การพัฒนาเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง

โดย
นายภากร กัทชลี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แบบ 1.1
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2567
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

การพัฒนาเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง

โดย
นายภากร กัทชลี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แบบ 1.1
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2567
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

DEVELOPING FRAMEWORK FOR CONSISTENCY EVALUATION OF ARTICLE WRITING STYLEUSING MACHINE LEARNING TECHNIQUES

By

MR. Pagon GATCHALEE

A Thesis Submitted in Partial Fulfillment of the Requirements

for Doctor of Philosophy INFORMATION TECHNOLOGY AND DIGITAL INNOVATION

Department of COMPUTER SCIENCE

Academic Year 2024

Copyright of Silpakorn University

หัวข้อ	การพัฒนาเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง
โดย	นายภากร กัทชลี
สาขาวิชา	เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แบบ 1.1
อาจารย์ที่ปรึกษาหลัก	อาจารย์ ดร. สัจจาภรณ์ ไวจรรยา
อาจารย์ที่ปรึกษาร่วม	ผู้ช่วยศาสตราจารย์ ดร. ณัฐโชติ พรหมฤทธิ์

คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

.....คณบดีคณะวิทยาศาสตร์
(ผู้ช่วยศาสตราจารย์ ดร. กรกช ชั้นจิรกุล)

พิจารณาเห็นชอบโดย

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. อนิราช มิ่งขวัญ)

.....อาจารย์ที่ปรึกษาหลัก
(อาจารย์ ดร. สัจจาภรณ์ ไวจรรยา)

.....อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร. ณัฐโชติ พรหมฤทธิ์)

.....ผู้ทรงคุณวุฒิภายใน
(รองศาสตราจารย์ ดร. ปานใจ ธารทัศนวงศ์)

.....ผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร. อรวรรณ เชาวลิต)

620730008 : เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แบบ 1.1
คำสำคัญ : การตลาดเชิงเนื้อหา, การเรียนรู้ของเครื่อง, การเรียนรู้เชิงลึก, สไตล์บทความ, เนื้อหาจีน

นาย ภากร กัทชลี: การพัฒนาเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : อาจารย์ ดร. สัจจาภรณ์ ไวจรรยา

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อพัฒนาและนำเสนอแนวทางการสร้างเครื่องมือประเมินความสอดคล้องของบทความตามหลักการตลาดเชิงเนื้อหา (Content Marketing) โดยเน้นเนื้อหาภาษาไทยที่เกี่ยวข้องกับประเทศจีน ซึ่งครอบคลุมการวิเคราะห์ใน 4 ด้าน ได้แก่ ความทันต่อเวลา (Timeliness) จุดประสงค์ในการสื่อสาร (Intention) การสื่อสารอารมณ์ (Emotion) และรูปแบบการเขียนการเล่าเรื่องหรืองานแปล (Storytelling vs. Translation) จากผลการวิจัยพบว่า โมเดล WangchanBERTa มีประสิทธิภาพสูงสุดในการจำแนกเนื้อหาตามกระแส (Timely–Timeless) โดยมี Accuracy บนชุดทดสอบ 93.00% และ F1-score อยู่ที่ 92.00% ด้วยความสามารถของ tokenizer ที่ผ่านการฝึกอบรมด้วยข้อมูลภาษาไทยขนาดใหญ่ช่วยเพิ่มความแม่นยำในการวิเคราะห์เนื้อหาที่ซับซ้อนส่วนการจำแนกจุดประสงค์ของเนื้อหาตามกรอบ PIE (Persuade, Inform, Entertain) โมเดล PhayaThaiBERT ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า Micro F1-score เท่ากับ 88.74%, Macro F1-score เท่ากับ 84.14% และ Hamming Loss ต่ำที่สุดที่ 12.14% แม้พบปัญหาการเกิด Overfitting แต่ยังคงทำงานได้ดีในการทดสอบกับข้อมูลใหม่ นอกจากนี้ยังพบว่า การเพิ่มฟีเจอร์อารมณ์ ในโมเดล PhayaThaiBERT ไม่มีผลต่อประสิทธิภาพอย่างชัดเจน แต่กลับส่งผลเชิงบวกต่อโมเดล WangchanBERTa แสดงถึงความสัมพันธ์ระหว่างบริบทอารมณ์และจุดประสงค์เนื้อหา โดยสำหรับ การจำแนกอารมณ์ พบว่า โมเดล WangchanBERTa มีศักยภาพสูงสุด ในการประมวลผลอารมณ์ 8 กลุ่ม โดยมีค่า Micro F1-score เท่ากับ 78.35%, Macro F1-score เท่ากับ 55.74% และ Hamming Loss ต่ำที่สุดที่ 10.27% ซึ่งแสดงถึงความแม่นยำสูงในการจัดการกับข้อมูลที่ไม่สมดุลทางอารมณ์ ส่วน การจำแนกรูปแบบงานแปลและการเล่าเรื่อง พบว่า โมเดล Random Forest ที่ใช้เทคนิค Ensemble และ Feature Engineering มีประสิทธิภาพสูงสุด โดยมีค่า Accuracy เท่ากับ 92.55% F1-score สำหรับคลาส Translation เท่ากับ 92.77% และสำหรับ Storytelling เท่ากับ 92.31% เมื่อนำโมเดลทั้งหมดมาพัฒนาเป็นเว็บแอปพลิเคชัน โดยทดสอบการใช้งานจริงได้รับการประเมินจากผู้เชี่ยวชาญและผู้มีส่วนร่วมในการผลิตสื่อเนื้อหา พบว่าระบบมีความเพึงพอใจและความเป็นไปได้ในการนำไปใช้จริง ได้คะแนนเฉลี่ย 4.0 จาก 5.0 คะแนน ทั้งสองด้าน ผู้เชี่ยวชาญยังแนะนำให้ประยุกต์ใช้ตามลักษณะเฉพาะของงานเขียน 4 ประเภท ได้แก่ งานเขียนทั่วไป งานเขียนแบบแปล งานเขียนข่าวสาร และงานเขียนที่ต้องการความทันต่อเหตุการณ์แบบเรียลไทม์ ทั้งนี้ งานวิจัยใน

อนาคตควรขยายขนาดและความหลากหลายของชุดข้อมูล รวมถึงการสร้างคลังข้อมูลเฉพาะทางเพื่อเพิ่มประสิทธิภาพในการจำแนกและวิเคราะห์เนื้อหาต่อไป

620730008 : Major INFORMATION TECHNOLOGY AND DIGITAL INNOVATION
Keyword : Content Marketing, Machine Learning, Deep Learning, Content Style, Chinese Content

MR. Pagon GATCHALEE : Developing Framework for Consistency Evaluation of Article Writing Styleusing Machine Learning Techniques Thesis advisor : Sajjaporn Waijanya, Ph.D.

This dissertation aims to develop and present a framework for evaluating the consistency of articles based on content marketing principles, focusing on Thai-language content related to China. The analysis covers four main aspects: Timeliness, Intention, Emotion, and Storytelling vs. Translation style. According to the results, the WangchanBERTa model best-classified content as Timely or Timeless, with 93.00% accuracy and an F1-score of 92.00%. The tokenizer, trained on large-scale Thai data, helped improve the model's ability to analyze complex content. For classifying content intentions based on the PIE framework (Persuade, Inform, Entertain), PhayaThaiBERT gave the best results, with a Micro F1-score of 88.74%, Macro F1-score of 84.14%, and the lowest Hamming Loss at 12.14%. Although there was minor overfitting, the model still worked well when tested on new data. It was also found that adding emotional context features did not clearly impact PhayaThaiBERT's performance but did improve WangchanBERTa, which suggests a link between emotional context and content intention. In emotion classification, WangchanBERTa showed the highest capability in handling 8 emotion categories, with a Micro F1 Score of 78.35%, a Macro F1 Score of 55.74%, and the lowest Hamming Loss at 10.27%, indicating strong performance with imbalanced data. For Translation classification, the Random Forest model, an ensemble learning model using feature engineering, performed best, with 92.55% accuracy, 92.77% F1-score for Translation, and 92.31% for Storytelling. All models were integrated into a web application. When tested in actual usage, feedback from experts and content creators showed that the system was usable and practical, scoring an average of 4.0 out of 5.0 in both areas. Experts also recommended applying the tool to four specific types of writing: general writing, translated content, news articles, and real-time content. Future research should expand the dataset, include more variety, and build domain-specific corpora

to further improve classification and content analysis performance.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จไปได้ด้วยดีนั้น ต้องขอกราบขอบพระคุณอาจารย์ ดร.สัจจา ภรณ์ ไวจรรยา อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และ ผู้ช่วยศาสตราจารย์ ดร. ณัฐโชติ พรหมฤทธิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งได้กรุณาให้ความรู้ คำแนะนำ และข้อเสนอแนะต่างๆ ด้วยความเมตตาอย่างยิ่ง ไม่เพียงเฉพาะในด้านวิชาการสำหรับการเรียนระดับปริญญาเอกและการจัดทำวิทยานิพนธ์เท่านั้น แต่ยังเป็นประโยชน์ต่อการดำเนินชีวิตในด้านอื่นอีกด้วย

นอกจากนี้ ขอขอบพระคุณประธานกรรมการและกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่ได้ให้คำแนะนำเพิ่มเติมและข้อเสนอแนะที่เป็นประโยชน์ในการพัฒนาวิทยานิพนธ์ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณ มหาวิทยาลัยเชียงใหม่ที่สนับสนุนให้โอกาสในการศึกษาต่อระดับปริญญาเอกที่มหาวิทยาลัยศิลปากร รวมทั้งขอขอบคุณผู้เชี่ยวชาญและผู้มีส่วนร่วมทุกท่านที่ให้ข้อมูลเชิงลึกผ่านการสัมภาษณ์เชิงลึก ซึ่งเป็นประโยชน์อย่างยิ่งสำหรับการพัฒนาโมเดลและการวิเคราะห์เนื้อหา โดยเฉพาะผู้บริหารสื่อ นักวางแผนสื่อ นักเขียน และบล็อกเกอร์ ที่ได้แบ่งปันข้อมูลและประสบการณ์อย่างเต็มที่

สุดท้ายนี้ ขอกราบขอบพระคุณครอบครัวกัทชลี ครอบครัวสุวรรณสิทธิ์ญาติมิตร และเพื่อนร่วมงานทุกท่านที่ให้การสนับสนุนทั้งด้านกำลังใจ คำแนะนำ และความช่วยเหลือในทุกๆ ด้านตลอดระยะเวลาการศึกษาวิจัย จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ภากร กัทชลี

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... ฉ

กิตติกรรมประกาศ ..... ซ

สารบัญ ..... ฌ

สารบัญตาราง ..... ฎ

สารบัญรูปภาพ ..... ฐ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5

2.1 เนื้อหาตามกระแสและเนื้อหาสดใหม่เสมอ (Timely and Timely Content) 5

2.2 จุดประสงค์ในการสื่อความของเนื้อหา (Text Intention) 6

2.3 การตรวจจับและวิเคราะห์อารมณ์ในข้อความ (Emotion Detection and Recognition in Text) 9

2.4 สไตล์การเขียนของบทความ 12

2.5 การเรียนรู้เชิงลึก (Deep Learning) 15

2.6 โครงข่ายประสาทเทียม (Artificial Neural Network) 15

2.7 การแบ่งกลุ่มข้อความ (Text Classification) 16

2.8 Convolutional Network (CNN) 17

2.9 การประมวลผลข้อความด้วย BERT 21

2.10 การใช้ Ensemble Learning ในการจำแนกประเภทข้อความ 23

2.11 อัลกอริทึมการวัดความคล้าย (Similarity Algorithm) 24

2.12 งานวิจัยที่เกี่ยวข้อง 26
บทที่ 3 ขั้นตอนและวิธีดำเนินการ 34
3.1 ภาพรวมขั้นตอนการดำเนินการวิจัย 34
3.2 การเตรียมข้อมูลและสกัดองค์ความรู้จากผู้เชี่ยวชาญ 35
3.3 การพัฒนาและทดลองหาโมเดลที่มีประสิทธิภาพดีที่สุด 40
3.4 การพัฒนาระบบ Web application 60
3.5 การทดสอบระบบโดยผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับข้อมูล 73
บทที่ 4 ผลการดำเนินงานวิจัย 77
4.1 ผลการทดลองที่ 1 โมเดลการจำแนกข้อความ Timely และ Timeless 77
4.2 ผลการทดลองที่ 2 โมเดลจำแนกเจตนาการสื่อความของเนื้อหา 78
4.3 ผลการทดลองที่ 3 โมเดลจำแนกเจตนาการสื่อความของเนื้อหา โดยทดสอบความสมดุลของข้อมูล 83
4.4 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์ 101
4.5 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง 108
4.6 สรุปโมเดลที่ดีที่สุดของแต่ละรูปแบบการประเมินความสอดคล้อง และผลการพัฒนา API 111
4.7 ผลทดสอบการใช้งานจริง Web application 113
4.8 ประเมินความพึงพอใจและความเป็นไปได้ในการใช้งานจริง 122
บทที่ 5 สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ 124
รายการอ้างอิง 128
ประวัติผู้เขียน 138

สารบัญตาราง

หน้า

ตารางที่ 2.1 รายละเอียดของจุดประสงค์การสื่อความเนื้อหาตามกรอบ PIE Framework 6

ตารางที่ 2.2 ตัวอย่างวิเคราะห์จุดประสงค์การสื่อความเนื้อหาตามกรอบ PIE Framework 7

ตารางที่ 2.3 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม CNN 28

ตารางที่ 2.4 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม Ensemble learning 29

ตารางที่ 2.5 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม Transformer 30

ตารางที่ 2.6 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับการประเมินความสอดคล้องเนื้อหา 32

ตารางที่ 3.1 จำนวนบทความในชุดข้อมูลแบ่งตามจุดประสงค์การสื่อความ 48

ตารางที่ 3.2 จำนวนบทความในชุดข้อมูลแบ่งตามจุดประสงค์การสื่อความหลังปรับเป็นการจำแนกแบบหลายป้ายกำกับ 48

ตารางที่ 3.3 จำนวนป้ายกำกับในแต่ละคลาสอารมณ์ ในรูปแบบหลายป้ายกำกับ 49

ตารางที่ 3.4 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมใกล้เคียงกัน (Dataset S) 50

ตารางที่ 3.5 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมเท่ากัน (Dataset A) 51

ตารางที่ 3.6 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมเท่ากัน (Dataset B) 52

ตารางที่ 3.7 ตารางสรุปการทดลองย่อยการทดลองที่ 3 โมเดลจำแนก Intention ทดสอบความสมดุลของข้อมูล 52

ตารางที่ 3.8 รูปแบบการจำแนกเนื้อหาและประเภทโมเดล 57

ตารางที่ 3.9 แสดงรายละเอียดคีย์และข้อมูลใน JSON 66

ตารางที่ 4.1 ตารางแสดงผลการทดลองที่ 1 โมเดลการจำแนกข้อความ Timely และ Timeless ... 78

ตารางที่ 4.2 ผลการทดลองที่ 2 โมเดลจำแนกการสื่อความของเนื้อหาแบบไม่ใช้ฟีเจอร์อารมณ์ 78

ตารางที่ 4.3 ผลการทดลองที่ 2 โมเดลจำแนกการสื่อความของเนื้อหาแบบใช้ฟีเจอร์อารมณ์ 82
ตารางที่ 4.4 ผลการทดลองย่อยที่1 ของการทดลองที่ 3..... 84
ตารางที่ 4.5 ผลการทดลองย่อยที่ 2 ของการทดลองที่ 3 86
ตารางที่ 4.6 ผลการทดลองย่อยที่ 3 ของการทดลองที่ 3 88
ตารางที่ 4.7 ผลการทดลองย่อยที่ 4 ของการทดลองที่ 3 90
ตารางที่ 4.8 ผลการทดลองย่อยที่ 5 ของการทดลองที่ 3 92
ตารางที่ 4.9 ผลการทดลองย่อยที่ 6 ของการทดลองที่ 3..... 94
ตารางที่ 4.10 ผลการทดลองย่อยที่ 7 ของการทดลองที่ 3 95
ตารางที่ 4.11 ผลการทดลองย่อยที่ 8 ของการทดลองที่ 3 97
ตารางที่ 4.12 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์แบบ 8 อารมณ์..... 102
ตารางที่ 4.13 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์แบบ 4 อารมณ์..... 105
ตารางที่ 4.14 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง แบบไม่ใช้ฟีเจอร์
..... 109
ตารางที่ 4.15 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง แบบใช้ฟีเจอร์... 111

มหาวิทยาลัยศิลปากร

สารบัญรูปภาพ

หน้า

ภาพที่ 2.1 Plutchik’s Wheel of Emotions 10

ภาพที่ 2.2 ตัวอย่างเนื้อหาบนเฟซบุ๊กเพจอ้ายจง 13

ภาพที่ 2.3 ตัวอย่างของโครงข่ายประสาทเทียม 16

ภาพที่ 2.4 ตัวอย่างโครงสร้างการทำงานการแบ่งกลุ่มข้อความ 17

ภาพที่ 2.5 ภาพรวมการวิเคราะห์ข้อมูลด้วย CNN 18

ภาพที่ 2.6 ตัวอย่างการทำ Padding ของ CNN 18

ภาพที่ 2.7 การทำ Word Embedding 19

ภาพที่ 2.8 โครงสร้างโมเดล CNN สำหรับตรวจสอบโครงสร้างกลอนแปดสุภาพภาษาไทย 20

ภาพที่ 2.9 แสดงผลลัพธ์ Accuracy การจัดกลุ่มและวิเคราะห์ Sentiment กลอนสุภาพภาษาไทย ด้วย CNN Model เทียบกับ SVM และ Naïve Bayes 20

ภาพที่ 2.10 แสดงผลลัพธ์ Accuracy ทำนายว่ากลอนที่แต่งขึ้นเป็นของใคร 21

ภาพที่ 2.11 โครงสร้างของ BERT (BERT Structure) 21

ภาพที่ 2.12 ตัวอย่างการประมวลผลข้อความด้วย Attention Mechanism หัวใจหลักของ BERT 22

ภาพที่ 3.1 ภาพรวมขั้นตอนการดำเนินการวิจัย 34

ภาพที่ 3.2 โมเดลจำแนกเนื้อหาทั้งสี่รูปแบบ 41

ภาพที่ 3.3 สถาปัตยกรรม Convolutional Neural Network Con1D 42

ภาพที่ 3.4 การพัฒนาโมเดลจำแนกเนื้อหาการแปลและเล่าเรื่อง 54

ภาพที่ 3.5 ภาพรวมของการพัฒนาระบบ Web application และโมเดลวิเคราะห์เนื้อหา 61

ภาพที่ 3.6 ส่วนนำเข้าข้อมูล 62

ภาพที่ 3.7 JavaScript กำหนดค่าตัวแปร presets 63

ภาพที่ 3.8 รูปแบบการส่งข้อมูลแบบ JSON 65

ภาพที่ 3.9 ตัวอย่าง JSON ที่ส่งออกโดย API 65
ภาพที่ 3.10 โครงสร้างการทำงานระหว่าง Web application และ API 67
ภาพที่ 3.11 JavaScript ข้อมูลจากส่วนนำเข้าข้อมูลไปยัง API server ในรูปแบบ JSON 68
ภาพที่ 3.12 Python code สร้าง endpoint /api/compare 68
ภาพที่ 3.13 Python code แสดงรายละเอียดเรียกใช้งานแต่ละโมเดลที่ดีที่สุด ใน API 69
ภาพที่ 3.14 Python code แสดงรายละเอียดขั้นตอน preprocessing ใน API 70
ภาพที่ 3.15 กระบวนการทำงานแสดงผลลัพธ์ ในภาพรวมตั้งแต่รับข้อมูล 71
ภาพที่ 3.16 ส่วนแสดงผลลัพธ์ (Output) 71
ภาพที่ 3.17 ส่วนแสดงผลลัพธ์กรณีเกิดข้อผิดพลาด (Output with Error Message) 73
ภาพที่ 3.18 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม 75
ภาพที่ 4.1 คลาสที่ทำนายผิดโมเดล PhayaThaiBERT ร่วมกับฟีเจอร์อารมณ์ 81
ภาพที่ 4.2 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ไม่ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Dataset A 99
ภาพที่ 4.3 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Dataset A 100
ภาพที่ 4.4 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Balanced แบบจำนวนป้ายกำกับใกล้เคียงกัน และสัดส่วน Train:Test 80:20 101
ภาพที่ 4.5 ตัวอย่างที่ 1 ทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีน 114
ภาพที่ 4.6 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่1 114
ภาพที่ 4.7 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่2 115
ภาพที่ 4.8 ตัวอย่างที่ 2 ทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีน 117
ภาพที่ 4.9 แสดงผลการทดสอบบทความตัวอย่างที่2 ส่วนที่1 118
ภาพที่ 4.10 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่2 119

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การตลาดด้วยเนื้อหา (Content Marketing) เป็นรูปแบบหนึ่งของเครื่องมือการตลาดยุคใหม่ โดยขับเคลื่อนด้วยเนื้อหา ซึ่งนักการตลาดยุคนี้เชื่อว่า เนื้อหาที่ดีมีค่าดั่งพระราชา ตามหลักกลยุทธ์ Content is King โดยหนึ่งในรูปแบบของเนื้อหาที่ได้รับความนิยมและทุกคนคุ้นเคยมากที่สุด ได้แก่ บทความ ทั้งลักษณะของบทความยาวและบทความสั้น รวมถึงลักษณะของข้อความไม่กี่บรรทัดที่เผยแพร่บนแพลตฟอร์มสังคมออนไลน์ อาทิ Facebook Twitter Instagram เป็นต้น (Diamond, 2016) จากข้างต้น แต่ละองค์กรทั้งภาครัฐ ภาคเอกชน และขนาดเล็กจนถึงรูปแบบองค์กรขนาดใหญ่ จึงให้ความสำคัญต่อการสร้างเนื้อหาบนสื่อช่องทางบนโลกออนไลน์ของตนเอง มีการจ้างพนักงานและสร้างทีมงานสำหรับจัดทำเนื้อหาโดยเฉพาะ แต่ปัญหาที่พบบ่อยที่สุด คือ เนื้อหาที่จัดทำขึ้นมีรูปแบบไม่สอดคล้องกับความต้องการหรือลักษณะบริบทขององค์กร เป็นผลโดยตรงทำให้การตลาดด้วยเนื้อหาไม่ประสบความสำเร็จตามเป้าหมาย แม้มีการวางแผนจัดทำเนื้อหา (Communication plan) และ การจัดทำเนื้อหาต้นแบบ เพื่อเป็นตัวอย่างสำหรับการจัดทำเนื้อหาให้เป็นไปตามเป้าประสงค์ โดยพบว่ามีสาเหตุจาก ขาดการวิเคราะห์และตรวจสอบว่าตรงตามแบบแผนและต้นแบบ รวมถึงเนื้อหาที่เคยมีการจัดทำมาก่อนหน้านั้นขององค์กรหรือไม่ (Müller & Christandl, 2019)

จุดเริ่มต้นของงานวิทยานิพนธ์ฉบับนี้ เกิดขึ้นจากประสบการณ์ตรงของผู้วิจัยในฐานะที่ปรึกษาด้านการตลาดด้วยเนื้อหา และเจ้าของเพจFacebook(Facebook Page) อ้ายจง ซึ่งมีผู้ติดตามจำนวน 280,000 ราย โดยผู้วิจัยเล็งเห็นถึงปัญหาดังกล่าวอย่างชัดเจน และพบว่าเนื้อหาภาษาไทยที่เกี่ยวข้องกับประเทศจีนมีแนวโน้มการเติบโตอย่างต่อเนื่องในประเทศไทย โดยเฉพาะผ่านช่องทางสื่อออนไลน์ ซึ่งมีอิทธิพลอย่างมากต่อการรับรู้เกี่ยวกับประเทศจีนของประชาชนชาวไทย ทว่าในปัจจุบัน ยังไม่มีงานวิจัยที่ครอบคลุมการวิเคราะห์เนื้อหาเกี่ยวกับประเทศจีนในบริบทของภาษาไทยอย่างเป็นระบบ ขณะที่เนื้อหาเกี่ยวกับประเทศจีนถือเป็นเนื้อหาเฉพาะด้าน จึงมีความน่าสนใจในการนำมาใช้เป็นกรณีศึกษา เพื่อจัดทำเฟรมเวิร์กสำหรับการวิเคราะห์การเขียนบทความให้สอดคล้องกับต้นแบบหรือเป้าหมายของสื่อหรือผู้ผลิตเนื้อหาในรูปแบบของบทความตัวอักษร

นอกจากนี้ การเติบโตของเนื้อหาเกี่ยวกับประเทศจีนในบริบทของภาษาไทย โดยเฉพาะในรูปแบบของ งานแปลจากภาษาจีน มีจำนวนเพิ่มขึ้นอย่างต่อเนื่อง (ปรียากร บุญธรรม, 2023) ซึ่งสอดคล้องกับงานวิจัยหลายฉบับที่ระบุว่า คนไทยมีแนวโน้มเรียนภาษาจีนเพิ่มขึ้น อย่างมีนัยสำคัญ

โดยมีปัจจัยสนับสนุนสำคัญมาจาก อิทธิพลของสื่อออนไลน์ ซึ่งทำให้ประชาชนสามารถเข้าถึงข้อมูล ข่าวสาร วัฒนธรรม และประเด็นต่าง ๆ ที่เกี่ยวข้องกับประเทศจีนได้ง่ายและหลากหลายมากยิ่งขึ้น (พัชชาชาติพาณิชย์ & ยุวรินธรไชยโชติช่วง, 2024) สื่อออนไลน์เหล่านี้ไม่เพียงแต่มีบทบาทในการถ่ายทอดเนื้อหาเท่านั้น แต่ยังทำให้เกิดการแปล และ การปรับบริบท ของเนื้อหาจากจีนสู่ไทยอย่างแพร่หลาย ส่งผลให้ปริมาณและประเภทของเนื้อหาภาษาไทยเกี่ยวกับจีนเพิ่มขึ้นในหลากหลายมิติ ทั้ง ข่าวสาร บทความเชิงวิเคราะห์ บทแปล บทวิจารณ์ หรือแม้แต่บทความที่มีลักษณะบันเทิงหรือความคิดเห็นส่วนตัว ดังนั้น การสร้างเครื่องมือหรือเฟรมเวิร์กที่สามารถวิเคราะห์และประเมินความสอดคล้องของบทความภาษาไทยที่เกี่ยวข้องกับประเทศจีน จึงมีความสำคัญอย่างยิ่ง ทั้งในมิติของคุณภาพเนื้อหา มาตรฐานการสื่อสาร และการทำความเข้าใจบริบทของการนำเสนอสารเกี่ยวกับจีนมายังคนไทย โดยเฉพาะจากสื่อจีนที่ปัจจุบันเข้ามามีอิทธิพลและสร้างการรับรู้เชิงบวกในไทยและเอเชียตะวันออกเฉียงใต้มากขึ้น (Chen, 2020)

สำหรับเฟรมเวิร์กในวิทยานิพนธ์นี้ สามารถวิเคราะห์และประเมินรูปแบบการเขียนบทความได้ทั้งหมด สี่รูปแบบที่เกี่ยวข้องกับการตลาดด้วยเนื้อหา ได้แก่

1) เนื้อหาตามกระแส และ เนื้อหาสดใหม่เสมอ (Timely and Evergreen content)

2) จุดประสงค์ในการสื่อความของเนื้อหา (Text Intention)

3) การสื่อสารอารมณ์ผ่านเนื้อหา (Emotional Content)

4) เนื้อหาการแปลและการเล่าเรื่อง (Translation and Storytelling)

จากการค้นคว้าศึกษางานวิจัยที่เกี่ยวข้องพบว่า ยังไม่เคยมีการวิจัยและพัฒนาเฟรมเวิร์กสำหรับประเมินการเขียนบทความแบบประกอบไปด้วยทั้ง 4 รูปแบบ มีแต่เพียงประกอบไปด้วยบางส่วนเท่านั้น

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอแนวทางการพัฒนาขั้นตอนวิธีการสร้างเครื่องมือประเมินการเขียนบทความที่สอดคล้องกับเนื้อหาต้นแบบ สำหรับนำไปใช้สนับสนุนการดำเนินการดำเนินงานด้านการตลาด

2. เพื่อวิจัยและพัฒนาโมเดลวิเคราะห์และประเมินเนื้อหาได้ว่าเป็นเนื้อหาตามกระแสหรือเนื้อหาสดใหม่เสมอ

3. เพื่อวิจัยและพัฒนาโมเดลวิเคราะห์และประเมินจุดประสงค์ในการสื่อความของเนื้อหา

4. เพื่อวิจัยและพัฒนาโมเดลวิเคราะห์และประเมินการสื่อสารอารมณ์ผ่านเนื้อหา

5. เพื่อวิจัยและพัฒนาโมเดลที่สามารถวิเคราะห์และประเมินได้ว่าเนื้อหานั้นเป็นบทความประเภทการแปลหรือการเล่าเรื่อง

1.3 ขอบเขตงานวิจัย

1. สร้างโมเดลการเรียนรู้ของเครื่องและโมเดลการเรียนรู้เชิงลึก เพื่อวิเคราะห์เนื้อหาโดยใช้คลังข้อมูลภาษาไทยจากเพจเฟซบุ๊กที่เกี่ยวข้องกับจีน รวมทั้งสิ้น 10 เพจ โดยแบ่งเป็น (1) เพจ Facebook ประเภทองค์กรสื่อ ได้แก่ China Xinhua News (ภาคภาษาไทย) China Report ASEAN – Thailand Jeenthainews China Face และ People's Daily Online Thai (2) เพจ Facebook ประเภทงานเขียนเชิงเล่าเรื่อง (Storytelling) ได้แก่ เพจอ้ายจง เพจ China Story เพจ Nihaosawadee เพจไทยคำจีนคำ และเพจ DomJeen ด้อมจีน ติ่งจีน ซีรี่ย์จีน ศิลปินจีน รักทุกอย่างที่เป็นจีน

2. ดำเนินการเก็บข้อมูลและสกัดองค์ความรู้จากผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับชุดข้อมูลเนื้อหาจีน เพื่อใช้ในการพัฒนาโมเดลและเฟรมเวิร์ก ตลอดจนประเมินความพึงพอใจและความเป็นไปได้ในการนำระบบไปใช้งานจริง โดยผู้ให้ข้อมูลประกอบด้วยจำนวนทั้งสิ้น 4 ท่าน ได้แก่

1) อาจารย์กอบกิจ ประดิษฐผลพานิช ผู้ดูแลงานแปลภาษาไทยของสำนักข่าว Xinhua และเจ้าของเพจ Jeenthainews Head of Corporate Communications และคณบดีกลุ่มวิชาการ (อดีตคณบดีคณะนิเทศศาสตร์) มหาวิทยาลัยธุรกิจบัณฑิตย์

2) คุณกิตติธัชช์ กิตติ์นัฏฐากูร ผู้เชี่ยวชาญไทยประจำสำนักข่าว China Radio International (CRI) ผู้ดูแลเนื้อหาเพจ China Face

3) ดร.เพ็ญพิสุทธิ์ สีกาแก้ว ผู้อำนวยการหลักสูตรศิลปศาสตรมหาบัณฑิต สาขาการสอนภาษาจีนสำหรับชาวต่างชาติ วิทยาลัยนานาชาติจีน มหาวิทยาลัยธุรกิจบัณฑิตย์ และผู้เชี่ยวชาญด้านวัฒนธรรมและการแปลจีนไทย (ปัจจุบันกำลังศึกษาต่อปริญญาโทใบที่สอง สาขาภาษาไทย)

4) คุณธัญญรัตน์ เลิศหิรัญวงศ์ นักแปลซีรีส์จีน-ไทย (แบบตอนสั้น), นักเขียนและนักแปลนิยายออนไลน์จีน-ไทย รวมถึงนักการตลาดออนไลน์ ซึ่งรับผิดชอบงานเขียนเนื้อหาภาษาไทยสำหรับแอปพลิเคชันประเภท Dating app ของจีน เพื่อดำเนินการตลาดในประเทศไทย

3. พัฒนาเฟรมเวิร์กเว็บแอปพลิเคชันต้นแบบ สำหรับนำเข้าเนื้อหาทางเว็บแอปพลิเคชันและประเมินเนื้อหาโดยใช้โมเดลการวิเคราะห์เนื้อหาทั้ง 4 รูปแบบพร้อมทั้งสามารถประเมินความคล้ายของแต่ละรูปแบบ และความสอดคล้องของสองบทความในภาพรวมได้

1.4 ขั้นตอนการดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้อง

2. รวบรวมข้อมูล เพื่อสร้างคลังข้อมูลของเนื้อหาตามจุดประสงค์การประเมินเนื้อหาทั้งสี่รูปแบบ โดยดำเนินการตามหลักกระบวนการจัดเตรียมข้อมูล (Preprocessing)

3. พัฒนาโมเดลสำหรับประเมินเนื้อหาทั้งสี่รูปแบบ

4. วัดผลและ ปรับพารามิเตอร์ต่าง ๆ ของแต่ละโมเดล เพื่อพัฒนาปรับปรุงให้ผลดีขึ้น โดยเทียบกับโมเดลที่เคยมีการนำเสนอในงานวิจัยอื่น ๆ (State-of-the-art Model)

5. พัฒนาเฟรมเวิร์กเว็บแอปพลิเคชันเพื่อเชื่อมกับโมเดลที่สร้างขึ้น

6. ทดสอบและวัดผลเฟรมเวิร์กเว็บแอปพลิเคชันที่พัฒนาขึ้น ประเมินโดยผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับชุดข้อมูลเนื้อหาเกี่ยวกับจีน ตามระบุในขอบเขตงานวิจัย

7. สรุปผลการวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เฟรมเวิร์กสามารถวิเคราะห์และประเมินบทความที่สอดคล้องกับเนื้อหาต้นแบบ เพื่อเพิ่มประสิทธิภาพการใช้เนื้อหาเพื่อผลทางการตลาดได้

2. สามารถวิเคราะห์และประเมินเนื้อหาตามจุดประสงค์รูปแบบต่าง ๆ ได้

3. สามารถเป็นแนวทางสำหรับการวิจัยและพัฒนาทางด้านการวิเคราะห์เนื้อหาภาษาไทยที่มีรูปแบบเฉพาะเจาะจง เช่น เนื้อหาเกี่ยวกับจีน เป็นต้น

1.6 สมมติฐานการวิจัย

1. โมเดลวิเคราะห์และประเมินเนื้อหา สามารถจำแนกข้อมูลทดสอบออกเป็นเนื้อหาตามกระแสหรือเนื้อหาสดใหม่เสมอ โดยมีความแม่นยำไม่น้อยกว่า 80%

2. โมเดลวิเคราะห์และประเมินเนื้อหา สามารถจำแนกข้อมูลทดสอบสำหรับการจำแนก Intention ของเนื้อหา โดยมีตัวชี้วัดประสิทธิภาพโมเดลสำหรับการจำแนกประเภทแบบหลายป้ายกำกับ (Multi-label Classification) ไม่น้อยกว่า 80%

3. โมเดลวิเคราะห์และประเมินเนื้อหา สามารถจำแนกข้อมูลทดสอบสำหรับการจำแนกสื่อสารอารมณ์ผ่านเนื้อหา โดยมีตัวชี้วัดประสิทธิภาพโมเดลสำหรับการจำแนกประเภทแบบหลายป้ายกำกับ ไม่น้อยกว่า 80%

4. โมเดลวิเคราะห์และประเมินเนื้อหา สามารถจำแนกข้อมูลทดสอบออกเป็นเนื้อหารูปแบบการเขียนแบบแปลและStorytelling โดยมีความแม่นยำไม่น้อยกว่า 80%

5. ได้รับการประเมินความสอดคล้องกับสไตล์ต้นแบบ โดยผู้ทดสอบจริง โดยมีความพึงพอใจไม่น้อยกว่า 4 จาก 5 คะแนนเต็ม ตามมาตรวัดแบบ Likert (Likert scale)

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทที่ 2 วิทยานิพนธ์นี้จะนำเสนอรายละเอียดทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพัฒนาเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง ซึ่งเป็นการประยุกต์ความรู้ในหลายศาสตร์ ทั้งทางด้านการตลาด สังคมศาสตร์ และเทคโนโลยีสารสนเทศ โดยมีรายละเอียดดังต่อไปนี้

2.1 เนื้อหาตามกระแสและเนื้อหาสดใหม่เสมอ (Timely and Timely Content)

เนื้อหาที่เกาะกระแส (Timely Content) และ เนื้อหาที่สดใหม่อยู่เสมอ (Evergreen Content) ถือเป็นหนึ่งในแนวทางการจำแนกประเภทเนื้อหาตามหลักการตลาดด้วยเนื้อหา (Content Marketing) โดยในทางการตลาดสมัยใหม่ มักเรียกเนื้อหาประเภท Evergreen ว่า Timeless Content เพื่อเน้นลักษณะที่ไม่ยึดโยงกับช่วงเวลาและสามารถนำกลับมาใช้ซ้ำได้ในระยะยาว

ในวิทยานิพนธ์นี้เลือกใช้คำว่า Timeless เพื่ออ้างถึงเนื้อหาประเภทEvergreen ทั้งหมด เพื่อให้เกิดความชัดเจนและสอดคล้องกับแนวคิดด้านการจำแนกเนื้อหาตามบริบทเวลา ซึ่งสามารถแบ่งความแตกต่างระหว่าง Timely และ Timeless ได้ดังนี้

1) เนื้อหาตามกระแส (Timely content)

เป็นเนื้อหาที่เขียนขึ้นตามเหตุการณ์ที่เกิดขึ้นและได้รับความสนใจจากคนกลุ่มหนึ่งภายใต้ความสนใจบางอย่างทั้งในระดับท้องถิ่น ระดับชาติ และระดับโลก โดยเนื้อหาตามกระแสที่เราพบได้บ่อยที่สุด คือ เนื้อหาประเภทข่าว และเนื้อหาไวรัลที่กำลังเป็นกระแสบนโลกออนไลน์ (Frith, 2020)

ตัวอย่างเช่น เนื้อหารายงานอัปเดตจำนวนผู้ติดเชื้อโควิดรายใหม่ในจีนประจำวันที่ 1 เมษายน พ.ศ. 2564 เมื่อผ่านไปหนึ่งสัปดาห์ ผู้อ่านอาจไม่สนใจแล้ว เพราะมีตัวเลขใหม่เกิดขึ้นในช่วงวันที่ผ่านไป

2) เนื้อหาสดใหม่เสมอ (Timely Content)

เป็นเนื้อหาที่ไม่มีเงื่อนไขของเวลามาเกี่ยวข้อง กล่าวคือไม่ว่าเวลาผ่านไปนานแค่ไหน เนื้อหาเหล่านั้นจะเป็นที่สนใจของคนกลุ่มใดกลุ่มหนึ่งอยู่เสมอ โดยส่วนใหญ่เป็นเนื้อหาที่จัดทำขึ้นเกี่ยวกับความสนใจพื้นฐานของคนกลุ่มนั้น (Goobich, 2020)

ตัวอย่างเช่น เนื้อหาอธิบายการเข้าชมสถานที่ท่องเที่ยวพระราชวังต้องห้ามในกรุงปักกิ่ง หลังจากเขียนเนื้อหาไปแล้วมากกว่าสามปี คนก็ยังสนใจอ่าน เพราะต้องการไปเที่ยวสถานที่ดังกล่าว

2.2 จุดประสงค์ในการสื่อความของเนื้อหา (Text Intention)

การสื่อความของเนื้อหาเป็นองค์ประกอบสำคัญในการวางแผนและพัฒนาเนื้อหาภายใต้หลักการตลาดด้วยเนื้อหา จุดประสงค์ในการสื่อความของเนื้อหามักถูกกำหนดโดยเจ้าของเนื้อหาหรือหน่วยงานองค์กรที่จัดทำเนื้อหา ซึ่งอาจแบ่งประเภทได้หลากหลาย เช่น ตามหมวดหมู่เนื้อหา (ข่าวทั่วไป ข่าวท่องเที่ยว ข่าวกีฬา ข่าวการเมือง ข่าวเศรษฐกิจ) หรือแบ่งตามลักษณะการนำเสนอ (บทความวิเคราะห์ บทความรีวิว บทความข่าว บทความแปล) ความหลากหลายนี้ส่งผลให้การวิเคราะห์และแบ่งประเภทจุดประสงค์ของเนื้อหามีความซับซ้อน เนื่องจากอาจขึ้นอยู่กับเจตนาแฝงของผู้สร้างเนื้อหา

ดังนั้น วิทยานิพนธ์นี้จึงเสนอแนวทางการกำหนดจุดประสงค์ในการสื่อความของเนื้อหาโดยพิจารณาเจตนาของผู้สร้างเนื้อหาโดยตรง ผ่านกรอบแนวคิด Persuade, Inform และ Entertain (PIE) Framework ซึ่งเป็นแนวทางที่ใช้กันอย่างแพร่หลายในด้านการสอนภาษาและการสื่อสารมวลชน (Hadi et al., 2024) โดยกรอบแนวคิดนี้สามารถประยุกต์ได้ทั้งเนื้อหาที่เป็นข้อความ รวมถึงสามารถขยายไปยังเนื้อหาในรูปแบบสื่อเสียงและสื่อภาพ (audio and visual media) เช่น สื่อภาพและเสียง ภาพนิ่ง และสื่อวิดีโอ (Grall & Finn, 2022)

รายละเอียดของแต่ละจุดประสงค์ สามารถสรุปได้ดังตารางที่ 2.1 ดังรายละเอียดต่อไปนี้

ตารางที่ 0.1 รายละเอียดของจุดประสงค์การสื่อความเนื้อหาตามกรอบ PIE Framework

Purpose	Description	Classification Points
โน้มน้าว (Persuade)	ผู้เขียนมุ่งหวังที่จะโน้มน้าวให้ผู้อ่านมีความคิดเห็น ความเชื่อ หรือกระทำบางอย่าง โดยใช้หลักฐานเชิงตรรกะ อารมณ์ และความคิดสร้างสรรค์	1. ข้อความพยายามโน้มน้าวให้ผู้อ่านมีความเชื่อหรือจุดยืนหรือไม่? 2. มีการใช้เทคนิคเช่น การเน้นย้ำ (repetition) หรือการกล่าวเกินจริง (hyperbole) หรือไม่? 3. มีการกระตุ้นให้ผู้อ่านลงมือทำหรือไม่ (Call to Action)?
ให้ข้อมูล (Inform)	ผู้เขียนต้องการให้ความรู้แก่ผู้อ่านโดยนำเสนอข้อเท็จจริงและคำอธิบายอย่างเป็นกลาง โดยไม่มุ่งหวังให้เกิดการโน้มน้าว	1. เนื้อหานำเสนอข้อเท็จจริงอย่างเป็นกลางและมีความตรงไปตรงมาหรือไม่? 2. ข้อเท็จจริงถูกใช้เพื่อการให้

Purpose	Description	Classification Points
	หรือจูงใจ	ความรู้ ไม่ใช่เพื่อการโน้มน้าวหรือไม่? 3. ไม่มีการกระตุ้นให้ผู้อ่านลงมือทำใช่หรือไม่?
ให้ความบันเทิง (Entertain)	ผู้เขียนต้องการสร้างความเพลิดเพลินหรือดึงดูดความสนใจของผู้อ่าน ผ่านตัวละคร โครงเรื่อง หรือบทสนทนา โดยกลยุทธ์ที่ใช้ มักเป็นการเล่าเรื่อง (storytelling)	1. ข้อความมีเป้าหมายเพื่อให้ความสนุกสนานหรือกระตุ้นอารมณ์ของผู้อ่านหรือไม่? 2. มีการใช้เทคนิคเช่น อารมณ์ขัน (humor) หรือการสร้างจุดพีคของเรื่องราว(cliffhangers) หรือไม่? 3. เป้าหมายหลักของเนื้อหาคือเพื่อสร้างความเพลิดเพลินใช่หรือไม่?

ตัวอย่างของการวิเคราะห์จุดประสงค์ในการสื่อความของข้อความ ตามหลักการเฟรมเวิร์ค PIE ซึ่งนำมาใช้กับบทความยาว ในวิทยานิพนธ์นี้ นำเสนอการวิเคราะห์โดยแบ่งเป็น ย่อหน้า (Paragraph) ดังรายละเอียดในตารางที่ 2.2

ตารางที่ 0.2 ตัวอย่างวิเคราะห์จุดประสงค์การสื่อความเนื้อหาตามกรอบ PIE Framework

ย่อหน้า	จุดประสงค์ในการสื่อความ	เหตุผล
นักศึกษาชาวจีนนิยมโพสต์แชร์ชีวิตเรียนต่างประเทศในแพลตฟอร์มโซเชียลยอดนิยมอย่าง 小红书 Xiaohongshu ตอกย้ำความเป็นโซเชียลมีเดียที่มีเป้าหมายคือกลุ่มคนจีนที่นิยมต่างประเทศและความพรีเมี่ยม	ให้ข้อมูล	ย่อหน้านี้ให้ข้อมูลเกี่ยวกับพฤติกรรมของนักศึกษาจีนที่ใช้ Xiaohongshu เพื่อแชร์ประสบการณ์เรียนต่อต่างประเทศ

ย่อหน้า	จุดประสงค์ในการสื่อความ	เหตุผล
โดยคำว่า 留学 หรือเรียนต่างประเทศ กลายมาเป็นคำค้นหายอดฮิตติดเทรนด์ใน Xiaohongshu ณ ขณะนี้ โดยมีจำนวนโพสต์มากกว่า 1แสนโพสต์ ซึ่งมีการจัดหมวดหมู่แบบ vlog ถ่ายทอดชีวิตนักเรียนนอกโดยเฉพาะด้วย	ให้ข้อมูล	ย่อหน้านี้ให้ข้อมูลจากการสังเกตเนื้อหาใน Xiaohongshu โดยระบุว่ามีโพสต์มากกว่า 100,000 โพสต์เกี่ยวกับชีวิตนักเรียนนอก
สำหรับ 泰国留学 หรือเรียนในไทย ก็มีนักศึกษาจีนมาโพสต์แชร์เรื่องราววิถีชีวิตในไทยมากกว่า 2,000 โพสต์ ซึ่งไหนๆ ก็พูดถึงเรื่องราวของการแชร์ชีวิตนักศึกษาจีนในไทย	ให้ข้อมูล	ย่อหน้านี้ให้ข้อมูลเกี่ยวกับจำนวนโพสต์ของนักศึกษาจีนที่แชร์ชีวิตในประเทศไทยโดยอ้างอิงจาก Xiaohongshu
เดี๋ยวพรุ่งนี้อ้ายจงจะรวบรวมมาบอกเล่าว่าคนจีนเขาเขียนถึงไทยอย่างไรกันบ้าง	โน้มน้าว	ย่อหน้านี้เป็นการให้สัญญาว่าจะนำเสนอข้อมูลเพิ่มเติมในวันพรุ่งนี้ เพื่อกระตุ้นให้ผู้อ่านติดตามเรื่องราวต่อไป
ใครที่กำลังอยากทำการตลาดอยากโปรโมทแบรนด์ในไทยให้คนจีนรู้จักผ่านสื่อออนไลน์ โดยเฉพาะต้องการเจาะกลุ่มคนจีนรุ่นใหม่ แพลตฟอร์ม Xiaohongshu ก็เป็นแพลตฟอร์มหนึ่งที่ตรงจุดมากๆ ครับ	โน้มน้าว	ย่อหน้านี้พยายามโน้มน้าวให้ผู้ที่ต้องการทำการตลาดเลือกใช้แพลตฟอร์ม Xiaohongshu
หากใครต้องการสอบถามข้อมูลการตลาดออนไลน์ในจีนเพิ่มเติม สามารถติดต่ออ้ายจงได้เลย	โน้มน้าว	ย่อหน้านี้มี Call to Action เพื่อกระตุ้นให้ผู้อ่านที่สนใจติดต่อผู้เขียนเกี่ยวกับการตลาดออนไลน์ในจีน

นอกจากนี้ยังพบอีกว่า คุณลักษณะเชิงอารมณ์ (Emotion) เป็นปัจจัยสำคัญในการวิเคราะห์และจำแนกวัตถุประสงค์ของผู้เขียน โดยเฉพาะในยุคที่มีข้อความจำนวนมหาศาลถูกเผยแพร่ผ่านสื่อสังคมออนไลน์อย่างต่อเนื่อง ซึ่งงานวิจัยก่อนหน้านี้ส่วนใหญ่มักมุ่งเน้นเฉพาะข้อความสั้นๆ หรือบทสนทนาในลักษณะโต้ตอบที่กระชับเป็นหลัก อย่างไรก็ตาม แนวทางที่นำเสนอในวิทยานิพนธ์นี้ได้มีการขยายขอบเขตการศึกษาให้ครอบคลุมเนื้อหาขนาดยาวขึ้น เช่น บทความเชิงลึกที่มีความซับซ้อนและหลากหลายในเชิงบริบททางอารมณ์และวัตถุประสงค์ของผู้เขียน (Chopra et al., 2024) ดังนั้น งานวิทยานิพนธ์นี้จึงนำเสนอแนวทางในการวิเคราะห์และจำแนกวัตถุประสงค์ของผู้เขียน โดยพิจารณาจากบริบททางอารมณ์เพื่อให้ครอบคลุมและสามารถประยุกต์ใช้ได้กับเนื้อหาที่หลากหลายมากยิ่งขึ้น

2.3 การตรวจจับและวิเคราะห์อารมณ์ในข้อความ (Emotion Detection and Recognition in Text)

Paul Ekman นักจิตวิทยาชาวอเมริกันได้นำเสนอทฤษฎีอารมณ์พื้นฐานของมนุษย์ที่สามารถรับรู้ด้วยการแสดงออกทางใบหน้า แบ่งออกเป็น 6 อารมณ์ (Ekman, 1993) ได้แก่

1) โกรธ (Angry)

2) ขยะแขยง (Disgust)

3) กลัว (Fear)

4) มีความสุข (Happiness)

5) เศร้า (Sadness)

6) ประหลาดใจ (Surprise)

โดย Paul Ekman ได้ระบุว่า นอกเหนือจากทั้ง 6 อารมณ์พื้นฐาน อาจมีอารมณ์อื่น ๆ เพิ่มเติมอีก เราจึงได้เห็นการแบ่งอารมณ์ของมนุษย์ที่มากกว่า 6 อารมณ์ เพิ่มเติมจากของ Paul Ekman อย่างเช่น วงล้ออารมณ์ของโรเบิร์ต พลัทชิค (Plutchik's Wheel of Emotions) ที่สร้างวงล้อหลัก 8 วง ซ้อนทับกัน และจัดวางแบบคู่ตรงข้าม สี่คู่ (Nielek et al., 2017) ดังภาพที่ 2.1 ได้แก่

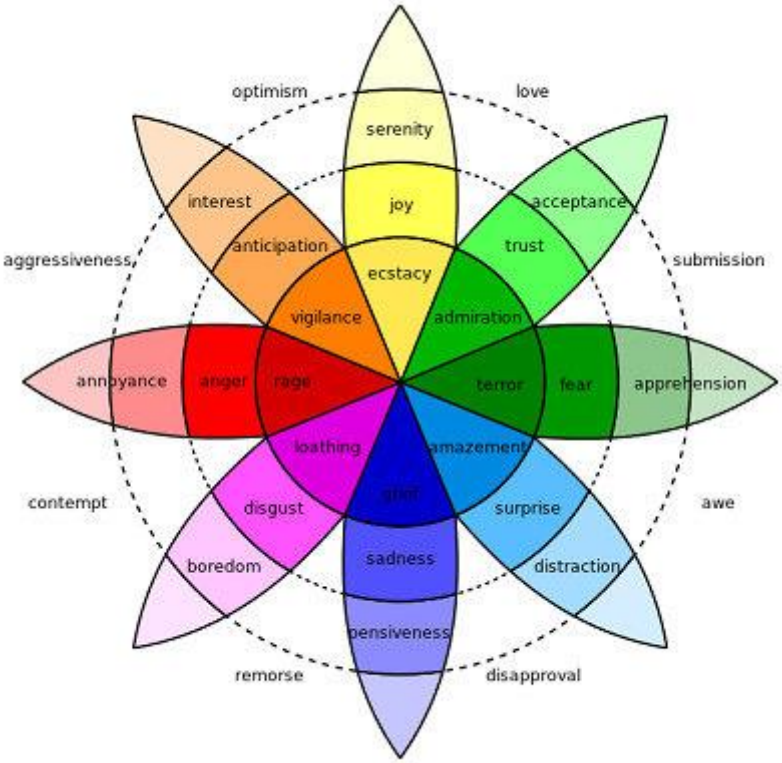
1) ความรื่นเริง (Joy) คู่ตรงข้ามกับ ความเศร้า (Sadness)

2) ความโกรธ (Anger) คู่ตรงข้ามกับ ความกลัว (Fear)

3) ความไว้เนื้อเชื่อใจ (Trust) คู่ตรงข้ามกับ ความรังเกียจ (Disgust)

4) ความประหลาดใจ (Surprise) คู่ตรงข้ามกับ ความคาดหวัง (Anticipation)

โดยในวงล้ออารมณ์ของโรเบิร์ต พลัทชิค ยังแบ่งเป็นอารมณ์แยกย่อยอีก รวมแล้วมีทั้งหมด 32 อารมณ์



ภาพที่ 0.1 Plutchik's Wheel of Emotions

ที่มา: Nielek et al. (2017)

ในงานวิจัยด้านภาษาศาสตร์ การวิเคราะห์อารมณ์ในข้อความมักอ้างอิงทฤษฎีที่แตกต่างกันไปตามลักษณะของข้อความ โดยเฉพาะในกรณีของข้อความขนาดยาว เช่น บทความ ข่าว นิทาน และเรื่องสั้น งานวิจัยหลายชิ้นทั้งในประเทศไทยและต่างประเทศได้นำทฤษฎีของ Ekman ซึ่งแบ่งอารมณ์ออกเป็น 6 ประเภทมาใช้เป็นแนวทางในการจำแนกอารมณ์ของข้อความ (Acheampong et al., 2020)

ตัวอย่างของการใช้ทฤษฎี 6 อารมณ์พื้นฐาน ได้แก่

Alm (2008) จัดหมวดหมู่ประโยคจากนิทานเป็น 6 หมวด ตามทฤษฎีของ Ekman งานวิจัยดังกล่าวเป็นวิทยานิพนธ์ระดับปริญญาเอกและได้รับการตีพิมพ์เป็นหนังสือในปี 2009 โดยเป็นหนึ่งในงานวิจัยแรก ๆ ที่ศึกษาการวิเคราะห์อารมณ์ในข้อความ และได้รับการอ้างอิงมากกว่า 112 ครั้ง (ข้อมูลจาก Semantic Scholar) ขณะที่ Buechel & Hahn (2017) ศึกษาอารมณ์ของบทความรีวิวภาพยนตร์ โดยเปรียบเทียบอารมณ์ของผู้เขียนบทความและความคิดเห็นของผู้อ่าน ด้าน Chaffar &

Inkpen (2011) นำเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้กับการวิเคราะห์อารมณ์ในประโยคภาษาอังกฤษจากนิทาน พาดหัวข่าว และบล็อก

สำหรับเนื้อหาภาษาไทยโดยเฉพาะ มีการนำทฤษฎี 6 อารมณ์พื้นฐานมาใช้เช่นกัน ตัวอย่างเช่น ในงานของ Inrak & Sinthupinyo (2010) วิเคราะห์ข้อความภาษาไทยจากอีเมล บล็อก และกระทู้บนเว็บบอร์ด โดยใช้เทคนิคการเรียนรู้ของเครื่อง ได้แก่ Naïve Bayes, Decision Tree และ Support Vector Machine (SVM) ขณะที่ Sarakit et al. (2015) ศึกษาความคิดเห็นของผู้ใช้ YouTube โดยใช้เทคนิคการเรียนรู้ของเครื่องเช่นเดียวกับงานก่อนหน้า แต่ปรับ Naïve Bayes เป็น Multinomial Naïve Bayes (MNB) และพบว่า SVM ให้ค่า Accuracy สูงสุดที่ 76.14% สำหรับคลิปโฆษณา ขณะที่ MNB ให้ค่า Accuracy สูงสุดที่ 84.48% สำหรับคลิปมิวสิกวิดีโอ

โดยการใช้ทฤษฎี 6 อารมณ์พื้นฐานแตกต่างชัดเจนจากกระบวนการ Sentiment Analysis โดย Yam (2015) นักวิทยาศาสตร์ข้อมูลจาก Microsoft อธิบายบน CSE Developer Blog ถึงความแตกต่างระหว่าง Sentiment Analysis และ Emotion Detection โดยชี้ให้เห็นว่า Sentiment Analysis เป็นการจำแนกข้อความออกเป็น เชิงบวก (Positive), เชิงลบ (Negative) และเป็นกลาง (Neutral) ขณะที่ Emotion Detection มุ่งเน้นไปที่การจำแนกอารมณ์ตามทฤษฎี เช่น โมเดลของ Ekman

แม้ว่าการวิเคราะห์อารมณ์ในข้อความจะเริ่มต้นจาก 6 อารมณ์พื้นฐานของ Ekman แต่มีงานวิจัยหลายชิ้นที่ปรับลดหรือเพิ่มจำนวนคลาสของอารมณ์เพื่อให้เหมาะสมกับบริบทของข้อความ ยกตัวอย่างเช่น Yam (2015) ปรับลดเหลือ 5 อารมณ์ โดยตัดอารมณ์ ความรังเกียจ (Disgust) ออกไป

ในปัจจุบัน งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์อารมณ์ในข้อความยังคงพัฒนาอย่างต่อเนื่อง โดยเน้นไปที่การใช้ โมเดลการเรียนรู้เชิงลึก (Deep Learning) และ Pre-trained Models ซึ่งช่วยเพิ่มประสิทธิภาพของการจำแนกอารมณ์ โดยค่า Accuracy ของโมเดลส่วนใหญ่อยู่ระหว่าง 70-87% (Alswaidan & Menai, 2020)

สำหรับวิทยานิพนธ์นี้ ได้เลือกใช้ 8 อารมณ์ของ Plutchik โดยจัดเป็น 4 คู่ตรงข้าม เนื่องจากมีข้อได้เปรียบเหนือกว่าโมเดลของ Ekman ดังนี้

1) Plutchik แสดงความสัมพันธ์ของอารมณ์ได้ดีกว่า

การจัดอารมณ์เป็นคู่ตรงข้ามช่วยให้สามารถวิเคราะห์ความสัมพันธ์ของอารมณ์ในข้อความได้เป็นระบบ ขณะที่โมเดลของ Ekman ไม่ได้กำหนดโครงสร้างความสัมพันธ์ของอารมณ์

2) Plutchik ครอบคลุมอารมณ์ที่เกี่ยวข้องกับการโน้มน้าวและการคาดการณ์

PIE Framework (Persuade, Inform, Entertain) มีองค์ประกอบของอารมณ์ที่ซับซ้อน เช่น Trust (ความไว้เนื้อเชื่อใจ) และ Anticipation (ความคาดหวัง) ซึ่งไม่มีอยู่ในโมเดลของ Ekman แต่เป็นส่วนสำคัญในการวิเคราะห์เนื้อหาที่มุ่งโน้มน้าวใจ

3) Plutchik รองรับระดับความเข้มข้นของอารมณ์

โมเดลของ Plutchik อนุญาตให้วิเคราะห์ ความเข้มข้นของอารมณ์ เช่น Anger (โกรธ) ⟶ Rage (เกรี้ยวกราด) หรือ Joy (รื่นเริง) ⟶ Ecstasy (ปลื้มปิติ) ซึ่งเป็นประโยชน์ในการวิเคราะห์ข้อความที่มีความแตกต่างกันในระดับอารมณ์

4) Plutchik เหมาะสมกับการวิเคราะห์อารมณ์ในโซเชียลมีเดีย

การศึกษาข้อความจากแพลตฟอร์มโซเชียลมีเดีย เช่น Xiaohongshu และ Twitter พบว่าการใช้ Plutchik's Model สามารถสะท้อนอารมณ์ของผู้ใช้ได้ละเอียดกว่า โดยเฉพาะเมื่อต้องวิเคราะห์อารมณ์ที่ซับซ้อน เช่น Trust (ความเชื่อมั่น) และ Surprise (ความประหลาดใจ) ที่ส่งผลต่อการโน้มน้าวใจ

