การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ

SENTIMENT ANALYSIS OF STUDENT FEEDBACK ON UNIVERSITY SERVICES USING NATURAL LANGUAGE PROCESSING TECHNIQUES

รัตนะ วงษ์บุญหนัก

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2567

การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ

รัตนะ วงษ์บุญหนัก

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567

SENTIMENT ANALYSIS OF STUDENT FEEDBACK ON UNIVERSITY SERVICES USING NATURAL LANGUAGE PROCESSING TECHNIQUES

RATTANA WONGBOONNAK

An Independent Study Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2024

สารนิพนธ์

เรื่อง

การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิคการประมวลผล

ภาษาธรรมชาติ

ของ

รัตนะ วงษ์บุญหนัก

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิยานนท์)

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.จันตรี ผลประเสริฐ)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ
ผู้วิจัย	รัตนะ วงษ์บุญหนัก
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2567
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. วราภรณ์ วิยานนท์

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ความรู้สึกของนักศึกษาที่มีต่อการให้บริการของมหาวิทยาลัย โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) จากความคิดเห็นเชิงข้อความจำนวน 1,000 รายการที่รวบรวมระหว่างปีการศึกษา 2564–2566 งานวิจัยได้เปรียบเทียบประสิทธิภาพของแบบจำลอง Naïve Bayes, Support Vector Machine (SVM) และ Random Forest โดยใช้การแทนค่าคำแบบ TF-IDF และ Word2Vec ร่วมกับเทคนิคจัดการข้อมูลไม่สมดุล ได้แก่ SMOTE, การสุ่มลดข้อมูล (Random Undersampling) และการถ่วงน้ำหนักคลาส (Class Weighting) ผลการทดลองพบว่าแบบจำลอง SVM ที่ใช้ TF-IDF ร่วมกับ SMOTE ให้ผลลัพธ์ดีที่สุด โดยมีค่า Accuracy 0.8636 และ F1-score 0.8630 นอกจากนี้ ยังได้ดำเนินการวิเคราะห์หัวข้อความคิดเห็นด้วยเทคนิค Latent Dirichlet Allocation (LDA) พบ 7 หัวข้อหลัก ได้แก่ (1) โครงสร้างพื้นฐานด้านเทคโนโลยีและพื้นที่การใช้งาน (2) สภาพแวดล้อมการเรียนรู้และสิ่งอำนวยความสะดวก (3) ระบบเอกสารและบริการการเงิน (4) กิจกรรมนักศึกษาและการพัฒนา (5) สภาพแวดล้อมทางกายภาพและการบริการ (6) ทรัพยากรการเรียนรู้และเครือข่าย และ (7) การเรียนการสอนและหลักสูตร การวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึกพบว่า หัวข้อที่ 5 มีสัดส่วนความรู้สึกเชิงบวกสูงสุด (74.19%) ขณะที่หัวข้อที่ 3 มีสัดส่วนความรู้สึกเชิงลบสูงสุด (46.15%) ผลการวิจัยนี้สามารถใช้เป็นแนวทางในการปรับปรุงคุณภาพการให้บริการของมหาวิทยาลัย โดยเฉพาะในประเด็นที่นักศึกษาแสดงความไม่พึงพอใจในระดับสูง

คำสำคัญ : การวิเคราะห์ความรู้สึก, การประมวลผลภาษาธรรมชาติ, การวิเคราะห์หัวข้อ, การเรียนรู้ของเครื่อง, การให้บริการของมหาวิทยาลัย

Title	SENTIMENT ANALYSIS OF STUDENT FEEDBACK ON UNIVERSITY SERVICES USING NATURAL LANGUAGE PROCESSING TECHNIQUES
Author	RATTANA WONGBOONNAK
Degree	MASTER OF SCIENCE
Academic Year	2024
Thesis Advisor	Waraporn Viyanon

This research analyzes student sentiments toward university services using Natural Language Processing techniques. The scope of analysis includes 1,000 textual feedback entries collected during the academic years 2021–2023. The research compares the performance of various models including Naïve Bayes, Support Vector Machine (SVM), and Random Forest, using TF-IDF and Word2Vec word representation techniques combined with imbalanced data handling methods: SMOTE, Random Undersampling, and Class Weighting. Results show that the SVM model combined with TF-IDF and SMOTE technique achieved the highest performance in sentiment analysis with 0.8636 Accuracy and 0.8630 F1-score. Topic modeling using LDA identified 7 main topics: 1) Technology Infrastructure and Usable Space, 2) Learning Environment and Facilities, 3) Documentation and Financial Services, 4) Student Activities and Development, 5) Physical Environment and Services, 6) Learning Resources and Networks, and 7) Teaching and Curriculum. Analysis of the relationship between topics and sentiments reveals that Topic 5 (Physical Environment and Services) had the highest proportion of positive sentiment (74.19%), while Topic 3 (Documentation and Financial Services) had the highest proportion of negative sentiment (46.15%). These findings can be used to improve the quality of university services, especially in areas with high negative sentiment.

Keyword : Sentiment Analysis Natural Language Processing Topic Modeling Machine Learning University Services

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. วราภรณ์ วิยานนท์ อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ที่ได้กรุณาให้คำแนะนำ คำปรึกษา และดูแลเอาใจใส่ในการทำวิจัยครั้งนี้ด้วยความเมตตาและอดทนอย่างยิ่ง รวมทั้งได้ให้ความรู้ ประสบการณ์ และแนวทางในการทำวิจัยที่มีคุณค่าตลอดระยะเวลาการศึกษา

ขอขอบพระคุณคณะกรรมการสอบ ผู้ช่วยศาสตราจารย์ ดร. จันตรี ผลประเสริฐ ประธานกรรมการและผู้ช่วยศาสตราจารย์ ดร. ศิริสรรพ เหล่าหะเกียรติ ที่ได้ให้ข้อเสนอแนะอันมีค่าเพื่อปรับปรุงงานวิจัยให้มีคุณภาพยิ่งขึ้น

ขอขอบคุณ คณาจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ได้ถ่ายทอดความรู้และสิ่งอำนวยความสะดวกในการทำวิจัย รวมทั้งเจ้าหนาที่ประจำหน่วยงานทุกท่านที่ได้ให้ความช่วยเหลือด้วยดี

สุดท้ายนี้ ขอกราบขอบพระคุณครอบครัว พ่อ แม่ และบุคคลสำคัญทุกท่าน ที่ได้ให้การสนับสนุน กำลังใจ และความเข้าใจตลอดระยะเวลาการศึกษา

หากงานวิจัยชิ้นนี้มีประโยชน์ต่อวงการวิชาการและสังคม ผู้วิจัยขอมอบเป็นกตัญญูกตเวทิตาคุณแก่บูรพาจารย์และผู้มีพระคุณทุกท่าน หากมีข้อผิดพลาดประการใด ผู้วิจัยขอรับผิดชอบแต่เพียงผู้เดียว

รัตนะ วงษ์บุญหนัก

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ในยุคดิจิทัลที่เทคโนโลยีมีการพัฒนาอย่างก้าวกระโดด สถาบันการศึกษาจำเป็นต้องปรับตัวและพัฒนาอย่างต่อเนื่องเพื่อรักษามาตรฐานการศึกษาและตอบสนองความต้องการของผู้เรียนที่เปลี่ยนแปลงไป การบริหารจัดการสิ่งอำนวยความสะดวกและทรัพยากรทางการศึกษาให้มีประสิทธิภาพจึงเป็นความท้าทายสำคัญของผู้บริหารสถาบันการศึกษา โดยเฉพาะอย่างยิ่งในด้านการประเมินคุณภาพการให้บริการและการพัฒนาสิ่งอำนวยความสะดวกให้สอดคล้องกับความต้องการของผู้ใช้งาน สถาบันการศึกษาในปัจจุบันได้พัฒนาระบบการรวบรวมข้อมูลป้อนกลับจากผู้ใช้บริการ โดยเฉพาะอย่างยิ่งจากนักศึกษาซึ่งเป็นผู้ใช้บริการหลัก ผ่านช่องทางที่หลากหลาย เช่น แบบสอบถามออนไลน์ ระบบประเมินการเรียนการสอน รวมถึงแพลตฟอร์มสื่อสังคมออนไลน์ ข้อมูลที่ได้รับประกอบด้วยทั้งข้อมูลเชิงปริมาณ เช่น คะแนนความพึงพอใจ และข้อมูลเชิงคุณภาพในรูปแบบของข้อความแสดงความคิดเห็นหรือข้อเสนอแนะ อย่างไรก็ตาม การวิเคราะห์ข้อมูลข้อความขนาดใหญ่ (Large-scale Text Analysis) ยังคงเป็นความท้าทายสำคัญด้วยหลายปัจจัย

ประการแรก ข้อมูลความคิดเห็นมีความซับซ้อนและหลากหลาย ทั้งในแง่ของภาษา การใช้คำ และบริบท การวิเคราะห์ด้วยวิธีการดั้งเดิมที่ใช้การอ่านและตีความโดยมนุษย์จึงใช้เวลามากและอาจเกิดความลำเอียงในการตีความ

ประการที่สอง การเพิ่มขึ้นอย่างต่อเนื่องของปริมาณข้อมูลส่งผลให้วิธีการประมวลผลแบบดั้งเดิมไม่สามารถรองรับได้อย่างมีประสิทธิภาพ และประการที่สาม การขาดแคลนเครื่องมือที่สามารถวิเคราะห์ความรู้สึกและอารมณ์ที่แฝงอยู่ในข้อความได้อย่างเหมาะสม อาจทำให้ประเด็นสำคัญบางประการที่ควรได้รับการแก้ไขอย่างเร่งด่วนถูกมองข้ามไป

นอกจากนี้ การพัฒนาเทคโนโลยีอย่างรวดเร็วยังส่งผลให้ความคาดหวังของนักศึกษาต่อสิ่งอำนวยความสะดวกและบริการทางการศึกษาเพิ่มสูงขึ้น สถาบันการศึกษาจึงจำเป็นต้องมีเครื่องมือที่มีประสิทธิภาพในการวิเคราะห์และประมวลผลข้อมูลป้อนกลับจากผู้ใช้บริการ เพื่อนำไปสู่การพัฒนาและปรับปรุงการให้บริการอย่างตรงจุดและทันท่วงที ด้วยเหตุนี้ การนำเทคโนโลยีที่เกี่ยวกับการวิเคราะห์ความรู้สึก (Sentiment Analysis) มาประยุกต์ใช้ในการประมวลผลข้อมูลความคิดเห็นจึงเป็นแนวทางที่มีความสำคัญ เนื่องจากสามารถช่วยจัดการกับความท้าทายดังกล่าวอย่างมีประสิทธิภาพ ผ่านการวิเคราะห์ข้อมูลเชิงข้อความจำนวนมากอย่างเป็นระบบ การจำแนกประเภทความรู้สึกและอารมณ์ที่แฝงอยู่ในข้อความ ตลอดจนการระบุ

ประเด็นสำคัญที่ต้องได้รับการพัฒนาปรับปรุง อันจะนำไปสู่การยกระดับคุณภาพการให้บริการและการพัฒนาสิ่งอำนวยความสะดวกทางการศึกษาที่ตอบสนองความต้องการของผู้ใช้งานได้อย่างแท้จริง

1.2 วัตถุประสงค์

1. เพื่อพัฒนาโมเดลสำหรับการวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิค NLP

2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดลต่างๆ ในการวิเคราะห์ความรู้สึกจากข้อความแสดงความคิดเห็นของนักศึกษา ได้แก่ Naïve Bayes, Support Vector Machine (SVM) และ Random Forest โดยใช้เทคนิคการแทนค่าคำแบบ TF-IDF และ Word2Vec

3. เพื่อค้นพบและจำแนกหัวข้อหลักในความคิดเห็นของนักศึกษาโดยใช้การวิเคราะห์หัวข้อด้วยวิธี Latent Dirichlet Allocation (LDA)

4. เพื่อวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึกของนักศึกษาเพื่อระบุประเด็นที่ต้องได้รับการปรับปรุง

1.3 ขอบเขตการวิจัย

ข้อมูลที่ใช้ในการวิเคราะห์เป็นความคิดเห็นของนักศึกษาต่อการให้บริการของมหาวิทยาลัยเกริกที่รวบรวมจากผลการประเมินความพึงพอใจโดยนักศึกษาในช่วงปีการศึกษา 2564-2566 จำนวน 1,000 รายการ

1. การจำแนกความรู้สึกแบ่งออกเป็น 3 ประเภท ได้แก่ เชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative) โดยใช้เกณฑ์จาก VADER (Valence Aware Dictionary and Sentiment Reasoner)

2. เทคนิค NLP และอัลกอริทึมการเรียนรู้ของเครื่องที่ใช้ในการสร้างโมเดลวิเคราะห์ความรู้สึก ประกอบด้วย

1) เทคนิคการแปลงข้อความเป็นข้อมูลเชิงตัวเลข ได้แก่ TF-IDF (Term Frequency-Inverse Document Frequency) และ Word2Vec
2) อัลกอริทึมการเรียนรู้ของเครื่อง เช่น Naïve Bayes (ทั้งแบบ MultinomialNB และ GaussianNB), Support Vector Machine (LinearSVC) และ Random Forest
3) เทคนิคสำหรับจัดการกับปัญหาข้อมูลไม่สมดุล SMOTE (Synthetic Minority Over-sampling Technique), Random Under Sampling และ Class Weighting

3. การวิเคราะห์หัวข้อใช้วิธี LDA (Latent Dirichlet Allocation) โดยหาจำนวนหัวข้อที่เหมาะสมด้วยการวิเคราะห์ค่า Coherence Score และ Perplexity

4. การประเมินประสิทธิภาพของโมเดลดำเนินการโดยใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall และ F1-Score

5. ภาษาที่ใช้ในการวิเคราะห์คือภาษาอังกฤษ โดยมีการแปลความคิดเห็นภาษาไทยเป็นภาษาอังกฤษด้วย Google Translate API

1.4 ขั้นตอนการดำเนินงาน

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความรู้สึกโดยใช้เทคนิค NLP และการเรียนรู้ของเครื่อง (Machine Learning) รวมถึงขั้นตอนการวิเคราะห์หัวข้อ (Topic Modeling)

2. รวบรวมและจัดเตรียมข้อมูลความคิดเห็นของนักศึกษาต่อการให้บริการของมหาวิทยาลัยเกริกโดยรวบรวมจากผลการประเมินความพึงพอใจ

3. เตรียมข้อมูลด้วยการทำความสะอาดข้อมูล (Data Cleaning) การตรวจสอบและแปลงภาษา (Language Identification and Translation) และการประมวลผลข้อความ (Text Preprocessing) ซึ่งประกอบด้วยการแบ่งคำ (Tokenization) การลบคำหยุด (Stop Words) และการลดรูปคำ (Stemming/Lemmatization)

4. วิเคราะห์ความรู้สึกเบื้องต้นด้วย VADER (Valence Aware Dictionary and Sentiment Reasoner) เพื่อจำแนกข้อความเป็น 3 ประเภท คือ เชิงบวก (Positive) เป็นกลาง (Neutral) และเชิงลบ (Negative)

5. สกัดคุณลักษณะสำคัญจากข้อความ (Feature Extraction) ด้วยเทคนิค TF-IDF และ Word2Vec เพื่อแปลงข้อความให้อยู่ในรูปแบบที่อัลกอริทึมการเรียนรู้ของเครื่องที่สามารถเข้าใจได้

6. การพัฒนาโมเดลสำหรับวิเคราะห์ความรู้สึก โดยใช้อัลกอริทึม Naïve Bayes, Support Vector Machine (SVM) และ Random Forest ร่วมกับเทคนิคการจัดการข้อมูลไม่สมดุล เช่น SMOTE (Synthetic Minority Over-sampling Technique)

7. ทดสอบและประเมินประสิทธิภาพของโมเดลด้วยเกณฑ์ Accuracy, Precision, Recall, F1-Score

8. วิเคราะห์หัวข้อด้วยเทคนิค LDA (Latent Dirichlet Allocation) เพื่อค้นหาหัวข้อหลักในความคิดเห็นของนักศึกษา โดยหาจำนวนหัวข้อที่เหมาะสมด้วยการวิเคราะห์ค่า Coherence Score และ Perplexity

9. วิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึก เพื่อระบุประเด็นที่นักศึกษามีความพึงพอใจและไม่พึงพอใจ

10. สรุปผลการวิจัย และให้ข้อเสนอแนะสำหรับการปรับปรุงคุณภาพการให้บริการของมหาวิทยาลัย

11. รายงานการวิจัยและนำเสนอผลการวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. พัฒนาโมเดลที่มีประสิทธิภาพในการวิเคราะห์ความรู้สึกของนักศึกษาที่มีต่อการให้บริการของมหาวิทยาลัย โดยสามารถจำแนกความรู้สึกออกเป็นสามประเภท ได้แก่ เชิงบวก เป็นกลาง และเชิงลบ ได้อย่างถูกต้องและแม่นยำ

2. มีข้อมูลเชิงลึกเกี่ยวกับหัวข้อหลักที่นักศึกษาให้ความสำคัญในการประเมินการให้บริการของมหาวิทยาลัย และความสัมพันธ์ระหว่างหัวข้อเหล่านั้นกับความรู้สึกของนักศึกษา

3. พัฒนาองค์ความรู้ในการประยุกต์ใช้เทคนิค NLP และการเรียนรู้ของเครื่องสำหรับการวิเคราะห์ความรู้สึกและการวิเคราะห์หัวข้อในบริบทของสถาบันการศึกษา

4. มหาวิทยาลัยสามารถนำผลการวิเคราะห์ไปใช้ในการปรับปรุงคุณภาพการให้บริการอย่างตรงจุด โดยมุ่งเน้นการแก้ไขประเด็นที่มีความรู้สึกเชิงลบ

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้นำเสนอทฤษฎีและเทคนิคสำคัญที่เกี่ยวข้องกับการวิเคราะห์ข้อความ ตั้งแต่หลักการพื้นฐานของการเรียนรู้ของเครื่องและการประมวลผลภาษาธรรมชาติ ไปจนถึงเทคนิคเฉพาะทางด้านการวิเคราะห์ความรู้สึกและการจำแนกหัวข้อ ทั้งนี้เพื่อวางรากฐานความเข้าใจในการพัฒนาระบบวิเคราะห์ความคิดเห็นของนักศึกษาที่มีต่อการให้บริการของมหาวิทยาลัย ซึ่งจะสามารถสนับสนุนให้สถาบันการศึกษานำข้อมูลที่ได้ไปใช้ในการปรับปรุงคุณภาพการให้บริการและยกระดับความพึงพอใจของนักศึกษาได้อย่างเป็นรูปธรรม

2.1 ทฤษฎีทีเกี่ยวข้อง

2.1.1 Machine Learning

การเรียนรู้ของเครื่อง (Machine Learning) เป็นแขนงหนึ่งของวิทยาการคอมพิวเตอร์ ที่มุ่งเน้นการพัฒนาระบบหรือแบบจำลองซึ่งสามารถเรียนรู้และปรับปรุงประสิทธิภาพของตนเองได้จากข้อมูลที่ได้รับ โดยไม่จำเป็นต้องอาศัยการเขียนโปรแกรมแบบกำหนดคำสั่งอย่างชัดเจน การเรียนรู้ของเครื่องสามารถแบ่งออกได้เป็นหลายรูปแบบ ได้แก่

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นวิธีที่ใช้ชุดข้อมูลฝึกฝนซึ่งมีการระบุคำตอบหรือเป้าหมาย (Labels) ไว้อย่างชัดเจน(Goodfellow et al., 2016) โดยโมเดลจะเรียนรู้ความสัมพันธ์ระหว่างข้อมูลนำเข้าและผลลัพธ์ที่ต้องการ เพื่อให้สามารถทำนายผลลัพธ์ของข้อมูลใหม่ได้อย่างแม่นยำ ตัวอย่างของวิธีนี้ ได้แก่ การจำแนกประเภท (Classification)

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ใช้ข้อมูลที่ไม่มีการระบุคำตอบหรือเป้าหมาย โดยโมเดลจะค้นหาโครงสร้างหรือรูปแบบที่ซ่อนอยู่ในข้อมูลด้วยตนเอง(Hastie et al., 2009) เช่น การจัดกลุ่ม (Clustering)

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นวิธีที่ตัวแทน (Agent) เรียนรู้จากปฏิสัมพันธ์กับสภาพแวดล้อมผ่านการทดลองและรับผลตอบแทน (Reward) หรือการลงโทษ (Penalty)(Sutton & Barto, 2018) วิธีการนี้มักใช้ในงานที่ต้องการการตัดสินใจแบบต่อเนื่อง เช่น การวางแผนเส้นทาง

การการเรียนรู้ของเครื่องถูกนำไปประยุกต์ใช้ในหลากหลายงาน เช่น การวิเคราะห์ข้อความ (Text Analysis), การตรวจจับความผิดปกติ (Anomaly Detection) และการวิเคราะห์เชิงคาดการณ์ (Predictive Analysis) สำหรับการวิจัยฉบับนี้ ได้ประยุกต์ใช้แนวคิดการเรียนรู้ของ

เครื่องในการวิเคราะห์ความคิดเห็นของนักศึกษา เพื่อระบุความรู้สึก (Sentiment) และหัวข้อสำคัญที่เกี่ยวข้องกับการให้บริการของมหาวิทยาลัย

2.1.2 Natural Language Processing

การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นแขนงหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นการพัฒนาความสามารถของคอมพิวเตอร์ในการเข้าใจ ตีความ และประมวลผลภาษามนุษย์ในรูปแบบที่ใกล้เคียงกับธรรมชาติของการสื่อสารของมนุษย์

NLP เป็นศาสตร์สหวิทยาการที่ผสมผสานองค์ความรู้จากหลากหลายสาขา ได้แก่ วิทยาการคอมพิวเตอร์ ภาษาศาสตร์ และสถิติ เพื่อสร้างแบบจำลองที่สามารถวิเคราะห์และเข้าใจภาษามนุษย์ได้อย่างมีประสิทธิภาพ โดยกระบวนการพื้นฐานของ NLP ประกอบด้วยขั้นตอนสำคัญหลายประการ เช่น การตัดคำ (Tokenization), การแปลงคำเป็นราก (Stemming หรือ Lemmatization), การลบคำที่ไม่สำคัญ (Stop Word Removal) และการแปลงข้อความเป็นเวกเตอร์เชิงตัวเลข(Goldberg, 2017)ดังนี้

1. การประมวลผลข้อความเบื้องต้น (Text Preprocessing) เป็นขั้นตอนแรกในการเตรียมข้อมูลสำหรับการวิเคราะห์ ประกอบด้วย

1) การทำความสะอาดข้อมูล (Data Cleaning) เช่น การกำจัดอักขระพิเศษ แท็ก HTML หรือสัญลักษณ์ที่ไม่เกี่ยวข้อง
2) การแบ่งคำ (Tokenization)คือการแยกข้อความออกเป็นหน่วยย่อยๆ เช่น คำ วลี หรือประโยค
3) การกำจัดคำที่ไม่มีความหมาย (Stop Words)เป็นการกำจัดคำที่พบบ่อยแต่ไม่มีนัยสำคัญต่อความหมาย เช่น "และ", "ที่", "เป็น"
4) การลดรูปคำ (Stemming/Lemmatization) คือการแปลงคำให้อยู่ในรูปแบบพื้นฐาน เช่น "running" เป็น "run

2. การวิเคราะห์ความหมายของคำ (Semantic Analysis) เป็นขั้นตอนที่พยายามเข้าใจความหมายและความสัมพันธ์ระหว่างคำในประโยค(Manning & Schutze, 1999)

1) การสร้างถุงคำ (Bag of Words) หรือ TF-IDF เพื่อแสดงความสำคัญของคำในข้อความ
2) การฝังคำ (Word Embeddings) เช่น Word2Vec, GloVe หรือ FastText ที่แสดงความสัมพันธ์ทางความหมายระหว่างคำในรูปแบบเวกเตอร์

3. การวิเคราะห์บริบท (Contextual Analysis) ซึ่งพิจารณาความหมายของคำในบริบท

1) โครงสร้างประโยค (Syntax Analysis)

2) ความสัมพันธ์ระหว่างคำ (Dependency Parsing)

3) ความหมายในระดับประโยค (Sentence-level Semantics)

เทคนิคการประมวลผลภาษาธรรมชาติ (NLP) ที่ได้รับความนิยมในปัจจุบันมีหลากหลายประเภท จากการศึกษาของ Min et al. (2021) พบว่า NLP มีการพัฒนาอย่างรวดเร็วในช่วงหลายปีที่ผ่านมา โดยเฉพาะการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก (Deep Learning) เช่น BERT, GPT และ Transformer ซึ่งสามารถเข้าใจบริบทและความหมายของข้อความได้ลึกซึ้งและแม่นยำยิ่งขึ้นเมื่อเทียบกับเทคนิคแบบดั้งเดิม อย่างไรก็ตาม งานวิจัยฉบับนี้มุ่งเน้นการประยุกต์ใช้เทคนิค NLP แบบดั้งเดิมร่วมกับอัลกอริทึมการเรียนรู้ของเครื่อง ในการวิเคราะห์ความรู้สึกและจำแนกหัวข้อจากข้อความแสดงความคิดเห็นของนักศึกษา

2.1.3 Naïve Bayes

Naïve Bayes เป็นอัลกอริทึมการเรียนรู้ของเครื่องที่อยู่ในกลุ่มการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้หลักการทางสถิติตามทฤษฎีของ Bayes' Theorem ในการคำนวณความน่าจะเป็นของเหตุการณ์ต่างๆ อัลกอริทึมนี้ได้รับความนิยมอย่างแพร่หลายในงานด้านการจำแนกประเภทข้อความ (Text Classification) และการวิเคราะห์ความรู้สึก (Sentiment Analysis) สามารถแสดงได้ด้วยสมการ (1)

$$P(A \mid B) = \frac{P(A) \times P(B|A)}{P(B)} \qquad (1)$$

โดยที่ ได้เกิดขึ้นแล้ว

$P(A|B)$ ความน่าจะเป็นการเกิดเหตุการณ์ A ภายใต้เงื่อนไขว่าเหตุการณ์ B ได้เกิดขึ้นแล้ว

$P(B|A)$ ความน่าจะเป็นการเกิดเหตุการณ์ B ภายใต้เงื่อนไขว่าเหตุการณ์ A ได้เกิดขึ้นแล้ว

$P(A)$ ความน่าจะเป็นของการเกิดเหตุการณ์ A โดยไม่ขึ้นกับเหตุการณ์อื่น

$P(B)$ ความน่าจะเป็นของการเกิดเหตุการณ์ B โดยไม่ขึ้นกับเหตุการณ์อื่น

ประเภทของ Naïve Bayes ที่ใช้ได้ดีกับปัญหาการจำแนกประเภทข้อความ (Text Classification) และการวิเคราะห์ความรู้สึก (Sentiment Analysis) มีดังนี้

1. Multinomial Naïve Bayes เป็นรูปแบบที่เหมาะสำหรับข้อมูลที่มีการนับความถี่ เช่น ความถี่ของคำในเอกสาร(Naulak, 2022)เหมาะสำหรับข้อมูลประเภท TF-IDF ที่เป็นการนับจำนวนคำในเอกสาร

2. Gaussian Naïve Bayes เป็นรูปแบบที่เหมาะกับข้อมูลต่อเนื่อง (Continuous Data) ที่มีการแจกแจงแบบปกติ (Normal Distribution) ซึ่งเหมาะกับข้อมูลที่แปลงเป็นเวกเตอร์แบบ Word2Vec ที่มีค่าต่อเนื่อง(Al-Saqqa & Awajan, 2019)

Naïve Bayes เป็นโมเดลที่มีจุดเด่นในด้านความรวดเร็วในการประมวลผล เข้าใจง่าย และสามารถให้ผลลัพธ์ที่ดีแม้มีข้อมูลฝึกอบรมน้อย อย่างไรก็ตาม ข้อจำกัดสำคัญของโมเดลนี้คือ การตั้งสมมติฐานว่าคุณลักษณะต่าง ๆ เป็นอิสระต่อกัน ซึ่งไม่สอดคล้องกับลักษณะข้อมูลจริง โดยเฉพาะในบริบทของข้อความที่มักมีความสัมพันธ์เชิงบริบทระหว่างคำ

