโน เจนพึ่งพร, และ ณัฐนี ณ เชียงใหม่. (2564). การพัฒนาสวัสดิการด้านการคลอดบุตรของกองทุนประกันสังคม. รายงานทีดีอาร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไอ, 180. https://tdri.or.th/2021/11/finding-the-appropriate-maternal-and-newborn-health-benefits-for-the-social-security-fund-of-thailand/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้