ด้วยเหตุผลเหล่านี้ โมเดลของ Plutchik จึงเป็นทางเลือกที่เหมาะสมกว่าในการวิเคราะห์อารมณ์ของข้อความผ่าน PIE Framework ซึ่งช่วยให้การจำแนก Persuade, Inform และ Entertain มีความแม่นยำและครอบคลุมมากขึ้น

2.4 สไตล์การเขียนของบทความ

สไตล์การเขียนของบทความ หมายถึง รูปแบบการเลือกใช้คำ สำนวนภาษา วิธีการเรียบเรียง และเล่าเรื่องตามแบบของผู้เขียนแต่ละคน ซึ่งถ้าเป็นการเขียนสำหรับองค์กรหรือหน่วยงาน รวมถึงการเขียนเชิงพาณิชย์ อย่างการเขียนเพื่อแบรนด์สินค้า สไตล์การเขียนถือว่าเป็นสิ่งสำคัญที่จะทำให้ผู้คนสามารถจดจำได้ และนำไปสู่เอกลักษณ์ของบทความและตัวเจ้าของบทความนั้น

ตัวอย่างเช่น เนื้อหาในเฟซบุ๊กเพจอ้ายจง ดังภาพที่ 2.2 มีสไตล์การเขียน ดังนี้

1) เน้นเขียนแบบ เล่า แสดงความคิดเห็น วิเคราะห์ สรุป อัปเดต

2) แบ่งโครงสร้างออกเป็น 4 ส่วนหลัก

- หัวข้อ
- เกริ่นนำ
- ส่วนเนื้อหา
- ส่วนสรุป

3) ใช้สไตล์การเล่า ไม่ใช่แบบเป็นทางการ ดังนั้น จึงปรากฏการใช้คำว่า อ้ายจง หรือ ผม เป็นตัวละครผู้เล่าเรื่องในบทความ

4) ความยาว โดยส่วนใหญ่ ไม่ต่ำกว่า 200 คำ สำหรับบทความที่เขียนขึ้นเอง

ถ้าเป็นการแชร์ โดยส่วนใหญ่ จะไม่แชร์โดยไม่มีแคปชัน หรือคำอธิบาย อย่างน้อยจะต้องมีเขียนแสดงความคิดเห็นตนเอง อย่างน้อย ไม่ต่ำกว่า 10 คำ

5) มีการอ้างอิงแหล่งที่มา และเล่าเรื่องแบบมีที่มาที่ไปทุกครั้ง

ตัวอย่างเช่น อ้ายจงเล่าเรื่องจาก

https://www.globaltimes.cn/page/202106/1225430.shtml

https://m.weibo.cn/status/4644624595157009

ภาพที่ 0.2 ตัวอย่างเนื้อหาบนเฟซบุ๊กเพจอ้ายจง

จากตัวอย่างข้างต้น สอดคล้องกับการวิเคราะห์สไตล์ของเนื้อหาผ่านกระบวนการวิจัยเชิงสำรวจ (Exploratory research) โดยใช้การสัมภาษณ์เชิงลึกและการวิเคราะห์เนื้อหาจากเพจ Facebook จำนวน 10 เพจในประเทศไทย ซึ่งมุ่งเน้นเนื้อหาที่เกี่ยวข้องกับประเทศจีนที่นำเสนอใน

ภาษาไทย การศึกษานี้ช่วยให้เข้าใจถึงกระบวนการคิดและสไตล์การสื่อสารของผู้สร้างเนื้อหาได้ลึกซึ้งขึ้นโดยผลการวิเคราะห์พบว่าสไตล์ของเนื้อหาสามารถแบ่งออกเป็นสองประเภทหลักได้แก่ Translation และ Storytelling

1) เนื้อหาประเภท Translation เป็นการแปลข่าวหรือบทความจากแหล่งข้อมูลต้นทางโดยไม่มีการเพิ่มความคิดเห็นหรือมุมมองของผู้เขียนเอง ลักษณะของเนื้อหากลุ่มนี้มักคงโครงสร้างตามต้นฉบับเดิม โดยมีลักษณะเฉพาะ ได้แก่ การใช้คำศัพท์เฉพาะทางและข้อมูลที่ชัดเจน เช่น ชื่อบุคคลสำคัญ สถานที่ วันที่ และแหล่งข่าวที่ถูกต้องตามต้นฉบับเดิม โครงสร้างของเนื้อหามักเป็นลักษณะ Step-by-Step Reporting โดยทุกส่วนของเนื้อหาจะมีความสอดคล้องกันตั้งแต่บทนำ (Introduction) ส่วนเนื้อหาหลัก (Body) ไปจนถึงบทสรุป (Conclusion) โดยไม่มีการปรับเปลี่ยนลำดับของข้อมูล ในเชิงภาษาศาสตร์ เนื้อหาประเภทนี้มักมี โทนกลาง (Neutral Tone) หรืออาจมีโทนรุนแรงในกรณีที่เนื้อหามาจากต้นฉบับที่มีการเน้นข้อมูลที่ส่งผลต่ออารมณ์โดยตรง

อย่างไรก็ตาม ความแตกต่างของ Translation กับ Storytelling ไม่ได้ขึ้นอยู่กับเนื้อหาที่เป็นเชิงลบหรือเชิงบวก แต่ขึ้นอยู่กับ ลักษณะของการเล่าเรื่องและความเป็นต้นฉบับของผู้เขียน นอกจากนี้ ในเชิงโครงสร้างข้อความ เนื้อหาแบบ Translation จะมีการใช้คำที่เป็นกลาง ไม่มีคำแสดงอารมณ์ของผู้เขียน และมักไม่มีคำสรรพนามบุรุษที่หนึ่ง เช่น ฉัน หรือ เรา

2) เนื้อหาประเภท Storytelling

ในทางตรงกันข้าม Storytelling เป็นเนื้อหาที่มีการนำเสนอข้อมูลผ่านการเล่าเรื่องโดยใช้อารมณ์และมุมมองของผู้เขียนเอง ซึ่งแตกต่างจาก Translation ที่เป็นการถ่ายทอดข้อมูลจากต้นฉบับโดยตรง ลักษณะสำคัญของ Storytelling คือการใช้ โทนเสียงที่มีอารมณ์และบุคลิกของผู้เขียน เช่น การใช้คำคุณศัพท์ คำวิเศษณ์ และคำสรรพนามบุรุษที่หนึ่ง นอกจากนี้ เนื้อหาในกลุ่มนี้ยังมีแนวโน้มที่จะใช้ โครงสร้างการเล่าเรื่องที่มีจุดพีค (Climax Point) โดยเฉพาะในบทสรุป ซึ่งอาจไม่จำเป็นต้องสอดคล้องกับข้อมูลที่กล่าวมาก่อนหน้านี้ในเนื้อหาหลัก ตัวอย่างเช่น การเขียนบทความเกี่ยวกับเหตุการณ์สำคัญของจีนโดยเริ่มจากข้อเท็จจริงที่เป็นกลาง แต่ลงท้ายด้วยมุมมองส่วนตัวของผู้เขียนเกี่ยวกับผลกระทบของเหตุการณ์นั้น ๆ ต่อประเทศไทยหรือความคิดเห็นของประชาชน นอกจากนี้ ในเชิงโครงสร้างของเนื้อหาแบบ Storytelling จะมีการใช้คำที่สะท้อนความรู้สึก เช่น น่าทึ่ง น่าสนใจ น่ากังวล หรือการตั้งคำถามปลายเปิดเพื่อกระตุ้นให้ผู้อ่านมีส่วนร่วม

จากการวิจัยเชิงสำรวจแบบหลายกรณีศึกษาและการวิเคราะห์เนื้อหา ยังพบว่า โครงสร้างของเนื้อหาสามารถแบ่งออกเป็น 3 ส่วนหลัก ได้แก่ บทนำ ส่วนเนื้อหาหลัก และบทสรุป ซึ่งเป็น

โครงสร้างสำคัญที่สามารถใช้เป็นปัจจัยในการจำแนกประเภทของเนื้อหา โดยในกรณีของเนื้อหาแบบ Translation บทสรุปมักเป็นการสรุปข้อมูลจากเนื้อหาหลักโดยตรง และมีความเชื่อมโยงกันอย่างสมบูรณ์ ขณะที่เนื้อหาแบบ Storytelling บทสรุปอาจมีการขยายความ หรือใช้การเล่าเรื่องที่สร้างจุดพีคซึ่งอาจไม่เกี่ยวข้องโดยตรงกับเนื้อหาหลัก แต่เป็นการเสริมบริบทหรือเพิ่มมุมมองของผู้เขียนเอง

การวิเคราะห์ลักษณะของทั้งสองสไตล์นี้ทำโดยใช้วิธีการ Feature Engineering ซึ่งพิจารณาคุณลักษณะสำคัญ ได้แก่ Named Entity Recognition (NER) และ Part of Speech (POS) เพื่อแยกแยะระหว่างคำที่เป็นชื่อเฉพาะและโครงสร้างทางไวยากรณ์ที่แตกต่างกันระหว่าง Translation และ Storytelling นอกจากนี้ ยังมีการใช้ Sentiment Analysis เพื่อตรวจสอบโทนอารมณ์ของเนื้อหา รวมถึงการวิเคราะห์ Similarity Features โดยใช้ Latent Dirichlet Allocation (LDA) เพื่อดูว่าบทสรุปของเนื้อหามีความเชื่อมโยงกับเนื้อหาหลักหรือไม่ หากมีความสอดคล้องกันสูงจะถูกจัดให้อยู่ในประเภท Translation ในขณะที่หากมีการเปลี่ยนแปลงโครงสร้างของเรื่องราวและมีการเพิ่มความคิดเห็นส่วนตัว จะถูกจัดเป็น Storytelling อีกทั้งยังมีการใช้ TF-IDF เพื่อตรวจสอบความสำคัญของคำศัพท์เฉพาะที่อาจบ่งบอกถึงสไตล์ของเนื้อหา เช่น การใช้ศัพท์ทางการหรือคำที่สะท้อนอารมณ์

ผลจากการวิเคราะห์นี้ช่วยให้สามารถจำแนกสไตล์เนื้อหาได้อย่างแม่นยำ และสามารถนำไปใช้เป็นแนวทางสำหรับนักเขียนหรือผู้ผลิตเนื้อหาในการตรวจสอบและปรับปรุงสไตล์ของเนื้อหาของตนให้เหมาะสมกับวัตถุประสงค์ ไม่ว่าจะเป็นการ แปลเนื้อหาโดยตรงให้คงความเป็นกลาง หรือการ เล่าเรื่องโดยมีเอกลักษณ์และมุมมองเฉพาะตัว ซึ่งจะช่วยให้การสื่อสารกับกลุ่มเป้าหมายมีประสิทธิภาพมากขึ้น

2.5 การเรียนรู้เชิงลึก (Deep Learning)

การเรียนรู้เชิงลึกเป็นรูปแบบหนึ่งของการเรียนรู้ของเครื่อง (Machine Learning) โดยมีจุดประสงค์มุ่งเน้นพัฒนาอัลกอริทึมทำงานเลียนแบบสมองของมนุษย์ เพื่อให้สามารถตัดสินใจและเรียนรู้รูปแบบของข้อมูลได้มีประสิทธิภาพมากขึ้นกว่าการเรียนรู้ของเครื่องแบบทั่วไป โดยการเรียนรู้เชิงลึกประกอบไปด้วยโครงข่ายประสาทเทียมเป็นเบื้องหลังสำคัญในการทำงาน ซึ่งแต่ละชั้นของแบบจำลองการเรียนรู้เชิงลึก ก็คือข่ายแต่ละข่ายของโครงข่ายประสาทเทียม ตั้งแต่ 3 ชั้นขึ้นไป (IBM, 2020)

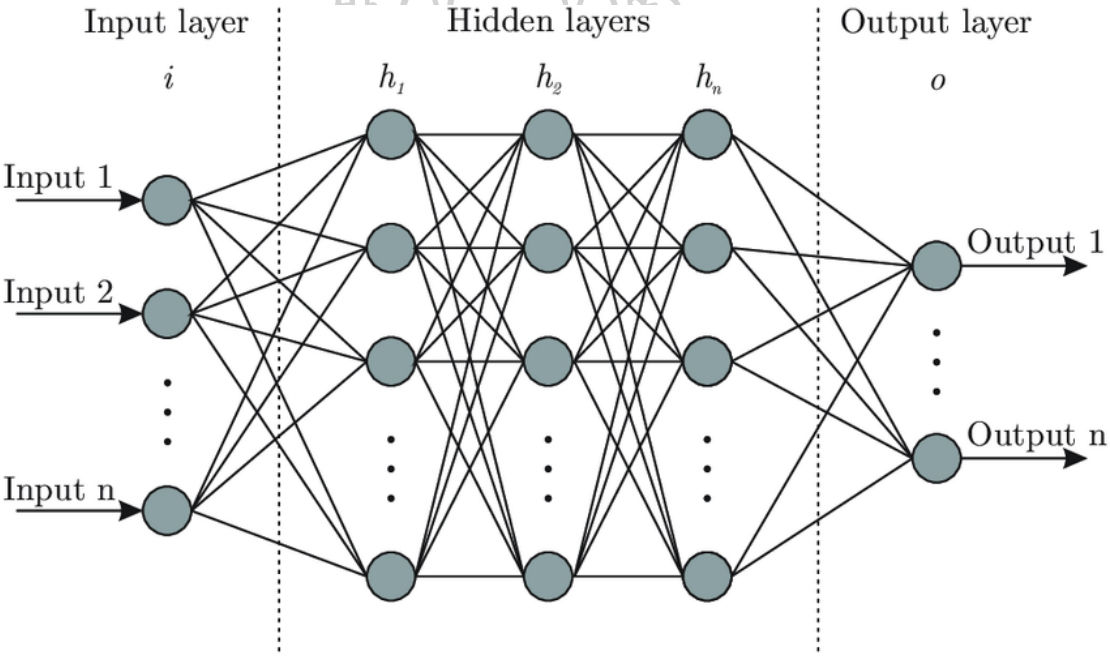
2.6 โครงข่ายประสาทเทียม (Artificial Neural Network)

เป็นการสร้างอัลกอริทึมบนพื้นฐานโมเดลทางคณิตศาสตร์เพื่อจำลองโครงสร้างระบบประสาทในสมองของมนุษย์ โดยแต่ละข่ายของโครงข่ายประสาทเทียมเปรียบได้กับเซลล์ประสาทที่

เชื่อมต่อกัน และรับส่งข้อมูลระหว่างเซลล์ในแต่ละชั้น โดยสามารถแบ่งชั้นแบบจำลองโครงข่ายประสาทเทียมได้เป็น สามชั้นหลัก ได้แก่ 1) ชั้นนำเข้าข้อมูล (Input Layer) 2) ชั้นประมวลผลข้อมูลหรือชั้นซ่อนของโครงข่าย (Hidden Layer) ชั้นนี้จะมีกี่ชั้นก็ได้ มีความสำคัญต่อการประมวลผล และเป็นส่วนในการเรียนรู้ของโครงข่ายประสาทเทียม 3) ชั้นส่งออกผลลัพธ์ (Output Layer) เป็นชั้นที่ส่งออกผลลัพธ์จากการประมวลผลในชั้นประมวลผลข้อมูล ออกไปยังเครื่องเพื่อให้นำผลลัพธ์ไปใช้ประโยชน์ต่อไป

สำหรับชั้นนำเข้าข้อมูล และ ชั้นส่งออกผลลัพธ์ จะมีกี่โหนดนำเข้า/ส่งออก ก็ได้ ขึ้นอยู่กับจำนวนข้อมูลที่นำเข้า และจำนวนผลลัพธ์ที่ส่งออก

จากภาพที่ 2.3 เป็นตัวอย่างโครงข่ายประสาทเทียมที่ประกอบไปด้วยสามชั้นหลักดังที่กล่าวไปข้างต้น



ภาพที่ 0.3 ตัวอย่างของโครงข่ายประสาทเทียม

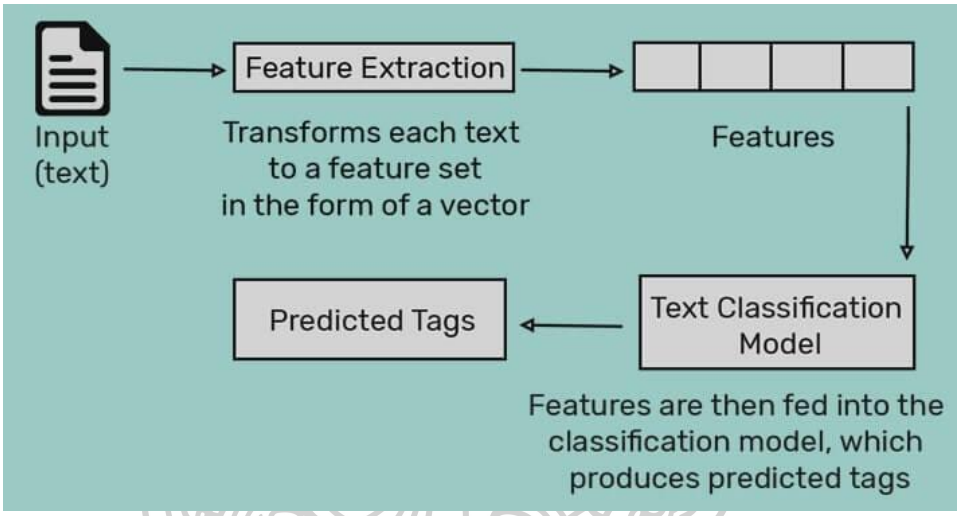
ที่มา: Bre et al. (2018)

2.7 การแบ่งกลุ่มข้อความ (Text Classification)

การแบ่งกลุ่มข้อความ เป็นหนึ่งในปัญหาการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติที่ได้รับความสนใจอย่างมาก มีจุดประสงค์แบ่งข้อความตามกลุ่มหรือประเภทที่ต้องการโดยสามารถแบ่งกลุ่มข้อความได้ตั้งแต่ข้อความสั้น เพียงไม่กี่ประโยค ไปจนถึงข้อความจำนวนมาก ในลักษณะของเอกสารหรือบทความ ซึ่งในปัจจุบัน การวิจัยทางด้านการแบ่งกลุ่มข้อความขนาด

ยาว หรือการแบ่งกลุ่มเอกสาร (Document Classification) เริ่มเป็นที่นิยมเพิ่มขึ้น ซึ่งส่วนใหญ่โฟกัสในการทำแบบจำลองการเรียนรู้เชิงลึก และใช้เทคนิคการสร้างโมเดลที่ฝึกสอนแล้วในภาษาต่าง ๆ มาประยุกต์ใช้

จากภาพที่ 2.4 เป็นตัวอย่างโครงสร้างการทำงานของการแบ่งกลุ่มข้อความ โดยมีการนำเข้าข้อความ และสกัดฟีเจอร์หรือปัจจัยที่มีผลต่อการแบ่งกลุ่มของข้อความจากแต่ละข้อความ จากนั้นนำไปประมวลผลโดยแบบจำลองการแบ่งกลุ่มข้อความ ซึ่งในตัวอย่างนี้ ได้นำผลลัพธ์ของการแบ่งกลุ่มข้อความไปใช้ในการพยากรณ์คำสำคัญ (Tag) ของข้อความนั้น

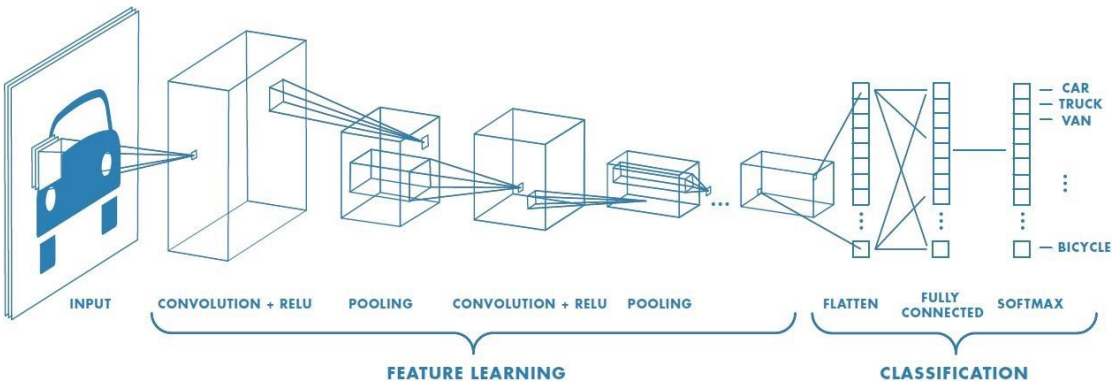


ภาพที่ 0.4 ตัวอย่างโครงสร้างการทำงานการแบ่งกลุ่มข้อความ

ที่มา: COGNITO (2021)

2.8 Convolutional Network (CNN)

Convolutional Network (CNN) ในภาษาไทยเรียกว่าโครงข่ายประสาทแบบคอนโวลูชันจัดเป็นโครงข่ายประสาทเทียมรูปแบบหนึ่งที่มีการจำลองสมองของมนุษย์ในการมองเห็นพื้นที่ขนาดใหญ่แบ่งเป็นพื้นที่ย่อย ๆ เพื่อวิเคราะห์คุณลักษณะโดยย่อ ก่อนที่จะนำมาพิจารณารวมกันเพื่อวิเคราะห์พื้นที่ขนาดใหญ่ทั้งหมด โดยส่วนสำคัญของ CNN คือ การทำ Feature Learning หรือการสกัดคุณลักษณะของข้อมูลนั้นออกมา เพื่อให้เข้าใจถึงข้อมูลนั้น ซึ่งเป็นการเรียนรู้ของเครื่องแบบการเรียนรู้เชิงลึก ดังปรากฏในภาพที่ 2.5



ภาพที่ 0.5 ภาพรวมการวิเคราะห์ข้อมูลด้วย CNN

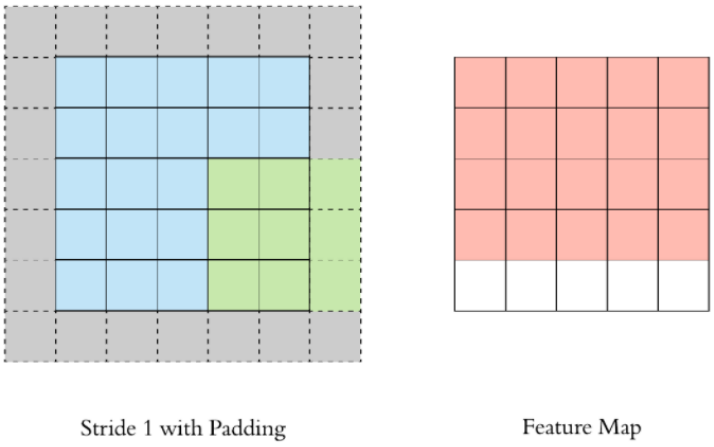
ที่มา: Saha (2018)

ในการทำ Feature Learning มีสิ่งที่เราต้องกำหนดดังนี้

1) การกำหนดค่าในตัวกรองที่เรียกว่า Filter หรือ Kernal โดยการสกัดคุณลักษณะของข้อมูลออกมาเราจำเป็นต้องมีหลายตัวกรอง เพื่อให้โมเดลสามารถเรียนรู้และเข้าใจคุณลักษณะนั้นได้

2) การกำหนดจำนวน Stride เป็นการกำหนดจำนวน Step ในการเลื่อนตัวกรอง ถ้าเรากำหนดค่าของ Stride มากขึ้น หมายความว่าเราต้องการคำนวณหาคุณลักษณะพื้นที่ทับซ้อนน้อยลง กล่าวคือเป็นการมองภาพรวมแบบคร่าว ๆ ยิ่งมากขึ้นเท่าไหร่ก็ยิ่งไม่ละเอียด

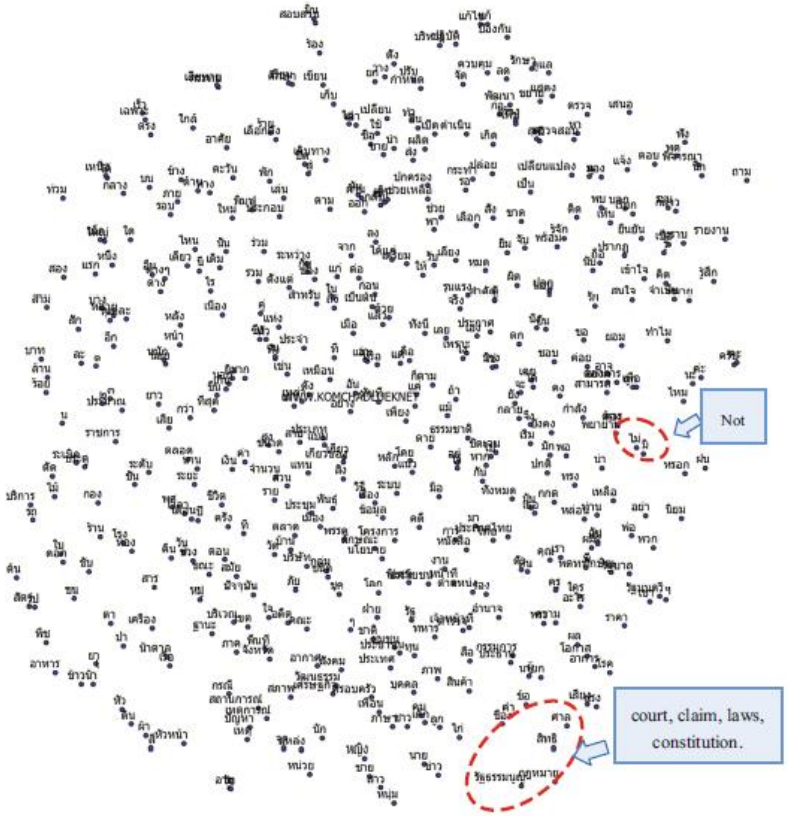
3) การกำหนดค่า Padding เป็นการเติมค่าอะไรบางอย่าง เช่น เลข 0 ลงไปในตารางข้อมูล เพื่อให้ขนาดของตารางคุณลักษณะ (Feature map) ที่ออกมา ยังคงมีขนาดเท่ากับ Input ถ้าดูตามภาพที่ 2.6 การทำ Padding คือการเติมพื้นที่สีเทา ล้อมรอบข้อมูล



ภาพที่ 0.6 ตัวอย่างการทำ Padding ของ CNN

ที่มา: Dertat (2017)

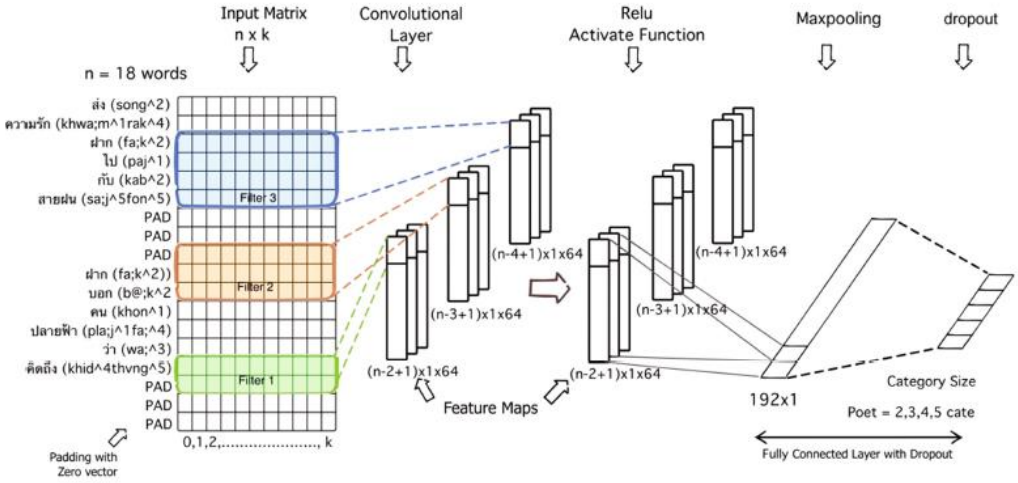
ปัจจุบันมีการนำ CNN มาใช้ในการจัดกลุ่มข้อมูล (Text Classification) มากขึ้น โดยการแปลงข้อความให้เป็นตัวเลขและจัดการให้อยู่ในรูปแบบของเวกเตอร์ที่มีมิติลดลง โดยเราเรียกกระบวนการนี้ว่า Word Embedding จากนั้นเรานำผลลัพธ์ที่ได้จากการทำ Word Embedding เข้าไปวิเคราะห์ด้วยเทคนิค CNN ได้ อย่างในการประมวลผลข้อความภาษาไทย มีการนำ CNN มาใช้เช่นกัน ตัวอย่างเช่น การแบ่งกลุ่มประเภทของกลอนสุภาพและการสื่อสารทัศนคติ หรือ Sentiment จากกลอนสุภาพที่แต่งขึ้น โดยงานวิจัยของ Promrit and Waijanya (2017) สร้างโมเดลการเรียนรู้เชิงลึก CNN มาวิเคราะห์ประเภทและการสื่อสารทัศนคติของกลอนสุภาพ โดยนำชุดข้อมูลมาทำ Word Embedding ด้วยเทคนิค Word2Vec ดังภาพที่ 2.7 โดยชุดข้อมูลภาษาไทยที่นำมาทำ Word Embedding ประกอบไปด้วยแหล่งออนไลน์ 5 แหล่ง ได้แก่ (1) BEST I Corpus by NECTEC (2) Contemporary Poets Society: www.kawethai.com (3) www.wannakadee.com (4) www.thaipoem. com และ (5) www.aromklon.com รวมจำนวนคำทั้งสิ้น 5.9 ล้านคำ จากนั้นนำมาทำ Bag of Words ได้ 101,432 คำ



ภาพที่ 0.7 การทำ Word Embedding

ที่มา: Promrit and Waijanya (2017)

เมื่อทำ Word Embedding แล้วจากนั้นนำเข้าสู่โมเดล CNN โดยมีโครงสร้างดังภาพที่ 2.8



ภาพที่ 0.8 โครงสร้างโมเดล CNN สำหรับตรวจสอบโครงสร้างกลอนแปดสุภาพภาษาไทย

ที่มา: Promrit and Waijanya (2017)

ในการจัดทำโมเดล CNN ในงานวิจัยดังกล่าว ใช้ความสามารถอันโดดเด่นของ CNN ในเรื่องของ Feature learning เรียนรู้จุดเด่นในองค์ประกอบของข้อมูล เพื่อให้โมเดลสามารถเรียนรู้และวิเคราะห์ได้ว่าส่วนใดสัมพันธ์กันบ้าง โดยโมเดลที่ทำขึ้นสามารถจัดกลุ่มกลอนสุภาพและวิเคราะห์ Sentiment ได้แม่นยำ โดยได้Accuracy 83% สำหรับการแบ่งกลุ่มกลอน และ 61% สำหรับการวิเคราะห์ Sentiment เป็นประสิทธิภาพที่ดีที่สุดเมื่อเทียบกับ SVM และ Naïve Bayes ดังปรากฏผลลัพธ์ในภาพที่ 2.9

Models	Accuracy of poem category classification	Accuracy of poem sentiment analysis
CNNs (k = 200, feature maps = 64)	83.00%	61.00%
SVM	16.36%	24.54%
Naïve Bayes	14.54%	41.36%

ภาพที่ 0.9 แสดงผลลัพธ์ Accuracy การจัดกลุ่มและวิเคราะห์ Sentiment กลอนสุภาพภาษาไทยด้วย CNN Model เทียบกับ SVM และ Naïve Bayes

ที่มา: Promrit and Waijanya (2017)

จากการพัฒนาโมเดล CNN เพื่อจัดกลุ่มและวิเคราะห์ Sentiment ของกลอนสุภาพภาษาไทย Waijanya and Promrit (2017) ได้ต่อยอดพัฒนาโมเดล CNN เพื่อทำนายว่า กลอนสุภาพ

ที่แต่งขึ้นเป็นของผู้แต่งท่านใด โดยผลการทดลองปรากฏว่า โมเดล CNN สามารถทำนายผู้แต่งกลอนได้อย่างถูกต้อง ได้ Accuracy 100% เมื่อมีกลอน 2 กลอนมาจับคู่ทำนาย แต่ Accuracy เริ่มลดลงเมื่อมีจำนวนผู้แต่งมากขึ้น โดยทดลองมากสุด 5 ตัวเลือก โดยมีผล Accuracy ดังภาพที่ 2.10

Number of poet	Accuracy (%)
2	100
3	80.55
4	72.92
5	57.32

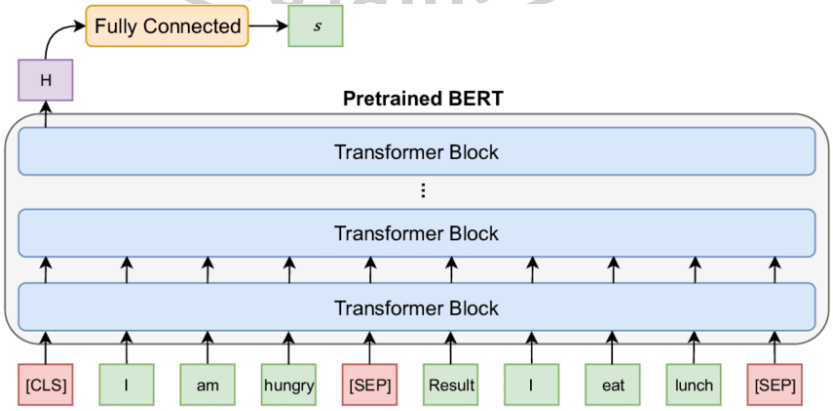
ภาพที่ 0.10 แสดงผลลัพธ์ Accuracy ทำนายว่ากลอนที่แต่งขึ้นเป็นของใคร ตามจำนวนตัวเลือกกลอน ระหว่าง 2 – 5 กลอน

ที่มา: Waijanya and Promrit (2017)

2.9 การประมวลผลข้อความด้วย BERT

BERT เป็นเทคนิคที่ใช้ในการประมวลผลข้อความที่ได้รับความนิยมเป็นอย่างมากในปัจจุบัน โดย BERT เป็นคำย่อของเทคนิคการประมวลผลข้อความด้วยการเรียนรู้ของเครื่องแบบ Bidirectional Encoder Representations from Transformer ซึ่งเป็นเทคนิคที่พัฒนาโดย Google ทำมาเพื่อประมวลผลข้อความโดยเฉพาะ โดยสามารถใช้ประมวลผลภาษาไทยได้ด้วยเช่นกัน (Devlin & Chang, 2018)

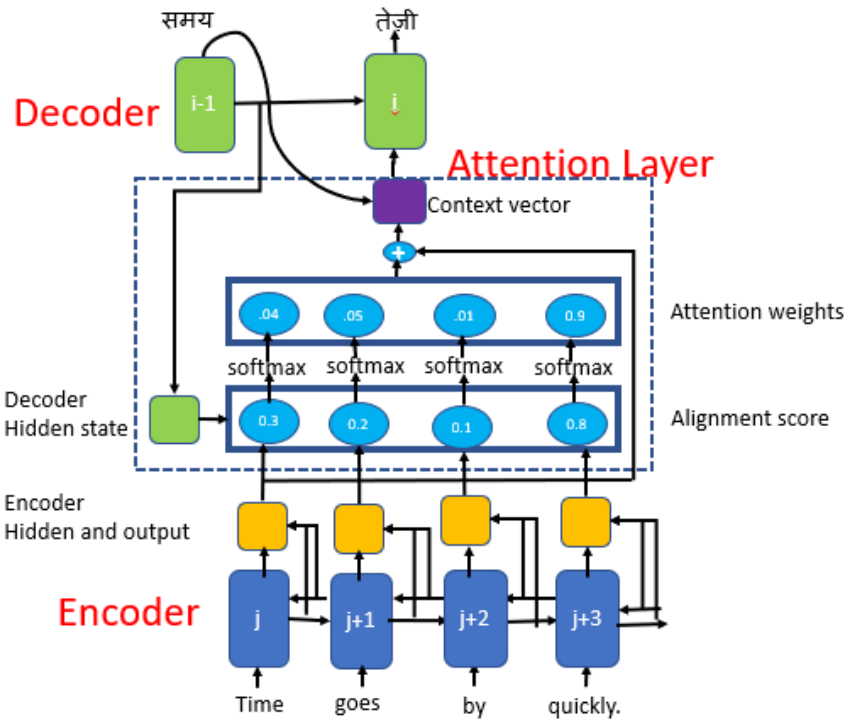
โครงสร้างของ BERT ปรากฏดังภาพที่ 2.11



ภาพที่ 0.11 โครงสร้างของ BERT (BERT Structure)

ที่มา: He et al. (2020)

จากโครงสร้างข้างต้น อธิบายได้ว่าสิ่งหลักที่ทำให้ BERT แตกต่างจากเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกแบบอื่น ได้แก่ การใช้การเรียนรู้แบบสองทิศทางของสถาปัตยกรรมแบบ Transformer ซึ่งมีหัวใจสำคัญคือ Attention mechanism โดยมีงานวิจัยระบุว่า Attention สามารถแก้ปัญหาคอขวด ข้อมูลบางอย่างตกหล่นในการส่งข้อมูลแบบเรียงแถวเพื่อไปประมวลผลข้อความที่เรานิยมทำกันแบบ Seq2Seq (Sequence to Sequence) ได้ เนื่องจาก Attention สามารถสร้าง Output ที่ไปโฟกัสในตำแหน่งใดของ Input ก็ได้ (Khandelwal, 2020)



ภาพที่ 0.12 ตัวอย่างการประมวลผลข้อความด้วย Attention Mechanism หัวใจหลักของ BERT

ที่มา: Khandelwal (2020)

นอกจากนี้จำนวนการ Train โดย BERT ต้อง Train 2 รอบ รอบแรกเป็นการนำคลังข้อมูลขนาดใหญ่เข้าไปสอน BERT เพื่อให้เรียนรู้เกี่ยวกับโครงสร้างและลักษณะของภาษาที่คลังข้อมูลนั้นใช้ โดยไม่ต้องมีการ Label ใด ๆ ขั้นตอนนี้เรียกว่าการทำ Pretrained Model หลังจากได้ Pretrained Model มาแล้ว BERT จะ Train โดยใช้ Pretrained Model ร่วมกับชุดข้อมูลที่มีการ Label ไว้แล้ว เพื่อเรียนรู้ข้อมูล และสามารถแยกแยะข้อมูลได้ (Zhou et al., 2020)

2.10 การใช้ Ensemble Learning ในการจำแนกประเภทข้อความ

โมเดล Ensemble Learning ได้รับความนิยมอย่างมากในงานจำแนกประเภทข้อความ เนื่องจากช่วยเพิ่มความแม่นยำโดยใช้ การรวมหลายโมเดลเข้าด้วยกัน เพื่อลดข้อผิดพลาดที่เกิดจาก Variance และ Bias ทำให้เหมาะสมอย่างยิ่งสำหรับงานวิเคราะห์รูปแบบเนื้อหา (Content Style Analysis)

ในการวิจัยนี้ เราใช้โมเดล Ensemble Learning ที่ได้รับการยอมรับ 4 แบบ ได้แก่

1) Random Forest

2) XGBoost

3) LightGBM

4) AdaBoost

การใช้ Ensemble Methods ช่วยให้การจำแนกประเภทข้อความมีความแม่นยำมากขึ้น ซึ่งมีงานวิจัยสนับสนุนแนวทางนี้ ตัวอย่างเช่น

งานวิจัยของ Phann et al. (2023) แสดงให้เห็นว่า การจำแนกข่าวหลายประเภทในภาษาเขมรโดยใช้ Ensemble Methods ให้ผลลัพธ์ที่ดีกว่าโมเดลเดี่ยว ซึ่งชี้ให้เห็นถึงประสิทธิภาพของแนวทางนี้ โดยเฉพาะในภาษาที่มีทรัพยากรจำกัด ขณะที่ Siriphanpornchana และ Kanjanawattana (2023) ศึกษาการใช้ Machine Learning สำหรับการตรวจจับข่าวปลอมในภาษาไทย พบว่า Ensemble Methods ให้ผลลัพธ์ที่ดีกว่าอัลกอริธึมเดี่ยว ด้าน Hou et al. (2023) นำเสนอ PromptBoosting ซึ่งเป็นการผสม AdaBoost กับตัวจำแนกประเภทอื่นๆ เพื่อให้สามารถจำแนกข้อความได้อย่างแม่นยำแม้ในสภาวะที่มีทรัพยากรจำกัด

นอกจากนี้ งานวิจัยหลายชิ้นยังแสดงให้เห็นถึงศักยภาพของ Random Forest ในงานวิเคราะห์ข้อความประเภทต่างๆ เช่น Taherkhani et al. (2023) ใช้ Random Forest วิเคราะห์ข้อมูลเพื่อตรวจจับลูกค้าที่มีแนวโน้มจะยกเลิกบริการในอุตสาหกรรมโรงแรม โดยใช้การทำเหมืองข้อความ (Text Mining) ส่วน Dewi et al. (2023) ปรับแต่ง Random Forest เพื่อจำแนกบทความจาก Cable News Network (CNN) และชี้ให้เห็นว่าการปรับแต่งโมเดลให้เข้ากับเนื้อหาประเภทเฉพาะสามารถเพิ่มประสิทธิภาพได้ ด้าน Wan et al. (2021) ศึกษาการวิเคราะห์เนื้อหาเอกสารโดยใช้ Random Forest ซึ่งแสดงให้เห็นถึงความสามารถในการจัดการกับชุดข้อมูลขนาดใหญ่

โดย Sameen et al. (2016) พัฒนากระบวนการคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้ Random Forests เพื่อเพิ่มประสิทธิภาพการจำแนกประเภทข้อความสั้น ซึ่ง

สอดคล้องกับ Mehta et al. (2021) ที่ใช้การสร้างคุณลักษณะ (Feature Engineering) ด้วย Word2Vec สำหรับการจำแนกข้อความร่วมกับอัลกอริธึม K-Nearest Neighbor (KNN) เพื่อเพิ่มความแม่นยำ

นอกจาก Random Forest และ AdaBoost แล้ว XGBoost และ LightGBM ยังเป็น Gradient Boosting Methods ที่ทรงพลังและได้รับความนิยมในงานจำแนกข้อความ ตัวอย่างเช่น Putra et al. (2022) ศึกษาการจัดประเภทข้อความของสิทธิบัตรในอุตสาหกรรมยานยนต์พลังงานใหม่ (New Energy Vehicles) และพบว่า XGBoost ช่วยเพิ่มประสิทธิภาพในการกรองคำ (Word Filtering) ด้าน Han (2023) ใช้ XGBoost ในการวิเคราะห์ Keystroke Authentication จากข้อความอิสระ ซึ่งแสดงให้เห็นถึงความสามารถของโมเดลในหลายๆ โดเมน ขณะที่ Irwanto & Goeirmanto (2023) เปรียบเทียบ CNN, XGBoost และ SVM ในการจำแนกความคิดเห็น และศึกษาการวิเคราะห์อารมณ์ (Sentiment Analysis) จากข้อความบน Twitter เกี่ยวกับวัคซีน COVID-19 โดยใช้ Naïve Bayes และ XGBoost

สำหรับ LightGBM ซึ่งเป็นอีกหนึ่งโมเดลที่มีประสิทธิภาพสูงสำหรับงานจำแนกข้อความ โดยมีหลายงานทีนำมาใช้ อาทิ Chandran et al. (2023) แนะนำ TopicStriker ซึ่งใช้ Topic Modeling ร่วมกับ LightGBM เพื่อเพิ่มประสิทธิภาพของการจำแนกประเภทข้อความ ด้าน Zhang et al. (2023) พัฒนาโมเดลจำแนกความคิดเห็นโดยใช้ LightGBM ซึ่งแสดงให้เห็นถึงศักยภาพของโมเดลในงานวิเคราะห์เนื้อหา ขณะที่ Taha et al. (2024) ศึกษาเทคนิคการจำแนกประเภทข้อความโดยเน้นไปที่ Gradient Boosting Methods ซึ่งพิสูจน์ให้เห็นว่า Ensemble Learning สามารถจัดการกับงานจำแนกข้อความที่หลากหลายได้ดี

2.11 อัลกอริทึมการวัดความคล้าย (Similarity Algorithm)

ในการวิเคราะห์ความคล้ายของเนื้อหาสองชุดในวิทยานิพนธ์นี้ ได้เลือกใช้ Hellinger Distance และ Hellinger Similarity ในการประเมินความเหมือนและความแตกต่างระหว่างเนื้อหา เนื่องจากข้อมูลที่ได้รับจากโมเดลมีลักษณะเป็นการแจกแจงความน่าจะเป็น (probability distributions)

Hellinger Distance เป็นวิธีวัดระยะทางระหว่าง probability distributions สองชุด ดังสมการที่ 2.1

$$H(p,q) = \frac{1}{\sqrt{2}}\sqrt{\sum\left(\sqrt{p_i} - \sqrt{q_i}\right)^2} \qquad (2.1)$$

โดยที่

- p_i และ q_i คือค่าความน่าจะเป็นของข้อมูลแต่ละหมวดหมู่

- ค่าของ Hellinger Distance มีช่วงระหว่าง 0 ถึง $1/\sqrt{2}$ โดย 0 แสดงถึงความเหมือนกัน และค่า $1/\sqrt{2}$ แสดงถึงความแตกต่างสูงสุด

จาก Hellinger Distance เราสามารถคำนวณค่า Hellinger Similarity เพื่อให้ง่ายต่อการตีความ ดังสมการ

$$\text{Hellinger Similarity} = [1 - (H(p,q)/(1/\sqrt{2}))] \times 100\%$$

ค่า similarity จะอยู่ในช่วง 0% ถึง 100% โดย:

- 100% แสดงว่าการแจกแจงทั้งสองเหมือนกันสมบูรณ์

- ค่าใกล้ 0% แสดงว่าการแจกแจงสองชุดแตกต่างกันมาก

สำหรับข้อดีและเหตุผลของการเลือกใช้ Hellinger Similarity สรุปได้ดังนี้

1) เหมาะสมกับข้อมูลแบบ Probability Distribution

วิธีที่นิยมใช้ทั่วไป เช่น Cosine Similarity หรือ Pearson Correlation ไม่เหมาะสมกับข้อมูลประเภทนี้ เนื่องจากไม่ได้คำนึงถึงการจำกัดผลรวมความน่าจะเป็นให้เท่ากับ 1 โดยตรง (Goussakov, 2020)

2) มีประสิทธิภาพในการจัดการข้อมูลที่เบาบาง (Sparse Data)

Hellinger Distance มีประสิทธิภาพสูงเมื่อใช้กับข้อมูลที่มีความเบาบาง (Sparse data) และสามารถแก้ปัญหาที่ต้องใช้ข้อมูลที่ผู้ใช้อาจไม่มีคะแนนร่วมกันได้ดีกว่าวิธีการอื่นๆ ที่อาศัยเฉพาะรายการที่ถูกให้คะแนนร่วมกันเท่านั้น (Wang et al., 2022)

3) เสถียรต่อการเปลี่ยนแปลงเล็กน้อยในข้อมูล (Robustness)

Hellinger Similarity มีความเสถียรต่อการเปลี่ยนแปลงเล็กๆ น้อยๆ ของข้อมูลมากกว่า Jensen-Shannon Divergence หรือ KL-Divergence ซึ่งอาจมีความไวต่อข้อมูลที่มีค่าความน่าจะเป็นต่ำจนเกินไป (Goussakov, 2020)

4) มีขอบเขตที่ชัดเจนและเข้าใจง่าย

ค่าของ Hellinger Similarity ถูกจำกัดอยู่ในช่วงที่เข้าใจง่าย ช่วยในการตีความผลลัพธ์ได้ชัดเจนและง่ายต่อการสื่อสาร

2.12 งานวิจัยที่เกี่ยวข้อง

ในงานวิทยานิพนธ์นี้ นำการวิเคราะห์เนื้อหาเชิงสำรวจ (Exploratory Content Analysis) มาใช้ร่วมด้วยกับวิธีทางวิทยาการคอมพิวเตอร์ โดยการวิเคราะห์เนื้อหาเชิงสำรวจเป็นวิธีที่ช่วยถอดองค์ความรู้จากผู้เชี่ยวชาญและผู้มีส่วนเกี่ยวข้องกับชุดข้อมูล อีกทั้งยังสามารถประยุกต์ใช้ได้ในหลากหลายสาขา เช่น ภาษาศาสตร์ การสื่อสาร และการตลาด โดยเฉพาะในยุคที่เนื้อหาดิจิทัลมีบทบาทสำคัญมากขึ้น การวิเคราะห์ลักษณะ รูปแบบ และธีมที่ซ่อนอยู่ในบริบทเฉพาะจึงเป็นสิ่งจำเป็นตัวอย่างเช่น Willoughby et al. (2024) ได้ใช้การวิเคราะห์เนื้อหาในการศึกษากลยุทธ์การสื่อสารด้านสุขภาพและการทำให้ร่างกายเป็นวัตถุผ่านโพสต์ของผู้มีอิทธิพลทางฟิตเนสบนโซเชียลมีเดีย ซึ่งชี้ให้เห็นถึงบทบาทของการวิเคราะห์เนื้อหาในการประเมินผลกระทบทางสุขภาพและการพิจารณาจริยธรรมในการสื่อสาร ในบริบทของข่าวปลอมระหว่างการแพร่ระบาดของโควิด-19 Anastasia et al. (2020) ได้วิเคราะห์เนื้อหาเชิงคุณภาพเพื่อจำแนกประเภทของข้อมูลที่ผิดและแหล่งที่มา ซึ่งถือเป็นแนวทางสำคัญในการพัฒนายุทธศาสตร์การสื่อสารสาธารณสุขที่มีประสิทธิภาพ

ในด้านการตลาด Ho et al. (2020) ใช้การสัมภาษณ์แบบกึ่งโครงสร้างกับผู้จัดการฝ่ายการตลาด เพื่อเสนอกรอบแนวคิดในการพัฒนาขีดความสามารถด้านการตลาดด้วยเนื้อหา โดยเน้นให้เห็นถึงสไตล์และกลยุทธ์ที่องค์กรนำมาใช้จริง ในด้านภาษาศาสตร์ Iyer และ Rose (2020) เสนอกรอบการทำงานด้วยวิธีการเรียนรู้ของเครื่องเพื่อระบุตัวตนของผู้เขียนจากข้อความ โดยอาศัยคุณลักษณะด้านสไตล์ เช่น รูปแบบการใช้คำและโครงสร้างประโยค ซึ่งสามารถประยุกต์ใช้ในงานวรรณกรรมเชิงนิติวิทยาศาสตร์ การตรวจสอบลิขสิทธิ์ และการวิเคราะห์เอกสารเชิงประวัติศาสตร์

Sabol และ Horák (2022) มุ่งเน้นการตรวจจับลักษณะการเขียนที่บิดเบือนในข่าวภาษาเช็กโดยใช้คุณลักษณะด้านสไตล์ร่วมกับโมเดลการเรียนรู้ของเครื่องเพื่อวิเคราะห์อคติ ขณะที่ Ríos-Toledo et al. (2022) ศึกษาการเปลี่ยนแปลงของสไตล์การเขียนในวรรณกรรมโดยใช้เทคนิค N-gram และการเรียนรู้แบบมีผู้สอนเพื่อสะท้อนการเปลี่ยนแปลงทางภาษาตามกาลเวลา ด้าน Onan (2018) เสนอการจำแนกประเภทของเนื้อหาตามประเภทวรรณกรรม โดยผสานการวิเคราะห์หน้าที่ของภาษา (Language Function Analysis) เข้ากับการออกแบบคุณลักษณะ (Feature Engineering) เพื่อเข้าใจความหลากหลายของสไตล์งานเขียน

ปัญหาหนึ่งที่พบได้บ่อยในการสร้างเนื้อหาเชิงการตลาด คือ ความไม่สอดคล้องระหว่างบทความที่ผลิตขึ้นกับรูปแบบหรือธีมที่กำหนดไว้ ซึ่งอาจมีสาเหตุมาจากทั้งรูปแบบทางสไตล์และโครงสร้างของเนื้อหา เช่น บทความข่าวที่มักประกอบด้วย 4 ส่วน ได้แก่ 1) หัวข้อข่าว (Headline)

เพื่อเรียกความสนใจ 2) คำนำ (Lead) เพื่อสรุปประเด็นสำคัญ 3) เนื้อหา (Body) ซึ่งขยายข้อมูลหลัก และ 4) ท้ายบทความ (Tail) ซึ่งใช้สรุปหรือเสริมข้อมูล (Conforti, Pilehvar & Collier, 2018)

หนึ่งปัจจัยสำคัญคือ รูปแบบเนื้อหาแบบ timely และ Timeless ซึ่งเป็นสไตล์ที่ใช้กันแพร่หลายในยุคของการตลาดดิจิทัล โดยเฉพาะบนเว็บไซต์ บล็อก และสื่อสังคมออนไลน์ (Hanlon, 2019; Busche, 2017; Flashman, 2020; Smith, 2020; Tuten, 2020) ความแตกต่างหลักอยู่ที่ความอ่อนไหวต่อเวลา (Time Sensitivity) และวัตถุประสงค์ของเนื้อหา (Content Objectives) โดยเนื้อหาแบบ timeless เป็นเนื้อหาที่ยังมีคุณค่าแม้ผ่านกาลเวลา เช่น บทความเกี่ยวกับประวัติศาสตร์จีน (Henricks & Shelton, 2016) ขณะที่เนื้อหาแบบ timely เช่น ข่าวเกี่ยวกับสถานการณ์โควิด-19 ในประเทศจีน มักล้าสมัยได้อย่างรวดเร็ว (Clampitt, 2017) นอกจากนี้งานวิจัยที่ผ่านมายังพบว่าวัตถุประสงค์ของเนื้อหา timeless มักใช้เพื่อกลยุทธ์ระยะยาว เช่น บทความแนะนำสถานที่ท่องเที่ยวเพื่อ SEO (Wilson, 2019) ขณะที่ timely ใช้ในแคมเปญระยะสั้น เช่น ข่าว การอัปเดตเหตุการณ์หรือบทความที่ใช้คำยอดนิยม (Beverland, 2021; Gonçalves, 2017)

การจำแนกจุดประสงค์ในการสื่อความ (Intention Classification) ซึ่งหมายถึงการระบุวัตถุประสงค์เบื้องหลังข้อความ มีความสำคัญในหลายบริบท เช่น การให้บริการลูกค้า และการจัดการปัญหา งานวิจัยหลายฉบับได้พัฒนาเทคนิคใหม่เพื่อรองรับบริบทที่มีข้อมูลจำกัดและการจำแนกแบบหลายป้ายกำกับ เช่น Damrongrat et al. (2022) ที่จำแนกเจตนาในข่าวไทยเป็น 3 ประเภท ได้แก่ แจ้งข้อมูล อธิบาย และเสนอแนวทาง พร้อมเสนอวิธี Non-Entity โดยใช้คำเชื่อมและคำหยุด เพื่อเพิ่มความแม่นยำในการจำแนกเจตนา

Vatathanavaro et al. (2020) แก้ปัญหาความไม่สมดุลของข้อมูลในระบบธนาคารผ่านมือถือด้วยโมเดล Bidirectional LSTM ร่วมกับ CNN ซึ่งเพิ่ม F1-score ได้ถึง 4.5% ส่วน Plubin et al. (2025) เสนอโมเดลลูกผสม Transformer-RNN ซึ่งใช้ BERT ฝังบริบท และ RNN จับลำดับ ทำให้ประสิทธิภาพสูงกว่าโมเดลเดิมในบริบทของการธนาคารไทย ขณะที่Sungthong et al. (2024) เสนอแนวทางสำหรับข้อมูลขนาดเล็ก โดยผสาน TF-IDF, Word2Vec และ BERT variants ทำให้สามารถจำแนก issue ticket ได้อย่างมีประสิทธิภาพ ส่วน Suramanka และ Hanskunatai (2024) ใช้ Classifier Chains ร่วมกับ Random Forest วิเคราะห์ความคิดเห็นของนักท่องเที่ยวต่างชาติ และได้ Accuracy สูงถึง 80% ในด้านเทคนิค Liu et al. (2021) แนะนำให้ใช้โมเดลลูกผสม CNN-RNN เพื่อจับความสัมพันธ์ของป้ายกำกับ ขณะที่ Chen et al. (2023) เสนอการเลือกเลเยอร์ของ BERT แบบ Adaptive เพื่อเพิ่มความยืดหยุ่นของโมเดล

การสื่อสารอย่างมีประสิทธิภาพต้องพิจารณาทั้งเจตนาและโทนของผู้เขียน โดยเจตนาบ่งชี้เหตุผลในการสื่อสาร ส่วนโทนสะท้อนทัศนคติ (Hall & Wallace, 2018) โดยเฉพาะอารมณ์ (Emotion) ซึ่งมีผลต่อการดึงดูดความสนใจและการเข้าใจของผู้อ่าน Chae (2020) พบว่าเนื้อหาที่มีโทนทางอารมณ์ ข้อความยาว และมีแฮชแท็ก ช่วยเพิ่ม engagement ได้ แม้โทนจะเป็นลบ Yeramosu et al. (2023) วิเคราะห์โพสต์เกี่ยวกับการผ่าตัดกระดูกสันหลัง พบว่า Instagram มักมีโทนบวก ส่วน Twitter มักสะท้อนเรื่องส่วนตัวเชิงลบ Leelawat et al. (2022) แสดงให้เห็นว่าปัจจัยภายนอก เช่น ความขัดแย้งทางการเมือง มีผลต่ออารมณ์ในข้อความเกี่ยวกับการท่องเที่ยวไทยในช่วงโควิด-19 Matook et al. (2022) วิเคราะห์ปรากฏการณ์บนโซเชียลมีเดีย และพบว่าโทนข้อความมีผลต่อการเกิดกระแสสนับสนุนหรือวิพากษ์ ขณะที่ Kim และ Masullo Chen (2020) พบว่า ความคิดเห็นเชิงสุภาพในการแก้ข่าวปลอมบนโซเชียลมีเดียถูกมองว่าน่าเชื่อถือ แม้ไม่ส่งผลต่อความน่าเชื่อถือของข่าวหลัก Rozado et al. (2022) วิเคราะห์พาดหัวข่าวกว่า 23 ล้านชิ้นในสื่ออเมริกัน และพบว่าโทนข่าวมีแนวโน้มเป็นลบมากขึ้น โดยเฉพาะในสื่อฝ่ายขวาที่เน้นความโกรธ กลัว และเศร้า ซึ่งสะท้อนเจตนาบรรณาธิการที่แตกต่างในแต่ละสื่อ

จากการทบทวนวรรณกรรม พบว่าสี่มิติสำคัญที่เกี่ยวข้องกับการกำหนดลักษณะของข้อความได้แก่ 1) การจำแนกประเภท Timely/Timeless 2) การจำแนกเจตนา 3) การจำแนกอารมณ์ และ 4) สไตล์การเขียนเชิงแปลหรือเล่าเรื่อง แม้ว่าจะมีงานวิจัยในระดับสากลมากพอสมควร แต่ในบริบทของภาษาไทยยังคงมีจำนวนน้อย อย่างไรก็ตาม แนวโน้มของงานวิจัยด้านการจำแนกข้อความภาษาไทยยังคงเติบโตอย่างต่อเนื่อง ซึ่งชี้ให้เห็นถึงความจำเป็นในการพัฒนาแนวทางที่เหมาะสมกับบริบทภาษาไทยที่มีลักษณะเฉพาะด้านวัฒนธรรมและรูปแบบการใช้ภาษา

ทั้งนี้วิทยานิพนธ์นี้ได้สรุปงานวิจัยที่เกี่ยวข้องตามแนวทางของการสร้างโมเดลการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกสำหรับงานภาษาศาสตร์ไว้ในตารางที่ 2.3 ถึง 2.5 และสรุปตามจุดประสงค์ของแต่ละงานไว้ในตารางที่ 2.6 พร้อมทั้งระบุความเกี่ยวข้องกับวิทยานิพนธ์นี้ (Our work) ในแต่ละตาราง

ตารางที่ 0.3 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม CNN

งานวิจัย	ปี	CNN	CNN Skipgram	CNN fastText
Promrit and Waijanya	2017	✓		
Waijanya and	2017	✓		