2.1.4 Support Vector Machine

Zhang and Wallace (2015)ได้ศึกษาการประยุกต์ใช้ SVM ในการวิเคราะห์ความรู้สึกจากข้อความบนสื่อสังคมออนไลน์ พบว่า SVM สามารถจำแนกความรู้สึกได้แม่นยำประมาณ 82-87% ในหลายภาษา การทำงานของ SVM อาศัยฟังก์ชันเคอร์เนล (Kernel Function) ในการแปลงข้อมูลไปยังมิติที่สูงขึ้น ทำให้สามารถแยกข้อมูลที่ไม่สามารถแบ่งแยกได้ในมิติเดิม(Chen & Liu, 2018)

สำหรับการวิเคราะห์ข้อความ LinearSVC เป็นรูปแบบของ SVM ที่มีประสิทธิภาพสูงและเหมาะสมกับข้อมูลที่มีมิติสูง เช่น ข้อมูลข้อความที่แปลงเป็นเวกเตอร์ด้วยเทคนิค TF-IDF หรือ Word2Vec(Kumar & Garg, 2019) LinearSVC ใช้ฟังก์ชันเคอร์เนลแบบเชิงเส้น (Linear Kernel) ซึ่งมีความเร็วในการประมวลผลสูงและประสิทธิภาพดีสำหรับข้อมูลข้อความ

Support Vector Machine (SVM) มีความสามารถในการจัดการกับข้อมูลที่มีมิติสูงได้อย่างมีประสิทธิภาพ และมีความทนทานต่อปัญหา Overfitting รวมถึงสามารถจำแนกประเภทข้อมูลได้อย่างแม่นยำ (Thanaki, 2017) อย่างไรก็ตาม ข้อจำกัดของ SVM คือความซับซ้อนในการปรับพารามิเตอร์ และอาจใช้เวลาในการประมวลผลค่อนข้างนานเมื่อทำงานกับชุดข้อมูลขนาดใหญ่

2.1.5 Random Forest

Random Forest เป็นอัลกอริทึมการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning Algorithm) ซึ่งได้รับการพัฒนาโดย Breiman (Breiman, 2001) ซึ่งสร้างต้นไม้ตัดสินใจหลายต้นที่มีความหลากหลาย โดยแต่ละต้นถูกสร้างจากชุดข้อมูลย่อยที่แตกต่างกัน (Bootstrap Sampling) และคุณลักษณะที่ถูกสุ่มเลือก ผลลัพธ์สุดท้ายได้จากการโหวตเสียงข้างมากของทุกต้นไม้ (การจำแนกประเภท) หรือค่าเฉลี่ย (การถดถอย)

กระบวนการทำงานหลัก Random Forest ทำงานใน 4 ขั้นตอนสำคัญ(Hastie, 2009)

1. ดำเนินการสุ่มเลือกข้อมูลโดยใช้เทคนิค Bootstrap Sampling โดยจากชุดข้อมูลต้นฉบับที่มีจำนวน N ตัวอย่าง จะทำการสุ่มเลือกตัวอย่างจำนวน N เช่นกัน ด้วยวิธีการสุ่มแบบมีการทดแทน (Sampling with Replacement) เพื่อสร้างชุดข้อมูลย่อยสำหรับใช้ในการฝึกแต่ละต้นไม้ในโมเดล Random Forest

2. การสร้างต้นไม้ตัดสินใจที่มีความหลากหลาย โดยสุ่มเลือกคุณลักษณะ ในการสร้างต้นไม้แต่ละต้น จะสุ่มเลือกคุณลักษณะจำนวน m คุณลักษณะจากคุณลักษณะทั้งหมด M คุณลักษณะ (โดยทั่วไป $m \approx \sqrt{M}$) เพื่อใช้เป็นเกณฑ์ในการแบ่งแยกข้อมูลที่แต่ละโหนด

1) การเติบโตของต้นไม้ (Tree Growing) ต้นไม้แต่ละต้นจะเติบโตจนถึงขนาดสูงสุดโดยไม่มีการตัดแต่งกิ่ง (Pruning)
2) รวมผลการทำนาย ผลการทำนายจากต้นไม้ทั้งหมดถูกนำมารวมกันเพื่อให้ได้ผลลัพธ์สุดท้าย

Biau and Scornet (2016)อธิบายว่าความหลากหลายของต้นไม้ในป่าช่วยลดความแปรปรวน (Variance) ของโมเดลโดยรวม ขณะที่ความลึกของต้นไม้แต่ละต้นช่วยลดความเอนเอียง (Bias) ทำให้ Random Forest มีความสมดุลระหว่างความเอนเอียงและความแปรปรวนและลดปัญหา Overfitting ได้อย่างมีประสิทธิภาพ

2.1.6 Term Frequency-Inverse Document Frequency

TF-IDF เป็นเทคนิคหนึ่งในกระบวนการแปลงข้อความ (Text Transformation) ให้กลายเป็นค่าตัวเลข เพื่อให้โมเดลการเรียนรู้ของเครื่องสามารถนำไปวิเคราะห์ได้ โดยเทคนิคนี้จะคำนวณค่าน้ำหนัก (Weight) ของแต่ละคำในเอกสาร เพื่อบ่งบอกถึงความสำคัญของคำนั้น ๆ ในเอกสารชุดนั้น ๆ ประกอบด้วยสองส่วนหลัก ได้แก่

1. Term Frequency (TF) คือความถี่ของคำที่ปรากฏในเอกสารใดเอกสารหนึ่ง ซึ่งคำนวณได้จากจำนวนครั้งที่คำปรากฏในเอกสารนั้น หารด้วยจำนวนคำทั้งหมดในเอกสาร

2. Inverse Document Frequency (IDF) คือการวัดความสำคัญของคำ โดยพิจารณาจากจำนวนเอกสารทั้งหมดที่คำ ๆ นั้นปรากฏอยู่ คำที่ปรากฏในหลายเอกสารจะได้รับค่าน้ำหนักต่ำกว่า ในขณะที่คำที่ปรากฏในเอกสารจำนวนน้อยจะได้รับค่าน้ำหนักสูง

การคำนวณ TF-IDF สามารถแสดงเป็นสมการ (2),(3),(4)

$$TF_{(ij)} = \frac{f_{ij}}{n_j} \quad (2)$$

โดยที่

f_{ij} แทนจำนวนความถี่ของคำ (term) i ในข้อความ j

n_j แทนจำนวนคำทั้งหมดในข้อความ j

$$IDF_{(i)} = 1 + \log\frac{N}{c_i} \quad (3)$$

โดยที่

N แทนจำนวนข้อความทั้งหมด

c_i แทนจำนวนข้อความที่มีคำ i ปรากฏอยู่ในข้อความ

$$TF - IDF_{(ij)} = TF_{(ij)} \times IDF_{(i)} \quad (4)$$

โดยที่

$TF - IDF_{(ij)}$ แทนค่าน้ำหนักของคำที่แยกตามความสำคัญ

Yao et al. (2019) TF-IDF สามารถระบุคำสำคัญได้อย่างมีประสิทธิภาพโดยพิจารณาจากความถี่ของคำในเอกสารเทียบกับคลังข้อมูล แต่ขาดการพิจารณาความสัมพันธ์เชิงความหมาย ในขณะที่ TextRank วิเคราะห์ความสัมพันธ์ของคำในเนื้อหา แต่อาจพลาดคำสำคัญที่ไม่ปรากฏ

โดยการผสมผสานวิธีการเหล่านี้ นักวิจัยได้พัฒนาระบบสกัดคำสำคัญที่มีประสิทธิภาพสูงขึ้น โดยเฉพาะสำหรับเนื้อหา ซึ่งวิธีการผสมผสานนี้เริ่มต้นด้วยการใช้เทคนิค TF-IDF เพื่อระบุ

คำสำคัญเบื้องต้น โดยอาศัยความถี่ทางสถิติเป็นเกณฑ์ในการคัดเลือกคำ จากนั้นใช้ TextRank เพื่อปรับผลลัพธ์โดยพิจารณาความสัมพันธ์เชิงความหมาย

การประเมินประสิทธิภาพของวิธีการผสมผสานนี้ พบว่ามีค่าความแม่นยำ (Precision), ค่าระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F1-score) สูงกว่าเมื่อเปรียบเทียบกับการใช้วิธีใดวิธีหนึ่งเพียงอย่างเดียว ส่งผลให้วิธีการดังกล่าวมีคุณค่าและเหมาะสมอย่างยิ่งสำหรับการวิเคราะห์เนื้อหา

2.1.7 Word2Vec

Word2Vec เป็นเทคนิคการเรียนรู้เชิงลึก เพื่อสร้างการแทนค่าคำในรูปแบบเวกเตอร์ (Word Embeddings) ที่สามารถจับความสัมพันธ์ทางความหมายระหว่างคำได้ เทคนิคนี้ถูกนำเสนอครั้งแรกโดย Mikolov et al. (2013) ในงานวิจัยเรื่อง "Efficient Estimation of Word Representations in Vector Space" Word2Vec ประกอบด้วยสองโมเดลหลักดังนี้

1. CBOW ทำนายคำเป้าหมายจากบริบทโดยรอบ โดยใช้หน้าต่างบริบท (Context Window) ที่กำหนด โดยหน้าต่างบริบทจะรวมคำที่อยู่ก่อนหน้าและหลังคำ

2. Skip-gram ทำงานในทางตรงกันข้ามกับ CBOW โดยทำนายบริบทจากคำเป้าหมาย ซึ่งมักมีประสิทธิภาพดีกว่า CBOW เมื่อมีข้อมูลฝึกฝนจำนวนมาก

การประยุกต์ใช้ Word2Vec ในการวิเคราะห์ความรู้สึกมีข้อดีหลายประการ

1. การจับความสัมพันธ์ทางความหมาย Li and Yang (2018) แสดงให้เห็นในงานวิจัย "Word Embedding for Understanding Natural Language: A Survey" ว่า Word2Vec สามารถเรียนรู้ความสัมพันธ์ระหว่างคำที่มีความหมายใกล้เคียงกัน เช่น "ดี" กับ "เยี่ยม" หรือ "แย่" กับ "ห่วย" ซึ่งเป็นประโยชน์อย่างมากในการวิเคราะห์ความรู้สึก

2. การจัดการกับคำที่ไม่เคยเห็น Camacho-Collados and Pilehvar (2018) พบว่าใน "From Word to Sense Embeddings: A Survey on Vector Representations of Meaning" ว่า แม้ Word2Vec จะมีข้อจำกัดในการจัดการกับคำที่ไม่ปรากฏในชุดข้อมูลฝึกฝน แต่สามารถนำไปต่อยอดด้วยเทคนิคอื่นๆ เพื่อประมาณความหมายของคำใหม่จากบริบทได้

3. การลดมิติข้อมูล Almeida and Xexéo (2019) อธิบายในงานวิจัย "Word Embeddings: A Survey" ว่า Word2Vec เป็นเทคนิคที่ช่วยลดมิติของข้อมูล โดยแปลงคำในข้อความให้อยู่ในรูปแบบเวกเตอร์ที่มีมิติต่ำกว่าการแทนค่าด้วย One-hot Encoding ส่งผลให้การประมวลผลข้อมูลมีประสิทธิภาพและรวดเร็วยิ่งขึ้น

Tang et al. (2014) เสนอในงานวิจัย "Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification" ว่าการใช้ Word2Vec ที่ปรับปรุงให้เฉพาะเจาะจงกับงานวิเคราะห์ความรู้สึกสามารถเพิ่มความแม่นยำในการจำแนกความรู้สึกได้มากกว่าการใช้ Word2Vec แบบทั่วไป

Sarzynska-Wawer et al. (2021) รายงานในงานวิจัย "Detecting Formal Thought Disorder by Word Embeddings" ว่า Word2Vec ร่วมกับโมเดลต้นไม้ตัดสินใจ (Decision Tree) และ Support Vector Machine (SVM) ให้ผลลัพธ์ที่ดีในการวิเคราะห์ข้อความที่มีความซับซ้อน

2.1.8 Sentiment Analysis

การวิเคราะห์ความรู้สึก (Sentiment Analysis) เป็นกระบวนการวิเคราะห์ข้อความเพื่อระบุความคิดเห็น อารมณ์ และทัศนคติที่แฝงอยู่ในเนื้อหา โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ(Liu, 2022) การวิเคราะห์ความรู้สึกมีบทบาทสำคัญในการทำความเข้าใจความคิดเห็นของผู้ใช้บริการในหลากหลายบริบท รวมถึงการศึกษาระดับอุดมศึกษา(Altrabsheh et al., 2014) ในบริบทของสถาบันการศึกษา การวิเคราะห์ความรู้สึกช่วยให้ผู้บริหารและผู้มีส่วนเกี่ยวข้องสามารถเข้าใจความพึงพอใจและความต้องการของนักศึกษาได้อย่างลึกซึ้ง(Rani & Kumar, 2017) โดยทั่วไป การวิเคราะห์ความรู้สึกสามารถทำได้โดยใช้สองวิธีหลัก คือ

1. วิธีการใช้พจนานุกรมคำศัพท์ (Lexicon-based approach) เช่น VADER ซึ่งอาศัยการเปรียบเทียบคำในข้อความกับพจนานุกรมที่มีการกำหนดค่าความรู้สึกไว้ล่วงหน้า (Taboada et al., 2011)

2. วิธีการเรียนรู้ของเครื่อง (Machine learning approach) ประกอบด้วยอัลกอริทึมต่างๆ เช่น Naïve Bayes, SVM, และ Random Forest ในการเรียนรู้รูปแบบจากข้อมูลที่มีการกำหนดประเภทความรู้สึกไว้แล้ว(Medhat et al., 2014)

ในงานวิจัยเกี่ยวกับการวิเคราะห์ความรู้สึกในสถาบันการศึกษา ผู้วิจัยมักจะผสมผสานวิธีการทั้งสองเข้าด้วยกัน โดยเริ่มจากการใช้วิธี Lexicon-based อย่างเช่น VADER สำหรับการวิเคราะห์ขั้นต้น และต่อยอดด้วยการใช้เทคนิคการเรียนรู้ของเครื่องในการวิเคราะห์อย่างละเอียดเพื่อเพิ่มความแม่นยำและความครอบคลุมของผลลัพธ์(Alessia et al., 2015)

VADER (Valence Aware Dictionary and sEntiment Reasoner)

VADER เป็นเครื่องมือวิเคราะห์ความรู้สึกแบบ Lexicon-based ที่ออกแบบมาเฉพาะสำหรับการวิเคราะห์ข้อความในสื่อสังคมออนไลน์และความคิดเห็นทั่วไป(Hutto & Gilbert, 2014)

VADER ใช้พจนานุกรมคำที่มีการให้ค่าความเข้มข้นของความรู้สึก (Sentiment Intensity) ซึ่งมีการคำนวณคะแนนความรู้สึกในหลายมิติ ได้แก่ Positive, Negative, Neutral และ Compound

VADER มีลักษณะเด่นที่สำคัญคือ

1. ความไวต่อความเข้มข้นของความรู้สึก เช่น "ดี" กับ "ดีมาก" จะได้คะแนนความเป็นบวกที่แตกต่างกัน

2. การรองรับข้อความที่มีอารมณ์ผสม โดยให้คะแนนทั้งด้านบวกและด้านลบในข้อความเดียวกัน

3. ความสามารถในการตีความเครื่องหมายวรรคตอนที่เน้นอารมณ์ เช่น ! หรือ !!!

4. การรับรู้การเปลี่ยนแปลงความหมายจากคำที่เป็นตัวปรับ (Modifiers) เช่น "ไม่" "มาก" "น้อย"

การทำงานของ VADER ในการจำแนกความรู้สึกมักใช้ค่า Compound Score ซึ่งเป็นค่าที่รวมผลจากทุกมิติและถูกนอร์มัลไลซ์ให้อยู่ในช่วง -1 (ความรู้สึกเชิงลบมากที่สุด) ถึง +1 (ความรู้สึกเชิงบวกมากที่สุด) โดยทั่วไปจะมีการกำหนดเกณฑ์ดังนี้

- ความรู้สึกเชิงบวก: Compound Score $\geq$ 0.05
- ความรู้สึกเป็นกลาง: -0.05 < Compound Score < 0.05
- ความรู้สึกเชิงลบ: Compound Score $\leq$ -0.05

Mäntylä et al. (2018) พบว่าการปรับค่าเกณฑ์ threshold ที่เหมาะสมกับข้อมูลเป็นสิ่งสำคัญในการเพิ่มประสิทธิภาพของ VADER ในการวิเคราะห์ความรู้สึก โดยค่า 0.05 มักเป็นค่าที่ให้ความสมดุลระหว่างความแม่นยำและความครอบคลุมในการจำแนกความรู้สึก

Hutto and Gilbert (2014)พบว่า VADER มีความแม่นยำสูงถึง 82-84% ในชุดข้อมูลสื่อสังคมออนไลน์

เปรียบเทียบ VADER กับเครื่องมือวิเคราะห์ความรู้สึกอื่นๆ

การเลือกใช้เครื่องมือที่เหมาะสมสำหรับการวิเคราะห์ความรู้สึก รวมถึงการเปรียบเทียบประสิทธิภาพและการพิจารณาความเหมาะสมของเครื่องมือแต่ละชนิด ถือเป็นปัจจัยสำคัญ นอกจาก VADER แล้วยังมีเครื่องมือและเทคนิคอื่น ๆ ที่ได้รับความนิยมในการวิเคราะห์ความรู้สึก โดยแต่ละเครื่องมือนั้นมีข้อดี ข้อจำกัด และความเหมาะสมในการใช้งานที่แตกต่างกัน ดังนี้

1. TextBlob เป็นไลบรารี Python ที่ออกแบบมาเพื่อการประมวลผลภาษาธรรมชาติ โดยมีฟังก์ชันการวิเคราะห์ความรู้สึกเป็นหนึ่งในความสามารถหลัก มีข้อดีคือใช้งานง่าย มีความยืดหยุ่นและรองรับหลายภาษา อย่างไรก็ตาม มีข้อจำกัดในด้านความแม่นยำเมื่อเทียบกับ VADER

โดยเฉพาะเมื่อวิเคราะห์ข้อความในสื่อสังคมออนไลน์ Loria (2018) พบว่า TextBlob มีความแม่นยำประมาณ 65-70% ในการวิเคราะห์ความรู้สึกจากข้อความทั่วไป

2. BERT (Bidirectional Encoder Representations from Transformers) เป็นโมเดลการเรียนรู้เชิงลึกที่สามารถปรับแต่งเพื่อใช้ในการวิเคราะห์ความรู้สึกได้ มีข้อดีคือมีความแม่นยำสูงมาก สามารถเข้าใจบริบทและความหมายที่ซับซ้อนได้ดี Sun et al. (2019) รายงานความแม่นยำสูงถึง 93-95% ในบางชุดข้อมูลมาตรฐาน อย่างไรก็ตาม BERT ต้องการทรัพยากรการประมวลผลสูง มีความซับซ้อนในการใช้งาน และต้องการเวลาในการฝึกฝนและประมวลผล

สามารถสรุปการเปรียบเทียบเครื่องมือทั้งสามได้ในตาราง 1 แสดงการเปรียบเทียบ VADER, TextBlob และ BERT ในเกณฑ์การประเมินหลัก 6 ด้าน ซึ่งจะช่วยในการตัดสินใจเลือกเครื่องมือที่เหมาะสมกับงานวิจัยนี้

ตาราง 1 การเปรียบเทียบ VADER กับเครื่องมือวิเคราะห์ความรู้สึกอื่นๆ

เกณฑ์	VADER	TextBlob	BERT
ความแม่นยำ	82-84%	65-70%	93-95%
ความเร็ว	สูง	กลาง	ต่ำมาก
ความซับซ้อน	ต่ำ	ต่ำ	สูงมาก
ทรัพยากรที่ใช้	น้อย	น้อย	มาก
การจัดการสื่อสังคม	ดีมาก	ปานกลาง	ดี
ต้องการข้อมูลฝึกฝน	ไม่ต้อง	ไม่ต้อง	ไม่ต้อง

จากการเปรียบเทียบข้างต้น VADER มีข้อดีที่ทำให้เหมาะสมกับงานวิเคราะห์ความรู้สึก

1. ทำงานได้เร็วมากเนื่องจากเป็น rule-based system ไม่ต้องการการฝึกฝนโมเดล เหมาะสำหรับการประมวลผลข้อมูลจำนวนมากในเวลาจำกัด

2. ออกแบบมาเฉพาะสำหรับข้อความที่มีลักษณะไม่เป็นทางการ (informal tone) ที่พบในสื่อสังคมออนไลน์ เช่น การใช้อิโมจิ เครื่องหมายวรรคตอนเพื่อเน้นอารมณ์ คำย่อ และการเขียนที่ไม่เป็นทางการ

3. เป็น lexicon-based approach ที่ไม่ต้องการชุดข้อมูลที่มีการติดฉลาก (labeled data) สำหรับการฝึกฝน

4. ผลลัพธ์สามารถตรวจสอบและอธิบายได้

5. ให้ความแม่นยำที่ยอมรับได้แต่ใช้ทรัพยากรน้อย

อย่างไรก็ตาม VADER มีข้อจำกัดสำคัญที่ต้องพิจารณาดังนี้

1. ข้อจำกัดด้านภาษา VADER ถูกพัฒนาสำหรับภาษาอังกฤษเป็นหลัก การใช้งานกับภาษาไทยหรือภาษาอื่นๆ อาจให้ผลลัพธ์ที่ไม่แม่นยำเท่าที่ควร เนื่องจากพจนานุกรมคำและกฎการวิเคราะห์ถูกออกแบบสำหรับภาษาอังกฤษโดยเฉพาะ

2. การจัดการกับข้อความที่ซับซ้อน VADER อาจมีปัญหาในการวิเคราะห์ข้อความที่มีโครงสร้างซับซ้อน การใช้ภาษาเปรียบเทียบ หรือความหมายที่ต้องอาศัยบริบทมาก เช่น การพูดแดกดัน (sarcasm) หรือการใช้ภาษาเชิงสัญลักษณ์

3. ประสิทธิภาพขึ้นอยู่กับความครอบคลุมและความถูกต้องของพจนานุกรม หากมีคำใหม่หรือคำสแลงที่ไม่อยู่ในพจนานุกรม อาจทำให้การวิเคราะห์ไม่แม่นยำ

งานวิจัยนี้เลือกใช้ VADER เป็นเครื่องมือหลักในการวิเคราะห์ความรู้สึก ด้วยเหตุผลสำคัญดังนี้

1. ความเหมาะสมกับลักษณะข้อมูล VADER ถูกออกแบบมาเฉพาะสำหรับการวิเคราะห์ข้อความในสื่อสังคมออนไลน์และข้อความสั้นๆ ซึ่งมีลักษณะคล้ายคลึงกับความคิดเห็นของนักศึกษาที่นำมาวิเคราะห์ในงานวิจัยนี้

2. ความสมดุลระหว่างประสิทธิภาพและความซับซ้อน VADER มีความแม่นยำซึ่งใกล้เคียงกับโมเดลที่ซับซ้อนกว่าอย่าง BERT แต่ใช้ทรัพยากรน้อยกว่า

3. ความไวต่อความเข้มข้นของความรู้สึก VADER สามารถจับความเข้มข้นของความรู้สึกได้ดี เช่น การใช้ตัวอักษรพิมพ์ใหญ่ เครื่องหมายวรรคตอนที่เน้นอารมณ์ หรือคำขยายที่เพิ่มความเข้มข้น ซึ่งพบได้บ่อยในความคิดเห็นของนักศึกษา

4. ไม่ต้องการข้อมูลฝึกฝน เนื่องจาก VADER เป็นวิธีแบบ Lexicon-based จึงไม่ต้องการข้อมูลฝึกฝน ซึ่งเป็นข้อได้เปรียบเมื่อมีข้อมูลที่ติดฉลากจำนวนจำกัด

5. การยืนยันประสิทธิภาพจากงานวิจัย Ribeiro et al. (2016) ในงานวิจัย "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods" ทำการเปรียบเทียบเครื่องมือวิเคราะห์ความรู้สึกอื่นๆ พบว่า VADER มีประสิทธิภาพดีกว่า TextBlob ในการวิเคราะห์ข้อความสั้นๆ โดยเปรียบเทียบจากความแม่นยำ

ด้วยเหตุผลเหล่านี้ VADER จึงเหมาะสมที่จะเป็นเครื่องมือสำหรับการวิเคราะห์ความรู้สึกจากความคิดเห็นของนักศึกษาในบริบทของการให้บริการของมหาวิทยาลัย ซึ่งมักประกอบด้วยข้อความสั้นๆ ที่มีการแสดงความรู้สึกที่หลากหลาย และต้องการการวิเคราะห์ที่รวดเร็วและแม่นยำ

2.1.9 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) เป็นเทคนิคหนึ่งในการทำการจัดกลุ่มหัวข้อ (Topic Modeling) ที่ใช้สำหรับการวิเคราะห์เอกสารจำนวนมากเพื่อค้นหาหัวข้อที่แฝงอยู่ในเอกสารนั้น ๆ LDA เป็นโมเดลแบบ Bayesian ที่มีการคำนวณจากความน่าจะเป็นเพื่อระบุหัวข้อและการกระจายตัวของคำในแต่ละหัวข้อ LDA ถูกพัฒนาขึ้นโดย Blei, Ng, และ Jordan ในปี 2003 และได้กลายเป็นหนึ่งในเทคนิคที่นิยมใช้ในการวิเคราะห์หัวข้อจากข้อมูลข้อความขนาดใหญ่ (Blei et al., 2003) ขั้นตอนการวิเคราะห์ด้วย LDA ประกอบด้วย

1. กำหนดจำนวนหัวข้อที่ต้องการค้นหา (Number of Topics) นี่เป็นพารามิเตอร์สำคัญที่ต้องกำหนดก่อนการวิเคราะห์ โดยอาจใช้เทคนิคต่าง ๆ เช่น การวัดค่า Coherence Score เพื่อหาจำนวนหัวข้อที่เหมาะสม(Röder et al., 2015)

2. ทำการสุ่มกำหนดหัวข้อให้กับคำในเอกสารเริ่มต้น

3. ปรับการกำหนดหัวข้อซ้ำ ๆ โดยใช้วิธีการสุ่มตัวอย่างแบบ Gibbs หรือการประมาณค่าแบบ Variational Inference จนกว่าการกระจายตัวของหัวข้อและคำจะมีความเสถียร(Blei et al., 2003; Griffiths & Steyvers, 2004)

การเปรียบเทียบ LDA กับ Clustering Algorithms อื่นๆ ในการค้นหาหัวข้อและกลุ่มความคิดเห็นจากข้อความ นอกจาก LDA แล้ว ยังมีอัลกอริทึมการจัดกลุ่ม (Clustering Algorithms) อื่นๆ ที่สามารถใช้ได้ การเปรียบเทียบระหว่างเทคนิคเหล่านี้ช่วยให้เข้าใจเหตุผลในการเลือกใช้ LDA สำหรับงานวิจัยนี้

1. K-Means Clustering เป็นอัลกอริทึมจัดกลุ่มแบบพาร์ทิชัน (Partitional Clustering) ที่แบ่งข้อมูลออกเป็น K กลุ่มโดยพยายามลดระยะห่างภายในกลุ่มให้น้อยที่สุด (MacQueen, 1967) ในบริบทของการวิเคราะห์ข้อความ K-Means มักถูกนำมาใช้กับข้อมูลที่แปลงเป็นเวกเตอร์ด้วยเทคนิค TF-IDF หรือ Word Embeddings

หลักการทำงานของ K-Means เริ่มต้นจากการกำหนดจำนวนกลุ่ม K ที่ต้องการล่วงหน้า จากนั้นจะสุ่มเลือกจุดศูนย์กลาง (Centroids) จำนวน K จุด แล้วจัดแต่ละตัวอย่างให้อยู่ในกลุ่มที่มีจุดศูนย์กลางใกล้ที่สุด ระบบจะคำนวณจุดศูนย์กลางใหม่สำหรับแต่ละกลุ่มและทำซ้ำกระบวนการนี้จนกว่าจุดศูนย์กลางจะไม่มีการเปลี่ยนแปลงหรือถึงจำนวนรอบที่กำหนด

ข้อดี Zhao and Karypis (2004)พบว่า K-Means มีประสิทธิภาพดีในการจัดกลุ่มเอกสารที่มีลักษณะคล้ายคลึงกัน และเกิดความซับซ้อนเชิงเวลาเป็น O(nkdi) เมื่อ n คือจำนวนตัวอย่าง, k

คือจำนวนกลุ่ม, d คือมิติของข้อมูล และ i คือจำนวนรอบ นอกจากนี้ยังง่ายต่อการใช้งานและทำความเข้าใจ รวมถึงมีประสิทธิภาพสูงสำหรับชุดข้อมูลขนาดใหญ่

