การจำแนกประเภทแบบหลายฉลากของความคิดเห็นนักท่องเที่ยวต่างชาติในการพัฒนาการท่องเที่ยวไทย

MULTI-LABEL CLASSIFICATION OF FOREIGN TOURISTS' OPINIONS ON THAILAND TOURISM DEVELOPMENT

ลลิตา สุระมรรคา

LALITA SURAMANKA

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2567

KMITL-2024-SC-M-002-024

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MULTI-LABEL CLASSIFICATION OF FOREIGN TOURISTS' OPINIONS ON THAILAND TOURISM DEVELOPMENT

LALITA SURAMANKA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024
KMITL-2024-SC-M-002-024

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2024
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การจำแนกประเภทแบบหลายฉลากของความคิดเห็นนักท่องเที่ยวต่างชาติในการพัฒนาการท่องเที่ยวไทย
ชื่อนักศึกษา	นางสาวลลิตา สุระมรรคา
รหัสประจำตัว	61605043
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2567
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.อนันตพร หรรษคุณาฒัย

บทคัดย่อ

การพัฒนาคุณภาพการท่องเที่ยวในประเทศไทยด้วยการทำความเข้าใจและใช้ประโยชน์จากความคิดเห็นของนักท่องเที่ยวต่างชาตินั้นเป็นเรื่องที่ท้าทายเนื่องจากปริมาณข้อมูลที่มีจำนวนมาก งานวิจัยนี้เสนอวิธีการสองส่วนเพื่อแก้ปัญหาดังกล่าว (1) การพัฒนาแบบจำลองการจำแนกประเภทความคิดเห็น และ (2) การวิเคราะห์ความคิดเห็นของนักท่องเที่ยวผ่านแดชบอร์ด ข้อมูลที่ใช้ในงานวิจัยนี้รวบรวมโดยการท่องเที่ยวแห่งประเทศไทย (ททท.) ซึ่งประกอบด้วยความคิดเห็นจากนักท่องเที่ยวต่างชาติเกี่ยวกับด้านที่การท่องเที่ยวของไทยควรปรับปรุง โดยรวบรวมความคิดเห็นได้ทั้งหมด 2,249 ความคิดเห็น ผลการทดลองแสดงให้เห็นว่าการใช้เทคนิคการเสริมข้อมูล การเลือกคุณลักษณะ การแปลงข้อมูลแบบหลายฉลาก ด้วยวิธี Classifier Chains และแบบจำลองการจำแนกประเภทด้วยวิธีการสุ่มแบบป่า กับชุดข้อมูลสำหรับการฝึกอบรม ให้ผลลัพธ์สูงที่สุด โดยมีความถูกต้อง 80% ความเที่ยงตรง 90% ความระลึก 81% F1-Score 85% และ Hamming loss 0.04 และผลการวิเคราะห์จากแดชบอร์ดแสดงให้เห็นประเด็นสำคัญสามอันดับแรกที่ควรปรับปรุง ได้แก่ การสื่อสาร/ภาษา การจราจร/ระบบขนส่งสาธารณะ และความสะอาด/สุขอนามัย

คำสำคัญ : การจำแนกหลายป้ายกำกับ การท่องเที่ยวไทย การทำเหมืองความคิดเห็น การเลือกคุณลักษณะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Multi-Label Classification of Foreign Tourists' Opinions on Thailand Tourism Development
Student Name	Ms. Lalita Suramanka
Student ID	61605043
Degree	Master of Science (Computer Science)
Department	Computer Science
Year	2024
Thesis Advisor	Asst.Prof. Dr. Anantaporn Hanskunatai

บทคัดย่อ

The enhancement of tourism quality in Thailand through the understanding and utilization of foreign tourists' opinions presents challenges due to the extensive volume of data involved. This research proposes a two-fold approach to address this issue: (1) the development of an opinion classification model, and (2) the analysis of tourists' opinions through a dashboard. A dataset, compiled by the Tourism Authority of Thailand (TAT) and consisting of opinions from foreign tourists regarding areas for improvement in Thai tourism, was utilized. A total of 2,249 comments were collected. Experimental results demonstrate that the use of data augmentation, feature selection, Multi-label transformation using Classifier Chains, and the Random Forest classification model on the training dataset yields promising results with an accuracy rate of 80%, precision of 90%, recall of 81%, F1-score of 85%, and Hamming loss of 0.04. Analysis from the dashboard revealed the top three key areas for improvement: communication/language, traffic/public transportation, and cleanliness/hygiene.

คำสำคัญ : Multi-label classification, Thai tourism, Opinion mining, Feature selection.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีจากแรงสนับสนุน คำแนะนำ ความช่วยเหลือ และกำลังใจจากบุคคลหลายฝ่าย ผู้วิจัยจึงใคร่ขอใช้เนื้อหาในส่วนของกิตติกรรมประกาศเพื่อขอขอบพระคุณเป็นอย่างสูงมา ณ ที่นี้

ข้าพเจ้าขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร. อนันตพร หรรษคุณาฒัย อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยให้คำแนะนำในการทำวิจัยเป็นอย่างดีจนทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์ฉบับนี้ได้สำเร็จลุล่วง

ข้าพเจ้าขอขอบคุณ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. อนุชิต จิตพัฒนกุล และกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. ปัทมา เจริญพร ที่ได้สละเวลาในการให้ข้อเสนอแนะ และนำสิ่งที่ควรเพิ่มเติมลงในวิทยานิพพนธ์ฉบับนี้จนเสร็จสมบูรณ์

ข้าพเจ้าขอขอบคุณ เพื่อน ๆ ทุกคน แฟนและครอบครัว ที่ได้ช่วยเหลือ ส่งเสริม สนับสนุน และเป็นกำลังใจตลอดช่วงเวลาการทำวิทยานิพนธ์

ข้าพเจ้าขอขอบคุณ อาจารย์ทุกท่าน ที่ได้สั่งสอนข้าพเจ้ามา ซึ่งความรู้แนวคิด และกระบวนการต่าง ๆ ที่ได้รับมา ล้วนมีประโยชน์ต่อวิทยานิพนธ์ฉบับนี้ทั้งสิ้น

ข้าพเจ้าขอขอบคุณ การท่องเที่ยวแห่งประเทศไทย (ททท.) สำหรับการสนับสนุนและการให้ข้อมูลที่เป็นประโยชน์อย่างยิ่งสำหรับการวิจัย

สุดท้ายนี้ข้าพเจ้าขอขอบคุณบิดา มารดาที่คอยอยู่เคียงข้างข้าพเจ้า และเป็นกำลังใจให้ข้าพเจ้าเสมอมา จนทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์ฉบับนี้ได้ลุล่วง และสำเร็จการศึกษาได้ในที่สุด

นางสาวลลิตา สุระมรรคา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของงานวิจัย	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ขั้นตอนการศึกษาและดำเนินการวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 การเสริมข้อมูล (Data Augmentation)	4
2.2 การเตรียมข้อมูลข้อความ (Text Preprocessing)	4
2.3 การเลือกคุณลักษณะของข้อมูล (Feature selection)	6
2.4 การเรียนรู้ของเครื่อง (Machine Learning Algorithm)	6
2.4.1 การจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification)	7
2.4.2 การจำแนกประเภท (Classification)	9
2.5 K-fold Cross Validation	12
2.6 การวัดประสิทธิภาพ	13
2.6.1 ความถูกต้อง(Accuracy)	13
2.6.2 ความเที่ยงตรง (Precision)	13
2.6.2 ความระลึก (Recall)	14
2.6.2 F1-Score	14
2.6.2 Hamming Loss	14
2.7 งานวิจัยที่เกี่ยวข้อง	14
บทที่ 3 วิธีดำเนินงานวิจัย	16
3.1 ชุดข้อมูล	17
3.2 การเสริมข้อมูล (Data Augmentation)	19
3.3 การเตรียมข้อมูล (Data Preprocessing)	20
3.3.1 การทำความสะอาดข้อมูล (Cleaning Data)	20
3.3.2 การแปลงข้อความเป็นตัวอักษรพิมพ์เล็ก (Lowercasing)	20
3.3.3 การลบคำที่ไม่มีความหมาย (Stop Word Removal)	20
3.3.4 การตัดคำ (Stemming)	20
3.3.5 TF-IDF (Term Frequency-Inverse Document Frequency)	20
3.3.6 การเลือกคุณลักษณะ (Feature Selection)	20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

3.4 การพัฒนาแบบจำลอง (Model Building) 21
3.5 การวัดประสิทธิภาพของแบบจำลอง (Model Evaluation) 22
3.6 การออกแบบส่วนประสานงานผู้ใช้ 23
บทที่ 4 ผลการวิจัยและการอภิปรายผล 24
4.1 ผลการวิจัย 24
4.1.1 เปรียบเทียบผลการทดสอบแบบจำลองด้วยชุดข้อมูลที่ต่างกัน 26
4.1.2 เปรียบเทียบผลการทดสอบแบบจำลองด้วยวิธีการแปลงข้อมูลและจำแนกข้อมูลที่ต่างกัน 32
4.1.3 ประเมินประสิทธิภาพของแบบจำลองบนชุดข้อมูลทดสอบ 34
4.2 การแสดงผลการจำแนกข้อความผ่านแดชบอร์ด 35
4.2.1 พัฒนาแดชบอร์ด 35
4.2.2 วิเคราะห์ผลการจำแนกผ่านแดชบอร์ด 37
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 39
5.1 สรุปผลการวิจัย 39
5.2 ข้อเสนอแนะ 40
เอกสารอ้างอิง 41
ภาคผนวก 43
ภาคผนวก ก งานวิจัยที่เผยแพร่ 44
ประวัติผู้เขียน 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
3.1	ตัวอย่างความคิดเห็นและหมวดหมู่ที่กำหนด	18
3.2	จำนวนความคิดเห็นของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ	19
3.3	จำนวนความคิดเห็นของชุดข้อมูลฝึกสอนหลังดำเนินการเสริมข้อมูล	20
3.4	การตั้งค่าพารามิเตอร์	21
4.1	ผลการทดลองการจำแนกประเภทข้อความในแต่ละวิธีบนชุดข้อมูลที่ต่างกัน	24
4.2	ผลการประเมินประสิทธิภาพของแบบจำลองบนชุดข้อมูลทดสอบ	34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	ตัวอย่างการเสริมข้อมูล (Data Augmentation)	4
2.2	Binary Relevance	7
2.3	Classifier Chains	8
2.4	Label Powerset	8
2.5	การถดถอยโลจิสติค (Logistic Regression : LR)	10
2.6	ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)	11
2.7	การสุ่มแบบป่า (Random Forest : RF)	12
2.8	ตัวอย่างของขั้นตอน K-fold Cross Validation	13
3.1	ขั้นตอนการดำเนินงานวิจัย	16
3.2	หมวดหมู่และจำนวนความคิดเห็น	17
3.3	ตัวอย่างโค้ดการพัฒนาแบบจำลอง	22
3.4	โครงสร้างของแอปพลิเคชัน	23
4.1	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน	26
4.2	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน	26
4.3	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน	27
4.4	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน	27
4.5	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน	28
4.6	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน	28
4.7	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน	29
4.8	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน	29
4.9	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน	30
4.10	เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน	30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

4.11 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน 31

4.12 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน 31

4.13 เปรียบเทียบค่าความถูกต้อง (Accuracy), ความเที่ยงตรง (Precision), ความระลึก (Recall), F1-Score และ Hamming Loss ด้วยวิธีการแปลงข้อมูลร่วมกับวิธีจำแนกข้อมูลแต่ละวิธี 32

4.14 รูปแบบไฟล์ที่กำหนด 35

4.15 หน้าต่างแอปพลิเคชันการนำเข้าข้อมูล 36

4.15 หน้าต่างเว็บไซต์แดชบอร์ดแสดงผลการจำแนกข้อความ 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

อุตสาหกรรมการท่องเที่ยวของไทยมีความสำคัญอย่างมากต่อเศรษฐกิจและการพัฒนาของประเทศไทย และเป็นแรงบันดาลใจให้เกิดการอนุรักษ์วัฒนธรรมและสิ่งแวดล้อม ซึ่งสนับสนุนให้ชุมชนท้องถิ่นมีส่วนร่วมและได้รับประโยชน์จากการท่องเที่ยวอย่างยั่งยืน โดยการท่องเที่ยวเป็นแหล่งรายได้หลักที่ส่งผลต่อการเติบโตและการกระจายรายได้ไปยังชุมชนท้องถิ่นผ่านการซื้อขายสินค้าและบริการที่หลากหลาย ซึ่งรายได้ส่วนใหญ่มาจากนักท่องเที่ยวต่างชาติ โดยเงินที่พวกเขาใช้จ่ายในประเทศไปสู่หลายภาคส่วน เช่น ค่าที่พัก อาหาร การเดินทาง และการซื้อของที่ระลึก ซึ่งส่งผลดีต่อเศรษฐกิจโดยรวมและทำให้เกิดโอกาสการสร้างงานให้กับประชากรในท้องถิ่น และด้วยจำนวนนักท่องเที่ยวต่างชาติที่เพิ่มขึ้นอย่างต่อเนื่อง ประเทศไทยจึงได้รับการยกย่องให้เป็นหนึ่งในจุดหมายปลายทางยอดนิยมระดับโลก (Statista Research Department, 2023) เกิดการกระตุ้นการลงทุนในธุรกิจที่เกี่ยวข้องกับการท่องเที่ยว เช่น โรงแรม รีสอร์ท ร้านอาหาร และบริการต่างๆ รวมถึงส่งเสริมการเรียนรู้และการแลกเปลี่ยนทางวัฒนธรรมระหว่างชาวต่างชาติและชาวไทย ทำให้เกิดความเข้าใจและการยอมรับทางวัฒนธรรมที่ดีขึ้น โดยรวมแล้วนักท่องเที่ยวต่างชาติมีส่วนช่วยให้อุตสาหกรรมการท่องเที่ยวของไทยเติบโตและพัฒนาอย่างต่อเนื่อง ซึ่งเป็นส่วนสำคัญของเศรษฐกิจของประเทศ ด้วยเหตุนี้การส่งเสริมและพัฒนาการท่องเที่ยวให้ตอบโจทย์นักท่องเที่ยวต่างชาติจึงมีความสำคัญอย่างยิ่งในการขับเคลื่อนเศรษฐกิจและสร้างความสัมพันธ์อันดีระหว่างประเทศ

การศึกษาความคิดเห็นของนักท่องเที่ยวต่างชาติหลังจากที่ได้ท่องเที่ยวประเทศไทยเป็นสิ่งสำคัญ เนื่องจากการทำความเข้าใจถึงความคิดเห็นเหล่านี้จะทำให้ทราบถึงสิ่งที่การท่องเที่ยวของไทยควรปรับปรุงให้ดียิ่งขึ้น ตลอดจนเป็นแนวทางในการพัฒนาการท่องเที่ยวให้ตรงกับความต้องการของนักท่องเที่ยวมากยิ่งขึ้น

ในปี 2563 และปี 2565 การท่องเที่ยวแห่งประเทศไทย (ททท.) ได้เก็บรวบรวมความคิดเห็นของนักท่องเที่ยวต่างชาติที่เดินทางเข้ามาท่องเที่ยวประเทศไทย โดยการสำรวจผ่านแบบสอบถามเกี่ยวกับความคิดเห็นสิ่งที่ต้องการให้การท่องเที่ยวของไทยปรับปรุงและพัฒนาให้ดีขึ้น ซึ่งมีจำนวน 2,249 ความคิดเห็น และ 1 ความคิดเห็น อาจมีสิ่งที่ต้องการให้พัฒนาหรือปรับปรุงมากกว่า 1 ด้าน เช่นต้องการให้พัฒนาในด้านของ อาหาร (food) และการจราจร/ขนส่งสาธารณะ (Traffic/Public Transportation) โดยที่การสำรวจความคิดเห็นนี้จะดำเนินการต่อเนื่องในทุกๆ ปี จึงทำให้ความคิดเห็นเพิ่มขึ้นเรื่อย ๆ และอาจมี หลากหลายด้านที่เกี่ยวกับการพัฒนาประเทศไทย ด้วยเหตุนี้ทำให้ผู้วิเคราะห์หรือผู้นำข้อมูลไปใช้งานเกิดความลำบาก หรือใช้เวลาอย่างมากในการอ่านและจำแนกข้อความความคิดเห็นตามด้านต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อแก้ไขปัญหานี้ผู้วิจัยจึงพัฒนาแบบจำลองในการจำแนกความคิดเห็นแบบอัตโนมัติและสรุปผลการจำแนกในรูปแบบของแดชบอร์ด

งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการวิเคราะห์เหมืองความคิดเห็น (Opinion Mining) ซึ่งเป็นเทคนิคการวิเคราะห์ความคิดเห็นจำนวนมาก และเป็นการเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูล เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่อยู่ในชุดข้อมูลความคิดเห็น รวมถึงนำเสนอวิธีการจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification) ในการจำแนกความคิดเห็นออกเป็นด้านต่าง ๆ วิธีการจำแนกนี้เหมาะกับชุดข้อมูลความคิดเห็น 1 ความคิดเห็นที่อาจมีหลากหลายด้าน (Tahsin Tasnia Khan. et al. 2023 ; Emre Deniz. et al. 2022) และผู้วิจัยมีการพัฒนาแบบจำลองให้เกิดความแม่นยำมากยิ่งขึ้นด้วยการใช้เทคนิคการเสริมข้อมูล (Data Augmentation) ร่วมกับการเลือกคุณลักษณะของข้อมูล (Feature selection) โดยจะทำการเปรียบเทียบประสิทธิภาพการจำแนกความคิดเห็นด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ระหว่าง 4 เทคนิค ได้แก่ นาอีฟเบย์ (Naïve Bayes: NB) การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression : LR) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และการสุ่มแบบป่า (Random Forest : RF) เพื่อหาอัลกอริทึมที่เหมาะสมกับการจำแนกความคิดเห็นมากที่สุด โดยผลลัพธ์ของการจำแนกความคิดเห็นจะนำเสนอผ่านแดชบอร์ด เพื่อให้ผู้ใช้งานข้อมูลสามารถใช้งานได้ง่าย และเพื่อการวิเคราะห์ความคิดเห็นที่สะดวก รวดเร็วยิ่งขึ้น อีกทั้งผลการวิเคราะห์ความคิดเห็นของนักท่องเที่ยวต่างชาติสามารถใช้เป็นแนวทางสำหรับการกำหนดนโยบายและการพัฒนาการท่องเที่ยวในประเทศไทย เพิ่มภาพลักษณ์ของประเทศจากมุมมองต่างๆ และตอบสนองความต้องการที่หลากหลายของนักท่องเที่ยว