งานวิจัย	ปี	CNN	CNN Skipgram	CNN fastText
Promrit				
Irwanto & Goeirmanto	2023	✓		
Vatathanavaro et al.	2020	✓		
Gadek & Guélorget	2020	✓		
Wang et al.	2020	✓		
Saedi & Dras	2021	✓		
Suramanka and Hanskunatai	2024	✓		
Liu et al.	2021	✓		
P. Li et al.	2021	✓		
WATJANAPRON	2024	✓		
Xu et al.	2019		✓	
Gatchalee et al.	2023	✓	✓	✓
Mestry et al.	2023			✓
Our work		✓	✓	✓

ตารางที่ 0.4 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม Ensemble learning

งานวิจัย	ปี	Random Forest	AdaBoost	XGBoost	LightGBM	SVM
Phann et al.	2023	✓	✓			
Taherkhani et al.	2023	✓				
Dewi et al.	2023	✓				
Wan et al.	2021	✓				
Sameen et al.	2016	✓				
Mehta et al.	2021	✓				
Halim et al.	2020	✓			✓	
Hou et al.	2023		✓			

งานวิจัย	ปี	Random Forest	AdaBoost	XGBoost	LightGBM	SVM
Putra et al.	2022			✓		
Han	2023			✓		
Irwanto & Goeirmanto	2023			✓		✓
Chandran et al.	2023				✓	
Zhang et al.	2023				✓	
Taha et al.	2024				✓	
Siriphanpornchana & Kanjanawattana	2023				✓	
Arreerard & Senivongse	2018				✓	
Our work		✓	✓	✓	✓	✓

ตารางที่ 0.5 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับโมเดลกลุ่ม Transformer

งานวิจัย	ปี	BERT	Transformer	ROBERTa	Wangchan BERTa	Phaya ThaiBERT
Moirangthem and Lee	2021					
Xiong et al.	2021					
Zheng et al.	2020	✓				
Liu et al.	2021	✓	✓			
Zhou et al.	2020	✓				
Saedi & Dras	2021	✓				

งานวิจัย	ปี	BERT	Transformer	ROBERTa	Wangchan BERTa	Phaya ThaiBERT
Chen et al.	2023	✓				
Gatchale e et al.	2023	✓		✓		
P. Li et al.	2021		✓			
L. Lowphan sirikul et al.	2021			✓		
Suraratch ai & Phoomv uthisarn	2024			✓		
Faengrit et al.	2024			✓		
Y. Liu et al.	2019		✓	✓		
Sungtho ng et al.	2024	✓		✓		✓
Sriwirote et al.	2023				✓	
WATJAN APRON	2024				✓	
Our work		✓	✓	✓	✓	✓

มหาวิทยาลัยศิลปากร

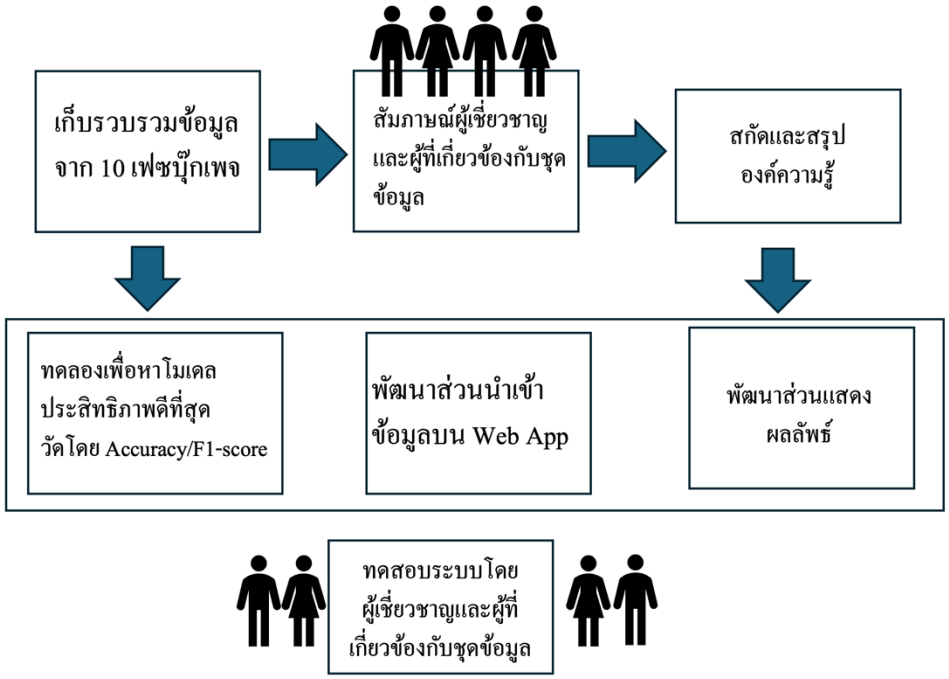
ตารางที่ 0.6 แสดงรายละเอียดงานวิจัยที่เกี่ยวข้องกับการประเมินความสอดคล้องเนื้อหา

งานวิจัย	Long text classification	Intention	Emotion	Style validation	Timeliness
Hadi et al. (2024)		✓			
Chopra and Verma (2023)		✓			
Yeramosu et al. (2023)			✓		
Leelawat et al. (2022)			✓		
G. Ríos-Toledo et al. (2022)				✓	
Matook et al. (2022)		✓			
R. Sabol and A. Horák (2022)				✓	
Rozado et al. (2022)			✓		
Xiong et al. (2021)	✓			✓	
Saedi & Dras (2021)	✓		✓		
P. Li et al. (2021)	✓	✓			
Liu et al. (2021)	✓	✓			
Moirangthem & Lee (2021)	✓				
Zhou et al. (2020)			✓		
Gadek & Guélorget (2020)			✓		
Wang et al. (2020)	✓	✓			
Zheng et al. (2020)	✓				
Qurashi et al. (2020)				✓	
Halim et al. (2020)			✓		
Chae (2020)			✓		

งานวิจัย	Long text classification	Intention	Emotion	Style validation	Timeliness
Fei et al. (2020)			✓		
Huang & Ren (2020)	✓	✓			
Amelin et al. (2018)	✓		✓		
Arreerard & Senivongse (2018)	✓				
Our work	✓	✓	✓	✓	✓

บทที่ 3
ขั้นตอนและวิธีดำเนินการ

3.1 ภาพรวมขั้นตอนการดำเนินการวิจัย



ภาพที่ 3.1 ภาพรวมขั้นตอนการดำเนินการวิจัย

ขั้นตอนการดำเนินงานวิจัยฉบับนี้ ดังนำเสนอในภาพที่ 3.1 ประกอบด้วย 3 ส่วนหลัก ได้แก่ 1) การเตรียมข้อมูลและวิเคราะห์ด้วยเนื้อหา 2) การพัฒนาและทดสอบระบบ และ 3) การประเมินผลจากผู้เชี่ยวชาญ โดยรายละเอียดในแต่ละขั้นตอนมีดังนี้

3.1.1 การเตรียมข้อมูลและสกัดองค์ความรู้จากผู้เชี่ยวชาญ

ขั้นตอนการเตรียมข้อมูลและสกัดองค์ความรู้จากผู้เชี่ยวชาญเป็นขั้นตอนแรกของการดำเนินการวิจัย ประกอบด้วย

1) การเก็บรวบรวมข้อมูลจาก 10 เพจเฟซบุ๊ก

ดำเนินการรวบรวมข้อมูลโพสต์จากเพจเฟซบุ๊กที่เผยแพร่เนื้อหาเกี่ยวกับประเทศจีน จำนวนทั้งสิ้น 10 เพจ เพื่อใช้เป็นฐานข้อมูลต้นทางในการวิเคราะห์ด้วยเนื้อหา

2) การสัมภาษณ์ผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับชุดข้อมูล

เพื่อให้เข้าใจบริบทของเนื้อหาได้อย่างถูกต้อง จึงมีการสัมภาษณ์ผู้เชี่ยวชาญและผู้มีประสบการณ์ในด้านการสื่อสารจีน-ไทย เพื่อเก็บข้อมูลเชิงคุณภาพประกอบการวิเคราะห์

3) การสกัดและสรุปองค์ความรู้

จากข้อมูลที่รวบรวมและความคิดเห็นจากผู้เชี่ยวชาญ จะมีการสังเคราะห์เพื่อจัดกลุ่มเนื้อหาทั้ง 4 รูปแบบ

3.1.2 การพัฒนาระบบ Web Application และโมเดลวิเคราะห์เนื้อหา

โดยถูกแบ่งเป็น 3 ส่วนหลัก

1) พัฒนาและทดลองฝึกโมเดลจำแนกเนื้อหาเพื่อหาโมเดลที่มีประสิทธิภาพดีที่สุดโดยวัดผลจากค่า Accuracy สำหรับโมเดลจำแนกหลายคลาส และ F1-score สำหรับโมเดล Multi-label

2) พัฒนาการนำเข้าข้อมูล Web Application และ API เชื่อมต่อกับโมเดล

3) พัฒนาส่วนแสดงผลลัพธ์การประเมินความสอดคล้องของบทความ

3.1.3 การทดสอบระบบโดยผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับข้อมูล

เพื่อประเมินความแม่นยำ ความเหมาะสมของผลลัพธ์ และการใช้งานจริงของระบบ Web Application ที่พัฒนาขึ้น โดยมีทั้งการให้คะแนนความพึงพอใจ ความเป็นไปได้ในการใช้งานจริง และการให้ข้อเสนอแนะสำหรับการพัฒนาต่อไป

3.2 การเตรียมข้อมูลและสกัดองค์ความรู้จากผู้เชี่ยวชาญ

วิทยานิพนธ์นี้รวบรวมข้อมูลจากเพจ Facebook จำนวน 10 เพจที่เน้นนำเสนอเนื้อหาที่เกี่ยวข้องกับประเทศจีนในรูปแบบภาษาไทย โดยมีจำนวนบทความรวมทั้งหมด 1,400 บทความ ซึ่งรวบรวมจากเพจที่มีเนื้อหาหลากหลายรูปแบบ เพื่อให้สามารถวิเคราะห์เนื้อหาได้อย่างครอบคลุมทั้งในแง่ของรายละเอียดและสไตล์การนำเสนอ โดยเนื้อหาที่รวบรวมมาส่วนใหญ่ประกอบด้วยเนื้อหาที่เกี่ยวข้องกับข่าวสาร สาระความรู้ ประวัติศาสตร์ การแปลเนื้อหาทางการ และการเล่าเรื่องที่มีลักษณะบอกเล่าประสบการณ์ โดยแพลตฟอร์มหลักในการนำเสนอเนื้อหาทั้งหมดคือ Facebook

3.2.1 ชุดข้อมูลจากเพจเฟซบุ๊กเนื้อหาจีนในภาษาไทย

ในชุดข้อมูลดังกล่าวมีการแบ่งประเภทเพจหลักๆ ออกเป็นสองรูปแบบ ได้แก่

3.2.1.1 เพจ Facebook ประเภทองค์กรสื่อ (Media Outlets)

ชุดข้อมูลประกอบด้วยเพจประเภทนี้จำนวน 5 เพจ โดยเนื้อหาส่วนใหญ่มุ่งเน้นการแปลและดัดแปลงเนื้อหาจากแหล่งที่เป็นทางการของจีนเพื่อนำเสนอในภาษาไทย โดยรักษาความถูกต้องและความเป็นทางการตามต้นฉบับ เนื้อหาจากเพจเหล่านี้จะมีลักษณะทางการ มีความถูกต้องและความ

ชัดเจนสูง เนื่องจากมาจากการแปลและเรียบเรียงจากแหล่งต้นทางที่มีความน่าเชื่อถือสูง เช่น สื่อจีนหรือหน่วยงานราชการจีน

1) China Xinhua News (ภาคภาษาไทย)

เป็นช่องทางสื่อสารอย่างเป็นทางการของ ซินหัว (Xinhua) สำนักข่าวของรัฐบาลกลางจีนในฉบับภาษาไทย เพจนี้นำเสนอข่าวที่มีลักษณะทางการสูง ครอบคลุมหัวข้อสำคัญทั้งด้านเศรษฐกิจ การเมือง วิทยาศาสตร์ นโยบายพัฒนา และโครงการระหว่างประเทศของจีน โดยเฉพาะนโยบายที่เกี่ยวข้องกับความร่วมมือไทย-จีนการดำเนินบทบาทเชิงนโยบายของเพจนี้ชัดเจนในฐานะสื่อกระบอกเสียงของรัฐจีนต่อผู้ชมชาวไทย

2) China Report ASEAN – Thailand

เพจภาคภาษาไทยของสื่อในเครือ China Report ASEAN สื่อประเทศจีน ซึ่งเป็นช่องทางสื่อสารเชิงยุทธศาสตร์ระหว่างจีนและประเทศในกลุ่มอาเซียน โดยเฉพาะไทย เนื้อหามักมาจากสำนักข่าวในจีน เช่น Xinhua หรือ CGTN แต่มีการปรับรูปแบบให้เข้ากับกลุ่มเป้าหมายคนไทยมากขึ้น โดยใช้สัดส่วนระหว่างข่าวจากจีนและเนื้อหาตามกระแส (viral content) ในสัดส่วนประมาณ 50:50 จุดเด่นของเพจนี้คือการนำเสนอข่าวสารที่เกี่ยวข้องกับภาพลักษณ์ความทันสมัยของจีน เช่น เมืองอัจฉริยะ หรือเทคโนโลยีใหม่ ซึ่งมักกลายเป็นไวรัลในไทยอย่างรวดเร็ว

3) Jeenthainews

Jeenthainews เป็นเพจที่แปลและเรียบเรียงข่าวจากสื่อของทางการจีนให้เป็นภาษาไทย โดยเน้นการสื่อสารเรื่องราวที่แสดงถึงความร่วมมือระหว่างไทยกับจีน ข่าวความสำเร็จทางเศรษฐกิจและบทวิเคราะห์เกี่ยวกับการพัฒนาของจีนในบริบทโลก การนำเสนอของเพจมีความเป็นทางการแต่ปรับให้อ่านง่าย เข้าถึงง่าย จึงเหมาะสำหรับผู้ที่สนใจติดตามแนวโน้มข่าวสารจากจีนในเชิงวิเคราะห์ โดยมีเป้าหมายชัดเจนในการส่งเสริมภาพลักษณ์และอิทธิพลเชิง soft power ของจีน

4) China Face

เป็นเพจภาคภาษาไทยของ China Radio International (CRI) ภายใต้เครือ China Media Group ซึ่งถือเป็นสื่อระดับชาติของจีน โดยเนื้อหาเน้นการแปลและเรียบเรียงข่าวจากแหล่งข่าวหลักของจีน โดยเฉพาะข่าวด้านการเมืองระหว่างประเทศ การทูต และนโยบายระดับรัฐ คิดเป็นสัดส่วนประมาณ 70% ของเนื้อหาทั้งหมด อีก 30% เป็นเนื้อหาเชิงสร้างภาพลักษณ์จีนในมุมสร้างสรรค์ เช่น เทคโนโลยี เมืองอัจฉริยะ และวัฒนธรรมยุคใหม่ จุดเด่นคือการใช้โทนภาพที่สวยงาม ภาษากระชับ และมีวิดีโอประกอบเพื่อสร้างอารมณ์เชิงบวกให้กับกลุ่มผู้ติดตามในไทย

5) People’s Daily Online Thai

เพจภาษาไทยของหนังสือพิมพ์ People’s Daily ซึ่งเป็นสื่อทางการหลักของพรรคคอมมิวนิสต์จีน เนื้อหาเน้นไปที่การถ่ายทอดนโยบายรัฐ ความก้าวหน้าทางวิทยาศาสตร์ เทคโนโลยี นวัตกรรม เศรษฐกิจ การทูต และกิจกรรมระหว่างประเทศของจีนในเชิงสถาบัน โดยเพจนี้มีลักษณะเป็นสื่อประชาสัมพันธ์เชิงโครงสร้าง โดยใช้ภาษาในลักษณะเป็นทางการสูง ตรงไปตรงมา และสอดคล้องกับจุดยืนของรัฐบาลจีนในการส่งเสริม soft power ผ่านช่องทางสื่อสารกับผู้ชมต่างประเทศ

3.2.1.2 เพจ Facebook ประเภทงานเขียนเชิงเล่าเรื่อง (Storytelling)

เพจที่อยู่ในกลุ่มนี้อีกจำนวน 5 เพจ โดยเพจเหล่านี้เน้นนำเสนอเนื้อหาในรูปแบบของการเล่าเรื่องที่มีความสร้างสรรค์และเป็นกันเอง มีการผสมผสานมุมมองส่วนตัวและอารมณ์ความรู้สึกของผู้เขียนเข้ามาในบทความ ส่งผลให้เนื้อหามีลักษณะเป็นกันเองและมีความน่าสนใจ ผู้เขียนเนื้อหาในเพจประเภทนี้มักใช้เทคนิคการเล่าเรื่องที่สร้างความสัมพันธ์กับกลุ่มผู้อ่านชาวไทย เพื่อให้เนื้อหาที่เกี่ยวข้องกับประเทศจีนเข้าถึงและสื่อสารกับกลุ่มเป้าหมายในประเทศไทยได้อย่างมีประสิทธิภาพ

1) เพจอ้ายจง

หนึ่งในสื่อออนไลน์ที่โดดเด่นด้านการนำเสนอเนื้อหาเกี่ยวกับประเทศจีนในบริบทภาษาไทย โดยเริ่มก่อตั้งเมื่อวันที่ 12 สิงหาคม พ.ศ. 2556 โดย นายภากร กัทชลี ขณะเป็นนักศึกษาระดับปริญญาโท ทุนรัฐบาลจีน ณ มหาวิทยาลัยเป่ยหัง (Beihang University) ประเทศจีน จุดประสงค์เริ่มต้นของเพจคือการเผยแพร่บทความและข่าวสารเกี่ยวกับประเทศจีนในรูปแบบที่เข้าใจง่ายสำหรับผู้ชมชาวไทย โดยลักษณะเฉพาะของเพจ คือ การนำเสนอเนื้อหาประเภท บทความเชิงลึก (long-form content) ซึ่งคิดเป็นสัดส่วนประมาณ 90% ของเนื้อหาทั้งหมด โดยบทความแต่ละชิ้นมีความยาวอย่างน้อย 250 ตัวอักษรขึ้นไป เพจไม่ได้เพียงรายงานข่าวสารทั่วไป แต่ยัง ตรวจสอบข้อเท็จจริง และนำเสนอเนื้อหาทางเลือกเกี่ยวกับจีนในมุมที่มักถูกสื่อไทยเข้าใจผิดหรือเสนออย่างไม่ถูกต้อง เช่น กรณี ทัวร์ศูนย์เหรียญ คนจีนแบนไทย และ เรือล่มที่จังหวัดภูเก็ต

เพจเริ่มได้รับความนิยมและการกล่าวถึงในวงกว้างในปี พ.ศ. 2559 และยังคงเป็นหนึ่งในเพจที่มีอิทธิพลด้านข่าวจีนในประเทศไทยอย่างต่อเนื่อง โดยปัจจุบัน (พ.ศ. 2567) มีผู้ติดตามมากกว่า 280,000 ราย และมีสถิติการเข้าถึงและการมีส่วนร่วมเฉลี่ยสูงถึง 1,000,000 ครั้งต่อเดือน และในช่วงวิกฤตการแพร่ระบาดของ COVID-19 ที่เริ่มต้นในจีนเมื่อปลายปี พ.ศ. 2562 เพจ อ้ายจง เป็นหนึ่งในสื่อแรก ๆ ที่รายงานสถานการณ์จากจีนอย่างใกล้ชิดและแม่นยำ ส่งผลให้ได้รับการเสนอชื่อเข้าชิง

รางวัล COVID-19 Content Creator จากสื่อ RAINMAKER ซึ่งมอบให้แก่ผู้ผลิตเนื้อหาออนไลน์ในสาขาต่าง ๆ ที่มีผลกระทบในวงกว้าง

อ้ายจง ถูกใช้เป็น ต้นแบบของแหล่งข้อมูล (corpus) สำหรับการวิจัยด้านการวิเคราะห์เนื้อหาเกี่ยวกับประเทศจีนในประเทศไทย โดยมีการรวบรวมข้อมูลจากเพจและสื่อออนไลน์รวม 10 แหล่งบนแพลตฟอร์ม Facebook ครอบคลุมช่วงเวลา ระหว่างปี พ.ศ. 2559–2567 เพื่อวิเคราะห์แนวโน้ม ประเภทเนื้อหา และบริบทของสื่อสารเชิงจีนในสังคมไทยอย่างเป็นระบบ

2) China Story

China Story เป็นเพจที่มีบทบาทเป็นช่องทางประชาสัมพันธ์ที่เน้นการนำเสนอเรื่องราวของประเทศจีนในเชิงบวก ด้วยเนื้อหาที่หลากหลาย เช่น ความสำเร็จทางเศรษฐกิจ เทคโนโลยี นโยบายด้านการศึกษา และวัฒนธรรมร่วมสมัย จุดมุ่งหมายหลักคือการสื่อสารเพื่อส่งเสริมภาพลักษณ์ของจีนต่อสาธารณชนชาวไทย โดยไม่จำกัดอยู่ที่การถ่ายทอดจากสื่อของรัฐเท่านั้น แต่ยังเลือกใช้เนื้อหาที่สอดคล้องกับความสนใจของกลุ่มเป้าหมายในไทย เช่น การท่องเที่ยวหรือวัฒนธรรมอาหาร เพจนี้จึงทำหน้าที่เป็นสื่อกลางในการแลกเปลี่ยนวัฒนธรรมเชิงประชาชนต่อประชาชน (people-to-people communication)

3) Nihaosawadee

เพจนี้เน้นการนำเสนอขาวสารจากจีนควบคู่กับเนื้อหาเชิงสร้างสรรค์ โดยจัดสัดส่วนระหว่างข่าวรายวัน (ประมาณ 40%) กับเนื้อหาที่สร้างความประทับใจหรือมุมมองเชิงบวกเกี่ยวกับจีน (ประมาณ 60%) อย่างลงตัว จุดเด่นของเพจคือการใช้ภาพถ่าย คลิปวิดีโอ และกราฟิกที่ช่วยสร้างความรู้สึกเชิงบวกต่อประเทศจีน โดยเฉพาะการนำเสนอเทคโนโลยี นวัตกรรม และโครงการพัฒนาในจีนที่มีความแตกต่างจากไทยอย่างชัดเจน จึงเป็นแหล่งข้อมูลที่สะท้อน soft power ในเชิงการสร้างภาพลักษณ์ที่น่าเชื่อถือของจีนต่อชาวไทย

4) ไทยคำจีนคำ

เพจไทยคำจีนคำ มีลักษณะเฉพาะที่เน้นการเชื่อมโยงเนื้อหาเกี่ยวกับประเทศจีนกับบริบทของสังคมไทย ผ่านการนำเสนอในรูปแบบที่ไม่จำกัดเฉพาะข่าวสาร แต่ครอบคลุมไปถึงเรื่องราวในเชิงความรู้และวัฒนธรรม เช่น ประวัติศาสตร์จีน บทสัมภาษณ์นักธุรกิจและผู้เชี่ยวชาญที่เกี่ยวข้องกับจีน รวมถึงประเด็นด้านไลฟ์สไตล์ที่สัมพันธ์กับทั้งจีนและไทย จุดเด่นของเพจนี้อยู่ที่ความสามารถในการสื่อสารเนื้อหาให้คงความสดใหม่เสมอ ทำให้เหมาะสำหรับผู้ติดตามที่ต้องการเรียนรู้มุมมองของจีนในเชิงลึก โดยไม่จำกัดเฉพาะประเด็นการเมืองหรือข่าวรายวันเท่านั้น

5) DomJeen ด้อมจีน ติ่งจีน ซีรี่ย์จีน ศิลปินจีน รักทุกอย่างที่เป็นจีน

เพจนี้เป็นตัวอย่างของการสื่อสารแบบกลุ่มแฟนคลับ (fan-based media communication) ที่สะท้อนถึงพลังของวัฒนธรรมป๊อปจีน (C-pop) ในสังคมไทย โดยมีการรวบรวมเนื้อหาด้านบันเทิงจีน ทั้งซีรีส์ ศิลปิน เพลง ข่าววงการบันเทิง และกิจกรรมของศิลปินที่ได้รับความนิยมในไทย ลักษณะการสื่อสารของเพจนี้เป็นไปในเชิงบันเทิง ใช้ภาษาที่เป็นกันเอง มีอารมณ์ร่วมสูง และอัปเดตอย่างรวดเร็ว เพจนี้จึงเหมาะสำหรับศึกษาบริบทของ soft power จากจีนผ่านวัฒนธรรมร่วมสมัย

3.2.2 การแบ่งชุดข้อมูล

สำหรับการแบ่งชุดข้อมูลที่ใช้ในการวิจัย เป็นแบบ Training, Validation และ Testing โดยชุดข้อมูลทั้งหมด 1,400 บทความถูกแบ่งในอัตราส่วน 70% สำหรับการฝึกโมเดล, 30% สำหรับการตรวจสอบความถูกต้อง (Validation) และชุดทดสอบ (Testing) แยกออกจากชุดข้อมูลสำหรับการฝึกอย่างชัดเจน เพื่อให้มั่นใจว่าการประเมินประสิทธิภาพของโมเดลจะสะท้อนถึงความสามารถที่แท้จริงในการจัดประเภทเนื้อหาที่ไม่เคยพบเห็นมาก่อนอย่างแท้จริง

3.2.3 การสัมภาษณ์และสกัดองค์ความรู้จากผู้เชี่ยวชาญ

เพื่อให้มั่นใจว่าการกำหนดป้ายกำกับ (Label) ในวิทยานิพนธ์นี้มีความเหมาะสมและสอดคล้องกับบริบททางวัฒนธรรมและการใช้ภาษาในเนื้อหาที่เกี่ยวข้องกับประเทศจีนที่เผยแพร่บนแพลตฟอร์ม Facebook ในภาษาไทย ทางวิทยานิพนธ์นี้ได้ดำเนินการสัมภาษณ์เชิงกึ่งโครงสร้าง (Semi-structured Interviews) กับผู้มีส่วนเกี่ยวข้องโดยตรงในด้านการผลิตและการวางแผนเนื้อหา จำนวน 4 ท่าน ประกอบด้วยเจ้าของสื่อออนไลน์ (Media Owner) บล็อกเกอร์ (Blogger) นักวางแผนสื่อ (Media Planner) และนักเขียนเนื้อหา (Content Writer) ที่มีประสบการณ์การทำงานเกี่ยวข้องโดยตรงกับเนื้อหาภาษาไทยที่เกี่ยวข้องกับจีน

กระบวนการสัมภาษณ์เชิงกึ่งโครงสร้างประกอบด้วยการพูดคุยแลกเปลี่ยนในหัวข้อสำคัญดังต่อไปนี้

3.2.3.1 กระบวนการสัมภาษณ์

ผู้มีส่วนร่วมแต่ละท่านได้ให้ข้อมูลผ่านการสัมภาษณ์แบบตัวต่อตัวโดยใช้เวลาประมาณหนึ่งชั่วโมง เพื่อสอบถามถึงประเด็นต่างๆ ได้แก่

1) การกำหนดวัตถุประสงค์หลักในการเผยแพร่เนื้อหาที่เกี่ยวข้องกับประเทศจีนบน Facebook ภาษาไทย

2) การอธิบายโทนอารมณ์ที่นิยมใช้ในการนำเสนอเนื้อหาดังกล่าว

3) ความท้าทายในการจำแนกและจัดหมวดหมู่เนื้อหาเกี่ยวกับจุดประสงค์และอารมณ์ที่เหมาะสมกับบริบทของไทย

4) การแนะนำหมวดหมู่จุดประสงค์และอารมณ์ที่เหมาะสม เพื่อให้สะท้อนถึงลักษณะและเจตนาของเนื้อหา

3.2.3.2 การวิเคราะห์ข้อมูล

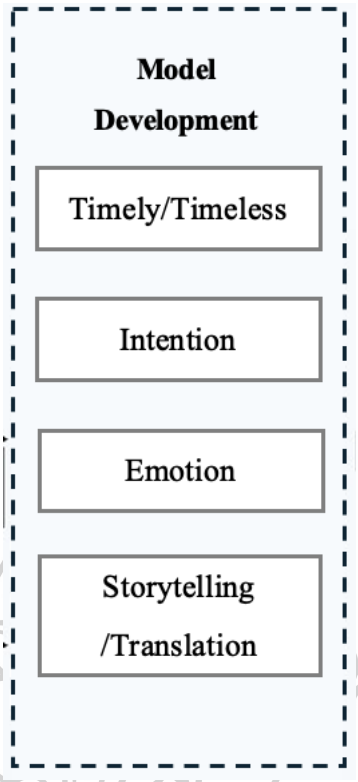
ข้อมูลที่ได้จากการสัมภาษณ์ถูกนำไปวิเคราะห์โดยวิธีการวิเคราะห์ Thematic Analysis เพื่อค้นหาประเด็นหลักที่เกิดขึ้นซ้ำๆ จากผู้ให้สัมภาษณ์

3.2.3.3 ผลลัพธ์ด้านองค์ความรู้

การสัมภาษณ์ผู้เชี่ยวชาญดังกล่าวนำไปสู่การกำหนดป้ายกำกับที่มีความแม่นยำและเหมาะสมทั้งในด้านวัฒนธรรมและภาษาสำหรับเนื้อหาในชุดข้อมูลรวมถึงการวิเคราะห์เทคนิคที่เหมาะสมรวมถึงสนับสนุนการใช้การจำแนกแบบหลายป้ายกำกับ เพื่อให้สามารถจับวัตถุประสงค์และอารมณ์ที่มีความซับซ้อนและผสมผสานกันภายในเนื้อหาเดียวกันได้อย่างครอบคลุมและมีประสิทธิภาพซึ่งสอดคล้องกับมุมมองร่วมกันของผู้เชี่ยวชาญเกี่ยวกับความหลากหลายของจุดประสงค์และอารมณ์ของผู้เขียน

3.3 การพัฒนาและทดลองหาโมเดลที่มีประสิทธิภาพดีที่สุด

สำหรับส่วนแบบจำลองการเรียนรู้ของเครื่องในเฟรมเวิร์กที่พัฒนาขึ้น ประกอบไปด้วย 4 รูปแบบการประเมินเนื้อหา ได้แก่ โมเดล จำแนก Timely และ Timeless โมเดลจำแนก Intention โมเดลจำแนกอารมณ์ (Emotion analysis) และโมเดลจำแนกเนื้อหาการแปลและการเล่าเรื่อง (Translation vs Storytelling) ดังภาพที่ 3.2 โดยจะทดลองเพื่อหาโมเดลที่เหมาะสมกับการจำแนกเนื้อหาทั้ง 4 รูปแบบ โดยการวัดประสิทธิภาพตามมาตรวัดที่กำหนดในแต่ละการทดลอง



ภาพที่ 3.2 โมเดลจำแนกเนื้อหาทั้งสี่รูปแบบ

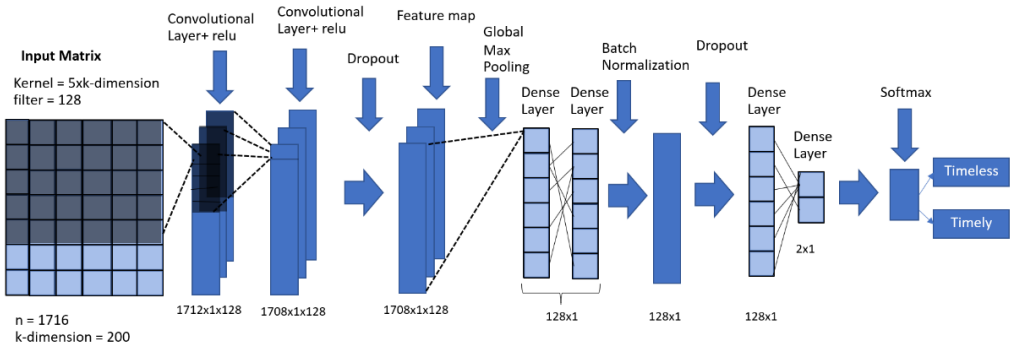
โดยรายละเอียดของการพัฒนาโมเดลมีรายละเอียดดังต่อไปนี้

3.3.1 โมเดลเทคนิคการเรียนรู้ของเครื่องและโมเดลเทคนิคการเรียนรู้เชิงลึก

ในวิทยานิพนธ์ฉบับนี้ ได้กำหนดโมเดลที่ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) สำหรับการประเมินเนื้อหาทั้ง 4 รูปแบบที่กำหนดไว้ โดยในแต่ละรูปแบบจะประกอบด้วยเทคนิคที่มีทั้งลักษณะที่เหมือนกันและแตกต่างกันบางส่วน ทั้งนี้เพื่อใช้ในการทดลองเปรียบเทียบประสิทธิภาพของแต่ละเทคนิค และคัดเลือกโมเดลที่ให้ผลลัพธ์ดีที่สุดในบริบทของแต่ละรูปแบบการประเมินเนื้อหา

3.3.1.1 CNN Conv1D

CNN แบบ Conv1D โดยมีโครงสร้างตามภาพที่ 3.3 ซึ่ง CNN นั้นเป็นเครือข่ายประสาทเทียมที่ได้รับแรงบันดาลใจจากสมองมนุษย์ มีความสามารถในการเรียนรู้รูปแบบจากข้อมูลข้อความได้ดี



ภาพที่ 3.3 สถาปัตยกรรม Convolutional Neural Network Con1D

ข้อมูลที่ใช้ในการป้อนเข้าสู่โมเดล CNN คือข้อความที่ผ่านการทำความสะอาดข้อมูล (cleaned text) แล้ว จากนั้นนำข้อมูลดังกล่าวเข้าสู่ CNN โดยแต่ละชั้นประกอบด้วย filter จำนวน 128 ตัวที่มีขนาด kernel เท่ากับ 5 จำนวนสองชั้น และเพิ่ม dropout ที่ระดับ 0.5 หลังชั้นที่สอง ก่อนที่จะใช้ Global Max Pooling ตามด้วย Dense layer เพื่อลดมิติข้อมูลและลดการ overfitting โมเดลสุดท้ายใช้ Softmax ในการจำแนกคลาสข้อความ และใช้ Adam optimizer เป็นตัวปรับแต่งน้ำหนัก

3.3.1.2 CNN Conv1D ร่วมกับ Skip-gram Classification Model

โมเดลนี้มีสถาปัตยกรรมเหมือน CNN Conv1D ดั้งเดิม แต่มีการปรับเปลี่ยนที่ Embedding Matrix โดยใช้ Skip-gram (Word2Vec) จากชุดข้อมูลที่เตรียมไว้เองเป็น Embedding Matrix โดยโมเดลจะใช้การสกัดคุณลักษณะเชิงลำดับจากลำดับของเวกเตอร์คำผ่านกระบวนการ convolution และ pooling ก่อนนำไปเข้าสู่ชั้น dense เพื่อจำแนกประเภท สำหรับจุดเด่นของโมเดลนี้คือการใช้ Embedding Matrix ที่ถูกฝึกจากชุดข้อมูลเฉพาะที่ผู้วิจัยเตรียมไว้เอง โดยใช้เทคนิค Skip-gram จาก Word2Vec ซึ่งเป็นวิธีการฝึกเวกเตอร์คำที่เรียนรู้จากบริบทของคำ (context-based learning) โดยพยายามทำนายคำรอบข้างจากคำเป้าหมาย (target word) เทคนิคนี้ช่วยให้เวกเตอร์ที่ได้สามารถเก็บความสัมพันธ์เชิงความหมายของคำในลักษณะ distributional semantics ได้ดี และมีความเหมาะสมกับลักษณะของข้อมูลที่เฉพาะเจาะจงในโดเมนของวิทยานิพนธ์นี้

3.3.1.3 CNN Conv1D ร่วมกับ FastText Skip-gram Classification Model

โมเดลนี้มีโครงสร้างโดยรวมเหมือนกับโมเดล CNN Conv1D ที่ใช้เวกเตอร์คำจากการฝึกแบบ Skip-gram ด้วย Word2Vec แต่มีการปรับเปลี่ยนในส่วนของ Embedding Layer โดยเลือกใช้เวกเตอร์คำแบบฝึกไว้ล่วงหน้า (pre-trained word embeddings) จาก FastText ซึ่งพัฒนาโดย Facebook AI Research (FAIR)

แม้ว่า FastText จะใช้หลักการฝึกแบบ Skip-gram เช่นเดียวกับ Word2Vec แต่มีข้อได้เปรียบเพิ่มเติม คือการฝึกจาก subword units (n-grams) แทนการใช้คำทั้งคำเพียงอย่างเดียว ซึ่งทำให้โมเดลสามารถสร้างเวกเตอร์แทนคำได้แม้ในกรณีที่คำเหล่านั้นไม่เคยปรากฏในชุดข้อมูลฝึกมาก่อน (out-of-vocabulary: OOV) จึงช่วยให้โมเดลมีความสามารถในการจัดการกับคำศัพท์ใหม่ คำสะกดผิด หรือคำเฉพาะทางที่ไม่อยู่ในพจนานุกรมมาตรฐานได้ดียิ่งขึ้น โดยการเลือกใช้เวกเตอร์จาก FastText เหมาะสมอย่างยิ่งในกรณีที่ชุดข้อมูลมีความ หลากหลายทางภาษา เช่น การผสมผสานระหว่างคำศัพท์ทั่วไปในภาษาไทยกับคำเฉพาะทางในบริบทเฉพาะ เช่น โดเมนภาษาจีนหรือศัพท์เฉพาะในแวดวงใดแวดวงหนึ่ง ทั้งนี้ เพื่อให้โมเดลสามารถเข้าใจบริบทและความหมายของคำได้ครอบคลุมยิ่งขึ้น

3.3.1.4 BERT Multilingual (mBERT)

BERT เป็นโมเดลประเภทTransformer ที่ Google พัฒนาและนิยมใช้ในการประมวลผลข้อความ มีการใช้เทคนิค Masked Language Modeling (MLM) เพื่อช่วยเรียนรู้โครงสร้างของภาษา มีข้อดีในการจับบริบทของข้อความอย่างมีประสิทธิภาพ สำหรับงานนี้ ใช้ BERT pre-trained และนำ token ที่ได้จากการประมวลผลไปใช้ในการจำแนกข้อความด้วย Softmax Classifier

โดย mBERT เป็นโมเดล BERT สำหรับหลายภาษา ซึ่งใช้วิธีการแบ่งคำแบบ WordPiece Tokenization โดยวิธีนี้จะตัดคำออกเป็นหน่วยย่อยที่เรียกว่า subword โดยอาศัย subword vocabulary ที่ได้จากการฝึกฝนกับชุดข้อมูลขนาดใหญ่ การใช้ subword ทำให้สามารถจัดการกับคำที่ไม่เคยปรากฏมาก่อนได้อย่างมีประสิทธิภาพ สำหรับภาษาไทย ในการศึกษานี้วิทยานิพนธ์นี้เลือกใช้โมเดลที่ได้รับการฝึกไว้ล่วงหน้า (pretrained) คือ bert-base-multilingual-cased จาก HuggingFace ซึ่งถูกฝึกด้วยข้อมูลจาก 104 ภาษา รวมถึงภาษาไทย โดยใช้ชุดข้อมูลจาก Wikipedia เป็นหลัก

3.3.1.5 XLM-RoBERTa

โมเดลนี้พัฒนาต่อจาก BERT แต่มีความแตกต่างในกระบวนการ tokenization และการ map token-index โดยใช้เทคนิค SentencePiece Tokenization ซึ่งมีข้อดีคือไม่ต้องทำ pre-tokenization และมีประสิทธิภาพสูงกว่า BERT ในการจัดการคำที่มีลักษณะซับซ้อน

XLM-RoBERTa ใช้ tokenizer แบบ SentencePiece Tokenizer ซึ่งแตกต่างจาก RoBERTa ดั้งเดิมที่ใช้วิธี Byte-Pair Encoding (BPE) ในการตัดคำ โดย BPE เป็นอัลกอริธึมที่ใช้สร้างหน่วย subword จากการจับคู่อักขระที่ปรากฏบ่อยที่สุดในชุดข้อมูล แล้วรวมกันเป็นคำใหม่อย่าง

ต่อเนื่องจนได้ชุด subword vocabulary ที่เหมาะสม วิธีการนี้ช่วยลดจำนวนคำในพจนานุกรมและสามารถจัดการกับคำใหม่ได้ในระดับหนึ่ง อย่างไรก็ตาม BPE มักอิงกับโครงสร้างของภาษาที่มีการเว้นวรรคอย่างชัดเจน เช่น ภาษาอังกฤษ

ในทางกลับกัน SentencePiece ที่ใช้ใน XLM-RoBERTa ได้รับการออกแบบให้สามารถทำงานกับภาษาที่ไม่มีการเว้นวรรคระหว่างคำ เช่น ภาษาไทย จีน หรือญี่ปุ่น ได้อย่างมีประสิทธิภาพ โดยไม่จำเป็นต้องมี tokenizer แยกต่างหากในแต่ละภาษา XLM-RoBERTa จึงสามารถรองรับการประมวลผลหลายภาษา (multilingual) ได้ภายในโมเดลเดียว

3.3.1.6 WangchanBERTa

สำหรับการใช้งานโมเดล Transformer ทั้งสาม (BERT, RoBERTa และ WangchanBERTa) วิทยานิพนธ์นี้ใช้ชุดไลบรารี Hugging Face Transformers เนื่องจากมีโมเดล pre-trained models หลากหลายพร้อมใช้งาน และรองรับงานจำแนกข้อความที่มีประสิทธิภาพสูง โดยเฉพาะสำหรับภาษาไทย

โมเดลนี้ใช้ SentencePiece Tokenizer เช่นเดียวกับ XLM-RoBERTa แต่มีการปรับแต่งเพิ่มเติมให้เหมาะกับลักษณะของภาษาไทยโดยเฉพาะ

สิ่งที่โดดเด่นใน WangchanBERTa คือการเพิ่มการจัดการช่องว่าง (spacing) ผ่านการใช้ token พิเศษ <_> เพื่อแทนตำแหน่งของช่องว่างระหว่างคำ เนื่องจากภาษาไทยมีลักษณะเฉพาะในการใช้ช่องว่าง โดยอาจใช้เพื่อแยกวลี หรือเพื่อเน้นความหมายในบางกรณี การใช้ token พิเศษนี้ช่วยให้โมเดลเข้าใจโครงสร้างของประโยคภาษาไทยได้ดีขึ้น (Lowphansirikul et al., 2021)

3.3.1.7 PhayathaiBERT

PhayathaiBERT เป็นโมเดลที่ได้รับการพัฒนาต่อยอดจาก WangchanBERTa โดยมีการเพิ่มขีดความสามารถในการรับรู้คำศัพท์ที่หลากหลายยิ่งขึ้น โดยเฉพาะอย่างยิ่งคำยืมที่ยังไม่ถูกกลมกลืนในภาษาไทย (unassimilated loanwords) อันเนื่องมาจากการฝึกด้วยคลังข้อมูลภาษาไทย (corpus) ที่มีขนาดใหญ่และหลากหลายกว่าเดิม ส่งผลให้โมเดลสามารถจัดการกับข้อความที่มีความหลากหลายทางภาษาได้อย่างมีประสิทธิภาพมากขึ้น เช่น ข้อความในสื่อสังคมออนไลน์ หรือบทความที่ใช้รูปแบบภาษาที่แตกต่างจากมาตรฐาน (Sriwirote et al., 2023)

ในส่วนของการตัดคำ (tokenization) PhayathaiBERT ยังคงใช้แนวทางเดียวกับ WangchanBERTa โดยอาศัย SentencePiece Tokenizer เป็นหลัก อย่างไรก็ตาม มีการปรับแต่ง tokenizer ดังกล่าวให้มีขอบเขตครอบคลุมที่กว้างขึ้น ด้วยการสร้าง expanded tokenizer ซึ่งเป็น

การขยายขนาดของ subword vocabulary เพื่อรองรับคำศัพท์ใหม่ ๆ หรือคำเฉพาะทางที่ปรากฏในบริบทของภาษาไทยสมัยใหม่ โดยยังคงลักษณะการจัดการช่องว่าง (spacing-aware) และการใช้ token พิเศษ <_> เพื่อแทนการเว้นวรรคในภาษาไทยไว้เช่นเดิม

3.3.1.8 Random Forest

Random Forest เป็นเทคนิคการเรียนรู้ของเครื่องแบบ ensemble ที่พัฒนาโดยการรวมกลุ่มของ decision trees เข้าด้วยกันโดยใช้วิธีการ bagging (Bootstrap Aggregating) ซึ่งเป็นการสุ่มตัวอย่างข้อมูลจำนวนหนึ่ง (พร้อมการทดแทน) เพื่อนำไปฝึกต้นไม้แต่ละต้นแบบเป็นอิสระต่อกัน โดยหลักการของ Random Forest คือ การสร้างต้นไม้หลายต้นและใช้การ vote แบบเสียงข้างมาก (majority voting) หรือการเฉลี่ยผลลัพธ์ (ในกรณี regression) เพื่อเพิ่มความแม่นยำ ลดความแปรปรวน (variance) และลดความเสี่ยงของ overfitting ที่มักเกิดใน decision tree เดี่ยว

3.3.1.9 XGBoost

XGBoost เป็นเทคนิคการเรียนรู้ในกลุ่ม ensemble แบบ gradient boosting ที่ได้รับการพัฒนาขึ้นเพื่อเพิ่มประสิทธิภาพทั้งด้านความเร็วและความแม่นยำ โดยใช้แนวคิดของการสร้างต้นไม้แบบต่อเนื่อง (sequential trees) ที่แต่ละต้นเรียนรู้จากข้อผิดพลาดของต้นก่อนหน้า โดยใช้regularization แบบ L1 และ L2 เพื่อป้องกัน overfitting และมีปรับปรุงการจัดการข้อมูลที่ขาดหาย หรือ missing values รวมทั้งมีการควบคุมความลึกของต้นไม้ (tree pruning) และการขนานข้อมูล (parallelization) เพื่อให้สามารถประมวลผลได้เร็วขึ้น

3.3.1.10 LightGBM

LightGBM เป็นเทคนิคการเรียนรู้ของเครื่องที่อยู่ในกลุ่มของ Ensemble Learning โดยเฉพาะในรูปแบบของ Gradient Boosting เช่นเดียวกับ XGBoost ซึ่งเป็นการรวมโมเดลย่อยหลายตัว (weak learners) เข้าด้วยกันแบบลำดับ (sequential) โดยที่แต่ละโมเดลจะถูกฝึกให้เรียนรู้จากข้อผิดพลาดของโมเดลก่อนหน้า เพื่อเพิ่มประสิทธิภาพในการทำนายผลลัพธ์ให้แม่นยำยิ่งขึ้นในทุก ๆ รอบการเรียนรู้

3.3.1.11 AdaBoost

AdaBoost (Adaptive Boosting) เป็นเทคนิคการเรียนรู้ของเครื่องที่อยู่ในกลุ่มของ Ensemble Learning แบบ Boosting โดยมีแนวคิดหลักในการผสานแบบจำลองย่อยที่มีประสิทธิภาพต่ำ (weak learners) หลายตัวเข้าด้วยกัน เพื่อสร้างแบบจำลองที่มีความแม่นยำสูงขึ้น (strong learner) โดยกระบวนการของ AdaBoost จะดำเนินการในลักษณะเชิงลำดับ (sequential

learning) โดยในแต่ละรอบการฝึกแบบจำลองจะให้น้ำหนักมากขึ้นกับตัวอย่างข้อมูลที่ถูกจำแนกผิดในรอบก่อนหน้า ทำให้แบบจำลองย่อยในรอบถัดไปสามารถ โฟกัส ไปยังตัวอย่างที่ยากในการเรียนรู้มากขึ้น ส่งผลให้ระบบสามารถแก้ไขข้อผิดพลาดอย่างต่อเนื่อง

เมื่อจบกระบวนการฝึก AdaBoost จะรวมผลลัพธ์ของ weak learners ทั้งหมดเข้าด้วยกัน โดยใช้วิธีการลงคะแนนแบบถ่วงน้ำหนัก (weighted voting) เพื่อให้ได้ผลลัพธ์สุดท้ายที่มีความแม่นยำและความสามารถในการจำแนกที่สูงกว่าแบบจำลองเดี่ยวอย่างมีนัยสำคัญ

ทั้งนี้ โมเดลในกลุ่ม Ensemble Learning ใช้ word_tokenize จากไลบรารี PyThaiNLP เป็น tokenizer ภาษาไทย

3.3.2 ออกแบบการทดลองเพื่อหาโมเดลที่มีประสิทธิภาพดีที่สุด

จากเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกทั้ง 11 โมเดลที่ได้ศึกษามาในวิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้ออกแบบการทดลองโดยคัดเลือกโมเดลแต่ละประเภทเพื่อนำมาเปรียบเทียบประสิทธิภาพการทำงานภายใต้บริบทของการประเมินความสอดคล้องของเนื้อหาทั้ง 4 รูปแบบที่กำหนดไว้ ทั้งนี้ เพื่อวัตถุประสงค์ในการค้นหาโมเดลที่ให้ผลการจำแนกที่มีประสิทธิภาพดีที่สุดและเหมาะสมต่อการประยุกต์ใช้กับชุดข้อมูล โดยมีรายละเอียดของโมเดลและแนวทางการประเมินดังต่อไปนี้

3.3.2.1 การทดลองที่ 1 โมเดลจำแนก Timely Timeless

การออกแบบทดลองสำหรับการทดลองที่ 1 เลือกใช้ โมเดล 5 รูปแบบสำหรับทดลองหาโมเดลที่มีประสิทธิภาพที่สุดสำหรับการจำแนก Timely และ Timeless ได้แก่

1) CNN 1D

2) CNN Conv1D ร่วมกับ Skip-gram Classification Model

3) CNN Conv1D ร่วมกับ FastText Skip-gram Classification Model

4) mBERT

5) XML-RoBERTa

6) WangchanBERTa

โดยในการกำหนดจำนวนรอบการฝึก (epochs) สำหรับโมเดลประเภท Convolutional Neural Network (CNN) ได้ฝึกระหว่าง 10,000 ถึง 20,000 epochs ขณะที่โมเดลประเภท Transformer ได้ฝึกเพียง 10 epochs ทั้งนี้เพื่อหลีกเลี่ยงปัญหา overfitting จากการฝึกมากเกินไป (overtraining) ระบบจึงมีการบันทึก checkpoint ของโมเดล ที่ให้ค่าความสูญเสียบนชุดข้อมูล

ตรวจสอบ (validation loss) ต่ำที่สุดไว้ จากนั้น เมื่อทดสอบกับชุดข้อมูลทดสอบ (test set) ระบบจะโหลดเฉพาะโมเดลที่มีประสิทธิภาพดีที่สุดในรอบการฝึกนั้นมาใช้

สำหรับ ค่า learning rate ของทั้งโมเดลในกลุ่ม CNN และ Transformer ได้ถูกกำหนดไว้ที่ 0.00005 โดยอิงจากการใช้งานฟังก์ชัน lr_find จากไลบรารี Ktrain ซึ่งช่วยให้สามารถเลือกค่า learning rate ที่เหมาะสมกับการฝึกได้อย่างอัตโนมัติ

นอกจากนี้ ยังมีการกำหนดค่า batch size ที่แตกต่างกันสำหรับแต่ละกลุ่มโมเดล โดยใช้ค่า 32 สำหรับโมเดล CNN และโมเดลกลุ่ม CNN variants และใช้ค่า 6 สำหรับโมเดล Transformer และกลุ่ม Transformer variants

3.3.2.2 การทดลองที่ 2 โมเดลจำแนก Intention

การออกแบบทดลองสำหรับการทดลองที่ 2 ใช้กลุ่มโมเดล Transformer ซึ่งให้ประสิทธิภาพสูงในการจำแนกเนื้อหาข้อความยาว ซึ่งช่วยแก้ไขข้อจำกัดของงานวิจัยที่ส่วนใหญ่เน้นแต่ข้อความที่มีความยาวสั้นเท่านั้น โดยในวิทยานิพนธ์นี้จึงนำโมเดล Transformer-based จำนวน 3 โมเดล ได้แก่ mBERT, WangchanBERTa และ PhayaThaiBERT มาประยุกต์ใช้กับงานจำแนก Intention ของข้อความภาษาไทย

เริ่มจากการกำหนดคลาส Intention จากในขั้นตอนการสัมภาษณ์และการถอดองค์ความรู้จากผู้เชี่ยวชาญ พบว่าการกำหนดจุดประสงค์ของเนื้อหามีความแตกต่างกันไปตามแต่ละผู้เขียนหรือแต่ละสื่อ อย่างไรก็ตาม วิทยานิพนธ์นี้พบว่าการระบุจุดประสงค์ของเนื้อหาที่มีความเฉพาะเจาะจงนั้นล้วนมีจุดร่วมที่สำคัญประการหนึ่ง คือ เป็นการนำเสนอเจตนาหรือจุดประสงค์ของผู้ส่งสารหรือผู้สร้างเนื้อหาโดยตรง ดังนั้น วิทยานิพนธ์นี้จึงได้เลือกใช้กรอบแนวคิดจากทางสายภาษาศาสตร์และการสื่อสารที่สอดคล้องกับข้อเสนอแนะของผู้เชี่ยวชาญ ได้แก่ PIE Framework (Persuade, Inform, Entertain) ซึ่งเน้นหนักที่จุดประสงค์ด้านการ Persuade (โน้มน้าว) และ Inform (แจ้งข้อมูล) สำหรับเนื้อหาข่าวและสื่อหลัก ขณะที่มีการให้ความสำคัญเพิ่มเติมกับจุดประสงค์ด้าน Entertain (บันเทิง) สำหรับสื่อรายบุคคล เช่น บล็อกเกอร์หรืออินฟลูเอนเซอร์

นอกจากนี้ ยังค้นพบอีกว่าอารมณ์มีส่วนเกี่ยวข้องโดยตรงกับจุดประสงค์ของเนื้อหาหรือการส่งสารจากผู้สร้างเนื้อหา จากการรวบรวมข้อมูลเชิงลึกของผู้เชี่ยวชาญ จึงได้นำกรอบแนวคิดอารมณ์พื้นฐาน 8 ประเภทของ Plutchik (ประกอบด้วยอารมณ์เชิงบวก 4 อารมณ์ และเชิงลบ 4 อารมณ์) มาใช้ในการวิเคราะห์ เพื่อจับความแตกต่างของการแสดงออกทางอารมณ์ในเนื้อหาได้อย่างละเอียดอ่อนและชัดเจนยิ่งขึ้น

วิทยานิพนธ์นี้ได้กำหนดให้การทดลองหาโมเดลประเมินจุดประสงค์ในการสื่อความของเนื้อหาเป็นการจำแนกข้อความแบบหลายป้ายกำกับ โดยชุดข้อมูลที่ใช้ในการจำแนกเจตนาของข้อความได้จากการเก็บรวบรวมและติดป้ายกำกับจากการให้คำปรึกษากับผู้เชี่ยวชาญ และวิเคราะห์เนื้อหาข้อความ ซึ่งมีจำนวนทั้งหมด 1,400 บทความ โดยแสดงจำนวนของแต่ละคลาสตามจุดประสงค์ของเนื้อหา (Intention Class) ดังในตารางที่ 3.1 พบว่าข้อมูลมีความไม่สมดุล (Imbalance) ค่อนข้างสูง

ตารางที่ 3.1 จำนวนบทความในชุดข้อมูลแบ่งตามจุดประสงค์การสื่อความ

จุดประสงค์การสื่อความ (Intention Class)	**จำนวน (บทความ)**
Inform (แจ้งข้อมูล)	1,089
Entertain (บันเทิง)	175
Persuade (โน้มน้าว)	136
รวม	**1,400**

หลังจากการปรึกษากับผู้เชี่ยวชาญเพิ่มเติมและการวิเคราะห์เนื้อหาข้อความอีกครั้ง พบว่าข้อความหนึ่งบทความสามารถมีจุดประสงค์ได้มากกว่า 1 อย่างพร้อมกัน จึงตัดสินใจใช้วิธีจำแนกแบบหลายป้ายกำกับ เพื่อให้สอดคล้องกับลักษณะของข้อมูล หลังจากนั้นได้นำชุดข้อมูลมาวิเคราะห์ใหม่อีกครั้ง เพื่อแสดงจำนวนของแต่ละคลาสตามจุดประสงค์หลังจากใช้วิธีจำแนกแบบหลายป้ายกำกับ ซึ่งแสดงไว้ในตารางที่ 3.2

ตารางที่ 3.2 จำนวนบทความในชุดข้อมูลแบ่งตามจุดประสงค์การสื่อความหลังปรับเป็นการจำแนกแบบหลายป้ายกำกับ

จุดประสงค์การสื่อความ (Intention Class)	**จำนวน (ป้ายกำกับ)**
Inform (แจ้งข้อมูล)	1,189
Entertain (บันเทิง)	373
Persuade (โน้มน้าว)	659
รวม	**2,221**

โดยหลังจากใช้เทคนิคการจำแนกแบบหลายป้ายกำกับ โดยจำนวนป้ายกำกับรวมจะมากกว่า 1,400 เนื่องจากตัวเลข 1,400 หมายถึงจำนวนบทความ ไม่ใช่จำนวนป้ายกำกับ เพราะหนึ่งบทความสามารถมีได้หลายป้ายกำกับพร้อมกัน

วิทยานิพนธ์นี้ยังได้นำวิธีการจำแนกแบบหลายป้ายกำกับไปใช้กับการติดป้ายกำกับด้านอารมณ์ ของบทความจำนวน 1,400 บทความเช่นกัน โดยผลการกระจายตัวของแต่ละอารมณ์แสดงไว้ในตารางที่ 3.3 ซึ่งพบว่าบางอารมณ์มีข้อมูลที่ไม่สมดุลสูงมาก ดังนั้น เพื่อแก้ไขปัญหานี้ วิทยานิพนธ์นี้จึงตัดสินใจนำคลาสอารมณ์ที่มีจำนวนน้อยกว่า 100 รายการออกจากคุณลักษณะด้านอารมณ์ในการวิเคราะห์ครั้งนี้

ตารางที่ 3.3 จำนวนป้ายกำกับในแต่ละคลาสอารมณ์ ในรูปแบบหลายป้ายกำกับ

อารมณ์ (Emotion Class)	**จำนวน (ป้ายกำกับ)**
Joy (ความสุข)	738
Trust (ความไว้วางใจ)	835
Fear (ความกลัว)	54 (ตัดออก)
Anger (ความโกรธ)	34 (ตัดออก)
Surprise (ความประหลาดใจ)	260
Anticipation (ความคาดหวัง)	944
Sadness (ความเศร้า)	73 (ตัดออก)
Disgust (ความรังเกียจ)	7 (ตัดออก)
รวม	**2,945**

จากรายละเอียดข้างต้น ได้แบ่งการทดลองออกเป็น 2 รูปแบบ สำหรับโมเดลจำแนก Intention ได้แก่ 1) จำแนก Intention โดยไม่ใช้อารมณ์ที่สื่อสารโดยผู้สร้างเนื้อหาเป็นฟีเจอร์ และ 2) จำแนก Intention โดยใช้อารมณ์ที่สื่อสารโดยผู้สร้างเนื้อหาเป็นฟีเจอร์

สำหรับจำนวนรอบในการเทรน กำหนดเท่ากันทั้ง 2 รูปแบบ และทั้ง 3 โมเดลในกลุ่ม Transformer (mBERT, WangchanBERTa, PhayaThaiBERT) ที่ 20 รอบ และ learning rate ที่ 0.00002

3.3.2.3 การทดลองที่ 3 โมเดลจำแนก Intention ทดสอบความสมดุลของข้อมูล

จากการทดลองที่ 2 ซึ่งเป็นการทดลองจำแนกเจตนา (Intention Classification) พบว่าชุดข้อมูลมีปัญหาความไม่สมดุลของข้อมูลอย่างชัดเจน โดยเฉพาะคลาส Inform ที่มักปรากฏเพียงลำพังในบทความเดียว และยังพบร่วมกับคลาสอื่นเกือบทุกรูปแบบ ทั้ง Inform+Persuade และ Inform+Entertain ทำให้เกิดสมมติฐานว่าคลาสนี้อาจเป็นตัวแปรสำคัญที่ก่อให้เกิด bias ในการเรียนรู้ของโมเดล ข้อมูลที่ไม่สมดุลดังกล่าวอาจส่งผลต่อประสิทธิภาพของการจำแนกเจตนาและ