ข้อเสีย Steinley (2006) ชี้ให้เห็นว่า K-Means มีความไวต่อการเลือกจุดศูนย์กลางเริ่มต้นและอาจติดกับดักค่าต่ำสุดเฉพาะที่ (Local Minima) อีกทั้งต้องกำหนดจำนวนกลุ่ม K ล่วงหน้า ซึ่งอาจเป็นเรื่องยากในการเลือกค่าที่เหมาะสม และไม่เหมาะกับข้อมูลที่มีรูปร่างซับซ้อนหรือมีความหนาแน่นไม่สม่ำเสมอ ในการวิเคราะห์ข้อความ K-Means ไม่สามารถจัดการกับความหลากหลายของบริบทได้ดีเท่า LDA

2. DBSCAN เป็นอัลกอริทึมการจัดกลุ่มข้อมูลแบบอิงความหนาแน่น (Density-based Clustering) ซึ่งสามารถค้นหากลุ่มข้อมูลที่มีรูปร่างหลากหลายและไม่แน่นอน รวมถึงสามารถแยกแยะข้อมูลรบกวน (Noise) ออกจากกลุ่มข้อมูลได้อย่างมีประสิทธิภาพ (Ester et al., 1996)

หลักการทำงานของ DBSCAN เริ่มต้นจากการกำหนดพารามิเตอร์สองตัวคือ **ε** (รัศมีของพื้นที่ใกล้เคียง) และ MinPts (จำนวนตัวอย่างขั้นต่ำที่ต้องมีในพื้นที่ใกล้เคียงเพื่อจัดเป็นจุดแกนหลัก) จากนั้นระบบจะระบุจุดแกนหลัก (Core Points) ซึ่งเป็นจุดที่มีจำนวนเพื่อนบ้านในระยะ **ε** มากกว่าหรือเท่ากับ MinPts ระบบจะเชื่อมจุดแกนหลักที่อยู่ในระยะ **ε** ของกันและกันเข้าด้วยกัน ส่วนจุดขอบ (Border Points) คือจุดที่ไม่ใช่จุดแกนหลักแต่อยู่ในระยะ **ε** ของจุดแกนหลัก ขณะที่จุดรบกวน (Noise Points) คือจุดที่ไม่ใช่ทั้งจุดแกนหลักและจุดขอบ

ข้อดี Ester et al. (1996) เสนอว่า DBSCAN สามารถค้นหากลุ่มข้อมูลที่มีรูปร่างไม่แน่นอนได้ดี ซึ่งเป็นประโยชน์ในการวิเคราะห์ข้อความที่มีความหลากหลาย อีกทั้งไม่ต้องกำหนดจำนวนกลุ่มล่วงหน้า สามารถตรวจจับข้อมูลรบกวนได้ และรองรับกลุ่มข้อมูลที่มีขนาดและความหนาแน่นไม่สม่ำเสมอ

ข้อเสีย Shah et al. (2012) ชี้ให้เห็นว่า DBSCAN มีความไวต่อการเลือกค่าพารามิเตอร์ **ε** และ MinPts โดยประสิทธิภาพจะลดลงเมื่อมิติของข้อมูลสูงขึ้น (Curse of Dimensionality) และข้อมูลที่มีความหนาแน่นแตกต่างกันมาก ในการวิเคราะห์ข้อความ DBSCAN ไม่สามารถจับความสัมพันธ์ทางความหมายในระดับลึกได้ดีเท่า LDA

การเปรียบเทียบประสิทธิภาพระหว่าง LDA, K-Means และ DBSCAN

จากการทดลองเปรียบเทียบประสิทธิภาพการวิเคราะห์หัวข้อจากความคิดเห็นของนักศึกษา พบความแตกต่างที่สำคัญดังนี้

1. ความสามารถในการระบุหัวข้อ

LDA ให้ผลลัพธ์ที่มีความหมายทางความรู้สึกมากกว่า โดยสามารถระบุหัวข้อที่สอดคล้องกับความเข้าใจของมนุษย์ได้ดีกว่า โดยเฉพาะในข้อมูลที่มีความหลากหลายของหัวข้อ

K-Means สามารถจัดกลุ่มเอกสารที่คล้ายคลึงกันได้ดี แต่ไม่สามารถแสดงความสัมพันธ์ระหว่างคำในหัวข้อเดียวกันได้ชัดเจนเท่า LDA

DBSCAN มีประสิทธิภาพในการค้นหากลุ่มความคิดเห็นที่มีลักษณะพิเศษ แต่มักมีปัญหาในการตีความหัวข้อที่พบ

2. การจัดการกับข้อความที่มีหลายหัวข้อ

LDA สามารถจัดการกับข้อความที่มีหลายหัวข้อได้ดี โดยแต่ละข้อความสามารถมีความน่าจะเป็นของการเป็นสมาชิกในหลายหัวข้อได้พร้อมกันส่วน K-Means และ DBSCAN จัดแต่ละข้อความให้อยู่ในกลุ่มเดียวเท่านั้น ซึ่งไม่สอดคล้องกับธรรมชาติของข้อความที่มักมีหลายประเด็น

3. ความแม่นยำในการจำแนกหัวข้อ

จากการทดลองของ Hong and Davison (2010) ในบทความ "Empirical Study of Topic Modeling in Twitter" พบว่า LDA มีค่า Topic Coherence Score เฉลี่ย 0.42 ซึ่งสูงกว่า K-Means (0.35) และ DBSCAN (0.31)

การเปรียบเทียบการค้นหาหัวข้อระหว่าง LDA กับ K-Means

LDA ถูกออกแบบมาโดยเฉพาะสำหรับการค้นหาหัวข้อในข้อความ โดยทำงานในเชิงความน่าจะเป็นที่สมมติว่าแต่ละเอกสารเป็นการผสมผสานของหลายหัวข้อ และแต่ละหัวข้อเป็นการกระจายความน่าจะเป็นของคำต่างๆ ลักษณะสำคัญของ LDA ในการค้นหาหัวข้อมีดังนี้

1. สร้างการกระจายความน่าจะเป็นของคำในแต่ละหัวข้อ ทำให้เราสามารถเห็นว่าคำใดมีความสำคัญมากที่สุดในหัวข้อนั้น เช่น หัวข้อ "การเรียนการสอน" อาจมี คำเกี่ยวกับ การสอน (0.15), อาจารย์ (0.12), หลักสูตร (0.10), ห้องเรียน (0.08), การเรียน (0.07)

2. เอกสารมีความน่าจะเป็นของการเป็นสมาชิกในแต่ละหัวข้อ เช่น เอกสารหนึ่งอาจเป็นสมาชิก หัวข้อ "การเรียนการสอน" 60%, หัวข้อ "สิ่งอำนวยความสะดวก" 30%, และ หัวข้อ "บริการนักศึกษา 10%"

3. การจัดการกับความซับซ้อนของภาษาธรรมชาติ LDA รองรับการที่คำเดียวกันสามารถมีความหมายต่างกันในบริบทต่างกัน และเอกสารหนึ่งสามารถมีได้หลายหัวข้อได้พร้อมกัน

K-Means ใช้แนวทางการแปลงข้อมูลข้อความให้เป็นเวกเตอร์ในหลายมิติ แล้วทำการจัดกลุ่มโดยใช้หลักการระยะห่างทางคณิตศาสตร์ โดยไม่ได้มุ่งเน้นที่การค้นหาหัวข้อเป็นหลัก

1. K-Means สร้างเพียง centroids ที่เป็นเวกเตอร์เฉลี่ยของกลุ่ม แต่ไม่สามารถแปลกลับเป็นคำที่มีความหมายได้อย่างชัดเจน เช่น centroid อาจมีค่า [0.23, 0.45, 0.12, ...] ในมิติต่างๆ แต่ไม่สามารถบอกได้ว่าหมายถึงหัวข้ออะไร

2. เอกสารถูกจัดให้อยู่ในกลุ่มเดียวเท่านั้น ไม่สามารถผสมผสานของหัวข้อได้

3. ประสิทธิภาพของ K-Means ขึ้นอยู่กับคุณภาพของการแปลงข้อความเป็นเวกเตอร์ และการเลือก distance metric ที่เหมาะสม

จากการเปรียบเทียบข้างต้น LDA จึงเป็นเทคนิคที่เหมาะสมสำหรับการวิเคราะห์หัวข้อจากความคิดเห็นของนักศึกษา เนื่องจากสามารถจัดการกับความซับซ้อนของภาษาธรรมชาติ รองรับข้อความที่มีหลายหัวข้อ และให้ผลลัพธ์ที่สอดคล้องกับการตีความของมนุษย์ได้ดีที่สุด

2.1.10 Imbalanced Data

ในงานวิเคราะห์ความรู้สึก ปัญหาข้อมูลไม่สมดุล (Imbalanced Data) เป็นความท้าทายที่พบได้บ่อย เนื่องจากข้อมูลความคิดเห็นมักมีการกระจายตัวไม่สม่ำเสมอระหว่างประเภทความรู้สึก(Krawczyk, 2016) ตัวอย่างเช่น ในความคิดเห็นของนักศึกษา อาจพบความคิดเห็นเชิงบวกมากกว่าเชิงลบอย่างมีนัยสำคัญ หรือในทางกลับกัน ข้อมูลไม่สมดุลส่งผลให้โมเดลมีแนวโน้มที่จะทำนายคลาสที่มีจำนวนมากกว่า (Majority Class) ได้ดีกว่า แต่ทำนายคลาสที่มีจำนวนน้อยกว่า (Minority Class) ได้แย่ลง (He & Garcia, 2009) เทคนิคการจัดการกับข้อมูลที่ไม่สมดุลที่นิยมใช้ในการวิเคราะห์ความรู้สึกมีดังนี้

1. การสุ่มเพิ่ม (Oversampling) การเพิ่มจำนวนตัวอย่างในคลาสที่มีข้อมูลน้อย วิธีที่นิยม คือ SMOTE (Synthetic Minority Over-sampling Technique) ซึ่งสร้างตัวอย่างใหม่ระหว่างตัวอย่างที่มีอยู่และเพื่อนบ้านใกล้ที่สุด(Chawla et al., 2002) ในขณะที่ Fernández et al. (2018)พบว่า SMOTE เพิ่มประสิทธิภาพการทำนายคลาสส่วนน้อยได้อย่างมีนัยสำคัญ โดยเฉพาะเมื่อใช้ร่วมกับ SVM และ Random Forest ทำให้ค่า F1-score เพิ่มขึ้น 8-12%

2. การสุ่มลด (Undersampling) การลดจำนวนตัวอย่างในคลาสที่มีข้อมูลมาก วิธีที่นิยม คือ Random Under Sampling (RUS) ซึ่งสุ่มลดจำนวนตัวอย่างในคลาสส่วนมากให้ใกล้เคียงกับคลาสส่วนน้อย อย่างไรก็ตาม Sarakit et al. (2015) แสดงให้เห็นว่า กับข้อมูลจำนวนมาก Undersampling ช่วยลดเวลาฝึกฝนโมเดลได้ดี โดยความแม่นยำลดลงเพียงเล็กน้อย

3. การใช้น้ำหนักคลาส (Class Weighting) การใช้น้ำหนักคลาสเป็นการกำหนดน้ำหนักให้กับแต่ละคลาสในระหว่างการฝึกฝนโมเดล โดยให้น้ำหนักมากกว่ากับคลาสที่มีจำนวนน้อยกว่าเพื่อให้โมเดลให้ความสำคัญกับการทำนายคลาสนั้นมากขึ้นKhoshgoftaar et al. (2007) แนะนำให้ใช้ Class Weighting ในกรณีที่มีข้อจำกัดด้านทรัพยากรคอมพิวเตอร์หรือเมื่อต้องการรักษาข้อมูลทั้งหมดไว้โดยไม่ต้องสร้างตัวอย่างเพิ่มหรือลดจำนวนตัวอย่าง ผลการศึกษาพบว่า การใช้ Class Weighting ร่วมกับ SVM ในการวิเคราะห์ความรู้สึกให้ผลลัพธ์ที่ใกล้เคียงกับการใช้ SMOTE แต่ใช้เวลาและทรัพยากรในการประมวลผลน้อยกว่า

4. การใช้เทคนิคผสม (Hybrid Approaches) เป็นการรวมหลายเทคนิคเข้าด้วยกัน เช่น การใช้ SMOTE ร่วมกับการคัดกรองตัวอย่างที่ไม่เหมาะสม (Tomek links) หรือการใช้ SMOTE ร่วมกับ Class Weighting Martinez นอกจากนี้ Galar et al. (2011)ยังได้นำเสนอการใช้เทคนิค Ensemble Learning ร่วมกับการจัดการข้อมูลไม่สมดุล ซึ่งพบว่ามีประสิทธิภาพสูงในการจำแนกข้อความที่มีความไม่สมดุลของคลาส โดยการสร้างโมเดลย่อยหลายๆ โมเดลที่เรียนรู้จากชุดข้อมูลที่ผ่านการปรับสมดุลด้วยเทคนิคต่างๆ แล้วนำผลลัพธ์มารวมกัน

ตามคำแนะนำของ López et al. (2013)ไม่มีเทคนิคใดเหมาะกับทุกสถานการณ์ การทดลองเปรียบเทียบกับชุดข้อมูลเฉพาะและประเมินจากหลายมิติจึงเป็นแนวทางที่ดีที่สุดในการเลือกเทคนิคที่เหมาะสม

2.1.11 University Service

การให้บริการของมหาวิทยาลัย หมายถึงบริการทุกประเภทที่มหาวิทยาลัยจัดเตรียมเพื่อสนับสนุนการเรียนรู้ การพัฒนา และประสบการณ์โดยรวมของนักศึกษา ครอบคลุมทั้งด้านวิชาการและไม่ใช่วิชาการ สามารถแบ่งได้เป็น 3 ประเภทหลัก ได้แก่

1. บริการด้านวิชาการ ได้แก่ การเรียนการสอน การให้คำปรึกษา และการเข้าถึงทรัพยากรการเรียนรู้ โดย Wen et al. (2014) พบว่านักศึกษาให้ความสำคัญกับคุณภาพการสอนอย่างมาก และมักแสดงความรู้สึกเชิงบวกต่อการให้คำปรึกษา ขณะที่มีความรู้สึกเชิงลบต่อกระบวนการวัดและประเมินผล

2. บริการสนับสนุนนักศึกษา ครอบคลุมการให้คำปรึกษาส่วนบุคคล การแนะแนวอาชีพ บริการด้านสุขภาพ และการสนับสนุนด้านการเงิน โดย Solinas et al. (2012) พบว่าการสนับสนุนนักศึกษาเป็นปัจจัยที่ส่งผลต่อความพึงพอใจโดยรวม

3. สิ่งอำนวยความสะดวกและโครงสร้างพื้นฐาน Khaiser et al. (2023) พบว่านักศึกษาให้ความสำคัญกับเทคโนโลยีรวมถึงสิ่งอำนวยความสะดวกต่างๆ เช่น ห้องสมุด และพื้นที่การเรียนรู้ที่ทันสมัย

คุณภาพการให้บริการมีความสัมพันธ์โดยตรงกับความพึงพอใจและความจงรักภักดีของนักศึกษา ซึ่งส่งผลต่อภาพลักษณ์และความสำเร็จของสถาบันในระยะยาว

2.1.12 การวัดประสิทธิภาพโมเดล

การวัดประสิทธิภาพของโมเดลการวิเคราะห์ความรู้สึกและการวิเคราะห์หัวข้อเป็นขั้นตอนสำคัญในการพัฒนาระบบวิเคราะห์ข้อความอัตโนมัติ โดยมีเกณฑ์การวัดประสิทธิภาพที่สำคัญดังนี้(Hossin & Sulaiman, 2015)

1. Accuracy คือการวัดความถูกต้องโดยรวมของผลการทำนาย โดยคำนวณจากสัดส่วนของจำนวนตัวอย่างที่ทำนายถูกต้องต่อจำนวนตัวอย่างทั้งหมด สามารถคำนวณได้จากสมการ (5)

$$\text{Accuracy} = \frac{TN+TP}{TP+TN+FP+FN} \quad (5)$$

โดยที่

TP (True Positive) คือจำนวนตัวอย่างเชิงบวกที่ทำนายถูกต้อง

TN (True Negative) คือจำนวนตัวอย่างเชิงลบที่ทำนายถูกต้อง

FP (False Positive) คือจำนวนตัวอย่างเชิงลบที่ทำนายว่าเป็นเชิงบวก

FN (False Negative) คือจำนวนตัวอย่างเชิงบวกที่ทำนายว่าเป็นเชิงลบ

2. Precision คือการวัดความแม่นยำของการทำนายในกลุ่มตัวอย่างที่ถูกจัดอยู่ในประเภทเชิงบวก โดยคำนวณจากสัดส่วนของจำนวนตัวอย่างที่ทำนายว่าเป็นบวกและเป็นค่าจริง (True Positive) ต่อจำนวนตัวอย่างทั้งหมดที่ทำนายว่าเป็นบวก (ทั้งที่ถูกและผิด) สามารถคำนวณได้จากสมการ (6)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

3. Recall คือการวัดความครบถ้วนของการทำนายเชิงบวก โดยคำนวณจากสัดส่วนของจำนวนตัวอย่างที่ทำนายว่าเป็นบวกและเป็นบวกจริง (True Positive) ต่อจำนวนตัวอย่างที่เป็นบวกทั้งหมดในความเป็นจริง (ทั้งที่ทำนายถูกและผิด) สามารถคำนวณได้จากสมการ (7)

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

4. F1- Score คือค่าความแม่นยำเชิงรวม (Harmonic Mean) ระหว่างค่า Precision และ Recall โดยให้ความสำคัญกับทั้งสองค่าในระดับที่เท่าเทียมกัน เหมาะสำหรับกรณีที่ต้องการสมดุลระหว่างความแม่นยำ (Precision) และความครบถ้วน (Recall) สามารถคำนวณได้จากสมการ (8)

$$\text{F1} - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

5. Confusion Matrix คือตารางที่แสดงจำนวนการทำนายในแต่ละกรณี (TP, TN, FP, FN) ซึ่งช่วยให้เห็นภาพรวมของประสิทธิภาพโมเดลและจุดที่โมเดลทำนายผิดพลาด

การวัดประสิทธิภาพสำหรับการวิเคราะห์หัวข้อด้วย LDA

1. Perplexity คือเกณฑ์ที่วัดความสามารถของโมเดลในการ "คาดเดา" หรือทำนายคำที่จะปรากฏในข้อความใหม่ที่โมเดลไม่เคยเห็นมาก่อน(Blei et al., 2003; Wallach et al., 2009) Perplexity สามารถเปรียบเทียบได้กับการที่เราให้คนอ่านข้อความครึ่งหนึ่ง แล้วให้เดาว่าคำต่อไปจะเป็นอะไร ยิ่งคาดเดาได้แม่นยำเท่าไหร่ แสดงว่าเข้าใจรูปแบบของภาษาได้ดีเท่านั้น ความหมายของ Perplexity ในการวัดการคาดเดาคำ Wallach et al. (2009) อธิบายว่า Perplexity สามารถมองได้เป็น "จำนวนคำที่โมเดลกำลังลังเลอยู่ในการเลือก" เมื่อทำนายคำถัดไป ยกตัวอย่างเช่น หากค่า Perplexity เท่ากับ 100 หมายความว่าโมเดลกำลังลังเลระหว่างคำประมาณ 100 คำในการเลือกคำถัดไป ซึ่งแสดงถึงความไม่แน่ใจของโมเดลในการทำนาย การตีความค่า Perplexity มีหลักเกณฑ์ดังนี้

1) ค่าต่ำ (เช่น 50-200): โมเดลสามารถคาดเดาคำได้ดี แสดงว่าโมเดลเข้าใจโครงสร้างและรูปแบบของข้อความได้ดี มีความไม่แน่นใจน้อยในการเลือกคำ
2) ค่าสูง (เช่น 1000+): โมเดลคาดเดาคำได้แย่ แสดงว่าโมเดลยังไม่เข้าใจรูปแบบของข้อความเท่าที่ควร มีความไม่แน่นใจสูงในการเลือกคำ

Perplexity คำนวณจากความน่าจะเป็นที่โมเดลจะสร้างข้อความในชุดข้อมูลทดสอบ สามารถคำนวณได้จากสมการ (9)

$$\text{Perplexity}(D_{\text{test}}) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right) \quad (9)$$

โดยที่

D_{test} = ชุดข้อมูลทดสอบ

M = จำนวนเอกสารในชุดทดสอบ

w_d = เอกสารที่ d

N_d = จำนวนคำในเอกสาร d

$p(w_d)$ = ความน่าจะเป็นของเอกสาร d ตามโมเดล LDA

Wallach et al. (2009)ในบทความ "Evaluation methods for topic models" ได้พัฒนาและปรับปรุงวิธีการคำนวณ Perplexity เพื่อประเมินประสิทธิภาพของโมเดล Topic Model โดยเฉพาะ LDA ค่า Perplexity ที่ต่ำกว่าบ่งชี้ว่าโมเดลสามารถคาดการณ์รูปแบบของภาษาได้ดีกว่า และมีความสามารถในการจำแนกหัวข้อที่ชัดเจนขึ้น

อย่างไรก็ตาม Chang et al. (2009)ในงานวิจัย "Reading Tea Leaves: How Humans Interpret Topic Models" ได้ทำการศึกษาที่สำคัญโดยเปรียบเทียบผลการประเมินจาก Perplexity กับการประเมินโดยมนุษย์ พบว่า Perplexity ไม่ได้สะท้อนความสามารถของมนุษย์ในการตีความหัวข้อเสมอไป เนื่องจาก Perplexity วัดความสามารถทางสถิติในการทำนายคำ แต่ไม่ได้วัดความมีความหมายหรือความสอดคล้องของหัวข้อที่มนุษย์รับรู้ได้

ตัวอย่างเช่น โมเดลที่มีค่า Perplexity ต่ำอาจสร้างหัวข้อที่มีคำที่เกี่ยวข้องกันทางสถิติ แต่ไม่มีความหมายที่สอดคล้องกันในทางความคิด เช่น หัวข้อที่ประกอบด้วยคำ "นักศึกษา", "คอมพิวเตอร์", "อาหาร", "กีฬา" ซึ่งแม้จะปรากฏร่วมกันบ่อยในข้อมูล แต่ไม่สามารถตีความเป็นหัวข้อที่มีความหมายได้

ดังนั้นการใช้ Coherence Score ร่วมกับ Perplexity จึงให้ภาพที่ครอบคลุมมากขึ้นในการประเมินประสิทธิภาพของโมเดล LDA โดย Perplexity วัดความสามารถทางสถิติในการทำนาย ขณะที่ Coherence Score วัดความมีความหมายและความสอดคล้องของหัวข้อที่มนุษย์สามารถตีความได้

1. Coherence Score คือเกณฑ์ที่วัดความสอดคล้องภายในของหัวข้อที่ค้นพบ โดยพิจารณาจากความสัมพันธ์ทางความหมายระหว่างคำในหัวข้อเดียวกัน(Mimno et al., 2011; Röder et al., 2015) ซึ่งแตกต่างจาก Perplexity ที่วัดความสามารถทางสถิติในการทำนายคำ Coherence Score มุ่งเน้นการวัดว่าหัวข้อที่ค้นพบมีความมีความหมายและสามารถตีความได้สำหรับมนุษย์ โดยประเภทของ Coherence Metrics ที่ใช้ในการประเมินคุณภาพหัวข้อ

Coherence Score มีหลายประเภทที่ใช้ในการประเมินคุณภาพของหัวข้อที่ค้นพบจาก LDA โดยแต่ละประเภทมีวิธีการคำนวณและข้อดีข้อเสียที่แตกต่างกัน (Röder et al., 2015) เมื่อเปรียบเทียบกับการประเมินโดยมนุษย์ พบว่า metrics ต่างๆ มีความสัมพันธ์ที่แตกต่างกันดังนี้

1) C_v (Coherence based on normalized pointwise mutual information and cosine similarity) เป็น metric ที่นิยมใช้มากที่สุดในปัจจุบัน โดยใช้ NPMI (Normalized Pointwise Mutual Information) ร่วมกับ cosine similarity ในการวัดความสัมพันธ์ระหว่างคำ ค่า C_v มีช่วงตั้งแต่ 0 ถึง 1 โดยค่าที่สูงกว่าแสดงถึงความสอดคล้องที่ดีกว่า C_v ได้รับการพัฒนาเพื่อให้สอดคล้องกับการประเมินของมนุษย์มากที่สุด และมักใช้เป็นมาตรฐานในการเปรียบเทียบประสิทธิภาพของโมเดล Topic Modeling
2) UMass Coherence (C_UMass) เป็น metric ที่ใช้การคำนวณ co-occurrence ของคำโดยตรงจากเอกสารในชุดข้อมูลเดียวกัน โดยพิจารณาความน่าจะเป็นร่วมของคำที่ปรากฏในเอกสารเดียวกัน พบว่า UMass มีความสัมพันธ์ที่ดีกับการตีความของมนุษย์ในบางกรณี โดยเฉพาะเมื่อข้อมูลมีขนาดใหญ่และมีความหลากหลายของหัวข้อ UMass มีข้อดีคือใช้เฉพาะข้อมูลภายในชุดข้อมูลต้นฉบับ ไม่ต้องอาศัย external corpus
3) UCI Coherence (C_UCI) ใช้ external corpus ตัวอย่างเช่น Wikipedia ในการคำนวณความสัมพันธ์ระหว่างคำ ทำให้สามารถจับความสัมพันธ์ที่อาจไม่ปรากฏในชุดข้อมูลต้นฉบับได้ UCI Coherence มีประโยชน์เมื่อชุดข้อมูลมีขนาดเล็กหรือเฉพาะเจาะจง เนื่องจากสามารถใช้ความรู้จาก external corpus ที่กว้างขวางกว่ามาช่วยในการประเมิน
4) NPMI (Normalized Pointwise Mutual Information) เป็น metric พื้นฐานที่หลาย coherence measures อื่นๆ ใช้เป็นส่วนประกอบ NPMI วัดความสัมพันธ์ระหว่างคำ

สองคำโดยพิจารณาจากความน่าจะเป็นที่คำทั้งสองจะปรากฏร่วมกันเทียบกับความน่าจะเป็นที่จะปรากฏแยกกัน

ดังนั้นงานวิจัยนี้จึงเลือกใช้ C_v เป็นหลักในการประเมินคุณภาพของหัวข้อ เนื่องจากมีความสัมพันธ์สูงสุดกับการประเมินของมนุษย์ และเป็นมาตรฐานที่ใช้กันอย่างแพร่หลายในงานวิจัย Topic Modeling

การตีความค่า Coherence Score ที่สูงแสดงว่าคำในหัวข้อนั้นมีความเกี่ยวข้องกันสูง และมนุษย์สามารถตีความได้ว่าเป็นหัวข้อที่มีความหมาย ซึ่งช่วยในการเลือกจำนวนหัวข้อที่เหมาะสมสำหรับ LDA โดยทั่วไป

1) ค่า C_v > 0.5 ถือว่าหัวข้อมีความสอดคล้องสูง และสามารถตีความได้ดี
2) ค่า C_v 0.3-0.5 ถือว่าหัวข้อมีความสอดคล้องปานกลาง อาจต้องปรับปรุง
3) ค่า C_v < 0.3 ถือว่าหัวข้อมีความสอดคล้องต่ำ และยากต่อการตีความ

ตัวอย่างเช่น หัวข้อที่ประกอบด้วยคำ "อาจารย์", "การสอน", "หลักสูตร", "ห้องเรียน", "การเรียน" จะมีค่า Coherence สูง เนื่องจากคำเหล่านี้มีความเกี่ยวข้องกันทางความหมายในบริบทของการศึกษา ในขณะที่หัวข้อที่ประกอบด้วยคำ "อาจารย์", "อาหาร", "กีฬา", "คอมพิวเตอร์", "สวน" จะมีค่า Coherence ต่ำ เนื่องจากคำเหล่านี้ไม่มีความเกี่ยวข้องกันทางความหมาย

การคำนวณ Coherence Score สามารถคำนวณได้จากสมการ (10)

$$C_V = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \mathrm{NPMI}(w_i, w_j) \qquad (10)$$