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อออกแบบ และพัฒนาแบบจำลองในการจำแนกความคิดเห็นแบบอัตโนมัติ สำหรับความคิดเห็นที่มีหลายแง่มุมในความคิดเห็นเดียว โดยใช้เทคนิคการจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification)

2) เพื่อแสดงผลการจำแนกความคิดเห็นของนักท่องเที่ยว และวิเคราะห์ผลผ่านแดชบอร์ด

1.3 ขอบเขตของงานวิจัย

1) งานวิจัยนี้รองรับการจำแนกประเภทของความคิดเห็นที่เป็นภาษาอังกฤษเท่านั้น

2) ขอบเขตของข้อมูลที่ศึกษาคือความคิดเห็นของนักท่องเที่ยวต่างชาติที่เดินทางเข้ามาท่องเที่ยวประเทศไทย ในปี 2563 และปี 2565 จำนวน 2,249 ความคิดเห็น

3) งานวิจัยนี้พัฒนาแบบจำลองในการจำแนกความคิดเห็นของนักท่องเที่ยวต่างชาติที่เดินทางเข้ามาท่องเที่ยวประเทศไทย ที่มาจากการเก็บรวบรวมโดยการท่องเที่ยวแห่งประเทศไทย (ททท.) เท่านั้น

4) งานวิจัยนี้ประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ได้แก่ นาอีฟเบย์ (Naïve Bayes: NB) การถดถอยโลจิสติค (Logistic Regression : LR) ซัพพอร์ตเวกเตอร์แมชชีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Support Vector Machine: SVM) และการสุ่มแบบป่า (Random Forest : RF) เพื่อหาวิธีที่เหมาะสมกับการจำแนกความคิดเห็นมากที่สุด

1.4 ขั้นตอนการศึกษาและดำเนินการวิจัย

1) ประสานงานกับหน่วยงานการท่องเที่ยวแห่งประเทศไทย (ททท.) เพื่อขอข้อมูลความคิดเห็น
2) ศึกษาทฤษฎี เทคนิคการเตรียมข้อมูลเพื่อการวิเคราะห์ข้อความ (Text Preprocessing)
3) ศึกษาทฤษฎี เทคนิคการจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification)
4) พัฒนาแบบจำลองการจำแนกความคิดเห็น
5) วิเคราะห์ผลและวัดประสิทธิภาพของแบบจำลอง
7) ออกแบบและพัฒนาแดชบอร์ดสำหรับการแสดงผลการจำแนกประเภทความคิดเห็น
8) สรุปผลการทดลอง
9) เรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ช่วยให้ผู้ใช้งานข้อมูลเข้าใจและวิเคราะห์ความคิดเห็นของนักท่องเที่ยวต่างชาติได้อย่างรวดเร็ว แม่นยำ และมีประสิทธิภาพยิ่งขึ้น

2) นำไปใช้เป็นแนวทางสำหรับการกำหนดนโยบายพัฒนาการท่องเที่ยวและตอบสนองความต้องการของนักท่องเที่ยวได้อย่างตรงจุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เสนอเรื่องการจำแนกความคิดเห็นของนักท่องเที่ยวต่างชาติ โดยในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องดังนี้

2.1 การเสริมข้อมูล (Data Augmentation)

การเสริมข้อมูลเป็นเทคนิคที่ใช้ในการเพิ่มจำนวนข้อมูลของข้อความฝึกอบรม (Training Data) โดยการแปลงข้อความที่มีอยู่ให้เป็นรูปแบบใหม่ๆ โดยไม่ต้องเขียนข้อความใหม่ หรือเก็บข้อมูลเพิ่ม เทคนิคนี้มีประโยชน์สำหรับงานประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) หลายประเภท โดยเฉพาะอย่างยิ่งงานที่ข้อมูลข้อความฝึกสอนมีจำนวนจำกัด โดยเทคนิคการเสริมข้อมูลมีหลายวิธี ขึ้นอยู่กับประเภทของข้อมูล ในงานวิจัยนี้เลือกใช้วิธีการแทนที่คำบางคำด้วยคำพ้องความหมาย (Synonym addition) โดยมีคลังคำศัพท์ที่สนับสนุนวิธีนี้เรียกว่า WordNet

WordNet เป็นฐานข้อมูลทางภาษาศาสตร์ที่จัดโครงสร้างของคำศัพท์ในภาษาอังกฤษโดยอิงจากความหมายและความเกี่ยวข้องภายในคำศัพท์เหล่านั้น ถูกใช้ในงานด้านการประมวลผลภาษาธรรมชาติ (NLP) เช่น การวิเคราะห์ความรู้สึก, การจำแนกประเภทข้อความ, และการค้นหาข้อมูล โดยมีคุณสมบัติการแทนที่ด้วยคำพ้องความหมายเพื่อสร้างข้อความใหม่ที่หลากหลาย ซึ่งมีการตรวจสอบเบื้องต้นในการแทนที่คำคือต้องเป็นคำที่มีคำพ้องความหมาย หากไม่มีคำพ้องความหมายคำนั้นจะไม่ถูกแทนที่ ดังรูปที่ 2.1

```
Original:
The quick brown fox jumps over the lazy dog
Augmented Text:
The quick brown fox parachute over the lazy blackguard
```

รูปที่ 2.1 ตัวอย่างการเสริมข้อมูล (Data Augmentation)

2.2 การเตรียมข้อมูลข้อความ (Text Preprocessing)

การเตรียมข้อมูลข้อความเป็นขั้นตอนสำคัญในการทำงานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing - NLP) โดยมีวัตถุประสงค์เพื่อทำความสะอาดและปรับโครงสร้างของข้อมูลข้อความให้พร้อมสำหรับการวิเคราะห์หรือการสร้างแบบจำลองในขั้นตอนต่อไป (Christine P. Chai. 2022 ; Anna Glazkova. 2023) วิธีการต่างๆ ใน การเตรียมข้อมูลข้อความมีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. **การทำความสะอาดข้อมูล (Cleaning Data) :** เป็นการลบส่วนที่ไม่จำเป็นออกจากข้อความ เช่น แท็ก HTML, สัญลักษณ์พิเศษ, เครื่องหมายวรรคตอนหรือตัวเลขที่ไม่มีความหมายในบริบทนั้นๆ
2. **การแปลงข้อความเป็นตัวอักษรพิมพ์เล็ก (Lowercasing) :** คือกระบวนการที่ใช้เพื่อเปลี่ยนตัวอักษรทั้งหมดในข้อความให้เป็นตัวพิมพ์เล็กทั้งหมด โดยไม่คำนึงถึงตำแหน่งหรือบทบาทของตัวอักษรในคำหรือประโยค เช่น การเปลี่ยน "Hello World" เป็น "hello world" หรือ "ThIs Is A TeSt" เป็น "this is a test"
3. **การลบคำที่ไม่มีความหมาย (Stop Word Removal) :** คือการลบคำที่มีความถี่สูงและไม่มีประโยชน์ในการวิเคราะห์ข้อความหรือคำที่ไม่มีความหมายออกจากข้อมูล เช่น คำเชื่อม (conjunctions) เช่น "and", "or", "but", คำเป็นที่นิยม (articles) เช่น "a", "an", "the", และคำเพื่อการวิเคราะห์ที่ไม่ได้มีความหมายสำคัญเช่น "is", "are", "was", "were"
4. **การตัดคำ (Stemming) :** เป็นกระบวนการที่ใช้ในการลบส่วนท้ายของคำเพื่อให้เหลือแต่คำหลัก (root word) การตัดคำนี้มักจะใช้กฎหรือหลักการง่าย ๆ เช่น การลบ -ing, -ed, -es, -s ออกจากคำ เพื่อให้คำที่ตัดไปยังคงมีความหมายอยู่
5. **TF-IDF (Term Frequency-Inverse Document Frequency) :** ซึ่งเป็นเทคนิคที่ใช้ในการประเมินความสำคัญของคำในเอกสารหรือกลุ่มของเอกสารที่เกี่ยวข้องกันในคลังของข้อความ โดยวิธีนี้เป็นที่นิยมใช้ในการค้นหาข้อมูลและการจัดการข้อมูลในภาษาธรรมชาติ (NLP) (En Yang and Zhaohua Long. 2023 ; Yushan Xu. et al. 2022 ; Mariem Bounabi. et al. 2019) โดยมีสองส่วนหลักได้แก่
   - Term Frequency (TF) : ความถี่ของคำในเอกสาร เป็นการจัดการกับความถี่ที่คำนั้นปรากฏขึ้นในเอกสารเฉพาะ คำที่ปรากฏบ่อยครั้งอาจมีความสำคัญสูงในเอกสารนั้น แต่อาจไม่ใช่ในคลังข้อความทั้งหมด การคำนวณ TF สามารถทำได้โดยการนับจำนวนครั้งที่คำปรากฏในเอกสารและแบ่งด้วยจำนวนคำทั้งหมดในเอกสารนั้น คำนวณได้จากสมการที่ 2.1

$$TF_{(term,document)} = \frac{f_{(term,document)}}{\sum_{(term' \in document)} f_{(term',document)}} \quad (2.1)$$

   - Inverse Document Frequency (IDF) : วัดความสำคัญของคำโดยพิจารณาจากจำนวนเอกสารในคลังข้อความทั้งหมดที่มีคำนั้นปรากฏอยู่ คำที่หายากมาก (ปรากฏในเอกสารไม่กี่เอกสาร) จะได้รับค่า IDF สูง ซึ่งสะท้อนถึงความสำคัญของคำนั้นในแง่ของการเป็นตัวบ่งชี้เนื้อหาที่เฉพาะเจาะจงในคลังข้อความ คำนวณได้จากสมการที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$IDF_{(term,allDocument)} = log\frac{N}{df(t)} \quad (2.2)$$

ดังนั้นสูตรสำหรับคำนวณ TF-IDF คือผลคูณของ TF และ IDF โดยคำนวณได้จากสมการที่ 2.3

$$TF\ IDF = TF_{(term,document)} \times IDF_{(term,allDocument)} \quad (2.3)$$

2.3 การเลือกคุณลักษณะของข้อมูล (Feature selection)

เป็นกระบวนการที่เลือกคุณลักษณะที่สำคัญที่สุดจากข้อมูลข้อความเพื่อใช้ในการสร้างโมเดลการจำแนกประเภทข้อความ วิธีนี้ช่วยลดขนาดของชุดข้อมูลและลดความซับซ้อนของโมเดลโดยการขจัดคุณลักษณะที่ไม่เกี่ยวข้องหรือซ้ำซ้อน ซึ่งอาจทำให้ประสิทธิภาพการจำแนกข้อความของแบบจำลองดีขึ้นและลดเวลาในการฝึกสอน นอกจากนี้ยังช่วยเพิ่มความเข้าใจเกี่ยวกับข้อมูลและการทำงานของแบบจำลอง

การใช้ไคสแควร์ (Chi-square) สำหรับการเลือกคุณลักษณะ (Feature selection) ในงานจำแนกประเภทข้อความ (text classification) เป็นเทคนิคสถิติที่ช่วยตัดสินใจว่าคุณลักษณะใดมีความสำคัญและมีความเกี่ยวข้องกับประเภทหรือคลาสของข้อมูลมากที่สุด โดยอาศัยการวัดความสัมพันธ์ระหว่างคุณลักษณะ (features) กับประเภทข้อมูล (classes) นั้นๆ วิธีนี้มีประโยชน์ในการกรองคุณลักษณะที่ไม่มีความเกี่ยวข้องออกหรือมีผลน้อยกับการจำแนกประเภท ซึ่งทำให้แบบจำลองที่ได้มีประสิทธิภาพสูงขึ้นและลดการเกิด overfitting โดยสูตรการคำนวณค่า Chi-square (χ^2) สำหรับการทดสอบความสัมพันธ์ระหว่างคุณลักษณะ (features) และประเภทของข้อมูล (classes) ในงานจำแนกประเภทข้อความ (text classification) สามารถเขียนได้ดังนี้

$$X^2 = \sum_{i=1}^{n}\frac{(O_i - E_i)^2}{E_i} \quad (2.4)$$

โดย O_i คือค่าที่สังเกตได้ (observed frequency) สำหรับคุณลักษณะในแต่ละคลาส

E_i คือค่าที่คาดหวัง (expected frequency) สำหรับคุณลักษณะนั้น ถ้าไม่มีความสัมพันธ์กับคลาส

n คือจำนวนของเซลล์ในตารางความถี่ (cells in the contingency table)

2.4 การจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification)

การจำแนกประเภทแบบหลายฉลาก เป็นหนึ่งในด้านการเรียนรู้ของเครื่อง (Machine Learning) ที่แบบจำลองการจำแนกประเภทจำเป็นต้องกำหนดหลายฉลากหรือหมวดหมู่ให้กับตัวอย่างเดียวกันได้พร้อมกัน ซึ่งแตกต่างจากการจำแนกประเภทแบบฉลากเดียว (Single-Label Classification) ที่ตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่ละตัวจะถูกจัดให้อยู่ในหมวดหมู่เดียวเท่านั้น เนื่องจากงานวิจัยนี้ 1 ความคิดเห็นของนักท่องเที่ยวอาจมีได้มากกว่า 1 ด้าน ดังนั้นจึงจำเป็นต้องใช้วิธีการจำแนกประเภทแบบหลายฉลาก โดยมีขั้นตอนดังนี้

2.4.1 การแปลงข้อมูลแบบหลายฉลาก (Multi-Label Transformation)

เป็นการแปลงปัญหาการจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification) ให้เป็นปัญหาการจำแนกประเภทแบบฉลากเดียว (Single-Label Classification) ช่วยให้สามารถใช้แบบจำลองการจำแนกประเภทแบบดั้งเดิม (Single-Label Classification Algorithms) กับงานประเภทแบบหลายฉลากได้ โดยมีเทคนิคการแปลงข้อมูลดังนี้

1. Binary Relevance

เป็นวิธีการแปลงปัญหาการจำแนกประเภทแบบหลายฉลากเป็นชุดปัญหาการจำแนกประเภทแบบไบนารี (binary) เพื่อระบุว่าแต่ละฉลากมีหรือไม่มีความเกี่ยวข้องกับตัวอย่างข้อมูลฉลากนั้น ๆ โดยข้อจำกัดของ Binary Relevance ไม่สามารถใช้ได้กับฉลากที่มีความสัมพันธ์กันในข้อมูลเพราะ แต่ละ Label เป็นอิสระต่อกันในการฝึกฝนของแบบจำลอง (Adil Yaseen Taha and S. Tiun. 2016.) ตามตัวอย่างในรูปที่ 2.2

X	Class1	Class2	Class3
X1	0	0	1
X2	0	1	0
X3	1	0	1

X	Class1
X1	0
X2	0
X3	1

X	Class2
X1	0
X2	1
X3	0

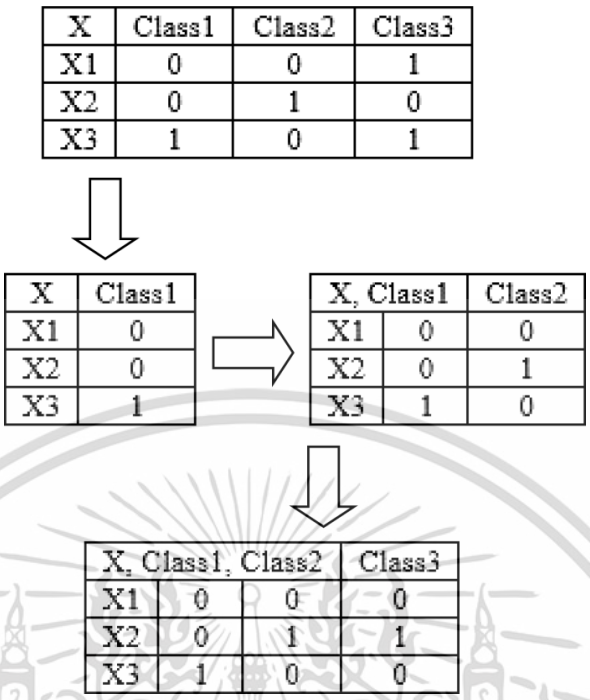
X	Class3
X1	1
X2	0
X3	1

รูปที่ 2.2 Binary Relevance

2. Classifier Chains

เป็นเทคนิคที่พัฒนามาจากวิธีการ Binary Relevance แต่ได้รับการออกแบบมาเพื่อจัดการกับข้อจำกัดหลักของ Binary Relevance คือแต่ละฉลากต่างๆ ไม่มีความสัมพันธ์กัน วิธีการนี้จะใช้การทำนายของฉลากก่อนหน้าเป็นข้อมูลสำหรับฉลากต่อไป ตามตัวอย่างในรูปที่ 2.3 และด้วยการนำความสัมพันธ์เหล่านี้มาพิจารณา ในบางงานวิจัย Classifier Chains ช่วยให้การทำนายมีความแม่นยำสูงขึ้น (Jesse Read. et al. 2021)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