อารมณ์ของระบบโมเดลภายหลัง ด้วยเหตุนี้จึงได้มีการออกแบบการทดลองที่ 3 โดยมุ่งเน้นไปที่การปรับชุดข้อมูลให้สมดุลในลักษณะต่าง ๆ เพื่อทดสอบประสิทธิภาพของโมเดลจำแนกอารมณ์ ซึ่งเป็นเป้าหมายใหม่ของการทดลอง

จากการวิเคราะห์ชุดข้อมูลเดิม พบว่ากลุ่มที่มีคลาสเดียว เช่น Inform แบบเดี่ยว ปรากฏมากที่สุดคือ 441 ตัวอย่าง ขณะที่กลุ่มที่มีคลาส Entertain แบบเดี่ยวมีเพียง 69 ตัวอย่าง และ Entertain+Persuade มีเพียง 4 ตัวอย่าง เมื่อพิจารณาจำนวนรวมของป้ายกำกับ (label tags) ทั้งหมด พบว่า Inform มีจำนวนมากที่สุดถึง 1,189 ป้าย รองลงมาคือ Persuade จำนวน 659 ป้าย และ Entertain จำนวน 373 ป้าย สถานการณ์นี้ชี้ให้เห็นถึงความไม่สมดุลเชิงปริมาณทั้งในระดับของคลาสเดี่ยวและระดับรวมของฉลากในแต่ละบทความ

1) การสร้างชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมใกล้เคียงกัน

รูปแบบนี้เรียกว่า Dataset S โดย S ย่อมาจาก Similar ซึ่งออกแบบโดยใช้จำนวนของคลาสที่มีน้อยที่สุดในชุดข้อมูลเดิม คือคลาส Entertain ที่มีเพียง 373 ป้ายกำกับ เป็นฐานในการควบคุมไม่ให้เกิด oversampling ที่มากเกินไป วิธีการคือ ลดจำนวนของบทความในคลาสที่มีความหนาแน่นสูง ได้แก่ Inform แบบปรากฏคลาสเดี่ยว และคัดเลือกเฉพาะบทความที่มีการกระจายคลาสในลักษณะที่หลากหลาย ได้แก่ บทความที่มีสองคลาสหรือสามคลาสร่วมกัน ส่งผลให้ชุดข้อมูลที่ได้ยังคงสะท้อนบริบทจริงบางส่วน พร้อมกับมีการกระจายของคลาสที่ใกล้เคียงกัน แม้จะยังไม่เท่ากันแบบสมบูรณ์ก็ตาม โดยรายละเอียดของชุดข้อมูล Dataset S มีรายละเอียดดังตารางที่ 3.4

ตารางที่ 3.4 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมใกล้เคียงกัน (Dataset S)

ป้ายกำกับ	จำนวน (ป้ายกำกับ)
Persuade	145
Inform	138
Entertain	107
Inform, Persuade	138
Inform, Entertain	73
Inform, Entertain, Persuade	24
Entertain, Persuade	4
Total Persuade	311
Total Inform	373

ป้ายกำกับ	จำนวน (ป้ายกำกับ)
Total Entertain	208

2) การสร้างชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับของแต่ละคลาสเท่ากันทั้งหมด

รูปแบบนี้แบ่งออกเป็นสองชุดย่อย คือ Dataset A และ Dataset B โดยมีหลักเกณฑ์ในการจัดการคลาส Inform แบบแตกต่างกัน Dataset A ตัดบทความที่มีเพียงคลาส Inform ออกทั้งหมด เนื่องจากคลาส Inform แบบเดี่ยวไม่เพียงมีจำนวนมากที่สุดในชุดข้อมูลดั้งเดิม แต่ยังพบว่าเป็นคลาสที่เกิดร่วมกับคลาสอื่นได้ทั้งหมด ทำให้เกิดความเป็นไปได้สูงที่จะสร้าง bias ต่อโมเดลอย่างไม่รู้ตัว การตัดคลาสนี้ออกจะช่วยให้บทความที่เหลืออยู่แสดงถึงเจตนารมณ์ร่วมของคลาสต่าง ๆ ได้อย่างชัดเจนมากขึ้น โดยเน้นที่บทความที่มีการเกิดคลาสร่วม ได้แก่ Inform+Entertain, Inform+Persuade, Entertain+Persuade และบทความที่มีครบทั้งสามคลาส ซึ่งช่วยลดอิทธิพลของคลาส Inform ที่มากเกินไปได้ โดยมีรายละเอียดดังตารางที่ 3.5

ตารางที่ 3.5 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมเท่ากัน (Dataset A)

ป้ายกำกับ	จำนวน (ป้ายกำกับ)
Inform, Entertain	121
Persuade	118
Inform, Entertain, Persuade	69
Inform, Persuade	63
Entertain	60
Entertain, Persuade	3
Total Persuade	**253**
Total Inform	**253**
Total Entertain	**253**

ในขณะที่ Dataset B ยังคงเก็บบทความที่มีคลาส Inform แบบเดี่ยวไว้ แต่ควบคุมจำนวนให้เหมาะสมเพื่อหลีกเลี่ยงความลำเอียงจากความหนาแน่นของคลาสนี้ การคัดเลือกบทความในชุดนี้ยังคงเน้นการมีคลาสร่วมแบบเดียวกันกับ Dataset A และยังคงควบคุมจำนวนป้ายกำกับของแต่ละคลาสให้เท่ากันทั้งหมดที่ 253 ตัวอย่างต่อคลาส การเปรียบเทียบระหว่าง Dataset A และ Dataset

B จะช่วยให้เข้าใจได้ว่าการคงไว้หรือตัดคลาส Inform เดี่ยวมีผลต่อการเรียนรู้และประสิทธิภาพของโมเดลในลักษณะใด ตามรายละเอียดในตารางที่ 3.6

ตารางที่ 3.6 รายละเอียดการกระจายของข้อมูลใน ชุดข้อมูลแบบ Balanced Dataset ที่จำนวนป้ายกำกับรวมเท่ากัน (Dataset B)

ป้ายกำกับ	จำนวน (ป้ายกำกับ)
Entertain	106
Persuade	106
Inform, Entertain	75
Inform, Persuade	75
Inform, Entertain, Persuade	69
Inform	34
Entertain, Persuade	3
Total Persuade	253
Total Inform	253
Total Entertain	253

การทดลองทั้งหมดถูกออกแบบโดยใช้ชุดข้อมูลทั้งสาม ได้แก่ Dataset S, Dataset A และ Dataset B และแบ่งการทดลองย่อยตามเงื่อนไขการใช้ฟีเจอร์อารมณ์ (emotion features) เช่นเดียวกับการทดลองก่อนหน้า โดยในแต่ละชุดข้อมูลมีการกำหนดสัดส่วนของชุดฝึกและชุดทดสอบในรูปแบบต่าง ๆ เพื่อประเมินผลของสมดุลข้อมูลร่วมกับการใช้ฟีเจอร์อารมณ์อย่างเป็นระบบ ตามรายละเอียดในตารางที่ 3.7

ตารางที่ 3.7 ตารางสรุปการทดลองย่อยการทดลองที่ 3 โมเดลจำแนก Intention ทดสอบความสมดุลของข้อมูล

การทดลองย่อยที่	ชุดข้อมูล	รายละเอียด	เงื่อนไข
1	Dataset S	แบ่งชุดข้อมูล Train 70 : Test 30	ไม่ใช้ฟีเจอร์อารมณ์
2	Dataset S	แบ่งชุดข้อมูล Train 80 : Test 20	ไม่ใช้ฟีเจอร์อารมณ์
3	Dataset A	แบ่งชุดข้อมูล Train 70 :	ไม่ใช้ฟีเจอร์อารมณ์

การทดลองย่อยที่	ชุดข้อมูล	รายละเอียด	เงื่อนไข
		Test 30	
4	Dataset B	แบ่งชุดข้อมูล Train 70 : Test 30	ไม่ใช้ฟีเจอร์อารมณ์
5	Dataset S	แบ่งชุดข้อมูล Train 70 : Test 30	ใช้ฟีเจอร์อารมณ์
6	Dataset S	แบ่งชุดข้อมูล Train 80 : Test 20	ใช้ฟีเจอร์อารมณ์
7	Dataset A	แบ่งชุดข้อมูล Train 70 : Test 30	ใช้ฟีเจอร์อารมณ์
8	Dataset B	แบ่งชุดข้อมูล Train 70 : Test 30	ใช้ฟีเจอร์อารมณ์

3.3.2.4 การทดลองที่ 4 โมเดลจำแนกอารมณ์ (Emotions Classification)

จากขั้นตอนการวิจัย วิทยานิพนธ์นี้ได้ดำเนินการพัฒนาโมเดลจำแนก อารมณ์ (Emotion Classification) ควบคู่ไปกับโมเดลจำแนกจุดประสงค์ของเนื้อหา (Intention Classification) โดยกำหนดให้การจำแนกอารมณ์ใช้เทคนิคแบบหลายป้ายกำกับเช่นเดียวกัน เนื่องจากหนึ่งบทความสามารถมีการแสดงออกทางอารมณ์ได้มากกว่าหนึ่งประเภทพร้อมกัน

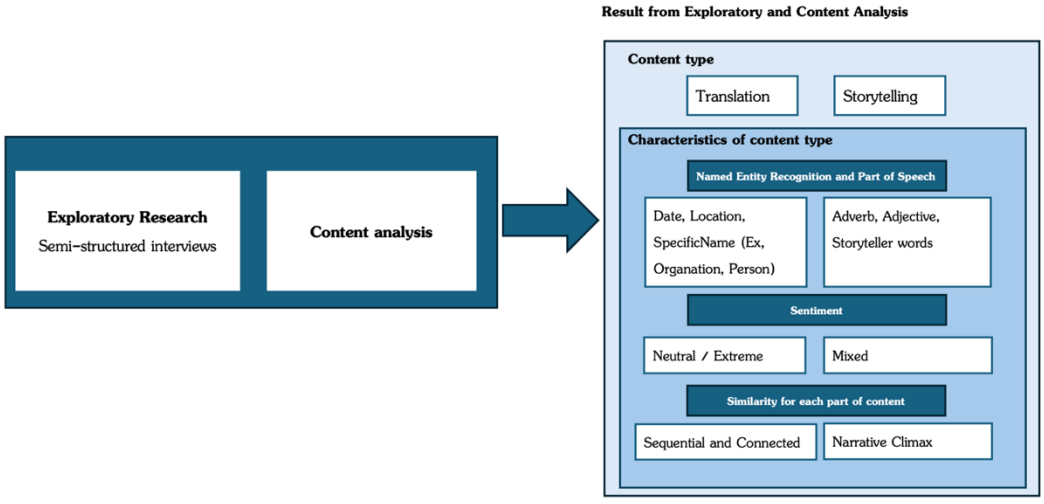
ในการพัฒนาโมเดลจำแนกอารมณ์นี้ วิทยานิพนธ์นี้เลือกใช้โมเดล Transformer จำนวน 3 โมเดล ได้แก่ mBERT, WangchanBERTa และ PhayaThaiBERT เช่นเดียวกับโมเดลจำแนก Intention ของเนื้อหา เพื่อให้เกิดความสอดคล้องในการวิเคราะห์และเปรียบเทียบประสิทธิภาพของโมเดลทั้งสองแบบ นอกจากนี้ยังดำเนินการทดลองและพัฒนาไปพร้อมกัน โดยใช้ชุดข้อมูลเดียวกันและตั้งค่าพารามิเตอร์ต่างๆ ในการฝึกโมเดลอย่างสอดคล้องกันด้วย โดยใช้จำนวนรอบการเรียนรู้ที่ 50 รอบ และ learning rate ที่ 0.00002

3.3.2.5 การทดลองที่ 5 โมเดลจำแนกงานแปลและการเล่าเรื่อง (Translation vs Storytelling)

สำหรับการพัฒนาโมเดลเพื่อจำแนกสไตล์การเขียนบทความ วิทยานิพนธ์นี้ได้สัมภาษณ์ผู้เชี่ยวชาญและถอดองค์ความรู้ พบว่าปัจจัยสำคัญที่เกี่ยวข้องโดยตรงกับการกำหนดสไตล์การเขียนเนื้อหาที่เกี่ยวข้องกับประเทศจีน ได้แก่ งานแปล (Translation) และ การเล่าเรื่อง (Storytelling)

เพื่อให้ได้ข้อมูลเชิงลึกจากผู้สร้างสรรค์เนื้อหา ทางวิทยานิพนธ์นี้ได้ดำเนินการสัมภาษณ์แบบกึ่งโครงสร้าง (Semi-structured Interviews) เพื่อเป็นการวิจัยเชิงสำรวจเบื้องต้น ซึ่งช่วยให้สามารถตั้งคำถามที่เจาะจงและสำรวจประเด็นที่เกิดขึ้นในระหว่างการพูดคุยได้อย่างละเอียด วิธีนี้ทำให้วิทยานิพนธ์นี้เข้าใจมุมมอง กลยุทธ์ และกระบวนการตัดสินใจของผู้สร้างเนื้อหาเกี่ยวกับประเทศจีนได้ชัดเจนยิ่งขึ้น และสามารถรวบรวมข้อมูลที่หลากหลาย รวมถึงประสบการณ์ของแต่ละบุคคลได้อย่างครอบคลุม นอกจากการสัมภาษณ์แล้ว วิทยานิพนธ์นี้ยังได้วิเคราะห์เนื้อหาโพสต์จากเพจ Facebook เน้นการวิเคราะห์ด้านการเลือกใช้ภาษา โทนในการเขียน ความยาวของเนื้อหา และเทคนิคการเล่าเรื่อง เป้าหมายเพื่อค้นหารูปแบบและแนวโน้มในการตัดสินใจของผู้สร้างเนื้อหา โดยสรุปผลจากการสัมภาษณ์และวิเคราะห์เนื้อหาดังกล่าว

หลังจากเสร็จสิ้นกระบวนการสำรวจเชิงลึกและการวิเคราะห์เนื้อหาแล้ว วิทยานิพนธ์นี้ได้นำองค์ความรู้จากการวิเคราะห์นี้ มาใช้ในการสร้างคุณลักษณะ (Feature Engineering) ที่เกี่ยวข้องกับข้อมูลชุดเนื้อหาภาษาไทยที่มีประเด็นเกี่ยวข้องกับประเทศจีน โดยนำมาพัฒนาเป็นโมเดลจำแนกสไตล์เนื้อหา โดยใช้วิธีการเรียนรู้ของเครื่องแบบ Ensemble Learning ร่วมกับ Feature Engineering ซึ่งแบ่งส่วนของเนื้อหาออกเป็น 3 ส่วนหลักคือ ส่วนบทนำ ส่วนเนื้อหา และสรุป โดยคุณลักษณะเฉพาะที่นำมาพัฒนา ตามระบุในกรอบย่อย Characteristics of content type ในภาพที่ 3.4 มีรายละเอียดดังต่อไปนี้



ภาพที่ 3.4 การพัฒนาโมเดลจำแนกเนื้อหาการแปลและเล่าเรื่อง

สำหรับ Feature Engineering ของโมเดลจำแนกเนื้อหาการแปลและเล่าเรื่อง แบ่งออกเป็น 6 คุณลักษณะ ดังนี้

1) คุณลักษณะด้าน Named Entity Recognition (NER) และ Part of Speech (POS) การระบุสไตล์เนื้อหาว่าเป็น Translation (งานแปล) จะพิจารณาจากการมีการกล่าวถึงข้อมูลที่ชัดเจน เช่น วันที่ เวลา สถานที่ และชื่อองค์กรต่างๆ ที่มักพบในเนื้อหาที่แปลมาจากต้นฉบับภาษาจีน ในขณะที่สไตล์ Storytelling (การเล่าเรื่อง) จะพบการใช้คำคุณศัพท์ (adjectives) คำกริยาวิเศษณ์ (adverbs) และสรรพนามบุรุษที่หนึ่ง เพื่อสื่อสารลักษณะการเล่าเรื่องจากมุมมองของผู้เขียน ดังนั้น วิทยานิพนธ์นี้จึงพัฒนาคุณลักษณะเหล่านี้โดยใช้เครื่องมือ ThaiNameTagger และ pos_tag จาก PyThaiNLP Python library เพื่อทำ NER และ POS Tagging ตามลำดับ

2) คุณลักษณะด้านความรู้สึก (Sentiment Features) วิทยานิพนธ์นี้ได้ใช้โมดูล TextBlob ในการวิเคราะห์ความรู้สึกที่ปรากฏในเนื้อหา จากผลการวิเคราะห์เบื้องต้นพบว่า หากเนื้อหามีความรู้สึกเป็นกลาง (neutral) หรือชัดเจนไปในทิศทางใดทิศทางหนึ่ง (extreme) มักจะถูกจัดเป็นประเภท Translation ในทางกลับกัน หากเนื้อหามีอารมณ์หรือความรู้สึกที่ผสมกันหลายทิศทาง (mixed sentiment) มักสะท้อนสไตล์ที่มีความเป็นส่วนตัวและมีการเล่าเรื่องที่ละเอียดอ่อน (nuanced) ซึ่งจัดเป็นสไตล์ Storytelling

3) คุณลักษณะด้านความคล้ายคลึง (Similarity Features) วิทยานิพนธ์นี้ได้ใช้เทคนิค Topic Modeling แบบ Latent Dirichlet Allocation (LDA) เพื่อวิเคราะห์ความเชื่อมโยงระหว่างแต่ละส่วนของเนื้อหา (Introduction, Body, Conclusion) โดยจากการสัมภาษณ์และวิเคราะห์เนื้อหาพบว่าหากส่วนต่างๆ โดยเฉพาะสรุป (Conclusion) และเนื้อหา (Body) มีความเชื่อมโยงกันอย่างชัดเจน และให้ข้อมูลที่เสริมกันเป็นลำดับขั้น (step-by-step) จะถูกจัดเป็นประเภท Translation ในทางกลับกัน หากส่วนสรุปมีลักษณะการเล่าเรื่องที่มีจุดไคลแมกซ์ (climax) และไม่ได้เชื่อมโยงกับส่วนก่อนหน้าโดยตรง จะถูกจัดอยู่ในประเภท Storytelling

4) คุณลักษณะด้านความยาวของเนื้อหา (Length Features) เป็นการวัดจำนวนคำในส่วนต่างๆ ของข้อความได้แก่ ส่วนบทนำ ส่วนเนื้อหา และสรุป เพื่อวิเคราะห์ความซับซ้อน ความละเอียดและความลึกของเนื้อหา ส่วนที่มีจำนวนคำมากอาจแสดงถึงเนื้อหาที่ละเอียดและมีความซับซ้อนหรือให้รายละเอียดที่มากกว่า

5) คุณลักษณะ TF-IDF โดย TF-IDF เป็นคุณลักษณะที่ใช้ในการวิเคราะห์ความสำคัญของคำศัพท์ในเอกสารแต่ละฉบับ เทียบกับชุดเอกสารทั้งหมด โดยให้คะแนนสูงกับคำศัพท์ที่ปรากฏบ่อยในเอกสารหนึ่งๆ แต่ไม่ค่อยปรากฏในเอกสารอื่นๆ คุณลักษณะประเภทนี้ช่วยในการแยกแยะสไตล์การเขียนของเนื้อหาโดยการระบุคำสำคัญที่เฉพาะเจาะจงต่อสไตล์นั้นๆ

6) โมเดลที่นำมาใช้ในการทดลอง ได้แก่ 4 โมเดลในกลุ่ม Ensemble learning (Random Forest XGBoost LightGBM และ AdaBoost) และ โมเดลในกลุ่ม Transformer-based Deep Learning (mBERT WangchanBERTa และ PhayaThaiBERT) โดยสาเหตุที่เลือกโมเดลเหล่านี้ เนื่องจากวิทยานิพนธ์นี้มุ่งเน้นที่การวิเคราะห์คุณลักษณะ (Feature Engineering) ในการจำแนกประเภทข้อความ ดังนั้น วิทยานิพนธ์นี้จึงเลือกใช้เทคนิคการเรียนรู้ของเครื่องแบบผสมผสาน (Ensemble Machine Learning) ซึ่งเป็นวิธีการที่รวมผลการทำนายจากหลายโมเดลย่อย (base models) เข้าด้วยกัน เพื่อสร้างผลลัพธ์ที่มีความแม่นยำและน่าเชื่อถือสูงขึ้น โดยแต่ละโมเดลในกลุ่ม Ensemble learningมีจุดแข็งเฉพาะด้าน เช่น ความสามารถในการลดความแปรปรวน ลดความเอนเอียง (bias) หรือทั้งสองอย่างพร้อมกัน ซึ่งช่วยให้สามารถรับมือกับลักษณะข้อมูลที่มีความซับซ้อนและไม่เป็นเชิงเส้นได้ดี อีกทั้งยังเหมาะสมกับข้อมูลเชิงโครงสร้างที่ได้จากการวิศวกรรมคุณลักษณะ เช่น จำนวนคำ ความยาว โครงสร้างไวยากรณ์ คะแนนความรู้สึก (sentiment) และจำนวน named entities เป็นต้น ซึ่งเป็นลักษณะที่สามารถแยกแยะสไตล์ของข้อความ เช่น แปลภาษา หรือ การเล่าเรื่อง ได้อย่างมีประสิทธิภาพ

อย่างไรก็ตาม เพื่อประเมินความแม่นยำของโมเดลอย่างรอบด้าน วิทยานิพนธ์นี้ได้ทดลองเปรียบเทียบกับโมเดลในกลุ่ม Transformer-based Deep Learning ได้แก่ mBERT WangchanBERTa และ PhayaThaiBERT ซึ่งเป็นโมเดลที่ใช้กันแพร่หลายในงานประมวลผลภาษาธรรมชาติ (NLP) ที่เกี่ยวข้องกับภาษาไทย

3.3.3 การจัดการข้อมูลก่อนเข้าสู่โมเดล (Pre-Processing)

เริ่มต้นจากการทำความสะอาดข้อมูล (Cleansing) ประกอบไปด้วยขั้นตอนดังนี้

1) ลบอักขระพิเศษทั้งหมดออก

2) ลบตัวอักษรภาษาอังกฤษและภาษาจีนออก คงไว้เฉพาะภาษาไทยเท่านั้นเนื่องจากเนื้อหาที่ใช้เป็นบทความเกี่ยวกับจีนที่เขียนเป็นภาษาไทย กลุ่มผู้อ่านหลักคือคนไทย ดังนั้น การจำแนกบทความว่าเป็น Timely หรือ Timeless จึงพิจารณาเฉพาะข้อมูลที่เป็นภาษาไทย

3) ลบตัวเลขออก

แม้ว่าหลายการศึกษาเกี่ยวกับการวิเคราะห์เนื้อหาจะรวมปัจจัยด้านเวลา เช่น ตัวเลขที่บอกเวลาในอนาคตเข้ามาด้วย แต่การวิจัยนี้ยังไม่ได้รวมปัจจัยเรื่องเวลา จึงเลือกที่จะลบตัวเลขออก คงไว้เพียงตัวอักษรภาษาไทย

4) ลบลิงก์เว็บไซต์และแฮชแท็ก

เนื่องจากต้องการวิเคราะห์เฉพาะเนื้อหาของบทความเท่านั้น

โดยหลังจากทำความสะอาดข้อมูลแล้ว จะเข้าสู่กระบวนการตัดคำแต่ละคำออกจากประโยค โดยใช้เทคนิค Tokenization ตามที่ระบุรายละเอียดในแต่ละโมเดลในหัวข้อ 3.3.1

3.3.4 การออกแบบการประเมินผลโมเดล (Model Validation)

ในการพัฒนาแต่ละโมเดล จะมีการประเมินผลโมเดลเพื่อวัดประสิทธิภาพ ซึ่งแตกต่างกันตามลักษณะของโมเดล โดยในวิทยานิพนธ์ฉบับนี้ มี 2 ประเภทหลัก คือ โมเดลแบบจำแนกหลายคลาส และ โมเดลแบบจำแนกหลายป้ายกำกับ ดังตารางที่ 3.8 และมีรายละเอียดดังนี้

ตารางที่ 3.8 รูปแบบการจำแนกเนื้อหาและประเภทโมเดล

รูปแบบการจำแนกเนื้อหา	ประเภทโมเดล
Timely/Timeless	โมเดลแบบจำแนกหลายคลาส
Intention	โมเดลแบบจำแนกหลายป้ายกำกับ
Emotion	โมเดลแบบจำแนกหลายป้ายกำกับ
Translation vs Storytelling	โมเดลแบบจำแนกหลายคลาส

3.3.4.1 การประเมินประสิทธิภาพโมเดลสำหรับโมเดลแบบจำแนกหลายคลาส

การประเมินประสิทธิภาพโมเดลสำหรับโมเดลประเภทนี้ มีรายละเอียดดังนี้

1) ประเมิน ค่า Precision

ค่า Precision คือจำนวนผลลัพธ์ที่โมเดลทายออกมาว่าถูกและตรงกับข้อมูลทีจำแนกผลไว้เรียกว่า True Positive โดยนำมาหารด้วยผลรวมระหว่าง True Positive และ สิ่งที่โมเดลทำนายว่าถูกแต่แท้จริงแล้วชุดข้อมูลที่จำแนกไว้ระบุว่าผิด หรือที่เรียกว่า False Negative

Precision สามารถคำนวณได้ดังสมการที่ (3.1)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3.1)$$

2) ประเมิน ค่า Recall

ค่า Recall คือ จำนวนผลลัพธ์ True Positive หารด้วยผลรวมระหว่าง True Positive และสิ่งที่มีโมเดลทำนายว่าไม่ถูก แต่แท้จริงแล้วข้อมูลที่จำแนกไว้คือถูก เรียกว่า False Negative

Recall สามารถคำนวณได้ดังสมการที่ (3.2)

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3.2$$

3) ประเมินค่า Accuracy

ค่า Accuracy คือผลลัพธ์ True Positive รวมกับ True Negative คือโมเดลทำนายว่าไม่ถูกต้องและข้อมูลที่จำแนกไว้ก็ถูกกำหนดไว้ว่าไม่ถูกต้อง กล่าวคือ เป็นการรวมผลลัพธ์จำนวนครั้งที่ทำนายได้ถูกต้องในทุกคลาส จากนั้นนำไปหารด้วย ทุกเคส ทั้ง True Positive (ทำนายว่าถูก จำแนกผลไว้คือถูกเช่นกัน) True Negative (ทำนายว่าผิด จำแนกผลไว้คือผิดเช่นกัน) False Positive (ทำนายว่าถูก แต่จำแนกผลไว้คือไม่ถูก) และ False Negative (ทำนายว่าไม่ถูก แต่จำแนกผลไว้คือถูก)

Accuracy สามารถคำนวณได้ดังสมการที่ (3.3)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (3.3$$

4) ประเมินค่า F1-Score

ค่า F1-Score คือค่าเฉลี่ยในรูปแบบ Harmonic mean ระหว่าง Precision และ Recall เวลาประเมินประสิทธิภาพโมเดล เราเลยสามารถดูแค่ค่า F1-Score และ Accuracy สองค่าก็ได้ เพราะ F1-Score เป็นการเฉลี่ยค่า Precision และ Recall แล้ว

F1-Score สามารถคำนวณได้ดังสมการที่ (3.4)

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall}\right) \quad (3.4)$$

3.3.4.2 การประเมินประสิทธิภาพโมเดลสำหรับโมเดลแบบหลายป้ายกำกับ

การวัดประสิทธิภาพของโมเดลในการจำแนกแบบหลายป้ายกำกับ ได้แก่ โมเดลประเมินจุดประสงค์ในการสื่อความของเนื้อหา และโมเดลประเมินจุดประสงค์ในการสื่ออารมณ์ของเนื้อหา สามารถวัดได้ด้วยเมตริกที่มีลักษณะเฉพาะ เพื่อประเมินความสามารถของโมเดลได้อย่างเหมาะสม โดยแต่ละตัวชี้วัดประสิทธิภาพ (evaluation metric) จะคำนวณจากค่าผลบวกแท้จริง (True positives: tp_j), ผลลบแท้จริง (True negatives: tn_j), ผลบวกเทียม (False positives: fp_j) และผลลบเทียม (False negatives: fn_j) ของแต่ละป้ายกำกับ y_j โดยที่ $j = [1,...,m]$

สำหรับตัวชี้วัด Macro F1 Score นั้นจะคำนวณโดยใช้ค่าเฉลี่ยฮาร์มอนิก (harmonic mean) ของ Precision และ Recall ซึ่งจะคำนวณแยกกันในแต่ละป้ายกำกับก่อน แล้วนำค่าที่ได้จากทุกป้ายกำกับมาหาค่าเฉลี่ยรวมกัน ข้อสังเกตสำคัญของตัวชี้วัด Macro คือ ปฏิบัติต่อทุกป้ายกำกับอย่างเท่าเทียมกันโดยไม่มีการถ่วงน้ำหนัก จึงไม่ได้สะท้อนถึงความไม่สมดุลของข้อมูลในแต่ละป้ายกำกับ ในทางกลับกัน ตัวชี้วัด Micro F1 Score จะถูกคำนวณโดยใช้ค่าเฉลี่ยฮาร์โมนิกของ Micro

Precision และ Micro Recall ที่ได้จากการรวมค่าผลบวกแท้จริง ผลลบแท้จริง ผลบวกเทียม และผลลบเทียมของทุกป้ายกำกับเข้าด้วยกันก่อน แล้วจึงนำไปคำนวณ Precision และ Recall

นอกจากนี้ยังมีการใช้ Hamming Loss ซึ่งเป็นตัวชี้วัดประสิทธิภาพที่สำคัญอีกตัวหนึ่งในการประเมินการจำแนกแบบหลายป้ายกำกับ โดยแสดงถึงสัดส่วนของป้ายกำกับที่ถูกทำนายผิดพลาดทั้งหมด (ทั้งผลบวกเทียมและผลลบเทียม) ต่อจำนวนป้ายกำกับทั้งหมด

สามารถสรุปเมตริกที่ใช้ดังนี้

1) Macro Precision (Macro P)

เป็นการวัดค่าความแม่นยำโดยเฉลี่ยจากทุกๆ คลาส โดยคำนวณค่าความแม่นยำ (Precision) ของแต่ละคลาส แล้วนำมาเฉลี่ยรวมกัน สามารถคำนวณได้จากสมการที่ (3.5)

$$\text{Macro Precision (Macro P)} = \frac{1}{m}\sum_{j=1}^{m}\frac{tp_j}{tp_j+fp_j} \qquad (3.5)$$

2) Macro Recall (Macro R)

เป็นการวัดค่าความครอบคลุมโดยเฉลี่ยของทุกคลาส โดยคำนวณค่าความครอบคลุม (Recall) ของแต่ละคลาส แล้วนำมาเฉลี่ยรวมกัน สามารถคำนวณได้จากสมการ (3.6)

$$\text{Macro Recall (Macro R)} = \frac{1}{m}\sum_{j=1}^{m}\frac{tp_j}{tp_j+fn_j} \qquad (3.6)$$

3) Macro F1

เป็นการวัดค่าความถูกต้องโดยรวมของโมเดล โดยเฉลี่ย F1 ของทุกคลาส ซึ่งคำนึงถึงความสมดุลระหว่าง Precision และ Recall ในแต่ละคลาส สามารถคำนวณได้จากสมการ (3.7)

$$\text{Macro F1} = \frac{1}{m}\sum_{j=1}^{m}\frac{2\times R_j\times P_j}{R_j+P_j} \qquad (3.7)$$

4) Micro Precision (Micro P)

เป็นการวัดค่าความแม่นยำโดยรวมโดยนับผลรวมจำนวนครั้งที่ทำนายถูกทั้งหมด หารด้วยผลรวมของจำนวนครั้งที่ทำนายทั้งหมด โดยไม่แยกแต่ละคลาส สามารถคำนวณได้จากสมการ (3.8)

$$\text{Micro Precision (Micro P)} = \frac{\sum_{j=1}^{m} tp_j}{\sum_{j=1}^{m} tp_j+\sum_{j=1}^{m} fp_j} \qquad (3.8)$$

5) Micro Recall (Micro R)

เป็นการวัดค่าความครอบคลุมโดยรวมของทุกคลาสแบบรวมกันทั้งหมด โดยนับผลรวมจำนวนครั้งที่ทำนายถูกต้อง หารด้วยผลรวมของจำนวนครั้งที่มีการกำกับจริงทั้งหมด สามารถคำนวณได้จากสมการ (3.9)

$$\text{Micro Recall (Micro R)} = \frac{\sum_{j=1}^{m} tp_j}{\sum_{j=1}^{m} tp_j + \sum_{j=1}^{m} fn_j} \quad (3.9)$$

6) Micro F1

เป็นการวัดค่าความถูกต้องโดยรวมของโมเดลที่คำนึงถึงความสมดุลระหว่าง Micro Precision และ Micro Recall ในภาพรวมทั้งหมดโดยไม่แยกแต่ละคลาส สามารถคำนวณได้จากสมการ (3.10)

$$\text{Micro F1} = \frac{2 \times Micro\ R \times Micro\ P}{Micro\ R + Micro\ P} \quad (3.10)$$

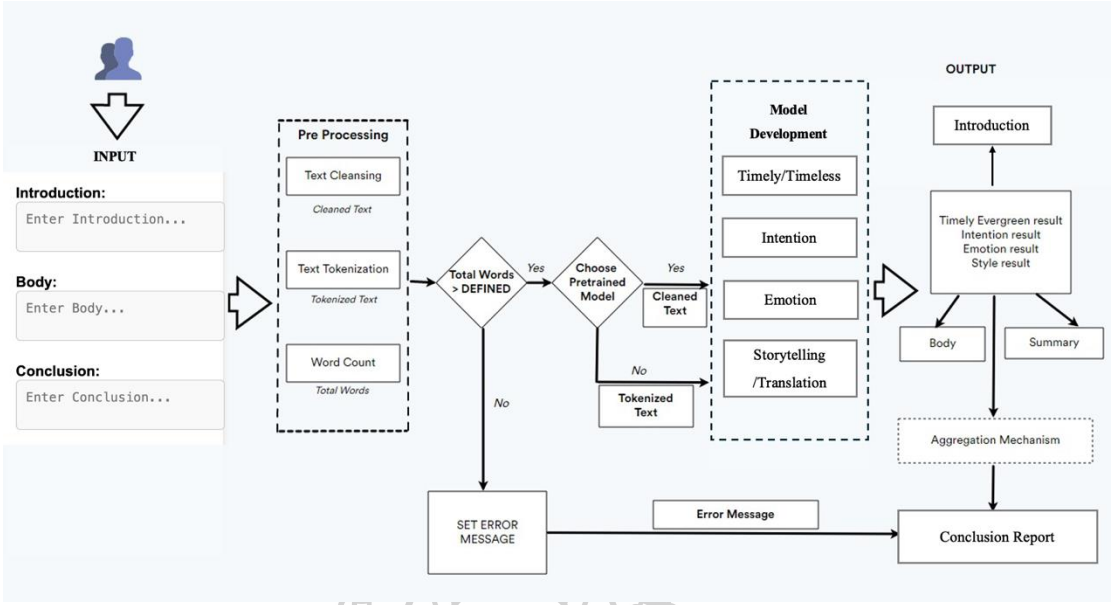
7) Hamming Loss

เป็นการวัดค่าความผิดพลาดของการจำแนกแบบหลายป้ายกำกับ โดยคิดจากสัดส่วนของป้ายกำกับที่ทำนายผิดทั้งหมด หารด้วยจำนวนรวมของป้ายกำกับที่เป็นไปได้ทั้งหมด สามารถคำนวณได้จากสมการ (3.11)

$$\text{Hamming Loss} = \frac{1}{|N| \cdot |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} xor\left(y_{ij}, \widehat{y_{iJ}}\right) \quad (3.11)$$

3.4 การพัฒนาระบบ Web application

สำหรับการพัฒนาระบบ Web application เพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษา เนื้อหาที่เกี่ยวข้องกับจีน ประกอบไปด้วย 3 ส่วนหลัก ได้แก่ 1) ส่วนนำเข้า (Input) 2) การพัฒนาส่วนผลรวมผลทำนายจากโมเดลเพื่อวิเคราะห์ความสอดคล้องของสไตล์ (Aggregation Mechanism) 3) ส่วนแสดงผลลัพธ์ (Output) ดังภาพที่ 3.5 ซึ่งจะมีการเชื่อมต่อระหว่าง ส่วนนำเข้า และส่วนแสดงผลลัพธ์ ด้วย API



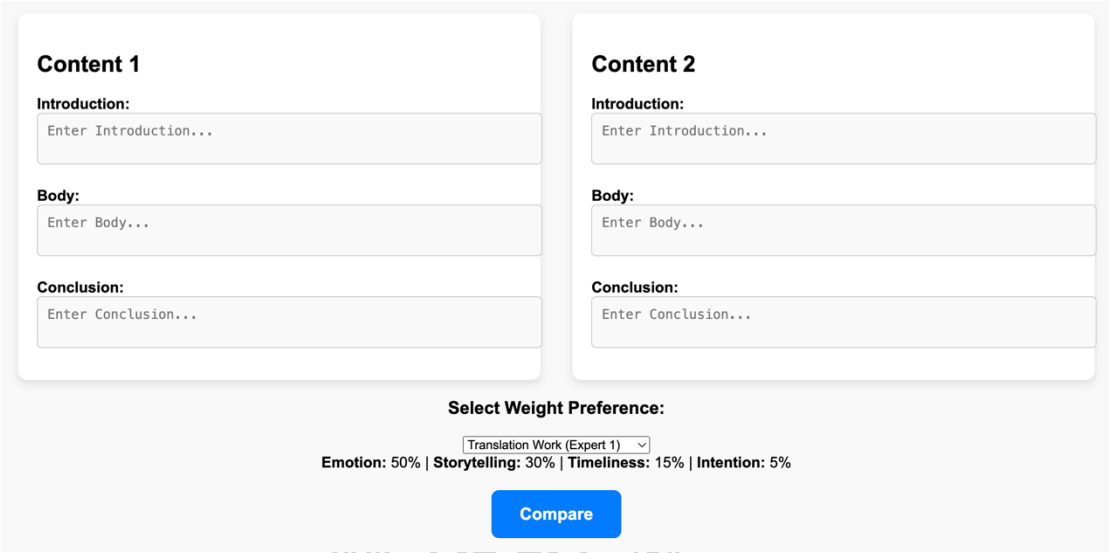
ภาพที่ 3.5 ภาพรวมของการพัฒนาระบบ Web application และโมเดลวิเคราะห์เนื้อหา

3.4.1 การพัฒนาส่วนนำเข้า (Input)

ส่วนนำเข้าของเฟรมเวิร์กเพื่อประเมินการเขียนบทความที่สอดคล้องกับต้นแบบด้วยเทคนิคการเรียนรู้ของเครื่อง ถูกจัดทำขึ้นในลักษณะของเว็บแอปพลิเคชัน โดยออกแบบส่วนรับข้อมูลสอดคล้องกับไตล์การเขียนของเฟซบุ๊กเพจอ้ายจงที่แบ่งบทความออกเป็นสี่ส่วน ตามที่ปรากฏรายละเอียดในหัวข้อย่อย สไตล์การเขียนของบทความของบทที่ 2

ระบบต้นแบบในวิทยานิพนธ์นี้ได้รับการพัฒนาในรูปแบบ Web-based Application โดยใช้โครงสร้างหลักของหน้าเว็บด้วยภาษา HTML และ CSS เพื่อจัดวางองค์ประกอบของระบบในรูปแบบอินเทอร์เฟซแบบโต้ตอบ (interactive interface) ร่วมกับ JavaScript โดยในส่วนของการป้อนข้อความเปรียบเทียบ ได้ออกแบบให้เป็น โครงสร้างแบบ 2 คอลัมน์ (2-column layout) ภายใน แต่ละฝั่ง แสดงตำแหน่งของเนื้อหาใน 3 ส่วน ได้แก่ Introduction, Body และ Conclusion ซึ่งควบคุมการแสดงผลด้วย CSS เพื่อให้ผู้ใช้งานสามารถป้อนข้อความได้อย่างเป็นระบบและเข้าใจง่าย

โดยส่วนนำเข้า ประกอบไปด้วยส่วนต่างๆ ดังภาพที่ 3.6 โดยมีรายละเอียดแต่ละส่วน ดังนี้



ภาพที่ 3.6 ส่วนนำเข้าข้อมูล

3.4.1.1 ส่วนเกริ่นนำ (Introduction)

ใส่รายละเอียดของบทความในส่วนเกริ่นนำ

3.4.1.2 ส่วนเนื้อหา (Body)

ใส่รายละเอียดของบทความในส่วนของเนื้อหาบทความ

3.4.1.3 ส่วนสรุปท้ายบทความ (Conclusion)

ใส่รายละเอียดส่วนสรุปท้ายบทความ

3.4.1.4 ส่วนของการเลือกค่าถ่วงน้ำหนัก (Select Weight Preference)

สำหรับเลือกค่าถ่วงน้ำหนักของแต่ละฟีเจอร์ที่ใช้ในการประเมินความสอดคล้องของบทความ โดยประมวลผลผ่านปุ่มดำเนินการประเมินความสอดคล้องของบทความ ซึ่งระบบจะเชื่อมต่อกับแบบจำลองการเรียนรู้ของเครื่อง เพื่อวิเคราะห์และประเมินเนื้อหาของทั้งสี่ส่วน คือ ส่วนเกริ่นนำ ส่วนเนื้อหา สรุปท้ายบทความ และค่าถ่วงน้ำหนัก

โดยการเลือกค่าถ่วงน้ำหนัก มีให้เลือก 4 รูปแบบ โดยได้มาจากการสัมภาษณ์ถอดองค์ความรู้จากผู้เชี่ยวชาญ ได้แก่

1) สำหรับงานแปล (Translation work) โดยกำหนดความสำคัญของ Emotion มากที่สุด ที่สัดส่วน 50% ตามด้วย Storytelling/Translation 30% Timeliness 15% และ Intention 5% ซึ่งผู้เชี่ยวชาญระบุว่า สำหรับงานแปล สไตล์ของเนื้อหาจะมาจากอารมณ์ (Emotion) มากที่สุด

2) สำหรับงานข่าว (News outlet) ให้ความสำคัญกับ Intention มากที่สุดที่ 50% ตามมาด้วย Emotion 30% และ Storytelling กับ Timeliness เท่ากันที่ 10%

3) สำหรับงานทั่วไป (Balanced) โดยเน้นความสมดุลของแต่ละฟีเจอร์ โดยกำหนด Emotion: 30% Timeliness: 30% Storytelling: 20% และ Intention: 20%

4) สำหรับงานที่มุ่งเน้นความชัดเจนของเวลา (Timeliness-focused) เป็นหลัก

ตัวอย่างเช่น การกำหนดชัดเจนสำหรับเนื้อหาที่เป็นตามกระแสหรือไวรัล และบทความที่สดใหม่เสมอ แต่ยังคงให้ความสำคัญกับด้านอารมณ์ เพราะถูกระบุว่า มีความเกี่ยวข้องกับความชัดเจนของเวลา ดังนั้น ค่าถ่วงน้ำหนัก มีรายละเอียดดังนี้ Timeliness: 40% Emotion: 30% Storytelling: 20% | Intention: 10%

และเมื่อเลือกรูปแบบค่าถ่วงน้ำหนักที่ช่อง Dropdown รายละเอียดของค่าถ่วงน้ำหนักจะถูกแสดงโดยทันที (Realtime) ซึ่งควบคุมด้วย JavaScript โดยใช้ฟังก์ชัน updatePresetDetails() ทำหน้าที่ดึงค่าที่ผู้ใช้เลือก และต่อจากนั้นจะนำมาคำนวณแสดงผลบนหน้าจอทันที โดยไม่ต้องรีเฟรชหน้าเว็บ (asynchronous update) ผ่านการเปลี่ยนตัวเลือกใน dropdown ด้วย onchange event ฟังก์ชัน ซึ่งทำงานโดยเข้าถึงค่าที่เลือกใน document.getElementById('weightPreset').value แล้วนำไปใช้เป็น key เพื่อดึงค่า weights ที่สัมพันธ์กันจาก object presets ซึ่งกำหนดไว้ล่วงหน้าดังภาพที่ 3.7

```
const presets = {
    translation: {emotion:0.50, storytelling:0.30, timeliness:0.15, intention:0.05},
    news: {emotion:0.30, storytelling:0.10, timeliness:0.10, intention:0.50},
    balance: {emotion:0.30, storytelling:0.20, timeliness:0.30, intention:0.20},
    timeliness: {emotion:0.30, storytelling:0.20, timeliness:0.40, intention:0.10}
};
function updatePresetDetails() {
    const preset = document.getElementById('weightPreset').value;
    const weights = presets[preset];
    document.getElementById('presetDetails').innerHTML = `
        <strong>Emotion:</strong> ${weights.emotion*100}% |
        <strong>Storytelling:</strong> ${weights.storytelling*100}% |
        <strong>Timeliness:</strong> ${weights.timeliness*100}% |
        <strong>Intention:</strong> ${weights.intention*100}%
    `;
}
```

ภาพที่ 3.7 JavaScript กำหนดค่าตัวแปร presets

3.4.2 การพัฒนาส่วนผลรวมผลทำนายจากโมเดลเพื่อวิเคราะห์ความสอดคล้องของสไตล์ (Aggregation Mechanism)

การแสดงผลเปอร์เซ็นต์ความคล้ายของรูปแบบบทความทั้งสองฉบับ จะใช้สเกลตั้งแต่ 0 ถึง 100 เปอร์เซ็นต์ โดยอาศัยวิธีการคำนวณ Hellinger Similarity จาก Hellinger Distance ดังที่อธิบายรายละเอียดไว้ในบทที่ 2

โดยสามารถแสดงความคล้ายได้ในทั้งสี่รูปแบบฟีเจอร์ ได้แก่ Timeliness, Translation vs Storytelling, Intention และ Emotion โดยแต่ละฟีเจอร์จะมีการคำนวณค่า Hellinger Similarity เป็นรายฟีเจอร์ แล้วจึงนำผลที่ได้ไปคำนวณผลรวมความคล้ายโดยรวม (Overall Similarity)

การคำนวณ Overall Similarity จะใช้ค่าถ่วงน้ำหนักที่ระบุรายละเอียดไว้ในส่วนของการเลือกค่าถ่วงน้ำหนัก หัวข้อย่อย 3.4.1.4 โดยมีสมการคำนวณดังนี้

Overall Similarity = (W_timeliness × S_timeliness) + (W_storytelling × S_storytelling) + (W_intention × S_intention) + (W_emotion × S_emotion) โดยที่

W_timeliness, W_storytelling, W_intention, W_emotion คือค่าถ่วงน้ำหนักที่กำหนดโดยผู้เชี่ยวชาญ

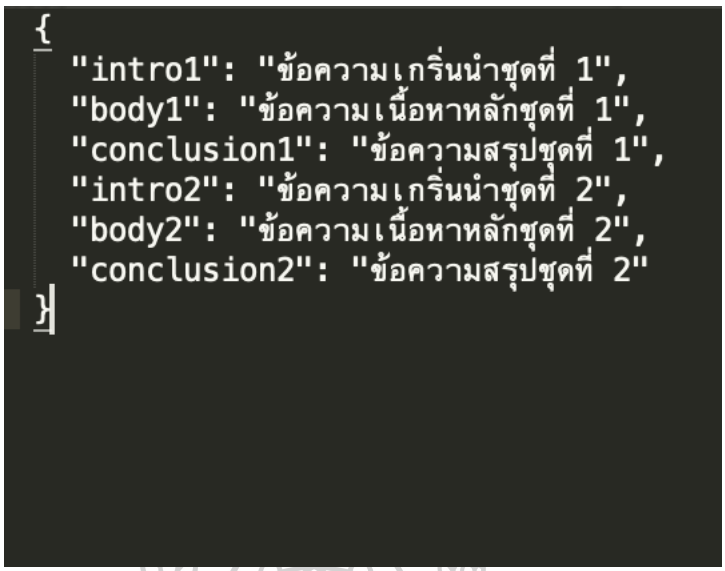
S_timeliness, S_storytelling, S_intention, S_emotion คือค่าความคล้ายที่คำนวณได้จาก Hellinger Similarity ของแต่ละฟีเจอร์

3.4.3 การพัฒนา API

วิทยานิพนธ์นี้ พัฒนาระบบต้นแบบโดยเชื่อมกับโมเดลที่ผ่านการทดสอบประสิทธิภาพตามแต่ละจุดประสงค์ของโมเดลดังที่กำหนดไว้ในหัวข้อ 3.3 การพัฒนาและทดลองหาโมเดลที่มีประสิทธิภาพดีที่สุด ผ่านการพัฒนา API

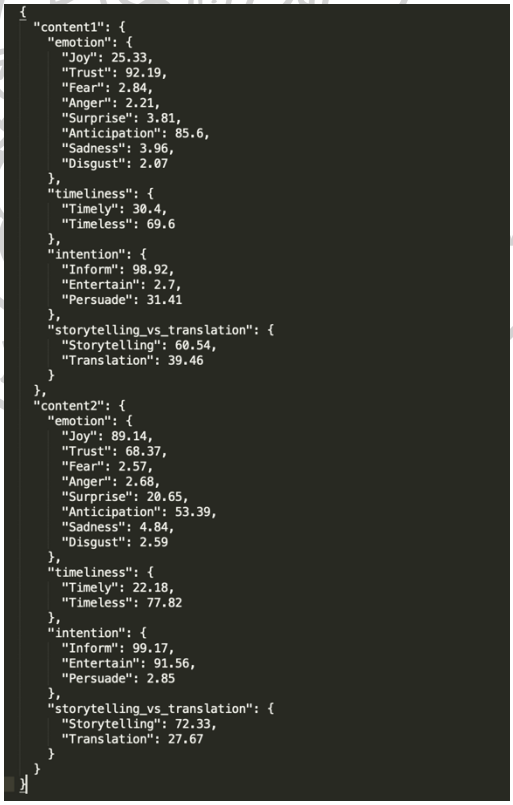
3.4.3.1 กำหนดรูปแบบการรับและส่งข้อมูลระหว่างส่วนนำเข้าข้อมูลบน Web application และ API โดยมีรายละเอียดดังนี้

1) รูปแบบการส่งข้อมูล รูปแบบ JSON หรือ content-type:application/json โดยมีรูปแบบดังภาพที่ 3.8



ภาพที่ 3.8 รูปแบบการส่งข้อมูลแบบ JSON

2) รูปแบบข้อมูลตอบรับจาก API โดยอยู่ในรูปแบบ JSON เช่นกัน ดังปรากฏในภาพที่ 3.9



ภาพที่ 3.9 ตัวอย่าง JSON ที่ส่งออกโดย API

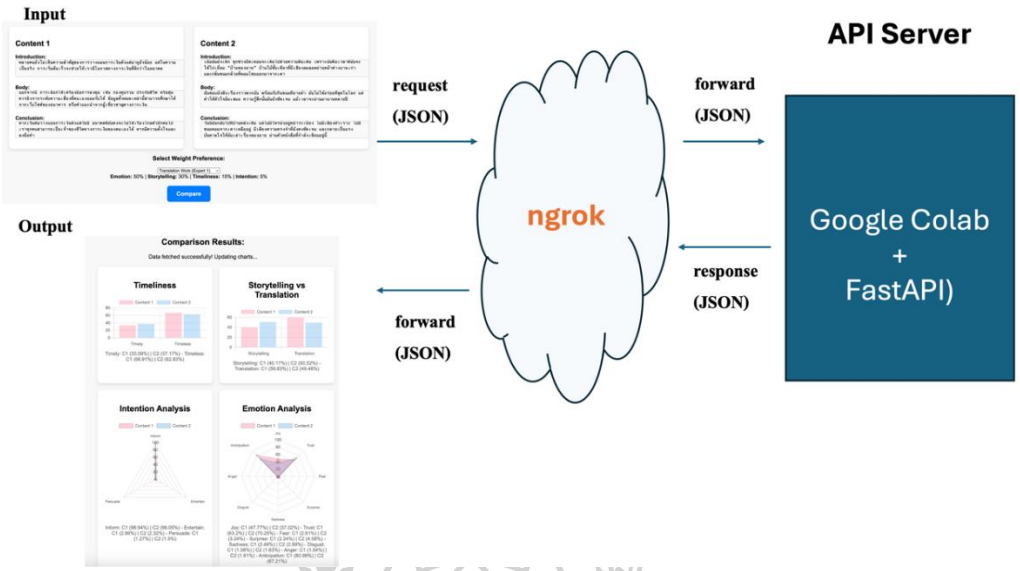
โดยสรุปโครงสร้าง JSON ที่ตอบรับและส่งคืนจาก API ดังตารางที่ 3.9

ตารางที่ 3.9 แสดงรายละเอียดคีย์และข้อมูลใน JSON

คีย์หลักย่อยของ Content	คีย์ย่อย	รายละเอียด
timeliness	Timely, Timeless	จำแนกเนื้อหาว่าเป็นเนื้อหาเฉพาะช่วงเวลา หรือเป็นเนื้อหาที่ยังคงใช้ได้ตลอดเวลา
intention	Inform, Entertain, Persuade	จำแนกเจตนาในการเขียน เช่น ให้ข้อมูล ความบันเทิง หรือเพื่อชักจูง
emotion	Joy, Trust, Fear, Anger, Surprise, Anticipation, Sadness, Disgust	จำแนกอารมณ์จากข้อความ
storytelling_vs_translation	Storytelling, Translation	จำแนกว่าเนื้อหามีลักษณะเป็นการเล่าเรื่อง หรือการแปล/สรุปความ

3.4.3.2 โครงสร้างการทำงานของ API

วิทยานิพนธ์นี้พัฒนา Web API สำหรับรองรับการจำแนกข้อความภาษาไทย โดยใช้เฟรมเวิร์ก FastAPI ซึ่งออกแบบมาให้สามารถทำงานร่วมกับส่วนหน้าของระบบ (frontend) ได้ในลักษณะ asynchronous และ interactive ผ่านการรับ-ส่งข้อมูลในรูปแบบ JSON จาก input และ output บน web application บนเครื่องคอมพิวเตอร์ localhost ไปยัง Server ที่ติดตั้ง API โดยระบบต้นแบบได้ใช้ Google Colab เป็น Server สำหรับประมวลผลโดยใช้ FastAPI ซึ่งการร้องขอและตอบรับ ระหว่าง web application และ API บน Google Colab กระทำผ่าน ngrok ที่ทำหน้าที่เป็นช่องทางส่งต่อ หรือ Tunnel ดังภาพที่ 3.10



ภาพที่ 3.10 โครงสร้างการทำงานระหว่าง Web application และ API

API ดังกล่าวสามารถรับข้อความจำนวนสองชุดในแต่ละครั้ง ซึ่งประกอบด้วยสามส่วน ได้แก่ ส่วนบทนำ ส่วนเนื้อหา และ บทสรุป ต่อหนึ่งชุดข้อมูล โดยข้อมูลที่ส่งเข้ามาจะถูกประมวลผลภายใน API และส่งผลลัพธ์การจำแนกออกมาในหลายมิติ ได้แก่ การจำแนก Timely และ Timeless การจำแนก Intention การจำแนกอารมณ์ การจำแนก Storytelling vs Translation)

โครงสร้างการทำงานของ API ประกอบด้วยขั้นตอนสำคัญ ได้แก่

1) การจัดการคำร้องและการเชื่อมต่อแบบออนไลน์ระหว่าง Web application และ API

API รองรับการร้องขอผ่าน endpoint /api/compare ซึ่งสามารถรับข้อมูลสองชุดพร้อมกัน และส่งคืนผลการเปรียบเทียบของแต่ละชุดโมเดลที่กล่าวข้างต้น นอกจากนี้ วิทยานิพนธ์นี้ ได้ใช้ ngrok เพื่อสร้าง URL แบบสาธารณะ ซึ่งช่วยให้สามารถเข้าถึง API บน Google Colab ได้ ดังภาพที่ 3.11 ภาพ JavaScript ในการส่งข้อมูลจากส่วนนำเข้าข้อมูลไปยัง API server ในรูปแบบ JSON และ ภาพที่ 3.12 แสดงรายละเอียด Python code ที่สร้าง endpoint /api/compare เพื่อประมวลผลทั้ง 4 รูปแบบการประเมินความสอดคล้องเนื้อหา

```
<script>
    let chartInstances = {};

    async function compareText() {
        const intro1 = document.getElementById("intro1").value;
        const body1 = document.getElementById("body1").value;
        const conclusion1 = document.getElementById("conclusion1").value;
        const intro2 = document.getElementById("intro2").value;
        const body2 = document.getElementById("body2").value;
        const conclusion2 = document.getElementById("conclusion2").value;

        if (!intro1 || !body1 || !conclusion1 || !intro2 || !body2 || !conclusion2) {
            document.getElementById("result").innerHTML = "<p style='color: red;'>Please fill in all fields for both
                content sections.</p>";
            return;  // Stops execution if any field is empty
        }

        let response = await fetch("https://2a2f-35-233-181-165.ngrok-free.app/api/compare", {
            method: "POST",
            headers: { "Content-Type": "application/json" },
            body: JSON.stringify({ intro1, body1, conclusion1, intro2, body2, conclusion2 }),
        });

        let textResponse = await response.text();

        try {
            let data = JSON.parse(textResponse);
            document.getElementById("result").innerHTML = "Data fetched successfully! Updating charts...";
            updateCharts(data);
            computeOverallHellingerSimilarity(data.content1, data.content2);
        } catch (jsonError) {
            console.error("Invalid JSON response:", textResponse);
            document.getElementById("result").innerHTML = "<p style='color: red;'>Error: Invalid JSON received from
                API.</p>";
        }
    }
```

ภาพที่ 3.11 JavaScript ข้อมูลจากส่วนนำเข้าข้อมูลไปยัง API server ในรูปแบบ JSON

```
# API Endpoint for Comparing Two Contents in One Request
@app.post("/api/compare")
def compare_texts(input: CompareTextInput):
    text1 = input.intro1 + " " + input.body1 + " " + input.conclusion1
    text2 = input.intro2 + " " + input.body2 + " " + input.conclusion2

    return {
        "content1": {
            "emotion": predict_emotion(text1),  # Added Emotion Prediction
            "timeliness": predict_timeliness(text1),
            "intention": predict_intention(text1),
            "storytelling_vs_translation": predict_storytelling(input.intro1, input.body1, input.conclusion1),
        },
        "content2": {
            "emotion": predict_emotion(text2),  # Added Emotion Prediction
            "timeliness": predict_timeliness(text2),
            "intention": predict_intention(text2),
            "storytelling_vs_translation": predict_storytelling(input.intro2, input.body2, input.conclusion2),
        }
    }

ngrok_tunnel = ngrok.connect(8000)
print("🚀 Public ngrok URL:", ngrok_tunnel.public_url)

nest_asyncio.apply()
uvicorn.run(app, host="0.0.0.0", port=8000)
```

ภาพที่ 3.12 Python code สร้าง endpoint /api/compare

2) การเตรียมและโหลดโมเดลที่ฝึกไว้ล่วงหน้า

ภายในโค้ดมีการโหลดโมเดลที่ได้รับการฝึกบนชุดข้อมูลเฉพาะไว้ล่วงหน้า โดยยังไม่ระบุรายละเอียดของโมเดลที่ใช้สำหรับแต่ละรูปแบบของการประเมินสไตล์ ซึ่ง API ถูกออกแบบให้