โดยที่

N = จำนวนคำที่พิจารณาในหัวข้อ (top N words)

w_i, w_j = คำที่ i และ j

NPMI(w_i, w_j) = Normalized Pointwise Mutual Information คำระหว่าง wi และ wj

การใช้ Coherence Score ร่วมกับ Perplexity ช่วยให้การประเมินประสิทธิภาพของโมเดล LDA มีความครอบคลุมทั้งในด้านความสามารถทางสถิติ (Perplexity) และความมีความหมายที่มนุษย์สามารถตีความได้ (Coherence Score) ซึ่งเป็นสิ่งสำคัญสำหรับการวิเคราะห์ความคิดเห็นของนักศึกษา

Topic Diversity คือเกณฑ์ที่วัดความหลากหลายระหว่างหัวข้อที่ค้นพบ ช่วยประเมินความซ้ำซ้อนของหัวข้อ(Dieng et al., 2020)หากค่า Topic Diversity สูง แสดงว่าหัวข้อที่ค้นพบมีความแตกต่างกันมาก ไม่ซ้ำซ้อน สามารถคำนวนได้จากสมการ (11)

$$\mathrm{PUW} = \frac{|V_{\mathrm{unique}}|}{|V_{\mathrm{total}}|} \qquad (11)$$

โดยที่

V_{unique} = จำนวนคำที่ไม่ซ้ำกันทั้งหมดในทุกหัวข้อ

V_{total} = จำนวนคำทั้งหมดในทุกหัวข้อ (รวมคำซ้ำ)

2.2 งานวิจัยที่เกี่ยวข้อง

การศึกษานี้ได้ทบทวนงานวิจัยที่เกี่ยวข้องกับแนวทางและวิธีการวิเคราะห์ความรู้สึกของนักศึกษาที่มีต่อการให้บริการของมหาวิทยาลัย โดยมุ่งเน้นการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) สามารถสรุปได้ดังนี้

1. Sentiment Analysis of Students' Feedback on Institutional Facilities Using Text Based Classification and Natural Language Processing (NLP)

Khaiser et al. (2023) ได้ศึกษาการวิเคราะห์ความรู้สึก (Sentiment Analysis) จากความคิดเห็นของนักศึกษาเกี่ยวกับสิ่งอำนวยความสะดวกในสถาบันการศึกษา โดยมีวัตถุประสงค์เพื่อจำแนกข้อความแสดงความคิดเห็นออกเป็นความรู้สึกเชิงบวก เชิงลบ หรือเป็นกลาง งานวิจัยนี้มุ่งเน้นการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) แทนวิธีการใช้พจนานุกรมคำศัพท์ (Lexicon-based) เนื่องจากมีความยืดหยุ่นสูงกว่าในการปรับให้เข้ากับบริบทเฉพาะของการศึกษา กระบวนการวิจัยประกอบด้วยขั้นตอนหลัก 4 ขั้นตอน ได้แก่ 1) การเก็บรวบรวมข้อมูล 2) การประมวลผลข้อความเบื้องต้น (Pre-processing) 3) การสกัดคุณลักษณะ (Feature Extraction) โดยใช้ TF-IDF และ 4) การจำแนกประเภทความรู้สึก โดยใช้อัลกอริทึม Naïve Bayes และ Support Vector Machine (SVM) ดังแสดงในภาพประกอบ 1

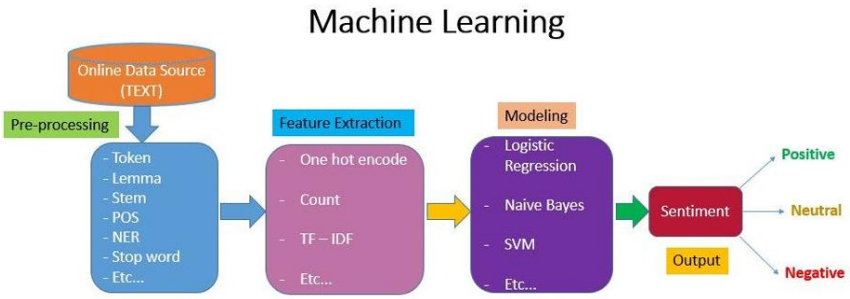


Figure 1: Sentiment Analysis Framework using Machine Learning

ภาพประกอบ 1 ขั้นตอนการวิเคราะห์ความรู้สึก

ที่มา:(Khaiser et al., 2023)

ผลการวิจัยพบว่า อัลกอริทึม Naïve Bayes ให้ประสิทธิภาพสูงกว่า SVM โดยมีค่า F1-Score เท่ากับ 92.00% ในขณะที่ SVM มีค่า F1-Score เท่ากับ 88.00% ดังแสดงในตาราง 1 นอกจากนี้ ผู้วิจัยยังพบว่าการใช้ TF-IDF ในการสกัดคุณลักษณะช่วยเพิ่มประสิทธิภาพของการจำแนกความรู้สึกได้ดีกว่าการใช้เพียง Bag of Words แบบดั้งเดิม

งานวิจัยนี้มีความสอดคล้องกับการวิจัยปัจจุบันในแง่ของการประยุกต์ใช้เทคนิค NLP ในการวิเคราะห์ความคิดเห็นของนักศึกษา อย่างไรก็ตาม งานวิจัยนี้มุ่งเน้นเฉพาะสิ่งอำนวยความสะดวกในสถาบัน ในขณะที่การวิจัยปัจจุบันมีขอบเขตกว้างกว่าโดยครอบคลุมการให้บริการทุกด้านของมหาวิทยาลัย

2. Sentiment analysis in MOOC discussion forums: What does it tell us?

Wen et al. (2014) นำเสนอแนวทางการวิเคราะห์ข้อมูลป้อนกลับของนักศึกษาโดยผสมผสานเทคนิคการวิเคราะห์ความรู้สึกร่วมกับการจัดกลุ่มหัวข้อ (Topic Modeling) โดยใช้ข้อมูลจากการประเมินรายวิชาและอาจารย์ผู้สอนของมหาวิทยาลัยในออสเตรเลีย จำนวน 4,812 ข้อความ ผู้วิจัยใช้ VADER (Valence Aware Dictionary and sEntiment Reasoner) ในการวิเคราะห์ความรู้สึกเบื้องต้น เนื่องจากมีประสิทธิภาพสูงสำหรับข้อความสั้นและไม่เป็นทางการ จากนั้นใช้ Latent Dirichlet Allocation (LDA) ในการจัดกลุ่มหัวข้อ

ผลการวิจัยพบว่าสามารถระบุหัวข้อสำคัญ 7 หัวข้อจากข้อมูลป้อนกลับ ได้แก่ 1) คุณภาพการสอน 2) เนื้อหารายวิชา 3) การวัดและประเมินผล 4) ทรัพยากรการเรียนรู้ 5) การให้คำปรึกษาและการสนับสนุน 6) การบริหารจัดการรายวิชา และ 7) สภาพแวดล้อมการเรียนรู้ ผู้วิจัยพบความสัมพันธ์ระหว่างความรู้สึกและหัวข้อ โดยหัวข้อด้านคุณภาพการสอนและการให้คำปรึกษา

จะมีความรู้สึกเชิงบวก ในขณะที่หัวข้อด้านการวัดและประเมินผลและการบริหารจัดการรายวิชามักมีความรู้สึกเชิงลบ ซึ่งให้ข้อมูลเชิงลึกที่มีค่าสำหรับการพัฒนาคุณภาพการศึกษา

3. Sentimental Analysis of Student Feedback using Machine Learning Techniques

Dsouza et al. (2019) ได้มีการเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องต่าง ๆ ในการวิเคราะห์ความรู้สึกจากข้อมูลป้อนกลับในรูปแบบข้อความของนักศึกษา โดยงานวิจัยนี้ได้รวบรวมข้อมูลผ่านแบบฟอร์ม Google Forms และนำข้อมูลที่ได้มาวิเคราะห์ด้วยอัลกอริทึมการเรียนรู้ของเครื่องจำนวน 3 ประเภทได้แก่ Support Vector Machine (SVM), Multinomial Naive Bayes Classifier (MNBC) และ Random Forest (RF) โดยมีการเปรียบเทียบประสิทธิภาพผ่านค่าความแม่นยำ (Accuracy)

ผลการวิจัยพบว่า SVM ให้ผลลัพธ์ดีที่สุด โดยมีความแม่นยำประมาณ 80% ซึ่งสูงกว่า Multinomial Naive Bayes Classifier (MNBC) และ Random Forest อย่างมีนัยสำคัญ ผู้วิจัยอธิบายว่า SVM มีประสิทธิภาพดีในการจำแนกข้อความสั้นๆ ซึ่งเป็นลักษณะทั่วไปของข้อมูลป้อนกลับจากนักศึกษา นอกจากนี้ยังพบว่า การประมวลผลข้อความเบื้องต้น (Pre-processing) โดยเฉพาะการกำจัดคำที่ไม่มีความหมาย (Stop Words) และการลดรูปคำ (Stemming) มีผลต่อประสิทธิภาพของการวิเคราะห์อย่างมาก

งานวิจัยนี้มีความสำคัญต่อการวิจัยปัจจุบันในแง่ของการเปรียบเทียบอัลกอริทึมสำหรับการวิเคราะห์ความรู้สึกจากข้อมูลป้อนกลับของนักศึกษา ซึ่งช่วยในการเลือกอัลกอริทึมที่เหมาะสมสำหรับการวิจัยปัจจุบัน

4. Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study

Kastrati et al. (2021) ได้มีการศึกษาและเปรียบเทียบประสิทธิภาพการทำงานของอัลกอริทึมการเรียนรู้ของเครื่องสำหรับการวิเคราะห์ความรู้สึกจากข้อมูลป้อนกลับทางการศึกษา โดยการศึกษานี้ใช้ชุดข้อมูลมากกว่า 24,000 ข้อความ ซึ่งได้จากความคิดเห็นของนักศึกษาเกี่ยวกับระบบการเรียนการสอนออนไลน์และการประเมินรายวิชาผู้วิจัยได้ทดสอบและเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องทั้งแบบดั้งเดิมและแบบการเรียนรู้เชิงลึก ได้แก่ Naïve Bayes, Support Vector Machine (SVM), Random Forest, Convolutional Neural Network (CNN),

Long Short-Term Memory (LSTM) และโมเดลที่ใช้ transformer เช่น BERT (Bidirectional Encoder Representations from Transformers) ผลการวิจัยพบว่า BERT ให้ประสิทธิภาพสูงสุด โดยมีค่า F1-Score 91.4% ตามด้วย LSTM (87.6%) และ CNN (85.3%) ในขณะที่วิธีแบบดั้งเดิม อย่าง SVM, Random Forest และ Naïve Bayes มีค่า F1-Score 83.1%, 81.7% และ 77.5% ตามลำดับ การศึกษานี้ยังเปรียบเทียบเทคนิคการแทนค่าคำ (Word Representation) แบบต่างๆ ได้แก่ TF-IDF, Word2Vec และ GloVe โดยพบว่า Word2Vec และ GloVe มีประสิทธิภาพดีกว่าเมื่อใช้กับโมเดลการเรียนรู้เชิงลึก ขณะที่ TF-IDF ยังคงเป็นทางเลือกที่เหมาะสมสำหรับอัลกอริทึมแบบดั้งเดิม อย่างไรก็ตาม ผู้วิจัยได้เน้นย้ำว่าแม้ BERT จะมีประสิทธิภาพสูงสุด แต่ก็ต้องการทรัพยากรการประมวลผลมากกว่าวิธีอื่นๆ อย่างมีนัยสำคัญ

5. Using sentiment analysis to evaluate qualitative students' responses

งานวิจัยของ Dake and Gyimah (2023) ได้ประยุกต์อัลกอริทึมสำหรับแยกประเภทได้แก่ Naïve Bayes, Support Vector Machine, J48 Decision Tree และ Random Forest เพื่อวิเคราะห์ความรู้สึกจากข้อความตอบกลับของนักศึกษาที่มหาวิทยาลัย University of Education, Winneba โดยการวิจัยแสดงผลว่าอัลกอริทึม SVM มีประสิทธิภาพสูงสุดด้วยค่าความแม่นยำ 63.79% การศึกษานี้มีความสำคัญอย่างยิ่งต่อการพัฒนาการวิเคราะห์ข้อความในบริบทของระบบการศึกษาสมัยใหม่ โดยสามารถช่วยให้สถาบันอุดมศึกษาประเมินคุณภาพการให้บริการได้อย่างแม่นยำ และนำข้อมูลที่ได้ไปใช้ในการพัฒนา นวัตกรรมด้านการเรียนการสอนเพื่อตอบสนองผู้เรียนได้อย่างมีประสิทธิภาพ

2.3 สรุปการทบทวนงานวิจัย

จากการทบทวนงานวิจัยที่เกี่ยวข้อง พบว่า งานวิจัยที่มุ่งเน้นด้านการวิเคราะห์ความรู้สึกของนักศึกษาที่มีต่อการให้บริการทางการศึกษา โดยอาศัยเทคนิคการประมวลผลภาษาธรรมชาติ (NLP) ร่วมกับอัลกอริทึมการเรียนรู้ของเครื่อง ยังมีประเด็นที่ควรได้รับการพัฒนาเพิ่มเติม โดยเฉพาะการผสมผสานเทคนิคต่าง ๆ อย่างเหมาะสม เพื่อให้ได้ข้อมูลเชิงลึกที่สามารถนำไปใช้ประโยชน์ได้อย่างมีประสิทธิภาพยิ่งขึ้น

งานวิจัยในบทที่ 3 ได้นำแนวทางและเทคนิคจากงานวิจัยที่เกี่ยวข้องมาปรับใช้อย่างเป็นระบบ โดยใช้ VADER (Valence Aware Dictionary and sentiment Reasoner) ในการวิเคราะห์ความรู้สึกเบื้องต้น เลือกใช้ TF-IDF (Term Frequency-Inverse Document Frequency) และ

Word2Vec ในการสกัดคุณลักษณะสำคัญจากข้อความ และใช้อัลกอริทึม Support Vector Machine (SVM), Naïve Bayes และ Random Forest ในการสร้างรูปแบบจำลองสำหรับการจำแนกความรู้สึก

เนื่องจากข้อมูลความคิดเห็นมีลักษณะไม่สมดุล (Class Imbalance) จึงใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) และ Random Undersampling ในการจัดการปัญหาดังกล่าว นอกจากนี้ยังนำ LDA (Latent Dirichlet Allocation) มาช่วยวิเคราะห์หัวข้อร่วมกับการวิเคราะห์ความรู้สึก เพื่อให้ได้ข้อมูลเชิงลึกที่สามารถนำไปใช้ในการปรับปรุงคุณภาพการให้บริการของมหาวิทยาลัยได้อย่างมีประสิทธิภาพ โดยครอบคลุมทั้งด้านการระบุความรู้สึกของนักศึกษาและการค้นหาประเด็นสำคัญที่นักศึกษาให้ความสนใจ

บทที่ 3
วิธีการวิจัย

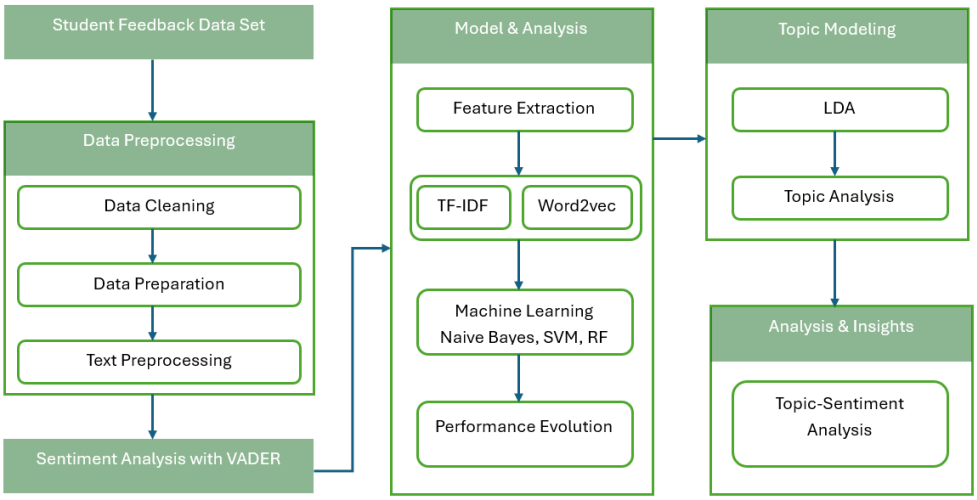
งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อตรวจสอบและประเมินทัศนคติของนักศึกษาที่มีต่อคุณภาพการบริการของสถาบันการศึกษาระดับอุดมศึกษา ทั้งนี้ได้นำเทคโนโลยีการประมวลผลภาษาธรรมชาติ (Natural Language Processing - NLP) ซึ่งเป็นสาขาย่อยภายใต้เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) มาใช้เป็นเครื่องมือหลักในการดำเนินการศึกษา ในบทนี้จะอธิบายถึงกระบวนการดำเนินงานวิจัยอย่างครบถ้วน ครอบคลุมตั้งแต่หลักการพื้นฐานและโครงสร้างทางทฤษฎีของการวิจัย กระบวนการจัดหาและรวบรวมข้อมูล ขั้นตอนการจัดการและวิเคราะห์ข้อมูล ไปจนถึงวิธีการนำเสนอผลการศึกษาในรูปแบบที่ชัดเจนและเข้าใจง่าย

3.1 แนวคิดและกรอบการวิจัย

การศึกษานี้มุ่งวิเคราะห์ทัศนคติของนักศึกษาต่อบริการมหาวิทยาลัยผ่านเทคนิคการประมวลผลภาษาธรรมชาติ โดยพัฒนาแบบจำลองเพื่อจำแนกความรู้สึกจากความคิดเห็นของนักศึกษาและค้นหาประเด็นหลักที่ใช้ประเมินคุณภาพการบริการ เพื่อนำผลไปปรับปรุงการให้บริการของมหาวิทยาลัย

งานวิจัยนี้ดำเนินการอย่างเป็นระบบตั้งแต่การเตรียมข้อมูลจนถึงการสรุปผลการศึกษา โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติเป็นหลัก ประกอบด้วย 2 ส่วนสำคัญ คือ การวิเคราะห์ความรู้สึก (Sentiment Analysis) เพื่อแยกความคิดเห็นเป็นแง่บวก แง่ลบ หรือเป็นกลาง และการวิเคราะห์หัวข้อ (Topic Modeling) เพื่อค้นหาประเด็นหลักที่ปรากฏในข้อมูลความคิดเห็น

กระบวนการวิจัยประกอบด้วยสี่ขั้นตอนหลัก ได้แก่ ขั้นตอนการเตรียมข้อมูล (Data Preprocessing) ที่รวมถึงการทำความสะอาดข้อมูล การเตรียมข้อมูล และการประมวลผลข้อความขั้นต้น ขั้นตอนการวิเคราะห์ความรู้สึกเบื้องต้นด้วย VADER ขั้นตอนการวิเคราะห์โมเดล (Model & Analysis) ที่ประกอบด้วยการสกัดลักษณะข้อมูล การแปลงข้อความด้วยเทคนิค TF-IDF และ Word2Vec การนำไปใช้กับอัลกอริทึมการเรียนรู้ของเครื่อง และการประเมินประสิทธิภาพ ขั้นตอนการวิเคราะห์หัวข้อ (Topic Modeling) ที่ใช้เทคนิค LDA เพื่อวิเคราะห์หัวข้อและสรุปผลการวิเคราะห์ความรู้สึกร่วมกับหัวข้อ ดังแสดงในภาพประกอบ 2



ภาพประกอบ 2 ขั้นตอนการดำเนินงาน

กรอบแนวคิดของงานวิจัยประกอบด้วย 4 ส่วนหลัก

1. การเตรียมข้อมูล (Data Set & Data Preprocessing)
2. การวิเคราะห์ความรู้สึกเบื้องต้นด้วย VADER (Sentiment Analysis with VADER)
3. การสร้างรูปแบบจำลองและวิเคราะห์ขั้นสูง (Modeling & Analysis)
4. การวิเคราะห์หัวข้อและการสรุปข้อมูลเชิงลึก (Topic Modeling & Analysis & Insights)

3.2 ขั้นตอนการดำเนินการวิจัย

3.2.1 การรวบรวมข้อมูล

ข้อมูลความคิดเห็นถูกเก็บรวบรวมจากระบบประเมินความพึงพอใจออนไลน์ของมหาวิทยาลัยในช่วงปีการศึกษา 2564-2566 โดยใช้วิธีการสุ่มแบบเจาะจง เพื่อให้ได้ข้อมูลที่ครอบคลุมหัวข้อสำคัญ เช่น การบริการนักศึกษา สิ่งอำนวยความสะดวกในสถานศึกษา และคุณภาพการสอน โดยมีจำนวนข้อมูลความคิดเห็นทั้งหมด 1,000 รายการ ข้อมูลถูกเก็บในรูปแบบไฟล์ CSV ที่ประกอบด้วยข้อความความคิดเห็น (Review) และข้อมูลอื่นๆ ที่เกี่ยวข้อง ดังแสดงในภาพประกอบ 3

	comment_id	date	year	department	major	gpa	topic	subtopic	Description
0	F0001	9/12/2020	ชั้นปีที่ 4	คณะนิติศาสตร์	นิติศาสตร์	2.06	อาคารสถานที่	โรงอาหาร	ไม่พอใจกับโรงอาหารมีขนาดที่เล็ก
1	F0002	15/8/2021	ชั้นปีที่ 3	คณะบริหารธุรกิจ	การจัดการทรัพยากรมนุษย์	2.56	อาคารสถานที่	ห้องเรียน	ขอชมเชยห้องเรียน เป็นเรื่องที่นักศึกษาพูดถึงเส...
2	F0003	11/2/2021	ชั้นปีที่ 1	คณะนิติศาสตร์	นิติศาสตร์	3.46	บริการนักศึกษา	การขอเอกสาร	การขอเอกสารข้ามากและยุ่งยากซับซ้อนพอสอบถามเจ้า...
3	F0004	2/9/2020	ชั้นปีที่ 2	คณะศิลปศาสตร์	การโรงแรม	3.41	บริการนักศึกษา	การยื่นคำร้อง	ขั้นตอนการยื่นคำร้องยุ่งยากและซับซ้อนเกินไป ทำ...
4	F0005	18/6/2022	ชั้นปีที่ 4	คณะนิติศาสตร์	นิติศาสตร์	3.27	สิ่งอำนวยความสะดวก	ห้องคอมพิวเตอร์	ไม่เหมาะสมเรื่องห้องคอมพิวเตอร์

ภาพประกอบ 3 การแสดงตัวอย่างชุดข้อมูล

3.2.2 การเตรียมข้อมูล

การเตรียมข้อมูลเป็นขั้นตอนสำคัญที่จะส่งผลต่อคุณภาพของการวิเคราะห์ ประกอบด้วย 3 ขั้นตอนย่อยดังนี้

3.2.2.1 การทำความสะอาดข้อมูล (Data Cleaning)

ขั้นตอนนี้ ผู้วิจัยดำเนินการกำจัดข้อมูลที่ไม่เกี่ยวข้องหรือไม่สมบูรณ์ เช่น

1) ตรวจสอบและจัดการข้อมูลที่ขาดหาย (missing values) โดยใช้คำสั่ง df.isnull().sum() และ df.isna().sum() เพื่อตรวจสอบจำนวนข้อมูลที่ขาดหาย
2) ในกรณีที่พบข้อมูลขาดหาย ผู้วิจัยใช้วิธีลบแถวที่มีข้อมูลขาดหายออกโดยใช้คำสั่ง df.dropna()
3) ปรับชื่อคอลัมน์ให้เป็นมาตรฐานและสื่อความหมาย โดยเปลี่ยนจาก 'Description' เป็น 'Review' ด้วยคำสั่ง df.rename(columns=(Alpaydin))

3.2.2.2 การตรวจสอบและแปลงภาษา (Language Identification and Translation)

เนื่องจากข้อมูลความคิดเห็นมีทั้งภาษาไทยและภาษาอังกฤษ จึงต้องมีการระบุภาษาและแปลงให้เป็นภาษาเดียวกัน ดังนี้

1) ตรวจสอบภาษาของแต่ละความคิดเห็นด้วยไลบรารี LangDetect ผ่านฟังก์ชัน detect_language ที่สร้างขึ้น ดังแสดงในภาพประกอบ 4

```python
def detect_language(text):
    if isinstance(text, str):
        if text.strip() == '':
            return 'unknown'
        try:
            return detect(text)
        except:
            return 'unknown'
    else:
        return 'unknown'
```

ภาพประกอบ 4 การตรวจสอบภาษา

2) แปลความคิดเห็นภาษาไทยเป็นภาษาอังกฤษด้วย Google Translate API โดยใช้ฟังก์ชัน translate_df แบบ asynchronous เพื่อจัดการกับข้อจำกัดของ API

3) ตรวจสอบคุณภาพของการแปลโดยใช้คำสั่ง df['Translated_Language'] = df['Translated_Review'].apply(detect_language) และกรองเฉพาะความคิดเห็นที่แปลเป็นภาษาอังกฤษได้อย่างถูกต้อง

3.2.2.3 การประมวลผลข้อความ (Text Preprocessing)

หลังจากได้ข้อความภาษาอังกฤษแล้ว ผู้วิจัยได้ดำเนินการประมวลผลข้อความดังนี้

1) แบ่งข้อความเป็นคำ (Tokenization) ด้วย NLTK word_tokenize(text)

2) ลบคำหยุด (Stop Words) ที่ไม่มีความหมายเชิงวิเคราะห์ เช่น '.', ',', '?', '!', ':', ';', '-', '(', ')', '[', ']', "", """] ด้วยคำสั่ง stop_words.update()

3) ลดรูปคำ (Stemming) ด้วย Porter Stemmer() เพื่อลดความซับซ้อนของข้อมูล

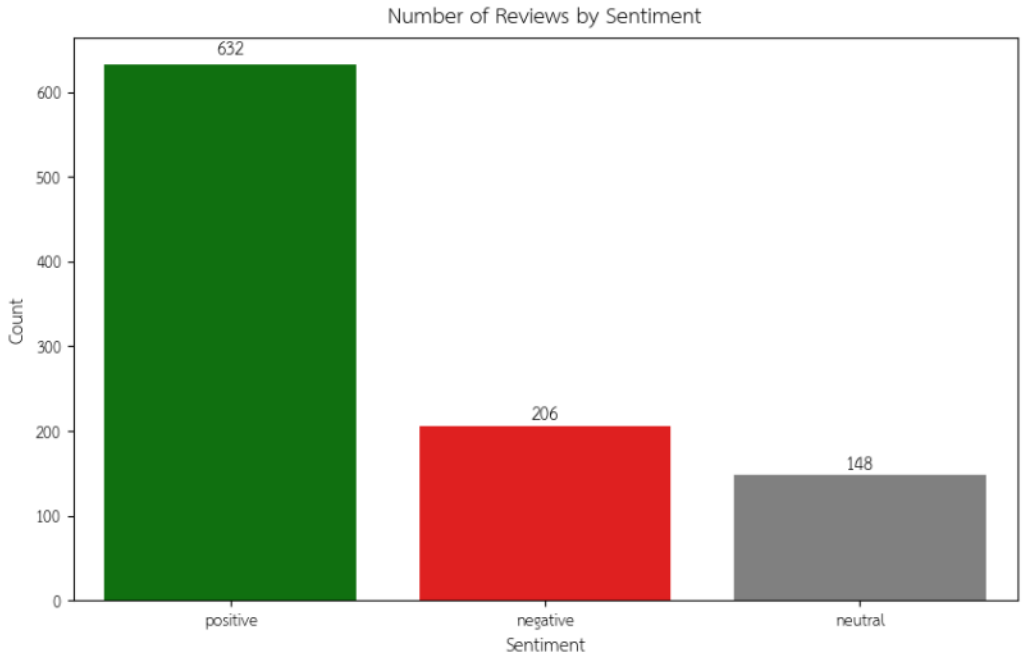
ผลลัพธ์จากขั้นตอนนี้คือข้อมูลความคิดเห็นภาษาอังกฤษที่ผ่านการทำความสะอาดและพร้อมสำหรับขั้นตอนต่อไป โดยสร้างคอลัมน์ 'Cleaned_Review' ที่เก็บข้อความที่ผ่านการประมวลผลแล้ว ดังแสดงในภาพประกอบ 5

	Translated_Review	Cleaned_Review
0	Not satisfied with the cafeteria that is small	Not satisfied cafeteria small
1	Compliment the classroom Is a story that stude...	Compliment classroom Is story students always ...
2	Requesting documents is very slow and complica...	Requesting documents slow complicated When ask...
3	The process of filing the petition is too comp...	The process filing petition complicated compli...
4	Not suitable for computer room	Not suitable computer room