X	Class1	Class2	Class3
X1	0	0	1
X2	0	1	0
X3	1	0	1

X	Class1
X1	0
X2	0
X3	1

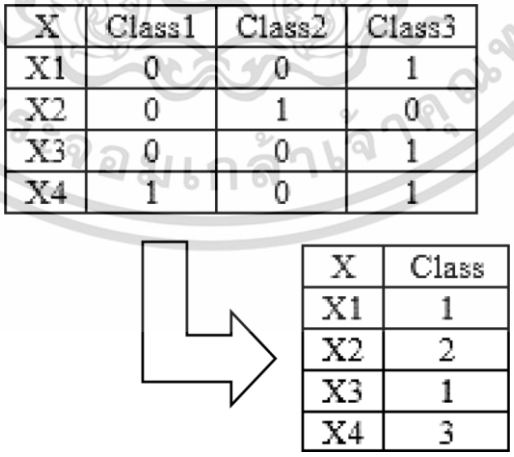
X, Class1		Class2
X1	0	0
X2	0	1
X3	1	0

X, Class1, Class2			Class3
X1	0	0	0
X2	0	1	1
X3	1	0	0

รูปที่ 2.3 Classifier Chains

3. Label Powerset

เป็นวิธีการที่พิจารณาทุกชุดฉลากที่ปรากฏในชุดข้อมูลการฝึกหัดเป็นหมวดหมู่เดียว นั่นคือแต่ละชุดฉลากที่แตกต่างกันถูกมองว่าเป็นคลาสเฉพาะในการจำแนกประเภทใหม่ (Bassam Al-Salemi. et al. 2019) เช่น X1 และ X3 มีคลาสที่เกิดขึ้นคือ 001 เหมือนกัน จึงถูกจำแนกเป็น Class 1 ดังแสดงตัวอย่างในรูปที่ 2.4



X	Class1	Class2	Class3
X1	0	0	1
X2	0	1	0
X3	0	0	1
X4	1	0	1

X	Class
X1	1
X2	2
X3	1
X4	3

รูปที่ 2.4 Label Powerset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2 การจำแนกประเภท (Classification)

เป็นวิธีการที่ดำเนินการหลังจากมีการแปลงข้อมูลแบบหลายฉลากเรียบร้อยแล้ว โดยประยุกต์ใช้การเรียนรู้ของเครื่อง (Machine Learning) ในการจำแนกประเภท เทคนิคเหล่านี้จะเป็นส่วนหนึ่งของการเรียนรู้แบบมีผู้สอน(Supervised Learning) ที่แบบจำลองเรียนรู้จากชุดข้อมูลฝึกสอนที่มีฉลากแล้ว และใช้ความรู้นั้นในการจำแนกข้อมูลใหม่ สำหรับเทคนิคการจำแนกประเภทข้อความที่ใช้กันอย่างแพร่หลายได้แก่ นาอีฟเบย์ (Naïve Bayes: NB) การถดถอยโลจิสติค (Logistic Regression : LR) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และการสุ่มแบบป่า (Random Forest : RF) (Roberto Carlos Morales-Hernández. et al. 2022)

1. นาอีฟเบย์ (Naïve Bayes: NB)

เป็นเทคนิคการจำแนกประเภทที่อาศัยหลักการของทฤษฎีความน่าจะเป็นโดยใช้ทฤษฎีเบย์ (Bayes' Theorem) ซึ่งเป็นวิธีการคำนวณความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้นโดยพิจารณาจากข้อมูลที่มีอยู่ก่อนและข้อมูลใหม่ที่ได้รับ สูตรของทฤษฎีเบย์คือ

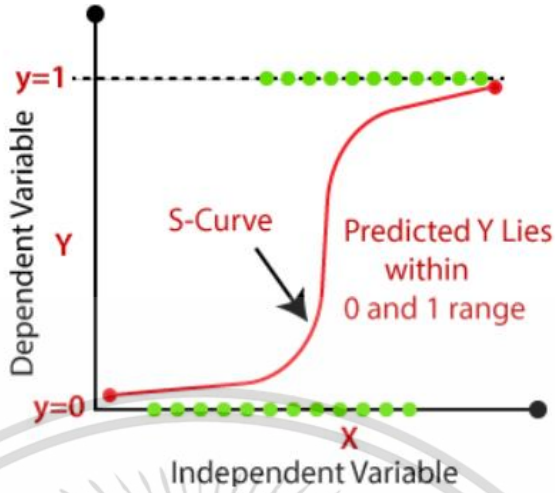
$$P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)} \quad (2.5)$$

โดยที่ P(Y|X): ความน่าจะเป็นที่จะเกิดเหตุการณ์ Y โดยมี X เป็นเงื่อนไข
P(X|Y): ความน่าจะเป็นที่จะเกิดเหตุการณ์ X โดยมี Y เป็นเงื่อนไข
P(Y): ความน่าจะเป็นที่เหตุการณ์ Y จะเกิดขึ้น (Prior Probability)
P(X): ความน่าจะเป็นที่เหตุการณ์ X จะเกิดขึ้น (Prior Probability)

2. การถดถอยโลจิสติค (Logistic Regression : LR)

เป็นวิธีที่อาศัยแนวคิดจากสถิติและความน่าจะเป็น เพื่อทำนายหมวดหมู่หรือป้ายกำกับ (label) ของข้อมูลที่เข้ามา สมการ Logistic Regression สามารถแสดงให้อยู่ในรูปของสมการ Sigmoid ซึ่งเป็นฟังก์ชันที่มีลักษณะ S-shape และสามารถแปลงค่าของตัวแปรอิสระ (independent variable) ให้อยู่ในช่วงค่าระหว่าง 0 ถึง 1 ซึ่งเหมาะสำหรับการแสดงผลลัพธ์เป็นความน่าจะเป็น (probability) ดังแสดงในรูปที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 การถดถอยโลจิสติค (Logistic Regression : LR)

สมการ Sigmoid จะมีรูปแบบดังนี้

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \quad (2.6)$$

โดยที่

P(Y=1|X) คือความน่าจะเป็นที่ตัวแปรตาม (dependent variable) จะมีค่าเป็น 1 โดยที่ตัวแปรอิสระ (independent variable) X มีค่าเป็น X

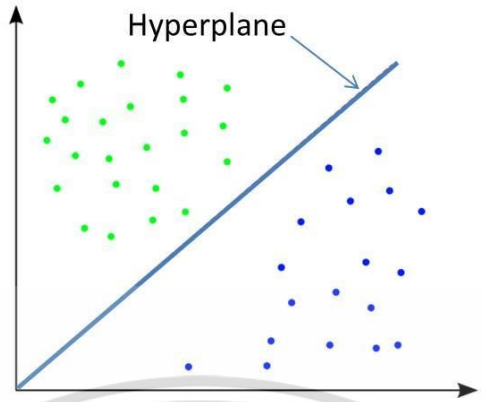
e คือค่าของค่าคงที่ Euler's number (มีค่าประมาณ 2.71828)

β_0 และ β_1 คือพารามิเตอร์ของโมเดล Logistic Regression ซึ่งเป็นค่าที่ต้องปรับให้เหมาะสมกับข้อมูล

3. ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

ออกแบบมาเพื่อหาขอบเขตการตัดสินใจ (decision boundary) ที่ดีที่สุดในการแบ่งข้อมูลออกเป็นสองกลุ่ม หรือมากกว่านั้น โดยที่ขอบเขตนี้จะเรียกว่า hyperplane ในพื้นที่คุณลักษณะ (feature space) ที่มีมิติสูง การจำแนกประเภทด้วย SVM นั้นมีหลักการคือการหา hyperplane ที่สามารถแยกข้อมูลออกเป็นสองกลุ่มได้ด้วยระยะห่างที่มากที่สุดระหว่างจุดข้อมูลที่ใกล้ที่สุดจากแต่ละกลุ่ม (ที่เรียกว่า support vectors) กับ hyperplane เอง โดยเป้าหมายคือการสร้างระยะห่างนี้ให้มากที่สุด เพื่อเพิ่มความแม่นยำในการจำแนกประเภทของโมเดล แสดงในรูปที่ 2.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM)

ฟังก์ชันเคอร์เนล (Kernel function) ใน SVM ช่วยให้การจำแนกข้อมูลที่ไม่เป็นเชิงเส้นเป็นไปได้โดยการแปลงข้อมูลไปยังพื้นที่ที่มีมิติสูง ซึ่งทำให้ข้อมูลที่ซับซ้อนสามารถถูกจำแนกโดย hyperplane เชิงเส้นได้ นี่คือภาพรวมของฟังก์ชันเคอร์เนลที่ใช้บ่อยใน SVM

- เคอร์เนลเชิงเส้น (Linear Kernel)

$$K(x, x') = x \cdot x' \tag{2.7}$$

- โพลินอเมียล เคอร์เนล (Polynomial Kernel)

$$K(x, x') = (\gamma x \cdot x' + r)^d \tag{2.8}$$

- RBF (Radial Basis Function) หรือ Gaussian Kernel

$$K(x, x') = \exp\left(-\gamma ||x - x'||^2\right) \tag{2.9}$$

4. การสุ่มแบบป่า (Random Forest : RF)

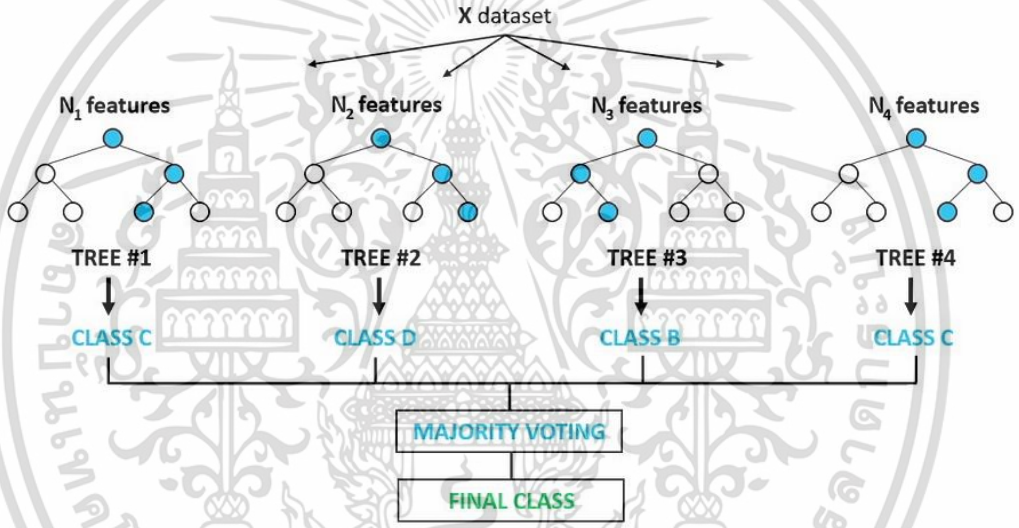
วิธีการสุ่มแบบป่ามีหลักการคือการผสมผสานของต้นไม้ตัดสินใจหลายต้น (Decision Trees) มาทำนายผลลัพธ์ที่ต้องการ เปรียบเสมือนป่าไม้ที่มีต้นไม้นานาพันธุ์ แต่ละต้นมีความคิดเห็นของตัวเอง โดยมีลักษณะเด่นที่ใช้ มีขั้นตอนดังนี้

- สร้างต้นไม้ตัดสินใจ: อัลกอริทึมจะสุ่มเลือกชุดข้อมูลย่อย (Subset) จากข้อมูลทั้งหมดมาสร้างต้นไม้ตัดสินใจแต่ละต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- จำกัดความลึกของต้นไม้: ต้นไม้แต่ละต้นจะมีความลึกจำกัด เพื่อป้องกันปัญหา Overfitting
- สุ่มเลือกฟีเจอร์: ในแต่ละโหนดของต้นไม้ อัลกอริทึมจะสุ่มเลือกฟีเจอร์ย่อย (Subset) มาใช้ในการตัดสินใจ
- โหวตหาผลลัพธ์: เมื่อป้อนข้อมูลใหม่ แต่ละต้นไม้จะทำการคาดการณ์ผลลัพธ์ ในกรณี Classification ผลลัพธ์ที่มีการโหวตมากที่สุดจะเป็นผลลัพธ์สุดท้าย และกรณี Regression จะคำนวณค่าเฉลี่ยของผลลัพธ์จากทุกต้นไม้

หลักการทำงานของ Random Forest แสดงในรูปที่ 2.7

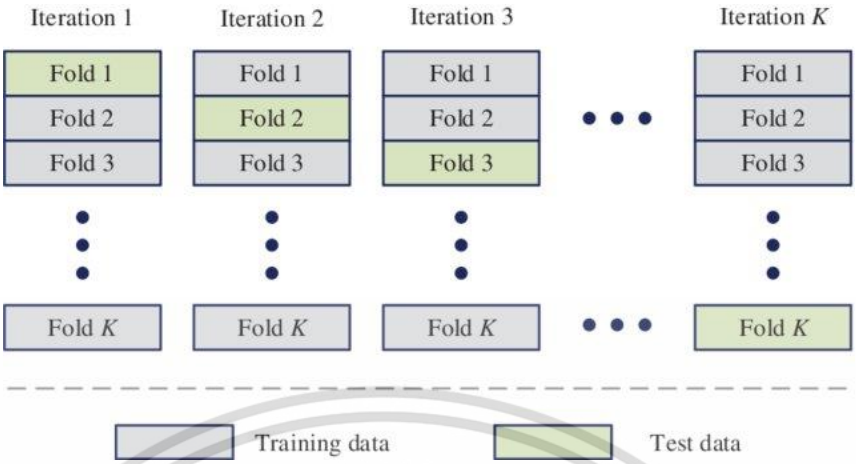


รูปที่ 2.7 การสุ่มแบบป่า (Random Forest : RF)

2.5 K-fold Cross Validation

เป็นเทคนิคหนึ่งในการประเมินประสิทธิภาพของแบบจำลองเทคนิคการเรียนรู้ของเครื่อง (Machine Learning Models) และเพื่อตรวจสอบว่าแบบจำลองที่พัฒนาขึ้นนั้นมีความเที่ยงตรงและสามารถทำนายผลได้ดีเมื่อใช้กับข้อมูลนอกชุดที่ใช้ในการฝึกฝน วิธีนี้ช่วยลดปัญหา overfitting และ underfitting โดยการแบ่งชุดข้อมูลออกเป็น 'k' ส่วน (folds) ที่มีขนาดเท่าๆ กัน และทำการฝึกฝนและทดสอบโมเดล 'k' ครั้ง โดยในแต่ละครั้งจะเลือกหนึ่งใน 'k' ส่วนเป็นชุดข้อมูลทดสอบ (test set) และส่วนที่เหลือเป็นชุดข้อมูลฝึกฝน (training set) แล้วจึงนำค่าความแม่นยำมาหาเฉลี่ยหรือค่าเฉลี่ยเพื่อประเมินประสิทธิภาพของแบบจำลองโดยรวม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.8 ตัวอย่างของขั้นตอน K-fold Cross Validation

2.6 การวัดประสิทธิภาพของแบบจำลอง (Model Evaluation)

การประเมินประสิทธิภาพโมเดล Multi-label classification นั้นมีความซับซ้อนกว่าโมเดล Classification ทั่วไป เนื่องจากโมเดล Multi-label classification ทำนายผลได้มากกว่า 2 คลาส ดังนั้นจึงไม่สามารถใช้วิธีวัดประสิทธิภาพแบบง่ายๆ เช่น ความแม่นยำ (Accuracy) เพียงอย่างเดียวได้ (Mohammadreza Heydarian. et al. 2022 ; Nawal Aljedani. et al. 2020 ; Chintrai Puttipornchai. et al. 2022) โดยวิธีการวัดประสิทธิภาพที่นิยมใช้สำหรับแบบจำลองการจำแนกประเภทแบบหลายฉลาก (Multi-label classification) มีดังนี้

2.6.1 ความถูกต้อง (Accuracy)

เป็นการวัดสัดส่วนของการจำแนกที่ถูกต้องทั้งหมดเมื่อเทียบกับข้อมูลทดสอบทั้งหมด หมายถึงผลการจำแนกของข้อความนั้นๆ จะต้องเหมือนกับฉลากที่กำกับจริงทุกประการ โดยที่ I (True) = 1 และ I (False) = 0 สามารถคำนวณได้ตามสมการที่ 2.9

$$Accuracy = \frac{1}{m}\sum_{i=1}^{m} I|Z_i = Y_i| \qquad (2.9)$$

2.6.2 ความเที่ยงตรง (Precision)

คือสัดส่วนของฉลากที่ทำนายได้ถูกต้องเทียบกับจำนวนฉลากที่ทำนายทั้งหมด จากนั้นจะเฉลี่ยค่านี้จากทุกตัวอย่างในชุดข้อมูลทดสอบ สามารถคำนวณได้ตามสมการที่ 2.10

$$Precision = \frac{1}{m}\sum_{i=1}^{m} \frac{|Z_i \cap Y_i|}{|Z_i|} \qquad (2.10)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.3 ความระลึก (Recall)

คือสัดส่วนของฉลากที่ทำนายได้ถูกต้องเทียบกับจำนวนฉลากที่กำกับจริงทั้งหมด จากนั้นจะเฉลี่ยค่านี้จากทุกตัวอย่างในชุดข้อมูลทดสอบ สามารถคำนวณได้ตามสมการที่ 2.11

$$Recall = \frac{1}{m}\sum_{i=1}^{m}\frac{|Z_i \cap Y_i|}{|Y_i|} \quad (2.11)$$

2.6.4 F1-Score

คือค่าเฉลี่ยระหว่างความเที่ยงตรง (Precision) และ ความระลึก (Recall) สามารถคำนวณได้ตามสมการที่ 2.12

$$F1 - Score = \frac{1}{m}\sum_{i=1}^{m}\frac{2|Z_i \cap Y_i|}{|Z_i| + |Y_i|} \quad (2.12)$$