สามารถรองรับโมเดลที่มีโครงสร้างแตกต่างกันได้ตามผลลัพธ์ที่ได้จากการทดลอง ด้วยเหตุนี้ ใน API จึงแยกการประมวลผลของแต่ละมิติออกจากกันอย่างเป็นอิสระ และสามารถโหลดโมเดลที่ผ่านการฝึกในแต่ละรูปแบบได้จากไฟล์แยกกันต่างหาก โดยไม่จำกัดว่าจะต้องเป็นโมเดลที่อยู่ในรูปแบบเดียวกัน เช่น บางโมเดลอาจจัดเก็บในรูปแบบ .pkl หากเป็นโมเดลที่ผ่านการฝึกด้วย scikit-learn สำหรับกลุ่มโมเดลเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม อย่างเช่นกลุ่ม Ensemble learning model ขณะที่บางโมเดลอาจอยู่ในรูปแบบ .bin หากเป็นโมเดลจากการฝึกด้วย PyTorch หรือ Hugging Face Transformers ดังภาพที่ 3.13

```python
from fastapi import FastAPI
from fastapi.middleware.cors import CORSMiddleware
from pydantic import BaseModel
import joblib
import numpy as np
import pandas as pd
import re
from pythainlp.tokenize import word_tokenize
from pythainlp.corpus.common import thai_stopwords
import torch
import torch.nn.functional as F
from transformers import AutoTokenizer, AutoModel
import uvicorn
from pyngrok import ngrok
import nest_asyncio
import torch.nn as nn

# ✅ Load Models
MODEL_PATH_STORY = "/content/drive/MyDrive/Research-data/model-api-resources/best_ensemble_model.pkl"  # Storytelling vs. Translation
MODEL_PATH_TIMELY = "/content/drive/MyDrive/Research-data/model-api-resources/best_wangchanberta_model.bin"  # Timely vs. Timeless
MODEL_PATH_INTENTION = "/content/drive/MyDrive/Research-data/model-api-resources/best_intention_model.bin"  # Intention Classification
TFIDF_PATH = "/content/drive/MyDrive/Research-data/model-api-resources/tfidf_vectorizer.pkl"
PCA_PATH = "/content/drive/MyDrive/Research-data/model-api-resources/pca_transform.pkl"
SCALER_PATH = "/content/drive/MyDrive/Research-data/model-api-resources/scaler.pkl"
```

ภาพที่ 3.13 Python code แสดงรายละเอียดเรียกใช้งานแต่ละโมเดลที่ดีที่สุด ใน API

3) การประมวลผลข้อความและการเตรียมฟีเจอร์

เมื่อ API ได้รับ JSON ที่ส่งจาก Web application ผ่าน ngrok จากนั้นจะเข้าสู่ขั้นตอน preprocessing เช่นเดียวกับขั้นตอนการพัฒนาโมเดลจากนั้นจะเข้าสู่กระบวนการแปลงเป็นเวกเตอร์หรือฟีเจอร์ที่ใช้ในการจำแนกสำหรับแต่ละโมเดล ดังรายละเอียดใน Python code ดังภาพที่ 3.14

```python
class CompareTextInput(BaseModel):
    intro1: str
    body1: str
    conclusion1: str
    intro2: str
    body2: str
    conclusion2: str

# ✅ Preprocessing Functions
def preprocess_text(text):
    text = text.lower()
    text = re.sub(r"[^\n-๙\s]", "", text)
    text = re.sub(r"[a-zA-Z\u4e00-\u9fff]", "", text)
    text = re.sub(r"\d+", "", text)
    text = re.sub(r"http\S+|www\S+", '', text, flags=re.MULTILINE)
    text = re.sub(r"#\S+", '', text)
    tokens = word_tokenize(text)
    tokens = [word for word in tokens if word not in thai_stopwords()]
    return ' '.join(tokens)

# ✅ Feature Extraction for Storytelling vs. Translation
def extract_features(introduction, body, conclusion):
    full_text = preprocess_text(introduction) + " " + preprocess_text(body) + " " + preprocess_text(conclusion)
    tfidf_features = tfidf_vectorizer.transform([full_text]).toarray()

    # ✅ Convert engineered features into DataFrame
    engineered_features = pd.DataFrame([[
        len(introduction.split()),
        len(body.split()),
        len(conclusion.split())
    ]], columns=["intro_length", "body_length", "conclusion_length"])
```

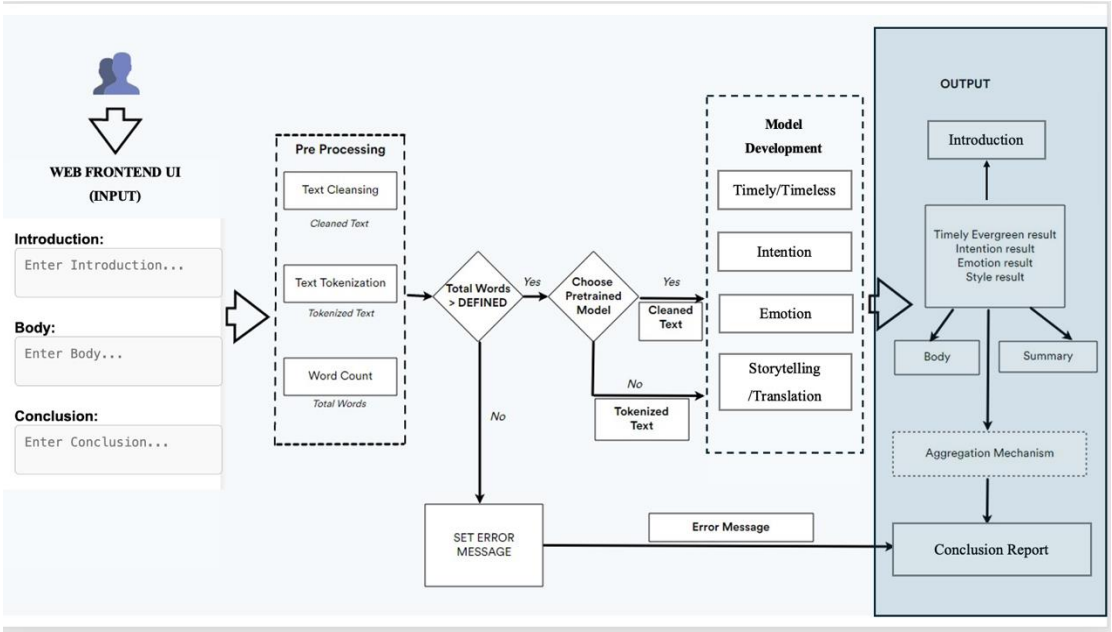
ภาพที่ 3.14 Python code แสดงรายละเอียดขั้นตอน preprocessing ใน API

4) การประเมินความน่าจะเป็นในแต่ละโมเดล

API จะประเมินข้อความแต่ละชุดโมเดล ทั้ง 4 โมเดล โดยโมเดล Timeliness และ Storytelling/Translation จะคืนค่าความน่าจะเป็นของแต่ละคลาส ด้วยผลรวม 100% ขณะที่โมเดล Intention และ Emotion ซึ่งเป็นลักษณะหลายป้ายกำกับ ผลรวมของคลาสจะไม่เท่ากับ 100% เนื่องจากบางคลาสอาจมีความน่าจะเป็นใกล้เคียงกัน โดยจะส่งผลลัพธ์กลับไปยัง Web application ในรูปแบบ JSON

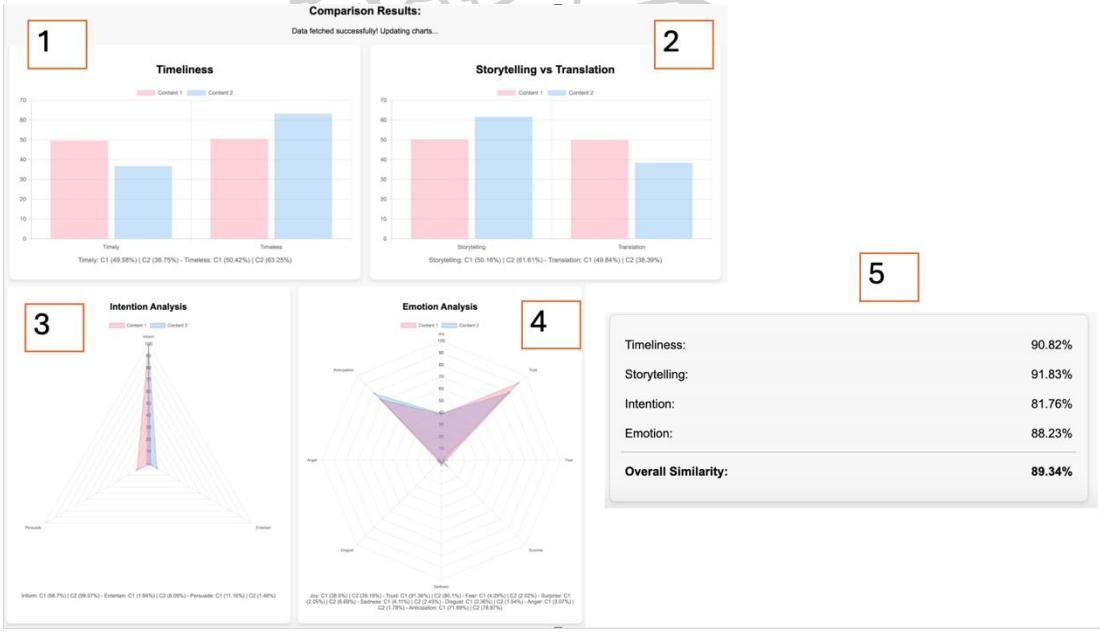
3.4.4 การพัฒนาส่วนแสดงผลลัพธ์ (Output)

สำหรับส่วนแสดงผลลัพธ์ จะเป็นลักษณะของเว็บแอปพลิเคชันเช่นเดียวกับส่วนนำเข้าข้อมูล โดยเมื่อนำเข้าข้อมูล และกดปุ่มเพื่อดำเนินการประเมินผลลัพธ์ด้วยโมเดลทั้ง 4 โมเดลที่อธิบายไปข้างต้น ระบบจะแสดงผลลัพธ์ ออกมา โดยมีกระบวนการทำงานดังภาพที่ 3.15 ซึ่งเป็นเส้นทางการทำงานตั้งแต่กระบวนการ Pre Processing จนถึงกระบวนการแสดงผลลัพธ์



ภาพที่ 3.15 กระบวนการทำงานแสดงผลลัพธ์ ในภาพรวมตั้งแต่รับข้อมูล

ผลลัพธ์จะถูกแสดงผล ดังภาพที่ 3.16 โดยมีรายละเอียดของแต่ละส่วน ดังนี้



ภาพที่ 3.16 ส่วนแสดงผลลัพธ์ (Output)

1) ส่วนหมายเลข 1

แสดงกราฟแท่งเปรียบเทียบคลาส Timely และ Timeless ในฟีเจอร์ Timeliness เพื่อแสดงค่าความน่าจะเป็นในการเป็นสไตล์เนื้อหาตามกระแส หรือ เนื้อหาสดใหม่เสมอ สำหรับสองบทความ

2) ส่วนหมายเลข 2

แสดงกราฟแท่งเปรียบเทียบคลาส Translation และ Storytelling เพื่อแสดงความน่าจะเป็นในการเป็นเนื้อหางานแปล หรือแบบเล่าเรื่อง สำหรับสองบทความ

3) ส่วนหมายเลข 3

เนื่องจากเป็นโมเดลจำแนกแบบหลายป้ายกำกับ ดังนั้นจึงแสดงด้วยแผนภูมิเรดาร์ (Radar chart) เพื่อแสดงความน่าจะเป็นของคลาส Inform Persuade และ Entertain ในฟีเจอร์ Intention จุดประสงค์การสื่อความ โดยเปรียบเทียบทั้งสองบทความในแผนภูมิเดียวกัน โดยแบ่งแยกสี เพื่อให้เห็นการซ้อนทับและความแตกต่างของสไตล์

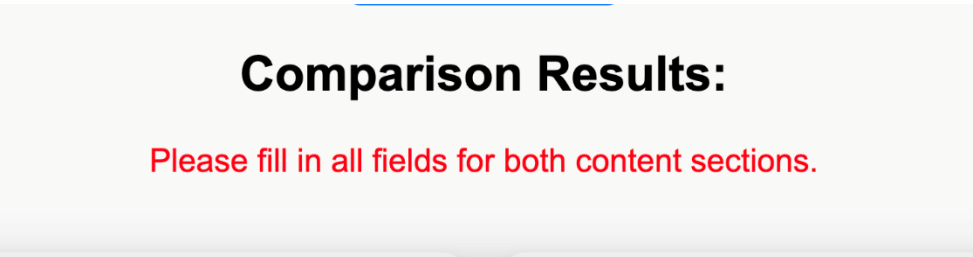
4) ส่วนหมายเลข 4

เนื่องจากเป็นโมเดลจำแนกแบบหลายป้ายกำกับ ดังนั้นจึงแสดงด้วยแผนภูมิเรดาร์ (Radar chart) เพื่อแสดงความน่าจะเป็นของคลาสอารมณ์ทั้งแปดอารมณ์ โดยเปรียบเทียบทั้งสองบทความในแผนภูมิเดียวกัน โดยแบ่งแยกสี เพื่อให้เห็นการซ้อนทับและความแตกต่างของสไตล์ ซึ่งการแสดงผลป้ายกำกับคลาสบนแผนภูมิเรดาร์ จะสอดคล้องกับคู่อารมณ์ของพลัทชิค ได้แก่ ความรื่นเริง (Joy) คู่ตรงข้ามกับ ความเศร้า (Sadness) ความโกรธ (Anger) คู่ตรงข้ามกับ ความกลัว (Fear) ความไว้เนื้อเชื่อใจ (Trust) คู่ตรงข้ามกับ ความรังเกียจ (Disgust) และ ความประหลาดใจ (Surprise) คู่ตรงข้ามกับ ความคาดหวัง (Anticipation)

5) ส่วนหมายเลข 5

แสดงเปอร์เซ็นต์ความคล้ายของรูปแบบทั้งสองบทความ ในสเกล 0 ถึง 100 เปอร์เซ็นต์ ตามรายละเอียดที่ระบุในการพัฒนาส่วนผลรวมผลทำนายจากโมเดลเพื่อวิเคราะห์ความสอดคล้องของสไตล์ (Aggregation Mechanism)

สำหรับในกรณีที่เกิดข้อผิดพลาดเกิดขึ้นในขั้นตอนการนำเข้าข้อมูลและการจัดการข้อมูลก่อนเข้าสู่โมเดล ดังรายละเอียดในหัวข้อการพัฒนาส่วนจัดการข้อมูลก่อนเข้าสู่โมเดล (Pre Processing) ส่วนแสดงผลลัพธ์จะปรากฏข้อความแจ้งข้อผิดพลาด ดังภาพที่ 3.17 ส่วนแสดงผลลัพธ์กรณีเกิดข้อผิดพลาด (Output with Error Message)



ภาพที่ 3.17 ส่วนแสดงผลลัพธ์กรณีเกิดข้อผิดพลาด (Output with Error Message)

โดยส่วนแสดงผลจะทำหน้าที่แสดงผลลัพธ์ที่ได้จากการประมวลผลด้วยฟังก์ชัน JavaScript compareText() ซึ่งจะถูกเรียกใช้เมื่อผู้ใช้กดปุ่ม Compare ในส่วนของการรับข้อมูลข้อความจากผู้ใช้ โดยฟังก์ชันนี้จะรับค่าจาก input ทั้ง 6 ช่อง ได้แก่ intro1, body1, conclusion1, intro2, body2 และ conclusion2 ซึ่งเป็นข้อความจากสองบทความที่ต้องการเปรียบเทียบ จากนั้นจะตรวจสอบว่าผู้ใช้กรอกข้อมูลครบถ้วนในทุกช่องหรือไม่ หากไม่ครบ ระบบจะแจ้งเตือนและยุติกระบวนการก่อนส่งข้อมูล

หากข้อมูลครบถ้วน ฟังก์ชัน compareText() จะส่งข้อมูลดังกล่าวไปยัง API ของโมเดลที่ได้รับการคัดเลือกว่ามีประสิทธิภาพดีที่สุดจากการทดลองที่ 1 ถึงการทดลองที่ 4 ซึ่งประกอบด้วยโมเดลในกลุ่ม Ensemble Learning ได้แก่ Random Forest, XGBoost, LightGBM และ AdaBoost โดยใช้บริการ ngrok เพื่อเชื่อมต่อกับโมเดลที่รันอยู่บน Google Colab ด้วยภาษา Python ให้สามารถเข้าถึงได้จากเครือข่ายภายนอกในลักษณะออนไลน์ทั่วโลก และเมื่อ API ประมวลผลเสร็จจะคืนค่าผลลัพธ์กลับมาในรูปแบบของข้อมูล JSON

หลังจากได้รับผลลัพธ์ JSON แล้ว ระบบจะเรียกใช้งานฟังก์ชัน updateCharts() ซึ่งทำหน้าที่แสดงผลลัพธ์ที่ได้ในรูปแบบกราฟหรือแผนภูมิตามประเภทที่กำหนดไว้ โดยจะแสดงผลเป็นกราฟแท่ง (Bar Chart) สำหรับมิติ Timeliness และ Storytelling/Translation และจะแสดงผลเป็นกราฟเรดาร์ (Radar Chart) สำหรับมิติ Intention และ Emotion Analysis โดยการแสดงผลทั้งหมดนี้จะเกิดขึ้นแบบอินเทอร์แอคทีฟภายในหน้าเว็บเดียวกัน โดยไม่จำเป็นต้องรีเฟรชหรือโหลดหน้าเว็บใหม่

3.5 การทดสอบระบบโดยผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับข้อมูล

3.5.1 โดยผู้เชี่ยวชาญ

ในการประเมินความพึงพอใจและความเป็นไปได้ในการนำระบบต้นแบบไปใช้งานจริง ได้มีการสัมภาษณ์และทดสอบระบบกับผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับเนื้อหาในชุดข้อมูล โดยดำเนินการ

ผ่านช่องทางออนไลน์ ได้แก่ แอปพลิเคชัน Zoom, Google Meet และ Messenger Call เพื่อให้เหมาะสมกับบริบทของการทำงานจริงและสะดวกต่อผู้เข้าร่วม โดยมีคำถามสำคัญที่ใช้ในการสัมภาษณ์และเก็บข้อมูล ดังนี้

3.5.1.1 คำถามด้านความพึงพอใจในการใช้งานระบบ

1) หลังจากได้ทดลองใช้งานระบบต้นแบบ ท่านมีความคิดเห็นอย่างไรเกี่ยวกับความสะดวกในการใช้งาน ความชัดเจนของผลลัพธ์ และรูปแบบการแสดงผลโดยรวม

2) หากให้คะแนนความพึงพอใจจากคะแนนเต็ม 5 (1 = น้อยที่สุด, 5 = มากที่สุด) ท่านให้คะแนนเท่าใด

3.5.1.2 คำถามด้านความเป็นไปได้ในการนำไปใช้งานจริง

1) ผลลัพธ์ที่ระบบวิเคราะห์มีความสอดคล้องกับเนื้อหาและลักษณะงานของท่านหรือไม่

2) ท่านคิดว่าระบบนี้สามารถนำไปประยุกต์ใช้ในกระบวนการทำงานจริงของท่านได้ในลักษณะใดบ้าง

3) หากให้คะแนนความเป็นไปได้ในการนำไปใช้งานจริงจากคะแนนเต็ม 5 (1 = น้อยที่สุด, 5 = มากที่สุด) ท่านให้คะแนนเท่าใด

3.5.2 แบบสอบถามแบบสอบถามความพึงพอใจและความเป็นไปได้ในการใช้งานระบบ

วิทยานิพนธ์ฉบับนี้ได้จัดทำแบบสอบถามขึ้นเพื่อใช้ในการประเมิน ระดับความพึงพอใจ และความเป็นไปได้ในการนำระบบวิเคราะห์และเปรียบเทียบสไตล์ข้อความภาษาไทย ไปใช้งานในบริบทของ งานด้าน Content Marketing และสื่อออนไลน์ โดยกลุ่มเป้าหมายที่ตอบแบบสอบถามทั้งหมดเป็นกลุ่มผู้มีความเกี่ยวข้องกับงานเขียนและการผลิตเนื้อหาในช่องทางสื่อดิจิทัล เช่น นักเขียนบทความ นักการตลาดดิจิทัล บรรณาธิการ เจ้าของสื่อ Blogger และ Content Creator ซึ่งเป็นผู้มีบทบาทสำคัญในกระบวนการผลิตและสื่อสารเนื้อหาผ่านแพลตฟอร์มออนไลน์ อาทิ เว็บไซต์ เพจ Facebook หรือแพลตฟอร์มขององค์กร

โดยกลุ่มตัวอย่างทั้งหมดเป็นผู้ที่ ไม่เคยมีประสบการณ์ใช้งานระบบมาก่อน จึงสามารถสะท้อนความคิดเห็นจากมุมมองของ “ผู้ใช้งานใหม่” ได้อย่างแท้จริง โดยเฉพาะในแง่ของการเข้าถึงระบบ การตีความผลลัพธ์ และการประเมินศักยภาพของระบบในบริบทการทำงานจริง

โดยแบ่งโครงสร้างแบบสอบถาม ดังรายละเอียดต่อไปนี้

3.5.2.1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

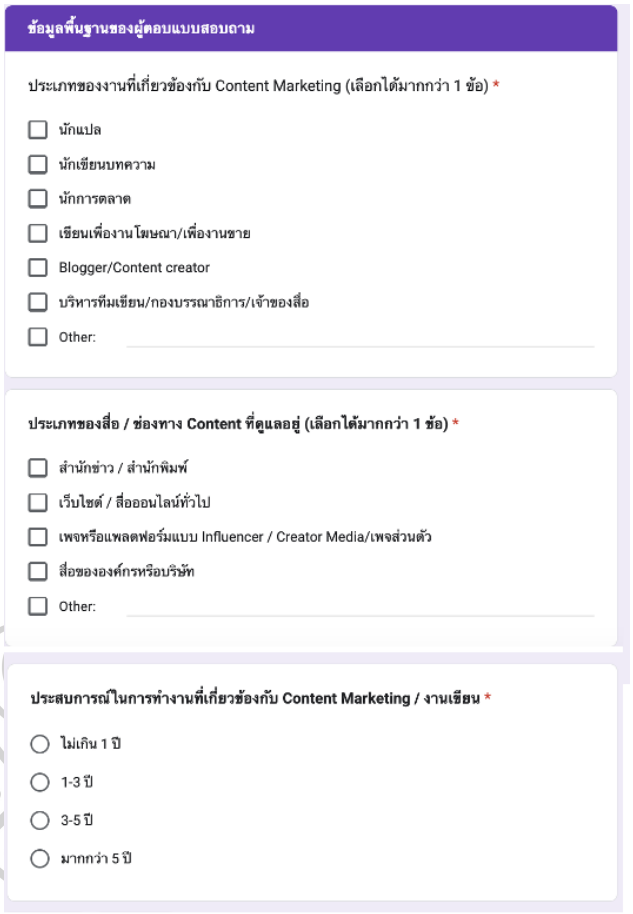
ประกอบด้วยคำถามแบบเลือกตอบเกี่ยวกับลักษณะงานของผู้ตอบ ได้แก่

1) ประเภทของงานที่เกี่ยวข้องกับ Content Marketing (เลือกได้มากกว่า 1 ข้อ)

2) ประเภทของช่องทางหรือแพลตฟอร์ม Content ที่ใช้งานอยู่

3) ระยะเวลาประสบการณ์ในการทำงานด้าน Content Marketing / งานเขียน

ดังปรากฏในภาพที่ 3.18



ภาพที่ 3.18 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

3.5.2.2 แบบประเมินความพึงพอใจและการใช้งานระบบ

ผู้ตอบแบบสอบถามให้คะแนนในระดับ 1 ถึง 5 โดย 1 หมายถึง เห็นด้วย“น้อยที่สุด” และ 5 หมายถึง “มากที่สุด” ในหัวข้อต่อไปนี้

1) ความสะดวกในการใช้งาน

พิจารณาว่า ระบบใช้ง่ายหรือซับซ้อนเพียงใด

2) ความชัดเจนของผลลัพธ์ที่แสดง

พิจารณาว่า การแสดงผลเข้าใจง่ายหรือไม่

3) รูปแบบการแสดงผลโดยรวม

พิจารณาว่า กราฟ ตาราง หรือข้อมูลที่เหมาะสมหรือไม่

4) ความยาก-ง่ายในการใช้งาน

พิจารณาว่า การเข้าถึงฟังก์ชันต่างๆ และการเรียนรู้ระบบ

5) ความถูกต้องของผลลัพธ์เทียบกับความคาดหวัง

พิจารณาว่า ผลการวิเคราะห์ตรงกับความรู้สึกหรือประสบการณ์จริงของผู้ใช้หรือไม่

6) ระดับความพึงพอใจโดยรวม

พิจารณาว่า เมื่อมองภาพรวม ผู้ตอบแบบสอบถามมีระดับความพึงพอใจในระดับใด

3.5.2.2 การประเมินศักยภาพความเป็นไปได้ในการประยุกต์ใช้จริง

พิจารณาความเป็นไปได้ในการประยุกต์ใช้จริง หลังจากผู้ตอบแบบสอบถามได้ทดลองใช้ระบบจริง โดยให้คะแนนในระดับ 1 ถึง 5 โดย 1 หมายถึง เห็นด้วย“น้อยที่สุด” และ 5 หมายถึง “มากที่สุด” ในหัวข้อต่อไปนี้

1) ความสอดคล้องของระบบกับเนื้อหาและลักษณะงาน

2) แนวทางที่ระบบสามารถประยุกต์ใช้ในการทำงานจริง

3) ระดับความเป็นไปได้ในการนำไปใช้ในงานจริง

มหาวิทยาลัยศิลปากร

บทที่ 4
ผลการดำเนินงานวิจัย

4.1 ผลการทดลองที่ 1 โมเดลการจำแนกข้อความ Timely และ Timeless

จากผลการทดลองที่นำเสนอในตารางที่ 4.1 ซึ่งเปรียบเทียบประสิทธิภาพของโมเดลในกลุ่ม CNN และ Transformer สำหรับงานจำแนกข้อความระหว่างกลุ่ม Timely และ Timeless แสดงให้เห็นว่าทุกโมเดลมีประสิทธิภาพในแง่ของ Accuracy สูงเกิน 80% ตามที่กำหนดไว้ในสมมติฐาน

โดยโมเดลที่ใช้สถาปัตยกรรมแบบ CNN เมื่อทำงานร่วมกับ embedding จาก fastText แบบ Skip-gram มีประสิทธิภาพสูงในการจำแนกข้อความ โดยโมเดล CNN ที่ใช้ fastText Skip-gram มี Accuracy บนชุดทดสอบสูงถึง 90.00% และค่า F1-score อยู่ที่ 89.93% ซึ่งถือว่าสูงที่สุดในกลุ่มโมเดล CNN ทั้งหมด ขณะที่โมเดล CNN ที่ใช้ embedding จาก Skip-gram แบบทั่วไป มี Accuracy และค่า F1-score ต่ำกว่าขณะเทรนเล็กน้อย โดยมีค่าอยู่ที่ 88.33% และ 88.28% ตามลำดับ ส่วน CNN1D แบบดั้งเดิม มีค่า Accuracy และ F1-score ต่ำสุดในกลุ่ม CNN แต่มีค่าทั้งสองสูงขึ้นบนชุดทดสอบ ที่ 89.17% และ 89.16% ตามลำดับ

เมื่อพิจารณากลุ่มโมเดล Transformer พบว่า WangchanBERTa เป็นโมเดลที่ให้ผลลัพธ์ดีที่สุด โดยมี Accuracy บนชุดทดสอบสูงถึง 93.00% และ F1-score อยู่ที่ 92.00% ซึ่งเหนือกว่าโมเดล mBERT และ RoBERTa อย่างชัดเจน แสดงให้เห็นถึงความสำคัญของการใช้โมเดลที่ถูกฝึกด้วยข้อมูลภาษาไทยโดยเฉพาะ

ในทางตรงกันข้าม โมเดล mBERT ซึ่งเป็นโมเดล multilingual ที่ไม่ได้ถูกฝึกเฉพาะกับข้อมูลภาษาไทย ให้ผลลัพธ์ต่ำที่สุดในกลุ่ม Transformer โดยมี Accuracy และ F1-score อยู่ที่เพียง 80.00% ซึ่งสะท้อนข้อจำกัดของการใช้โมเดลข้ามภาษากับงานที่ต้องอาศัยความเข้าใจเชิงลึกในภาษาไทยอย่างเฉพาะเจาะจง สำหรับ RoBERTa ซึ่งเป็นโมเดล pretrained ที่มีโครงสร้างใกล้เคียงกับ BERT แต่ได้รับการฝึกด้วย corpus ที่กว้างและเข้มข้นกว่า ให้ผลลัพธ์ใกล้เคียงกับ CNN ที่ใช้ fastText โดยมี Accuracy และ F1-score อยู่ที่ 90.00%

อย่างไรก็ตาม เมื่อเปรียบเทียบกับ WangchanBERTa ซึ่งถูกออกแบบมาสำหรับภาษาไทยโดยตรง จะเห็นได้ว่า RoBERTa ยังมีข้อจำกัดในการเรียนรู้บริบทของภาษาไทยเช่นเดียวกับ mBERT

ตารางที่ 4.1 ตารางแสดงผลการทดลองที่ 1 โมเดลการจำแนกข้อความ Timely และ Timeless

Model	Accuracy (Validation)	Accuracy (Test)	Precision (Test)	Recall (Test)	F1-Score (Test)
CNN Conv1D	86.80%	89.17%	89.26%	89.17%	89.16%
CNN Conv1D Skip-gram	88.89%	88.33%	89.03%	88.33%	88.28%
CNN fastText Skip-gram	88.19%	90.00%	91.14%	90.00%	89.93%
BERT	81.00%	80.00%	80.00%	80.00%	80.00%
RoBERTa	88.89%	90.00%	90.00%	90.00%	90.00%
WangchanBERTa	**93.06%**	**93.00%**	**93.00%**	**93.00%**	**92.00%**

4.2 ผลการทดลองที่ 2 โมเดลจำแนกเจตนาการสื่อความของเนื้อหา

การทดลองในส่วนนี้ได้เปรียบเทียบประสิทธิภาพของโมเดล Transformer 3 แบบ คือ mBERT, WangchanBERTa และ PhayaThaiBERT โดยได้ทดลองทั้งแบบไม่ใช้และแบบที่มีการเพิ่มบริบททางอารมณ์ เข้าไป เพื่อศึกษาว่าคุณลักษณะด้านอารมณ์จะช่วยพัฒนาความสามารถของโมเดลในการจำแนกเจตนาได้ดีขึ้นหรือไม่

4.2.1 ผลการทดลองของโมเดลที่ไม่ใช้ฟีเจอร์อารมณ์

ผลจากการทดลองดังปรากฏในตารางที่ 4.2 พบว่า โมเดล PhayaThaiBERT มีประสิทธิภาพสูงที่สุดในทุกตัวชี้วัดทั้ง Precision, Recall, F1-score และมี Hamming Loss ต่ำที่สุดทั้งในชุดข้อมูล Validation และ Test รองลงมาคือ WangchanBERTa และ mBERT ตามลำดับ โดยเฉพาะในคลาส Entertain และ Persuade โมเดล PhayaThaiBERT ทำได้ดีกว่าอย่างชัดเจน แสดงให้เห็นถึงความสามารถของโมเดลที่ถูกออกแบบมาเฉพาะภาษาไทย โดยพบว่าคลาส Entertain มีคะแนนต่ำที่สุดในทุกโมเดล ซึ่งสะท้อนว่าเป็นคลาสที่มีความท้าทายในการจำแนก แต่โมเดลภาษาไทยโดยเฉพาะเช่น WangchanBERTa และ PhayaThaiBERT สามารถจำแนกได้ดีกว่า mBERT อย่างชัดเจน

ตารางที่ 4.2 ผลการทดลองที่ 2 โมเดลจำแนกการสื่อความของเนื้อหาแบบไม่ใช้ฟีเจอร์อารมณ์

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT	Inform	94.52%	98.22%	96.34%	16.57%

Model	Class	Precision	Recall	F1-Score	Hamming Loss
(Validation)	Entertain	61.43%	55.84%	58.50%	
	Persuade	76.65%	73.56%	75.07%	
	Micro	84.50%	84.02%	84.26%	
	Macro	77.53%	75.88%	76.64%	
mBERT (Test)	Inform	96.73%	96.34%	96.54%	16.79%
	Entertain	68.42%	65.82%	67.10%	
	Persuade	75.44%	65.65%	70.20%	
	Micro	86.21%	82.24%	84.18%	
	Macro	80.20%	75.94%	77.95%	
WangchanBERTa (Validation)	Inform	92.28%	97.86%	94.99%	15.18%
	Entertain	75.71%	68.83%	72.11%	
	Persuade	75.14%	78.16%	76.62%	
	Micro	84.52%	87.22%	85.85%	
	Macro	81.04%	81.62%	81.24%	
WangchanBERTa (Test)	Inform	94.88%	97.97%	96.40%	17.98%
	Entertain	74.29%	65.82%	69.80%	
	Persuade	67.48%	63.36%	65.35%	
	Micro	84.12%	82.46%	83.28%	
	Macro	78.88%	75.72%	77.18%	
PhayaThaiBERT (Validation)	Inform	95.53%	98.93%	97.20%	**11.21%**
	Entertain	75.95%	77.92%	76.92%	
	Persuade	81.22%	84.48%	82.82%	
	Micro	88.02%	91.17%	**89.57%**	
	Macro	84.23%	87.11%	**85.65%**	
PhayaThaiBERT (Test)	Inform	97.23%	100.00%	98.60%	**12.14%**
	Entertain	80.00%	75.95%	77.92%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Persuade	78.69%	73.28%	75.89%	
	Micro	**89.33%**	**88.16%**	**88.74%**	
	Macro	**85.31%**	**83.08%**	**84.14%**	

4.2.2 ผลการทดลองของโมเดลที่ใช้ฟีเจอร์อารมณ์

หลังจากเพิ่มบริบททางอารมณ์ในการจำแนกเจตนาของข้อความ พบว่าโมเดลทั้งสาม (mBERT, WangchanBERTa และ PhayaThaiBERT) มีประสิทธิภาพที่ดีขึ้นอย่างชัดเจน โดยเฉพาะในคลาสที่เดิมมีความท้าทายในการจำแนกมากที่สุด เช่น Entertain และ Persuade ดังที่แสดงไว้ในตารางที่ 4.3 ซึ่งสะท้อนว่าการเพิ่มบริบททางอารมณ์ช่วยให้โมเดลจำแนกได้แม่นยำมากขึ้น โดยเฉพาะในคลาสที่มีคะแนนต่ำก่อนหน้านี้

โมเดล PhayaThaiBERT ยังคงมีประสิทธิภาพดีที่สุดในทุกตัวชี้วัดโดยรวม ทั้ง Hamming Loss, Macro F1 และ Micro F1 อย่างไรก็ตาม ประสิทธิภาพของ PhayaThaiBERT ในตอนนี้กลับใกล้เคียงกับ WangchanBERTa มากขึ้น โดยเฉพาะในชุดข้อมูลทดสอบ (Test set) พบว่า WangchanBERTa มีประสิทธิภาพดีกว่าผลจากการฝึก (Training set) ของตัวเอง ซึ่งแตกต่างจากแนวโน้มที่พบในอีกสองโมเดล

ประเด็นที่น่าสนใจคือคะแนน Macro F1 ซึ่งเป็นตัวชี้วัดที่คำนึงถึงปัญหาข้อมูลไม่สมดุล พบว่า WangchanBERTa ได้คะแนน Macro F1 สูงกว่า PhayaThaiBERT เล็กน้อยในชุดข้อมูลทดสอบ ซึ่งแสดงว่า WangchanBERTa จัดการกับข้อมูลที่มีความไม่สมดุลของแต่ละคลาสได้ดีกว่าเล็กน้อย แม้ว่าโดยรวมแล้ว PhayaThaiBERT จะยังคงมีประสิทธิภาพดีที่สุด แต่ผลที่ได้นี้แสดงให้เห็นว่า WangchanBERTa มีศักยภาพที่น่าสนใจ โดยเฉพาะในด้านการปรับใช้กับข้อมูลใหม่ที่ไม่เคยพบมาก่อน (generalization)

ผลลัพธ์เหล่านี้เน้นให้เห็นว่าการนำบริบททางอารมณ์มาใช้มีประโยชน์อย่างชัดเจน โดยช่วยให้โมเดลมีประสิทธิภาพดีขึ้นและลดข้อผิดพลาดในการจำแนกประเภทของข้อความ โดยเฉพาะในคลาสที่เคยจำแนกได้ไม่ดีนักก่อนหน้านี้ และแม้ว่า PhayaThaiBERT จะยังคงดีที่สุดโดยรวม แต่การพัฒนาของ WangchanBERTa โดยเฉพาะในคะแนน Macro F1 ในชุดข้อมูลทดสอบ ก็แสดงให้เห็นถึงความสามารถที่ดีในการจัดการกับข้อมูลที่มีความไม่สมดุล

เมื่อวิเคราะห์การทำนายผิดพลาดของโมเดล PhayaThaiBERT โดยใช้ฟีเจอร์อารมณ์ ที่ให้ประสิทธิภาพที่ดีที่สุด ของชุดข้อมูลแบบดั้งเดิมที่ยังไม่มีการปรับสมดุลของจำนวนคลาส ตามปรากฏในภาพที่ 4.1 พบว่า โมเดลมีแนวโน้มที่จะทำนายไม่ถูกต้องในหลายกรณี ทั้งในลักษณะที่ทำนายป้ายกำกับไม่ครบถ้วน และการทำนายเกินความจำเป็น โดยข้อผิดพลาดที่พบบ่อยที่สุดคือ การที่โมเดลมีแนวโน้มจะให้ป้ายกำกับเฉพาะ Inform เพียงอย่างเดียว แม้ว่าในความเป็นจริง ข้อความเหล่านั้นจะมีลักษณะของการให้ความบันเทิงหรือมีเจตนาในการโน้มน้าวด้วย ตัวอย่างเช่น ข้อความที่กล่าวถึงสถานที่ท่องเที่ยวหรือวัฒนธรรมในประเทศจีน ซึ่งมีทั้งเนื้อหาข้อมูลและการใช้ภาษาที่ชักชวนให้ผู้อ่านสนใจ แต่มักถูกโมเดลตีความว่าเป็นเพียงข้อความเชิงสาระเท่านั้น สะท้อนให้เห็นว่าโมเดลอาจมีอคติในการเลือกป้ายกำกับ inform มากเกินไป หรือยังไม่สามารถเข้าใจบริบทของภาษาที่ซับซ้อนในเชิงอารมณ์ ความสนุก หรือความโน้มน้าวได้ดีพอ

text	true_labels	predicted_labels
กาแฟ ลาเต้ หมี แพนด้า เคเอฟซี ประเทศ จีน การตลาด จีน การนำ ความน่ารัก สัตว์ ผสมผสาน ผลิตภัณฑ์ บริการ ตึงตูด ครองใจ ผู้บริโภค แก้ว อุ้งเท้า แมว พลุ แตก ทำ แก้ว อุ้งเท้า หมา ตามมา วิเคราะห์ แคมเปญ แก้ว อุ้งเท้า แมว อุ้งเท้า หมา พูดถึง ค้น กา ข้อมูล โลก ออนไลน์ จีน สรุป คนจีน พูดถึง แมว อันดับ โดยเฉพาะ สาว แมว น่ารัก หมา สัตว์เลี้ยง คนจีน ชอบ พูดถึง สำหรับ หมี แพนด้า เป็นหนึ่ง สัตว์ สัญลักษณ์ ประจำ ประเทศ จีน คนจีน ภาพ เชิง บวก หมี แพนด้า ความน่ารัก น่าเอ็นดู น่าแปลกใจ ทำ ลาเต้ อาร์ต หมี แพนด้า อ้าย อ่าน ความคิดเห็น คนจีน ลิ้มลอง กาแฟ หมี แพนด้า เคเอฟซี โพสต์ แบ่ง ชอบ น่ารัก รสชาติ ไม่เลว ดื่ม ชม ชอบ ลาย เหมือน หมี แพนด้า แก้ว ดู ลาย ขึ้นอยู่กับ คน ทำ แต่ละคน บางคน ตอนที่ แก้ว ลาย เหมือน แพนด้า เตี้ยวเตียว ลาย เปลี่ยนไป ที่อยู่ จีน อย่า ลืม ลอง แชร์ รสชาติ ชอบ ชอบ	1, 2, 3	1, 3
ยล ทิวทัศน์ ธรรมชาติ แม่น้ำ ยาร์ ลุง ซัง โป มหา นที ดินแดน หลังคา โลก ชม ความงดงาม ภาพ มุม แม่น้ำ ยาร์ ลุง ซัง โป แม่น้ำ พรหมบุตร เมือง ชาน หนาน เขตปกครองตนเอง ทิเบต ตะวันตกเฉียงใต้ จีน เป็นหนึ่ง แหล่งน้ำ เขตแดน สาย ทวีปเอเชีย สำหรับ แหล่งท่องเที่ยว ห้าม พลาด ยาร์ ลุง ซัง โป แกรนด์ แคนยอน หุบเขา ลึก โลก ขึ้นทะเบียน แหล่งท่องเที่ยว ระดับ	1, 3	1
ปักกิ่ง คาเฟ่ ธีม แพนด้า กอด แพนด้า จำลอง น่ารัก จนใจ ละลาย	1, 2	1, 3
เรื่อง น่าสนใจ ในประเทศ จีน เดินทาง การนัด ตรวจ โรงพยาบาล ชำระเงิน สแกน ชีวิตประจำวัน รูปแบบ ต่าง ๆ ล้วนแล้วแต่ การนำ เทคโนโลยี ประโยชน์ ทันสมัย แน่นอน ที่จะ เข้าถึง เทคโนโลยีสมัยใหม่ โดยเฉพาะ ผู้สูงอายุ สถานการณ์ อุปกรณ์ อัจฉริยะ สาม กีดกัน ออกจาก ยุค ดิจิทัล ตัว ปัญหา ทางการ จีน นึงนอนใจ สั่งการ สำนักงาน กิจการ พลเรือน นคร หนาน จิง พิธี ชั้นเรียน สมาร์ท โฟน สำหรับ ผู้สูงอายุ หมิ่น ชื่อว่า สอน วิธีใช้ โทรศัพท์ ท่าน เชต ฉิน หวาย ทำการ ฝึกอบรม สมาร์ท โฟน ผู้สูงอายุ ชุด จำนวน หมิ่น คน ทั่วเมือง ดำเนิน มาตรการ รูปธรรม ขอความร่วมมือ ศูนย์ บริการ ดูแล ผู้สูงอายุ บ้าน ศูนย์ บริการ เบ็ดเสร็จ ชุมชน จุด ฝึกอบรม สมาร์ท โฟน สำหรับ ผู้สูงอายุ จำนวน ฝึกอบรม ผู้สูงอายุ สมัครใจ ค่าใช้จ่าย บริเวณ ที่พัก อาศัยที่ สะดวกสบาย	1, 3	1
ตะลุย งาน จัดแสดง น้ำแข็ง หิมะ โลก ส่วนหนึ่ง เทศกาล น้ำแข็ง แกะสลัก หิมะ นานาชาติ เข้ามา งาน เจอ หอคอย น้ำแข็ง ประโยค น่าสนใจ รูป ประโยค ประธานาธิบดี สี จิ้น ผิง มีความหมาย ประชาชน ช่วยกัน อนุรักษ์ สิ่งแวดล้อม ในประเทศ แชร์ ทรัพยากร น้ำ ใส ป่าไม้ อุดมสมบูรณ์ คู่ เมือง จีน สร้าง มูลค่า ประเทศ จีน ช่วยกัน คนละไม้คนละมือ แล้วก็ บอกต่อ พวกเรา คนไทย ควรจะ อนุรักษ์ ทรัพยากรธรรมชาติ บ้าน เหมือนกัน ช่วยกัน เนอะ งาน สถาปัตยกรรม สร้าง น้ำแข็ง ยุโรป จีน ดู สวยงาม แกะสลัก น้ำแข็ง ประเทศ เข้าร่วม ดู เดิน ดู เพลิน เล่น สไลเดอร์ น้ำแข็ง เมตร น่าสนใจ คน แถว แถว ชั่วโมง สัมผัส หนาว ยะเยือก ดูงาน แกะสลัก น้ำแข็ง หิมะ เดือน หน้า กุมภาพันธ์ น่าดู เกาะ พระอาทิตย์	1, 2, 3	1, 3
การเยือน พระราชวัง ต้องห้าม ปักกิ่ง รู้สึก บรรยากาศ ความลึกลับ น่าตื่นเต้น เดิน ห้อง ต่าง ๆ ประวัติศาสตร์ จีน เหมือน หลุด เข้าไป	2, 3	1, 2
จีน งาน บูรณะ รูปปั้น พระอรหันต์ พัน ปี ครั้งแรก รอบ ปี รูปปั้น พระอรหันต์ องค์ ลักษณะ ท่าทาง สีหน้า ท่า นั่ง แต่งกาย สีผิว แตก ต่างกัน ท่า นั่ง เหมือนกัน ดู เด็ดเดี่ยว อาจหาญ หน้าตา โมโห โกรธา เปี่ยม เมตตา หลับตา ท่าสมาธิ รายละเอียด เครื่องแต่งกาย ความประณีต งดงาม ละเอียดอ่อน องค์ ปล่อย ชาย เสื้อคลุม แขน เผย ข้อศอก เผย เท้าเปล่า	1	1, 2, 3
โรงแรม มหานคร เซี่ยงไฮ้ ยกเลิก จัดวาง ของใช้ ทิ้ง เอาไว้ ลูกค้า พัก กฎหมาย เซี่ยงไฮ้ ห้าม โรงแรม ของใช้ ทิ้ง ลูกค้า สนองนโยบาย รัฐบาล จีน การรักษา สิ่งแวดล้อม ลด ขยะ ของใช้ ทิ้ง รัฐบาล เซี่ยงไฮ้ ห้าม โรงแรม เตรียม ห้องพัก ลูกค้า แปรงสีฟัน หวี ฟองน้ำ สำหรับ ถู ตัว เวลา อาบน้ำ มีดโกน หนวด ที่ลับ เล็บ ชัด รองเท้า จุดประสงค์ พัก สิ่งของ ซ้ำ โรงแรม เตรียม แน่นอน ทิ้ง เตรียม คน น พัก ต่อไป ตามกฎหมาย ข้อ ระบุ ข้อยกเว้น ในกรณีที่ พัก ร้องขอ พนักงาน โรงแรม ง เตรียม เอาไว้ ล่วงหน้า เอาไว้ ห้องพัก เป็นมา มีความผิด ข้อหา ฝ่าฝืน กฎ โทษปรับ หยวน บาท ข้อมูล สื่อ จีน กฎหมาย บังคับใช้ กรกฎาคม	1, 3	1
ศิลปิน จีน ชื่อดัง เล่น คอนเสิร์ต ปัง ไหว คู ล คร่า ๆๆ ดู ภาพ บางส่วน ที่นี่	1, 2	1, 2, 3
จีน ฟัง แนวคิด น โยบาย หยุด ฟุ่มเฟือย ที่จะ วิถีชีวิต ประชาชน จีน	1, 3	1
ทีม นักวิจัย สถาบัน โบราณคดี มณฑล ซาน ซี ขุด ค้นพบ เครื่องประดับ ลักษณะ อุ้งเท้า แมว ทองคำ ประดับ บริเวณ เอว เจ้าของ หลุมศพ เพศชาย ยุค ราชวงศ์ โจว พื้นที่ สุสาน เปย ไป๋ เอ่อ มณฑล ซาน ซี ประกาศ ค้นพบ เครื่องประดับ แสน น่ารัก เน็ต จิน เห็นภาพ ระบุ หลักฐาน ทาส แมว อย่างไรก็ตาม ผู้เชี่ยวชาญ ฟัง ธง เครื่องประดับ อุ้งเท้า แมว เป็นต้นแบบ ตรวจสอบ ปริมาณ ทองคำ เครื่องประดับ ปริมาณ หลุมศพ เปย ไป๋ เอ่อ เป็นที่ตั้ง สุสาน ราชวงศ์ โจว ขุด ค้นพบ เดือน เมษายน ปี ค้นพบ โบราณวัตถุ ชิ้น	1, 2	1
ข้อมูล เชิง ลึก การวิจัย ชม ความสวยงาม ทุ่ง ดอก ทิวลิป บานสะพรั่ง ต้อนรับ บรรดา นักท่องเที่ยว จีน สารทิศ ไม่ต้อง ประเทศ เนเธอร์แลนด์ ตำบล หลิว เจีย อำเภอ หย่ง จิ่ง มณฑล กาน ซู่ จีน ทางการ ท้องถิ่น เนรมิต ร ทุ่ง ทุ่ง ดอก ทิวลิป ในประเทศ เนเธอร์แลนด์ ป้องกัน ความเสียหาย นักท่องเที่ยว ฝ่าฝืน ถ่ายรูป ดอกไม้ แปลง โดน โทษปรับ	1, 3	1

ภาพที่ 4.1 คลาสที่ทำนายผิดโมเดล PhayaThaiBERT ร่วมกับฟีเจอร์อารมณ์

นอกจากนี้ยังพบความสับสนระหว่างป้ายกำกับ Inform กับ Persuade ซึ่งเป็นผลจากความใกล้เคียงของโครงสร้างภาษาที่ใช้ในการให้ข้อมูลเกี่ยวกับสินค้า บริการ หรือสถานที่ท่องเที่ยว ซึ่งมักมีลักษณะโน้มน้าวในตัวแต่ไม่ได้ใช้ภาษาที่ชัดเจน เช่น คำสั่งหรือการชักจูงโดยตรง ส่งผลให้โมเดล

จำแนกผิดว่าเป็นข้อความให้ข้อมูลเพียงอย่างเดียว ทั้งที่ควรถูกจัดอยู่ในหมวด persuade ด้วย ในบางกรณี ข้อความที่มีเนื้อหาเพื่อความบันเทิงหรือการเล่าเรื่องก็ถูกมองว่าเป็นการให้ข้อมูลเช่นกัน ซึ่งสะท้อนให้เห็นถึงความท้าทายของโมเดลในการแยกแยะจุดประสงค์แฝงของข้อความที่ใช้ภาษาธรรมดาแต่มีบริบทเชิงอารมณ์หรือเชิงโน้มน้าว

สามารถอธิบายได้ว่าข้อผิดพลาดของโมเดลในงานจำแนกข้อความหลายป้ายกำกับนี้เกิดจากความสามารถจำกัดของโมเดลในการตีความเจตนาที่หลากหลายของข้อความ โดยเฉพาะเมื่อต้องแยกความแตกต่างระหว่างข้อความเชิงข้อมูล บันเทิง และการชักจูงที่มีลักษณะทับซ้อนกันในระดับภาษาหรือบริบท การปรับปรุงโมเดลในลักษณะนี้จึงควรคำนึงถึงการฝึกด้วยข้อมูลที่มีความหลากหลายในเชิงวาทกรรม การเน้นการเรียนรู้บริบทจากภาษาธรรมชาติ และการกำกับด้วยป้ายกำกับที่ชัดเจนและสม่ำเสมอมากขึ้น ซึ่งเป็นผลมาจากชุดข้อมูลที่ขาดความสมดุลของจำนวนคลาส และมีจำนวนคลาส Inform ที่เด่นชัดที่สุดมากเกินไป จึงทำให้ผลทำนายส่วนใหญ่ทำนายไปที่ Inform เป็นหลัก

ตารางที่ 4.3 ผลการทดลองที่ 2 โมเดลจำแนกการสื่อความของเนื้อหาแบบใช้ฟีเจอร์อารมณ์

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBert (Validation)	Inform	94.50%	97.86%	96.15%	15.77%
	Entertain	79.15%	61.04%	65.28%	
	Persuade	74.03%	77.01%	75.49%	
	Micro	84.60%	85.71%	85.15%	
	Macro	79.56%	78.64%	78.97%	
mBERT (Test)	Inform	96.44%	99.19%	97.80%	15.24%
	Entertain	75.41%	58.23%	65.71%	
	Persuade	74.22%	72.52%	73.36%	
	Micro	87.10%	84.43%	85.75%	
	Macro	82.02%	76.64%	78.96%	
WangchanBERTa (Validation)	Inform	96.56%	100.00%	98.25%	12.10%
	Entertain	70.00%	81.82%	75.45%	
	Persuade	80.84%	77.59%	79.18%	
	Micro	87.41%	90.04%	88.70%	
	Macro	82.47%	86.47%	84.29%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
WangchanBERTa (Test)	Inform	97.22%	99.59%	98.39%	11.67%
	Entertain	80.25%	82.28%	81.25%	
	Persuade	82.57%	68.70%	75.00%	
	Micro	**90.50%**	**87.72%**	**89.09%**	
	Macro	**86.68%**	**83.52%**	**84.88%**	
PhayaThaiBERT (Validation)	Inform	95.90%	100.00%	97.91%	11.31%
	Entertain	78.38%	75.32%	76.82%	
	Persuade	83.65%	76.44%	79.88%	
	Micro	89.73%	88.72%	**89.22%**	
	Macro	85.98%	83.92%	**84.87%**	
PhayaThaiBERT (Test)	Inform	97.22%	99.59%	98.39%	11.90%
	Entertain	86.30%	79.75%	82.89%	
	Persuade	80.95%	64.89%	72.03%	
	Micro	**91.40%**	**86.18%**	**88.71%**	
	Macro	**88.16%**	**81.41%**	**84.44%**	

4.3 ผลการทดลองที่ 3 โมเดลจำแนกเจตนาการสื่อความของเนื้อหา โดยทดสอบความสมดุลของข้อมูล

การทดลองในส่วนนี้ได้เปรียบเทียบประสิทธิภาพของโมเดล Transformer 3 แบบ คือ mBERT, WangchanBERTa และ PhayaThaiBERT กับชุดข้อมูลที่ปรับความสมดุล โดยได้ทดลองทั้งแบบไม่ใช้และแบบที่มีการเพิ่มฟีเจอร์ เข้าไป เพื่อศึกษาว่าคุณลักษณะด้านอารมณ์จะช่วยพัฒนาความสามารถของโมเดลในการจำแนกเจตนาได้ดีขึ้นหรือไม่

4.3.1 ผลการทดลองย่อยที่ 1

จากตารางผลการทดลองย่อยที่ 1 ในตารางที่ 4.4 ผลจากการทดลองแสดงให้เห็นว่าโมเดล PhayaThaiBERT มีประสิทธิภาพสูงที่สุดในทุกตัวชี้วัดทั้ง Precision, Recall, F1-score และมีค่า Hamming Loss ต่ำ ทั้งในชุดข้อมูล Validation และ Test รองลงมาคือ WangchanBERTa และ mBERT ตามลำดับ โดยเฉพาะในคลาส Entertain และ Persuade โมเดล PhayaThaiBERT ทำได้

ดีกว่าอย่างชัดเจน แสดงให้เห็นถึงศักยภาพของโมเดลที่ถูกออกแบบมาโดยเฉพาะสำหรับภาษาไทยโดยเฉพาะ

เมื่อพิจารณาในเชิงของ ความสมดุลของคลาส พบว่า คลาส Inform มีคะแนน Precision และ Recall สูงในทุกโมเดล โดยเฉพาะใน PhayaThaiBERT และ WangchanBERTa ที่ F1-score สูงกว่า 92% อย่างสม่ำเสมอ สะท้อนว่าเป็นคลาสที่มีลักษณะภาษาชัดเจน หรือมีจำนวนตัวอย่างในชุดฝึกที่มากกว่า ทำให้โมเดลสามารถเรียนรู้ได้แม่นยำ ในทางกลับกัน คลาส Entertain เป็นคลาสที่มีคะแนนต่ำที่สุดในทุกโมเดล โดยเฉพาะ Recall ที่ต่ำกว่า 75% ในหลายกรณี โดยอาจเกิดจากข้อมูลไม่สมดุล อย่างไรก็ตาม โมเดลภาษาไทยโดยเฉพาะอย่าง WangchanBERTa และ PhayaThaiBERT สามารถจัดการกับคลาสนี้ได้ดีกว่า mBERT อย่างชัดเจน

สำหรับ คลาส Persuade พบว่า PhayaThaiBERT มีค่า Recall สูงมากถึง 91.40% ใน Test set แต่ Precision ยังต่ำอยู่เล็กน้อย แสดงว่าโมเดลสามารถจับข้อความโน้มน้าวใจได้เกือบครบ แต่ยังมีการทำนายเกินบ้างในคลาสอื่น ซึ่งอาจเกิดจากความคล้ายคลึงกันของภาษาระหว่างคลาส Persuade และ Inform โดยรวมแล้ว แสดงให้เห็นว่าแม้จะมีปัญหาความไม่สมดุลระหว่างคลาส แต่โมเดลที่ถูกออกแบบมาสำหรับภาษาไทยสามารถจัดการกับความแตกต่างนี้ได้ดีกว่าโมเดลข้ามภาษา และมีศักยภาพในการนำไปใช้งานจริงในระบบจำแนกเจตนาในบริบทภาษาไทย

ตารางที่ 4.4 ผลการทดลองย่อยที่1 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	84.71%	92.31%	88.34%	22.73%
	Entertain	72.09%	70.45%	71.26%	
	Persuade	64.18%	66.15%	65.15%	
	Micro	74.87%	78.07%	76.44%	
	Macro	73.66%	76.31%	74.92%	
mBERT (Test)	Inform	84.80%	94.64%	89.45%	22.22%
	Entertain	72.22%	62.90%	67.24%	
	Persuade	65.00%	69.89%	67.36%	
	Micro	75.27%	78.65%	76.92%	
	Macro	74.01%	75.81%	74.68%	
WangchanBERTa	Inform	90.36%	96.15%	93.17%	14.65%

Model	Class	Precision	Recall	F1-Score	Hamming Loss
(Validation)	Entertain	82.93%	77.27%	80.00%	
	Persuade	76.12%	78.46%	77.27%	
	Micro	83.77%	85.56%	84.66%	
	Macro	83.14%	83.96%	83.48%	
WangchanBERTa (Test)	Inform	90.00%	96.43%	93.10%	12.35%
	Entertain	94.00%	75.81%	83.93%	
	Persuade	78.79%	83.87%	81.25%	
	Micro	**86.62%**	**87.27%**	**86.94%**	
	Macro	**87.60%**	**85.37%**	**86.09%**	
PhayaThaiBERT (Validation)	Inform	91.25%	93.59	92.41%	17.17%
	Entertain	74.47%	79.55%	76.92%	
	Persuade	68.75%	84.62%	75.86%	
	Micro	78.74%	87.17%	82.74%	
	Macro	78.16%	85.92%	81.73%	
PhayaThaiBERT (Test)	Inform	92.24%	95.54%	93.86%	13.76%
	Entertain	88.46%	74.19%	80.70%	
	Persuade	71.43%	91.40%	80.19%	
	Micro	**82.93%**	**89.14%**	**85.92%**	
	Macro	**84.04%**	**87.04%**	**84.92%**	

4.3.2 ผลการทดลองย่อยที่ 2

เมื่อเปรียบเทียบผลลัพธ์ระหว่างการทดลองสองชุดที่มีการแบ่งข้อมูลต่างกัน คือ ชุดใช้สัดส่วนการแบ่งข้อมูลเป็น 70:30 ในการทดลองย่อยที่ 1 และชุดข้อมูลที่ใช้สัดส่วน 80:20 ในการทดลองย่อยที่ 2 นี้ ตามปรากฏในตารางที่ 4.5 จะเห็นได้ว่า การเพิ่มปริมาณข้อมูลสำหรับฝึกโมเดล (Training set) ส่งผลให้ประสิทธิภาพของโมเดลโดยรวมดีขึ้น โดยเฉพาะสำหรับโมเดลที่ออกแบบมาเฉพาะสำหรับภาษาไทยอย่าง PhayaThaiBERT และ WangchanBERTa ซึ่งสามารถเรียนรู้ลักษณะภาษาไทยได้ลึกซึ้งยิ่งขึ้นเมื่อมีข้อมูลมากพอ ในขณะที่โมเดลข้ามภาษาอย่าง mBERT ได้ประโยชน์จากข้อมูลที่เพิ่มขึ้นน้อยกว่าเมื่อเทียบกับโมเดลภาษาไทย

สำหรับ PhayaThaiBERT ผลลัพธ์จากชุดข้อมูล 80:20 แสดงให้เห็นถึงการพัฒนาอย่างชัดเจน โดยค่า F1-score ของคลาส Inform บนชุดทดสอบเพิ่มขึ้นจาก 93.86% (ในชุด 70:30) เป็น 96.00% ซึ่งถือว่าใกล้เคียงกับค่าที่สมบูรณ์แบบ เช่นเดียวกับคลาส Persuade ที่ค่า F1-score เพิ่มจาก 80.19% เป็น 82.44% นอกจากนี้ค่า Macro F1-score โดยรวมของโมเดลก็เพิ่มขึ้นจาก 84.92% เป็น 86.33% และ Hamming Loss ลดลงจาก 13.76% เป็น 11.38%