ภาพประกอบ 5 การแสดงตัวอย่างข้อความที่ผ่านการประมวลผล

3.2.3 การวิเคราะห์ความรู้สึกเบื้องต้น (Sentiment Analysis with VADER)

ในขั้นต้นก่อนพัฒนาแบบจำลอง การศึกษานี้ประยุกต์ใช้ VADER (Valence Aware Dictionary and sentiment Reasoner) เป็นเครื่องมือวิเคราะห์ความรู้สึกพื้นฐานที่มีประสิทธิภาพสำหรับข้อความสื่อสังคมและความคิดเห็นออนไลน์ VADER ประเมินความรู้สึกโดยใช้พจนานุกรมคำที่มีการให้ค่าความเข้มข้นของความรู้สึก ซึ่งมีการคำนวณคะแนนความรู้สึกในมิติต่างๆ ได้แก่ Positive, Negative, Neutral และ Compound ดังแสดงในภาพประกอบ 6



ภาพประกอบ 6 การจำแนกประเภทความรู้สึกด้วย VADER

กำหนดเกณฑ์การแบ่งประเภทความรู้สึกโดยใช้ค่า Compound Score แบ่งเป็น 3 กลุ่ม ดังนี้ Positive: Compound Score $\geq$ 0.05, Negative: Compound Score $\leq$ -0.05, Neutral: -0.05 < Compound Score < 0.05

การเลือกค่า threshold ที่ 0.05 เป็นผลจากการทดลองกับค่า threshold หลายค่า (0.05, 0.1, 0.2) และพบว่าค่า 0.05 ให้การกระจายของความรู้สึกที่สมดุลที่สุด โดยไม่ทำให้ความคิดเห็นเป็นกลางมากเกินไปหรือสูญเสียความคิดเห็นด้านอื่นๆ ดังแสดงในภาพประกอบ 7

```
Threshold = 0.05: positive    632
negative    206
neutral     148
Name: count, dtype: int64
Threshold = 0.1: positive    620
negative    196
neutral     170
Name: count, dtype: int64
Threshold = 0.2: positive    594
neutral     221
negative    171
Name: count, dtype: int64
```

ภาพประกอบ 7 การทดลองแบ่งประเภทความรู้สึกของแต่ละ threshold

คะแนน Compound เป็นค่าที่รวมผลจากทุกมิติและถูกนอร์มัลไลซ์ให้อยู่ในช่วง -1 (ความรู้สึกเชิงลบ) ถึง +1 (ความรู้สึกเชิงบวก) ซึ่งเป็นค่าที่ใช้ในการตัดสินประเภทความรู้สึกโดยรวมของข้อความ ดังแสดงในภาพประกอบ 8

```
--- ตัวอย่างความคิดเห็น positive ---
1. Impressed with fast and friendly service
   คะแนน: {'neg': 0.0, 'neu': 0.388, 'pos': 0.612, 'compound': 0.743}
2. Quite normal about the exam If developed, it will create a lot of impression.
   คะแนน: {'neg': 0.0, 'neu': 0.733, 'pos': 0.267, 'compound': 0.4588}
3. Feel good about student data systems Is something that students want a lot
   คะแนน: {'neg': 0.0, 'neu': 0.704, 'pos': 0.296, 'compound': 0.4939}
4. Feel good about the sports center Would like to continue to develop Which is related to wifi
   คะแนน: {'neg': 0.0, 'neu': 0.735, 'pos': 0.265, 'compound': 0.6597}
5. Very headache when requesting documents Should be improved Especially when filing a petition
   คะแนน: {'neg': 0.0, 'neu': 0.78, 'pos': 0.22, 'compound': 0.4767}

--- ตัวอย่างความคิดเห็น neutral ---
1. Enter the sports center Makes living in the university more convenient
   คะแนน: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
2. The application process is quite complex. But can still understand
   คะแนน: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
3. The collection of fund application history is not yet as it should be. Must fill out the information repeatedly in each application
   คะแนน: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
4. May have to develop the modernity of the course In terms of access to teachers
   คะแนน: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
5. The exam format is too obsession. Would like to have a subjective exam that has expressed more ideas
   คะแนน: {'neg': 0.121, 'neu': 0.754, 'pos': 0.126, 'compound': 0.0258}
```

ภาพประกอบ 8 การแสดงตัวอย่างคะแนนแต่ละความรู้สึก

3.2.4 การสร้างแบบจำลองและการวิเคราะห์ (Modeling & Analysis)

หลังจากข้อมูลผ่านการเตรียมและวิเคราะห์ความรู้สึกเบื้องต้นแล้ว นำข้อมูลไปสร้างแบบจำลองเพื่อจำแนกความรู้สึกโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ด้วยวิธีการดังนี้

3.2.4.1 การสกัดคุณลักษณะสำคัญจากข้อความ(Feature Extraction)

เป็นขั้นตอนการจัดเตรียมและประมวลผลข้อมูลเพื่อใช้กับโมเดลเรียนรู้ของเครื่อง โดยประยุกต์ใช้วิธีการหลัก 2 แบบ

1. TF-IDF (Term Frequency-Inverse Document Frequency) เป็นเทคนิคในการแปลงข้อความเป็นเวกเตอร์ที่แทนความสำคัญของคำแต่ละคำในเอกสาร โดยค่า TF (Term Frequency) แสดงความถี่ของคำในเอกสาร และค่า IDF (Inverse Document Frequency) แสดงความสำคัญของคำที่ปรากฏในเอกสารทั้งหมด ดังแสดงในภาพประกอบ 9

```
[ ] # TF-IDF Vectorization
    tfidf_vectorizer = TfidfVectorizer(max_features=5000, min_df=1, max_df=0.8)
    tfidf_matrix = tfidf_vectorizer.fit_transform(df['Cleaned_Review'])
    print(f"รูปร่างของ TF-IDF Matrix: {tfidf_matrix.shape}")

→ รูปร่างของ TF-IDF Matrix: (986, 882)
```

ภาพประกอบ 9 การแปลงข้อความเป็นเวกเตอร์ด้วยวิธี TF-IDF

ได้กำหนดพารามิเตอร์ดังนี้

1) max_features=5000: จำกัดจำนวนมิติของเวกเตอร์ให้ไม่เกิน 5,000 มิติ
2) min_df=1: คำนั้นต้องปรากฏอย่างน้อย 1 ครั้งในเอกสาร
3) max_df=0.8: คำนั้นต้องปรากฏไม่เกิน 80% ของเอกสารทั้งหมด

2. Word2Vec (Word-to-Vector Embeddings) เป็นเทคนิคการเรียนรู้การแทนค่าคำในรูปแบบเวกเตอร์ (Word Embeddings) ที่สามารถจับความสัมพันธ์เชิงความหมายและบริบทระหว่างคำได้อย่างมีประสิทธิภาพ แตกต่างจาก TF-IDF ที่เน้นความถี่ของคำ Word2Vec สามารถเรียนรู้ความหมายเชิงลึกของคำจากบริบทที่ปรากฏร่วมกัน การพัฒนาโมเดล Word2Vec เฉพาะโดเมน ในงานวิจัยนี้ได้สร้างโมเดล Word2Vec เฉพาะโดเมน (Domain-Specific Word2Vec Model) โดยเทรนด้วยข้อมูลความคิดเห็นของนักศึกษาที่เก็บรวบรวมได้ เพื่อให้ได้เวกเตอร์คำที่สะท้อนบริบทและความหมายเฉพาะของศัพท์ที่ใช้ในการประเมินการให้บริการของมหาวิทยาลัย ดังแสดงในภาพประกอบ 10

```python
# แบ่งข้อความเป็นคำ (tokenize)
tokenized_texts = [word_tokenize(text) for text in df['Cleaned_Review'] if isinstance(text, str)]

# สร้างโมเดล Word2Vec
word2vec_model = Word2Vec(
    sentences=tokenized_texts,
    vector_size=300,  # ขนาดมิติของเวกเตอร์
    window=5,         # จำนวนคำรอบข้างที่พิจารณา
    min_count=2,      # จำนวนครั้งขั้นต่ำที่คำปรากฏ
    workers=4,        # จำนวน CPU cores ที่ใช้
    sg=0              # ใช้โมเดล Skip-gram (ถ้าเป็น 0 จะใช้ CBOW)
)
```

ภาพประกอบ 10 การแปลงข้อความเป็นเวกเตอร์ด้วยวิธี Word2Vec

พารามิเตอร์ที่กำหนด

1) vector_size=300 เลือกขนาด 300 มิติซึ่งเป็นมาตรฐานที่ให้ความสมดุลระหว่างความสามารถในการแทนความหมายและประสิทธิภาพการคำนวณ การทดลองเปรียบเทียบกับขนาด 100, 200 และ 400 มิติ พบว่า 300 มิติให้ผลลัพธ์ที่ดีที่สุดในงานจำแนกความรู้สึก
2) window=5 กำหนดให้พิจารณาบริบทของคำรอบข้าง 5 คำ (ทั้งซ้ายและขวา) เหมาะสมสำหรับการจับความสัมพันธ์ในประโยคความคิดเห็นที่มีความยาวเฉลี่ย 10-15 คำ การทดลองกับ window size 3, 5, 7 และ 10 พบว่าค่า 5 ให้ความสมดุลที่ดีระหว่างการจับบริบทใกล้และไกล
3) min_count=2 กรองคำที่ปรากฏน้อยกว่า 2 ครั้ง เพื่อลดสัญญาณรบกวนจากคำที่หายากและลดขนาด vocabulary ที่จำเป็นต้องเรียนรู้ จากการวิเคราะห์พบว่าการใช้ min_count=1 ทำให้ vocabulary มีขนาดใหญ่เกินไปและส่งผลต่อประสิทธิภาพ

4) sg=0 คือการใช้อัลกอริทึม CBOW (Continuous Bag of Words) ซึ่งทำนายคำเป้าหมายจากคำบริบทรอบข้าง เหมาะสมกับข้อมูลขนาดกลาง (1,000 รายการ) และให้ผลลัพธ์ที่เสถียรกว่า Skip-gram (sg=1) ในกรณีของข้อมูลที่มีจำนวนจำกัด
5) epochs=10 เป็นจำนวนรอบการเรียนรู้ทั้งหมด การทดลองกับ 5, 10, 15 และ 20 epochs พบว่า 10 epochs ให้ความสมดุลที่ดีระหว่างการลู่เข้าของโมเดลและเวลาการประมวลผล
6) workers=4 เป็นจำนวน thread ที่ใช้ในการประมวลผลแบบขนาน เพื่อเพิ่มความเร็วในการเทรนโมเดล
7) negative=5 เป็นจำนวน negative samples ที่ใช้ในการเทรน ซึ่งช่วยเพิ่มประสิทธิภาพของการเรียนรู้โดยการเปรียบเทียบกับคำที่ไม่เกี่ยวข้อง

3.2.4.2 การจัดการกับข้อมูลที่ไม่สมดุล (Imbalanced Data Handling)

เนื่องจากข้อมูลมีความไม่สมดุล (Class Imbalance) งานวิจัยนี้ได้ทดลองใช้เทคนิคการแก้ปัญหาความไม่สมดุลของคลาส 3 แบบ ได้แก่ 1. Oversampling ด้วย SMOTE เพิ่มจำนวนตัวอย่างในคลาสที่มีน้อย 2. Undersampling ด้วย Random Under Sampling ลดจำนวนตัวอย่างในคลาสที่มีมาก 3. Class Weighting กำหนดน้ำหนักให้แต่ละคลาสตามสัดส่วนที่แตกต่างกัน ปัญหาความไม่สมดุลของคลาส (Class Imbalance) พบได้บ่อยในงานจำแนกข้อความ โดยเฉพาะในกรณีที่บางคลาสมีจำนวนตัวอย่างมากกว่าคลาสอื่นอย่างมีนัยสำคัญ ซึ่งอาจส่งผลให้แบบจำลองมีอคติในการทำนายคลาสที่มีจำนวนตัวอย่างมาก ในงานวิจัยนี้ได้ทดลองใช้เทคนิคการแก้ปัญหาความไม่สมดุลของคลาส 3 แบบ

1. การเพิ่มข้อมูล (Oversampling) ด้วย SMOTE (Synthetic Minority Over-sampling Technique) เป็นเทคนิคที่สร้างตัวอย่างสังเคราะห์สำหรับคลาสส่วนน้อย โดยมีขั้นตอนดังนี้

1) หาเพื่อนบ้านที่ใกล้ที่สุดสำหรับแต่ละตัวอย่างในคลาสส่วนน้อย
2) สร้างตัวอย่างสังเคราะห์โดยการสุ่มเลือกจุดระหว่างตัวอย่างและเพื่อนบ้านที่เลือก
3) เพิ่มตัวอย่างสังเคราะห์เหล่านี้ในชุดข้อมูลจนกว่าคลาสจะมีความสมดุล

การใช้ SMOTE มีข้อดีคือไม่สูญเสียข้อมูลใดๆ จากชุดข้อมูลเดิม และยังเพิ่มความหลากหลายของตัวอย่างในคลาสส่วนน้อย อย่างไรก็ตาม อาจมีความเสี่ยงที่จะสร้างตัวอย่างที่ไม่เป็นตัวแทนที่ดีของคลาสนั้นๆ หากข้อมูลมีลักษณะซับซ้อน ในการวิจัยนี้ ได้ปรับใช้ SMOTE

กับข้อมูลที่แปลงด้วยทั้ง TF-IDF และ Word2Vec เพื่อให้จำนวนตัวอย่างในแต่ละคลาสมีความสมดุลกัน ซึ่งช่วยให้แบบจำลองสามารถเรียนรู้ลักษณะเฉพาะของแต่ละคลาสได้อย่างเท่าเทียมกัน ดังแสดงในภาพประกอบ 11

```
จำนวนข้อมูลก่อน SMOTE: 788
จำนวนข้อมูลหลัง SMOTE: 1515
การกระจายตัวของคลาสก่อน SMOTE:
sentiment
positive    505
negative    165
neutral     118
Name: count, dtype: int64
การกระจายตัวของคลาสหลัง SMOTE:
sentiment
neutral     505
positive    505
negative    505
Name: count, dtype: int64
```

ภาพประกอบ 11 การกระจายของคลาสต่างๆด้วยวิธี SMOTE

2. การลดข้อมูล (Undersampling) ด้วย Random Under Sampling เป็นการลดจำนวนตัวอย่างในคลาสส่วนมาก โดยมีขั้นตอนดังนี้

1) สุ่มเลือกตัวอย่างจากคลาสส่วนมากเท่ากับจำนวนตัวอย่างในคลาสส่วนน้อย
2) นำตัวอย่างที่เลือกมารวมกับคลาสส่วนน้อยเพื่อสร้างชุดข้อมูลที่มีความสมดุล
3) กระบวนการนี้อาจสูญเสียข้อมูลบางส่วนจากคลาสส่วนมาก

ข้อดีของวิธีนี้คือทำให้ขนาดของชุดข้อมูลเล็กลง ซึ่งช่วยลดเวลาในการฝึกแบบจำลอง แต่ข้อเสียคือข้อมูลบางส่วนถูกตัดทิ้งไป ซึ่งอาจทำให้สูญเสียข้อมูลที่มีประโยชน์ต่อการเรียนรู้ของแบบจำลอง ในการวิจัยนี้ ได้ทดลองใช้ Random Under Sampling กับข้อมูลที่แปลงด้วยทั้ง TF-IDF และ Word2Vec เพื่อเปรียบเทียบกับเทคนิคอื่นๆ ดังแสดงในภาพประกอบ 12

```
จำนวนข้อมูลก่อน Undersampling: 788
จำนวนข้อมูลหลัง Undersampling: 354
การกระจายตัวของคลาสก่อน Undersampling:
sentiment
positive    505
negative    165
neutral     118
Name: count, dtype: int64
การกระจายตัวของคลาสหลัง Undersampling:
sentiment
negative    118
neutral     118
positive    118
Name: count, dtype: int64
```

ภาพประกอบ 12 การกระจายของคลาสต่างๆด้วยวิธี Undersampling

1. การปรับค่าน้ำหนักคลาส (Class Weighting) เป็นเทคนิคที่ไม่มีการปรับข้อมูล แต่ปรับน้ำหนักในฟังก์ชันการสูญเสีย (Loss Function) ของแบบจำลอง โดยมีรายละเอียดดังนี้

1) คำนวณน้ำหนักแบบ 'balanced' โดยอัตโนมัติจากความถี่ของแต่ละคลาส
2) น้ำหนักของคลาส = จำนวนตัวอย่างทั้งหมด / (จำนวนคลาส * จำนวนตัวอย่างในคลาสนั้น)
3) คลาสที่มีจำนวนตัวอย่างน้อยจะได้รับน้ำหนักมากกว่า

จุดเด่นของแนวทางนี้อยู่ที่การคงสภาพข้อมูลต้นฉบับไว้ครบถ้วน ไม่ทำให้เกิดการสูญหายของข้อมูล และยังช่วยลดความยุ่งยากในกระบวนการสร้างตัวอย่างข้อมูลใหม่ อย่างไรก็ตาม วิธีนี้อาจไม่มีประสิทธิภาพเท่าที่ควรหากความไม่สมดุลของคลาสมีความรุนแรงมาก ในการวิจัยนี้ ได้ใช้ Class Weighting กับแบบจำลองที่รองรับการกำหนดน้ำหนักคลาส เช่น SVM และ Random Forest เพื่อดูว่าวิธีนี้จะช่วยปรับปรุงประสิทธิภาพได้มากน้อยเพียงใดเมื่อเทียบกับ SMOTE และ Random Under Sampling ดังแสดงในภาพประกอบ 13

ภาพประกอบ 13 การกำหนด Class Weight ในโมเดล SVMและ Random Forest

3.2.4.3 โมเดลการเรียนรู้ของเครื่อง (Machine Learning Models)

ในการศึกษานี้ได้นำแบบจำลองการเรียนรู้ของเครื่องหลากหลายรูปแบบมาทดสอบเพื่อวิเคราะห์เปรียบเทียบความสามารถในการทำงาน โดยแบ่งตามเทคนิคการสกัดคุณลักษณะและการแก้ปัญหาความไม่สมดุลของคลาสในแต่ละโมเดลดังนี้

1. การสร้างระบบแบบจำลองโดยอาศัยเทคนิค TF-IDF
   1) การฝึกโมเดล Naive Bayes โดยใช้ค่าเริ่มต้นดังนี้:"nb_model_over = MultinomialNB()" ซึ่ง MultinomialNB เป็นอัลกอริทึมที่เหมาะสมสำหรับข้อมูลเชิงจำนวนนับ เช่น ความถี่คำในเอกสาร วิธีการนี้จะทำการจัดแบ่งข้อมูลเป็น 2 ชุด ได้แก่ ข้อมูลชุดฝึกสอน (Training Set) และข้อมูลชุดทดสอบ (Test Set)แล้วใช้ชุดข้อมูลที่ผ่านการแปลงด้วย TF-IDF เป็นอินพุตให้กับโมเดล

2) การฝึกโมเดล Support Vector Machine โดยใช้ค่าเริ่มต้นดังนี้: "svm_model_over = LinearSVC(C=1.0, max_iter=1000,random_state=42" LinearSVC ใช้ kernel เชิงเส้นสำหรับการจำแนกข้อมูล โดยมีพารามิเตอร์และขั้นตอนโดย กำหนด C=1.0: พารามิเตอร์ควบคุมระดับการลงโทษ (Penalty) สำหรับการจำแนกผิดพลาด, max_iter=1000: จำนวนรอบสูงสุดในการฝึกโมเดล ,random_state=42: ค่าเริ่มต้นสำหรับการสุ่มเพื่อให้ผลลัพธ์คงที่ใช้ชุดข้อมูลที่ผ่านการแปลงด้วย TF-IDF เป็นอินพุตให้กับโมเดล

3) การฝึกโมเดล Random Forest โดยใช้ค่าเริ่มต้นดังนี้:"rf_model_over = RandomForestClassifier(n_estimators=100, max_depth=None, random_state=42" Random Forest ประกอบด้วยต้นไม้ตัดสินใจหลายต้น โดยมีพารามิเตอร์และขั้นตอนดังนี้ n_estimators=100: จำนวนต้นไม้ตัดสินใจในแบบจำลอง, max_depth=None: ความลึกสูงสุดของแต่ละต้นไม้ โดยค่า None หมายถึงต้นไม้จะแตกกิ่งจนกว่าจะถึงใบที่มีข้อมูลเดียว, random_state=42: ค่าเริ่มต้นสำหรับการสุ่ม โดยแต่ละต้นไม้ถูกฝึกด้วยข้อมูลชุดย่อยที่สุ่มเลือกจากชุดข้อมูลฝึก ผลการทำนายมาจากการลงคะแนนเสียงข้างมากจากทุกต้นไม้

2. การสร้างระบบแบบจำลองโดยอาศัยเทคนิค Word2Vec

1) การฝึกโมเดล Naive Bayes โดยใช้ค่าเริ่มต้นดังนี้: “nb_model_w = GaussianNB()" สำหรับข้อมูลที่แปลงด้วย Word2Vec เราใช้ GaussianNB ซึ่งเหมาะกับข้อมูลต่อเนื่องที่มีการกระจายแบบปกติ โดยมีขั้นตอนดังนี้ รับข้อมูลที่แปลงด้วย Word2Vec เป็นอินพุต คำนวณค่าเฉลี่ยและค่าความผันแปรของคุณสมบัติทุกตัวใน แต่ละคลาส

2) การฝึกโมเดล Support Vector Machine โดยใช้ค่าเริ่มต้นดังนี้: “svm_model_over = LinearSVC(C=1.0, max_iter=1000, random_state=42)" กระบวนการฝึกเหมือนกับที่ใช้กับ TF-IDF แต่อินพุตคือข้อมูลที่แปลงด้วย Word2Vec

3) การฝึกโมเดล Random Forest โดยใช้ค่าเริ่มต้นดังนี้: “rf_model_over = RandomForestClassifier(n_estimators=100, max_depth=None, random_state=42)" ขั้นตอนการฝึกเหมือนกับที่ใช้กับ TF-IDF แต่ใช้ข้อมูลที่

แปลงด้วย Word2Vec สร้างต้นไม้จำนวน 100 ต้นโดยสุ่มตัวอย่างและคุณลักษณะที่แตกต่างกัน แต่ละต้นไม้เรียนรู้การแบ่งข้อมูลที่แตกต่างกัน

3. การหาค่าที่เหมาะสมสำหรับพารามิเตอร์ของ Linear Support Vector Classifier (LinearSVC) งานวิจัยนี้ใช้วิธี Grid Search Cross-Validation เพื่อกำหนดค่าพารามิเตอร์ที่ให้ผลการทำงานที่ดีที่สุด โดยปรับปรุงพารามิเตอร์สำคัญจำนวน 5 ค่า มีรายละเอียดดังนี้
    1) Regularization Parameter (C) ค่าที่ทดสอบ [0.01, 0.1, 0.5, 1.0, 5.0, 10.0] ซึ่งค่า C จะควบคุมระดับความซับซ้อนของแบบจำลอง ค่าต่ำ (0.01, 0.1) จะให้โมเดลที่เรียบง่ายแต่อาจ underfit ค่าสูง (5.0, 10.0) จะให้โมเดลที่ซับซ้อนแต่อาจ overfit ค่ากลาง (0.5, 1.0) มักให้ผลลัพธ์ที่สมดุล โดยผลการทดลอง C=5 ให้ประสิทธิภาพดีที่สุดในข้อมูลนี้
    2) Loss Function ค่าที่ทดสอบ ['hinge', 'squared_hinge'] ค่า'hinge' เป็น loss function มาตรฐานของ SVM ที่ทนทานต่อ outliers ส่วน 'squared_hinge' มีความไวต่อ outliers มากกว่าแต่อาจให้ผลลัพธ์ที่ smoot กว่า โดยผลการทดลอง 'hinge' ให้ประสิทธิภาพที่ดีกว่า
    3) Class Weight ค่าที่ทดสอบ [None, 'balanced'] ค่าNone คือข้อมูลมีความไม่สมดุลระหว่างคลาส (Positive: 45%, Negative: 30%, Neutral: 25%) การใช้ 'balanced' จะปรับน้ำหนักอัตโนมัติตามสัดส่วนผกผันของแต่ละคลาส โดยผลการทดลองได้ค่า None เนื่องจากข้อมูลมีการปรับสมดุลไปแล้ว
    4) Dual Optimization ค่าที่ทดสอบ [True, False] ค่าdual=True ใช้ dual formulation เหมาะกับกรณีที่ จำนวนข้อมูลมากกว่าจำนวน features, dual=False ใช้ primal formulation เหมาะกับกรณีที่ จำนวน featuresมากกว่าจำนวนข้อมูล ซึ่งในงานนี้ จำนวนข้อมูล =1000, n_features=5000 (TF-IDF) จึงทดสอบทั้งสองค่า โดยผลการทดลอง dual= True ให้ความเร็วในการประมวลผลดีกว่า
    5) Tolerance ค่าที่ทดสอบ [1e-4, 1e-3] เป็นการบอกโมเดลว่า "ผิดพลาดแค่ไหนถึงจะหยุดเรียนรู้" โดยสองค่าที่เปรียบเทียบคือ 1e-4 (0.0001) เป็นการยอมรับความผิดพลาดได้ต่ำ ความแม่นยำสูงแต่ใช้เวลานาน, 1e-3 (0.001) เป็นการยอมรับ

ผิดพลาดได้มากกว่า ความแม่นยำน้อยกว่าแต่ใช้เวลาน้อย โดยผลการทดลองเลือก 1e-4 ให้ประสิทธิภาพดีที่สุดในข้อมูลนี้ ดังแสดงในภาพประกอบ 14

```python
# กำหนดช่วงของพารามิเตอร์ที่ต้องการทดสอบ
param_grid = {
    'C': [0.01, 0.1, 0.5, 1.0, 5.0, 10.0],  # ค่า regularization
    'loss': ['hinge', 'squared_hinge'],  # ฟังก์ชันสูญเสีย
    'class_weight': [None, 'balanced'],  # การถ่วงน้ำหนักคลาส
    'dual': [True, False],  # การแก้ปัญหาด้วยรูปแบบ dual หรือ primal
    'tol': [1e-4, 1e-3]  # ค่าความคลาดเคลื่อนที่ยอมรับได้
}

# สร้าง GridSearchCV object
grid_search = GridSearchCV(
    LinearSVC(max_iter=10000, random_state=42),
    param_grid,
    cv=5,  # 5-fold cross-validation
    scoring='f1_weighted',  # ใช้ f1 weighted เป็นเกณฑ์ในการเลือกโมเดลที่ดีที่สุด
    n_jobs=-1,  # ใช้ทุก CPU cores
)

# ทำการ fit ข้อมูล
grid_search.fit(X_train_resampled, y_train_resampled)
```

```
Fitting 5 folds for each of 96 candidates, totalling 480 fits
Best parameters: {'C': 5.0, 'class_weight': None, 'dual': True, 'loss': 'hinge', 'tol': 0.0001}
Best score: 0.9535699684778418
Best SVM Accuracy: 0.8737
Classification Report for Best SVM:
              precision    recall  f1-score   support

    negative       0.86      0.78      0.82        41
     neutral       0.68      0.63      0.66        30
    positive       0.92      0.96      0.94       127

    accuracy                           0.87       198
   macro avg       0.82      0.79      0.80       198
weighted avg       0.87      0.87      0.87       198
```

ภาพประกอบ 14 การทำ hyper-parameter tuning

จากการดำเนินการด้วย Grid Search พบว่าพารามิเตอร์ที่ให้ประสิทธิภาพดีที่สุดคือ C=5.0, loss='hinge', class_weight=None, dual=True, และ tol=1e-4 โดยให้ค่า F1-weighted score สูงสุดที่ 0.9536 และความแม่นยำโดยรวม (accuracy) ที่ 0.8737 การใช้ค่า C=5.0 แสดงว่าโมเดลต้องการระดับ regularization ที่ไม่เข้มงวดมากเพื่อจับรูปแบบที่ซับซ้อนในข้อมูล ขณะที่การใช้ loss='hinge' ซึ่งเป็นฟังก์ชันสูญเสียมาตรฐานของ SVM ให้ผลลัพธ์ที่ดีกว่า 'squared_hinge' ในกรณีของข้อมูลนี้ ส่วนการที่ class_weight=None ให้ผลดีที่สุดแสดงว่าการปรับน้ำหนักคลาสไม่ช่วยปรับปรุงประสิทธิภาพเนื่องจากมีการปรับความสมดุลของคลาสไปด้วยวิธี SMOTE ไปก่อนหน้านี้