2.6.5 Hamming Loss

คือการคำนวณฉลากที่ทำนายไม่ถูกต้องในกลุ่มตัวอย่าง จากนั้นจึงหาค่าเฉลี่ยของกลุ่มตัวอย่างทั้งหมดในชุดข้อมูลทดสอบ ดังนั้นหากค่านี้มีค่าต่ำมากที่สุดแสดงว่าผลการทำนายดีที่สุด คำนวณได้ตามสมการที่ 2.13

$$Hamming\ Loss = \frac{1}{m}\sum_{i=1}^{m}|Z_i \neq Y_i| \quad (2.13)$$

2.7 งานวิจัยที่เกี่ยวข้อง

การจำแนกข้อความความคิดเห็นจำเป็นต้องใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) วิเคราะห์ข้อมูลข้อความจำนวนมากที่เกี่ยวข้องกับการท่องเที่ยว เช่น รีวิวจากนักท่องเที่ยว บทความบล็อก โพสต์โซเชียลมีเดีย เว็บไซต์การท่องเที่ยว เอกสารทางการท่องเที่ยว ฯลฯ เพื่อค้นหารูปแบบ ความรู้ และข้อมูลเชิงลึกที่มีประโยชน์ต่อธุรกิจการท่องเที่ยว อีกทั้งยังมีความสำคัญต่อธุรกิจการท่องเที่ยวมากมาย เช่น ทำให้เข้าใจความต้องการของลูกค้า และสามารถนำมาพัฒนาผลิตภัณฑ์และบริการ และยังมีส่วนช่วยในการสร้างกลยุทธ์การตลาดการท่องเที่ยว โดยมีงานวิจัยที่เกี่ยวข้องดังนี้

งานวิจัยของ Puteri Prameswari. et al. (2017) ได้ทำการศึกษาเกี่ยวกับการใช้เทคนิคการทำเหมืองข้อมูลข้อความเพื่อสกัดความคิดเห็นและจัดหมวดหมู่ความคิดเห็นของนักท่องเที่ยวเกี่ยวกับบริการของโรงแรมในอินโดนีเซีย จากการตรวจสอบความคิดเห็นทั้งหมด นักวิจัยได้ระบุหัวข้อสำคัญในการจำแนกตามด้านต่าง ๆ เช่น การเข้าถึง (Accessibility) กิจกรรมและความบันเทิง (Activities and

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Entertainment) อาหารและเครื่องดื่ม (Food and Beverage) ทรัพยากรมนุษย์ (Human Resources) และสภาพแวดล้อมทางกายภาพ (Physical Environment) ผลของการวิจัยมีส่วนช่วยเสริมสร้างมาตรฐานการบริการและปรับปรุงการบริการให้ดียิ่งขึ้น เพื่อสร้างความพึงพอใจของลูกค้า

งานวิจัยของ Muhammad Afzaal and Muhammad Usman (2015) ได้รวบรวมข้อมูลความคิดเห็นผ่าน Twitter โดยจำแนกหมวดหมู่ความคิดเห็นตามสถานที่นั้นๆ เช่น สถานที่ท่องเที่ยว สวนสาธารณะ โรงแรม และร้านอาหาร โดยใช้เทคโนโลยีการทำเหมืองข้อมูลโดยมุมมอง (aspect-based opinion mining) ซึ่งเป็นส่วนหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) พวกเขาจัดหมวดหมู่ความคิดเห็นลงในหมวดหมู่ที่แตกต่างกัน เช่น ศิลปะ สิ่งอำนวยความสะดวก อาหาร และร้านหนังสือ การวิเคราะห์ความคิดเห็นของงานวิจัยนี้ช่วยให้นักท่องเที่ยวตัดสินใจได้ง่ายขึ้นสำหรับการเลือกสถานที่ท่องเที่ยวต่าง ๆ

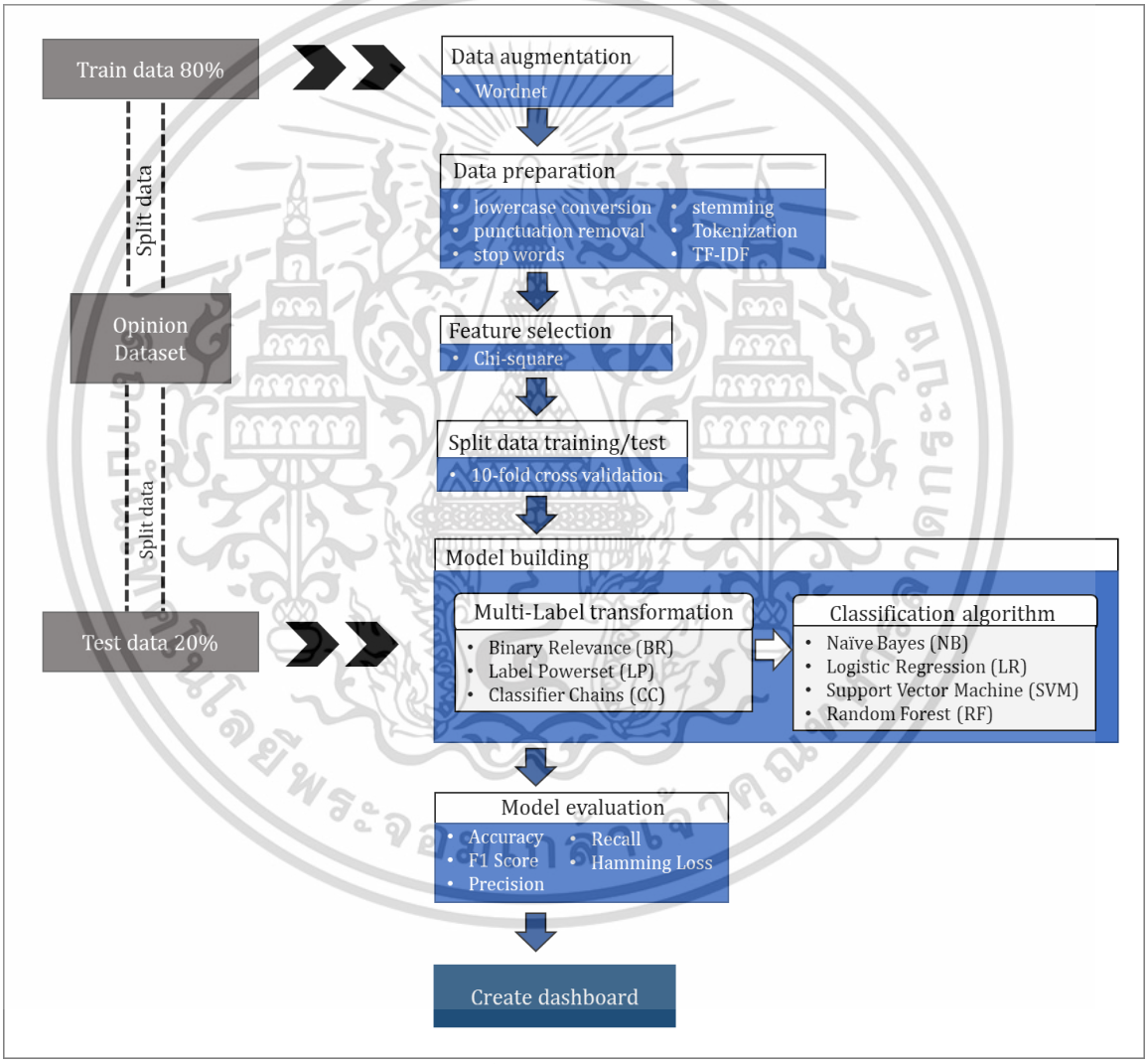
นอกจากเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) แล้ว ผู้วิจัยยังศึกษาวิธีการเพิ่มประสิทธิภาพการจำแนกข้อความให้มีความแม่นยำมากยิ่งขึ้น โดยงานวิจัยของ Uchenna Cosmas Ugwuoke. et al. (2022) ได้ทดลองใช้การเสริมข้อมูลข้อความ โดยใช้เทคนิคการแทนที่ด้วยคำพ้องความหมาย (Synonym Replacement) และใช้คลังศัพท์ที่เรียกว่า WordNet ในการหาคำพ้องความหมายที่เหมาะสมแทนที่คำต้นฉบับ ประยุกต์ใช้กับข้อมูลข้อความที่สกัดมาจากบทความข่าว BBC ผลการทดลองแสดงให้เห็นว่าเทคนิคดังกล่าวสามารถเพิ่มประสิทธิภาพในการจำแนกประเภทข้อความ และงานวิจัยของ Yujia Zhai. et al. (2018) ได้เสนอวิธีการเลือกคุณลักษณะโดยใช้เทคนิคไค-สแควร์ (Chi-square) บนชุดข้อมูลรีวิวคอมพิวเตอร์และหนังสือ ผลลัพธ์แสดงให้เห็นว่าวิธีการดังกล่าวสามารถเพิ่มประสิทธิภาพการจำแนกความข้อความให้มีความแม่นยำมากยิ่งขึ้น ดังนั้นงานวิจัยนี้จึงเลือกใช้วิธีการเพิ่มจำนวนข้อมูลของข้อความฝึกอบรมด้วยการแทนที่ด้วยคำพ้องความหมาย โดยใช้คลังศัพท์ที่เรียกว่า WordNet ร่วมกับเทคนิคเลือกคุณลักษณะของข้อมูล (Feature selection) ด้วยวิธีไค-สแควร์ (Chi-square) เพื่อเพิ่มประสิทธิภาพการจำแนกข้อความของแบบลำจอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีดำเนินงานวิจัย

ในเนื้อหาของบทนี้เป็นการอธิบายถึงขั้นตอนการทำงานที่ใช้สำหรับการพัฒนาแบบจำลองการจำแนกประเภทแบบหลายฉลากของความคิดเห็นนักท่องเที่ยวต่างชาติ โดยขั้นตอนทั้งหมดแสดงในรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนการดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.1 ขั้นตอนแรกผู้วิจัยแบ่งข้อมูลเป็นชุดฝึกสอนและชุดทดสอบ และดำเนินการเสริมข้อมูล (Data Augmentation) ในชุดฝึกสอน ต่อจากนั้นเข้าสู่ขั้นตอนการเตรียมข้อมูลข้อความ (Text Preprocessing) รวมถึงการเลือกคุณลักษณะของข้อมูล (Feature selection) เพื่อพร้อมต่อการสร้างแบบจำลอง เมื่อผ่านกระบวนการเตรียมข้อมูลแล้วจะต้องนำข้อมูลมาแบ่งเป็น 10 ชุด จากนั้นจึงนำข้อมูลมาสร้างแบบจำลองการจำแนกประเภทความคิดเห็น สุดท้ายนำเสนอผลลัพธ์การจำแนกผ่านแดชบอร์ด

3.1 ชุดข้อมูล

งานวิจัยนี้ใช้ชุดข้อมูลความคิดเห็นของนักท่องเที่ยวหลังจากที่ได้เดินทางเข้ามาท่องเที่ยวประเทศไทยในปี 2563 และปี 2565 เก็บรวบรวมโดยการท่องเที่ยวแห่งประเทศไทย (ททท.) รวมทั้งสิ้น 2,249 ความคิดเห็น ชุดข้อมูลนี้เป็นข้อมูลความคิดเห็นของนักท่องเที่ยวเกี่ยวกับสิ่งที่ต้องการให้การท่องเที่ยวของไทยปรับปรุงให้ดียิ่งขึ้น โดยเก็บรวบรวมผ่านแบบสอบถามด้วยคำถาม “In your opinion, what could be improved to make Thailand tourism better” ซึ่งความคิดเห็นจากคำถามนี้ถูกจัดหมวดหมู่เป็น 8 หมวดหลักโดยมีรายละเอียดจำนวนความคิดเห็นและหมวดหมู่การจำแนกแสดงในรูปที่ 3.2

หมวดหมู่หลัก	จำนวน
Communication/Language	486
Traffic/Public Transportation	478
Cleanliness/Hygiene	463
Honesty/Friendly/Scamming	276
Service/Facilities	268
Value for money	194
Food	84
Other	430

หมวดหมู่อื่นๆ (Other)	จำนวน
Availability of tourist information	55
Safety	54
Infrastructure	47
Pollution	46
Recycling/Reduce plastic	38
Attraction/Activitiy	32
Security	29
COVID-19 situation	23
Environmental	20
Animal conservation/protection	19
Natural resources	11
Tradition/Culture	7
Prostitution	6
Clamorous, Air	5
Government support	4
Airport, Politics, Police duty, Convenience of immigration and visa procedures, Direct flight, Racism, Hard selling	2
Natural energy, Narcotics, Cater to backpackers, E-Cigarettes, Treatment of tourists, Airline, Homeless person, Flood, Stress, Foreign Business License, Travel agency, Business, War, Cannabis, Health services	1

รูปที่ 3.2 หมวดหมู่และจำนวนความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานวิจัยนี้นำเสนอการจำแนกความคิดเห็นออกเป็น 8 หมวดหลักได้แก่ 1) การสื่อสาร/ภาษา 2) การจราจร/การขนส่งสาธารณะ 3) ความสะอาด/สุขอนามัย 4) ความซื่อสัตย์/เป็นมิตร/การหลอกลวง 5) บริการ/สิ่งอำนวยความสะดวก 6) ความคุ้มค่าเงิน 7) อาหาร 8) หมวดอื่นๆ (การท่องเที่ยวแห่งประเทศไทย (ททท.) 2566) โดยที่หมวดอื่นๆ ประกอบไปด้วยหมวดย่อยดังนี้ ความพร้อมของข้อมูลการท่องเที่ยว ความปลอดภัย โครงสร้างพื้นฐาน มลพิษ รีไซเคิล/ลดการใช้พลาสติก เป็นต้น โดยหมวดย่อยนี้จะไม่ถูกจัดเป็นหมวดหลักเนื่องจากมีจำนวนความคิดเห็นที่น้อยและมีจำนวนที่ใกล้เคียงกัน ผู้วิจัยจึงจัดหมวดย่อยดังกล่าวให้รวมอยู่ในหมวดที่ 8 อื่นๆ

จากการสำรวจความคิดเห็นของนักท่องเที่ยว 1 คน สามารถจัดหมวดหมู่ได้มากกว่า 1 หมวด โดยแสดงตัวอย่างความคิดเห็นและหมวดหมู่ที่กำหนดในตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างความคิดเห็นและหมวดหมู่ที่กำหนด

ลำดับ	ความคิดเห็น	หมวดหมู่	จำนวนหมวดหมู่
1	More expensive	Value for money	1
2	The language problem	Communication/Language	1
3	Better traffic, less dirty streets	Traffic/Public Transportation, Cleanliness/Hygiene	2
4	Hygiene, English speaking	Cleanliness/Hygiene, Communication/Language	2
5	Communication and add international food	Communication/Language, Food	2
6	Better transport infrastructure, English language skills	Traffic/Public Transportation, Communication/Language	2
7	Cleaning, Tourist facilities, Price	Cleanliness/Hygiene, Service/Facilities, Value for money	3
8	Improved level of English, cleanliness, sincerity	Communication/Language, Cleanliness/Hygiene, Honesty/Friendly/Scamming	3
9	Clean better the ocean/streets public transportation better bathrooms less people with bad intentions about money/unfair charges	Cleanliness/Hygiene, Traffic/Public Transportation, Service/Facilities, Honesty/Friendly/Scamming	4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูลความคิดเห็นของนักท่องเที่ยวทั้ง 8 หมวดหมู่หลักได้ถูกแบ่งออกเป็นชุดข้อมูลสำหรับการฝึกสอน 80% และชุดข้อมูลสำหรับการทดสอบ 20% ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 จำนวนความคิดเห็นของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ

ลำดับ	หมวดหมู่	ชุดข้อมูลฝึกสอน	ชุดข้อมูลทดสอบ	รวม
1	Communication/Language	389	97	486
2	Traffic/Public Transportation	382	96	478
3	Cleanliness/Hygiene	370	93	463
4	Honesty/Friendly/Scamming	221	55	276
5	Service/Facilities	214	54	268
6	Value for money	155	39	194
7	Food	67	17	84
8	Other	344	86	430

3.2 การเสริมข้อมูล (Data Augmentation)

จากการศึกษาวิจัยที่เกี่ยวข้องเกี่ยวกับการเพิ่มประสิทธิภาพของแบบจำลองการจำแนกประเภทข้อความ โดยใช้เทคนิคการเสริมข้อมูล (Data Augmentation) ในการเพิ่มชุดข้อมูลฝึกสอนโดยใช้ WordNet ทำให้ได้ชุดข้อมูลที่หลากหลายยิ่งขึ้น โดย 1 ความคิดเห็นสามารถเสริมข้อมูลเพิ่มขึ้นได้ 1 ความคิดเห็น กระบวนการเสริมข้อมูลนี้นำไปใช้กับข้อมูลฝึกสอนในหมวดที่ 2 การจราจร/การขนส่งสาธารณะ (Traffic/Public Transportation) ถึงหมวดที่ 8 อื่น ๆ (Other) โดยมีวัตถุประสงค์เพื่อให้จำนวนข้อมูลของแต่ละหมวดหมู่มีจำนวนที่เท่ากันหรือใกล้เคียงกับหมวดหมู่ที่ 1 ซึ่งเป็นหมวดที่มีจำนวนสูงที่สุด หลังจากการเสริมข้อมูลแล้วหากหมวดหมู่ใดหมวดหมู่หนึ่งมีจำนวนเกินหมวดหมู่ที่ 1 จะมีการสุ่มเลือกบางความเห็นที่เสริมมาเท่านั้น วิธีนี้ช่วยให้การกระจายข้อมูลในทุกหมวดหมู่มีความสมดุลกัน ดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 จำนวนความคิดเห็นของชุดข้อมูลฝึกสอนหลังดำเนินการเสริมข้อมูล