ในกรณีของ WangchanBERTa แม้จะมีประสิทธิภาพดีอยู่แล้วในชุด 70:30 แต่เมื่อเพิ่มข้อมูลฝึกเป็น 80% โมเดลยังสามารถทำคะแนนได้ดีขึ้นในหลายตัวชี้วัด โดยเฉพาะในคลาส Persuade ที่ค่า F1-score เพิ่มขึ้นจาก 81.25% เป็น 85.71% บนชุดทดสอบ เช่นเดียวกับคลาส Entertain ที่ยังคงมีค่า F1-score สูงกว่า 80% ซึ่งแสดงให้เห็นว่าโมเดลสามารถเข้าใจเจตนาในข้อความที่ซับซ้อนได้ดีขึ้น ขณะที่ค่า Macro F1-score เพิ่มขึ้นเล็กน้อยจาก 86.09% เป็น 86.91% และ Hamming Loss ลดลงอย่างเห็นได้ชัด จาก 12.35% เป็น 11.38%

ในทางกลับกัน mBERT แม้จะได้รับข้อมูลฝึกเพิ่มขึ้นเช่นเดียวกัน แต่กลับไม่สามารถแสดงพัฒนาการได้ชัดเจนเท่ากับสองโมเดลภาษาไทย โดยค่า F1-score ของคลาส Entertain เพิ่มขึ้นเพียงเล็กน้อย จาก 67.24% เป็น 67.57% และ Persuade จาก 67.36% เป็น 72.34% ขณะที่ค่า Macro F1-score โดยรวมเพิ่มขึ้นเพียงเล็กน้อยจาก 74.68% เป็น 76.46% เท่านั้น และ Hamming Loss ยังคงอยู่ในระดับสูงกว่าโมเดลอื่น โดยมีค่ามากกว่า 20%

เมื่อวิเคราะห์ตามคลาส จะเห็นว่า คลาส Inform ยังคงเป็นคลาสที่จำแนกได้ดีที่สุดในทุกโมเดล ซึ่งอาจเกิดจากลักษณะของข้อความที่ตรงไปตรงมา เป็นข้อเท็จจริง และมีจำนวนข้อมูลมาก ขณะที่ คลาส Entertain ยังคงมีผลลัพธ์ต่ำกว่าคลาสอื่น แม้จะมีการเพิ่มข้อมูลฝึกก็ตาม

ตารางที่ 4.5 ผลการทดลองย่อยที่ 2 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	84.54%	92.13%	88.17%	20.53%
	Entertain	85.71%	60.00%	70.59%	
	Persuade	64.58%	83.78%	72.94%	
	Micro	76.32%	81.69%	78.91%	
	Macro	78.28%	78.64%	77.23%	
mBERT (Test)	Inform	88.31%	90.67%	89.47%	20.90%

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Entertain	75.76%	60.98%	67.57%	
	Persuade	65.38%	80.95%	72.34%	
	Micro	76.60%	80.45%	78.47%	
	Macro	76.48%	77.53%	76.46%	
WangchanBERTa (Validation)	Inform	92.39%	95.51%	93.92%	14.35%
	Entertain	85.00%	68.00%	75.56%	
	Persuade	76.92%	81.08%	78.95%	
	Micro	85.24%	84.04%	84.63%	
	Macro	84.77%	81.53%	82.81%	
WangchanBERTa (Test)	Inform	91.03%	94.67%	92.81%	**11.38%**
	Entertain	93.75%	73.17%	82.19%	
	Persuade	81.43%	90.48%	85.71%	
	Micro	**87.78%**	**88.27%**	**88.02%**	
	Macro	**88.73%**	**86.10%**	**86.91%**	
PhayaThaiBERT	Inform	94.44%	95.51%	94.97%	12.36%
	Entertain	83.33%	80.00%	81.63%	
	Persuade	80.82%	79.73%	80.27%	
	Micro	87.20%	86.38%	86.79%	
	Macro	86.20%	85.08%	85.63%	
PhayaThaiBERT (Test)	Inform	96.00%	96.00%	96.00%	**11.38%**
	Entertain	93.55%	70.73%	80.56%	
	Persuade	79.41%	85.71%	82.44%	
	Micro	**89.08%**	**86.59%**	**87.82%**	
	Macro	**89.65%**	**84.15%**	**86.33%**	

4.3.3 ผลการทดลองย่อยที่ 3

ผลการทดลองย่อยที่ 3 โดยใช้ Dataset A ชุดข้อมูลที่มีจำนวนป้ายกำกับรวมทั้งหมดของแต่ละคลาสเท่ากันทุกคลาส โดยตัดข้อมูลที่ปรากฏคลาส Inform แบบคลาสเดี่ยวออกไป ดังปรากฏผล

การทดลองในตารางที่ 4.6 แสดงให้เห็นว่า ประสิทธิภาพของโมเดลโดยรวมดีขึ้นอย่างชัดเจนในหลายมิติ โดยเฉพาะในคลาส Entertain และ Persuade ซึ่งเดิมมักเป็นคลาสที่มีคะแนนต่ำที่สุดในหลายรอบการทดลองก่อนหน้านี้

โมเดลที่โดดเด่นที่สุดจากการทดลองนี้คือ PhayaThaiBERT ให้ค่า Macro F1-score บนชุดทดสอบสูงถึง 90.10% และ Hamming Loss ต่ำที่สุดที่ 11.49% โดยสามารถจำแนกคลาส Entertain ได้แม่นยำมากที่สุดในทุกโมเดล โดยมี F1-score 94.74% แสดงให้เห็นถึงความสามารถของโมเดลในการจัดการกับข้อมูลของคลาสที่พบน้อยในชุดข้อมูลดั้งเดิม ได้ดีขึ้น

ในส่วนของ WangchanBERTa คลาส Entertain มีค่า Recall สูงมาก ที่ 0.9821 และมีค่า F1-score 91.67% ขณะที่คลาส Inform มี F1-score 94.44% แสดงให้เห็นถึงความเสถียรของโมเดลในการจัดการกับคลาสหลัก อย่างไรก็ตาม ข้อจำกัดของ WangchanBERTa คือคลาส Persuade ที่แม้จะมี Precision สูงถึง 91.67% แต่ Recall ต่ำที่สุดที่ 71.74% โดย Macro F1 รวมลดลงเล็กน้อยเป็น 88.87% โดยมีประสิทธิภาพต่ำกว่า PhayaThaiBERT

ขณะที่ mBERT มีผลลัพธ์ที่ดีขึ้นจาก Dataset A เช่นกัน แม้จะได้ประโยชน์จากการ โดยเฉพาะในคลาส Inform และ Entertain ซึ่งได้ F1-score สูงถึง 92.04% และ 84.91% ตามลำดับ แต่ในคลาส Persuade โมเดลยังคงมีข้อจำกัดโดยให้ค่า F1 เพียง 74.00% เท่านั้น ส่งผลให้ค่า Macro F1 โดยรวมอยู่ที่เพียง 83.65% และ Hamming Loss สูงที่สุดในกลุ่ม ที่ 19.54% ซึ่งแสดงถึงการทำนายผิดคลาสบ่อยกว่าทุกโมเดล

ตารางที่ 4.6 ผลการทดลองย่อยที่ 3 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	85.07%	96.61%	90.48%	13.65%
	Entertain	88.33%	89.83%	89.08%	
	Persuade	83.33%	88.71%	85.94%	
	Micro	85.49%	91.67%	88.47%	
	Macro	85.58%	91.72%	88.5%	
mBERT (Test)	Inform	89.66%	94.55%	92.04%	19.54%
	Entertain	90.0%	80.36%	84.91%	
	Persuade	68.52%	80.43%	74.0%	
	Micro	82.72%	85.35%	84.01%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Macro	82.72%	85.11%	83.65%	
WangchanBERTa (Validation)	Inform	86.36%	96.61%	91.2%	15.87%
	Entertain	82.61%	96.61%	89.06%	
	Persuade	81.36%	77.42%	79.34%	
	Micro	83.51%	90.0%	86.63%	
	Macro	83.44%	90.21%	86.53%	
WangchanBERTa (Test)	Inform	96.23%	92.73%	94.44%	12.26%
	Entertain	85.94%	98.21%	91.67%	
	Persuade	91.67%	71.74%	80.49%	
	Micro	90.85%	88.54%	89.68%	
	Macro	91.28%	87.56%	88.87%	
PhayaThaiBERT (Validation)	Inform	90.48%	96.61%	93.44%	10.16%
	Entertain	87.88%	98.31%	92.8%	
	Persuade	86.15%	90.32%	88.19%	
	Micro	88.14%	95.0%	91.44%	
	Macro	88.17%	95.08%	91.48%	
PhayaThaiBERT (Test)	Inform	96.15%	90.91%	93.46%	11.49%
	Entertain	93.10%	96.43%	94.74%	
	Persuade	79.59%	84.78%	82.11%	
	Micro	89.94%	91.08%	90.51%	
	Macro	89.62%	90.71%	90.10%	

4.3.4 ผลการทดลองย่อยที่ 4

จากผลการทดลองโดยใช้ Dataset B ตามตารางที่ 4.7 พบว่า โมเดลที่มีประสิทธิภาพดีที่สุดในชุดทดลองนี้คือ PhayaThaiBERT โดยให้ค่า Macro F1-score บนชุดทดสอบสูงสุดที่ 88.19% และ Micro F1-score ที่ 88.29% ที่มีค่าใกล้เคียงกัน ขณะที่คลาสที่เคยมีปัญหาในการทดลองที่ 2 ในชุดข้อมูลดั้งเดิม ได้แก่ Persuade ได้ ค่า F1-score ถึง 89.11% ซึ่งเป็นค่าที่สูงที่สุดในทุกโมเดลในชุด

ข้อมูลนี้ ขณะที่คลาส Inform และ Entertain มี F1-score 91.26% และ 84.21% ตามลำดับ แม้ค่า Entertain จะลดลงเล็กน้อยจากการทดลองย่อยที่ 3 ซึ่งใช้ Dataset A

ขณะที่ WangchanBERTa แสดงผลได้ใกล้เคียงกับ PhayaThaiBERT อย่างมาก โดยเฉพาะในคลาส Persuade ที่ให้ F1-score สูงถึง 0.8738 และในคลาส Entertain ให้ค่า F1-score 87.50% เทียบเท่ากันใน Precision และ Recall ขณะที่คลาส Inform ยังคงมี F1 สูงสุด 92.16% โดยรวม WangchanBERTa ให้ Macro F1-score อยู่ที่ 0.8901 และ Hamming Loss ต่ำที่สุดในชุดทดสอบที่ 11.70%

สำหรับ mBERT เมื่อใช้ Dataset B โมเดลสามารถทำงานได้ดีในคลาส Inform และ Entertain โดยให้ค่า F1-score อยู่ที่ 85.71% และ 79.17% ตามลำดับ ส่วนคลาส Persuade มีค่า F1-score ต่ำที่สุดที่ 78.43% ทั้งนี้ค่า Macro และ Micro F1-score มีค่าต่ำที่สุดเช่นกัน 81.10% และ 0.8108 ตามลำดับ

ตารางที่ 4.7 ผลการทดลองย่อยที่ 4 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	86.11%	95.38%	90.51%	19.47%
	Entertain	83.33%	89.29%	86.21%	
	Persuade	65.28%	79.66%	71.76%	
	Micro	77.94%	88.33%	82.81%	
	Macro	78.24%	88.11%	82.82%	
mBERT (Test)	Inform	87.5%	84.0%	85.71%	19.86%
	Entertain	79.17%	79.17%	79.17%	
	Persuade	75.47%	81.63%	78.43%	
	Micro	80.54%	81.63%	81.08%	
	Macro	80.71%	81.6%	81.1%	
WangchanBERTa (Validation)	Inform	87.14%	93.85%	90.37%	14.16%
	Entertain	84.13%	94.64%	89.08%	
	Persuade	77.61%	88.14%	82.54%	
	Micro	83.0%	92.22%	87.37%	
	Macro	82.96%	92.21%	87.33%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
WangchanBERTa (Test)	Inform	90.38%	94.0%	92.16%	**11.70%**
	Entertain	87.5%	87.5%	87.5%	
	Persuade	83.33%	91.84%	87.38%	
	Micro	87.01%	91.16%	89.04%	
	Macro	87.07%	91.11%	89.01%	
PhayaThaiBERT (Validation)	Inform	89.55%	92.31%	90.91%	**11.21%**
	Entertain	85.25%	92.86%	88.89%	
	Persuade	89.66%	88.14%	88.89%	
	Micro	**88.17%**	**91.11%**	**89.62%**	
	Macro	**88.15%**	**91.1%**	**89.56%**	
PhayaThaiBERT (Test)	Inform	88.68%	94.0%	91.26%	12.41%
	Entertain	85.11%	83.33%	84.21%	
	Persuade	86.54%	91.84%	89.11%	
	Micro	**86.84%**	**89.8%**	**88.29%**	
	Macro	**86.77%**	**89.72%**	**88.19%**	

4.3.5 ผลการทดลองย่อยที่ 5

ผลการทดลองครั้งนี้ใช้ชุดข้อมูลที่แบ่งเป็น Train 70 : Test 30 และ เพิ่มฟีเจอร์อารมณ์ ตามตารางที่ 4.8 ผลลัพธ์ที่ได้ พบว่า mBERT มีความสามารถโดดเด่นที่สุดในคลาส Inform โดยสามารถทำค่า F1-score ได้ถึง 89.08% บนชุดทดสอบ และเมื่อพิจารณาคลาส Entertain และ Persuade โมเดลยังคงให้ค่า F1-score ที่ต่ำที่สุด ที่ 64.91% และ 72.54% ตามลำดับ ซึ่งบ่งชี้ว่าแม้จะมีฟีเจอร์อารมณ์เข้ามาช่วย แต่โมเดล mBERT ยังไม่สามารถใช้ประโยชน์จากบริบทเชิงอารมณ์ได้เต็มที่ นอกจากนี้ ค่า Hamming Loss ที่สูงสุดในกลุ่ม ที่ 20.81% แสดงให้เห็นถึงทำนายผิดมากที่สุด

ในขณะที่ WangchanBERTa แสดงผลลัพธ์ดีขึ้นเล็กน้อยจากการทดลองย่อยที่ 1 ซึ่งไม่ใช้ฟีเจอร์อารมณ์ โดยMicro และ Macro F1-score ของชุดทดสอบ อยู่ที่ 83.52% และ 82.53% ตามลำดับ โดยค่า Hamming Loss อยู่ที่ 15.52%

PhayaThaiBERT ชุดทดสอบ มีประสิทธิภาพสูงสุดในแง่ของ Mirco และ Macro F1-score แต่มีค่า Hamming Loss สูงกว่า WangchanBERTa และมีประสิทธิภาพลดลงเมื่อเทียบกับไม่ใช้ฟีเจอร์อารมณ์ ซึ่งสอดคล้องกับในผลการทดลองที่ 2 แต่เป็นที่น่าสังเกตว่า มี Hamming Loss ต่ำที่สุดที่ 13.40%

ตารางที่ 4.8 ผลการทดลองย่อยที่ 5 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	85.71%	84.62%	85.16%	23.23%
	Entertain	70.0%	63.64%	66.67%	
	Persuade	68.18%	69.23%	68.7%	
	Micro	75.96%	74.33%	75.14%	
	Macro	74.63%	72.49%	73.51%	
mBERT (Test)	Inform	87.18%	91.07%	89.08%	20.81%
	Entertain	71.15%	59.68%	64.91%	
	Persuade	70.0%	75.27%	72.54%	
	Micro	77.7%	78.28%	77.99%	
	Macro	76.11%	75.34%	75.51%	
WangchanBERTa (Validation)	Inform	88.75%	91.03%	89.87%	18.43%
	Entertain	72.34%	77.27%	74.73%	
	Persuade	73.13%	75.38%	74.24%	
	Micro	79.38%	82.35%	80.84%	
	Macro	78.07%	81.23%	79.61%	
WangchanBERTa (Test)	Inform	90.18%	90.18%	90.18%	15.52%
	Entertain	80.0%	77.42%	78.69%	
	Persuade	77.89%	79.57%	78.72%	
	Micro	83.52%	83.52%	83.52%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Macro	82.69%	82.39%	82.53%	
PhayaThaiBERT (Validation)	Inform	91.46%	96.15%	93.75%	**15.40%**
	Entertain	83.33%	79.55%	81.40%	
	Persuade	74.19%	70.77%	72.44%	
	Micro	**83.87%**	**83.42%**	**83.65%**	
	Macro	**83.00%**	**82.16%**	**82.53%**	
PhayaThaiBERT (Test)	Inform	89.83%	94.64%	92.17%	**13.40%**
	Entertain	93.88%	74.19%	82.88%	
	Persuade	78.12%	80.65%	79.37%	
	Micro	**86.31%**	**85.02%**	**85.66%**	
	Macro	**87.28%**	**83.16%**	**84.81%**	

4.3.6 ผลการทดลองย่อยที่ 6

จากผลการทดลองโดยใช้ชุดข้อมูลที่แบ่งเป็น Train 80: Test 20 และมีการเพิ่ม ฟีเจอร์อารมณ์ ตามผลลัพธ์ในตารางที่ 4.9 พบว่าโมเดลทั้งสาม ได้แก่ mBERT WangchanBERTa และ PhayaThaiBERT ต่างมีการตอบสนองต่อข้อมูลและฟีเจอร์ที่เพิ่มขึ้นในรูปแบบที่น่าสนใจ และแต่ละโมเดลก็แสดงศักยภาพที่แตกต่างกันออกไปตามลักษณะของเจตนา

mBERT มีประสิทธิภาพสูงขึ้นจากการทดลองย่อยที่ 2 ซึ่งไม่ใช้ฟีเจอร์อารมณ์ โดยคลาส Inform มีค่า F1-score อยู่ที่ 88.89% บนชุดทดสอบ แต่เมื่อพิจารณาคลาส Entertain และ Persuade พบว่าค่า F1-score ยังคงต่ำที่สุดที่ 71.43% และ 76.19% ตามลำดับ ขณะที่ Micro และ Macro F1-score อยู่ที่ ต่ำกว่า 80% ต่ำที่สุดในผลการทดลองย่อยที่ 6 เช่นกัน

WangchanBERTa แสดงให้เห็นถึงประสิทธิภาพที่เพิ่มขึ้นอย่างชัดเจนเมื่อมีฟีเจอร์อารมณ์และข้อมูลฝึกที่มากขึ้น โดยเฉพาะในคลาส Entertain และ Persuade ที่มีค่า F1-score สูงถึง 83.78% และ 83.76% ตามลำดับ โดยมีMacro F1-score อยู่ที่ 0.8722 และ Micro F1-score ที่ 88.37%

PhayaThaiBERT ยังคงเป็นโมเดลที่แสดงผลลัพธ์ดีที่สุด และดีขึ้นจากการทดลองย่อยที่ 2 ซึ่งไม่ใช้ฟีเจอร์อารมณ์ โดยเฉพาะในคลาส Inform ที่ได้ค่า F1-score สูงสุดถึง 95.30% และในคลาส Entertain และ Persuade 87.18% และ 83.33% ตามลำดับ ขณะที่ Micro และ Macro F1-score

89.14% และ 88.60% ตามลำดับ โดยแตกต่างจากการทดลองที่ 2 และการทดลองย่อยที่ 5 ของการทดลองย่อยที่ 3 ซึ่งผลลัพธ์ PhayaThaiBERT ลดต่ำลง เมื่อใช้ฟีเจอร์อารมณ์

ตารางที่ 4.9 ผลการทดลองย่อยที่ 6 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	88.64%	87.64%	88.14%	18.76%
	Entertain	73.91%	68.0%	70.83%	
	Persuade	77.94%	71.62%	74.65%	
	Micro	81.68%	77.46%	79.52%	
	Macro	80.16%	75.75%	77.87%	
mBERT (Test)	Inform	92.75%	85.33%	88.89%	18.52%
	Entertain	69.77%	73.17%	71.43%	
	Persuade	76.19%	76.19%	76.19%	
	Micro	81.14%	79.33%	80.23%	
	Macro	79.57%	78.23%	78.84%	
WangchanBERTa (Validation)	Inform	91.49%	96.63%	93.99%	14.13%
	Entertain	81.82%	72.00%	76.60%	
	Persuade	86.44%	68.92%	76.69%	
	Micro	87.82%	81.22%	84.39%	
	Macro	86.58%	79.18%	82.43%	
WangchanBERTa (Test)	Inform	92.31%	96.0%	94.12%	10.58%
	Entertain	93.94%	75.61%	83.78%	
	Persuade	90.74%	77.78%	83.76%	
	Micro	92.12%	84.92%	88.37%	
	Macro	92.33%	83.13%	87.22%	
PhayaThaiBERT (Validation)	Inform	95.60%	97.75%	96.67%	**10.82%**
	Entertain	82.35%	84.0%	83.17%	
	Persuade	81.58%	83.78%	82.67%	
	Micro	**87.61%**	**89.67%**	**88.63%**	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Macro	**86.51%**	**88.51%**	**87.50%**	
PhayaThaiBERT (Test)	Inform	95.95%	94.67%	95.30%	**10.32%**
	Entertain	91.89%	82.93%	87.18%	
	Persuade	79.71%	87.30%	83.33%	
	Micro	**88.89%**	**89.39%**	**89.14%**	
	Macro	**89.18%**	**88.30%**	**88.60%**	

4.3.7 ผลการทดลองย่อยที่ 7

จากผลการทดลองย่อยที่ 7 ซึ่งใช้ Dataset A และเพิ่มฟีเจอร์อารมณ์ ตามตารางที่ 4.10 พบว่าโมเดลทั้งสาม มีแนวโน้มการจำแนกเจตนาที่ดีขึ้น

mBERT มีค่า Micro และ Macro F1-score ทั้งชุด Validation และชุดทดสอบต่ำกว่าในผลการทดลองย่อยที่ 3 ซึ่งไม่ใช้ฟีเจอร์อารมณ์ ซึ่งเป็นลักษณะแตกต่างจากผลการทดลองที่ 2 และผลการทดลองย่อยอื่นๆของผลการทดลองที่ 3 ที่ใช้ฟีเจอร์อารมณ์ ขณะที่ WangchanBERTa มีค่า Micro และ Macro F1-score ซึ่งมีค่าเท่ากันที่ 86.79% โดยต่ำกว่าแบบที่ไม่ใช้ฟีเจอร์อารมณ์ ในชุด Validation แต่มีประสิทธิภาพดีขึ้นเล็กน้อยสำหรับชุดทดสอบ ที่ 89.81% และ 89.24% ตามลำดับ

โดย PhayaThaiBERT ยังคงแสดงประสิทธิภาพสูงที่สุดในทุกโมเดล และมีประสิทธิภาพลดลงเมื่อใช้ฟีเจอร์อารมณ์ เทียบกับไม่ใช้ฟีเจอร์อารมณ์ ทั้งชุด Validation และ Test โดยมีค่า Micro และ Macro F1-score ได้แก่ 89.81% 89.70% และ 90.03% และ 89.69% ตามลำดับ

ตารางที่ 4.10 ผลการทดลองย่อยที่ 7 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	86.15%	94.92%	90.32%	14.6%
	Entertain	87.93%	86.44%	87.18%	
	Persuade	84.13%	85.48%	84.8%	
	Micro	86.02%	88.89%	87.43%	
	Macro	86.07%	88.95%	87.43%	
mBERT (Test)	Inform	90.74%	89.09%	89.91%	19.54%

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Entertain	90.2%	82.14%	85.98%	
	Persuade	69.81%	80.43%	74.75%	
	Micro	83.54%	84.08%	83.81%	
	Macro	83.58%	83.89%	83.55%	
WangchanBERTa (Validation)	Inform	87.5%	94.92%	91.06%	15.56%
	Entertain	83.33%	93.22%	88.0%	
	Persuade	81.97%	80.65%	81.3%	
	Micro	84.29%	89.44%	86.79%	
	Macro	84.27%	89.59%	86.79%	
WangchanBERTa (Test)	Inform	92.59%	90.91%	91.74%	12.26%
	Entertain	88.71%	98.21%	93.22%	
	Persuade	87.8%	78.26%	82.76%	
	Micro	89.81%	89.81%	89.81%	
	Macro	89.70%	89.13%	89.24%	
PhayaThaiBERT (Validation)	Inform	91.67%	93.22%	92.44%	**11.75%**
	Entertain	85.51%	100.00%	92.19%	
	Persuade	90.74%	79.03%	84.48%	
	Micro	**89.07%**	**90.56%**	**89.81%**	
	Macro	**89.30%**	**90.75%**	**89.70%**	
PhayaThaiBERT (Test)	Inform	95.92%	85.45%	90.38%	**11.88%**
	Entertain	91.53%	96.43%	93.91%	
	Persuade	84.78%	84.78%	84.78%	
	Micro	**90.91%**	**89.17%**	**90.03%**	
	Macro	**90.74%**	**88.89%**	**89.69%**	

4.3.8 ผลการทดลองย่อยที่ 8

จากผลการทดลองโดยใช้ Dataset B และเพิ่มฟีเจอร์อารมณ์ ตามตารางที่ 4.11 พบว่า mBERT มีประสิทธิภาพที่ดีขึ้นอย่างเห็นได้ชัดเมื่อเทียบกับผลการทดลองย่อยที่ 4 ซึ่งไม่ใช้ฟีเจอร์

อารมณ์ โดยสามารถทำงานได้ดีในคลาส Inform และ Entertain โดยมีค่า F1-score บนชุดทดสอบอยู่ที่ 88.00% และ 84.54% ตามลำดับ ขณะที่คลาส Persuade มีค่า F1-score ต่ำที่สุดที่ 80.39% และภาพรวม mBERT มี Macro F1-score ที่ 84.31% และ Micro F1-score ที่ 84.28%

ขณะที่ WangchanBERTa มีประสิทธิภาพลดลงทั้งชุด Validation และชุดทดสอบ Macro F1-score เท่ากับ 87.36% และ Micro F1-score ที่ 87.33% ในชุดทดสอบ และมีค่า Hamming Loss สูงขึ้นเช่นกัน ที่ 13.48% แสดงถึงข้อผิดพลาดที่เพิ่มขึ้น

ด้าน PhayaThaiBERT มีประสิทธิภาพสูงที่สุด แต่มีประสิทธิภาพลดลงจากผลการทดลองย่อยที่ 4 ซึ่งไม่ใช้ฟีเจอร์อารมณ์ เช่นกัน ทั้งชุด Validation และชุดทดสอบ โดย Macro F1-score เท่ากับ 87.87% และ Micro F1-score เท่ากับ 87.92% ในชุดทดสอบ

ตารางที่ 4.11 ผลการทดลองย่อยที่ 8 ของการทดลองที่ 3

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Inform	86.11%	95.38%	90.51%	18.88%
	Entertain	84.48%	87.5%	85.96%	
	Persuade	66.67%	81.36%	73.28%	
	Micro	78.71%	88.33%	83.25%	
	Macro	70.9%	88.08%	83.25%	
mBERT (Test)	Inform	88.0%	88.0%	88.0%	16.67%
	Entertain	83.67%	85.42%	84.54%	
	Persuade	77.36%	83.67%	80.39%	
	Micro	82.89%	85.71%	84.28%	
	Macro	83.01%	85.7%	84.31%	
WangchanBERTa (Validation)	Inform	89.86%	95.38%	92.54%	14.45%
	Entertain	84.13%	94.64%	89.08%	
	Persuade	77.05%	79.66%	78.33%	
	Micro	83.94%	90.00%	86.86%	
	Macro	83.68%	89.90%	86.65%	
WangchanBERTa (Test)	Inform	89.8%	88.00%	88.89%	13.48%
	Entertain	86.00%	89.58%	87.76%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Persuade	81.48%	89.8%	85.44%	
	Micro	85.62%	89.12%	87.33%	
	Macro	85.76%	89.13%	87.36%	
PhayaThaiBERT (Validation)	Inform	90.91%	92.31%	91.6%	11.5%
	Entertain	85.94%	98.21%	91.67%	
	Persuade	84.75%	84.75%	84.75%	
	Micro	87.30%	91.67%	89.43%	
	Macro	87.20%	91.76%	89.34%	
PhayaThaiBERT (Test)	Inform	88.46%	92.0%	90.2%	12.77%
	Entertain	85.42%	85.42%	85.42%	
	Persuade	86.27%	89.8%	88.0%	
	Micro	86.75%	89.12%	87.92%	
	Macro	86.72%	89.07%	87.87%	

4.3.9 วิเคราะห์ภาพรวมผลการทดลองโมเดลจำแนกเจตนาการสื่อความของเนื้อหาโดยใช้ Balanced Data

จากผลการทดลองจำแนกเจตนาโดยใช้ชุดข้อมูลแบบ Balanced ทั้งจำนวนป้ายกำกับทุกคลาสใกล้เคียงกัน และเท่ากันทุกคลาส ส่งผลต่อประสิทธิภาพของโมเดลอย่างมีนัยสำคัญ โดยเฉพาะในกรณีของ Dataset A ซึ่งเป็นชุดข้อมูลแบบ Balanced ที่ออกแบบให้ ไม่มีคลาส Inform ที่ปรากฏแบบเดี่ยว ส่งผลให้โมเดลสามารถเรียนรู้ความสัมพันธ์ระหว่างป้ายกำกับได้ดียิ่งขึ้น และลดอคติที่อาจเกิดจากคลาส Inform ซึ่งมักเกิดขึ้นในชุดข้อมูลที่ไม่สมดุล ผลลัพธ์แสดงให้เห็นว่า Dataset A ให้ค่า Precision ค่า Recall และ ค่า F1 ทั้งแบบ Micro และ Macro สูงกว่า ชุดข้อมูล Balanced ทั่วไป และ Imbalanced อย่างชัดเจน โดยเฉพาะในโมเดลที่มีศักยภาพสูง เช่น WangchanBERTa และ PhayaThaiBERT แม้ว่า Dataset A จะมีขนาดเล็กกว่าชุดอื่น แต่สามารถแสดงผลลัพธ์ที่เหนือกว่าในหลายมิติ โดยเฉพาะเรื่อง ความใกล้เคียงกันของผลลัพธ์ในแต่ละคลาส ซึ่งแสดงให้เห็นว่าโมเดลสามารถจำแนกคลาสทั้ง 3 ได้สมดุลมากขึ้น ไม่เอนเอียงไปที่คลาสใดคลาสหนึ่งอย่างชัดเจน จึงตั้งข้อสังเกตได้ว่า คุณภาพของโครงสร้างข้อมูลสำคัญกว่าปริมาณของข้อมูล

เมื่อมีการเพิ่มฟีเจอร์อารมณ์ เพื่อช่วยในการแยกเจตนา ผลลัพธ์มีแนวโน้มดีขึ้น โดยเฉพาะเมื่อใช้ร่วมกับโมเดลขนาดใหญ่ อย่างไรก็ตาม ประสิทธิภาพของใช้ร่วมกับฟีเจอร์อารมณ์มีแนวโน้มลดลงในโมเดล PhayaThaiBERT ซึ่งเป็นแนวโน้มเดียวกันกับชุดข้อมูลเดิม นอกจากนี้ ยังพบข้อมูลสำคัญว่า ชุดข้อมูลแบบ Balanced ที่มีจำนวนป้ายกำกับใกล้เคียงกันแต่ไม่เท่ากันทั้งหมด กลับมีผลลัพธ์ ต่ำกว่าชุดข้อมูลแบบไม่สมดุล (Imbalanced) ในการฝึกแบบมาตรฐาน (เช่น 70:30) แต่เมื่อมีการเพิ่มสัดส่วนของข้อมูลฝึกเป็น 80:20 ผลลัพธ์กลับดีขึ้นชัดเจน แต่ยังคงต่ำกว่าในชุดข้อมูลเดิม

เมื่อวิเคราะห์คลาสที่ทำนายผิดในโมเดล PhayaThaiBERT ที่ให้ผลลัพธ์ที่ดีที่สุดทั้งใช้และไม่ใช้ฟีเจอร์อารมณ์ร่วมกับชุดข้อมูล Dataset A ตามที่ปรากฏในภาพที่ 4.2 และ ภาพที่ 4.3 พบว่าโมเดลมีแนวโน้มที่สามารถทำนายคลาส Entertain และ Persuade ได้มากขึ้นเมื่อเปรียบเทียบกับโมเดลที่ใช้ชุดข้อมูลเดิม ที่มักทำนายได้เพียงคลาส Inform เท่านั้น อย่างไรก็ตาม แม้จะสามารถตรวจจับคลาสที่มีเจตนาอื่นนอกเหนือจาก Inform มากขึ้น แต่ยังมีแนวโน้มที่จะทำนายป้ายกำกับเกินความเป็นจริง ตัวอย่างเช่น ในบางกรณีที่เนื้อหามีเจตนา Inform แต่กลับได้รับการจัดประเภทว่าเป็น Entertain หรือ Persuade ร่วมด้วย ในขณะเดียวกัน โมเดลก็ยังคงมีปัญหาในการทำนายไม่ครบถ้วนในกรณีที่ข้อความมีลักษณะผสมผสานหลายเจตนา

text	true_labels	predicted_labels
ซี รีย์ ฮา ไหว ว่าด้วย ความกลัว ตอน กินข้าว เฮีย ต้าเว่ย ไฟ ดับ คน ชน กลัว ไหม พ รี่ แจ ค ผม กลัว ปาร์ตี้ ชุดนอน ไฟ ดับ พ รี เ จ ค ทำ กอด แน่น ร่าง ไห นพ รี่ กลัว ค ร้า ??? เดิ่ง กอด ผิดคน หยาง มี เจ็ แจ ค คน กลัว กอด กอด กอด เดิ่ง รอด ไหม	1, 2	1, 2, 3
ขอแสดงความยินดี ไป่ จิง ยิง เก็ โต เดิม ความหวาน ปล่อยตัว ชม ติด ชาร์ต อันดับ ประจำ วีค ฟัง	1, 2, 3	1, 2
เถา เป่า เว็บไซต์ ช้อปปิ้ง ออนไลน์ ยักษ์ จีน รายงาน ความคืบหน้า การประชุม งบประมาณ ปี ปี จำนวน ผู้ค้า ลงทะเบียน ร้านค้า ออนไลน์ เพิ่มขึ้น ต่อเนื่อง โดยเฉพาะอย่างยิ่ง เดือน มีนาคม ส ถา ร การณ์ แพร่ระบาด เชื้อไวรัส โควิด จีน เบาบาง ลงทะเบียน ร้านค้า ออนไลน์ เฉลี่ย หมื่น อย่างไรก็ตาม ปริมาณ ผู้ค้า ออนไลน์ เถา เป่า มากขึ้น ยอด ขายของ ร้าน เดิบโต แพ้ เถา เป่า บรรดา ร้านค้า ขายดี ประจำปี ปี ติดต่อกัน ร้านค้า แพ ลท ฟอร์ม ได้มาตรฐาน ผู้ค้า พื้นที่ เดิบโต ช้อมูล	1, 2	1, 3
ห้าง ปักกิ่ง แจก คูปอง ส่วนลด สำหรับ สินค้า ส่งท้าย ปี ตื่นเต้น รีบ ช้อป	2	1, 3
ข่าว กวางโจว วิ่ง ระดมทุน ช่วยเหลือ เด็ก พื้นที่ ชนบท	1, 3	3
ทะเลสาบ เฉียน เต่า ทะเลสาบ ผิมู่ อนุษย์ ที่เกิด สร้าง เชื่อม กั้น แม่น้ำ ชิ น โรง ผลิต กระแสไฟฟ้า พลังงาน น้ำ ปี ล้อมรอบ ห้า สิงโต ตั้งอยู่ มณฑล เจ้อเจียง ขนาด ครอบคลุม พื้นที่ ตารางกิโลเมตร ปริมาณ น้ำ มหาศาล ไหล เข้ามา เชื่อม พื้นที่ โดยรอบ จม บาดาล ยอดเขา ความสูง ไผล่ เหนือ น้ำ กลาย สภาพ เกาะ แก่ง ทะเลสาบ เฉียน เต่า ทะเลสาบ พัน เกาะ กลายเป็น สถานที่ท่องเที่ยว สวยงาม จีน พื้นที่ เป็นที่ตั้ง เมือง โบราณ สอง เมือง ชื่อ เจิง เมือง สิงโต สร้าง ปี ราชวงศ์ ถัง เมือง เมือง เห อเจิง เมือง โบราณ สร้าง ปี ราชวงศ์ ชิ่น ทั้งสอง เมือง ชม น้ำ ระดับน้ำ ที่สูง เรื่อย ๆ เมือง บาดาล กลายเป็น จุด ท่องเที่ยว น้ำ ตำน้ำ ตำ เยี่ยมชม สภาพ เมือง ยังมี ความสมบูรณ์ ช้างมาก สภาพ น้ำ เอื้อ อำนวย การเจริญเติบโต สิ่งมีชีวิต เมือง บาดาล กลายเป็น สถานที่ ศึกษา ประวัติศาสตร์ อีกด้วย ข้อมูล แหล่งข่าว	1, 2	1, 2, 3
จีน ผม บันทึก ประสบการณ์ จีน วิธีการ รับมือ ดูถูก เหตุการณ์ ดี เกิดขึ้น ต่างแดน อ่าน โพสต์ ขออนุญาต ข้อ ชี้แจง โพสต์ เล่า เคส ดี เจอ ใช้ชีวิต จีน จุดประสงค์ ต้องการ ข้อคิด เหมา ระบุ ชัดเจน อยู่แล้ว เป็นเรื่อง เจอ บางคน เล่า ข้อคิด เจอ แบบนี้ ความคิด ความรู้สึก คนจีน ดี ดี เพจ เล่าเรื่อง คนจีน บันทึก ปี ตอนนั้น จีน เป็นเวลา ปี ที่อยู่ จีน เจอ เรื่องราว คนจีน พูดถึง ไทย ดี ดี เป็นเรื่อง ธรรมดา มุม จีน ประเทศ ประชากร หลัก พัน ล้าน คน แน่นอน ค วา ? หลากหลาย ความคิด นิสัยใจคอ คน เจอ ข้อดี ข้อ ผม ค้นพบ จีน จีน ผม วัน หนึ่ง อ้าย ซื้อ ร้านค้า ร้าน แถว มหาวิทยาลัย ที่อยู่ ที่พัก อ้าย เจ้าของร้าน จีน ถาม คน อ้าย ตอบ คนไทย คนจีน คน เงยหน้า สัก แป๊บ ออกมา ประเทศ ไทย ประเทศ ความรู้สึก ตอนนั้น ยิ้ม ออกมา คิดใน ใจ หมาย ถึงว่า พื้นที่ ประเทศ ไทย ละ มั้ง คนจีน คน หยุด ทำ เสียง ดูถูก อันนี้ ได้คิด ถาม จีน กี่ ปี ความจริง ปี เจ้าของร้าน จบ แฟน เจ้าของร้าน แก่กว่า ปี เรียน ป โท ป เอก ละ อ้าย บอ เรียน โท ปักกิ่ง ได้ยิน ปักกิ่ง ตัว เจ้าของร้าน น้ำเสียง ออกมา ทำนอง ดีกว่า เดิม ชื่นชม คนจีน บางคน ดูถูก คนจีน ด้วยซ้ำ ยกย่อง คนเมือง ครั้งแรก เจอ ทำนอง เจอ คนจีน คน กะ บ๊ะ ผม มาจาก ประเทศ ไทย ประเทศ จีน หรอ เจอ เจอ คนจีน พูดจา ล้อเลียน สำเนียง ชัด อ้าย ตรงนี้ มีปัญหา ติดอ่าง ชัด มาโดยตลอด โดน เหยียบ ย่ำ บางคน ซื้อ ร้าน ขายของ คนจีน ทำ ทำทาง ติดอ่าง ชัด ล้อเลียน เวลา เจอ เรื่อง ดี ความรู้สึก แย่ เฉยอะ ขอบคุณ ผม มีพลัง เอาชนะ ติดอ่าง คน เจอ เหตุการณ์ ผม เล่า อาการ ไมโห แน่นอน ผม ไมโห นึ่ง ติดได้ เผลอ ดูถูก คนอื่น โดยไม่รู้ตัว เหมือน กัน ประเทศ เพื่อนบ้าน ดังนั้น ทนทาน มอง โลก มากขึ้น ข้อผิดพลาด รู้ ห้าม ความคิด ความคิด ตัวเรา ที่จะ ความคิด ดี เหมือน เหรียญ สอง คนจีน ดี ดี เคารพ ผู้อื่น ดี พวกเรา แตก ต่างกัน โลก ใบ ทำความเข้าใจ เคารพ ตอนนั้น ใช้ชีวิต จีน ผม ดี กลายเป็น แรงบันดาลใจ ที่จะ คน มาจาก ประเทศ ประเทศ ทำ ตอบได้ คน ความคิด ดี การกระทำ เป็นประโยชน์ ปราศจาก ความรุนแรง การทำร้าย ความคิด ทางกาย วาจา หล่อหลอม อ้าย ผม ชอบ แบ่งปัน ประสบการณ์ ส่งต่อ ความรู้ คน คน ทำได้ ผู้อื่น ที่จะ ทำ ผิดผล ดี คนอื่น ฝ่าก เอาไว้ ดูถูก คนอื่น รู้สึก ดี แย่ แย่ คนอื่น ทำต่อ ดังนั้น อย่า ทำต่อ	1, 3	1, 2, 3
หนุ่ม ดารา ชาย จีน ฮอต ยุค ฝ่าก แฟน	1, 2	2
ตรุษจีน ปีใหม่ คนจีน แปล ความหมาย ภาษาจีน หมายถึง เริ่มต้น ฤดูใบไม้ผลิ ดอกไม้ บาน สะ พร่าง ดอก เหมย ดอก บ๊วย เป็นหนึ่ง ดอกไม้จีน มีชื่อเสียง เป็นอย่างมาก พิชช่า เอา ใจ คนไทย เชื่อสาย จีน คนไทย ชื่นชอบ จีน ผสมผสาน พิชช่า ตะวันตก ลงตัว พิชช่า จักรพรรดิ รูปทรง ดอก เหมย เสริม มงคล ตรุษจีน สื่อ ดอก เหมย เบ่งบาน สดใส รุ่งเรือง เหมาะสม เริ่มต้น ตรุษจีน อ้าย ถือโอกาส รีวิว พิชช่า จักรพรรดิ ต้นตำรับ จีน ต้อนรับ ตรุษจีน พิชช่า จักรพรรดิ เลือก ลิ้มลอง รสชาติ หน้า หมูแดง ช่องกง หน้า ไก่ ผัด เม็ด มะม่วง ลอง เสี่ย รววกับ พิชช่า น่าเข้า จีน โฆษณา รสชาติ ไกล้เคียง รสชาติ จีน ไก่ ผัด เม็ด มะม่วง หมูแดง ช่องกง หอม ความมัน อารมณ์ กลิ่นอาย เมือง จีน โดยเฉพาะ ซอบ พิชช่า อัดแน่น ชิ้น เนื้อ หมูแดง ย่าง หอม ไก่ ผัด เม็ด มะม่วง รอส ฉ่ำ เข้าไป อร่อย เข้มข้น กินกัน ปรบ อย่า อ้าย ลิ้มลอง ด้วยตัวเอง รอ เล่า ลอง พิชช่า จักรพรรดิ โดยพลัน ของใหว้ บรรพบุรุษ อ้าย เชิญชวน ชัก อร่อย ตรุษจีน ปี ทาน พิชช่า จักรพรรดิ ความหมาย มงคล ดี กุมภา ราคา เริ่มต้น บาท รีบ ลิ้มลอง รสชาติ จีน ต้นตำรับ สาขา สะดวก ร้าน โทร สั่ง สั่งออนไลน์ อร่อย ความหมาย มงคล เหมือนกัน	1, 2, 3	1, 2
วิเคราะห์ กลยุทธ์ การตลาด ออนไลน์ กระแส แก้ว อุ้งเท้า แมว จีน ฟีเวอร์ ปี ที่ผ่านมา คอลเลคชั่น แก้ว อุ้งเท้า แมว จีน กระแส ไวรัล มูลค่า แก้ว ตอนนั้น ราคา ปกติ หยวน คน สน ใจ กัน เอา ชาย มูลค่า หยวน ปี จีน คอลเลคชั่น แมว อีกแล้ว อุ้งเท้า แมว แก้ว หาง แมว ย้อนรอย กระแส ไวรัล เจ้า แก้ว อุ้งเท้า แมว สีชมพู สร้าง ปราก ฏ การณ์ ไวรัล จีน ปี เพราะเหตุ ใด กลายเป็น กระแส ไวรัล วงกว้าง ก่อให้เกิด ยอดขาย ถล่มทลาย ย้อน เดือน มีนาคม ปี ที่ผ่านมา เป็นช่วง คนจีน ตามหา แก้ว อุ้งเท้า แมว จีน กระแส โลก สังคม ออนไลน์ จีน สื่อ สื่อ จีน บรรดา ผู้ใช้งาน ทั่วไป โพสต์ รูป แก้ว ที่มา รูป อุ้งเท้า แมว สีชมพู อ่อน กระแส ความต้องการ อยากได้ แก้ว คอลเลคชั่น ครอบครอง สาขา คน เข้าคิว รอ ซื้อ เข้ามิด ทะเลาะ แย่ง แก้ว ใบ ครอง ราคา ปกติ ใบ หยวน เป็นที่ต้องการ ของตลาด ตอนนี้ คน ชาย ออนไลน์ ราคา หยวน คนซื้อ ชาย ปี จีน คลอด คอลเลคชั่น อุ้งเท้า แมว ออกมา อีกแล้ว ตอกย้ำ ความนิยม อ้าย เคส วิเคราะห์ เรื่อง การตลาด เคส น่าสน ใจ ซื้อเสียง ความนิยม ตัว แบรนด์ หมู่ คนจีน อันดับ จีน โลก สังคม ออนไลน์ มีส่วน ท่าให้เกิด ปัง กระแส อย่างไรบ้าง อ้าย วิเคราะห์ เลือก กระแส เหมาะสม คนจีน นิยม เสพ สื่อ ออนไลน์ โดยเฉพาะ สังคม ออนไลน์ ต่อเนื่อง กระแส เกิดขึ้น โลก ออนไลน์ จีน ปฏิเสธ ยุท เชีย ล มีเดีย เบอร์ จีน คนจีน นิยม ใช้บน มือถือ โพสต์ เวลา เจอ น่าสน ใจ ส่งต่อ ข่าง ผู้ใช้งาน อายุ ต่ำกว่า ปี ถือเป็น วัยรุ่น วัย ความสน ใจ สิ่งใหม่ กระแส ผู้ใช้งาน รูปแบบ โปรโมท		

ภาพที่ 4.2 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ไม่ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Dataset A

วิเคราะห์ กลยุทธ์ การตลาด ออนไลน์ กระแส แก้ว อุ้งเท้า แมว จีน ฟีเวอร์ ปี ที่ผ่านมา คอลเลคชั่น แก้ว อุ้งเท้า แมว จีน กระแส ไวรัล มูลค่า แก้ว ตอนนั้น ราคา ปกติ หยวน คน สนใจ กันเอา ขาย มูลค่า หยวน ปี จีน คอลเลคชั่น แมว อีกแล้ว อุ้งเท้า แมว แก้ว หาง แมว ย้อนรอย กระแส ไวรัล เจ้า แก้ว อุ้งเท้า แมว สีชมพู สร้าง ปรากฏ การณ์ ไวรัล จีน ปี เพราะเหตุใด กลายเป็น กระแส ไวรัล วงกว้าง ก่อให้เกิด ยอดขาย ถล่มทลาย ย้อน เดือน มีนาคม ปี ที่ผ่านมา เป็นช่วง คนจีน ตามหา แก้ว อุ้งเท้า แมว จีน กระแส โลก สังคม ออนไลน์ จีน สื่อ สื่อ จีน บรรดา ผู้ใช้งาน ทั่วไป โพสต์ รูป แก้ว ที่มา รูป อุ้งเท้า แมว สีชมพู อ่อน กระแส ความต้องการ อยากได้ แก้ว คอลเลคชั่น ครอบครอง สาขา คน เข้าคิว รอ ซื้อ เช้ามืด ทะเลาะ แย่ง แก้ว ใบ ครอง ราคา ปกติ ใบ หยวน เป็นที่ต้องการ ของตลาด ตอนนี้ คน ขาย ออนไลน์ ราคา หยวน คนซื้อ ขาย ปี จีน คลอด คอลเลคชั่น อุ้งเท้า แมว ออก มา อีกแล้ว ตอกย้ำ ความนิยม อ้าย เคส วิเคราะห์ เรื่อง การตลาด เคส น่าสนใจ ชื่อเสียง ความนิยม ตัว แบรนด์ หมู่ คนจีน อันดับ จีน โลก สังคม ออนไลน์ มีส่วน ทำให้เกิด ปัง กระแส อย่างไรบ้าง อ้าย วิเคราะห์ เลือก กระแส เหมาะสม คนจีน นิยม เสพ สื่อ ออนไลน์ โดยเฉพาะ สังคม ออนไลน์ ต่อเนื่อง กระแส เกิดขึ้น โลก ออนไลน์ จีน ปฏิเสธ ฌช เชีย ล มีเดีย เบอร์ จีน คนจีน นิยม ใช้บน มือถือ โพสต์ เวลา เจอ น่าสนใจ ส่งต่อ ย่าง ผู้ใช้งาน อายุ ต่ำกว่า ปี ถือเป็น วัยรุ่น วัย ความสนใจ สิ่งใหม่ กระแส ผู้ใช้งาน รูปแบบ โปรโมท ขาย สินค้า ขับเคลื่อนด้วย เนื้อหา มากยิ่งขึ้น ขาย ได้ๆๆ แพ ลท ฟอร์ม ไฟกัส สร้าง แพ ลท ฟอร์ม สร้าง เนื้อหา น่าสนใจ กระแส การรับรู้ ขาย แพ ลท ฟอร์ม เจ้าที่ มาแรง จีน ฟัง ก์ชั่น ขาย สินค้า แพ ลท ฟอร์ม ทำ ให้การ ขาย เกิดขึ้น ควบคู่ กระแส รับรู้ วงกว้าง สี่ ประสาน พลัง ไชเชียล จีน วงกว้าง เจาะลึก กลุ่มเป้าหมาย เมือง จีน สังคม ออนไลน์ แพ ลท ฟอร์ม จำนวนมาก เล่า โพสต์ จบ มีอิทธิพล กระแส สังคม ออนไลน์ จีน สอง ปี หนี พ้น เจาะ ผู้ใช้งาน ออนไลน์ จีน วงกว้าง ข้อมูล อ้าย เล่า กลุ่มอายุ ปี หลัก กระแส แพ ลท ฟอร์ม แหล่ง สื่อ จีน แทบ สำนัก นำเสนอ ข่าว กระแส ออกมา ครอบคลุม แทบ วัย แฮ ช แท็ก มีพลัง แฮ ช แท็ก ค้นหา ติดอันดับ ค้นหา ยอดฮิต กระแส ปัง การันตี ภาษาจีน แอพ วัยรุ่น ตอนต้น ยัน ตอนปลาย วัย ชื่นชอบ ดู เนื้อหา คลิป ฮา ไวรัล สังคม ออนไลน์ เจาะ วง แคบ มาจาก กลุ่มคน ชอบ นึกถึง เพี้ย ภาพ ชัดเจน ซื้อ สินค้า รีวิว สินค้า ผู้ใช้งาน แรกเริ่ม ปี ผู้หญิง ชนชั้นกลาง ชอบ สิน ค้าความ สวย ความงาม ท่องเที่ยว ช่วงอายุ คน อายุ ปี อาศัย เมือง ชั้น จีน ถือเป็น แพ ลท ฟอร์ม เติบโต ไว เวลา ปี ผู้ใช้งาน ล้าน คน หมู่ นักลงทุน บริษัท ยักษ์ จีน เจ้าของ แจ็ค หม่า ซื้อ หุ้น ลงทุน ปี พุดถึง ทำ ตลาด ออนไลน์ คน นึกถึง ติด ภาพ บ้าน แนะนำ ทำ ได้ผล ข้อจำกัด เอาไว้ เล่า โอกาส ต่อไป ก่อให้เกิด กระแส ตลาด ออนไลน์ จีน แชท มือถือ อันดับ จีน ดังนั้น กระแส น่าสนใจ ดี ดี พฤติกรรม คนเรา บอกต่อ เพื่อน คนจีน แชท ก่อให้เกิด การตลาด ผู้ใช้งาน บอกต่อ จำเป็นต้อง คน ติดตาม ลอง นึก ถึงว่า คนจีน คน คน คน ข้อความ น่าสนใจ เข้าไป ดู จำนวนมาก ยังมี ฟีเจอร์ ฟีเจอร์ นำมาใช้ ทำ การตลาด ออนไลน์ ดี ฟีเจอร์ คน ใช้งาน โพสต์ ปุ่ม กด แฮร์ หลายครั้ง กระแส ฟีเจอร์ โพสต์ แคป เจอร์ ต่อไป บอ แหล่ง สร้าง กระแส เจาะ วง แคบ บุคคล มูลค่า หวัง ได้ประโยชน์ บทความ วิเคราะห์ กระแส จีน ควบคู่ การตลาด ออนไลน์ จีน อ้าย วิเคราะห์ ข้อสงสัย สอบถาม เข้ามา	1, 3	1, 2, 3
ร้านอาหาร ติด ดาว แอ ป เหมย ถวน แอ ป รีวิว อาหาร ชื่อดัง จีน ที่นี่ คำตอบ	3	1, 3
ปล่อย คลิป โปรโมท คอน เส ริ์ต พบกัน วันที่ พฤษภาคม พบกัน ธีม เยาว์วัย กลับมา พบกัน อีกครั้ง ภูเขา แม่น้ำ บรรจบ เสียงเพลง อา จื่อ อา เวิน ลอย สายน้ำ เลิกรา	1, 2, 3	1, 2
ชาวโพลน ทั่งเมือง หิมะ ปี เยือน ฮอต ฮิต ติด เทรนด์ ปี หิมะ ตกใน ปักกิ่ง เน็ต พลาด ที่จะ แชะ ภาพ สวย แฮร์ ติด แฮ ช แท็ก	1, 2	1, 2, 3
เรื่องราว ทำความรู้จัก ศาสตร์ หอม วิธีการ กิน ข้าวแช่ ต้ารับ ชาววัง ร้านอาหาร สีลม วิลเลจ	1, 3	2
ดารา จีน ชื่อดัง อัพ โพส แคป ชั่น วันหยุด ผม เลิกงาน ทานข้าว พรุ่งนี้ ทำงานต่อ	1, 2	2
ภาพ เล่าเรื่อง ย้อน จีน ยุค สถานที่ท่องเที่ยว มีชื่อเสียง เมือง ไห โซ่ว มณฑล ไห่ หนาน คนไทย เกาะ ไหหลำ สร้าง ปี สถานที่ ถ่ายหนัง รองรับ อุตสาห กรรมการ สร้าง หนัง จีน เติบโต เป็นอย่างมาก ส่วนหนึ่ง สนามกอล์ฟ รีสอร์ท ชื่อดัง ลงทุน เฝิง เสี่ยว กัง ผู้กำกับ หนัง ชื่อดัง จีน บริษัทผู้ผลิต หนัง ไฮไลท์ พูดถึง โซน จำลอง วิถีชีวิต สถาปัตยกรรม ตึกราม บ้านช่อง คนจีน ยุค สาธารณรัฐ แรงบันดาลใจ หนังจีน เชิง ประวัติศาสตร์ ชื่อดัง เฝิง เสี่ยว กัง เรื่อง เล่า เรื่องราว ความอดอยาก คนจีน คนตาย ล้าน คนใน เหอหนาน ปี ๑๙ ที่อยู่ สงคราม จีน ญี่ปุ่น อ้าย มีโอกาส เยือน ที่นี่ กด ชัตเตอร์ ภาพ สไตล์ ขาว ดำ ย้อนยุค ภาพ ภาพ ภาพ อ้าย ถ่าย ดังนั้น ชอบ ขอให้ กด แฮร์ แทน รูป โพสต์ ที่มา ภาพ ขอบคุณ	1, 2	1, 2, 3
เรื่อง น่าสนใจ โอว หยาง จิน หยวน ดีไซเนอร์ เสื้อผ้า วัย ปี นคร จงซิ่ง ผัน ตัว ฮิต อาชีพ ดีไซเนอร์ หมั่นโถว พลิก โฉม อาหาร ชั้นพื้นฐาน หน้าตา แสน ธรรมดา จีน กลายเป็น ศิลปะ มูลค่า มากขึ้น ออกแบบ หมั่นโถว รูปแบบ สวยงาม เดิม ประยุกต์ ศิลปะ ประสบการณ์ ออกแบบ เสื้อผ้า สิบ ปี เช้ากับ ทำ ขนม ลูกค้า ซื้อ วับประทาน ยังมี สั่งทำ ประดับ ตกแต่ง โต๊ะอาหาร งานเลี้ยง พิธีการ ต่าง ๆ อีกด้วย ทำให้เกิด อาชีพ น่าสนใจ โด่งดัง โลก ออนไลน์ จีน ปี โอว หยาง จิน หยวน ทำ หมั่นโถว ชิ้น รูปทรง ไส่ รูปทรง สัตว์ ตัว การ์ตูน ส่งผล สินค้า เสียง ชื่นชม ต่างประเทศ หลักสูตร ทำ หมั่นโถว แบ่งปัน เทคนิค วิธี ต่าง ๆ แม่บ้าน รัก ทำ ขนม	1, 2, 3	1, 2
อาหาร ชุด สาขา พระราม อร่อย ฟิน ไม่ต้อง บิน ฮ่องกง	1, 3	3
นวัตกรรม เก๋ ประเทศ จีน เตา ปิ้ง ย่าง หมา ล่า มงคล งาน จัดแสดง สินค้า ธุรกิจ โรงแรม มหานคร เซี่ยงไฮ้ ประเทศ จีน สำหรับ น้ำเต้า ถือเป็น มงคล ความเชื่อ คนจีน ว่า ภาชนะ บรรจุ น้ำ อมตะ ติดตัว สุขภาพ แข็งแรง ปัดเป่า ดี ดังนั้น คน กิน ปิ้ง ย่า าง หมา ล่า เตา อันนี้ มงคล ร้อย แน่นอน ฮ่า	1, 2	1, 2, 3
เมนู โฮมเมด ยอดฮิต เมือง จีน ประจำปี บ้าน สังเกต สถานการณ์ แพร่ระบาด โรค ผลกระทบ วิถีชีวิต ผู้คน ออกมา ใช้ชีวิต ที่มา หา กิจกรรม ทำ ผัน ตัว จิตรกร เป็นยอด นัก ประดิษฐ์ คนสวน มือโปร ขาดไม่ได้ การผัน ตัว เชฟ กระทะ เหล็ก ชวน ดู เมนู โฮมเมด ยอดฮิต เมือง จีน เน็ต แฮร์ สูตร รวมไปถึง คลิป สอน ทำอาหาร เล่น คนดู น้ำลายสอ เดิน เข้าครัว เตรียม วัตถุดิบ แปลงร่าง แตจังกึม ทำอาหาร กิน บ้าน ถือเป็น ทางเลือก ฆ่าเวลา ว่าง สะอาด ปลอดภัย มั่นใจ ทำ ลำบาก เย็น รสชาติ ช่าง อภิรมย์ แอดมิน แนะนำ สั่ง กิน ตอบ โจทย์ ช่วยเหลือ ร้านอาหาร รายได้ หดหาย หน้า ร้าน ตัว รายละเอียด เมนู ยอดฮิต คลิก ดู รูปภาพ เลีย ยยย	1, 2	1, 2, 3
ท่อง โลก อินเตอร์เน็ต แอบ อัลบั้ม รูป เมือง เส้า ชิง ฝาก เพื่อน ๆ ช่อง คอมเม้นท์ ตั้งอยู่ มณฑล เจ้อเจียง ตะวันออก จีน หังโจว แอด มีโอกาส เยือน ปีก่อน ดู รูป ราวกับ ย้อน อนุรักษ์ ดี เมือง มีอายุ เก่าแก่ ปี ดู คลอง อ่อนช้อย วัฒนธรรม สมญานาม เวนิส บูรพา	1, 2, 3	1, 2