จากผลการประเมินโมเดล SVM ที่ได้รับการปรับแต่งพารามิเตอร์ พบว่าประสิทธิภาพในการจำแนกแต่ละคลาสดีขึ้น โดยคลาส Positive มีประสิทธิภาพสูงสุดด้วยค่า Precision 0.92, Recall 0.96 และ F1-score 0.94 เนื่องจากมีข้อมูลตัวอย่างมากที่สุด (127 ตัวอย่าง) และมีลักษณะเฉพาะที่ชัดเจน ขณะที่คลาส Negative มีประสิทธิภาพปานกลางด้วยค่า Precision 0.86, Recall 0.78 และ F1-score 0.82 จากข้อมูล 41 ตัวอย่าง

อย่างไรก็ตาม คลาส Neutral มีประสิทธิภาพต่ำที่สุดด้วยค่า Precision 0.68, Recall 0.63 และ F1-score 0.66 จากข้อมูลเพียง 30 ตัวอย่าง ซึ่งเป็นผลมาจากการมีข้อมูลน้อยที่สุดและลักษณะของความคิดเห็นเป็นกลางที่อาจมีความคล้ายคลึงกับคลาสอื่นๆ ทำให้โมเดลมีความยากลำบากในการจำแนกประเภทนี้ ค่า Weighted Average ที่ได้คือ Precision 0.87, Recall 0.87 และ F1-score 0.87 ซึ่งสะท้อนประสิทธิภาพโดยรวมของโมเดลที่คำนึงถึงสัดส่วนของข้อมูลในแต่ละคลาส

3.2.5 การวิเคราะห์หัวข้อ (Topic Modeling)

การวิเคราะห์หัวข้อเป็นเทคนิคในการค้นหาโครงสร้างความสัมพันธ์เชิงความหมายที่แฝงอยู่ภายในชุดข้อมูลเอกสารขนาดใหญ่ งานวิจัยนี้ใช้อัลกอริทึม LDA (Latent Dirichlet Allocation) ซึ่งเป็นโมเดลความน่าจะเป็นเชิงสถิติที่สามารถระบุกลุ่มของคำที่มักปรากฏร่วมกันเป็นหัวข้อ (topic) โดยแต่ละเอกสารประกอบด้วยหลายหัวข้อในสัดส่วนที่แตกต่างกัน

3.2.5.1 การจัดเตรียมข้อมูลสำหรับการวิเคราะห์หัวข้อ

การประมวลผลข้อมูลเข้าสู่ระบบวิเคราะห์ LDA ต้องการการจัดรูปแบบข้อมูลให้เหมาะสม รวมถึงการดำเนินการทำความสะอาดข้อมูลเพิ่มเติมที่ปรับแต่งเฉพาะสำหรับงานวิเคราะห์หัวข้อ ดังแสดงในภาพประกอบ 15

```
    # คำที่อาจพบในหลายหัวข้อและไม่ช่วยจำแนกหัวข้อ
    'impression', 'benefit', 'direct', 'effect', 'life', 'factor', 'compared',
    'various', 'school', 'involves', 'term', 'subject',

    # เพิ่มคำอื่นๆ ที่พบในผลการวิเคราะห์แต่ไม่มีความหมายเฉพาะเจาะจง
    'review', 'point', 'allowing', 'choosing', 'workload', 'modernity', 'slow',
    'experience', 'university'
    ])
    # รวมกับคำหยุดมาตรฐาน
    stop_words = set(stopwords.words('english'))
    if custom_stop_words:
        stop_words.update(custom_stop_words)
    # Set up lemmatizer
    lemmatizer = WordNetLemmatizer()
    def clean_and_tokenize(text):
        """Clean and tokenize a single text document"""
        if not isinstance(text, str) or text.strip() == '':
            return []
        # Convert to lowercase and remove punctuation
        text = re.sub(r"\W+", " ", text.lower())
        # Tokenize
        tokens = word_tokenize(text)
        # Remove stopwords and lemmatize
        tokens = [lemmatizer.lemmatize(w) for w in tokens if w not in stop_words and len(w) > 2]
        return tokens

    # Apply preprocessing
    print("กำลังตัดคำและทำความสะอาดข้อความ...")
    processed_df["Cleaned_Tokens"] = processed_df[text_column].apply(clean_and_tokenize)
    processed_df["Cleaned_Review"] = processed_df["Cleaned_Tokens"].apply(lambda x: " ".join(x))
```

ภาพประกอบ 15 การทำความสะอาดข้อมูลเพิ่มเติมเฉพาะสำหรับการวิเคราะห์หัวข้อ

สำหรับการวิเคราะห์หัวข้อ มีการกำหนดรายการคำหยุด (stop words) เพิ่มเติมที่เฉพาะเจาะจงกับบริบทของงานวิจัยนี้ เพื่อกรองคำที่ไม่มีความหมายเฉพาะเจาะจง นอกจากนี้ยัง

ใช้ WordNetLemmatizer ในการลดรูปคำเพื่อรวมคำที่มีรากศัพท์เดียวกัน หลังจากนั้น จึงสร้าง Document-Term Matrix สำหรับการวิเคราะห์หัวข้อด้วย LDA ดังแสดงในภาพประกอบ 16

```python
def create_document_term_matrix(processed_df, min_df=5, max_df=0.7, max_features=5000):

    print("\nกำลังสร้าง Document-Term Matrix...")

    # Create vectorizer
    vectorizer = CountVectorizer(
        max_features=max_features,
        min_df=min_df,
        max_df=max_df
    )

    # Transform documents to document-term matrix
    doc_term_matrix = vectorizer.fit_transform(processed_df['Cleaned_Review'])

    # Get feature names
    feature_names = vectorizer.get_feature_names_out()
```

ภาพประกอบ 16 การเตรียมข้อมูลสำหรับ LDA

3.2.6 การค้นหาจำนวนหัวข้อที่เหมาะสม

สำหรับการสร้างโมเดล LDA ในขั้นตอนสุดท้าย งานวิจัยนี้ได้ทำการทดลองเพื่อกำหนดจำนวนหัวข้อที่เหมาะสมผ่านการใช้เกณฑ์วัดผล 2 วิธี 1. Perplexity คือ วัดความสามารถในการทำนายข้อมูลใหม่ (ค่ายิ่งต่ำยิ่งดี) และ 2. Coherence Score คือ วัดความสอดคล้องภายในของหัวข้อ (ค่ายิ่งสูงยิ่งดี) ดังแสดงในภาพประกอบ 17

```python
# Train LDA model
lda_model = LatentDirichletAllocation(
    n_components=k,
    random_state=42,
    max_iter=50,
    learning_method='online',
    learning_decay=0.7,
    evaluate_every=10,
    n_jobs=-1
)

lda_model.fit(doc_term_matrix)
models.append(lda_model)

# Calculate perplexity
perplexity = lda_model.perplexity(doc_term_matrix)
perplexity_scores.append(perplexity)

# Calculate coherence
topic_words = []
for topic in lda_model.components_:
    top_words = [feature_names[i] for i in topic.argsort()[:-11:-1]]
    topic_words.append(top_words)

cm = CoherenceModel(
    topics=topic_words,
    texts=token_lists,
    dictionary=id2word,
    coherence='c_v'
)

coherence = cm.get_coherence()
coherence_scores.append(coherence)
```

ภาพประกอบ 17 การทดสอบเพื่อหาจำนวนหัวข้อที่เหมาะสม

หลังจากได้ค่า Coherence Score และ Perplexity สำหรับจำนวนหัวข้อต่างๆ แล้ว งานวิจัยนี้ได้วิเคราะห์เพื่อเลือกจำนวนหัวข้อที่เหมาะสมโดยพิจารณาจากทั้งสองเกณฑ์ร่วมกัน ดังแสดงในภาพประกอบ 18

```
def analyze_model_quality(coherence_scores, perplexity_scores, best_k_coherence,
    # วิเคราะห์และเปรียบเทียบผลลัพธ์
    # ...
    # สรุปจำนวนหัวข้อที่เหมาะสม
    if abs(best_k_coherence - best_k_perplexity) <= 2:
        optimal_topics = min(best_k_coherence, best_k_perplexity)
    else:
        optimal_topics = best_k_coherence

    return optimal_topics
```

ภาพประกอบ 18 การสรุปหัวข้อที่เหมาะสม

3.2.6.1 การสร้างโมเดล LDA และการตีความหัวข้อ

เมื่อได้จำนวนหัวข้อที่เหมาะสมแล้ว จึงสร้างโมเดล LDA ด้วยพารามิเตอร์ ดังแสดงในภาพประกอบ 19

```
def create_final_lda_model(doc_term_matrix, optimal_num_topics):

    print(f"\nกำลังสร้างโมเดลสุดท้ายที่มี {optimal_num_topics} หัวข้อ...")

    final_lda_model = LatentDirichletAllocation(
        n_components=optimal_num_topics,
        random_state=42,
        max_iter=100,  # More iterations for final model
        learning_method='online',
        learning_decay=0.7,
        n_jobs=-1
    )

    final_lda_model.fit(doc_term_matrix)
    print("สร้างโมเดลสุดท้ายเสร็จสิ้น")

    return final_lda_model
```

ภาพประกอบ 19 การสร้างโมเดล LDA

โมเดล LDA จะแสดงคำสำคัญในแต่ละหัวข้อพร้อมค่าน้ำหนัก ซึ่งสามารถนำมาตีความและกำหนดชื่อหัวข้อได้ตามบริบทของข้อมูล ดังแสดงในภาพประกอบ 20

```python
def extract_topic_keywords(final_lda_model, feature_names, num_keywords=15):

    print("\nคำสำคัญของแต่ละหัวข้อ:")
    topic_keywords = []

    for topic_idx, topic in enumerate(final_lda_model.components_):
        # Sort keywords by weight and get top n
        top_keywords_idx = topic.argsort()[:-num_keywords-1:-1]
        top_keywords = [feature_names[i] for i in top_keywords_idx]
        topic_keywords.append(top_keywords)

        # Display topic keywords
        print(f"\nหัวข้อที่ {topic_idx+1}:")
        print(", ".join(top_keywords))

    return topic_keywords
```

ภาพประกอบ 20 การแสดงคำสำคัญในแต่ละหัวข้อพร้อมค่าน้ำหนัก

3.2.6.2 การระบุหัวข้อหลักให้กับแต่ละความคิดเห็น

หลังจากสร้างโมเดล LDA แล้ว งานวิจัยนี้ได้ระบุหัวข้อหลักให้กับแต่ละความคิดเห็นโดยพิจารณาจากค่าความน่าจะเป็นสูงสุด ดังแสดงในภาพประกอบ 21

```python
def assign_topics_to_documents(final_lda_model, doc_term_matrix, processed_df, topic_names=None):

    print("\nกำลังจัดสรรหัวข้อให้กับเอกสาร...")

    # Transform documents to get topic distribution
    doc_topic_dist = final_lda_model.transform(doc_term_matrix)

    # Create a copy of the dataframe
    df_topics = processed_df.copy()

    # Assign dominant topic and probability
    df_topics['Dominant_Topic'] = doc_topic_dist.argmax(axis=1)
    df_topics['Dominant_Topic_Prob'] = doc_topic_dist.max(axis=1)

    # Save full topic distribution for each document
    for i in range(final_lda_model.n_components):
        df_topics[f'Topic_{i+1}_Prob'] = doc_topic_dist[:, i]

    # Assign topic names if provided
    if topic_names:
        # Adjust topic_names if necessary
        if len(topic_names) < final_lda_model.n_components:
            topic_names.extend([f"หัวข้อที่ {i+1}" for i in range(len(topic_names), final_lda_model.n_components)])
        elif len(topic_names) > final_lda_model.n_components:
            topic_names = topic_names[:final_lda_model.n_components]

        df_topics['Topic_Name'] = df_topics['Dominant_Topic'].apply(lambda x: topic_names[x])
    else:
        df_topics['Topic_Name'] = df_topics['Dominant_Topic'].apply(lambda x: f"หัวข้อที่ {x+1}")

    return df_topics
```

ภาพประกอบ 21 การระบุหัวข้อหลักให้กับแต่ละความคิดเห็น

3.2.6.3 การศึกษาความเชื่อมโยงระหว่างหัวข้อและความรู้สึก

เพื่อให้เข้าใจความสัมพันธ์ระหว่างหัวข้อและความรู้สึก งานวิจัยนี้ได้วิเคราะห์การกระจายตัวของความรู้สึกในแต่ละหัวข้อ ดังแสดงในภาพประกอบ 22

```python
def analyze_topic_sentiment(df_topics, original_df):
    # ตรวจสอบว่ามีข้อมูลความรู้สึกหรือไม่
    if 'sentiment' not in original_df.columns:
        print("⚠️ ไม่พบข้อมูลความรู้สึก (sentiment) ในชุดข้อมูล")
        return

    # สร้าง DataFrame ใหม่โดยรวมข้อมูลหัวข้อและความรู้สึก
    df_analysis = df_topics.copy()
    df_analysis['sentiment'] = original_df.loc[df_topics.index, 'sentiment'].values

    # 1. สร้างตารางความถี่ระหว่างหัวข้อและความรู้สึก
    topic_sentiment = pd.crosstab(df_analysis['Topic_Name'], df_analysis['sentiment'])

    # คำนวณร้อยละของแต่ละความรู้สึกในแต่ละหัวข้อ
    topic_sentiment_pct = topic_sentiment.div(topic_sentiment.sum(axis=1), axis=0) * 100

    # คำนวณคะแนนความรู้สึกเฉลี่ยของแต่ละหัวข้อ
    sentiment_score_map = {'positive': 1, 'neutral': 0, 'negative': -1}
    df_analysis['sentiment_score'] = df_analysis['sentiment'].map(sentiment_score_map)
    topic_sentiment_score = df_analysis.groupby('Topic_Name')['sentiment_score'].agg(['me

    return topic_sentiment_score
```

ภาพประกอบ 22 การตรวจนับและประมวลผลร้อยละของความเห็นในแต่ละกลุ่ม

3.3 การแสดงผลข้อมูล

การแสดงผลข้อมูลเป็นส่วนสำคัญที่ช่วยสื่อสารผลการวิจัยให้เข้าใจง่ายและชัดเจน ในงานวิจัยนี้ได้ใช้เทคนิค Visualization หลายรูปแบบดังนี้

1. แผนภูมิแท่ง (Bar Chart) ใช้สำหรับแสดงการเปรียบเทียบเชิงปริมาณ เช่น การกระจายของความรู้สึกในข้อมูล และการกระจายตัวของหัวข้อในความคิดเห็นทั้งหมด

2. แผนภูมิเส้น (Line Chart) ประยุกต์ใช้เพื่อแสดงผลค่า Coherence Score ในจำนวนหัวข้อแตกต่างกัน เพื่อกำหนดจำนวนหัวข้อที่เหมาะสมที่สุดสำหรับการวิเคราะห์ LDA

3. แผนภูมิความร้อน (Heatmap) ใช้สำหรับแสดง Confusion Matrix ของโมเดลวิเคราะห์ความรู้สึก ทำให้เห็นรายละเอียดของการทำนายในแต่ละประเภทความรู้สึก

4. WordCloud ใช้สำหรับแสดงคำที่ปรากฏบ่อยในความรู้สึกเชิงบวกและเชิงลบ โดยขนาดของคำแสดงถึงความถี่ที่ปรากฏในข้อมูล

การสร้างภาพข้อมูลทั้งหมดใช้ไลบรารี Matplotlib และ Seaborn ในภาษา Python เพื่อให้ได้ภาพที่มีคุณภาพสูงและปรับแต่งได้ตามต้องการ

3.4 การประเมินผล

โมเดลที่สร้างขึ้นถูกประเมินด้วยชุดข้อมูลทดสอบเพื่อวัดความสามารถในการทำนาย โดยใช้ตัวชี้วัดต่างๆ ดังนี้

1. Accuracy วัดสัดส่วนของการทำนายที่ถูกต้องทั้งหมด ซึ่งให้ภาพรวมของประสิทธิภาพโมเดล

2. Precision วัดระดับความแม่นยำในการทำนายแต่ละคลาส (สัดส่วนของผลทำนายในคลาสนั้นที่มีความถูกต้อง) ซึ่งมีความสำคัญต่อการระบุความรู้สึกที่เป็นจริง

3. Recall วัดความครบถ้วนของการทำนายแต่ละคลาส (สัดส่วนของกรณีจริงในคลาสนั้นที่ถูกทำนายถูกต้อง) ซึ่งสำคัญในการระบุปัญหาที่ต้องได้รับการแก้ไข

4. F1-Score ค่าเฉลี่ยแบบฮาร์โมนิกระหว่าง Precision กับ Recall ซึ่งสะท้อนมุมมองที่เป็นสมดุลของความสามารถในการทำงานของโมเดล

5. Confusion Matrix แสดงรายละเอียดของการทำนายในแต่ละประเภทความรู้สึก ทำให้เห็นจุดแข็งและจุดอ่อนของโมเดลในการจำแนกแต่ละประเภท

การวิเคราะห์หัวข้อผ่าน LDA การกำหนดจำนวนหัวข้อที่เหมาะสมเป็นขั้นตอนสำคัญที่ส่งผลต่อคุณภาพของการวิเคราะห์ งานวิจัยนี้ใช้เกณฑ์การประเมิน 2 ตัวชี้วัดหลักในการกำหนดจำนวนหัวข้อที่เหมาะสม ได้แก่ Coherence Score และ Perplexity โดยมีรายละเอียดดังนี้

1. Coherence Score เป็นตัวชี้วัดความสอดคล้องเชิงความหมายภายในแต่ละหัวข้อ โดยประเมินว่าคำที่อยู่ในหัวข้อเดียวกันมีความเกี่ยวข้องกันมากน้อยเพียงใด การคำนวณ Coherence Score ใช้หลักการวัดความถี่ของการปรากฏร่วมกันของคำในเอกสารเดียวกัน ยิ่งคำในหัวข้อหนึ่งปรากฏร่วมกันบ่อยในเอกสารต่างๆ แสดงว่าหัวข้อนั้นมีความสอดคล้องสูง ในงานวิจัยนี้ใช้ Coherence Score แบบ 'c_v' ซึ่งเป็นมาตรฐานที่นิยมใช้กันอย่างแพร่หลาย

2. Perplexity เป็นตัวชี้วัดความสามารถของโมเดลในการทำนายข้อมูลใหม่ โดยคำนวณจากค่า log-likelihood ของข้อมูลทดสอบ ค่า Perplexity ที่ต่ำกว่าแสดงว่าโมเดลสามารถทำนายข้อมูลได้ดีกว่า อย่างไรก็ตาม การใช้ Perplexity เพียงอย่างเดียวอาจไม่เพียงพอ เนื่องจากค่าที่ต่ำอาจไม่ได้หมายความว่าหัวข้อที่ได้มีความหมายที่สามารถตีความได้ จึงจำเป็นต้องใช้ร่วมกับ Coherence Score

3. การหาจำนวนหัวข้อที่เหมาะสม ในงานวิจัยนี้ได้ทำการทดสอบจำนวนหัวข้อตั้งแต่ 2 ถึง 8 หัวข้อ โดยคำนวณค่า Coherence Score และ Perplexity สำหรับแต่ละจำนวนหัวข้อ จากนั้นพิจารณาจุดที่ให้ค่า Coherence Score สูงสุดและ Perplexity ต่ำสุด ในกรณีที่สองค่าขัดแย้งกันจะให้น้ำหนักมากกว่ากับ Coherence Score เนื่องจากสะท้อนคุณภาพของหัวข้อที่สามารถตีความได้

บทที่ 4
ผลการวิจัย

การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์ทัศนคติของนักศึกษาต่อบริการมหาวิทยาลัยผ่านเทคนิคการประมวลผลภาษาธรรมชาติซึ่งเป็นเทคโนโลยีปัญญาประดิษฐ์ บทนี้นำเสนอผลการวิจัยตามวัตถุประสงค์ที่กำหนด ซึ่งครอบคลุมการวิเคราะห์ความรู้สึกจากความคิดเห็นของนักศึกษาและการวิเคราะห์หัวข้อสำคัญ โดยแบ่งการนำเสนอออกเป็น 4 ส่วนหลัก ดังนี้

1. ผลการวิเคราะห์ข้อมูลเบื้องต้น
2. ผลการวิเคราะห์ความรู้สึกด้วย VADER
3. ผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดลด้วยวิธีการที่แตกต่างกัน
4. ผลการศึกษาหัวข้อและความสัมพันธ์กับด้านความรู้สึก

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

4.1.1 ลักษณะของข้อมูล

ผู้ดำเนินการวิจัยได้รวบรวมข้อมูลความคิดเห็นผ่านแพลตฟอร์มประเมินความพึงพอใจทางออนไลน์ของมหาวิทยาลัยตั้งแต่ปีการศึกษา 2564 ถึง 2566 โดยมีจำนวนข้อความความคิดเห็นทั้งหมด 1,000 รายการ ข้อมูลถูกเก็บในรูปแบบไฟล์ CSV ที่ประกอบด้วยข้อความความคิดเห็น (Review) และข้อมูลอื่นๆ ที่เกี่ยวข้อง ในการเตรียมข้อมูลเบื้องต้น ผู้วิจัยได้ตรวจสอบภาษาที่ใช้ในความคิดเห็นพบว่าข้อมูลมีทั้งภาษาไทยและภาษาอังกฤษ จึงได้ทำการแปลความคิดเห็นภาษาไทยเป็นภาษาอังกฤษด้วย Google Translate API เพื่อให้สามารถวิเคราะห์ด้วยเครื่องมือที่รองรับภาษาอังกฤษได้อย่างมีประสิทธิภาพ

จากการตรวจสอบคุณภาพข้อมูล ไม่พบค่าที่ขาดหาย (Missing Values) ในชุดข้อมูลดังนี้

4.1.2 การแปลข้อมูลและการตรวจสอบภาษา

ผู้วิจัยได้ใช้ LangDetect เพื่อตรวจสอบภาษาของแต่ละความคิดเห็น จากนั้นได้ทำการแปลความคิดเห็นภาษาไทยทั้งหมดเป็นภาษาอังกฤษโดยใช้ Google Translate API

ผลการตรวจสอบภาษาหลังการแปลพบว่า 14 แถวของข้อมูลที่ไม่สามารถแปลเป็นภาษาอังกฤษ ซึ่งจะทำการลบแถว ที่ไม่เป็นภาษาอังกฤษทั้งหมดเพื่อทำให้ชุดข้อมูลพร้อมสำหรับการวิเคราะห์ต่อไป

4.1.3 การจัดเตรียมข้อมูลสำหรับการวิเคราะห์

ผู้ดำเนินการวิจัยได้ทำการปรับปรุงข้อมูลให้เหมาะสมต่อการวิเคราะห์ ตามขั้นตอนดังต่อไปนี้

1. การทำความสะอาดข้อมูล (Data Cleaning)
    1) ลบแถวที่มีค่าว่าง
    2) ตรวจสอบและแก้ไขความซ้ำซ้อนของข้อมูล
    3) แก้ไขชื่อคอลัมน์ให้เป็นมาตรฐาน
2. การประมวลผลข้อความ (Text Preprocessing)
    1) แบ่งข้อความเป็นคำ (Tokenization) ด้วย NLTK word_tokenize
    2) ลบคำหยุด (Stop Words) ที่ไม่มีความหมายเชิงวิเคราะห์

ตัวอย่างข้อมูลหลังการเตรียมข้อมูลแสดงในตาราง 2

ตาราง 2 การแสดงตัวอย่างข้อมูลก่อนและหลังการเตรียมข้อมูล

Translated_Review	Cleaned_Review
Not satisfied with the cafeteria that is small	not satisfied cafeteria small
Compliment the classroom Is a story that stude...	compliment classroom Is story students always ...
Requesting documents is very slow and complica...	requesting documents slow complicated When ask...
Not suitable for computer room	not suitable computer room

จากกระบวนการเตรียมข้อมูล คณะผู้วิจัยได้ข้อมูลที่มีความพร้อมสำหรับการดำเนินการวิเคราะห์ความรู้สึกและการวิเคราะห์หัวข้อในขั้นตอนถัดไป

4.2 ผลการวิเคราะห์ความรู้สึกด้วย VADER

ผู้ดำเนินการวิจัยได้เลือกใช้เครื่องมือ VADER (Valence Aware Dictionary and Sentiment Reasoner) ซึ่งเป็นระบบวิเคราะห์อารมณ์ที่เหมาะสมสำหรับการประมวลผลข้อความในสื่อสังคมออนไลน์และความคิดเห็นทั่วไป ในการวิเคราะห์ความรู้สึกเบื้องต้นจากข้อความแสดงความคิดเห็นของนักศึกษา

4.2.1 ผลการจำแนกความรู้สึก

ผลการศึกษาความรู้สึกโดยใช้ VADER นำเสนอการกระจายของอารมณ์ที่ปรากฏในข้อความความคิดเห็น ดังแสดงในตาราง 3

ตาราง 3 การจำแนกประเภทความรู้สึกด้วย VADER

ประเภทความรู้สึก	จำนวน	ร้อยละ
Positive	632	64.1%
Neutral	148	15.0%
Negative	206	20.9%
รวม	986	100%

จากตารางที่ 3 สามารถสังเกตได้ว่าข้อความความคิดเห็นเกินครึ่งมีลักษณะเชิงบวก (64.1%) รองลงมาคือความรู้สึกเชิงลบ (20.9%) และความรู้สึกเป็นกลาง (15.0%) ตามลำดับ ซึ่งแสดงให้เห็นว่านักศึกษาส่วนใหญ่มีความรู้สึกเชิงบวกต่อการให้บริการของมหาวิทยาลัย

การเลือกค่า threshold ที่ 0.05 สำหรับการจำแนกความรู้สึกด้วย VADER เป็นผลจากการทดลองกับค่า threshold หลายค่า (0.05, 0.1, 0.2) ดังแสดงในตาราง 4

ตาราง 4 การเปรียบเทียบผลการจำแนกความรู้สึกที่ threshold ต่างๆ

Threshold	Positive	Neutral	Negative
0.05	632	148	206
0.10	620	170	196
0.20	594	221	171

จากตารางที่ 4 พบว่าเมื่อเพิ่มค่า threshold การจำแนกความรู้สึกเป็นกลางมีแนวโน้มเพิ่มขึ้น ในขณะที่การจำแนกความรู้สึกเชิงบวกและเชิงลบลดลง โดยที่ threshold 0.05 ให้การกระจายของความรู้สึกที่สมดุล

4.2.2 ตัวอย่างข้อความแสดงความคิดเห็นและคะแนนความรู้สึก

แสดงตัวอย่างข้อความแสดงความคิดเห็นในแต่ละประเภทความรู้สึกพร้อมคะแนนที่ได้จากการวิเคราะห์ด้วย VADER ดังแสดงในตาราง 5

ตาราง 5 การแสดงตัวอย่างข้อความคิดเห็นและคะแนนความรู้สึกด้วยVADAR

ประเภทความรู้สึก	ข้อความแสดงความคิดเห็น	คะแนนความรู้สึก			
		Positive	Negative	Neutral	Compound
Positive	Impressed with fast and friendly service	0.612	0.0	0.388	0.743
Positive	Feel good about student data systems Is something that students want a lot	0.296	0.0	0.704	0.4939
Neutral	The application process is quite complex. But can still understand	0.0	0.0	1.0	0.0
Neutral	May have to develop the modernity of the course In terms of access to teachers	0.0	0.0	1.0	0.0
Negative	Not impressed with the library	0.0	0.39	0.61	-0.3724
Negative	Very bad about signals, especially at the library.	0.0	0.351	0.649	-0.5849

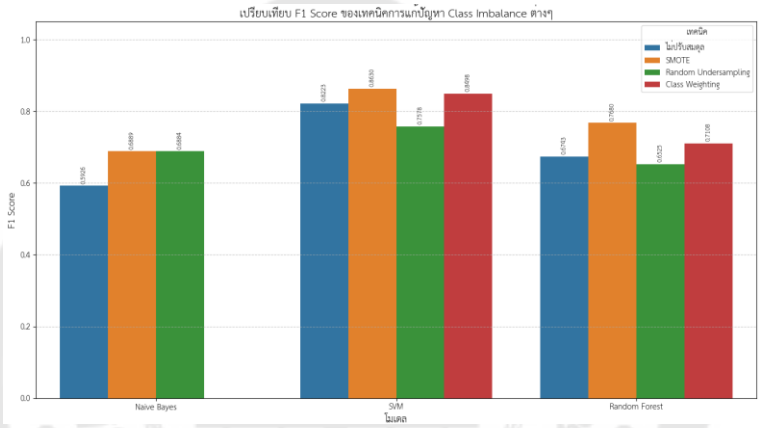
จากตารางที่ 5 แสดงให้เห็นว่า VADER สามารถให้คะแนนความรู้สึกที่สอดคล้องกับเนื้อหาของข้อความแสดงความคิดเห็น โดยข้อความที่มีคำที่แสดงความรู้สึกเชิงบวกอย่างชัดเจน เช่น "Impressed", "Feel good" จะได้คะแนน Compound สูง ในขณะที่ข้อความที่มีคำที่แสดงความรู้สึกเชิงลบ เช่น "Not impressed", "Very bad" จะได้คะแนน Compound ต่ำ