ลำดับ	หมวดหมู่	ชุดข้อมูลเดิม	ชุดข้อมูลที่ผ่านการเสริม
1	Communication/Language	389	389
2	Traffic/Public Transportation	382	389
3	Cleanliness/Hygiene	370	389
4	Honesty/Friendly/Scamming	221	389

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับ	หมวดหมู่	ชุดข้อมูลเดิม	ชุดข้อมูลที่ผ่านการเสริม
5	Service/Facilities	214	389
6	Value for money	155	310
7	Food	67	134
8	Other	344	389

3.3 การเตรียมข้อมูล (Data Preprocessing)

หลังจากดำเนินการเสริมข้อมูลแล้วต่อไปเป็นขั้นตอนการเตรียมข้อมูลเพื่อให้นำไปพัฒนาแบบจำลอง โดยมีขั้นตอนดำเนินการดังนี้

3.3.1 การทำความสะอาดข้อมูล (Cleaning Data)

ลบส่วนที่ไม่จำเป็นออกจากข้อความ เช่น สัญลักษณ์พิเศษและเครื่องหมายวรรคตอนที่ไม่มีความหมายในบริบทนั้นๆ

3.3.2 การแปลงข้อความเป็นตัวอักษรพิมพ์เล็ก (Lowercasing)

เปลี่ยนตัวอักษรทั้งหมดในข้อความให้เป็นตัวพิมพ์เล็กทั้งหมด โดยไม่คำนึงถึงตำแหน่งหรือบทบาทของตัวอักษรในคำหรือประโยค สำหรับวิธีนี้ใช้ฟังก์ชัน .lower() ในภาษาไพธอน

3.3.3 การลบคำที่ไม่มีความหมาย (Stop Word Removal)

ลบคำที่ไม่มีประโยชน์ในการวิเคราะห์ข้อความหรือคำที่ไม่มีความหมายออกจากข้อมูล เช่น คำเชื่อม (conjunctions) เช่น "and", "or", "but", คำเป็นที่นิยม (articles) เช่น "a", "an", "the", และคำเพื่อการวิเคราะห์ที่ไม่ได้มีความหมายสำคัญเช่น "is", "are", "was", "were" สำหรับวิธีนี้ใช้ฟังก์ชัน stopwords ไลบรารี NLTK ในภาษาไพธอน

3.3.4 การตัดคำ (Stemming)

ลบส่วนท้ายของคำเพื่อให้เหลือแต่คำหลัก (root word) เช่น การลบ -ing, -ed, -es, -s ออกจากคำ วิธีนี้ใช้ฟังก์ชัน SnowballStemmer ไลบรารี NLTK ในภาษาไพธอน

3.3.5 TF-IDF (Term Frequency-Inverse Document Frequency)

ใช้วิธีสกัดข้อความเพื่อทำ TF-IDF วิธีนี้ใช้ฟังก์ชัน TfidfVectorizer ไลบรารี NLTK ในภาษาไพธอน ในการคำนวณความถี่คำที่เกิดขึ้นในเอกสารทั้งหมด

3.3.6 การเลือกคุณลักษณะ (Feature Selection)

ใช้ TF-IDF และวิธีการทดสอบไคสแควร์สำหรับการเลือกคุณลักษณะ กำหนดค่า k=800 ซึ่งหมายถึงจำนวนคุณลักษณะ 800 รายการที่เกี่ยวข้องกับข้อมูลและมีผลต่อการจำแนก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความคิดเห็นมากที่สุด วิธีการนี้เริ่มต้นด้วยการสร้างเมทริกซ์ TF-IDF เพื่อประเมินคำสำคัญตามด้วยการทดสอบไคสแควร์เพื่อกำหนดความสัมพันธ์ของแต่ละเทอมกับหมวดหมู่เป้าหมาย ส่งผลให้มีการเลือกคำศัพท์ 800 อันดับแรกตามคะแนนไคสแควร์ วิธีการนี้จะเพิ่มประสิทธิภาพการทำงานของแบบจำลองโดยลดความซับซ้อนของข้อมูลและมุ่งเน้นไปที่คุณลักษณะที่สำคัญที่สุด

3.4 การพัฒนาแบบจำลอง (Model Building)

ผู้วิจัยจะนำข้อมูลของความคิดเห็นชุดข้อมูลฝึกสอนมาแบ่งออกเป็นเซตย่อย โดยใช้เทคนิคการตรวจสอบด้วยวิธีการตรวจสอบไขว้ (K-fold cross-validation) กำหนดค่า k=10 จากนั้นข้อมูลเซตย่อยของชุดข้อมูลฝึกสอนจะถูกนำไปฝึกฝนเพื่อพัฒนาแบบจำลอง โดยการพัฒนาแบบจำลองจะถูกแบ่งออกเป็นสองขั้นตอนดังนี้

1. **การแปลงข้อมูลแบบหลายฉลาก (Multi-label transformation) :** นำชุดข้อมูลฝึกสอนมาแปลงประเภทข้อมูลแบบหลายฉลากให้เป็นประเภทแบบฉลากเดียว (Single-Label Classification) โดยการศึกษานี้ได้นำเสนอ 3 วิธีได้แก่ BR (Binary Relevance), LP (Label Powerset), และ CC (Classifier Chains) ซึ่งผู้วิจัยใช้ไลบรารี ‘skmultilearn’ ในภาษาไพธอนสำหรับขั้นตอนนี้
2. **การจำแนกประเภท (Classification) :** หลังจากการแปลงข้อมูลแล้วจะนำชุดข้อมูลมาจำแนกประเภท โดยใช้วิธีการจำแนก 4 แบบคือ NB (Naive Bayes), LR (Logistic Regression), SVM (Support Vector Machine), และ RF (Random Forest) ซึ่งมีการกำหนดค่าพารามิเตอร์ในแต่ละวิธีดังแสดงในตารางที่ 3.4

ตารางที่ 3.4 การตั้งค่าพารามิเตอร์

อัลกอริทึม	การตั้งค่าพารามิเตอร์
Naïve Bayes	Alpha = 1.0
Logistic Regression	C=1.0
Support Vector Machine	kernel='linear', C=1
Random Forest	N =200

โดยข้อมูลที่ได้จากการแปลงข้อมูลแบบหลายฉลาก (Multi-label transformation) 1 วิธี จะดำเนินการจำแนกประเภท (Classification) กับ 4 วิธีในขั้นถัดไป ด้วยพารามิเตอร์ที่ตั้งค่าไว้ โดยแสดงตัวอย่างโค้ดการพัฒนาแบบจำลองในรูปที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```python
# นำเข้าไลบรารีและโมดูล
from skmultilearn.problem_transform import BinaryRelevance
from skmultilearn.problem_transform import LabelPowerset
from skmultilearn.problem_transform import ClassifierChain
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier

# แปลงข้อมูลด้วย BinaryRelevance และจำแนกประเภททั้ง 4 วิธี
classifier1 = BinaryRelevance(MultinomialNB(alpha=1.0))
classifier2 = BinaryRelevance(LogisticRegression(C=1.0))
classifier3 = BinaryRelevance(SVC(kernel='linear', C=1))
classifier4 = BinaryRelevance(RandomForestClassifier(n_estimators=200))

# แปลงข้อมูลด้วย LabelPowerset และจำแนกประเภททั้ง 4 วิธี
classifier5 = LabelPowerset(MultinomialNB(alpha=1.0))
classifier6 = LabelPowerset(LogisticRegression(C=1.0))
classifier7 = LabelPowerset(SVC(kernel='linear', C=1))
classifier8 = LabelPowerset(RandomForestClassifier(n_estimators=200))