ภาพที่ 4.3 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Dataset A

ขณะที่เมื่อวิเคราะห์การทำนายผิดของโมเดล PhayaThaiBERT ใช้ฟีเจอร์อารมณ์ ในชุดข้อมูล Balanced แบบจำนวนป้ายกำกับแต่ละคลาสใกล้เคียงกัน และสัดส่วน Training ต่อ Test data เป็น 80 ต่อ 20 ซึ่งได้ประสิทธิภาพสูงสุดในกลุ่มนี้ ตามปรากฏในภาพที่ 4.4 พบว่า แม้จะมีข้อผิดพลาดทั้งทำนายไม่ครบและทำนายเกิน แต่โมเดลแสดงให้เห็นถึงความสมดุลของการทำนายแต่ละคลาสมากขึ้น โดยไม่เอนเอียงไปยังคลาสใดคลาสหนึ่ง โดยโมเดลสามารถเลือกป้ายกำกับได้อย่างสอดคล้องกับบริบทของข้อความมากยิ่งขึ้น มีแนวโน้มที่จะทำนายคลาสได้ตรงกับเจตนาของข้อความโดยรวม และลดการทายซ้ำซ้อนหรือลักษณะสุ่ม

text	true_labels	predicted_labels
อ้าย ชม ถ้ำ จีอ จิน อุทยาน ธรณี โลก ถ้ำ ชื่อว่า สวย จีน ถ้ำ ได้ยิน ชื่อ คน จินตนาการ สถานที่ ความลึกลับ เดา ช้างใน คน เข้าไป ถ้ำ ความกล้า เข้าไป สถานที่ มืดมิด ค้นหา โดยเฉพาะ คนใน ผู้ค้นพบ ถ้ำ คน สถานที่ท่องเที่ยว ดังเช่น ถ้ำ ถ้ำ จีอ จิง ซ่อนตัว เขตภูเขา อำเภอ จีอ จิน เมือง ปี้ เจี๋ย มณฑล กุ้ย โจว เป็นเวลา ทีม สำรวจ การท่องเที่ยว อำเภอ ปี้ เจี๋ย ค้นพบ ปี คศ ในขณะนั้น ถ้ำ ถ้ำ ต่า จี ค้นพบ การท่องเที่ยว ท้องถิ่น เข้ามา เตรียม ความพร้อม สถานที่ท่องเที่ยว ทางธรรมชาติ นักท่องเที่ยว เข้าชม เป็นครั้งแรก ปี ชื่อ ถ้ำ ถ้ำ จีอ จิน สถานที่ท่องเที่ยว นักท่องเที่ยว จีน ต่างประเทศ ให้ความสนใจ ชม ความงาม ถ้ำ ถ้ำ ตาย ยังมี หินงอก หินย้อย จุดเด่น ที่นี่ หินงอก หินย้อย เสา หิน ปี ะ ติ กรรม ทางธรรมชาติ ดู วิจิตร งดงาม ชวน จินตนาการ รูปร่าง สิ้นสุด อาทิ พระราชวัง เสา หิน พระพุทธรูป สำหรับ ขนาด ถ้ำ พื้นที่ ตารางกิโลเมตร บริเวณ โถง ขนาดใหญ่ ถ้ำ กิน พื้นที่ ตารางเมตร ถ้ำ กิโลเมตร เดิน เข้าชม ตอนนี้ กิโลเมตร เข้าชม ความงาม เนรมิต โดยธรรมชาติ การันตี นิตยสาร จัดอันดับ ถ้ำ สวย จีน ปี ถ้ำ จีอ จิน เป็นหนึ่ง สถานที่ท่องเที่ยว มณฑล กุ้ย โจว การพัฒนา ส่วนหนึ่ง แก้ ยกระดับ เศรษฐกิจ มณฑล ดีขึ้น อุตสาหกรรม การท่องเที่ยว ประสบความสำเร็จ ดีเยี่ยม ผลักดัน ถ้ำ สถานที่ท่องเที่ยว ระดับ สำเร็จ ปี ขึ้นทะเบียน อุทยาน ธรณี โลก ปี ที่ผ่านมา อีกด้วย โดยส่วนตัว อ้าย ชอบ ดู หินงอก หินย้อย สร้าง โดยธรรมชาติ ถ้ำ เป็นอย่างมาก สอน รู้ ทุกสิ่งทุกอย่าง มุมมอง มุม เกิดขึ้น ใย ห้วง ความคิด แปร ความหมาย เดิม บรรดา หินงอก หินย้อย เสา หิน รูปร่าง แปลกตา ถ้ำ เดินทาง เริ่มต้น เมือง กุ้ย หยาง มณฑล กุ้ย โจว บิน สุวรรณภูมิ นั่ง รถบัส สถานี สถานี รถบัส ทางไกล สำหรับ สถาน ที่เที่ยว มณฑล กุ้ย โจว บริเวณ ประตู ตะวันออกเฉียงเหนือ สนามกีฬา มณฑล กุ้ย โจว	1, 2, 3	1
หยาดเหงื่อ รอยยิ้ม คราบ น้ำตา ภาพ บอกเล่า เรื่องราว แดน มังกร ประจำปี กี่ ชั่วโมง ผ่านพ้น ปี ต้อนรับ เข้าสู่ ปี ปี ถือเป็น ปี เรื่องราว รสชาติ หยาดเหงื่อ หยด น้ำตา รอยยิ้ม แอด ประมวล ภาพ ภาพ บอกเล่า เรื่องราว แดน มังกร ทั้งปี ที่ผ่านมา คนรับ ชม คลิก รูปภาพ ย้อน ดู เรื่องราว ส่งท้าย ปี	1	1, 3
เชี่ยงไฮ้ สร้าง รถ พลังงาน ไฮโดรเจน ประหยัด พลังงาน	1, 3	1
ฮาร์ บิน จัดแสดง น้ำแข็ง แกะสลัก ละเอียด โลก	1, 3	1
อาหารจีน ชนิด ภูมิภาค เอกลักษณ์ เฉพาะตัว	2	3
ติด เทรนด์ เช้านี้ เป้าหมาย เป็นอิสระจาก รายการ เมื่อคืนนี้ หลังจากที่ น้อง ประโยค เด็ด สำหรับ เรียนจบ สำหรับ ผม เลิกงาน ขอให้ พวกเรา มีความสุข ขอบคุณ สนับสนุน รายการ ประกาศ คะแนน จบ น้อง คน เวที ภาพ ท่า มีม น้อง บิน เลิกงาน ปลผล โหวต น้อง	1, 2	1, 2, 3
เฉิน เซียง พยาบาล แผนก สูติ นารี แพทย์ โรงพยาบาล กลาง เมือง ปา มณฑล เสฉวน คนไข้ บกพร่อง ทางการ ได้ยิน เตรียมตัว ผ่า คลอด เร่งด่วน ภาพวาด การ์ตูน สื่อสาร ตรงกัน พยาบาล ลูกสาว คนไข้ สามี วาดภาพ ต่าง ๆ ที่อยู่ บ้าน คนไข้ เตรียมตัว การสื่อสาร ภาพ การผ่า คลอด เป็นไป ง่ายดาย ทารก คุณแม่ ปลอดภัย ทั้งคู่ เผยแพร่ เรื่อง ทำให้เกิด กระแส ชื่นชม ไหวพริบ การแก้ปัญหา นางพยาบาล ท่าน โลก อินเตอร์เน็ต ล้นหลาม ประกาศ	1	1, 3
เกิดขึ้น บทเรียน ดรามา่ แบรนด์ กลายเป็น แบรนด์ พัง แดน มังกร ปฏิเสธ จีน เป็นหนึ่ง ในประเทศ แบรนด์ ต่างชาติ แบรนด์ เข้าไป ทำ การตลาด จีน ประชากร อันดับ โลก ยังมี กำลังซื้อ ผู้ประกอบการ เจ้า เจาะ ลูกค้า จีน ตีตลาด ความรู้ ความเข้าใจ วัฒนธรรม การเมือง ลักษณะ สังคม ท้องถิ่น ก้าวพลาด ธุรกิจ พัง ไม่เป็นท่า เหมือนกับ เหตุการณ์ ดรามา่ เช้าตา คนจีน พลิก โฉม แบรนด์ กลายเป็น แบรนด์ พัง ชั่วพริบตา	1, 3	3
องค์การ อวกาศ แห่งชาติ จีน เผย ผลสำเร็จ ยิง จรวด ขนส่ง ลอง มาร์ช ขนส่ง ดาวเทียม เชิงพาณิชย์ ดวง อวกาศ เข้าสู่ วงโคจร เรียบร้อย ภารกิจ การสำรวจ ระยะไกล การสำรวจ ที่ดิน ก่อสร้าง เมือง การเกษตร ป่าไม้ พลังงาน การป้องกัน ลด ภัยพิบัติ มณฑล ชาน ตง ตะวันออก จีน ภารกิจ ดาวเทียม ดวง ทดสอบ เทคโนโลยี ออกแบบ ฐาน ดาวเทียม การถ่ายภาพ สังเกตการณ์ เรียลไทม์ ข้อมูล สังเกตการณ์ วัตถุ ท้องฟ้า ขนาดเล็ก ให้บริการ สำรวจ ระยะไกล ดาวเทียม ดวง ปล่อย ศูนย์ ปล่อย ดาวเทียม ไท่ หยวน มณฑล ชาน ซี ตอนเหนือ ประเทศ น ตามเวลา ปักกิ่ง ภารกิจ เที่ยวบิน จรวด ขนส่ง ตระกูล ลอง มาร์ช	1	1, 3
สถานการณ์ ในประเทศ จีน ผู้ป่วย เพิ่มขึ้น กลับบ้าน รักษาตัว โรงพยาบาล เสียชีวิต ยอด ผู้ป่วย สะสม ผู้ป่วย วัยหนุ่มสาว คน ทำงาน ลด กิจกรรม ทาง สังคม เว้น ระยะห่าง ดูแล สุขภาพ	1, 3	1
ส่อง เทศกาล วิ่งวัว กลาง ปลัก โคลน อินโดฯ วิ่งวัว หาง วัว วิ่ง ลื่นไถล ปลัก โคลน เทศกาล ปา จา วี การแข่งขัน วิ่งวัว ตั้งเดิม ระทึก อินโดนีเซีย ย่าน ทา นาห์ ดา ตาร์ เกาะ สุมาตรา ตะวันตก ประเทศ อินโดนีเซีย	1, 2	1
กาแฟ ลาเต้ หมี แพนด้า แก้ว โลก จีน กาแฟ แพนด้า แก้ว เคเอฟซี ประเทศ จีน แคมเปญ การตลาด สร้าง ความประหลาดใจ ความน่ารัก สัตว์ ผสมผสาน ผลิตภัณฑ์ บริการ ดึงดูด ครองใจ ผู้บริโภค แก้ว อุ้งเท้า แมว พลุ แตก ทำ แก้ว อุ้งเท้า หมา ตามมา วิเคราะห์ แคมเปญ แก้ว อุ้งเท้า แมว อุ้งเท้า หมา พูดถึง คัน กา ข้อมูล โลก ออนไลน์ จีน สรุป คนจีน พูดถึง แมว อันดับ โดยเฉพาะ สาว แมว น่ารัก หมา สัตว์เลี้ยง คนจีน ชอบ พูดถึง สำหรับ หมี แพนด้า เป็นหนึ่ง สัตว์ สัญลักษณ์ ประจำ ประเทศ จีน คนจีน ภาพ เชิง บวก หมี แพนด้า ความน่ารัก น่าเอ็นดู น่าแปลกใจ ทำ ลาเต้ อาร์ต หมี แพนด้า อ้าย อ่าน ความคิดเห็น คนจีน ลิ้มลอง กาแฟ หมี แพนด้า เคเอฟซี โพสต์ แบ่ง ชอบ น่ารัก รสชาติ ไม่เลว ดื่ม ขม ชอบ ลาย เหมือน หมี แพนด้า แก้ว ดูลาย ขึ้นอยู่กับ คน ทำ แต่ละคน บางคน ตอนที่ แก้ว ลาย เหมือน แพนด้า เดี๋ยวเดียว ลาย เปลี่ยนไป ที่อยู่ จีน อย่า ลืม ลอง แชร์ รสชาติ ชอบ ชอบ	1, 2, 3	1, 3

ภาพที่ 4.4 คลาสที่ทำนายผิดของโมเดล PhayaThaiBERT ใช้ฟีเจอร์อารมณ์ ของชุดข้อมูล Balanced แบบจำนวนป้ายกำกับใกล้เคียงกัน และสัดส่วน Train:Test 80:20

4.4 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์

จากตารางผลการทดลองการจำแนกอารมณ์ 8 กลุ่ม ดังปรากฏในตารางที่ 4.12 ได้แก่ Anger Anticipation Disgust Fear Joy Sadness Surprise และ Trust โดยใช้โมเดล mBERT, PhayaThaiBERT และ WangchanBERTa ทั้งในช่วง Validation และ Test พบว่าประสิทธิภาพของแต่ละโมเดลมีความแตกต่างกันทั้งในภาพรวมและรายคลาส โดยเฉพาะเมื่อนำไปประเมินกับชุดทดสอบจริง

โมเดล mBERT แสดงให้เห็นถึงความเสถียรในระดับหนึ่งระหว่างชุด Validation และ Test โดยมีค่า F1-score แบบ Micro ที่คงที่ใกล้เคียงกัน (70.87% และ 73.53% ตามลำดับ) อย่างไรก็ตาม ค่าความแม่นยำเฉลี่ยแบบ Macro ซึ่งไม่ถ่วงน้ำหนักตามจำนวนข้อมูลในแต่ละคลาสกลับอยู่ในระดับต่ำ โดยเฉพาะในชุดทดสอบที่มีค่าเพียง 38.53% สะท้อนว่าโมเดลไม่สามารถจำแนกอารมณ์ที่มีตัวอย่างจำนวนน้อยได้อย่างมีประสิทธิภาพ เช่น Disgust Sadness และ Fear ที่หลายคลาสมีค่า F1 เท่ากับศูนย์หรือใกล้ศูนย์ ขณะที่ PhayaThaiBERT มีค่า Micro F1-score สูงกว่า mBERT อย่าง

ต่อเนื่องทั้งใน Validation (76.00%) และ Test (77.96%) และแสดงให้เห็นถึงความสามารถในการจำแนกคลาสที่สำคัญ เช่น Trust และ Anticipation ได้ดีขึ้น แต่ในขณะเดียวกันก็ยังไม่สามารถจัดการกับคลาสที่มีข้อมูลจำกัดอย่าง Anger หรือ Disgust ได้อย่างมีประสิทธิภาพ

ในทางตรงกันข้าม WangchanBERTa ให้ผลลัพธ์ที่ดีที่สุดในทุกตัวชี้วัด โดยมี Micro F1-score อยู่ที่ 78.35% และ Macro F1-score ที่ 55.74% ในชุดทดสอบ ซึ่งสูงที่สุดในบรรดาทุกโมเดล นอกจากนี้ ยังสามารถรักษาความแม่นยำในคลาสที่มีปริมาณข้อมูลน้อยได้ดีกว่า เช่น fear, anger และ sadness แม้ค่าจะยังไม่สูงมาก แต่มีแนวโน้มดีกว่าระบบอื่นอย่างชัดเจน ส่งผลให้ค่าเฉลี่ย Macro ที่ได้มีความสมดุลมากขึ้น และใกล้เคียงกับค่าความแม่นยำในคลาสที่พบได้บ่อยอย่าง anticipation หรือ trust

ค่า Hamming Loss ซึ่งใช้วัดความผิดพลาดเฉลี่ยในระดับ label ต่อข้อความ โดยค่าที่ต่ำกว่าจะสะท้อนถึงความแม่นยำที่สูงขึ้น พบว่า WangchanBERTa มีค่า Hamming Loss ต่ำที่สุดที่ 10.27% ขณะที่ PhayaThaiBERT และ mBERT มีค่าอยู่ที่ 10.74% และ 12.76% ตามลำดับ แสดงให้เห็นว่า WangchanBERTa มีความแม่นยำในการจำแนกอารมณ์โดยรวมสูงที่สุด ทั้งในแง่ของการทำนายแบบครบถ้วนและลดข้อผิดพลาดในระดับข้อความ

เมื่อพิจารณาร่วมกับบริบทของปัญหาคลาสไม่สมดุลในชุดข้อมูล โดยเฉพาะกลุ่มอารมณ์เชิงลบหรือเชิงดราม่า ได้แก่ Anger Disgust Fear และ Sadness ซึ่งมีจำนวนตัวอย่างน้อยมากเมื่อเทียบกับคลาสอื่น เช่น Trust และ Anticipation ทำให้ค่า Macro F1-score มีความอ่อนไหวและลดลงตามประสิทธิภาพของโมเดลในคลาสเหล่านี้ การที่ WangchanBERTa สามารถรักษาค่า Macro F1 ได้สูงกว่าระบบอื่นสะท้อนถึงศักยภาพของโมเดลที่สามารถจัดการกับข้อจำกัดด้านโครงสร้างข้อมูลได้ดีกว่า และเหมาะสมที่จะนำไปใช้งานจริงในบริบทที่หลากหลายมากขึ้น

ตารางที่ 4.12 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์แบบ 8 อารมณ์

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Validation)	Anger	100.00%	22.22%	36.36%	13.75%
	Anticipation	76.63%	71.21%	73.82%	
	Disgust	0.00%	0.00%	0.00%	
	Fear	66.67%	18.18%	28.57%	
	Joy	72.66%	73.81%	73.23%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Sadness	40.00%	10.53%	16.67%	
	Surprise	56.79%	71.88%	63.45%	
	Trust	71.94%	79.66%	75.60%	
	Micro	71.29%	70.46%	70.87%	
	Macro	60.59%	43.44%	45.96%	
mBERT (Test)	Anger	0.00%	0.00%	0.00%	12.76%
	Anticipation	73.95%	85.78%	79.42%	
	Disgust	0.00%	0.00%	0.00%	
	Fear	33.33%	6.67%	11.11%	
	Joy	74.27%	74.71%	74.49%	
	Sadness	0.00%	0.00%	0.00%	
	Surprise	70.42%	62.50%	66.23%	
	Trust	76.44%	77.48%	76.96%	
	Micro	73.68%	73.38%	73.53%	
	Macro	41.05%	38.39%	38.53%	
PhayaThaiBERT (Validation)	Anger	80.00%	44.44%	57.14%	11.29%
	Anticipation	83.87%	78.79%	81.25%	
	Disgust	0.00%	0.00%	0.00%	
	Fear	71.43%	45.45%	55.56%	
	Joy	78.79%	82.54%	80.62%	
	Sadness	66.67%	21.05%	32.00%	
	Surprise	62.50%	62.50%	62.50%	
	Trust	73.71%	80.79%	77.09%	
	Micro	76.77%	75.25%	76.00%	
	Macro	64.62%	51.95%	55.77%	
PhayaThaiBERT (Test)	Anger	0.00%	0.00%	0.00%	10.74%
	Anticipation	78.31%	86.67%	82.28%	
	Disgust	0.00%	0.00%	0.00%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
	Fear	50.00%	6.67%	11.76%	
	Joy	77.27%	80.00%	78.61%	
	Sadness	40.00%	11.76%	18.18%	
	Surprise	90.16%	68.75%	78.01%	
	Trust	74.23%	86.94%	80.08%	
	Micro	**77.29%**	**78.65%**	**77.96%**	
	Macro	**51.25%**	**42.60%**	**43.62%**	
WangchanBERTa (Validation)	Anger	0.00%	0.00%	0.00%	**10.28%**
	Anticipation	79.83%	84.44%	82.07%	
	Disgust	0.00%	0.00%	0.00%	
	Fear	50.00%	40.00%	44.44%	
	Joy	82.24%	73.53%	77.64%	
	Sadness	80.00%	23.53%	36.36%	
	Surprise	80.77%	78.75%	79.75%	
	Trust	79.39%	81.53%	80.44%	
	Micro	79.80%	76.89%	78.32%	
	Macro	56.53%	47.72%	50.09%	
WangchanBERTa (Test)	Anger	100.00%	33.33%	50.00%	**10.27%**
	Anticipation	81.73%	81.31%	81.52%	
	Disgust	0.00%	0.00%	0.00%	
	Fear	66.67%	36.36%	47.06%	
	Joy	75.00%	80.95%	77.86%	
	Sadness	66.67%	21.05%	32.00%	
	Surprise	77.05%	73.44%	75.20%	
	Trust	78.46%	86.44%	82.26%	
	Micro	78.48%	78.22%	78.35%	
	Macro	68.20%	51.61%	55.74%	

จากผลลัพธ์ข้างต้น พบว่าชุดข้อมูลเนื้อหาเกี่ยวกับจีน มีความไม่สมดุลและเนื้อหาปริมาณน้อยมากสำหรับ 4 อารมณ์เชิงลบหรือดราม่า ได้แก่ Anger Disgust Fear และ Sadness ซึ่งส่งผลต่อประสิทธิภาพของโมเดล แต่จากการสังเกตพบว่า อีก 4 อารมณ์ที่เหลือ มีค่า Precision ค่า Recall และ ค่า F1 score ที่สูงเกินราว 70% ดังนั้น วิทยานิพนธ์นี้จึงทดลองเพิ่มเติม โดยตัดเหลือ 4 อารมณ์ ได้แก่ Anticipation Joy Surprise และ Trust จากนั้นทดสอบโมเดลอีกครั้ง

หลังจากทดลองจำแนกอารมณ์ในกลุ่ม 4 อารมณ์ ข้างต้น พบว่าโมเดลทั้ง 3 มีระดับประสิทธิภาพที่ดีขึ้นจาก 8 อารมณ์ โดยเฉพาะเมื่อเปรียบเทียบระหว่างเทรนด้วยชุด Validation กับผลลัพธ์บนชุด Test dataset ดังตารางที่ 4.13

ตารางที่ 4.13 ผลการทดลองที่ 4 โมเดลจำแนกอารมณ์แบบ 4 อารมณ์

Model	Class	Precision	Recall	F1-Score	Hamming Loss
mBERT (Test)	Anticipation	79.29%	79.29%	79.29%	22.41%
	Joy	72.58%	71.43%	72.00%	
	Surprise	62.69%	65.62%	64.12%	
	Trust	72.96%	80.79%	76.68%	
	Micro	73.85%	76.46%	75.13%	
	Macro	71.88%	74.28%	73.02%	
mBERT (Validation)	Anticipation	76.07%	79.11%	77.56%	23.11%
	Joy	70.83%	70.00%	70.41%	
	Surprise	77.19%	55.00%	64.23%	
	Trust	73.44%	84.68%	78.66%	
	Micro	73.99%	75.90%	74.93%	
	Macro	74.38%	72.20%	72.72%	
PhayaThaiBERT (Validation)	Anticipation	81.62%	84.89%	83.22%	17.30%
	Joy	82.14%	81.18%	81.66%	
	Surprise	77.65%	82.50%	80.00%	
	Trust	79.72%	77.93%	78.82%	
	Micro	80.68%	81.49%	81.08%	
	Macro	80.28%	81.62%	80.92%	

Model	Class	Precision	Recall	F1-Score	Hamming Loss
PhayaThaiBERT (Test)	Anticipation	81.44%	79.80%	80.61%	19.83%
	Joy	75.35%	84.92%	79.85%	
	Surprise	67.69%	68.75%	68.22%	
	Trust	75.96%	78.53%	77.22%	
	Micro	76.71%	79.29%	77.98%	
	Macro	75.11%	78.00%	76.45%	
WangchanBERTa (Test)	Anticipation	80.20%	79.80%	80.00%	18.65%
	Joy	76.19%	76.19%	76.19%	
	Surprise	71.23%	81.25%	75.91%	
	Trust	78.17%	87.01%	82.35%	
	Micro	**77.57%**	**81.42%**	**79.45%**	
	Macro	**76.45%**	**81.06%**	**78.61%**	
WangchanBERTa (Validation)	Anticipation	80.42%	85.78%	83.01%	**16.84%**
	Joy	82.50%	77.65%	80.00%	
	Surprise	82.28%	81.25%	81.76%	
	Trust	78.99%	84.16%	81.74%	
	Micro	**80.61%**	**82.93%**	**81.75%**	
	Macro	**81.05%**	**82.34%**	**81.63%**	

จากผลการทดลองเปรียบเทียบประสิทธิภาพของโมเดลในการจำแนกอารมณ์หลัก 4 ประเภท ได้แก่ Anticipation Joy Surprise และ Trust พบว่าโมเดลทั้งสามชุดที่ใช้ในการจำแนกอารมณ์ ได้แก่ mBERT, PhayaThaiBERT และ WangchanBERTa มีลักษณะการทำงานที่แตกต่างกันอย่างชัดเจนในแต่ละช่วงของการประเมิน ทั้งในชุด Validation และ Test โดยสามารถวิเคราะห์ได้ทั้งเชิงรายคลาสและเชิงภาพรวม สำหรับโมเดล mBERT มีค่า F1-score โดยรวมแบบ Micro ในชุด Validation เท่ากับ 74.93% และในชุด Test อยู่ที่ 75.13% ซึ่งถือว่ามีความเสถียรระดับหนึ่งเมื่อประเมินจากความแตกต่างระหว่างช่วงการฝึกและทดสอบจริง อย่างไรก็ตาม ค่า Macro F1-score ของ mBERT กลับลดลงเล็กน้อยในชุด Test (จาก 72.72% เหลือ 73.02%) สะท้อนให้เห็นว่าประสิทธิภาพของโมเดลยังขึ้นอยู่กับลักษณะข้อมูลในแต่ละคลาส โดยเฉพาะคลาส surprise ที่มีค่า

Precision และ Recall ไม่สมดุลในช่วง Validation และลดลงต่อเนื่องใน Test ขณะที่ anticipation และ trust ยังคงมีค่า F1 ที่สูงและเสถียรตลอดช่วงการประเมิน

PhayaThaiBERT แสดงผลลัพธ์ที่ดีขึ้นอย่างเห็นได้ชัดในทุกมิติ โดยในช่วง Validation มีค่า Micro F1-score สูงถึง 81.08% และ Macro F1-score เท่ากับ 80.92% ซึ่งถือว่าเป็นค่าที่สอดคล้องกันอย่างมาก และสะท้อนว่าโมเดลสามารถจัดการกับคลาสทั้ง 4 ได้อย่างสมดุล โดยเฉพาะ anticipation และ joy ที่มีค่า F1-score มากกว่า 81% ทั้งคู่ ขณะที่ในชุด Test แม้ค่าจะลดลงเล็กน้อย แต่ยังคงแสดงความแม่นยำที่สูงกว่ามาตรฐานในทุกคลาส โดยเฉพาะ trust ที่มีค่า F1-score อยู่ที่ 77.22% และ joy ที่เพิ่มขึ้นเป็น 79.85% แสดงให้เห็นถึงความสามารถในการ generalize ของโมเดลบนข้อมูลใหม่ได้ดี

ด้าน WangchanBERTa ถือเป็นโมเดลที่มีประสิทธิภาพสูงสุดในภาพรวม ทั้งในช่วง Validation และ Test โดยมีค่า Micro F1-score สูงสุดในชุด Test ที่ 79.45% และ Macro F1-score เท่ากับ 78.61% ซึ่งสูงกว่าโมเดลอื่นอย่างมีนัยสำคัญ อีกทั้งยังรักษาค่าความแม่นยำในแต่ละคลาสได้อย่างสมดุล ทั้ง Anticipation Joy Surprise และ Trust โดยเฉพาะในคลาส Trust ที่มีค่า Recall สูงถึง 87.01% และ F1-score ที่ 82.35% แสดงถึงความสามารถในการตรวจจับอารมณ์ที่มีนัยต่อบริบทเชิงความมั่นใจได้อย่างแม่นยำ นอกจากนี้ ในชุด Validation WangchanBERTa ยังให้ค่า F1 ในแต่ละคลาสสูงเกิน 80% เกือบทั้งหมด และค่า Macro F1-score สูงถึง 81.63% ซึ่งบ่งชี้ว่าโมเดลสามารถเรียนรู้จากข้อมูลฝึกได้ลึก และส่งต่อประสิทธิภาพไปยังการทดสอบจริงได้อย่างเสถียร

ในแง่ของค่า Hamming Loss ซึ่งใช้วัดอัตราความผิดพลาดเฉลี่ยต่อ label โดยยิ่งมีค่าน้อยแสดงว่าโมเดลมีความแม่นยำในการจำแนกได้มากขึ้น พบว่า WangchanBERTa มีค่า Hamming Loss ต่ำที่สุดทั้งใน Validation (16.84%) และ Test (18.65%) ตามลำดับ รองลงมาคือ PhayaThaiBERT (17.30% และ 19.83%) และ mBERT ซึ่งมีค่า Hamming Loss สูงที่สุดที่ 22.41% ใน Test และ 23.11% ใน Validation สะท้อนว่าประสิทธิภาพของ WangchanBERTa มีความเสถียรทั้งเชิงความแม่นยำและความผิดพลาดโดยเฉลี่ย

จากการวิเคราะห์ภาพรวมทั้งหมดสามารถสรุปได้ว่า WangchanBERTa มีความสามารถสูงสุดในการจำแนกอารมณ์ 4 กลุ่มในบริบทภาษาไทยทั้งในชุดข้อมูลฝึกและทดสอบ รองลงมาคือ PhayaThaiBERT ซึ่งยังคงรักษาระดับความแม่นยำได้อย่างดี ส่วน mBERT แม้จะมีความสามารถในการจำแนกคลาสหลักได้ในระดับหนึ่ง แต่ยังมีข้อจำกัดในด้านความแม่นยำในคลาสเฉพาะ และมีความ

เปลี่ยนแปลงของค่าเมตริกที่สูงกว่าระบบอื่นเมื่อเปลี่ยนจาก Validation ไปสู่ Test ซึ่งสามารถเป็นแนวทางสำหรับวิจัยเพื่อพัฒนาให้มีประสิทธิภาพดียิ่งขึ้นได้ในอนาคต

4.5 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง

สำหรับการผลการทดลองที่ 5 แบ่งออกเป็นผลการทดลองย่อย 2 ผลการทดลองได้แก่ ผลการทดลองโดยไม่ใช้ Feature Engineering และผลการทดลองที่ใช้ Feature Engineering ดังรายละเอียดต่อไปนี้

4.5.1 ผลการทดลองโดยไม่ใช้ Feature Engineering

ผลการทดลองนี้แสดงให้เห็นถึงประสิทธิภาพของโมเดลต่างๆ ในการจำแนกประเภทข้อความระหว่าง Translation และ Storytelling โดยไม่ใช้ feature engineering ซึ่งพบว่าโดยภาพรวมโมเดลที่ไม่ใช่กลุ่ม Transfomer มีประสิทธิภาพดีที่สุดในแง่ของ Accuracy ดังปรากฏในตารางที่ 4.14

โดย Random Forest ให้ผลลัพธ์โดยรวมดีที่สุดด้วย Accuracy 88.20% และ F1-score ที่สูงทั้งในคลาส Translation (89.14%) และ Storytelling (87.07%) สะท้อนให้เห็นว่าโมเดลนี้สามารถแยกแยะลักษณะของข้อความสองประเภทได้อย่างมีประสิทธิภาพ โดยเฉพาะในคลาส Translation ที่ได้ Recall สูงถึง 91.76% หมายความว่าสามารถตรวจจับข้อความแปลได้อย่างครอบคลุม แม้ Precision จะต่ำกว่าเล็กน้อยเมื่อเทียบกับคลาส Storytelling ขณะที่ XGBoost, LightGBM และ AdaBoost มี Accuracy อยู่ในช่วงประมาณ 82% ถึง 83% โดยแต่ละโมเดลมีจุดแข็งแตกต่างกัน เช่น LightGBM แม้จะมี Precision ไม่สูงที่สุด แต่สามารถตรวจจับข้อความในคลาส Translation ได้มากที่สุด (Recall 89.41%) เช่นเดียวกับ AdaBoost ที่มีลักษณะใกล้เคียงกันและมี F1-score ในคลาส Translation สูงถึง 84.92% ซึ่งสะท้อนว่า Boosting methods เหล่านี้มีความสามารถในการเรียนรู้ feature pattern ที่ซับซ้อนและครอบคลุมบริบทได้ดีพอสมควร โดยเฉพาะในกลุ่มข้อมูลที่มีลักษณะโครงสร้างชัดเจนอย่างข้อความแปล

เมื่อพิจารณาโมเดลในกลุ่ม Transformer พบว่า WangchanBERTa มีความแม่นยำโดยรวมสูงที่สุดในกลุ่มโมเดลภาษา (80.75%) โดยมี F1-score ที่ดีทั้งในคลาส Translation (81.21%) และ Storytelling (80.25%) ซึ่งบ่งชี้ว่าการใช้โมเดล pretrained ที่ถูกฝึกจากภาษาไทยโดยเฉพาะสามารถจับความหมายและบริบทของข้อความได้ดีกว่าโมเดล multilingual อย่าง mBERT ที่มี Accuracy ต่ำเพียง 59.01% และมี Recall ต่ำมากในคลาส Translation (31.76%) สะท้อนถึงข้อจำกัดของการใช้โมเดลข้ามภาษาโดยไม่ fine-tune ด้วยข้อมูลภาษาไทยด้าน PhayaThaiBERT ซึ่งเป็นโมเดล pretrained ที่เทรนจากชุดข้อมูลที่ใหญ่กว่า WangchanBERTa และถูกคาดหวังว่า

สามารถจับบริบทของข้อความได้มากที่สุด ให้ผลลัพธ์ที่ต่ำกว่า WangchanBERTa ซึ่งอาจเกี่ยวข้องกับชุดข้อมูลในการทดลองเป็นชุดข้อมูลขนาดเล็ก โดยสอดคล้องกับผลการทดลอง 4.1 ที่ WangChanBERTa มีประสิทธิภาพสูงที่สุด และเป็นชุดข้อมูลขนาดเล็กเช่นกัน โดยแสดงให้เห็นว่าการเลือกโมเดล pretrained ที่มีฐานข้อมูลใกล้เคียงกับลักษณะของข้อความที่วิเคราะห์ จะช่วยเพิ่มประสิทธิภาพการจำแนกได้อย่างมีนัยสำคัญ

ตารางที่ 4.14 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง แบบไม่ใช้ฟีเจอร์

Model	Accuracy	Class	Precision	Recall	F1-score
Random Forest	**88.20%**	**Translation**	**86.67%**	**91.76%**	**89.14%**
		Storytelling	**90.14%**	**84.21%**	**87.07%**
XGBoost	81.99%	Translation	81.82%	84.71%	83.24%
		Storytelling	82.19%	78.95%	80.54%
LightGBM	82.61%	Translation	80.00%	89.41%	84.44%
		Storytelling	86.36%	75.00%	80.28%
AdaBoost	83.23%	Translation	80.85%	89.41%	84.92%
		Storytelling	86.57%	76.32%	81.12%
mBERT	59.01%	Translation	77.14%	31.76%	45.00%
		Storytelling	53.97%	89.47%	67.33%
PhayaThaiBERT	76.40%	Translation	80.52%	72.94%	76.54%
		Storytelling	72.62%	80.26%	76.25%
WangchanBERTa	80.75%	Translation	83.75%	78.82%	81.21%
		Storytelling	77.78%	82.89%	80.25%

4.5.2 ผลการทดลองของโมเดลที่มี Feature Engineering

จากผลการทดลองที่ปรากฏในตารางที่ 4.15 ซึ่งเป็นการประเมินประสิทธิภาพของโมเดลในการจำแนกประเภทข้อความระหว่าง Translation และ Storytelling โดยใช้วิธีการ feature engineering ตามที่อธิบายไว้ในบทที่ 3 สามารถวิเคราะห์เชิงเปรียบเทียบกับกรณีที่ไม่ใช้ feature engineering ได้ดังนี้

ผลการทดลองแสดงให้เห็นว่า การเพิ่มฟีเจอร์ เช่น ความยาวของข้อความ จำนวนคำประเภทต่าง ๆ (POS tags) การตรวจจับชื่อเฉพาะ (NER) และค่าความรู้สึก (sentiment score) ส่งผลต่อการเพิ่มประสิทธิภาพของโมเดลอย่างชัดเจน โดยเฉพาะในกลุ่มโมเดลที่ไม่ใช่ Transformer

Random Forest เป็นโมเดลที่มีประสิทธิภาพสูง วัดด้วยค่า Accuracy โดยเพิ่มขึ้นจาก 88.20% (ไม่ใช้ฟีเจอร์) เป็น 92.55% เมื่อใช้ feature engineering และมีค่า F1-score สำหรับคลาส Translation และ Storytelling สูงถึง 92.77% และ 92.31% ตามลำดับ แสดงให้เห็นถึงความสามารถของโมเดลในการจำแนกข้อความสองประเภทนี้ได้อย่างแม่นยำและสมดุล นอกจากนี้ค่า Precision และ Recall ในทั้งสองคลาสก็อยู่ในระดับสูงอย่างสม่ำเสมอ สะท้อนถึงประโยชน์ของฟีเจอร์ที่ช่วยเสริมความเข้าใจเชิงโครงสร้างของโมเดล

ขณะที่ XGBoost และ LightGBM มีแนวโน้มที่คล้ายกันกับข้างต้น โดย Accuracy เพิ่มขึ้นจากประมาณ 82% เป็น 87.58% และ 86.96% ตามลำดับ พร้อมกับค่า F1-score ที่สูงขึ้นในทั้งสองคลาสเช่นกัน แสดงให้เห็นว่า boosting-based models ได้รับประโยชน์จากฟีเจอร์ที่ออกแบบมาอย่างมีระบบเช่นเดียวกับ Random Forest โดยเฉพาะในส่วนของ Recall ของคลาส Translation ซึ่งสูงเกิน 89% ในทั้งสองโมเดล สำหรับ AdaBoost แม้จะมี Accuracy ต่ำสุดในโมเดลที่ไม่ใช่ Transformer โดยมี accuracy 80.75% แต่ยังคงแสดงให้เห็นว่า feature engineering ช่วยทำให้ F1-score ของคลาส Translation ได้ถึง 83.60% ซึ่งสูงกว่าการทดลองในกรณีที่ไม่มีฟีเจอร์

ในส่วนของโมเดลกลุ่ม Transformer พบว่าแม้การเพิ่มฟีเจอร์ จะช่วยให้ผลลัพธ์ดีขึ้นแต่ยังคงมีประสิทธิภาพด้อยกว่าบางโมเดลในกลุ่มที่ไม่ใช่ Transformer โดยเฉพาะในกรณีของ mBERT ที่มี Accuracy อยู่ที่ 62.11% แม้จะเพิ่มจากไม่ใช้ฟีเจอร์ และ Recall ในคลาส Storytelling เพิ่มเป็น 77.63% จากเดิมที่ต่ำที่สุด ราว 33% แต่ยังมีค่า Accuracy และ F1-score โดยรวมต่ำกว่ากลุ่มโมเดลอื่นอย่างมีนัยสำคัญ สะท้อนข้อจำกัดของการใช้โมเดล multilingual ที่ไม่ได้เทรนเฉพาะกับภาษาไทย WangchanBERTa มี Accuracy เพิ่มขึ้นเป็น 85.09% พร้อมค่า F1-score สำหรับคลาส Storytelling สูงถึง 85.88% และ Recall สูงถึง 96.05% ส่วน PhayaThaiBERT มี Accuracy และ F1 score ของทั้งสองคลาสสูงกว่า 80% แต่ยังคงมีประสิทธิภาพด้อยกว่า WangchanBERTa เช่นเดียวกับแบบไม่ใช้ฟีเจอร์ อย่างไรก็ตาม Accuracy ของ WangchanBERTa และ PhayaThaiBERT ที่สูงกว่า mBERT อย่างชัดเจน แสดงให้เห็นถึงความสามารถของโมเดลที่ได้รับการ pretrain ด้วยข้อมูลภาษาไทยโดยตรง และสามารถตรวจจับลักษณะการเล่าเรื่องซึ่งมักมีบริบทซับซ้อนได้อย่างแม่นยำ

จากผลลัพธ์ที่กล่าวมา สามารถสรุปได้ว่า การใช้กระบวนการ feature engineering ส่งผลอย่างมีนัยสำคัญต่อประสิทธิภาพของโมเดลจำแนกข้อความ โดยเฉพาะในกลุ่มโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิมที่ไม่ได้อยู่ในกลุ่ม Transformer ซึ่งอาศัยข้อมูลเชิงโครงสร้างอย่างชัดเจนเป็นหลัก

ในทางตรงกันข้าม โมเดลแบบ Transformer มีความสามารถในการเรียนรู้บริบทจากข้อความดิบโดยตรง และจึงได้รับผลกระทบน้อยกว่าเมื่อไม่มีฟีเจอร์เสริม อย่างไรก็ตาม ผลการทดลองชี้ให้เห็นถึงความสำคัญของการวิจัยเชิงสำรวจเพื่อสกัดองค์ความรู้จากผู้เชี่ยวชาญ เพื่อนำไปออกแบบฟีเจอร์ที่สะท้อนโครงสร้างของเนื้อหาได้อย่างมีประสิทธิภาพ ซึ่งสามารถช่วยยกระดับผลลัพธ์ของโมเดลแบบดั้งเดิมให้ใกล้เคียงหรือในบางกรณีเหนือกว่าโมเดลแบบ Transformer ที่ต้องพึ่งพาข้อมูลขนาดใหญ่ในการฝึก

ตารางที่ 4.15 ผลการทดลองที่ 5 โมเดลจำแนกเนื้อหาการแปล และการเล่าเรื่อง แบบใช้ฟีเจอร์

Model	Accuracy	Class	Precision	Recall	F1-score
Random Forest	**92.55%**	**Translation**	**95.06%**	**90.59%**	**92.77%**
		Storytelling	**90.00%**	**94.74%**	**92.31%**
XGBoost	87.58%	Translation	87.36%	89.41%	88.37%
		Storytelling	87.84%	85.53%	86.67%
LightGBM	86.96%	Translation	86.36%	89.41%	87.86%
		Storytelling	87.67%	84.21%	85.91%
AdaBoost	80.75%	Translation	75.96%	92.94%	83.60%
		Storytelling	89.47%	67.11%	76.69%
mBERT	62.11%	Translation	70.69%	48.24%	57.34%
		Storytelling	57.28%	77.63%	65.92%
PhayaThaiBERT	80.75%	Translation	86.49%	75.29%	80.50%
		Storytelling	75.86%	86.84%	80.98%
WangchanBERTa	85.09%	Translation	95.52%	75.29%	84.21%
		Storytelling	77.66%	96.05%	85.88%

4.6 สรุปโมเดลที่ดีที่สุดของแต่ละรูปแบบการประเมินความสอดคล้อง และผลการพัฒนา API

จากการทดลองเปรียบเทียบประสิทธิภาพของโมเดลในแต่ละรูปแบบของการจำแนกบทความ พบว่าโมเดลที่ให้ผลลัพธ์ดีที่สุดในแต่ละมิติมีลักษณะและเทคนิคที่แตกต่างกัน ซึ่งส่งผลต่อแนวทางการเรียกใช้งานภายในระบบ API ตามที่ได้อธิบายไว้ในบทที่ 3 โดยรายละเอียดของโมเดลที่ได้รับการคัดเลือกว่าให้ประสิทธิภาพสูงสุดในแต่ละรูปแบบ และถูกนำมาใช้งานจริงในระบบ API ดังต่อไปนี้

4.6.1 การจำแนก Timely และ Timeless

ผลการทดลองที่ 1 พบว่าโมเดลที่พัฒนาด้วยสถาปัตยกรรม Transformer อย่างโมเดล WangchanBERTa มีประสิทธิภาพมากที่สุด โดยโมเดลถูกจัดเก็บในรูปแบบ .bin และโหลดด้วย PyTorch ซึ่งภายใน API จะใช้ไฟล์ best_wangchanberta_model.bin ร่วมกับ tokenizer ของ WangchanBERTa (airesearch/wangchanberta-base-att-spm-uncased) ในการประมวลผลและให้ผลลัพธ์ในลักษณะ binary classification ระหว่างเนื้อหา Timely และ Timeless

4.6.2 การจำแนก Translation และ Storytelling

ผลการทดลองที่ 5 พบว่าโมเดลที่ให้ผลลัพธ์ประสิทธิภาพดีที่สุดคือโมเดลที่พัฒนาด้วย RandomForest ซึ่งเป็นโมเดลประเภท ensemble ที่เรียนรู้จากฟีเจอร์ที่ได้จากการทำ TF-IDF ร่วมกับ Feature engineering ซึ่งหลังจากผ่านกระบวนการลดมิติฟีเจอร์และปรับสเกล โมเดลดังกล่าวถูกบันทึกในรูปแบบ .pkl และโหลดใช้งานผ่านไลบรารี joblib ประกอบด้วยไฟล์สำคัญ ได้แก่ best_ensemble_model.pkl, tfidf_vectorizer.pkl, scaler.pkl และ pca_transform.pkl

โดยสาเหตุที่โมเดลซึ่งพัฒนาด้วยอัลกอริธึม Random Forest ไม่สามารถใช้งานได้จากเพียงไฟล์เดียว (best_ensemble_model.pkl) แต่จำเป็นต้องใช้ไฟล์ประกอบอื่นร่วมด้วยนั้น เป็นเพราะต้องดำเนินการร่วมกับกระบวนการเตรียมข้อมูล (preprocessing) ที่มีความสอดคล้องกับขั้นตอนการฝึกโมเดล โดยเฉพาะการทำ feature engineering ที่ถูกออกแบบไว้ล่วงหน้า เพื่อให้โมเดลสามารถประมวลผลข้อมูลใหม่ได้อย่างถูกต้องและสอดคล้องกับข้อมูลที่เทรน
แต่ละไฟล์ มีรายละเอียดและหน้าที่ดังนี้

1) best_ensemble_model.pkl โมเดลหลัก

2) tfidf_vectorizer.pkl

เป็นไฟล์เก็บอ็อบเจกต์ TF-IDF Vectorizer ซึ่งถูกฝึกด้วยข้อความชุดเดียวกับตอนฝึกโมเดลหลัก โดยมีหน้าที่แปลงข้อความที่ผ่านการ preprocessing แล้ว ให้กลายเป็นเวกเตอร์เชิงความถี่ของคำในรูปแบบ sparse vector ก่อนเข้าสู่โมเดล จำเป็นต้องใช้ vectorizer ตัวเดียวกันกับตอน train เพื่อให้การแปลงเวกเตอร์ในตอนทำนายมีความสอดคล้อง

3) scaler.pkl

เป็นไฟล์เก็บอ็อบเจกต์ของ StandardScaler จาก scikit-learn ซึ่งใช้ในการปรับค่าฟีเจอร์เชิงตัวเลข เช่น ความยาวของข้อความ หรือฟีเจอร์อื่น ๆ ให้อยู่ในช่วงมาตรฐาน เพื่อช่วยให้โมเดลสามารถเรียนรู้จากฟีเจอร์ต่างชนิดกันได้อย่างมีประสิทธิภาพ

4) pca_transform.pkl

เป็นไฟล์เก็บอ็อบเจกต์ PCA สำหรับการลดมิติของเวกเตอร์ TF-IDF เพื่อช่วยลดภาระในการประมวลผลและลดความเสี่ยงจากปัญหา overfitting ขณะเดียวกันก็รักษาความสำคัญของข้อมูลไว้ได้

4.6.3 การจำแนก Intention ของเนื้อหา

ผลการทดลองที่ 2 พบว่าโมเดลที่ให้ผลลัพธ์ดีที่สุดคือโมเดลที่ใช้ WangchanBERTa ร่วมกับฟีเจอร์อารมณ์ โดยเก็บโมเดลที่ดีที่สุดในรูปแบบbest_intention_model.bin และใช้ tokenizer clicknext/phayathaibert เช่นเดียวกับขณะเทรนและทดสอบ

ขณะที่ผลการทดลองที่ 3 ทดสอบโมเดลการจำแนก Intention ของเนื้อหาแบบทดสอบความสมดุลของป้ายกำกับ ได้นำโมเดลที่ดีที่สุดจากรูปแบบจำนวนป้ายกำกับของแต่ละคลาสใกล้เคียงกันแบบ 80:20 และ โมเดลที่ดีที่สุดจากรูปแบบจำนวนป้ายกำกับของแต่ละคลาสจำนวนเท่ากันแบบ Dataset A ไปเชื่อมกับระบบต้นแบบแยกเฉพาะเจาะจง เพื่อทดสอบการใช้งานจริง เพื่อให้ทราบผลลัพธ์ความแตกต่างระหว่างชุดข้อมูลดั้งเดิมและชุดข้อมูลที่แก้ไขความสมดุลในแต่ละรูปแบบ

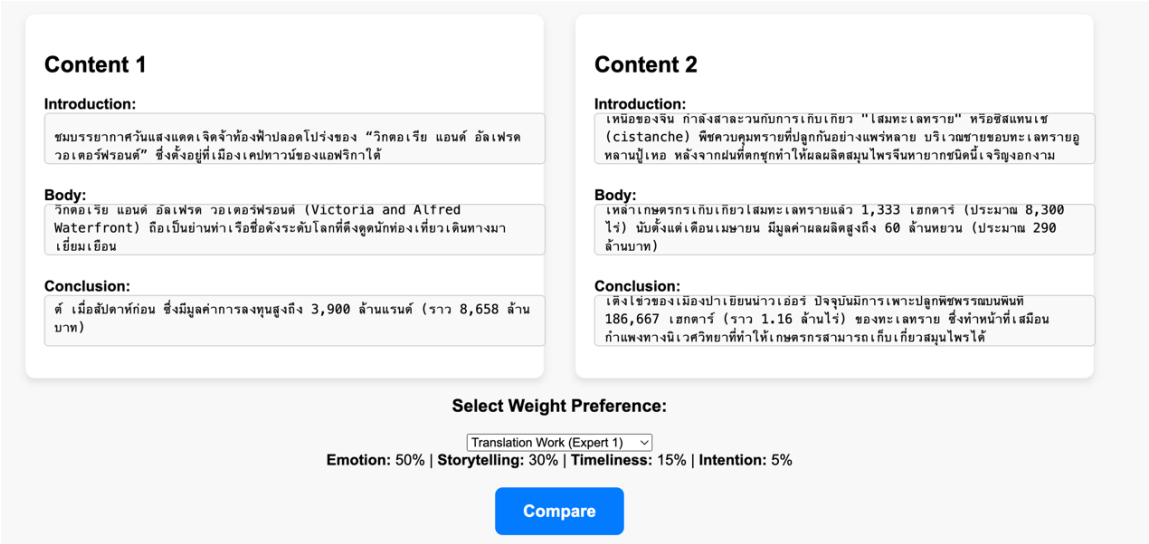
4.6.4 การจำแนกอารมณ์ของเนื้อหา

แม้ในการทดลองที่ 4 พบว่า โมเดลที่วัดอารมณ์ 8 อารมณ์ มีปัญหาเรื่องข้อมูลน้อยและไม่สมดุลในบางอารมณ์ โดยเฉพาะอารมณ์ที่สื่อถึงเชิงลบ หรือ ดราม่า ได้แก่ Anger Disgust Sadness และ Fear แต่จากการปรึกษาและสัมภาษณ์ผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับชุดข้อมูล พบว่า การนำเสนอครบทั้ง 8 ข้อมูลในระบบต้นแบบ จะทำให้เห็นภาพรวมได้ดีที่สุด ดังนั้น วิทยานิพนธ์นี้ยังคงเลือกใช้โมเดลจำแนกอารมณ์แบบ 8 อารมณ์

โดยโมเดลที่มีประสิทธิภาพที่ดีที่สุดสำหรับจำแนกอารมณ์ 8 อารมณ์ ได้แก่ โมเดล WangchanBERTa ซึ่งจัดเก็บในไฟล์ best_emotion_model.bin และใช้ร่วมกับ tokenizer ของ WangchanBERTa ใน API

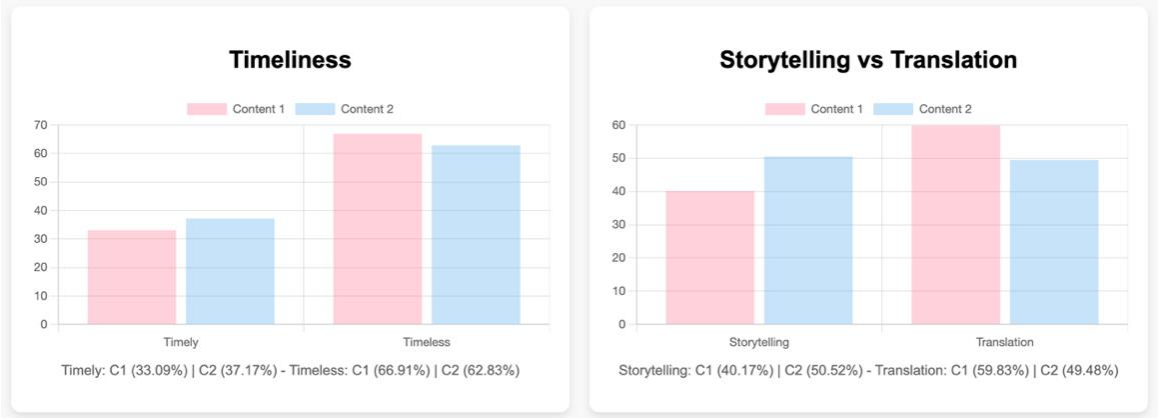
4.7 ผลทดสอบการใช้งานจริง Web application

4.7.1 ตัวอย่างที่ 1 ทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีน



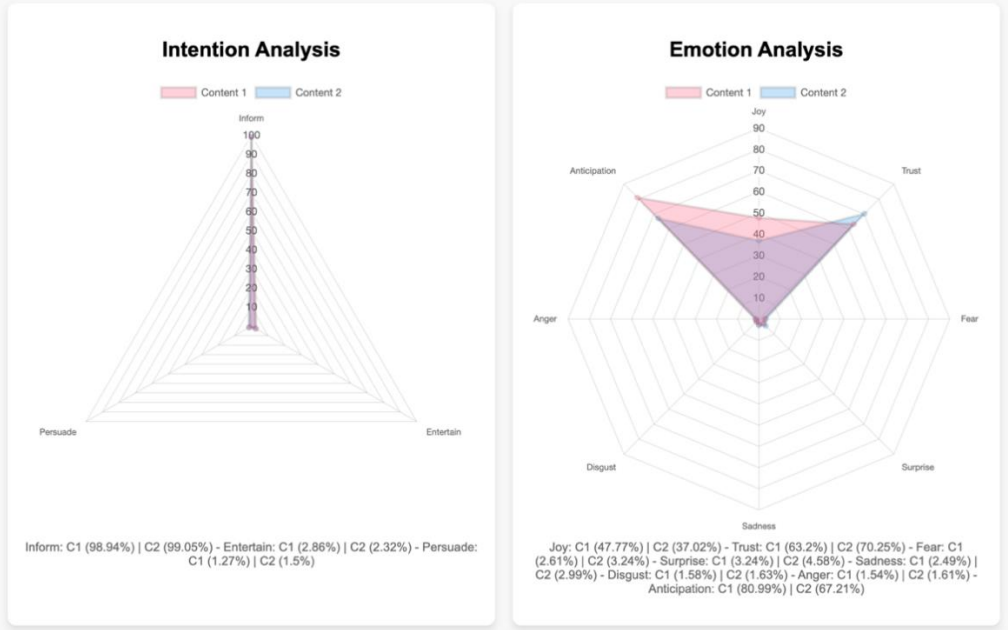
ภาพที่ 4.5 ตัวอย่างที่ 1 ทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีน

จากการทดลองใช้งานระบบต้นแบบในการเปรียบเทียบบทความสองชิ้นที่มีเนื้อหาเกี่ยวข้องกับจีน ดังภาพที่ 4.5 พบว่าทั้งสองบทความมีระดับความคล้ายคลึงกันในหลากหลายมิติ ได้แก่ ด้านเวลา (Timeliness) รูปแบบภาษา (Storytelling vs. Translation) เจตนาในการสื่อสาร (Intention) และอารมณ์ของเนื้อหา (Emotion) โดยจากภาพที่ 4.6 ในมิติ Timeliness พบว่าเนื้อหาทั้งสองบทความมีแนวโน้มไปทาง Timeless อย่างชัดเจน โดย บทความที่1 มีคะแนน Timeless สูงถึง 66.91% และ บทความที่ 2 ที่ 62.83% สอดคล้องกับลักษณะของบทความเชิงวิเคราะห์ที่ไม่ได้เน้นการรายงานสถานการณ์ล่าสุด ส่วนในมิติ Storytelling vs. Translation พบว่า บทความที่1 มีลักษณะของการแปลสูงกว่า (59.83%) ในขณะที่ บทความที่ 2 มีค่าความเป็น Storytelling ที่สูงกว่า (50.52%) ซึ่งอาจสะท้อนความแตกต่างในการใช้สำนวนหรือโครงสร้างภาษา



ภาพที่ 4.6 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่1

ด้านเจตนาในการสื่อสาร ทั้งสองบทความแสดงค่าในหมวด Inform อย่างชัดเจน (C1 = 98.94%, C2 = 99.05%) ขณะที่เจตนาอื่น ๆ อย่าง Entertain และ Persuade มีค่าในระดับต่ำมาก สะท้อนว่าเนื้อหามีลักษณะเชิงสาระความรู้ ไม่เน้นความบันเทิงหรือการชักจูง ส่วนในมิติ Emotion พบว่าอารมณ์เด่นร่วมกันคือ Trust (C1 = 63.2%, C2 = 70.25%) และ Anticipation (C1 = 80.99%, C2 = 67.21%) ซึ่งบ่งชี้ถึงการนำเสนอข้อมูลที่ให้ความรู้สึกมั่นใจและมุ่งคาดการณ์อนาคต ขณะที่อารมณ์เชิงลบหรือดราม่า เช่น Fear, Disgust และ Sadness มีค่าในระดับต่ำมาก ดังภาพที่ 4.7



ภาพที่ 4.7 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่2

เพื่อประเมินระดับความคล้ายคลึงโดยรวมของทั้งสองบทความ ระบบได้คำนวณค่า Overall Similarity ด้วยวิธี Hellinger Similarity ร่วมกับ การถ่วงน้ำหนัก (weighting scheme) ตามมุมมองของผู้เชี่ยวชาญ 4 รูปแบบ ได้แก่ แบบเน้นการแปลภาษา (Translation-focused), แบบเน้นข่าว (News-focused), แบบสมดุล (Balanced) และแบบเน้นความเร่งด่วน (Timeliness-focused) จะพบว่า ค่าความคล้ายคลึงในแต่ละมิติ ที่ได้จากการประมวลผลของโมเดลยังคงมีค่า เท่ากันทุกครั้ง โดยไม่ขึ้นกับน้ำหนักที่เลือก ซึ่งแสดงถึงความถูกต้องและเสถียรของระบบ ดังนี้

1) Timeliness Similarity: 96.98%

2) Storytelling Similarity: 92.64%

3) Intention Similarity: 98.62%

4) Emotion Similarity: 90.82%

เมื่อพิจารณาค่าความคล้ายคลึงโดยรวม (Overall Similarity) ที่ได้จากการถ่วงน้ำหนักตามมุมมองของผู้เชี่ยวชาญในแต่ละรูปแบบ จะได้ผลลัพธ์ดังนี้