4.3 ผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดลด้วยวิธีการที่แตกต่างกัน

4.3.1 การเปรียบเทียบวิธีการแก้ไขปัญหา Class Imbalance

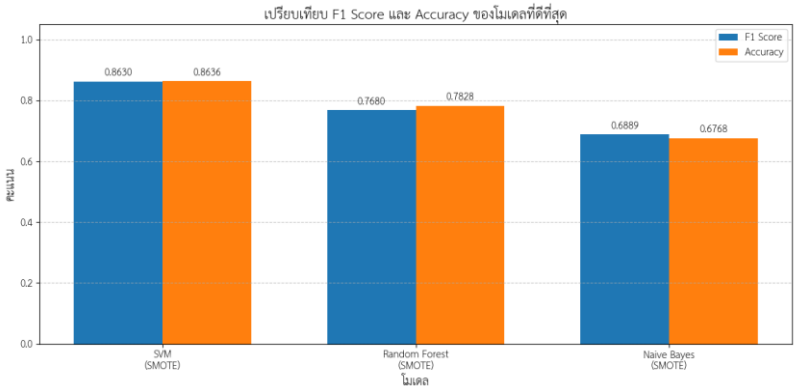
เนื่องจากข้อมูลมีความไม่สมดุล (Class Imbalance) งานวิจัยนี้ได้ทดลองใช้เทคนิคการแก้ปัญหาความไม่สมดุลของคลาส 3 แบบ ได้แก่ 1. Oversampling ด้วย SMOTE เพิ่มจำนวนตัวอย่างในคลาสที่มีน้อย 2. Undersampling ด้วย Random Under Sampling ลดจำนวนตัวอย่างในคลาสที่มีมาก 3. Class Weighting กำหนดน้ำหนักให้แต่ละคลาสตามสัดส่วนที่แตกต่างกันเปรียบเทียบทุกโมเดลโดยแบ่งตามคุณลักษณะและวัดผลด้วยค่า F1-Score

1. การสกัดคุณลักษณะสำคัญจากข้อความด้วยวิธี TF-IDF ดังแสดงในภาพประกอบ 23



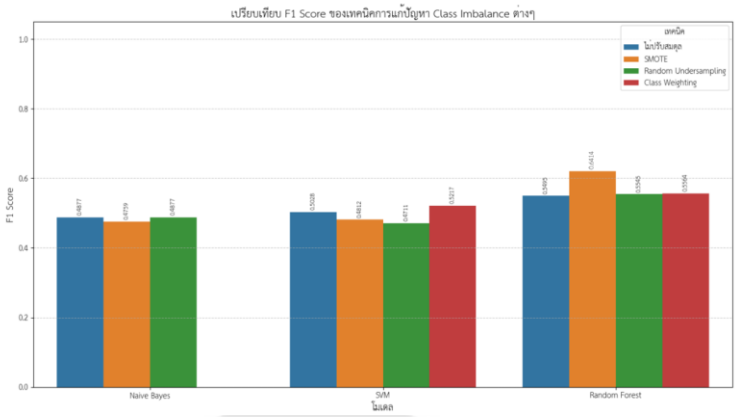
ภาพประกอบ 23 การเปรียบเทียบ F1-Score การแก้ปัญหา Class Imbalance (TF-IDF)

จากการเปรียบเทียบพบว่า SMOTE ให้ผลดีที่สุดในการปรับสมดุลเมื่อเปรียบเทียบกับเทคนิคอื่นๆ ดังแสดงในภาพประกอบ 24



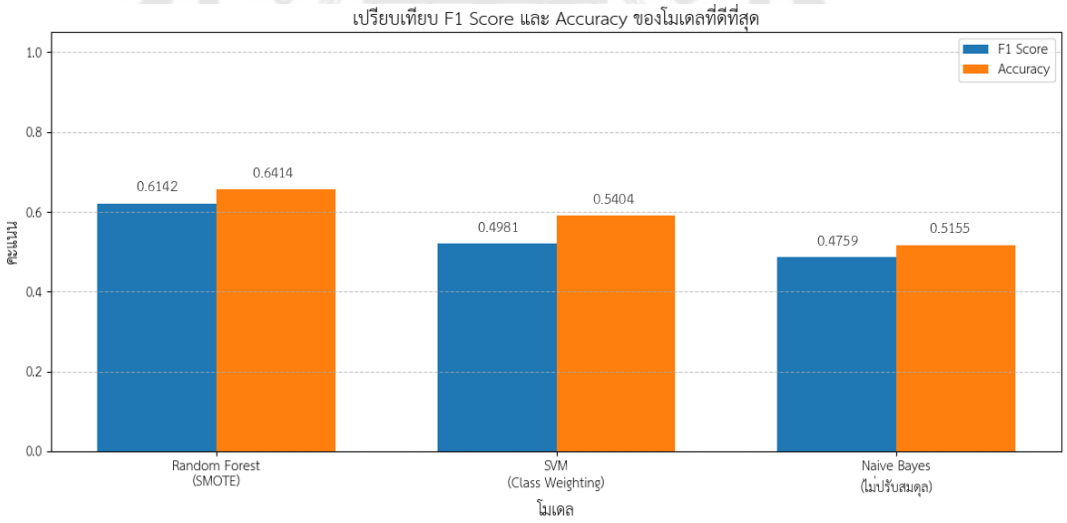
ภาพประกอบ 24 การเปรียบเทียบเทคนิคค่า F1-Scoreและ Accuracy (TF-IDF)

2. การสกัดคุณลักษณะสำคัญจากข้อความด้วยวิธี Word2Vec แสดงในภาพประกอบ 25



ภาพประกอบ 25 การเปรียบเทียบ F1-Score การแก้ปัญหา Class Imbalance (Word2Vec)

จากการเปรียบเทียบ พบว่า SMOTE ให้ผลดีที่สุดในการปรับสมดุลเมื่อเปรียบเทียบกับเทคนิคอื่นๆ ดังแสดงในภาพประกอบ 26



ภาพประกอบ 26 การเปรียบเทียบเทคนิคค่า F1-ScoreและAccuracy (Word2Vec)

ผลการเปรียบเทียบประสิทธิภาพของเทคนิคการแก้ปัญหา Class Imbalance ทั้ง 3 วิธี กับแบบจำลองการเรียนรู้ของเครื่อง 3 แบบ ได้แก่ Naive Bayes, SVM และ Random Forest โดยใช้ค่า F1-Score ในการประเมิน ผู้วิจัยได้เลือกการสกัดคุณลักษณะสำคัญจากข้อความด้วยวิธี TF-IDF ซึ่งเป็นวิธีที่มีประสิทธิภาพสูงสุด ดังแสดงในตาราง 6

ตาราง 6 การเปรียบเทียบค่า F1-Score ของเทคนิคการแก้ปัญหา Class Imbalance

โมเดล	ไม่ปรับสมดุล	SMOTE	Random Undersampling	Class Weighting
Naive Bayes	0.5926	0.6889	0.6884	-
SVM	0.8223	0.8630	0.7578	0.8498
Random Forest	0.6743	0.7680	0.6525	0.7108

จากตารางที่ 6 พบว่าเทคนิค SMOTE มีค่า F1-Score ที่สูง ซึ่งโมเดล SVM ที่ประยุกต์ใช้ SMOTE แสดงประสิทธิภาพที่ดีที่สุดด้วยค่า 0.8630 ในขณะที่ Random Forest ที่ใช้ SMOTE มีประสิทธิภาพรองลงมาที่ 0.7680

4.3.2 การเปรียบเทียบเทคนิคการสกัดคุณลักษณะสำคัญจากข้อความ

งานวิจัยนี้ได้ดำเนินการศึกษาวิธีการแยกคุณลักษณะสำคัญจากข้อความ 2 รูปแบบ ประกอบด้วย TF-IDF (Term Frequency-Inverse Document Frequency) และ Word2Vec เพื่อประเมินและเปรียบเทียบความสามารถร่วมกับแบบจำลองการเรียนรู้ของเครื่องต่างๆ ผลการเปรียบเทียบปรากฏดังนี้ ดังแสดงในตาราง 7

ตาราง 7 การเปรียบเทียบประสิทธิภาพระหว่าง TF-IDF และ Word2Vec (SMOTE)

เทคนิคการสกัดคุณลักษณะ	อัลกอริทึม	Accuracy	F1-Score
TF-IDF	Naive Bayes	0.6768	0.6889
TF-IDF	SVM	0.8636	0.8630
TF-IDF	Random Forest	0.7828	0.7680
Word2Vec	Naive Bayes	0.5505	0.5055
Word2Vec	SVM	0.5404	0.4981
Word2Vec	Random Forest	0.6414	0.6142

จากตารางที่ 7 พบว่าการสกัดคุณลักษณะสำคัญจากข้อความด้วยวิธี TF-IDF ให้ผลลัพธ์ดีกว่า Word2Vec โดยโมเดลที่มีประสิทธิภาพสูงที่สุดคือ SVM ที่ใช้ TF-IDF (Accuracy = 0. 8636, F1-Score = 0. 8630) รองลงมาคือ Random Forest ที่ใช้ TF-IDF (Accuracy = 0. 7828, F1-Score = 0. 7680)

จากการเปรียบเทียบ TF-IDF อาจเหมาะสมกว่า Word2Vec สำหรับข้อมูลนี้

1. ข้อมูลความคิดเห็นมักมีคำเฉพาะที่บ่งบอกความรู้สึกชัดเจน (เช่น 'excellent', 'Impressed')

2. TF-IDF ให้น้ำหนักกับคำที่พบบ่อยในเอกสารหนึ่งแต่พบน้อยในเอกสารอื่น ซึ่งเหมาะกับการจำแนกความรู้สึก

3. Word2Vec อาจต้องการข้อมูลมากกว่านี้เพื่อสร้างเวกเตอร์ของคำที่มีความหมาย

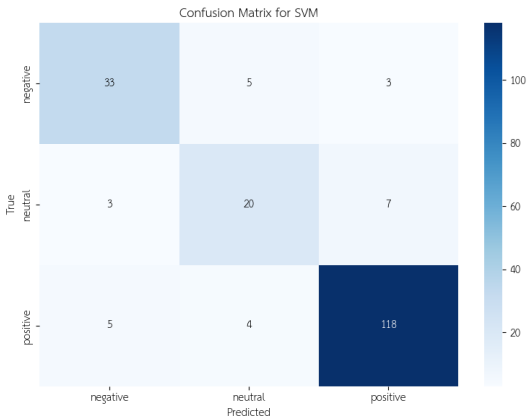
4.3.3 การเปรียบเทียบโมเดลการจำแนกความรู้สึก

จากการดำเนินการศึกษาโมเดลการเรียนรู้ของเครื่อง 3 แบบ ได้แก่ Naive Bayes, SVM และ Random Forest ผลการสำรวจเปรียบเทียบประสิทธิภาพแบบรายละเอียดของแต่ละแบบจำลองโดยนำเทคนิค SMOTE ร่วมกับการดึงคุณลักษณะสำคัญจากข้อความผ่านวิธี TF-IDF ดังแสดงในตาราง 8

ตาราง 8 การเปรียบเทียบความสามารถของโมเดล (TF-IDF + SMOTE)

โมเดล	Accuracy	F1- weighted
Naive Bayes	0.6768	0.6889
SVM	0.8636	0.8630
Random Forest	0.7828	0.7680

จากตารางที่ 8 พบว่าโมเดล SVM มีประสิทธิภาพสูงที่สุดในทุกเมตริก ตามด้วย Random Forest และ Naive Bayes ตามลำดับ ซึ่งสอดคล้องกับสมมติฐานของงานวิจัยที่คาดว่า SVM จะมีประสิทธิภาพดีในการจำแนกข้อความ รายละเอียดผลการจำแนกความรู้สึกของโมเดล SVM ซึ่งมีประสิทธิภาพดีที่สุดแสดงใน Confusion Matrix ดังแสดงในภาพประกอบ 27



ภาพประกอบ 27 การแสดงผล Confusion Matrix of SVM

4.3.4 การวิเคราะห์ความผิดพลาดจาก Confusion Matrix และแนวทางปรับปรุงโมเดล

การศึกษารูปแบบข้อผิดพลาดที่ปรากฏใน confusion matrix ถือเป็นกระบวนการที่มีความจำเป็นอย่างยิ่งเพื่อให้เกิดความเข้าใจถึงขีดจำกัดในการทำงานของแบบจำลองและผลที่ตามมาเมื่อนำไปประยุกต์ใช้ในสภาพแวดล้อมจริง ดังที่ปรากฏในภาพประกอบ 28 ซึ่งแสดงลักษณะของข้อผิดพลาดที่เกิดขึ้นระหว่างการจัดประเภทอารมณ์ความรู้สึกทั้ง 3 กลุ่ม (เชิงบวก เป็นกลาง และเชิงลบ) การวิจัยนี้ได้ทำการวิเคราะห์ข้อผิดพลาดที่สำคัญ 5 ประเภท โดยนำเสนอกรณีตัวอย่าง ปัจจัยที่ก่อให้เกิดข้อผิดพลาด ผลกระทบเมื่อนำไปใช้ในทางปฏิบัติ ซึ่งมีรายละเอียดดังนี้

4.3.4.1 ความผิดพลาดในการจำแนกความรู้สึกเชิงลบเป็นความรู้สึกเป็นกลาง (False Neutral for Negative)

ตัวอย่างข้อความที่เกิดข้อผิดพลาดประเภทนี้ คือ "ระบบ WiFi ใช้งานได้ปกติในบางพื้นที่ แต่บางจุดยังมีปัญหา" ซึ่งโมเดลทำนายเป็น "เป็นกลาง" แต่ที่จริงแล้วเป็น "เชิงลบ" สาเหตุหลักเกิดจากข้อความมีทั้งส่วนบวก ("ใช้งานได้ปกติ") และส่วนลบ ("มีปัญหา") ทำให้โมเดลเกิดความสับสน

ผลกระทบต่อการนำไปใช้งาน อาจส่งผลให้สถาบันการศึกษาอาจละเลยประเด็นปัญหาที่สำคัญของผู้เรียน เพราะระบบวิเคราะห์ไม่สามารถระบุความไม่พอใจที่แฝงอยู่ในข้อความที่ปรากฏเป็นความคิดเห็นแบบกลางๆ ได้ ทำให้ปัญหาและส่งผลเสียต่อชื่อเสียงขององค์กรในอนาคตได้

แนวทางการแก้ไขมีดังนี้ ประการแรก เพิ่มข้อมูลฝึกสอนที่มีลักษณะผสมระหว่างข้อความเชิงบวกและเชิงลบ ประการที่สอง พัฒนา feature ที่จับความหนักแน่นของข้อความเชิงลบที่อยู่ในประโยคเดียวกับข้อความเชิงบวก และประการสุดท้าย ปรับค่าถ่วงน้ำหนักให้คำที่แสดงปัญหามีผลต่อการจำแนกมากขึ้น

4.3.4.2 ความผิดพลาดในการจำแนกความรู้สึกเชิงบวกเป็นความรู้สึกเป็นกลาง (False Neutral for Positive)

ตัวอย่างข้อความที่เกิดข้อผิดพลาด คือ "ห้องเรียนสะอาดและมีอุปกรณ์พร้อมใช้งาน" ซึ่งโมเดลทำนายเป็น "เป็นกลาง" แต่ที่จริงแล้วเป็น "เชิงบวก" สาเหตุเกิดจากข้อความใช้คำบรรยายทั่วไปโดยไม่มีคำแสดงความรู้สึกเชิงบวกที่ชัดเจน

ผลกระทบต่อการนำไปใช้งาน ข้อผิดพลาดในลักษณะนี้ทำให้สถาบันการศึกษาพลาดโอกาสในการค้นหาจุดเด่น อาจส่งผลให้ไม่สามารถนำข้อได้เปรียบดังกล่าวไปใช้ประโยชน์ในการสื่อสารเชิงการตลาด อีกทั้งยังอาจส่งผลให้การวัดผลความสำเร็จของแผนงานพัฒนาต่างๆ

แนวทางการแก้ไขประกอบด้วย การเพิ่ม feature ที่จับบริบทและความหมายแฝงของคำ การปรับ threshold ในการจำแนกระหว่างคลาสเป็นกลางกับคลาสเชิงบวกให้เหมาะสมยิ่งขึ้น และการเพิ่มข้อมูลฝึกสอนที่มีการแสดงความรู้สึกเชิงบวกโดยนัย

4.3.4.3 ความผิดพลาดในการจำแนกความรู้สึกเป็นกลางเป็นความรู้สึกเชิงบวก (False Positive for Neutral)

ตัวอย่างข้อความที่เกิดข้อผิดพลาด คือ "กระบวนการลงทะเบียนมีขั้นตอนปกติตามระเบียบ" ซึ่งโมเดลทำนายเป็น "เชิงบวก" แต่ที่จริงแล้วเป็น "เป็นกลาง" สาเหตุเกิดจากคำว่า "ปกติ" มักปรากฏในข้อความเชิงบวก ทำให้โมเดลเกิดอคติ นอกจากนี้ ข้อความดังกล่าวเป็นการบอกเล่าข้อเท็จจริงที่ไม่มีความคิดเห็นแทรกอยู่

ผลกระทบต่อการนำไปใช้งาน ข้อผิดพลาดดังกล่าวอาจก่อให้เกิดการตีความที่คลาดเคลื่อนโดยผู้กำกับดูแล ซึ่งอาจเข้าใจว่าผู้เรียนมีระดับความพอใจที่ดี แม้ว่าในความเป็นจริงผู้เรียนอาจมีความรู้สึกแบบไม่แสดงออกหรือรับได้แต่ไม่ประทับใจ การวิเคราะห์ที่ผิดพลาดนี้อาจส่งผลต่อการวางแผนที่ไม่เหมาะสม อาทิ การลดระดับความสำคัญของการพัฒนาระบบการลงทะเบียน หรือการโอนงบประมาณไปใช้ในส่วนอื่นที่อาจมีความจำเป็นน้อยกว่า

แนวทางการแก้ไขประกอบด้วย การปรับ threshold ในการจำแนกระหว่างคลาสเป็นกลางกับคลาสเชิงบวกให้เหมาะสมยิ่งขึ้น การพัฒนา feature ที่แยกแยะระหว่างข้อเท็จจริงกับความคิดเห็น และการเพิ่มข้อมูลฝึกสอนที่เป็นข้อความบอกเล่าข้อเท็จจริงในคลาสเป็นกลางให้มากขึ้น

4.3.4.4 ความผิดพลาดในการจำแนกความรู้สึกเชิงลบเป็นความรู้สึกเชิงบวก (False Positive for Negative)

ตัวอย่างข้อความที่เกิดข้อผิดพลาด คือ "ดีมากที่ทำให้นักศึกษารอนาน" ซึ่งโมเดลทำนายเป็น "เชิงบวก" แต่ที่จริงแล้วเป็น "เชิงลบ" สาเหตุเกิดจากประโยคแสดงการประชดประชัน ซึ่งโมเดลไม่สามารถจับนัยได้ นอกจากนี้ คำว่า "ดีมาก" มีน้ำหนักเชิงบวกสูง ทำให้โมเดลให้ความสำคัญกับส่วนนี้

ผลกระทบต่อการนำไปใช้งาน ข้อผิดพลาดประเภทนี้ถือว่ามีความร้ายแรงสูงสุด เพราะทำให้สถาบันเกิดความเข้าใจที่ตรงกันข้ามอย่างสิ้นเชิง โดยคิดว่าผู้เรียนมีความพอใจ แท้จริงแล้วผู้เรียนกำลังสื่อสารความไม่พอใจอย่างรุนแรงด้วยวิธีการเสียดสีหรือประชดประชัน ประเด็นวิกฤตที่ผู้เรียนต้องการให้ดำเนินการแก้ไขโดยเร่งด่วนอาจถูกละเลย ซึ่งจะทำให้ระดับความไม่พอใจขยายตัวมากขึ้น และอาจพัฒนาไปสู่การแสดงออกในรูปแบบอื่น เช่น การร้องเรียนผ่านช่องทางที่เป็นทางการ หรือการสร้างกระแสความคิดเห็นเชิงลบบนแพลตฟอร์มสื่อสังคม

แนวทางการแก้ไขประกอบด้วย การพัฒนา feature สำหรับตรวจจับประโยคประชดประชัน การใช้ deep learning models ที่สามารถเข้าใจบริบทได้ดีกว่า เช่น BERT หรือ Thai-RoBERTa และการเพิ่มข้อมูลฝึกสอนที่มีการประชดประชันให้มากขึ้น

4.3.4.5 ความผิดพลาดในการจำแนกความรู้สึกเป็นกลางเป็นความรู้สึกเชิงลบ (False Negative for Neutral)

ตัวอย่างข้อความที่เกิดข้อผิดพลาด คือ "ห้องสมุดเปิดทุกวันจันทร์ถึงศุกร์" ซึ่งโมเดลทำนายเป็น "เชิงลบ" แต่ที่จริงแล้วเป็น "เป็นกลาง" สาเหตุเกิดจากข้อความเป็นข้อเท็จจริงล้วนๆ ไม่มีการแสดงความคิดเห็น และขาดบริบทว่าเวลาเปิดดังกล่าวดีหรือไม่ดี

ผลกระทบต่อการนำไปใช้งาน การวิเคราะห์ที่คลาดเคลื่อนในกรณีนี้อาจก่อให้เกิดความกังวลเกินความจำเป็น โดยผู้ดูแลระบบอาจเข้าใจผิดว่าผู้เรียนมีความไม่พอใจต่อช่วงเวลาการให้บริการของห้องสมุด แม้ว่าในความเป็นจริงเป็นเพียงการแจ้งข้อมูลเท่านั้น ซึ่งอาจส่งผลให้เกิดการใช้งบประมาณเพื่อแก้ไขสิ่งที่ไม่ใช่ปัญหาจริง

แนวทางการแก้ไขประกอบด้วย การเพิ่มข้อมูลฝึกสอนที่เป็นข้อเท็จจริงในคลาสเป็นกลาง การพัฒนา feature ที่แยกแยะระหว่างประโยคบอกเล่าข้อเท็จจริงกับประโยคแสดงความคิดเห็น และการปรับค่า class_weight ให้คลาสเป็นกลางมีน้ำหนักมากขึ้น เนื่องจากมีตัวอย่างน้อย

4.3.5 บทสรุปผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดล

จากผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดลที่สามารถสรุปได้ดังนี้

1. เทคนิคการแก้ปัญหา Class Imbalance: SMOTE ให้ผลดีที่สุดสำหรับทุกโมเดลที่ทดสอบ

2. เทคนิคการสกัดคุณลักษณะสำคัญจากข้อความด้วยวิธี TF-IDF ให้ผลดีกว่า Word2Vec

3. โมเดลการเรียนรู้ของเครื่องแบบ SVM มีประสิทธิภาพสูงสุด ตามลำดับด้วย Random Forest และ Naive Bayes

โมเดลนี้จะได้รับการนำไปประยุกต์ใช้ในการวิเคราะห์ความรู้สึกของนักศึกษาเกี่ยวกับการให้บริการของมหาวิทยาลัยในขั้นตอนถัดไป เพื่อค้นหาหัวข้อที่นักศึกษาให้ความใส่ใจและมีทัศนคติเชิงบวกหรือเชิงลบ ซึ่งจะเป็นแนวทางในการพัฒนาคุณภาพการให้บริการของมหาวิทยาลัยในอนาคต

4.4 ผลการวิเคราะห์หัวข้อและความสัมพันธ์กับความรู้สึก

เพิ่มเติมจากการวิเคราะห์ความรู้สึก ผู้ดำเนินการวิจัยได้ทำการศึกษาหัวข้อ (Topic Modeling) เพื่อค้นหาประเด็นหลักที่ปรากฏในข้อความแสดงความเห็นของนักศึกษา ผ่านการใช้อัลกอริทึม LDA (Latent Dirichlet Allocation)

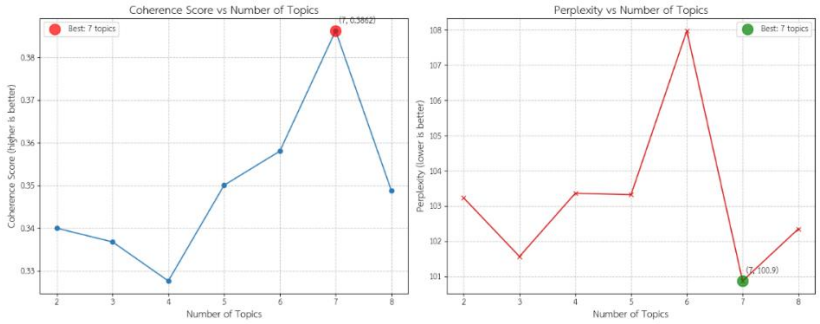
4.4.1 การระบุจำนวนหัวข้อที่เหมาะสม

การคัดเลือกปริมาณหัวข้อที่พอเหมาะนับเป็นกระบวนการที่มีนัยสำคัญในการดำเนินการวิเคราะห์หัวข้อด้วยเทคนิค LDA ทั้งนี้เพราะปริมาณหัวข้อที่มากหรือน้อยเกินความพอดีจะกระทบต่อคุณภาพการจัดหมวดหมู่หัวข้อ ผู้วิจัยจึงดำเนินการทดสอบแบบจำลอง LDA โดยกำหนดปริมาณหัวข้อตั้งแต่ 2 จนถึง 8 หัวข้อ พร้อมทั้งวัดผลสัมฤทธิ์ของแบบจำลองผ่านดัชนี Coherence Score และค่า Perplexity อันเป็นมาตรวัดที่ได้รับการยอมรับในการตรวจสอบคุณภาพแบบจำลองการวิเคราะห์หัวข้อ ดังปรากฏในตาราง 9

ตาราง 9 การแสดงค่า Coherence Score และ Perplexity ที่จำนวนหัวข้อต่างๆ

จำนวนหัวข้อ	Coherence Score	Perplexity
2	0.3400	103.23
3	0.3368	101.57
4	0.3276	103.36
5	0.3501	103.32
6	0.3580	107.97
7	0.3862	100.87
8	0.3488	102.35

เพื่อแสดงให้เห็นภาพการเปลี่ยนแปลงของดัชนีชี้วัดทั้งสองอย่างเด่นชัด ผู้วิจัยได้จัดทำการนำเสนอผลในลักษณะกราฟเส้น ดังปรากฏในภาพประกอบ 28



ภาพประกอบ 28 การแสดงค่า Coherence Score และ Perplexity ที่จำนวนหัวข้อต่างๆ

การพิจารณาและคัดเลือกปริมาณหัวข้อที่พอเหมาะ จากข้อมูลในตารางที่ 9 และภาพประกอบ 28 สามารถวิเคราะห์ได้ดังนี้

1. ดัชนี Coherence Score ค่าดังกล่าวสะท้อนถึงระดับความเชื่อมโยงของคำศัพท์ภายในหัวข้อเดียวกัน ค่าที่สูงขึ้นหมายถึงคำศัพท์ในหัวข้อนั้นมีความเกี่ยวข้องกัน จากผลการศึกษาพบว่าดัชนี Coherence Score แสดงความผันแปรในช่วงปริมาณหัวข้อ 2-6 หัวข้อ โดยค่าสูงที่สุดปรากฏที่ปริมาณ 7 หัวข้อ (0.3862) ซึ่งโดดเด่นกว่าจุดอื่นๆ อย่างเห็นได้ชัดเจนและเมื่อเพิ่มปริมาณเป็น 8 หัวข้อCoherence Score ปรับตัวลดลง

2. ดัชนี Perplexity ค่านี้ใช้วัดแบบจำลองในการคาดการณ์ข้อมูล โดยค่าที่ต่ำลงบ่งชี้ถึงประสิทธิภาพที่ดีขึ้น ผลการศึกษาแสดงให้เห็นว่าค่า Perplexity มีเสถียรภาพในช่วงปริมาณ 2-5 หัวข้อ (101.57-103.36) และเกิดการเพิ่มขึ้นอย่างผิดปกติที่ปริมาณ 6 หัวข้อ (107.97) หลังจากนั้นค่า Perplexity จะกลับมาต่ำที่สุดที่ปริมาณ 7 หัวข้อ (100.87)