# แปลงข้อมูลด้วย ClassifierChain และจำแนกประเภททั้ง 4 วิธี
classifier9 = ClassifierChain(MultinomialNB(alpha=1.0))
classifier10 = ClassifierChain(LogisticRegression(C=1.0))
classifier11 = ClassifierChain(SVC(kernel='linear', C=1))
classifier12 = ClassifierChain(RandomForestClassifier(n_estimators=200))
```

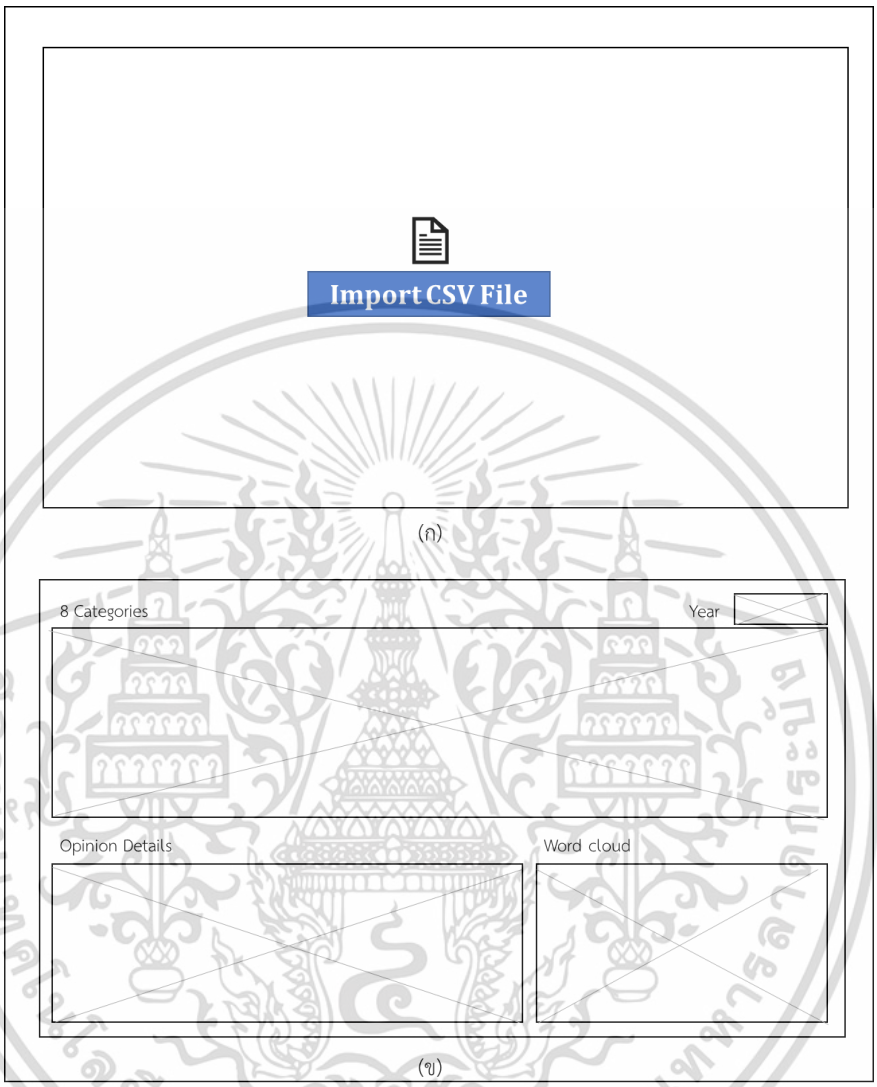
รูปที่ 3.3 ตัวอย่างโค้ดการพัฒนาแบบจำลอง

3.5 การวัดประสิทธิภาพของแบบจำลอง (Model Evaluation)

การประเมินแบบจำลองการจำแนกประเภทหลายฉลากของความคิดเห็นนักท่องเที่ยว จะใช้ข้อมูลทดสอบ (Testing data) ที่ได้แบ่งไว้ 20% ในขั้นตอนแรก มาทดสอบกับแบบจำลองที่มีประสิทธิภาพมากที่สุดที่ได้จากขั้นตอนการพัฒนาแบบจำลอง โดยจะพิจารณาประสิทธิภาพจากค่าต่างๆ ได้แก่ ความถูกต้อง (Accuracy), ความเที่ยงตรง (Precision), ความระลึก (Recall), F1-score และ Hamming loss

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6 การออกแบบส่วนประสานงานผู้ใช้



รูปที่ 3.4 โครงสร้างของแอปพลิเคชัน

รูปที่ 3.4 แสดงโครงสร้างของแอปพลิเคชัน ประกอบไปด้วย 2 ส่วน ได้แก่ หน้าที่ 1 ดังรูปที่ 3.4 (ก) ของแอปพลิเคชันที่ให้ผู้ใช้งานได้นำเข้า (Import) ชุดข้อมูลความคิดเห็นที่อยู่ในรูปแบบไฟล์ CSV และหน้าที่ 2 ดังรูปที่ 3.4 (ข) เมื่อผู้ใช้นำเข้า (Import) ชุดข้อมูลเรียบร้อยแล้ว ผลลัพธ์การจำแนกความคิดเห็นที่ได้นำเข้าจะถูกแสดงผลผ่านแดชบอร์ด โดยมีส่วนประกอบ 4 ส่วนคือ

- ฟังก์ชันการเลือกปีของข้อมูล
- กรอบแสดงผลลัพธ์จำนวนความคิดเห็นของหมวดต่าง ๆ ทั้งหมด 8 หมวด
- กรอบแสดงรายละเอียดความคิดเห็นต้นฉบับ
- กรอบแสดง Word cloud

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

ในบทนี้นำเสนอผลการทดลองและผลการเปรียบเทียบการพัฒนาแบบจำลองการจำแนกความคิดเห็นด้วยวิธีต่างๆ เพื่อให้แบบจำลองมีประสิทธิภาพสูงที่สุด สามารถจำแนกข้อความได้ถูกต้องที่สุด พร้อมทั้งแสดงผลการจำแนกผ่านแดชบอร์ด โดยมีผลการวิจัยและอภิปรายดังนี้

4.1 ผลการวิจัย

ผู้วิจัยได้นำชุดข้อมูลสำหรับการฝึกสอนมาพัฒนาแบบจำลองและทำการเปรียบเทียบวิธีการแปลงข้อมูลแบบหลายฉลาก (Multi-Label Transformation) 3 วิธีคือ Binary Relevance, Classifier Chain และ Label-Power Set ร่วมกับการจำแนกประเภท (Classification) 4 วิธีคือ นาอีฟเบย์ (Naïve Bayes), การถดถอยโลจิสติค (Logistic Regression), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และการสุ่มแบบป่า (Random Forest) ด้วยพารามิเตอร์ที่กำหนด ดังแสดงในตารางที่ 3.4 ซึ่งในแบบจำลองแต่ละวิธีจะทำการทดลองทั้งการใช้ชุดข้อมูลเดิม ชุดข้อมูลที่ผ่านการเสริมข้อมูล ชุดข้อมูลที่ผ่านการเลือกคุณลักษะข้อมูล และชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล โดยผลการทดลองในแต่ละวิธีที่ผ่านวิธีการตรวจสอบไขว้ (K-fold cross-validation) แสดงในตารางที่ 4.1

ตารางที่ 4.1 ผลการทดลองการจำแนกประเภทข้อความในแต่ละวิธีบนชุดข้อมูลที่ต่างกัน

ชุดข้อมูล	การแปลงข้อมูล	การจำแนกข้อมูล	Accuracy	Precision	Recall	F1-Score	Hamming Loss
ชุดข้อมูลเดิม	BR	NB	0.53	0.98	0.41	0.52	0.07
		LR	0.60	0.96	0.53	0.67	0.06
		SVM	0.74	0.93	0.74	0.82	0.04
		RF	0.72	0.92	0.73	0.80	0.04
	LP	NB	0.69	0.88	0.60	0.66	0.06
		LR	0.74	0.91	0.70	0.78	0.05
		SVM	0.76	0.90	0.74	0.81	0.05
		RF	0.76	0.88	0.73	0.80	0.05
	CC	NB	0.60	0.95	0.50	0.62	0.07
		LR	0.70	0.90	0.63	0.71	0.07
		SVM	0.79	0.90	0.78	0.83	0.04
		RF	0.77	0.90	0.76	0.81	0.04

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

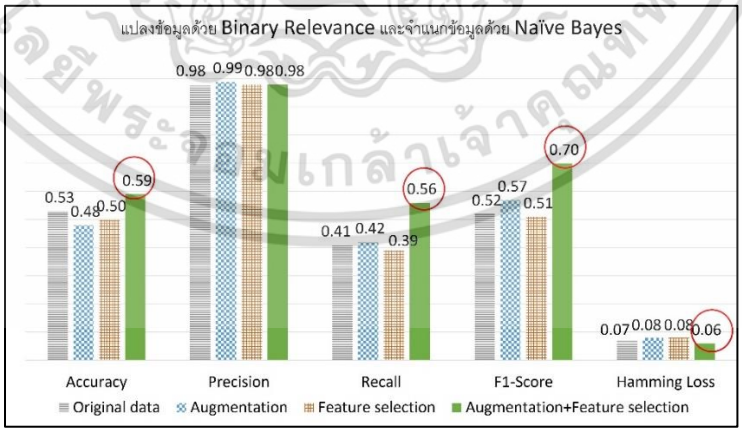
ชุดข้อมูล	การแปลงข้อมูล	การจำแนกข้อมูล	Accuracy	Precision	Recall	F1-Score	Hamming Loss
ชุดข้อมูลที่ผ่านการเสริมข้อมูล	BR	NB	0.48	0.99	0.42	0.57	0.08
		LR	0.58	0.97	0.56	0.69	0.06
		SVM	0.76	0.94	0.80	0.86	0.04
		RF	0.75	0.95	0.80	0.87	0.03
	LP	NB	0.75	0.89	0.71	0.77	0.05
		LR	0.78	0.91	0.76	0.83	0.05
		SVM	0.81	0.93	0.81	0.87	0.04
		RF	0.81	0.92	0.83	0.87	0.04
	CC	NB	0.56	0.97	0.52	0.66	0.07
		LR	0.70	0.89	0.70	0.77	0.06
		SVM	0.81	0.92	0.84	0.88	0.04
		RF	0.78	0.95	0.81	0.88	0.03
ชุดข้อมูลที่ผ่านเลือกคุณลักษณะข้อมูล	BR	NB	0.50	0.98	0.39	0.51	0.08
		LR	0.58	0.97	0.51	0.65	0.06
		SVM	0.72	0.92	0.71	0.80	0.04
		RF	0.71	0.91	0.72	0.80	0.04
	LP	NB	0.69	0.88	0.59	0.64	0.07
		LR	0.74	0.91	0.69	0.77	0.05
		SVM	0.76	0.91	0.73	0.8	0.05
		RF	0.74	0.89	0.71	0.78	0.05
	CC	NB	0.58	0.96	0.49	0.62	0.07
		LR	0.69	0.90	0.62	0.71	0.07
		SVM	0.78	0.90	0.76	0.82	0.05
		RF	0.76	0.89	0.74	0.80	0.05

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูล	การแปลงข้อมูล	การจำแนกข้อมูล	Accuracy	Precision	Recall	F1-Score	Hamming Loss
ชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล	BR	NB	0.59	0.98	0.56	0.70	0.06
		LR	0.60	0.96	0.58	0.71	0.06
		SVM	0.75	0.94	0.80	0.86	0.04
		RF	0.80	0.92	0.87	0.89	0.03
	LP	NB	0.76	0.90	0.73	0.79	0.05
		LR	0.78	0.92	0.77	0.83	0.04
		SVM	0.80	0.92	0.81	0.86	0.04
		RF	0.81	0.92	0.82	0.87	0.04
	CC	NB	0.66	0.96	0.64	0.77	0.05
		LR	0.71	0.89	0.71	0.78	0.06
		SVM	0.79	0.90	0.83	0.86	0.04
		RF	0.83	0.92	0.87	0.89	0.03

4.1.1 เปรียบเทียบผลการทดสอบแบบจำลองด้วยชุดข้อมูลที่ต่างกัน

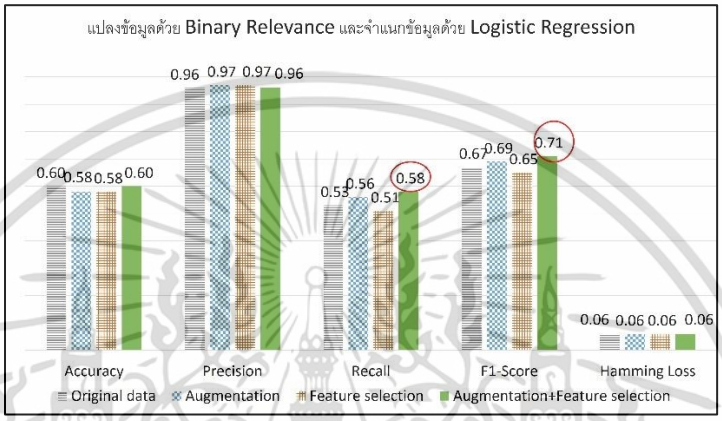
โดยผู้วิจัยได้นำเสนอกราฟแผนภูมิแท่งซึ่งสรุปผลมาจากตารางที่ 4.1 สำหรับการเปรียบเทียบผลการทดสอบแบบจำลองด้วยชุดข้อมูลที่ต่างกัน ได้แก่ 1) ชุดข้อมูลเดิม 2) ชุดข้อมูลที่ผ่านการเสริมข้อมูล 3) ชุดข้อมูลที่ผ่านการเลือกคุณลักษณะข้อมูล และ 4) ชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล ด้วยวิธีการแปลงข้อมูลและจำแนกข้อมูลแต่ละวิธี ดังแสดงในรูปที่ 4.1 - 4.12



รูปที่ 4.1 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน

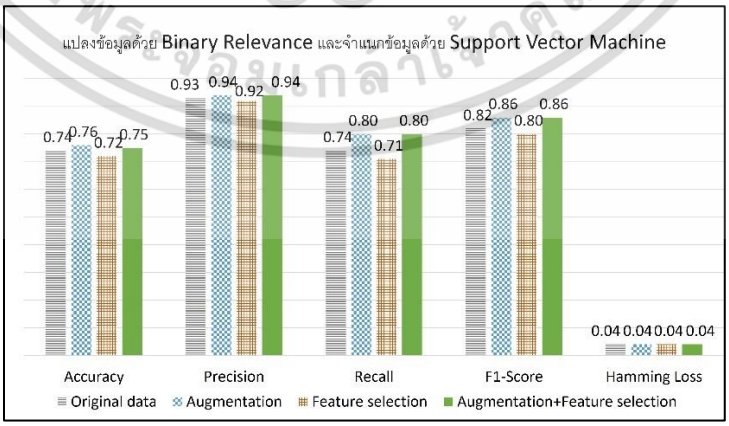
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.1 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Naïve Bayes บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.59, Recall เท่ากับ 0.56, F1-Score เท่ากับ 0.70 และ Hamming loss เท่ากับ 0.06 ในส่วนของค่า Precision เท่ากับ 0.98 ซึ่งมีค่าใกล้เคียงกับข้อมูลชุดอื่นๆ



รูปที่ 4.2 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน

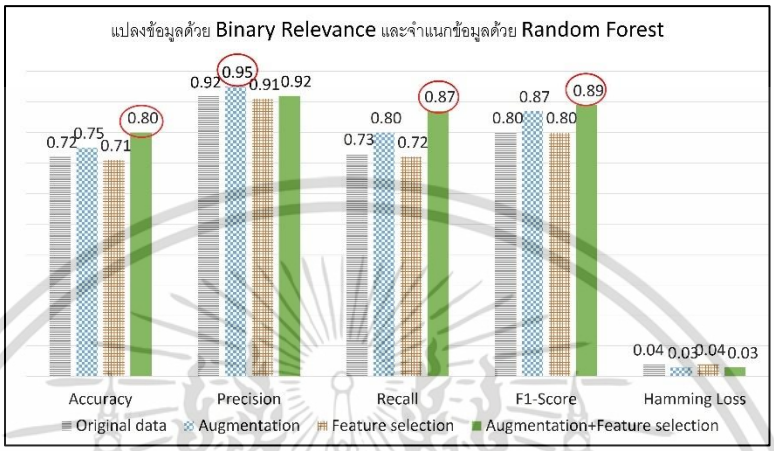
จากรูปที่ 4.2 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Logistic Regression บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะมีค่า Recall เท่ากับ 0.58 และ F1-Score เท่ากับ 0.71 ซึ่งมีค่าสูงที่สุดเมื่อเทียบกับชุดข้อมูลอื่น ในส่วนของค่า Accuracy ที่เท่ากับ 0.60, Precision เท่ากับ 0.96 และค่า Hamming loss เท่ากับ 0.06 ซึ่งมีความใกล้เคียงกับข้อมูลชุดอื่นๆ



รูปที่ 4.3 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน

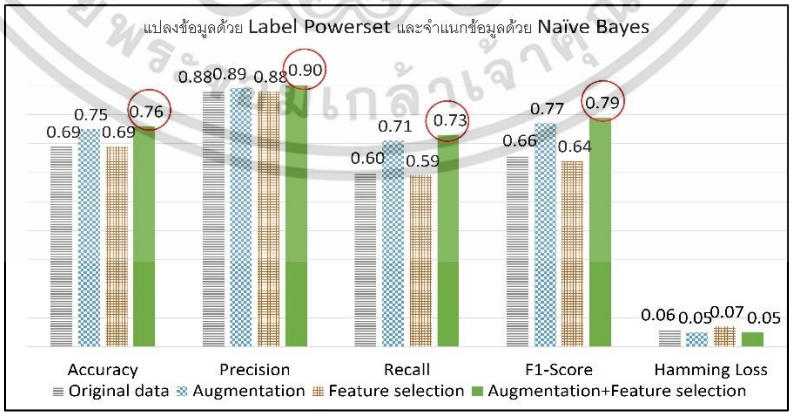
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.3 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Support Vector Machine บนชุดข้อมูลทั้ง 3 ชุดมีค่า Accuracy, Precision, Recall, F1-Score และ Hamming loss ใกล้เคียงกัน



รูปที่ 4.4 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน

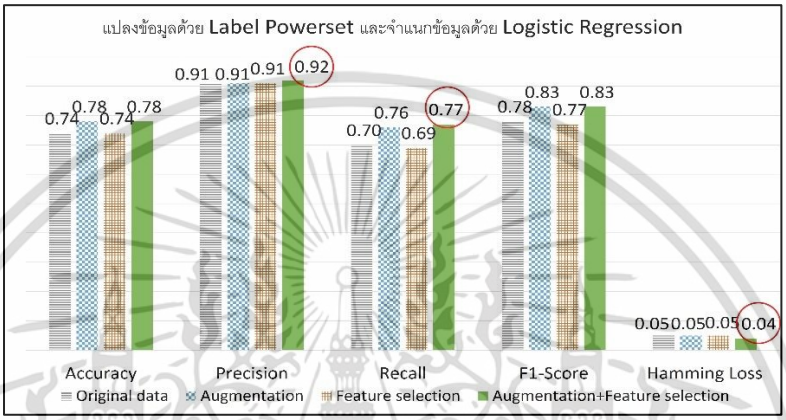
จากรูปที่ 4.4 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Binary Relevance และจำแนกข้อมูลด้วย Random Forest บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.80, Recall เท่ากับ 0.87, F1-Score เท่ากับ 0.89 และ Hamming loss เท่ากับ 0.03 มีเฉพาะในส่วนของค่า Precision เท่ากับ 0.92 เท่านั้นที่มีค่าน้อยกว่าการทดลองในชุดข้อมูลที่ผ่านการเสริมข้อมูล



รูปที่ 4.5 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน

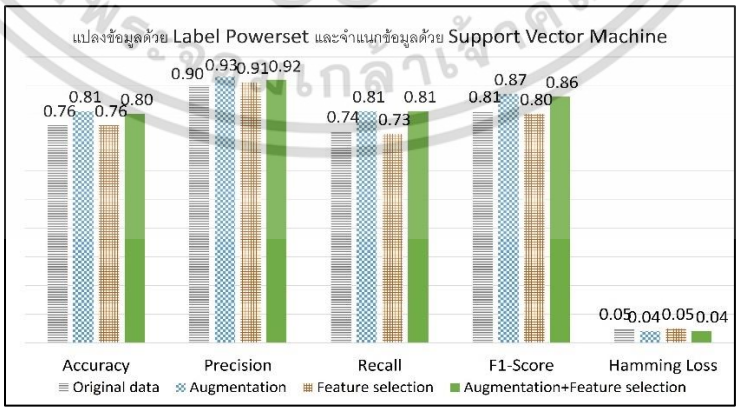
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.5 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Naïve Bayes บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.76, Precision เท่ากับ 0.90, Recall เท่ากับ 0.73, F1-Score เท่ากับ 0.79 และ Hamming loss เท่ากับ 0.05



รูปที่ 4.6 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน

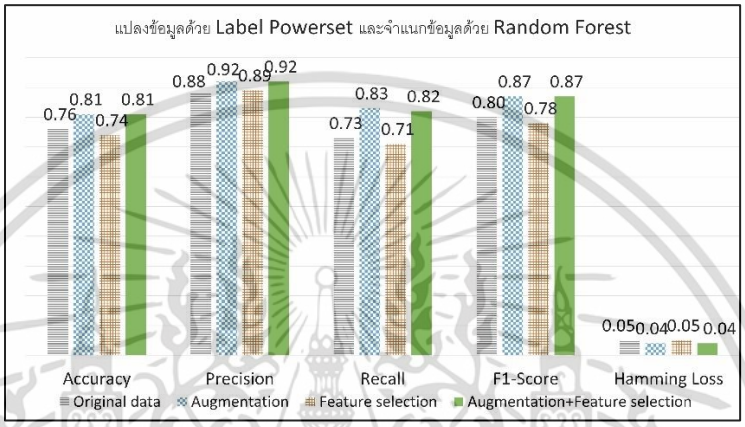
จากรูปที่ 4.6 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Logistic Regression บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายโดยมีค่า Precision เท่ากับ 0.92, Recall เท่ากับ 0.77 และ Hamming loss เท่ากับ 0.04 เป็นค่าสูงสุดเมื่อเปรียบเทียบกับชุดข้อมูลอื่น ในส่วนของค่า Accuracy ที่เท่ากับ 0.78 และ F1-Score เท่ากับ 0.83 ซึ่งมีค่าใกล้เคียงกับชุดข้อมูลอื่น



รูปที่ 4.7 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน

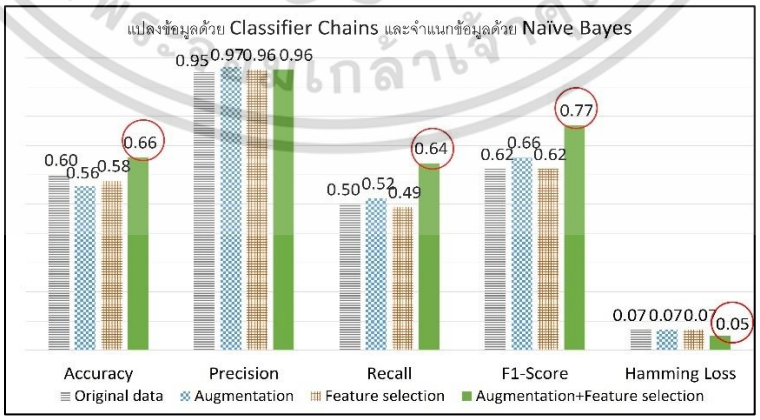
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.7 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Support Vector Machine บนชุดข้อมูลที่ผ่านการเสริมข้อมูล และชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล มีค่า Accuracy, Precision, Recall, F1-Score และ Hamming loss ใกล้เคียงกัน ซึ่งมีค่ามากกว่าการทดลองบนชุดข้อมูลเดิม และบนชุดข้อมูลที่ผ่านการเลือกคุณลักษณะข้อมูล



รูปที่ 4.8 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน

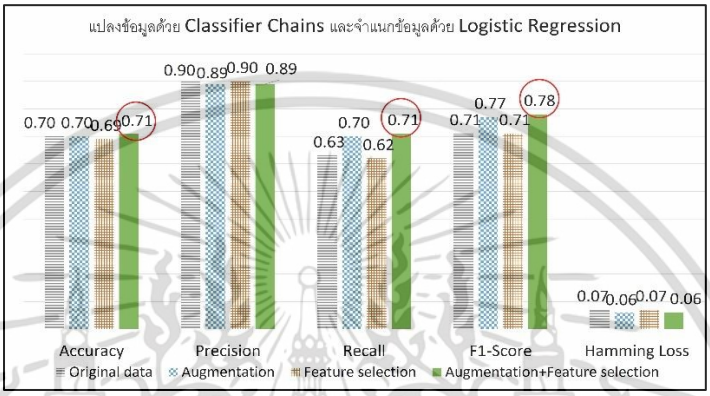
จากรูปที่ 4.8 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Label Powerset และจำแนกข้อมูลด้วย Random Forest บนชุดข้อมูลที่ผ่านการเสริมข้อมูล และชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล มีค่า Accuracy, Precision, Recall, F1-Score และ Hamming loss ใกล้เคียงกัน ซึ่งมีค่ามากกว่าการทดลองบนชุดข้อมูลเดิม และบนชุดข้อมูลที่ผ่านการเลือกคุณลักษณะข้อมูล



รูปที่ 4.9 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Naïve Bayes ในชุดข้อมูลที่ต่างกัน

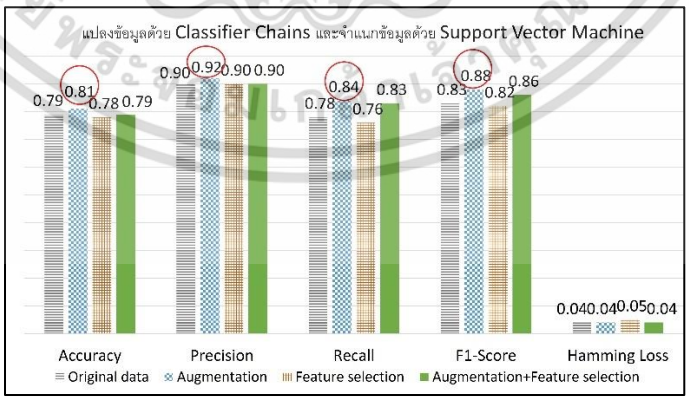