1) แบบที่ 1 มุ่งเน้นงานแปล: 92.68%

2) แบบที่ 2 มุ่งเน้นงานข่าว: 95.52%

3) แบบที่ 3 มุ่งเน้นงานทั่วไป: 94.59%

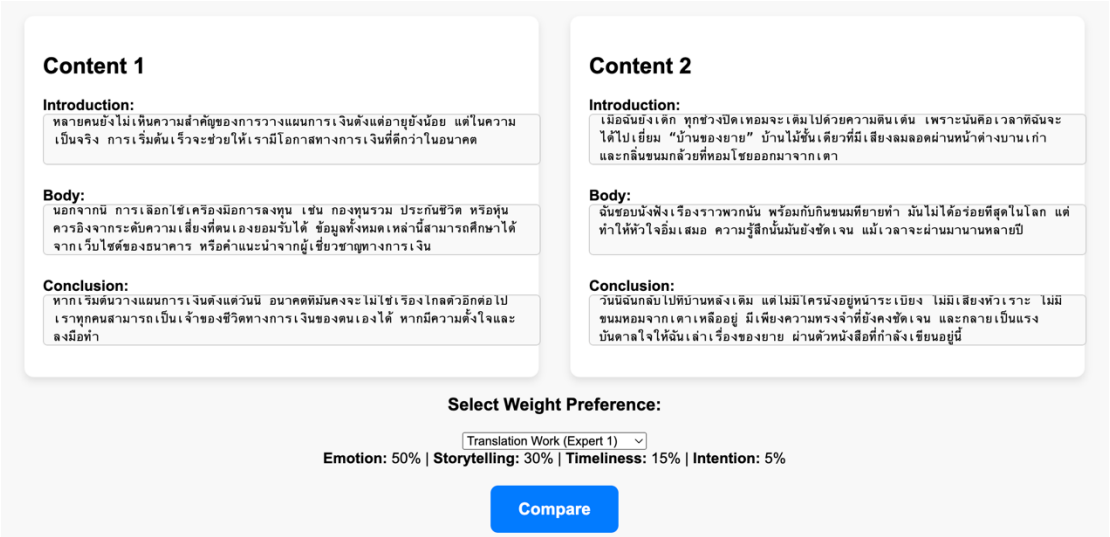
4) แบบที่ 4 มุ่งเน้นงานที่เกี่ยวข้องกับเวลา: 94.43%

ความคล้ายคลึงของสองบทความ ทั้งสี่รูปแบบ มีค่าสูงเกิน 90% ทำให้ค่าความคล้ายโดยรวมทั้งสี่รูปแบบค่าถ่วงน้ำหนัก มีค่าเกิน 90% เช่นกัน

ผลลัพธ์ดังกล่าวสะท้อนให้เห็นว่าแม้ค่าความคล้ายคลึงรายมิติจะคงที่จากการประมวลผลของโมเดล แต่ค่า Overall Similarity สามารถปรับเปลี่ยนได้ตามน้ำหนักที่กำหนดในแต่ละมิติ ซึ่งช่วยให้ผู้ใช้สามารถประเมินความสอดคล้องของเนื้อหาได้หลากหลายมุมมองตามวัตถุประสงค์หรือบริบทการใช้งานจริง เช่น หากให้ความสำคัญกับความเป็นข่าวมากขึ้น (News-focused) ความเหมือนระหว่างบทความจะถูกประเมินว่ามีค่าสูงกว่ารูปแบบอื่น ๆ ซึ่งจะเห็นได้ว่าระบบต้นแบบมีความยืดหยุ่นในการนำเสนอผลลัพธ์และรองรับการประเมินเชิงสไตล์ที่ปรับตามความต้องการของผู้ใช้หรือผู้เชี่ยวชาญในแต่ละด้านได้อย่างมีประสิทธิภาพ

4.7.2 ตัวอย่างที่ 2 ทดสอบระบบต้นแบบด้วยบทความที่ไม่เกี่ยวข้องกับจีน

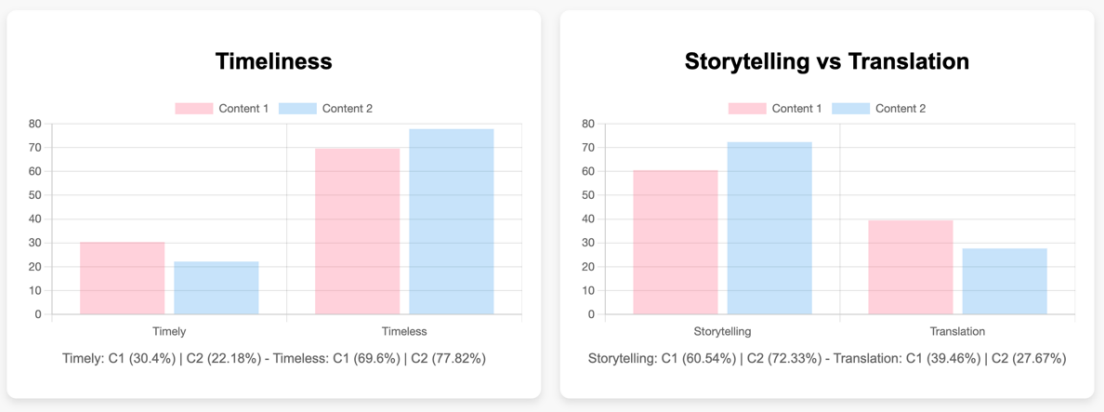
นอกจากการทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีนแล้ว การทดสอบด้วยบทความที่ไม่เกี่ยวข้องกับจีนได้ถูกจัดทำขึ้น ในฐานะของเนื้อหาที่ไม่เคยพบมาก่อน เพื่อทดสอบประสิทธิภาพของระบบ ดังภาพที่ 4.8



ภาพที่ 4.8 ตัวอย่างที่ 2 ทดสอบระบบต้นแบบด้วยบทความที่เกี่ยวข้องกับจีน

จากผลทดสอบดังปรากฏในภาพที่ 4.9 พบว่า ด้าน Timeliness ข้อมูลจากโมเดลแสดงให้เห็นว่าเนื้อหาทั้งสองบทความมีลักษณะโน้มเอียงไปทาง Timeless อย่างชัดเจน โดย บทความที่1 มีสัดส่วน Timeless อยู่ที่ 69.6% และ บทความที่ 2 สูงถึง 77.82% ซึ่งสอดคล้องกับลักษณะของข้อความที่เน้นการบรรยายหรือถ่ายทอดความคิดในเชิงแนวคิด มากกว่าการอ้างอิงสถานการณ์ปัจจุบัน ในขณะที่สัดส่วนของ Timely ในทั้งสองบทความอยู่ต่ำกว่า 31% แสดงให้เห็นว่าเนื้อหาทั้งสองไม่เน้นการรายงานเชิงทันเหตุการณ์

เมื่อพิจารณาลักษณะของภาษาในมิติ Storytelling vs. Translation พบว่า บทความที่ 2 มีความเป็น Storytelling สูงถึง 72.33% ขณะที่ บทความที่1 อยู่ที่ 60.54% แม้จะถือว่าเป็นสัดส่วนที่สูงเช่นกัน แต่ยังคงมีองค์ประกอบของภาษาที่มีลักษณะเป็นทางการหรือคล้ายภาษาที่ถูกดัดแปลงจากแหล่งอื่นมากกว่า บทความที่ 2 ที่มีโทนของการเขียนแบบ narrative ชัดเจน ในขณะที่สัดส่วน Translation ของ บทความที่ 2 ต่ำกว่า บทความที่1 อย่างชัดเจน (27.67% เทียบกับ 39.46%) ซึ่งสอดคล้องกับการใช้ภาษาที่ลื่นไหลและมีลักษณะเล่าเรื่องในบทความชุดที่สอง



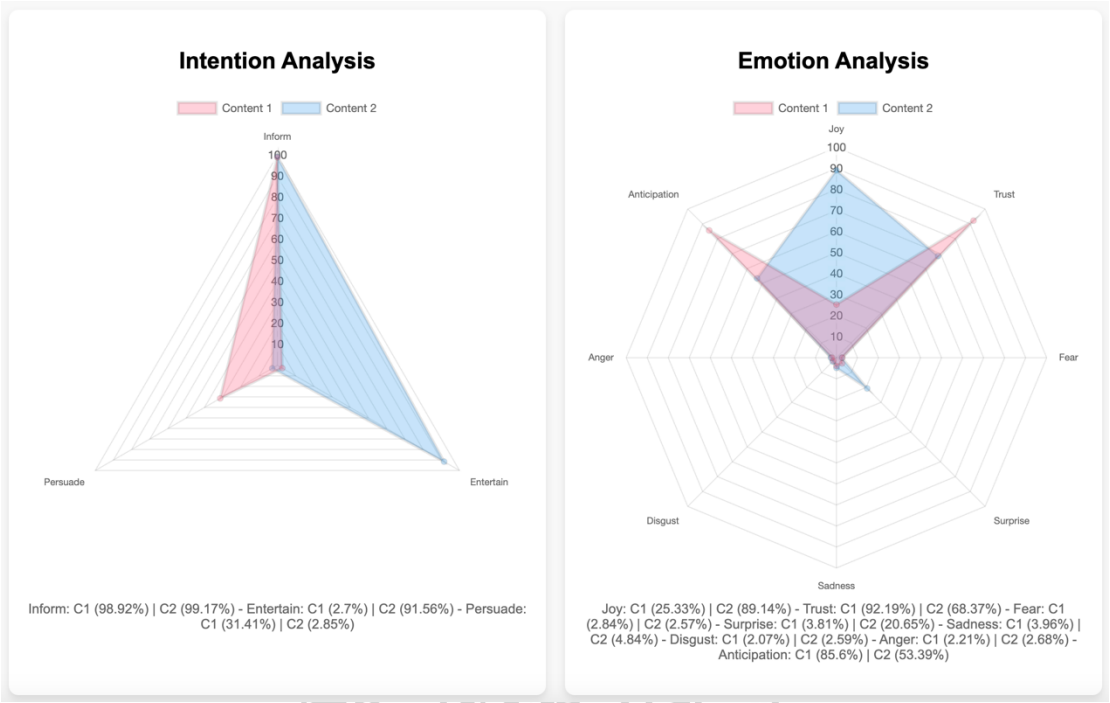
ภาพที่ 4.9 แสดงผลการทดสอบบทความตัวอย่างที่2 ส่วนที่1

ด้านเจตนาการสื่อสาร ดังปรากฏในภาพที่ 4.10 พบว่า โมเดลจำแนกได้อย่างชัดเจนว่า บทความทั้งสองอยู่ภายใต้กลุ่มเจตนา Inform โดย บทความที่1 มีสัดส่วน 98.92% และ บทความที่ 2 อยู่ที่ 99.17%

อย่างไรก็ตาม ความแตกต่างที่น่าสนใจปรากฏในเจตนา Entertain ซึ่ง บทความที่ 2 มีสัดส่วนสูงถึง 91.56% ขณะที่ บทความที่1 อยู่ที่เพียง 2.7% ซึ่งบ่งชี้ว่าบทความชุดที่สองมีโทนการนำเสนอที่เอื้อต่อการสร้างความเพลิดเพลินหรือความรู้สึกเชิงบันเทิงมากกว่า โดยบทความที่1 เน้นเนื้อหาสาระหรือข้อเท็จจริง ในขณะเดียวกัน เจตนา Persuade ปรากฏชัดใน บทความที่1 (31.41%) มากกว่า บทความที่ 2 (2.85%) สะท้อนถึงความพยายามโน้มน้าวหรือสร้างอิทธิพลต่อผู้อ่านในบทความแรก

ด้าน Emotion ระบบแสดงให้เห็นถึงอารมณ์เด่นที่แตกต่างกันอย่างชัดเจนระหว่างสองบทความ โดย บทความที่ 2 มีคะแนน Joy สูงมากถึง 89.14% ขณะที่ บทความที่1 อยู่ที่ 25.33% ซึ่งสะท้อนความรู้สึกเบิกบานหรือพึงพอใจที่ชัดเจนในเนื้อหาของ บทความที่ 2 ในขณะเดียวกัน อารมณ์ Trust ก็อยู่ในระดับสูงในทั้งสองบทความ โดย บทความที่1 อยู่ที่ 92.19% และ บทความที่ 2 ที่ 68.37%

นอกจากนี้ บทความที่ 2 ยังแสดงอารมณ์ Surprise และ Sadness สูงกว่า บทความที่1 อย่างมีนัยสำคัญ โดยเฉพาะ Surprise ซึ่งมีค่า 20.65% เมื่อเทียบกับ 3.81% ใน บทความที่1



ภาพที่ 4.10 แสดงผลการทดสอบความตัวอย่างที่1 ส่วนที่2

เมื่อประเมินระดับความคล้ายคลึงระหว่างบทความทั้งสองด้วยวิธี Hellinger Similarity โดยใช้รูปแบบการถ่วงน้ำหนักตามมุมมองของผู้เชี่ยวชาญ จะได้ค่าความคล้ายคลึงในแต่ละมิติดังนี้

1) Timeliness Similarity: 93.38%

2) Storytelling Similarity: 91.15%

3) Intention Similarity: 37.49%

4) Emotion Similarity: 60.11%

ค่าดังกล่าวเป็นค่าคงที่ที่ได้จากการประมวลผลของโมเดลในแต่ละมิติโดยตรง และไม่เปลี่ยนแปลงแม้จะมีการเปลี่ยนแปลงรูปแบบการถ่วงน้ำหนัก อย่างไรก็ตาม ค่า Overall Similarity ซึ่งเป็นผลลัพธ์สุดท้ายของการวิเคราะห์ความคล้ายคลึงเชิงภาพรวม จะเปลี่ยนไปตามสัดส่วนน้ำหนักที่กำหนดในแต่ละมิติ โดยผลลัพธ์ที่ได้จากทั้ง 4 รูปแบบมีดังนี้:

1) แบบที่ 1 มุ่งเน้นงานแปล: 73.28%

2) แบบที่ 2 มุ่งเน้นงานข่าว: 55.23%

3) แบบที่ 3 มุ่งเน้นงานทั่วไป: 71.78%

4) แบบที่ 4 มุ่งเน้นงานที่เกี่ยวข้องกับเวลา: 77.37%

ความแตกต่างอย่างชัดเจนของค่า Overall Similarity ตามแต่ละค่าถ่วงน้ำหนัก ซึ่งระบุถึงจุดประสงค์ของการให้ความสำคัญแต่ละรูปแบบ สอดคล้องกับค่า Similarity ของแต่ละส่วนโดยตรง ยกตัวอย่างเช่น แบบที่1 ให้ความสำคัญกับการจำแนกอารมณ์ 50% ดังนั้น เมื่อเนื้อหาทั้งสองบทความมีค่า Emotion Similarity ราว 60% จึงทำให้มีค่าความคล้ายโดยรวมราว 73% แม้ว่า Timeliness และ Storytelling Similarity จะเกิน 90% ขณะที่แบบที่ 2 ให้ความสำคัญกับ Intention มากที่สุด 50% รองลงมาคือ Emotion 30% ซึ่งสองรูปแบบนี้ มีความคล้ายคลึงต่ำกว่า 70% โดยเฉพาะ Intention ที่มีความคล้ายเพียงราว 37% ทำให้ค่าความคล้ายรวมของแบบที่ 2 อยู่ที่ราว 55% เท่านั้น ส่วนแบบที่ 4 ที่ให้ความสำคัญกับ Timeliness มากที่สุด มีความคล้ายโดยรวมมากที่สุด เนื่องจาก Timeliness Similarity ของทั้งสองบทความสูงที่สุด ที่ราว 93%

ผลลัพธ์ดังกล่าวแสดงให้เห็นว่า แม้บทความความทั้งสองจะไม่ได้มีที่มาจากแหล่งเดียวกัน ไม่ได้เกี่ยวกับจีน และเขียนขึ้นใหม่ทั้งหมด แต่ระบบยังสามารถวัดความคล้ายคลึงเชิงสไตล์และเนื้อหาได้อย่างมีเสถียรภาพ โดยเฉพาะในรูปแบบถ่วงน้ำหนักแบบที่ 1 และ 4 ซึ่งให้น้ำหนักกับมิติด้านอารมณ์และเวลา พบว่าความคล้ายคลึงโดยรวมยังคงอยู่ในระดับสูงเกิน 89%

ทั้งสองตัวอย่าง สะท้อนถึงความสามารถของระบบในการตรวจจับความสอดคล้องของบทความ ทั้งที่เกี่ยวข้องกับจีนและไม่เกี่ยวข้องได้อย่างชัดเจน

4.7.3 ทดสอบระบบต้นแบบด้วยชุดข้อมูลเดิมเทียบกับชุด Balanced ที่จำนวนป้ายกำกับเท่ากันทั้งหมดแบบ Dataset A

ในการทดลองนี้ ได้ทดสอบระบบต้นแบบจำแนกเจตนาของข้อความ ด้วยการเปรียบเทียบระหว่างชุดข้อมูลเดิม และชุดข้อมูลที่มีการปรับให้ป้ายกำกับของแต่ละคลาสเท่ากัน โดยเลือกใช้โมเดลที่มีประสิทธิภาพดีที่สุดจากผลการทดลองในหัวข้อ 4.3.7 ผลการทดลองย่อยที่ 7 ซึ่งใช้ Dataset A และเพิ่มฟีเจอร์อารมณ์ มาดำเนินการทดสอบ โดยภาพตัวอย่างทั้งสามชุดแสดงให้เห็นอย่างชัดเจนว่า เมื่อใช้ชุด Balanced ที่ป้ายกำกับแต่ละคลาสเท่ากัน ระบบสามารถทำนายคลาส Entertain และ Persuade ได้มากขึ้นอย่างมีนัยสำคัญ ต่างจากชุดข้อมูลเดิมที่ระบบมักทำนายเป็นคลาส Inform เกือบทั้งหมด ซึ่งเกิดจากความไม่สมดุลของข้อมูลที่มีคลาส Inform เป็นสัดส่วนหลัก

อย่างไรก็ตามเมื่อพิจารณาผลการจำแนกเจตนาในบทความข่าวจริงของแต่ละชุดพบข้อผิดพลาดสำคัญในชุด Balanced Dataset A โดยเฉพาะเมื่อระบบต้องประเมินบทความที่เป็น “ข่าวรายงาน” หรือ “ข่าวทั่วไป” ซึ่งตามเนื้อหามีลักษณะเป็น Inform ชัดเจนและไม่ได้มีลักษณะ

บันเทิงหรือโน้มน้าวใจอย่างเด่นชัด แต่ระบบกลับทำนายผิดเป็นคลาส Entertain หรือ Persuade แทน

ข้อผิดพลาดนี้สะท้อนผลกระทบจากการตัดบทความที่มีเพียงคลาส Inform เดี่ยวออกไปจากชุด Balanced ทำให้โมเดลขาดความสามารถในการแยกแยะลักษณะของข้อความที่เน้นให้ข้อมูลแบบรายงาน เช่น ข่าวการประชุมสภาของจีน หรือข่าวเศรษฐกิจที่มีลักษณะการนำเสนอเชิงข้อเท็จจริง

แม้การใช้ชุด Balanced Dataset A จะช่วยเพิ่มความสามารถของระบบในการทำนายคลาสที่มีสัดส่วนน้อยอย่าง Entertain และ Persuade แต่ก็ทำให้ความแม่นยำในการประเมินคลาส Inform ลดลงอย่างมีนัยสำคัญ โดยเฉพาะในบริบทที่บทความมีลักษณะสื่อสารข้อมูลเชิงสาระโดยไม่มีอารมณ์หรือการโน้มน้าว ทำให้เกิดผลลัพธ์คลาดเคลื่อนจากความจริงในเชิงเนื้อหา

4.7.4 ทดสอบระบบต้นแบบด้วยชุดข้อมูลเดิมเทียบกับชุด Balanced Dataset ที่จำนวนป้ายกำกับรวมใกล้เคียงกัน

จากการนำชุดข้อมูลแบบ Balanced ที่มีสัดส่วนป้ายกำกับใกล้เคียงกันในอัตรา 80:20 มาใช้กับระบบต้นแบบ แล้วเปรียบเทียบกับชุดข้อมูลเดิม พบความเปลี่ยนแปลงในเชิงคุณภาพของการจำแนกเจตนา โดยเฉพาะในกรณีที่บทความมีเจตนาซ้อนทับกันหลายมิติ

ในภาพตัวอย่างแรก บทความทั้งสองชิ้นเป็นข่าวการประชุมสภาที่ปรึกษาการเมืองแห่งชาติของจีน ซึ่งมีลักษณะการเขียนแบบข่าวทางการ แจ้งข้อมูลชัดเจนและมีน้ำเสียงแบบเป็นทางการ จึงจัดอยู่ในคลาส Inform อย่างชัดเจน อย่างไรก็ตาม เมื่อลองวิเคราะห์เชิงเนื้อหา พบว่าบทความมีการแฝงจุดประสงค์เชิงโน้มน้าว (Persuade) อยู่บางช่วง เช่น การเน้นย้ำความสำคัญของการประชุม หรือการอ้างอิงความสำเร็จของระบบการเมืองจีน ดังนั้น ระบบต้นแบบที่ใช้ชุด Balanced 80:20 จึงสามารถตรวจจับคลาส Persuade ได้แม่นยำขึ้น แต่ยังคงอยู่ในระดับที่เหมาะสม (ไม่เกิน 20%) โดยไม่ทำนายเกินจริงอย่างที่เกิดขึ้นกับชุด Balanced แบบป้ายกำกับเท่ากันทุกคลาส

ในภาพถัดมา ผลลัพธ์จากการเปรียบเทียบสองบทความโดยใช้ระบบต้นแบบกับชุดข้อมูลดังกล่าว พบว่า บทความที่ 1 เป็นข่าวประชาสัมพันธ์เกี่ยวกับกิจกรรมอบรมพิเศษ ซึ่งมีเป้าหมายชัดเจนในการส่งเสริมภาพลักษณ์และสร้างการรับรู้ต่อโครงการ เนื้อหามีลักษณะเป็นการโน้มน้าวอย่างชัดเจน แต่ในชุดข้อมูลเดิม ระบบมักทำนายเป็น Inform เนื่องจากเนื้อหามีลักษณะข่าว แต่ไม่ได้พิจารณาจากบริบทของจุดประสงค์ ดังนั้นเมื่อใช้ชุดข้อมูล Balanced แบบ 80:20 ระบบจึงสามารถจำแนกเจตนาเป็น “Persuade” ได้แม่นยำขึ้นตามบริบทจริงของบทความ ส่วนบทความที่ 2 เป็นตัวอย่างของโพสต์กึ่งข่าวสารที่มีโทนการเล่าเรื่องแบบให้ข้อมูล (Inform) แต่มีลักษณะความบันเทิง

แฝงอยู่ในบางท่อน ซึ่งระบบสามารถตรวจจับคลาส “Entertain” ได้ในระดับที่เหมาะสม โดยยังคงรักษาความแม่นยำในคลาสหลักคือ Inform ได้ดี และไม่ได้ทำนายคลาส Persuade เกินจริง

ระบบต้นแบบที่ใช้ชุดข้อมูล Balanced แบบจำนวนใกล้เคียงกันของแต่ละคลาส และแบ่งสัดส่วนใช้ Train และ Test แบบ 80:20 แสดงให้เห็นถึงความสามารถในการวิเคราะห์เจตนาเชิงบริบทได้แม่นยำขึ้น โดยเฉพาะในกรณีที่ข้อความมีเจตนาแฝงมากกว่าหนึ่งแบบ และไม่สามารถตัดสินได้จากคำหรือโครงสร้างภาษาตรงๆ เพียงอย่างเดียว ช่วยลดความเอนเอียงจากชุดข้อมูลเดิมที่มักประเมิน Inform เป็นหลัก และเพิ่มความแม่นยำในการตรวจจับคลาสรองอย่างมีเหตุผลสอดคล้องกับเนื้อหาจริง

4.8 ประเมินความพึงพอใจและความเป็นไปได้ในการใช้งานจริง

4.8.1 ประเมินจากผู้เชี่ยวชาญ

จากผลการประเมินของผู้เชี่ยวชาญและผู้ที่เกี่ยวข้องกับชุดข้อมูล ตามระบุในรายละเอียดหัวข้อ 3.4 พบว่าคะแนนเฉลี่ยจากผู้เชี่ยวชาญจำนวน 4 ท่าน สำหรับทั้ง 2 ด้าน ได้แก่ ความพึงพอใจในการใช้งาน และความเป็นไปได้ในการนำไปใช้งานจริง อยู่ที่ 4.0 คะแนน จากคะแนนเต็ม 5 คะแนน โดยผู้เชี่ยวชาญแต่ละท่านให้เหตุผลสนับสนุนที่แตกต่างกันตามลักษณะของงานและวัตถุประสงค์เฉพาะของตนเอง ซึ่งปรากฏรายละเอียดไว้ในบทวิเคราะห์รายบุคคล

นอกจากนี้ ผู้เชี่ยวชาญแต่ละท่านยังได้สะท้อนแนวทางการใช้งานจริงผ่านการเลือกค่าถ่วงน้ำหนักที่ต่างกันในระบบวิเคราะห์ ได้แก่ รูปแบบที่ 1 ถึงรูปแบบที่ 4 ซึ่งสะท้อนให้เห็นถึงความยืดหยุ่นของระบบในการตอบสนองต่อความต้องการที่หลากหลายของผู้ใช้ และเป็นอีกหนึ่งตัวชี้วัดที่สะท้อนความพร้อมในการนำไปใช้ประโยชน์จริงในหลากหลายบริบท

4.8.2 ผลประเมินจากผู้ตอบแบบสอบถาม

กลุ่มที่มี ประสบการณ์มากกว่า 5 ปี ให้คะแนนเฉลี่ย สูงที่สุดทุกด้าน (เต็ม 5.0 เกือบทุกหัวข้อ) ซึ่งสะท้อนว่าผู้ที่มีความเข้าใจเชิงลึกในงาน Content Marketing และมีประสบการณ์ตรงกับการวิเคราะห์เนื้อหา สามารถประเมินศักยภาพของระบบได้ชัดเจน และเห็นประโยชน์ในการประยุกต์ใช้งานจริงมากกว่า กลุ่มที่มี ประสบการณ์ 3–5 ปี และ 1–3 ปี ให้คะแนนอยู่ในระดับปานกลางถึงสูง (เฉลี่ยประมาณ 4.0–4.3) โดยมีความพึงพอใจต่อรูปแบบผลลัพธ์และความชัดเจน แต่ยังไม่ถึงระดับที่โดดเด่น ขณะที่กลุ่ม ไม่เกิน 1 ปี แม้จะให้คะแนนในด้านความง่ายในการใช้งานและผลลัพธ์อยู่ในระดับดี (4.0–4.25) แต่ให้คะแนนเรื่องความสอดคล้องกับลักษณะงาน และ การประยุกต์ใช้จริงต่ำกว่ากลุ่มอื่น สะท้อนว่าอาจยังขาดประสบการณ์ในการจับคู่ระบบกับบริบทการใช้งานจริง

ขณะที่กลุ่มที่ทำ งานบริหารทีมเขียน / เจ้าของสื่อ ให้คะแนน สูงที่สุด ทุกด้าน (5.0 เต็ม) เช่นเดียวกับกลุ่มมากกว่า 5 ปี แสดงให้เห็นว่าผู้ที่มองภาพรวมของกระบวนการผลิตเนื้อหาและการใช้เครื่องมือวิเคราะห์ มองเห็นคุณค่าของระบบได้อย่างครอบคลุม กลุ่ม เขียนเพื่องานโฆษณา/เพื่องานขาย และ นักเขียนบทความ ให้คะแนนในระดับดีทุกด้าน (เฉลี่ยประมาณ 4.0–4.3) โดยเฉพาะในด้านความสอดคล้องกับงาน และผลลัพธ์ชัดเจน ขณะที่กลุ่ม นักแปล และ นักการตลาด แม้จะมีความเข้าใจในการนำเสนอเนื้อหาเชิงกลยุทธ์ แต่ให้คะแนนอยู่ที่ระดับ 3.6–4.2 ซึ่งอาจเป็นผลมาจากความคาดหวังที่สูงหรือความท้าทายของงานที่เน้นความแม่นยำเชิงภาษา ส่วนกลุ่ม Blogger / Creator ให้คะแนนเฉลี่ยปานกลาง โดยเฉพาะด้านความสอดคล้องกับงาน ที่ยังต่ำกว่ากลุ่มอื่น (3.67) ซึ่งอาจสะท้อนถึงช่องว่างระหว่างผลลัพธ์จากระบบกับเนื้อหาเชิงสร้างสรรค์แบบเฉพาะบุคคล

สำหรับกลุ่มที่ดูแลเพจแบบ Influencer/Creator Media หรือเพจส่วนตัว ให้คะแนนเฉลี่ยค่อนข้างสูงเกือบทุกด้าน (4.0–4.67) โดยเฉพาะความพึงพอใจ และความเป็นไปได้ในการนำไปใช้งานจริง สะท้อนว่าระบบนี้สามารถตอบโจทย์งาน Content บนแพลตฟอร์มโซเชียลได้ดีพอสมควร กลุ่มที่ดูแล เว็บไซต์ / สื่อออนไลน์ทั่วไป ให้คะแนนในระดับปานกลาง (เฉลี่ย 3.75–4.25) ซึ่งอาจสะท้อนถึงความต้องการการวิเคราะห์เชิงเทคนิคหรือความแม่นยำมากขึ้น ส่วนกลุ่ม สื่อขององค์กรหรือบริษัท มีความพึงพอใจค่อนข้างสูง โดยเฉพาะรูปแบบการแสดงผล (4.67) และ ความชัดเจนของผลลัพธ์ (4.33) ซึ่งอาจเนื่องจากเนื้อหาประเภทองค์กรมีโครงสร้างที่สอดคล้องกับระบบวิเคราะห์ได้ง่าย

มหาวิทยาลัยศิลปากร

บทที่ 5
สรุปผลการดำเนินงานวิจัยและข้อเสนอแนะ

วิทยานิพนธ์นี้ดำเนินตามวัตถุประสงค์ในการพัฒนาและนำเสนอแนวทางในการสร้างเครื่องมือประเมินความสอดคล้องของบทความที่สอดรับกับหลักการตลาดด้วยบทความ โดยมุ่งเน้นการศึกษาและเลือกใช้เนื้อหาที่เกี่ยวข้องกับประเทศจีนในรูปแบบภาษาไทย เนื่องจากมีลักษณะเฉพาะที่สามารถวิเคราะห์ได้ชัดเจนและครอบคลุม 4 ด้านหลัก ได้แก่ 1) ความทันต่อเวลา (Timeliness) 2) จุดประสงค์ในการสื่อความ (Intention) 3) การสื่อสารอารมณ์ผ่านเนื้อหา (Emotion) และ 4) รูปแบบการเขียนที่แบ่งเป็นงานแปลและการเล่าเรื่อง (Storytelling vs. Translation)

จากการทดลองและเปรียบเทียบผลการทำงานของโมเดลต่างๆ ที่พัฒนา พบว่า สำหรับการทดลองที่ 1 โมเดล WangchanBERTa มีประสิทธิภาพสูงสุดในการจำแนกเนื้อหาตามกระแสหรือเนื้อหาสดใหม่ (Timely–Timeless) โดยสามารถบรรลุผลความแม่นยำในการทดสอบมากที่สุดเมื่อเปรียบเทียบกับโมเดลอื่นๆ เช่น CNN ซึ่งแม้จะมีความแม่นยำที่สูงแต่ก็ยังด้อยกว่า WangchanBERTa เล็กน้อย โมเดล WangchanBERTa ใช้ tokenizer ที่ผ่านการฝึกอบรมด้วยชุดข้อมูลภาษาไทยขนาดใหญ่ถึง 15 ล้านประโยค ทำให้สามารถจัดการกับเนื้อหาภาษาไทยที่ซับซ้อนและยาวได้อย่างมีประสิทธิภาพ

สำหรับการทดลองที่ 2 การจำแนกจุดประสงค์ของเนื้อหา (Intention Classification) ตามกรอบ PIE ได้แก่ Persuade, Inform และ Entertain พบว่า PhayaThaiBERT เป็นโมเดลที่ให้ผลลัพธ์ดีที่สุด โดยมีค่า Micro F1-score เท่ากับ 88.74% Macro F1-score เท่ากับ 84.14% และ Hamming Loss ต่ำที่สุดที่ 12.14% อย่างไรก็ตาม พบสัญญาณ Overfitting เล็กน้อย แต่ยังคงทำงานได้ดีเมื่อนำไปทดสอบกับข้อมูลใหม่ นอกจากนี้ ยังพบว่า การเพิ่มฟีเจอร์ด้านอารมณ์ (Emotion Features) ไม่ได้ช่วยเพิ่มประสิทธิภาพของ PhayaThaiBERT อย่างชัดเจน แต่กลับ ช่วยเพิ่มประสิทธิภาพของ WangchanBERTa ได้อย่างมีนัยสำคัญ ซึ่งสะท้อนถึงบทบาทของบริบททางอารมณ์ในบางกรณีของการจำแนกเจตนาเนื้อหา

การทดลองที่ 3 ซึ่งมุ่งแก้ไขปัญหาข้อมูลไม่สมดุล สำหรับการจำแนกจุดประสงค์ของเนื้อหา ยังแสดงให้เห็นอย่างชัดเจนว่า ขนาดของชุดข้อมูลฝึกมีอิทธิพลโดยตรงต่อประสิทธิภาพของโมเดล โดยเฉพาะเมื่อเปรียบเทียบสัดส่วน train:test ที่ 70:30 กับ 80:20 สำหรับชุดข้อมูลที่มีจำนวนป้าย

กำกับใกล้เคียงกัน พบว่าเมื่อใช้ข้อมูลฝึกในสัดส่วน 80% โมเดลทุกตัวมีผลการทำงานดีขึ้นในทุกตัวชี้วัด ไม่ว่าจะเป็น ค่า F1-score ค่า Exact Match Accuracy และ ค่า Hamming Loss โดยเฉพาะโมเดลขนาดใหญ่ เช่น PhayaThaiBERT และ WangchanBERTa ซึ่งต้องอาศัยข้อมูลจำนวนมากเพื่อเรียนรู้ได้อย่างมีประสิทธิภาพ โดยผลลัพธ์ชี้ให้เห็นว่า PhayaThaiBERT แสดงอาการ overfitting อย่างเด่นชัดในหลายกรณี โดยเฉพาะเมื่อใช้ข้อมูลที่ไม่สมดุลหรือเพิ่มฟีเจอร์อารมณ์ ในทางกลับกัน WangchanBERTa โดยส่วนใหญ่มีประสิทธิภาพที่ดีขึ้นเมื่อใช้ฟีเจอร์อารมณ์ ส่วน mBERT แม้มีประสิทธิภาพต่ำที่สุด แต่แสดงให้เห็นว่า มีประสิทธิภาพที่ดีขึ้น เมื่อเพิ่มปริมาณข้อมูลฝึก ซึ่งทั้งหมดนี้ยืนยันว่า การออกแบบระบบควรพิจารณาให้สมดุลระหว่างขนาดของข้อมูล ความซับซ้อนของโมเดล และการใช้ฟีเจอร์เสริม เพื่อให้ได้ผลลัพธ์ที่แม่นยำและเสถียรที่สุดในการใช้งานจริง

ขณะที่การทดลองที่ 4 การจำแนกอารมณ์ (Emotion Classification) พบว่า WangchanBERTa มีศักยภาพในการวิเคราะห์เนื้อหาที่แสดงออกถึงอารมณ์ในภาษาไทยทั้ง 8 กลุ่มได้อย่างแม่นยำที่สุด โดยให้ค่า Micro F1-score เท่ากับ 78.35% Macro F1-score เท่ากับ 55.74% และ Hamming Loss ต่ำที่สุดที่ 10.27% แสดงถึงความสามารถของโมเดลในการรับมือกับข้อมูลที่ไม่สมดุลในแต่ละอารมณ์ได้อย่างมีประสิทธิภาพ

ด้านการทดลองที่ 5 การจำแนกรูปแบบงานแปลและการเล่าเรื่อง (Storytelling vs. Translation) โมเดล Random Forest ซึ่งใช้เทคนิค Ensemble Model และมีการทำ Feature Engineering อย่างเหมาะสมด้วยฟีเจอร์ที่เกี่ยวข้อง ได้แก่ Named Entity Recognition (NER), Sentiment Analysis, Part of Speech Tagging (POS), Similarity Features, Length Features และ TF-IDF สามารถเพิ่มความแม่นยำจากเดิมที่ 85% เป็น 92% หลังใช้ Features อย่างมีนัยสำคัญ แสดงให้เห็นถึงความสำคัญของการคัดเลือกและพัฒนาฟีเจอร์ที่เหมาะสมสำหรับการจำแนกประเภทเนื้อหาอย่างชัดเจน

ในการพัฒนาระบบวิเคราะห์ความสอดคล้องของบทความ ทางวิทยานิพนธ์นี้ได้พัฒนาบนพื้นฐานของชุดข้อมูลและความรู้จากผู้เชี่ยวชาญและผู้เกี่ยวข้องจริงในการจัดทำสื่อเนื้อหาภาษาไทยที่เกี่ยวข้องกับภาษาจีนบนโซเชียลมีเดีย โดยเฉพาะแพลตฟอร์มFacebookซึ่งเป็นช่องทางหลักที่พบว่าใช้ในการนำเสนอเนื้อหาแบบข้อความมากที่สุด โดยผู้เชี่ยวชาญและผู้มีส่วนร่วมในสื่อดังกล่าวได้มีส่วนร่วมตั้งแต่การให้ข้อมูลก่อนการพัฒนา และการประเมินการใช้งานจริงของระบบที่พัฒนาขึ้น ซึ่งผลการประเมินพบว่าระบบมีความเหมาะสมในเชิงการใช้งาน (Usability) และสามารถนำไปใช้งานได้จริงด้วยคะแนนเฉลี่ย 4 คะแนนจากคะแนนเต็ม 5 ทั้งในแง่ของการออกแบบ UI/UX และการประยุกต์ใช้

งานจริง โดยได้รับข้อเสนอแนะจากผู้เชี่ยวชาญว่า เครื่องมือนี้สามารถนำไปใช้ได้ตามลักษณะเฉพาะของแต่ละประเภทงานเขียน ได้แก่

1) งานเขียนทั่วไป

2) งานเขียนที่เน้นการแปล เนื่องจากเนื้อหาส่วนใหญ่เป็นภาษาจีน

3) งานเขียนที่เกี่ยวข้องกับข่าวสาร

4) งานเขียนที่ต้องการความทันต่อเหตุการณ์หรือการจับกระแสเรียลไทม์ ตัวอย่างเช่น เนื้อหาด้านประวัติศาสตร์ องค์ความรู้เฉพาะเรื่องเกี่ยวกับประเทศจีน หรือการเล่าประสบการณ์เกี่ยวกับการท่องเที่ยว

ในส่วนของผลการประเมินจากผู้ใช้งานจริงผ่านแบบสอบถาม พบว่า กลุ่มผู้ที่มีประสบการณ์มากกว่า 5 ปี รวมถึงกลุ่มที่ทำหน้าที่บริหารหรือเจ้าของสื่อ ให้คะแนนเฉลี่ยสูงสุดในทุกด้าน แสดงให้เห็นถึงการมองเห็นคุณค่าของระบบในภาพรวมอย่างเด่นชัด ขณะที่กลุ่มผู้ใช้งานที่มีประสบการณ์น้อยกว่า 1 ปี หรือผู้ที่ทำงานสร้างสรรค์เชิงเฉพาะตัว เช่น Blogger หรือ Creator ให้คะแนนในระดับกลาง โดยเฉพาะในด้านความสอดคล้องกับลักษณะงานที่ต่ำกว่ากลุ่มอื่น ซึ่งอาจสะท้อนถึงช่องว่างระหว่างการวิเคราะห์เชิงเทคนิคของระบบกับความต้องการเนื้อหาเชิงสร้างสรรค์เฉพาะบุคคล อย่างไรก็ตาม กลุ่มที่ดูแล Content บนโซเชียลมีเดีย รวมถึง Influencer และทีมดูแลสื่อขององค์กร ให้คะแนนเฉลี่ยอยู่ในระดับค่อนข้างสูงทั้งในด้านความพึงพอใจ และ ความเป็นไปได้ในการใช้งานจริง

สำหรับประโยชน์ที่ได้รับ สามารถสรุปได้ดังนี้

1) นำเสนอเฟรมเวิร์กใหม่สำหรับประเมินความสอดคล้องของบทความภาษาไทยกับเนื้อหาที่เกี่ยวข้องกับจีนโดยเฉพาะ

2) พัฒนาต้นแบบการวิเคราะห์การสื่อสารของผู้จัดทำเนื้อหา โดยประยุกต์ใช้เฟรมเวิร์ก PIE ร่วมกับฟีเจอร์อารมณ์ เพื่อประเมินเจตนาในการสื่อสารได้อย่างเป็นรูปธรรม และใช้เป็นแนวทางในการวิเคราะห์การสื่อความของเนื้อหา

3) ผลการวิจัยแสดงให้เห็นแนวทางในการพัฒนาการวิเคราะห์เนื้อหาข้อความที่อยู่ในโดเมนเฉพาะและมีข้อจำกัดด้านขนาดของชุดข้อมูล โดยเฉพาะในประเด็นการแก้ไขปัญหาความไม่สมดุลของข้อมูล

4) พัฒนาแนวทางการวิเคราะห์เนื้อหาบทความภาษาไทย โดยใช้โมเดลการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก ซึ่งสามารถติดป้ายกำกับเนื้อหาในระดับบทความ โดยพิจารณาบริบทภาพรวมไม่จำกัดเพียงคำหรือโครงสร้างประโยค

ทั้งนี้ การพัฒนางานวิจัยในอนาคตนั้น ควรพิจารณาเพิ่มขนาดและความหลากหลายของชุดข้อมูลที่ใช้ในการฝึกโมเดล รวมถึงการสร้าง corpus ที่มีลักษณะเฉพาะเจาะจงมากขึ้นสำหรับการตลาดด้วยเนื้อหา เพื่อเพิ่มประสิทธิภาพในการจำแนกเนื้อหาตามจุดประสงค์ต่างๆ นอกจากนี้ยังสามารถขยายการศึกษาไปสู่ปัจจัยอื่นเพิ่มเติม เช่น การวิเคราะห์ช่วงเวลาการนำเสนอเนื้อหา หรือการแยกป้ายกำกับในแต่ละส่วนของบทความ (บทนำ, เนื้อหา, บทสรุป) เพื่อให้ได้ข้อมูลที่มีรายละเอียดลึกซึ้งยิ่งขึ้น ซึ่งจะเป็นการเพิ่มประสิทธิภาพของโมเดลให้สามารถทำงานได้อย่างมีประสิทธิผลมากยิ่งขึ้นในอนาคต

รายการอ้างอิง

Acheampong, F. A., Wenyu, C., & Nunoo‐Mensah, H. (2020). Text‐based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, *2*(7), e12189.

Alm, E. C. O. (2008). *Affect in* text and speech*. Citeseer.

Alswaidan, N., & Menai, M. E. B. (2020). A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 1-51.

Amelin, K., Granichin, O., Kizhaeva, N., & Volkovich, Z. (2018). Patterning of writing style evolution by means of dynamic similarity. *Pattern Recognition*, *77*, 45-64.

Anastasia, J., Wahbeh, H., Delorme, A., & Okonsky, J. (2020). A qualitative exploratory analysis of channeled content. *EXPLORE*, *16*(4), 231-236.

Arreerard, R., & Senivongse, T. (2018). Thai defamatory text classification on social media. 2018 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD),

Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, *158*, 1429-1441.

Buechel, S., & Hahn, U. (2017). Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation. Proceedings of the 11th Linguistic Annotation Workshop,

Busche, L. (2017). *Powering content: building a nonstop content marketing machine*. O'Reilly Media, Inc..

Chae, M. J. (2020). The effects of message tone and formats of CSR messages on engagement in social media. *The Journal of Asian Finance, Economics and Business*, *7*(10), 501-511.

Chaffar, S., & Inkpen, D. (2011). Using a heterogeneous dataset for emotion analysis in text. Canadian conference on artificial intelligence,

Chandran, N. V., Anoop, V. S., & Asharaf, S. (2023). Topicstriker: A topic kernels-powered approach for text classification. Results in Engineering, 17, 100949.

Chen, C. (2020). *China's influence on media in southeast Asia: a case study of the Philippines, Thailand and Cambodia* (Doctoral dissertation, Hong Kong Baptist University).

Chen, Z., Liu, H., & Zhang, Q. (2023, July). Adaptive Selection of BERT Layer for Multi-Label Text Classification. In *2023 International Conference on Machine Learning and Cybernetics (ICMLC)* (pp. 434-439). IEEE.

Chopra, S., & Verma, M. (2023). Fake Viral Vidoes: The Delusion of Infotainment. IIS. Univ. JA, 11, 242-255.

COGNITO. (2021). *Text Classification Dataset for Machine Learning Algorithms*. Retrieved 1 June from https://www.cogitotech.com/services/text-classification

Conforti, C., Pilehvar, M. T., & Collier, N. (2018, November). Towards automatic fake news detection: cross-level stance detection in news articles. In *Proceedings of the first workshop on fact extraction and VERification (FEVER)* (pp. 40-49).

Damrongrat, C., Sangkeettrakarn, C., Piyatumrong, A., Trankultaweekoon, K., Seekhem, P., & Chuangkrud, P. (2022, May). A Non-Entity Approach for Intent-Based Classification: A Case Study of Thai News. In *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1-4). IEEE.

Daradkeh, M., Mansoor, W., Atalla, S., Himeur, Y., & Kerdjidj, O. (2023, June). Lifelong machine learning for topic modeling based on Hellinger distance. 2023 International Joint Conference on Neural Networks (IJCNN) (pp. 01–08). IEEE.

Dertat, A. (2017). *Applied Deep Learning - Part 4: Convolutional Neural Networks*. Retrieved 1 June from https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Devlin, J., & Chang, M.-W. (2018). Open sourcing BERT: state-of-the-art pre-training for natural language processing. *Google AI Blog*, *2*.

Diamond, S. (2016). *Content Marketing Strategies for Dummies*. John Wiley & Sons.

Ekman, P. (1993). Facial expression and emotion. *American psychologist*, *48*(4), 384.

Fei, H., Zhang, Y., Ren, Y., & Ji, D. (2020). Latent emotion memory for multi-label emotion classification. Proceedings of the AAAI Conference on Artificial Intelligence,

Flashman, G. (2020). *Powerful B2B content: using brand journalism to create compelling and authentic storytelling*. Kogan Page Publishers.

Frith, J. (2020). *Timely vs evergreen content: dividing your content budget*. Retrieved 1 June from https://www.finextra.com/blogposting/18803/timely-vs-evergreen-content-dividing-your-content-budget

Gadek, G., & Guélorget, P. (2020). An interpretable model to measure fakeness and emotion in news. *Procedia Computer Science*, *176*, 78-87.

Gallagher, J. R. (2020). *Update culture and the afterlife of digital writing*. University Press of Colorado.

Grall, C., & Finn, E. S. (2022). Leveraging the power of media to drive cognition: A media-informed approach to naturalistic neuroscience. Social Cognitive and Affective Neuroscience, 17(6), 598-608.

Gonçalves, A. (2021). Social media analytics strategy: Using data to optimize business performance.

Goobich, J. (2020). *Why Evergreen Content Can Be A Golden Ticket To Growth*. Retrieved 1 June from https://www.forbes.com/sites/forbescommunicationscouncil/2020/12/01/why-evergreen-content-can-be-a-golden-ticket-to-growth/

Goussakov, R. (2020). Hellinger Distance-based Similarity Measures for Recommender Systems.

Hadi, S., Pianto, H. A., Hafidah, A. S., Putra, H. R., & Arsyad, A. (2024). Understanding the Author's Purpose in Interpretative Reading. Asian Journal of Multidisciplinary Research and Analysis, 2(1), 140–145.

Halim, Z., Waqar, M., & Tahir, M. (2020). A machine learning-based investigation utilizing the in-text features for the identification of dominant emotion in an email. *Knowledge-Based Systems*, *208*, 106443.

Hall, B., & Wallace, E. (2018). Purpose, Audience, Tone, and Content. *College ESL Writers: Mohawk College Edition*.

Han, T. (2023, October). Research on Chinese Patent Text Classification in the Field of New Energy Vehicles Based on the XGBoost Model. In Proceedings of the 2023 7th International Conference on Electronic Information Technology and

Computer Engineering (pp. 481-485).

Hanlon, A. (2021). Digital marketing: strategic planning & integration.

He, M., Song, Y., Xu, K., & Dong, Y. (2020). On the Role of Conceptualization in Commonsense Knowledge Graph Construction. *arXiv preprint arXiv:2003.03239*.

Henricks, B. C., & Shelton, R. (2016). *Mastering the new media landscape: Embrace the micromedia mindset*. Berrett-Koehler Publishers.

Ho, J., Pang, C., & Choy, C. (2020). Content marketing capability building: a conceptual framework. *Journal of Research in Interactive Marketing*, *14*(1), 133-151.

Hou, B., O'connor, J., Andreas, J., Chang, S., & Zhang, Y. (2023, July). mt. In International Conference on Machine Learning (pp. 13309-13324). PMLR.

Huang, Y., & Ren, W. (2020). A novel multidimensional analysis of writing styles of editorials from China Daily and The New York Times. *Lingua*, *235*, 102781.

IBM. (2020). *Deep Learning*. Retrieved 1 June from https://www.ibm.com/cloud/learn/deep-learning

Inrak, P., & Sinthupinyo, S. (2010). Applying latent semantic analysis to classify emotions in Thai text. 2010 2nd International Conference on Computer Engineering and Technology,

Irwanto, A., & Goeirmanto, L. (2023). Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia Using Naïve Bayes and XGboost Classifier Algorithm. Sinergi, 27(2), 145-152.

Iyer, R. R., & Rose, C. P. (2019). A machine learning framework for authorship identification from texts. *arXiv preprint arXiv:1912.10204*.

Khandelwal, R. (2020). *Attention: Sequence 2 Sequence model with Attention Mechanism*. Retrieved 1 June from https://towardsdatascience.com/sequence-2-sequence-model-with-attention-mechanism-9e9ca2a613a

Kim, J. W., & Masullo Chen, G. (2021). Exploring the influence of comment tone and content in response to misinformation in social media news. *Journalism Practice*, *15*(4), 456-470.

Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., ... & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine

learning. *Heliyon*, *8*(10).

Liu, N., Wang, Q., & Ren, J. (2021). Label-Embedding Bi-directional Attentive Model for Multi-label Text Classification. *Neural Processing Letters*, *53*(1), 375-389.

Lowphansirikul, L., Polpanumas, C., Jantrakulchai, N., & Nutanong, S. (2021). Wangchanberta: Pretraining transformer-based thai language models. arXiv preprint arXiv:2101.09635.

Maruf, S., Javed, K., & Babri, H. A. (2016). Improving text classification performance with random forests-based feature selection. Arabian Journal for Science and Engineering, 41, 951-964.

Matook, S., Dennis, A. R., & Wang, Y. M. (2022). User comments in social media firestorms: A mixed-method study of purpose, tone, and motivation. *Journal of Management Information Systems*, *39*(3), 673-705.

Mehta, S., Jain, T., & Aggarwal, N. (2021). Multilingual short text analysis of twitter using random forest approach. In Knowledge Graphs and Semantic Web: Third Iberoamerican Conference and Second Indo-American Conference, KGSWC 2021, Kingsville, Texas, USA, November 22–24, 2021, Proceedings 3 (pp. 84-92). Springer International Publishing.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning--based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, *54*(3), 1-40.

Moirangthem, D. S., & Lee, M. (2021). Hierarchical and lateral multiple timescales gated recurrent units with pre-trained encoder for long text classification. *Expert Systems with Applications*, *165*, 113898.

Müller, J., & Christandl, F. (2019). Content is king–But who is the king of kings? The effect of content marketing, sponsored content & user-generated content on brand responses. *Computers in Human Behavior*, *96*, 46-55.

Nielek, R., Ciastek, M., & Kope**Ć**, W. (2017). Emotions make cities live: towards mapping emotions of older adults on urban space. Proceedings of the International Conference on Web Intelligence,

Noormanshah, W. M., Nohuddin, P. N., & Zainol, Z. (2021). Document content analysis based on random Forest algorithm. In Intelligent Computing and Innovation on

Data Science: Proceedings of ICTIDS 2019 (pp. 485-494). Singapore: Springer Singapore.

Onan, A. (2018). An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science*, *44*(1), 28-47.

Phann, R., Soomlek, C., & Seresangtakul, P. (2023). Multi-Class Text Classification on Khmer News Using Ensemble Method in Machine Learning Algorithms. Acta Informatica Pragensia, 12(2), 243-259.

Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT?. arXiv preprint arXiv:1906.01502.

Plubin, S., Bunyatisai, W., Khamkong, M., Mouktonglang, T., & Plubin, B. (2025). Multilabel-Thai Text Classification with Transformer-Rnn in Thai Banking Classification. *Pakistan Journal of Life & Social Sciences*, *23*(1).

Promrit, N., & Waijanya, S. (2017). Convolutional neural networks for thai poem classification. International Symposium on Neural Networks,

Putra, S. J., Gunawan, M. N., & Hidayat, A. A. (2022, September). Feature engineering with Word2vec on text classification using the K-nearest neighbor algorithm. In 2022 10th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE.

Qurashi, A. W., Holmes, V., & Johnson, A. P. (2020). Document Processing: Methods for Semantic Text Similarity Analysis. 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA).

Ríos-Toledo, G., Posadas-Durán, J. P. F., Sidorov, G., & Castro-Sánchez, N. A. (2022). Detection of changes in literary writing style using N-grams as style markers and supervised machine learning. *Plos one*, *17*(7), e0267590.

Rozado, D., Hughes, R., & Halberstadt, J. (2022). Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models. *Plos one*, *17*(10), e0276367.

Sabol, R., & Horák, A. (2022). Manipulative Style Recognition of Czech News Texts using Stylometric Text Analysis. *RASLAN 2022 Recent Advances in Slavonic Natural Language Processing*, 191.

Saedi, C., & Dras, M. (2021). Siamese networks for large-scale author identification. *Computer Speech & Language*, *70*, 101241.

Saha, S. (2018). *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way*. Retrieved 1 June from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Sangkatip, W., Chomphuwiset, P., Bunluewong, K., Mekruksavanich, S., Okafor, E., & Surinta, O. (2024). Improving Neural Network-Based Multi-Label Classification With Pattern Loss Penalties. IEEE Access.

Saputro, D. R. S., & Sidiq, K. (2023). CABLE NEWS NETWORK (CNN) ARTICLES CLASSIFICATION USING RANDOM FOREST ALGORITHM WITH HYPERPARAMETER OPTIMIZATION. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 17(2), 0847-0854.

Sarakit, P., Theeramunkong, T., Haruechaiyasak, C., & Okumura, M. (2015). Classifying emotion in Thai youtube comments. 2015 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES),

Siriphanpornchana, P., & Kanjanawattana, S. (2023, July). Comparative Analysis of Machine Learning Algorithms for Classification of Thai Fake News. In 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT) (pp. 453-458). IEEE.

Smith, J. (2020). *Digital marketing for businesses in easy steps*. In Easy Steps Limited.

Sriwirote, P., Thapiang, J., Timtong, V., & Rutherford, A. T. (2023). Phayathaibert: Enhancing a pretrained thai language model with unassimilated loanwords. arXiv preprint arXiv:2311.12475.

Sungthong, C., Ruangtanusak, S., & Songmuang, P. (2024, March). Multi-Label Classification for Low-Resource Issue Tickets with BERT. In *2024 IEEE International Conference on Cybernetics and Innovations (ICCI)* (pp. 1-6). IEEE.

Suramanka, L., & Hanskunatai, A. (2024, May). Multi-Label Classification of Foreign Tourists' Opinions on Thailand Tourism Development. In *Proceedings of the 2024 9th International Conference on Big Data and Computing* (pp. 32-38).

Taha, K., Yoo, P. D., Yeun, C., & Taha, A. (2024). Text classification: A review, empirical, and experimental evaluation. arXiv preprint arXiv:2401.12982.

Taherkhani, L., Daneshvar, A., Amoozad Khalili, H., & Sanaei, M. R. (2023). Analysis of the customer churn prediction project in the hotel industry based on text mining and the random forest algorithm. Advances in Civil Engineering, 2023(1), 6029121.

Tuten, T., & Mintu-Wimsatt, A. (2018). Advancing our understanding of the theory and practice of social media marketing: Introduction to the special issue. *Journal of Marketing Theory and Practice*, *26*(1-2), 1-3.

Vatathanavaro, S., Pasupa, K., Sirirattanajakarin, S., & Suntisrivaraporn, B. (2020, September). Improved identification of imbalanced multiple annotation intent labels with a hybrid BLSTM and CNN model and hybrid loss function. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 355-368). Cham: Springer International Publishing.

Waijanya, S., & Promrit, N. (2017). The poet identification using convolutional neural networks. International Conference on Computing and Information Technology,

Wang, Y., Ma, K., Garcia-Hernandez, L., Chen, J., Hou, Z., Ji, K., Chen, Z., & Abraham, A. (2020). A CLSTM-TMN for marketing intention detection. *Engineering Applications of Artificial Intelligence*, *91*, 103595.

Wang, Y., Zhao, X., Zhang, Z., & Zhang, L. Y. (2022). A collaborative filtering algorithm based on item labels and Hellinger distance for sparse data. Journal of Information Science, 48(6), 749–766.

Willoughby, J. F., Couto, L., Kang, S., Randall, J., Kirkpatrick, A. W., Lee, D. K. L., ... & Domgaard, S. (2024). An exploratory content analysis of the use of health communication strategies and presence of objectification in fitness influencer social media posts. *Health Communication*, *39*(5), 888-895.

Wilson, L. (2019). *30-minute website marketing: a step by step guide*. Emerald Publishing Limited.

Xiong, J., Yu, L., Zhang, D., & Leng, Y. (2021). DNCP: An attention-based deep learning approach enhanced with attractiveness and timeliness of News for online news click prediction. *Information & Management*, *58*(2), 103428.

Yam, C.-Y. (2015). Emotion detection and recognition from text using deep learning. *Microsoft Developer Blog*.

Yeramosu, T., Dominy, C. L., Arvind, V., Isleem, U. N., & Cho, S. K. (2023). Scoliosis surgery: a social media analysis of content, Tone, and perspective. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, *31*(1), 26-33.

Zhang, T., Li, S., Jing, X., Song, J., Shi, L., & He, X. (2023, March). Negative comment recognition model based on lightGBM. In Second International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2022) (Vol. 12597, pp. 1027-1032). SPIE.

Zheng, J., Cai, F., Chen, H., & de Rijke, M. (2020). Pre-train, Interact, Fine-tune: a novel interaction representation for text classification. *Information Processing & Management*, *57*(6), 102215.

Zhou, D., Wu, S., Wang, Q., Xie, J., Tu, Z., & Li, M. (2020). Emotion Classification by Jointly Learning to Lexiconize and Classify. Proceedings of the 28th International Conference on Computational Linguistics.

ปรียา กร บุญธรรม. (2023). ความ นิยม ของ ผู้ เรียน ภาษา จีน ที่ มี ต่อ ผล งาน วรรณกรรม จีน แปล ไทย: ความ นิยม ของ ผู้ เรียน ภาษา จีน ที่ มี ต่อ ผล งาน วรรณกรรม จีน แปล ไทย. *วารสาร วิจัย และ พัฒนา อนุ ภูมิภาค ลุ่มน้ำ โขง*, *2*(3), 17-25.

พัช ชา ชาติ พาณิชย์, & ยุว ริน ธร ไชย โชติช่วง. (2024). ปัจจัย ที่ มี อิทธิพล ต่อ การ ตัดสินใจ เลือก เรียน ภาษา จีน ผ่าน ช่อง ทาง ออนไลน์ ของ Generation Y ใน ประเทศไทย. *วารสาร มณี เชษฐา ราม วัด จอม มณี*, *7*(6), 436-451.

มหาวิทยาลัยศิลปากร

ประวัติผู้เขียน

ชื่อ-สกุล	ภากร กัทชลี
วุฒิการศึกษา	Master of Engineering, Computer Software and Theory, Beihang University กรุงปักกิ่ง ประเทศจีน วิทยาศาสตรบัณฑิต วิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่
ผลงานตีพิมพ์	Gatchalee, P., Waijanya, S., & Promrit, P. (2023). Thai text classification experiment using cnn and transformer models for timely-timeless content marketing. ICIC Express Letters, 19, 91-101. Pagon, G., Waijanya, S., & Promrit, N. (2025). A hybrid approach to analyzing storytelling and translation about Chinese topics on Thai Facebook pages. ICIC Express Letters, Part B: Applications – An International Journal of Research and Surveys. (in press). Gatchalee, P., Waijanya, S., & Promrit, N. (2025). Intention classification of Chinese topics on Thai Facebook pages using transformer models with emotional features. International Conference, The 6th Asia Joint Conference on Computing (AJCC2025). 21st-24th April 2025, Osaka, Japan.
รางวัลที่ได้รับ	รางวัล Best Presentation Award ในงาน AJCC 2022: The Asia Joint Conference on Computing ระหว่างวันที่ 24-25 กุมภาพันธ์ 2565 ชื่อผลงาน: Thai Text Classification Experiment Using Transformer and 1D-CNN Models for Timely-Timeless Content Marketing