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.9 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Naïve Bayes บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.66, Recall เท่ากับ 0.64, F1-Score เท่ากับ 0.77 และ Hamming loss เท่ากับ 0.05 โดยที่มีส่วนของค่า Precision ที่เท่ากับ 0.96 เท่านั้นที่มีค่าใกล้เคียงกับชุดข้อมูลอื่น



รูปที่ 4.10 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Logistic Regression ในชุดข้อมูลที่ต่างกัน

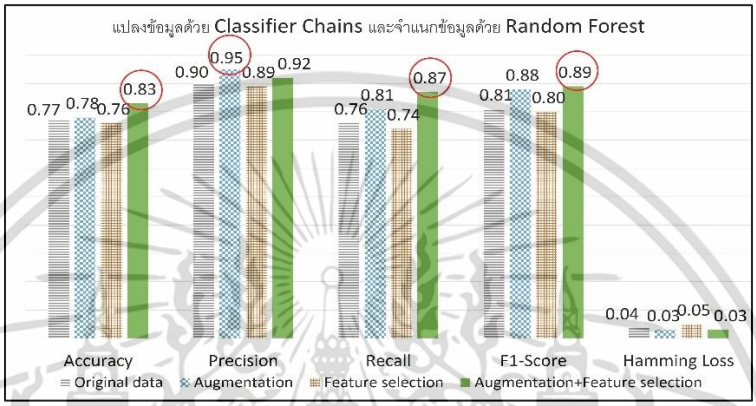
จากรูปที่ 4.10 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Logistic Regression บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.71, Recall เท่ากับ 0.71, F1-Score เท่ากับ 0.78 และ Hamming loss เท่ากับ 0.06 โดยที่มีส่วนของค่า Precision ที่เท่ากับ 0.89 เท่านั้นที่เป็นค่าใกล้เคียงกับชุดข้อมูลอื่น



รูปที่ 4.11 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Support Vector Machine ในชุดข้อมูลที่ต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.11 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Support Vector Machine บนชุดข้อมูลที่ผ่านการเสริมข้อมูลให้ประสิทธิภาพในการทำนายสูงสุด โดยมีค่า Accuracy เท่ากับ 0.81, Precision เท่ากับ 0.92, Recall เท่ากับ 0.84 และ F1-Score เท่ากับ 0.88 ในส่วนของค่า Hamming loss เท่ากับ 0.04 ซึ่งเป็นค่าใกล้เคียงกับชุดข้อมูลอื่น



รูปที่ 4.12 เปรียบเทียบผลการทดสอบแบบจำลองสำหรับการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Random Forest ในชุดข้อมูลที่ต่างกัน

จากรูปที่ 4.12 พบว่าผลการทดสอบแบบจำลองการจำแนกข้อความโดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Random Forest บนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล ให้ผลการทำนายสูงสุดโดยมีค่า Accuracy ที่เท่ากับ 0.83, Recall เท่ากับ 0.87, F1-Score เท่ากับ 0.89 และ Hamming loss เท่ากับ 0.03 โดยมีส่วนของค่า Precision ที่เท่ากับ 0.92 ที่มีค่าน้อยกว่าชุดข้อมูลที่ผ่านการเสริมข้อมูล

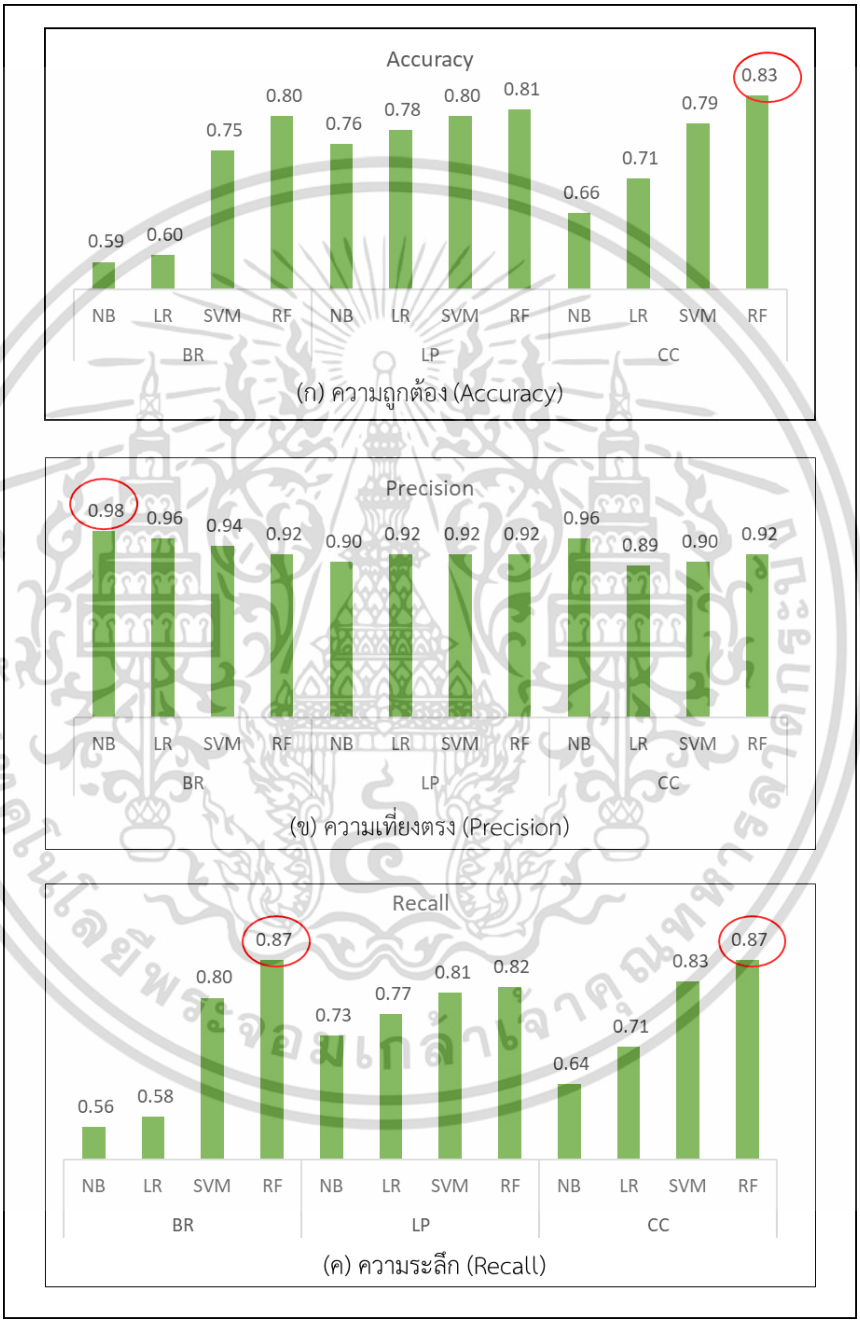
จากรูปที่ 4.1 - 4.12 แสดงให้เห็นว่าการทดสอบแบบจำลองการจำแนกข้อความส่วนใหญ่ได้ผลลัพธ์การจำแนกที่ดีบนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล ดังนั้นผู้วิจัยจะทำการเปรียบเทียบผลการทดสอบแบบจำลองบนชุดข้อมูลดังกล่าวกับแปลงข้อมูลด้วยวิธีการแปลงข้อมูลและจำแนกข้อมูลที่ต่างกัน ในขั้นต่อไป

4.1.2 เปรียบเทียบผลการทดสอบแบบจำลองด้วยวิธีการแปลงข้อมูลและจำแนกข้อมูลที่ต่างกัน

ผู้วิจัยได้นำเสนอกราฟแผนภูมิแท่งโดยสรุปผลจากตารางที่ 4.1 สำหรับการเปรียบเทียบผลการทดสอบแบบจำลองบนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล โดยพัฒนาแบบจำลองด้วยวิธีการแปลงข้อมูลทั้ง 3 วิธีคือ และจำแนกข้อมูลแต่ละวิธี Binary Relevance, Classifier Chain และ Label-Power Set ร่วมกับการจำแนกประเภท (Classification)

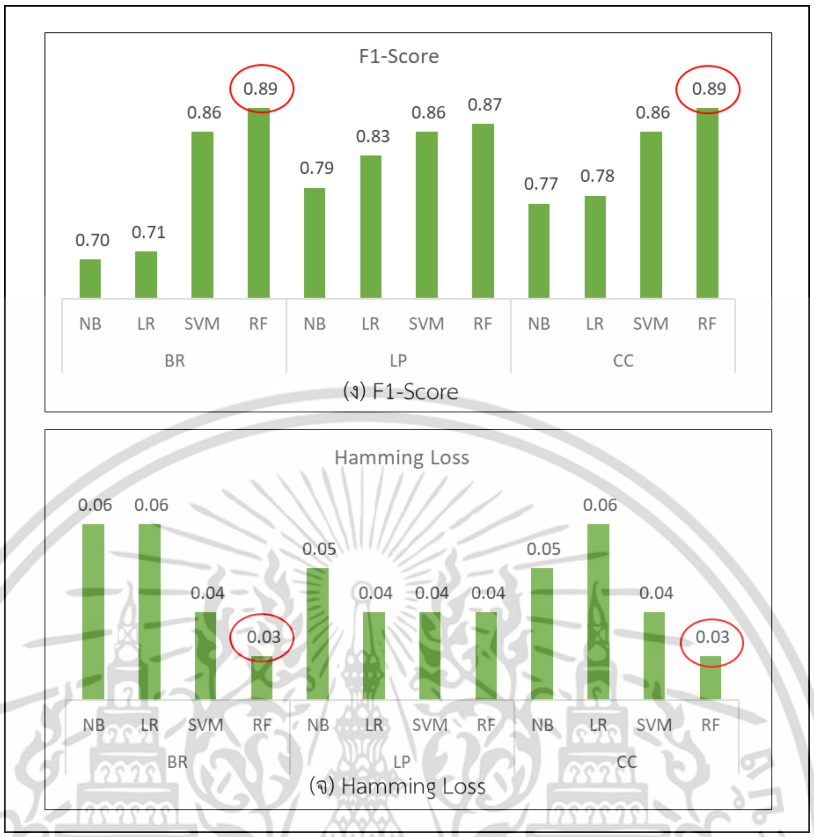
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4 วิธีคือ นาอีฟเบย์ (Naïve Bayes), การถดถอยโลจิสติค (Logistic Regression), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และการสุ่มแบบป่า (Random Forest) ด้วยพารามิเตอร์ที่กำหนด ผลการเปรียบเทียบแสดงในรูปที่ 4.13 (ก) – (จ)



รูปที่ 4.13 เปรียบเทียบค่าความถูกต้อง (Accuracy), ความเที่ยงตรง (Precision), ความระลึก (Recall), F1-Score และ Hamming Loss ด้วยวิธีการแปลงข้อมูล ร่วมกับวิธีจำแนกข้อมูลแต่ละวิธี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.13 (ต่อ) เปรียบเทียบค่าความถูกต้อง (Accuracy), ความเที่ยงตรง (Precision), ความระลึก (Recall), F1-Score และ Hamming Loss ด้วยวิธีการแปลงข้อมูล ร่วมกับวิธีจำแนกข้อมูลแต่ละวิธี

จากรูปที่ 4.13 พบว่าผลการทดสอบการสร้างแบบจำลองด้วยเทคนิควิธีการตรวจสอบไขว้ (K-fold cross-validation) k=10 ในแบบจำลองการจำแนกข้อความบนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล โดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Random Forest ให้ผลการทำนายสูงที่สุดเมื่อเปรียบเทียบกับวิธีอื่นๆ โดยมีค่า Accuracy เท่ากับ 0.83, Recall เท่ากับ 0.87 และ F1-Score เท่ากับ 0.89 และ Hamming loss เท่ากับ 0.03 และสำหรับค่า Precision ที่อาจมีค่าน้อยกว่าแบบจำลองอื่นๆ โดยมีค่าเท่ากับ 0.92 แต่หากพิจารณาภาพรวมของประสิทธิภาพแบบจำลองแล้ววิธีการดังกล่าวถือว่ามีประสิทธิภาพสูงที่สุด

4.1.3 ประเมินประสิทธิภาพของแบบจำลองบนชุดข้อมูลทดสอบ

ผู้วิจัยเลือกใช้วิธีสร้างแบบจำลองบนชุดข้อมูลที่ผ่านการเสริมข้อมูลร่วมกับการเลือกคุณลักษณะของข้อมูล โดยใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และจำแนกข้อมูลด้วย Random Forest ซึ่งเป็นวิธีที่มีประสิทธิภาพสูงที่สุด มาทดสอบกับข้อมูลชุดทดสอบเพื่อประเมิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสิทธิภาพของแบบจำลองในการนำไปจำแนกข้อความและแสดงผลผ่านแดชบอร์ด โดยผลการทดสอบแสดงดังตารางที่ 4.2

ตารางที่ 4.2 ผลการประเมินประสิทธิภาพของแบบจำลองบนชุดข้อมูลทดสอบ

Classifier Model	Accuracy	Precision	Recall	F1-Score	Hamming Loss
Train set	0.83	0.92	0.87	0.89	0.03
Test set	0.80	0.90	0.81	0.85	0.04

เมื่อพิจารณาตารางที่ 4.2 พบว่าประสิทธิภาพการจำแนกประเภทข้อความบนชุดข้อมูลการทดสอบมีค่า Accuracy, Precision, Recall, F1-Score และ Hamming Loss ไม่แตกต่างจากข้อมูลชุดฝึกสอน โดยมีค่า Accuracy เท่ากับ 0.80, Precision เท่ากับ 0.90, Recall เท่ากับ 0.81, F1-Score เท่ากับ 0.85 และ Hamming Loss เท่ากับ 0.04 จึงสามารถนำแบบจำลองไปใช้ได้อย่างมีประสิทธิภาพ

4.2 การแสดงผลการจำแนกข้อความผ่านแดชบอร์ด

จากแบบจำลองที่มีประสิทธิภาพสูงสุด ผู้วิจัยได้นำมาจัดทำเป็นแดชบอร์ดแสดงผลการจำแนกเพื่อให้ผู้ใช้งานสามารถวิเคราะห์ผลการจำแนกได้อย่างสะดวก รวดเร็ว และมีความถูกต้องมากยิ่งขึ้น ผู้วิจัยได้แบ่งออกเป็น 2 ส่วนได้แก่ 1) พัฒนาแดชบอร์ด 2) วิเคราะห์ผลการจำแนกผ่านแดชบอร์ด โดยมีรายละเอียดดังนี้

4.2.1 พัฒนาแดชบอร์ด

การนำเสนอผลลัพธ์การจำแนกข้อความผ่านแดชบอร์ดได้พัฒนาโดยใช้โปรแกรม Microsoft Power BI ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ธุรกิจที่ออกแบบมาเพื่อช่วยให้ผู้ใช้งานสามารถสร้างรายงานแดชบอร์ดสำหรับการวิเคราะห์ การประมวลผล ได้อย่างน่าสนใจ เพื่อประกอบการตัดสินใจ อีกทั้งยังสามารถอัพเดตข้อมูลได้อย่างทันที โดยข้อมูลที่เก็บรวบรวมจะถูกแปลงเป็นแผนภูมิหรือการแสดงผลที่เป็นภาพ ทำให้ผู้ใช้สามารถเข้าใจและวิเคราะห์ข้อมูลได้ง่ายยิ่งขึ้น

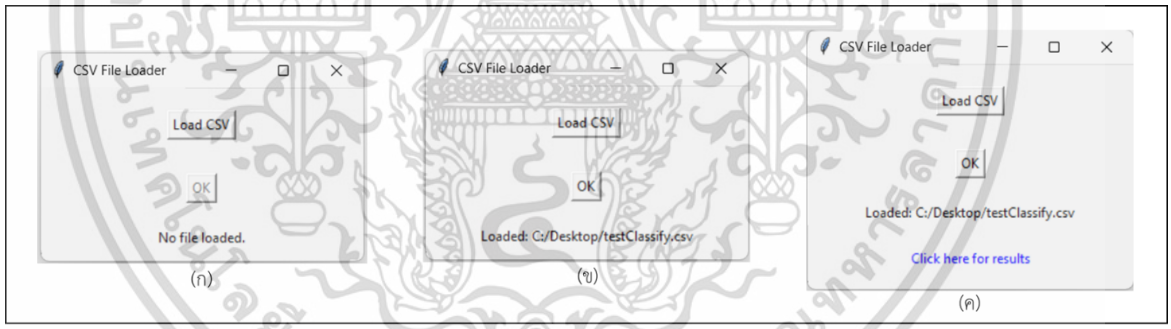
ในการเริ่มต้นกระบวนการจำแนกข้อความแบบอัตโนมัติ ผู้ใช้จำเป็นต้องเตรียมไฟล์ชุดข้อมูลความคิดเห็นในรูปแบบ CSV (ค่าที่คั่นด้วยเครื่องหมายจุลภาค) ตามรูปแบบที่กำหนดคือ คอลัมน์ A กำหนดให้เป็นข้อมูลปี และคอลัมน์ B เป็นข้อมูลของความคิดเห็น ดังแสดงในรูปที่ 4.14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A	B	C	D	E	F	G	H	I	J
1	Year	Text								
2	2563	safety , public transport								
3	2563	I want to find Egyptian food more								
4	2563	I speak french and the food								
5	2563	bad smell from Road water tunnels								
6	2563	the language problem								
7	2563	The price of goods in the airport is too expensive.								
8	2563	traffic so fast and taxi expensive								
9	2563	to stop promoting sex tourism								
10	2563	English								
11	2563	traffic not convenient taxi expensive								
12	2563	make the environment more clean better public facilities								
13	2563	public transport								

รูปที่ 4.14 รูปแบบไฟล์ที่กำหนด

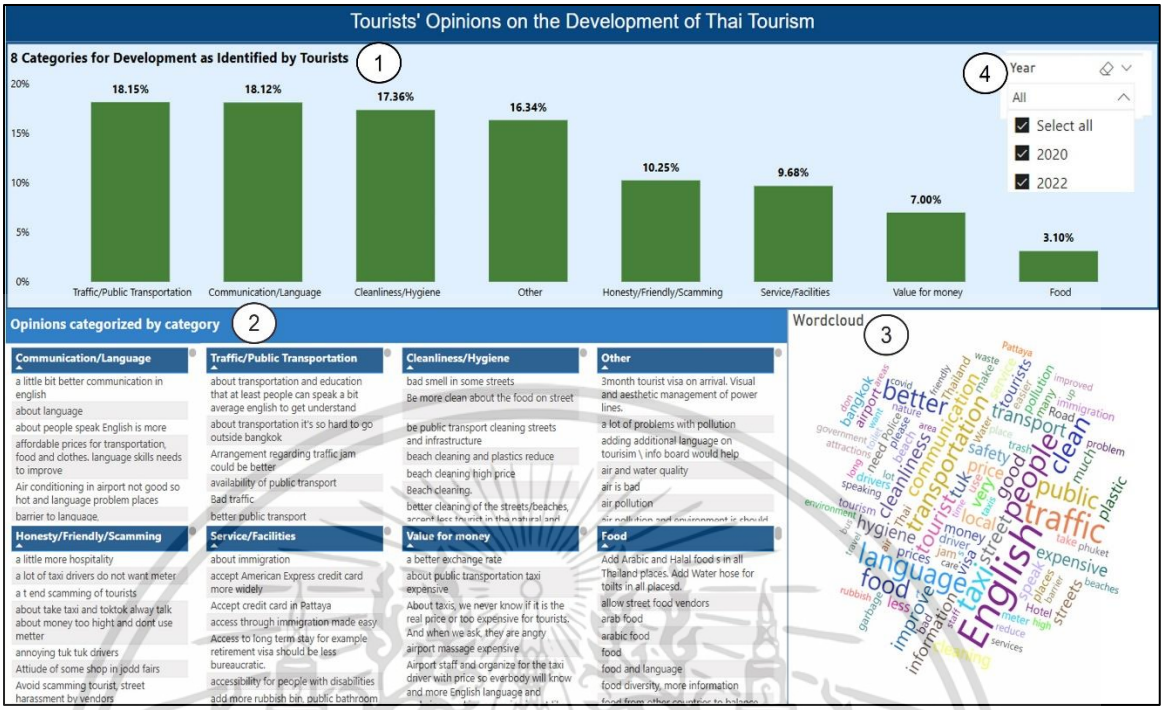
หลังจากเตรียมไฟล์เรียบร้อยแล้ว จึงนำเข้าข้อมูลผ่านหน้าต่างแอปพลิเคชันโดยกดปุ่ม ‘Load CSV’ เพื่อนำเข้าข้อมูล ดังแสดงในรูปที่ 4.15 (ก) หลังจากการนำเข้าแล้วหน้าต่างจะแสดงปุ่ม ‘OK’ และแสดงที่อยู่ของไฟล์ที่นำเข้า ดังแสดงในรูปที่ 4.15 (ข) หากที่อยู่ของไฟล์ถูกต้องกดปุ่ม ‘OK’ เพื่อดำเนินการจำแนกข้อความแบบอัตโนมัติ และเมื่อโปรแกรมจำแนกข้อความสำเร็จแล้วจะแสดงคำว่า ‘Click here for results’ ดังแสดงในรูปที่ 4.15 (ค) โดยสามารถคลิกดูผลลัพธ์การจำแนกข้อความผ่านแดชบอร์ดได้ที่คำนี้



รูปที่ 4.15 หน้าต่างแอปพลิเคชันการนำเข้าข้อมูล

ผู้ใช้สามารถกดดูผลลัพธ์การจำแนกข้อความผ่านแดชบอร์ดได้ที่คำว่า ‘Click here for results’ หลังจากกดแล้วผลการจำแนกข้อความในรูปแบบแดชบอร์ดจะแสดงผลผ่านทางหน้าเว็บไซต์ ดังแสดงในรูปที่ 4.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.16 หน้าต่างเว็บไซต์แดชบอร์ดแสดงผลการจำแนกข้อความ

จากรูปที่ 4.16 หน้าต่างเว็บไซต์แดชบอร์ดประกอบไปด้วย 4 ส่วน ได้แก่

1. **8 Categories for Development as Identified by Tourists :** แสดงสัดส่วนผลการจำแนกความคิดเห็นตาม 8 หมวดหลักที่ผ่านการจำแนกโดยใช้การเรียนรู้ของเครื่อง (Machine Learning Algorithm)
2. **Opinions categorized by category :** แสดงรายละเอียดของความคิดเห็นประโยคเต็มแบบต้นฉบับ โดยแบ่งตามหมวดต่างๆ ทำให้ผู้ใช้งานสามารถดูข้อความความคิดเห็นที่นักท่องเที่ยวได้ตอบแบบสอบถามมา
3. **Word cloud :** แสดงคำต่างๆ ที่มาจากความคิดเห็น นำมาจัดกลุ่มและขยายคำที่เป็นที่นิยม เพื่อให้เห็นภาพที่ชัดเจนมากยิ่งขึ้น
4. **Year :** เป็นฟังก์ชันการเลือกปีของข้อมูลที่ผู้ใช้สามารถเลือกได้ว่าต้องการดูข้อมูลของปีใด โดยในรูปที่ 4.16 ผู้วิจัยเลือกดูข้อมูลของปี 2563 และปี 2565

4.2.2 วิเคราะห์ผลการจำแนกผ่านแดชบอร์ด

จากการวิเคราะห์ความคิดเห็นของนักท่องเที่ยวต่างชาติที่นำเสนอในรูปแบบแดชบอร์ด ดังแสดงในรูปที่ 4.16 พบว่า 3 สิ่งแรกที่นักท่องเที่ยวต้องการให้ปรับปรุงในภาคการท่องเที่ยวของประเทศไทยคือ การสื่อสาร/ภาษา การจราจร/การขนส่งสาธารณะ และความสะอาด/สุขอนามัย เมื่อพิจารณาจากรายละเอียดข้อความทั้งหมดแบบประโยคเต็มที่เป็นต้นฉบับ พบว่าภายใต้หัวข้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสื่อสาร/ภาษา นักท่องเที่ยวต้องการให้คนไทยพูดภาษาอังกฤษได้มากขึ้น โดยระบุว่าอุปสรรคด้านภาษาเป็นอุปสรรคสำคัญ โดยในด้านของการจราจร/การขนส่งสาธารณะ รายละเอียดความคิดเห็นได้กล่าวถึงปัญหาการจราจรที่ติดขัดรุนแรง ในขณะที่ด้านความสะอาด/สุขอนามัย มีการแสดงความคิดเห็นเกี่ยวกับกลิ่นไม่พึงประสงค์ตามท้องถนนและชายหาดที่ไม่สะอาด ในส่วนของการวิเคราะห์ Word Cloud ระบุว่า "English" เป็นคำที่กล่าวถึงบ่อยที่สุด ซึ่งแสดงให้เห็นว่าการสื่อสารภาษาอังกฤษเป็นประเด็นหลักที่นักท่องเที่ยวต้องการให้การท่องเที่ยวของไทยพัฒนาในด้านนี้อย่างมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

ในเนื้อหาของบทนี้จะกล่าวสรุปผลการวิจัยและข้อเสนอแนะสำหรับงานวิจัยเรื่องการจำแนกประเภทแบบหลายฉลาก (Multi-Label Classification) ของความคิดเห็นนักท่องเที่ยวต่างชาติในการพัฒนาการท่องเที่ยวไทย

5.1 สรุปผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการจำแนกประเภทแบบหลายฉลาก เพื่อจำแนกความคิดเห็นของนักท่องเที่ยวต่างชาติเกี่ยวกับการพัฒนาการท่องเที่ยวไทย โดยใช้เทคนิคการเรียนรู้ของเครื่องและนำเสนอผลการจำแนกผ่านแดชบอร์ด โดยที่การท่องเที่ยวแห่งประเทศไทย (ททท.) ได้ทำการรวบรวมข้อมูลความคิดเห็นของนักท่องเที่ยวต่างชาติ ในปี 2563 และปี 25665 จำนวนทั้งหมด 2,249 ความคิดเห็น ซึ่งเป็นความคิดเห็นเกี่ยวกับด้านที่การท่องเที่ยวของประเทศไทยควรปรับปรุง ซึ่งข้อมูลความคิดเห็นนี้ถูกแบ่งออกเป็น 8 หมวดหมู่หลัก ได้แก่ 1) การสื่อสาร/ภาษา 2) การจราจร/การขนส่งสาธารณะ 3) ความสะอาด/สุขอนามัย 4) ความซื่อสัตย์/เป็นมิตร/การหลอกลวง 5) บริการ/สิ่งอำนวยความสะดวก 6) ความคุ้มค่าเงิน 7) อาหาร และ 8) อื่นๆ

ผู้วิจัยได้ทำการเพิ่มประสิทธิภาพของแบบจำลองด้วยวิธีการเสริมข้อมูล (Data Augmentation) ร่วมกับการเลือกคุณลักษณะของข้อมูล (Feature Selection) บนชุดข้อมูลฝึกสอน เพื่อให้ได้ชุดข้อมูลที่มีความหลากหลาย และลดความซับซ้อนของข้อมูล จากนั้นจึงนำชุดข้อมูลที่ผ่านการเตรียมข้อมูลแล้วมาแบ่งออกเป็น 10 ชุด (k=10) เพื่อนำไปพัฒนาแบบจำลองด้วยเทคนิคการตรวจสอบไขว้ (K-fold cross validation) ในการประเมินประสิทธิภาพของแบบจำลอง โดยที่ขั้นตอนของการพัฒนาแบบจำลองได้นำเสนอวิธีการแปลงข้อมูลแบบหลายฉลาก (Multi-Label Classification) 3 วิธี ได้แก่ Binary Relevance, Classifier Chain และ Label-Power Set ร่วมกับการจำแนกประเภท (Classification) 4 วิธี ได้แก่ นาอีฟเบย์ (Naïve Bayes), การถดถอยโลจิสติค (Logistic Regression), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และการสุ่มแบบป่า (Random Forest)

ผลการทดลองพบว่าแบบจำลองที่มีประสิทธิภาพสูงที่สุดคือแบบจำลองที่ใช้วิธีการแปลงข้อมูลด้วย Classifier Chains และใช้วิธีการจำแนกประเภทด้วยการสุ่มแบบป่า (Random Forest) และได้นำมาแบบจำลองวิธีดังกล่าวมาทดสอบกับชุดข้อมูลทดสอบ ผลการทดสอบพบว่ามีค่า Accuracy เท่ากับ 80%, Precision เท่ากับ 90%, Recall เท่ากับ 81%, F1-Score เท่ากับ 85% และ Hamming loss 0.04

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการวิเคราะห์ข้อมูลด้วยแดชบอร์ด พบว่า 3 ประเด็นแรกที่นักท่องเที่ยวต้องการให้มีการปรับปรุงมากที่สุด ได้แก่ การสื่อสาร/ภาษา การจราจร/การขนส่งสาธารณะ และความสะอาด/สุขอนามัย ซึ่งสอดคล้องกับผลการวิเคราะห์ Word Cloud ที่พบว่าคำว่า "English" เป็นคำที่ถูกกล่าวถึงบ่อยที่สุด แสดงให้เห็นว่านักท่องเที่ยวให้ความสำคัญกับการสื่อสารเป็นอย่างมาก

จากการวิเคราะห์ผลนี้ สามารถพัฒนาการท่องเที่ยวการท่องเที่ยวตามประเด็นที่สำคัญดังกล่าว ตัวอย่างเช่น ในด้านของการสื่อสาร/ภาษา หน่วยงานการท่องเที่ยวสามารถจัดโปรแกรมการฝึกอบรมภาษาอังกฤษโดยมีเป้าหมายไปที่ผู้ประกอบการในอุตสาหกรรม เช่น ร้านค้า ร้านอาหาร และโรงแรม โดยมีเป้าหมายเพื่อเพิ่มคุณภาพการบริการและยกระดับการสื่อสารกับนักท่องเที่ยว ในทำนองเดียวกัน ในด้านการจราจร/การขนส่งสาธารณะ การสร้างแอปพลิเคชันหรือระบบข้อมูลออนไลน์ที่เผยแพร่รายละเอียดที่เกี่ยวข้องเกี่ยวกับตารางเวลา ที่จอดรถ และตัวเลือกเส้นทางสามารถปรับปรุงการวางแผนการเดินทางสำหรับนักท่องเที่ยวได้ การพัฒนาดังกล่าวอาจช่วยเพิ่มความพึงพอใจให้แก่นักท่องเที่ยวในอนาคต

5.2 ข้อเสนอแนะ

จากการศึกษาและวิจัยในครั้งนี้ ผู้วิจัยมีข้อเสนอแนะดังต่อไปนี้

- เพิ่มเติมกรอบการวิเคราะห์เพื่อจัดหมวดหมู่ความคิดเห็นตามตัวแปรทางประชากรศาสตร์หรือสถานที่ท่องเที่ยวที่เฉพาะเจาะจง ซึ่งจะช่วยให้ผลการวิเคราะห์มีความละเอียดมากยิ่งขึ้นและได้ทราบว่านักท่องเที่ยวมีความคิดเห็นอย่างไรในสถานที่ที่แตกต่างกัน
- พัฒนาแบบจำลองให้สามารถจำแนกความคิดเห็นในหมวดย่อยของกลุ่มอื่นๆ เพิ่มเติมได้ หากจำนวนความคิดเห็นในกลุ่มอื่น ๆ มีจำนวนมากขึ้น
- เพิ่มความแม่นยำของแบบจำลองการจำแนกความคิดเห็น เพื่อให้มั่นใจถึงความแม่นยำสูงสุดที่เป็นไปได้ในการวิเคราะห์ในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

การท่องเที่ยวแห่งประเทศไทย (ททท.). 2566. **โครงการสำรวจเพื่อการวิเคราะห์พฤติกรรมนักท่องเที่ยวระหว่างประเทศ ปี 2566.** กรุงเทพฯ : บริษัท อินทัช รีเสิร์ช แอนด์ คอนซัลแทนซี่ จำกัด

Adil Yaseen Taha and S. Tiun. 2016. “Binary relevance (BR) method classifier of multi-label classification for arabic text.” J. Theoretical and Applied Information Technology. 84 : 414-422.

Anna Glazkova. 2023. “A comparison of text preprocessing techniques for hate and offensive speech detection in Twitter.” J. Springer. 13(155) : 1-28.

Bassam Al-Salemi, Masri Ayob, Graham Kendall, and Shahrul Azman Mohd Noah. 2019. “Multi-label Arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms.” J. Information Processing & Management. 56(1) : 212-227.

Chintrai Puttipornchai, Sapa Chanyachatchawan, and Nuengwong Tuaycharoen. 2022. “Multi-Label Classification for Articles in Thai Journal Database from Article's Abstract.” In Proceedings of **the 19th International Joint Conference on Computer Science and Software Engineering**. Bangkok : IEEE.

Christine P. Chai. 2023. “Comparison of text preprocessing methods.” Cambridge University Press. 29(3) : 509 - 553.

Emre Deniz, Hasan Erbay, and Mustafa Coşar. 2022. “Multi-Label Classification of E-Commerce Customer Reviews via Machine Learning.” J. Axioms. 11(9).

En Yang and Zhaohua Long. 2023. “Research on the Weighting Method Based on Tf-IDF and Apriori Algorithm.” 1003-1005. In Proceedings of **the 6th International Conference on Information Systems and Computer Aided Education**. Dalian : IEEE.

Jesse Read, Bernhard Pfahringer, Geoffrey Holmes, and Eibe Frank. 2021. “Classifier Chains: A Review and Perspectives.” J. Artificial Intelligence Research. 70 : 683-718.

Mariem Bounabi, Karim El Moutaouakil, and Khalid Satori. 2019. “Text classification using Fuzzy TF-IDF and Machine Learning Models.” 1-6. In Proceedings of **the 4th International Conference on Big Data and Internet of Things.** Rabat : ACM.

Mohammadreza Heydarian, Thomas E. Doyle, and Reza Samavi. 2022. “MLCM: Multi-Label Confusion Matrix.” J. IEEE Access. 10 : 19083-19095.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

Muhammad Afzaal and Muhammad Usman. 2015. “A novel framework for aspect-based opinion classification for tourist places.” 1-9. In Proceedings of **the Tenth International Conference on Digital Information Management**. Jeju : IEEE.

Nawal Aljedani, Reem Alotaibi, and Mounira Taileb. 2020. “Multi-Label Arabic Text Classification: An Overview.” J. IJACSA. 11(10) : 2020, 694-706.

Puteri Prameswari, Zulkarnain, Isti Surjandari, and Enrico Laoh. 2017. “Mining online reviews in Indonesia's priority tourist destinations using sentiment analysis and text summarization approach.” 121-126. In Proceedings of **the 8th International Conference on Awareness Science and Technology**. Taichung : IEEE.

Roberto Carlos Morales-Hernández, Joaquín Gutiérrez Jagüey, and David Becerra-Alonso. 2022. “A Comparison of Multi-Label Text Classification Models in Research Articles Labeled With Sustainable Development Goals.” J. IEEE Access. 10 : 123534-123548.

Statista Research Department. 2023**. Tourism Industry in Thailand.** [Online]. Available : https://www.statista.com/statistics/

Tahsin Tasnia Khan, Abid Hassan, Md Faysal Ahamed, and Samiul Islam. 2023. “Multi-label Bengali Abusive Comments Classification using Problem Transformation Method.” In Proceedings of **the 20th International Conference on Electrical Engineering, Computing Science and Automatic Control**. Mexico City : IEEE.

Uchenna Cosmas Ugwuoke, Enesi Femi Aminu, and Ayobami Ekundayo. 2022. “Performing Data Augmentation Experiment to Enhance Model Accuracy: A Case Study of BBC News’ Data.” In Proceedings of **International Conference on Information systems and Emerging Technologies**. SSRN.

Yujia Zhai, Wei Song, Xianjun Liu, Lizhen Liu, and Xinlei Zhao. 2018. “A Chi-Square Statistics Based Feature Selection Method in Text Classification.” 160-163. In Proceedings of **the 9th International Conference on Software Engineering and Service Science**. Beijing : IEEE.

Yushan Xu, Chu Zhang, and Wenyan Song. 2022. “Prioritizing Customer Requirements for Science and Technology Service Platform Based on Improved TF-IDF and Sentiment Analysis.” 210-214. In Proceedings of **the International Conference on Industrial Engineering and Engineering Management**. Kuala Lumpur : IEEE.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

งานวิจัยที่เผยแพร่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นางสาวลลิตา สุระมรรคา
วัน เดือน ปีเกิด	29 มิถุนายน พ.ศ. 2537
ที่อยู่ปัจจุบัน	1/509 การเคหะรามคำแหง ถนนรามคำแหง 190/2 เขตมีนบุรี แขวงมีนบุรี กรุงเทพมหานคร 10510
ประวัติการศึกษา	(2558) วิทยาศาสตรบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (2566) วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ผลงานทางวิชาการ	L. Suramanka and A. Hanskunatai, “Multi-Label Classification of Foreign Tourists’ Opinions on Thailand Tourism Development”, Proceedings of the 2024 9th International Conference on Big Data and Computing, May 24-26, 2024, Bangkok, Thailand

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้