เหตุผลสนับสนุนการคัดเลือกปริมาณหัวข้อ ผู้วิจัยตัดสินใจเลือกใช้ปริมาณ 7 หัวข้อสำหรับการวิเคราะห์ โดยมีเหตุผลประกอบดังนี้

1. ความเป็นเลิศของดัชนีชี้วัดทั้งคู่ เป็นจุดเดียวที่ทั้งดัชนี Coherence Score และค่า Perplexity แสดงผลลัพธ์ที่ดีเยี่ยมไปพร้อมกัน ซึ่งสะท้อนว่าแบบจำลองมีความสามารถในการจัดหมวดหมู่หัวข้ออย่างมีนัยสำคัญและมีความเที่ยงตรง

2. ความแจ่มชัดในการจำแนกประเด็น ปริมาณ 7 หัวข้อช่วยให้สามารถแบ่งแยกมุมมองต่างๆ ในข้อคิดเห็นของผู้เรียนได้อย่างละเอียดในระดับที่พอดี ไม่มากเกินไปจนสับสนหรือน้อยเกินไปจนผสมปนเปประเด็นที่ต่างกันเข้าด้วยกัน

3. ความเหมาะสมกับบริบทสถาบันอุดมศึกษา ปริมาณ 7 หัวข้อสามารถครอบคลุมมิติสำคัญของการบริการในสถาบันอุดมศึกษา อันประกอบด้วย สิ่งอำนวยความสะดวกพื้นฐาน สภาพแวดล้อม กระบวนการเอกสาร กิจกรรมเสริมหลักสูตร แหล่งเรียนรู้ กระบวนการเรียนการสอน และบริการสนับสนุนทั่วไป

4. การป้องกันภาวะ Overfitting แม้ว่าการขยายปริมาณหัวข้อเป็น 8 หรือมากกว่านั้นอาจช่วยลดค่า Perplexity ได้บ้าง แต่ดัชนี Coherence Score ที่ลดต่ำลงชี้ให้เห็นว่าเริ่มมีการแบ่งแยกหัวข้อที่ละเอียดเกินความจำเป็นจนสูญเสียความหมายที่ชัดเจน

ด้วยเหตุนี้ การตัดสินใจเลือกใช้ปริมาณ 7 หัวข้อจึงเป็นการดำเนินการที่อ้างอิงจากข้อมูลเชิงประจักษ์และมีความสอดรับกับหลักการทางทฤษฎีของการวิเคราะห์หัวข้อที่ระบุว่า ปริมาณหัวข้อที่เหมาะควรจะแสดงค่าดัชนี Coherence Score ในระดับสูงและค่า Perplexity ในระดับต่ำไปพร้อมๆ กัน อันจะทำให้ผลการวิเคราะห์หัวข้อมีความน่าเชื่อถือและสามารถนำไปประยุกต์ใช้ในการยกระดับคุณภาพการบริการของสถาบันอุดมศึกษาได้อย่างมีประสิทธิผล

4.4.2 ผลการวิเคราะห์หัวข้อด้วย LDA

ผู้วิจัยได้สร้างโมเดล LDA จำนวน 7 หัวข้อ พร้อมแสดงหัวข้อที่สำคัญด้วย WordCloud โดยแยกเป็นแต่ละหัวข้อ ดังแสดงในภาพประกอบ 29

ภาพประกอบ 29 การแสดงหัวข้อที่สำคัญแยกตามหัวข้อด้วย WordCloud

จากการวิเคราะห์ข้อความความคิดเห็นของนักศึกษาผ่านอัลกอริทึม LDA สามารถแยกประเภทได้ 7 หัวข้อสำคัญ โดยหัวข้อแต่ละหัวข้อประกอบด้วยกลุ่มคำสำคัญที่มีความเชื่อมโยงกันทางด้านความหมาย ผู้ดำเนินการวิจัยได้ดำเนินการตีความและกำหนดชื่อให้กับแต่ละหัวข้อที่ปรากฏในตาราง 10

ตาราง 10 การกำหนดชื่อแต่ละหัวข้อ

หัวข้อ	ชื่อ	คำสำคัญ
1	โครงสร้างพื้นฐานด้านเทคโนโลยีและพื้นที่การใช้งาน	make, computer, room, application, living, laboratory, place, printer, wifi, safety, sport, open, highlight, signal, causing
2	สภาพแวดล้อมการเรียนรู้และสิ่งอำนวยความสะดวก	learning, cafeteria, bathroom, equipment, enter, atmosphere, study, sport, equal, criterion, place, failure, application, admirable, clear
3	ระบบเอกสารและบริการการเงิน	document, work, fee, pay, registration, complicated, understand, request, take, officer, staff, answer, reporting, tuition, confused
4	กิจกรรมนักศึกษาและการพัฒนา	activity, sport, volunteer, camp, academic, student, capital, work, event, term, service, internet, fund, younger, standard
5	สภาพแวดล้อมทางกายภาพและการบริการ	affect, service, parking, image, area, cleanliness, system, landscape, building, resting, learning, conditioning, air, classroom, advice
6	ทรัพยากรการเรียนรู้และเครือข่าย	student, system, library, club, network, data, story, contest, shop, copy, really, consulting, highlight, printer, place
7	การเรียนการสอนและหลักสูตร	teaching, teacher, exam, textbook, section, document, university, medium, compliment, course, term, disadvantage, advantage, matter, used

4.4.3 การกระจายของหัวข้อในข้อความแสดงความคิดเห็น

หลังจากระบุหัวข้อหลักแล้ว การวิจัยนี้ได้ประเมินการกระจายของหัวข้อในความคิดเห็นของนักศึกษา โดยจัดสรรหัวข้อหลักให้แต่ละความคิดเห็นตามค่าความน่าจะเป็นที่สูงที่สุด ผลการวิเคราะห์แสดงให้เห็นว่าข้อมูลความคิดเห็นทั้งหมด 986 รายการมีการกระจายตัวไปยัง 7 หัวข้อหลัก ดังแสดงในตาราง 11

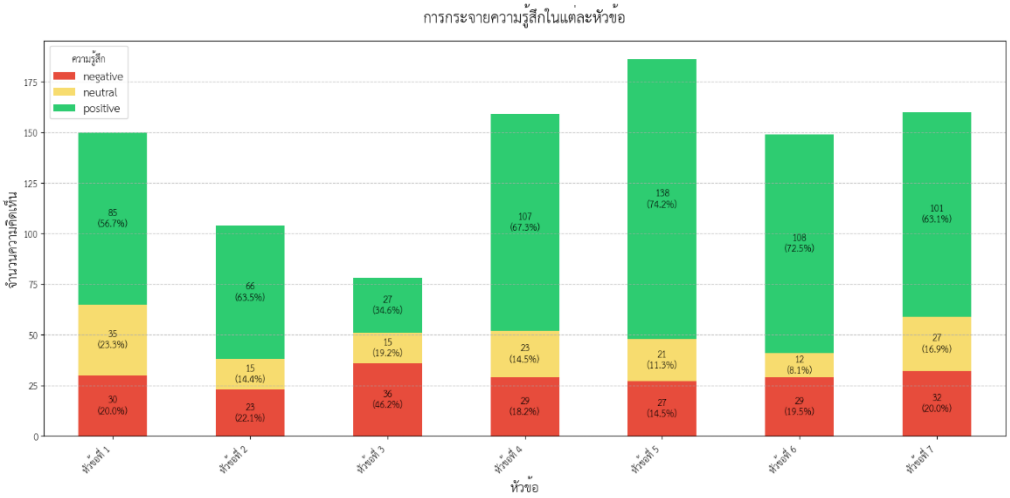
ตาราง 11 การกระจายของหัวข้อในข้อความแสดงความคิดเห็น

หัวข้อ	**จำนวน**	**ร้อยละ**
1	150	15.21%
2	104	10.54%
3	78	7.91%
4	159	16.12%
5	186	18.86%
6	149	15.11%
7	160	16.22%
รวม	986	100%

จากตารางที่ 11 พบว่าการกระจายของหัวข้อในข้อความแสดงความคิดเห็นพบว่า หัวข้อที่พบมากที่สุดคือ หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ) ร้อยละ 18.86% รองลงมาคือ หัวข้อที่ 7 (การเรียนการสอนและหลักสูตร) ร้อยละ 16.22% และหัวข้อที่ 4 (กิจกรรมนักศึกษาและการพัฒนา) ร้อยละ 16.12% ตามลำดับ ส่วนหัวข้อที่พบน้อยที่สุดคือ หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) ร้อยละ 7.91%

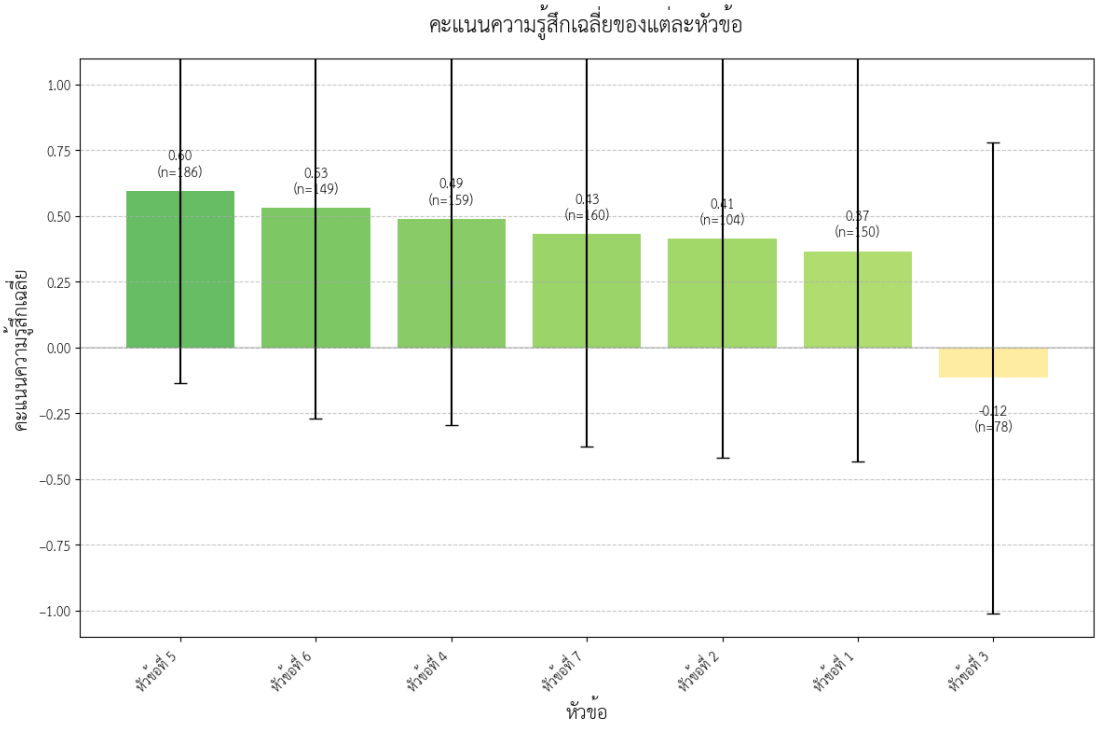
4.4.4 ความสัมพันธ์ระหว่างหัวข้อและความรู้สึก

การวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึกช่วยให้เข้าใจว่านักศึกษามีความรู้สึกอย่างไรต่อแต่ละหัวข้อ ผลการวิเคราะห์แสดงในภาพประกอบ 30



ภาพประกอบ 30 การกระจายความรู้สึกในแต่ละหัวข้อ

การวิเคราะห์ความรู้สึกเฉลี่ยของแต่ละหัวข้อแสดงให้เห็นถึงทัศนคติของนักศึกษาต่อประเด็นต่างๆ ในการให้บริการของมหาวิทยาลัย ดังแสดงในภาพประกอบ 31



ภาพประกอบ 31 การแสดงคะแนนความรู้สึกเฉลี่ยแต่ละหัวข้อ

จากภาพประกอบดังกล่าว พบว่าหัวข้อที่ 5 มีค่าเฉลี่ยความรู้สึกสูงที่สุดที่ 0.60 (n=186) แสดงให้เห็นถึงความรู้สึกเชิงบวกที่แข็งแกร่ง รองลงมาคือหัวข้อที่ 6 ที่มีค่าเฉลี่ย 0.53 (n=149)

และหัวข้อที่ 4 ที่มีค่าเฉลี่ย 0.49 (n=159) ส่วนหัวข้อที่มีความรู้สึกเชิงบวกน้อยที่สุดคือหัวข้อที่ 1 ที่มีค่าเฉลี่ย 0.37 (n=50) และหัวข้อที่ 2 ที่มีค่าเฉลี่ย 0.41 (n=58)

สิ่งที่น่าสังเกตคือหัวข้อที่ 3 เป็นหัวข้อเดียวที่มีค่าเฉลี่ยความรู้สึกเชิงลบที่ -0.12 (n=78) ซึ่งแสดงให้เห็นว่านักศึกษามีความรู้สึกไม่พึงพอใจในประเด็นนี้ ขณะที่หัวข้ออื่นๆ ทั้งหมดแสดงความรู้สึกเชิงบวกในระดับที่แตกต่างกัน การกระจายของข้อมูลในแต่ละหัวข้อ ซึ่งแสดงด้วยแท่งข้อผิดพลาด (Error bars) บ่งชี้ถึงความหลากหลายของความคิดเห็นภายในแต่ละหัวข้อ

ผลการวิเคราะห์คะแนนความรู้สึกในแต่ละหัวข้อแสดงในภาพประกอบที่ 31 และภาพประกอบที่ 32 ซึ่งแสดงสัดส่วนของความรู้สึกเชิงบวก เชิงลบ และเป็นกลางในแต่ละหัวข้อ การวิเคราะห์นี้ได้จำแนกข้อความแสดงความคิดเห็นของนักศึกษาออกเป็น 3 ประเภทความรู้สึกดังนี้

หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) มีสัดส่วนความรู้สึกเชิงลบสูงถึง 26.0% ซึ่งสูงที่สุดเมื่อเทียบกับหัวข้ออื่นๆ ความไม่พึงพอใจนี้อาจเกิดจากความซับซ้อนของระบบเอกสารและขั้นตอนทางการเงิน ซึ่งสอดคล้องกับคำสำคัญที่พบในหัวข้อนี้ เช่น "complicated" และ "understand"

หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ) มีสัดส่วนความรู้สึกเชิงบวกสูงถึง 74.19% ซึ่งสูงที่สุดในบรรดาทุกหัวข้อ ความพึงพอใจในระดับสูงนี้สะท้อนให้เห็นว่านักศึกษาประทับใจกับสภาพแวดล้อมทางกายภาพและการบริการของมหาวิทยาลัย ซึ่งรวมถึงความสะอาด ภูมิทัศน์ และสิ่งอำนวยความสะดวกต่างๆ

4.4.5 นัยสำคัญของผลการวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึก

จากผลการวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึก สามารถสรุปนัยสำคัญที่มีผลต่อการพัฒนาคุณภาพการบริการของมหาวิทยาลัยได้ดังนี้

1. จุดแข็งที่ควรรักษาและพัฒนาต่อ

หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ) เป็นจุดแข็งที่ได้รับความพึงพอใจสูงสุด สะท้อนให้เห็นความสำเร็จของมหาวิทยาลัยในการจัดการสภาพแวดล้อมทางกายภาพและการให้บริการ การรักษามาตรฐานและการพัฒนาต่อยอดในประเด็นนี้จะช่วยเสริมสร้างประสบการณ์ที่ดีให้กับนักศึกษา

หัวข้อที่ 7 (การเรียนการสอนและหลักสูตร) มีสัดส่วนความรู้สึกเชิงบวกในระดับสูง แสดงให้เห็นว่านักศึกษาพึงพอใจกับคุณภาพการเรียนการสอนและหลักสูตรของมหาวิทยาลัย

2. จุดที่ควรปรับปรุงเร่งด่วน

หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) มีสัดส่วนความรู้สึกเชิงลบสูงที่สุดและมีสัดส่วนความรู้สึกเชิงบวกต่ำ ควรได้รับการปรับปรุงอย่างเร่งด่วน โดยเฉพาะในประเด็นความซับซ้อนของระบบและกระบวนการ การปรับปรุงอาจรวมถึงการลดขั้นตอนที่ไม่จำเป็น การพัฒนาระบบออนไลน์ที่ใช้งานง่าย และการฝึกอบรมเจ้าหน้าที่ให้สามารถให้บริการได้อย่างมีประสิทธิภาพมากขึ้น

3. จุดที่ควรตรวจสอบความสม่ำเสมอ

หัวข้อที่ 1 (โครงสร้างพื้นฐานด้านเทคโนโลยีและพื้นที่การใช้งาน) มีสัดส่วนความรู้สึกเป็นกลางสูงที่สุด บ่งชี้ถึงความไม่สม่ำเสมอในคุณภาพ ควรมีการตรวจสอบและปรับปรุงให้มีมาตรฐานเพื่อให้นักศึกษาทุกคนได้รับบริการที่ดีอย่างเท่าเทียมกัน

4. ผลกระทบต่อความพึงพอใจโดยรวม

หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ)และหัวข้อที่ 7 (การเรียนการสอนและหลักสูตร) เป็นหัวข้อที่มีอิทธิพลสูงต่อความพึงพอใจโดยรวมของนักศึกษา เนื่องจากมีทั้งสัดส่วนที่สูงในข้อความแสดงความคิดเห็นและมีสัดส่วนความรู้สึกเชิงบวกสูงการรักษาความพึงพอใจในสองหัวข้อนี้ควบคู่ไปกับการปรับปรุง หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) จะช่วยเพิ่มความพึงพอใจโดยรวมของนักศึกษาได้อย่างมีประสิทธิภาพ

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

งานวิจัย "การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ" มีจุดมุ่งหมายเพื่อศึกษาทัศนคติของนักศึกษาต่อบริการมหาวิทยาลัยและระบุประเด็นสำคัญที่นักศึกษาให้ความสนใจ ผ่านเทคนิค NLP บทนี้เสนอสรุปผลการวิจัย การอภิปราย และข้อเสนอแนะจากการศึกษา โดยจัดเป็น 4 ส่วนหลัก คือ

1. สรุปผลการศึกษาวิจัย
2. ข้อจำกัดของงานวิจัย
3. ข้อเสนอแนะ
4. บทสรุปโดยรวม

5.1 สรุปการศึกษาวิจัย

การศึกษานี้วิเคราะห์ความคิดเห็นของนักศึกษา 1,000 รายการที่รวบรวมจากแพลตฟอร์มประเมินความพึงพอใจออนไลน์ของมหาวิทยาลัยระหว่างปีการศึกษา 2564-2566 ด้วยเทคนิคการประมวลผลภาษาธรรมชาติ ผลการศึกษาสรุปได้ดังต่อไปนี้

5.1.1 ผลการวิเคราะห์ความรู้สึกด้วย VADER

การวิเคราะห์ความรู้สึกด้วย VADER (Valence Aware Dictionary and Sentiment Reasoner) พบว่าข้อความแสดงความคิดเห็นของนักศึกษามีการกระจายของความรู้สึกดังนี้

1. ความรู้สึกเชิงบวก (Positive): 632 รายการ (64.1%)
2. ความรู้สึกเป็นกลาง (Neutral): 148 รายการ (15.0%)
3. ความรู้สึกเชิงลบ (Negative): 206 รายการ (20.9%)

ผลการวิเคราะห์แสดงให้เห็นว่า นักศึกษาส่วนใหญ่มีความรู้สึกด้านบวกต่อการบริการของมหาวิทยาลัยมากกว่าความรู้สึกด้านลบ ทั้งนี้ยังคงมีความคิดเห็นเชิงลบอยู่ที่ร้อยละ 20.9 ซึ่งถือเป็นสัดส่วนที่น่าสนใจและควรนำมาพิจารณาปรับปรุง

5.1.2 ผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดล

การศึกษาได้สร้างโมเดลการเรียนรู้ของเครื่องเพื่อวิเคราะห์ความรู้สึกจากข้อความความเห็น โดยทำการทดลองใช้วิธีการดึงคุณสมบัติ 2 รูปแบบ (TF-IDF และ Word2Vec) ร่วมกับอัลกอริทึมการเรียนรู้ 3 ประเภท (Naive Bayes, SVM, และ Random Forest) ผลการประเมินเปรียบเทียบประสิทธิภาพของโมเดลแสดงให้เห็นว่า

1. การเปรียบเทียบเทคนิคการแก้ปัญหาความไม่สมดุลของข้อมูล (Class Imbalance)

ทดลองใช้เทคนิคการแก้ปัญหาความไม่สมดุลของคลาส 3 วิธี ได้แก่ การเพิ่มข้อมูล (Oversampling) ด้วย SMOTE, การลดข้อมูล (Undersampling) ด้วย Random Under Sampling และการปรับค่าน้ำหนักคลาส (Class Weighting)

พบว่าเทคนิค SMOTE ให้ค่า F1-Score สูงที่สุดสำหรับทุกโมเดล โดยโมเดล SVM ที่ใช้ SMOTE มีประสิทธิภาพสูงสุดที่ 0.863 แบบจำลอง Random Forest ที่ใช้ Word2Vec มีประสิทธิภาพใกล้เคียงกับ SVM ที่ใช้ TF-IDF โดยมีค่า Accuracy 0.835 และ F1-Score 0.836

2. การเปรียบเทียบเทคนิคการสกัดคุณลักษณะสำคัญจากข้อความ

ศึกษาเทคนิคการสกัดคุณลักษณะสำคัญจากข้อความ 2 รูปแบบ คือ TF-IDF (Term Frequency-Inverse Document Frequency) และ Word2Vec

พบว่า TF-IDF ให้ผลลัพธ์ดีกว่า Word2Vec สำหรับทุกโมเดลที่ทดสอบ โดยโมเดล SVM ที่ใช้ TF-IDF มีประสิทธิภาพสูงสุด (Accuracy = 0.8636, F1-Score = 0.8630)

3. การประเมินเปรียบเทียบโมเดลการเรียนรู้ของเครื่อง

ทดสอบโมเดลการเรียนรู้ของเครื่อง 3 แบบ ได้แก่ Naive Bayes, SVM และ Random Forest พบว่าโมเดล SVM มีประสิทธิภาพสูงสุดในทุกเมตริก ตามด้วย Random Forest และ Naive Bayes ตามลำดับ ผู้วิจัยได้เสนอแนวทางปรับปรุงโมเดลโดยรวมดังนี้

1) การเพิ่มคุณภาพและปริมาณข้อมูลฝึกสอน

การปรับปรุงข้อมูลฝึกสอนเป็นปัจจัยสำคัญในการเพิ่มประสิทธิภาพโมเดล โดยเฉพาะการเพิ่มข้อมูลที่มีลักษณะที่ทำให้เกิดความผิดพลาดบ่อย การให้มนุษย์ติดป้ายข้อมูลที่โมเดลทำนายผิดพลาดเพื่อนำกลับมาฝึกโมเดลใหม่ และการสร้างความสมดุลของข้อมูลระหว่างคลาส โดยเฉพาะคลาสเป็นกลางที่มีจำนวนน้อย

2) การปรับปรุง Feature Engineering

การพัฒนา features เฉพาะทางมีความสำคัญต่อการเพิ่มประสิทธิภาพโมเดล ซึ่งรวมถึงการพัฒนา features เฉพาะทางสำหรับตรวจจับการประชดประชัน การสร้าง features ที่สามารถแยกข้อเท็จจริงจากความคิดเห็น และการพัฒนา features ที่จับความขัดแย้งในประโยคเดียวกัน

3) การปรับเปลี่ยนสถาปัตยกรรมโมเดล

การทดลองใช้โมเดลแบบผสมผสาน (ensemble) ร่วมกับ deep learning การใช้โมเดลภาษาขนาดใหญ่ที่ผ่านการ fine-tune ด้วยข้อมูลภาษาไทย และการทดลองใช้เทคนิค transfer learning จากโมเดลภาษาไทยที่ฝึกฝนมาแล้ว สามารถช่วยเพิ่มประสิทธิภาพการทำนายได้

4) การปรับปรุงการประมวลผลภาษาไทย

การพัฒนาระบบวิเคราะห์ภาษาไทยโดยตรงเพื่อลดความผิดพลาดจากการแปล การสร้างคลังคำแสดงความรู้สึกภาษาไทยเฉพาะทางการศึกษา และการพัฒนาระบบตัดคำภาษาไทยที่แม่นยำมากขึ้นสำหรับเอกสารในบริบทการศึกษา เป็นแนวทางที่สำคัญในการปรับปรุงประสิทธิภาพโมเดล

5) การปรับปรุงกระบวนการเลือกโมเดล

การทดลองใช้ scoring metric อื่นๆ นอกจาก f1_weighted เช่น balanced_accuracy การปรับค่า class_weight='balanced' เพื่อให้ความสำคัญกับคลาสเป็นกลางมากขึ้น และการเพิ่มพารามิเตอร์ในการค้นหา เช่น multi_class และ class_weight เพื่อหาค่าที่เหมาะสมมากขึ้น สามารถช่วยปรับปรุงประสิทธิภาพการทำนายของโมเดลได้อย่างมีนัยสำคัญ

5.1.3 ผลการประเมินหัวข้อ

การวิเคราะห์หัวข้อ (Topic Modeling) เป็นเทคนิคการประมวลผลภาษาธรรมชาติที่ใช้ในการค้นหาโครงสร้างความสัมพันธ์เชิงความหมายที่แฝงอยู่ในเอกสารขนาดใหญ่ งานวิจัยนี้ได้ประยุกต์ใช้อัลกอริทึม LDA (Latent Dirichlet Allocation) ซึ่งเป็นโมเดลความน่าจะเป็นเชิงสถิติในการวิเคราะห์ข้อความแสดงความคิดเห็นของนักศึกษา โดยมีสมมติฐานว่าแต่ละเอกสารประกอบด้วยหัวข้อหลายหัวข้อในสัดส่วนที่แตกต่างกัน และแต่ละหัวข้อมีการกระจายตัวของคำที่มีลักษณะเฉพาะ

การกระจายของหัวข้อในข้อความแสดงความคิดเห็นพบว่า หัวข้อที่พบมากที่สุดคือ หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ) ร้อยละ 18.86% รองลงมาคือ หัวข้อที่ 7 (การเรียนการสอนและหลักสูตร) ร้อยละ 16.22% และหัวข้อที่ 4 (กิจกรรมนักศึกษาและการพัฒนา) ร้อยละ 16.12% ตามลำดับ ส่วนหัวข้อที่พบน้อยที่สุดคือ หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) ร้อยละ 7.91%

5.1.4 การวิเคราะห์ความสัมพันธ์ระหว่างหัวข้อและความรู้สึก

หัวข้อที่ 5 (สภาพแวดล้อมทางกายภาพและการบริการ) มีสัดส่วนความรู้สึกเชิงบวกมากที่สุด (74.19%) แสดงให้เห็นว่านักศึกษาพึงพอใจต่อสภาพแวดล้อมทางกายภาพและการบริการของมหาวิทยาลัย

หัวข้อที่ 3 (ระบบเอกสารและบริการการเงิน) มีสัดส่วนความรู้สึกเชิงลบมากที่สุด (46.15%) และมีสัดส่วนความรู้สึกเชิงบวกต่ำ (34.62%) สะท้อนให้เห็นถึงความไม่พึงพอใจของนักศึกษาที่มีต่อระบบเอกสารและบริการการเงินของมหาวิทยาลัย

หัวข้อที่ 1 (โครงสร้างพื้นฐานด้านเทคโนโลยีและพื้นที่การใช้งาน) มีสัดส่วนความรู้สึกเป็นกลางมากที่สุด (23.33%) แสดงให้เห็นความคิดเห็นที่หลากหลายของนักศึกษาต่อโครงสร้างพื้นฐานด้านเทคโนโลยีและพื้นที่การใช้งาน

ผลการวิเคราะห์นี้ชี้ให้เห็นว่า โครงสร้างพื้นฐานด้านเทคโนโลยีเป็นหัวข้อที่นักศึกษาให้ความสำคัญมากที่สุด แต่กลับเป็นหัวข้อที่มีความรู้สึกเชิงลบสูงที่สุด ในขณะที่การบริการนักศึกษาและการเรียนการสอนมีความรู้สึกเชิงบวกในสัดส่วนที่สูง

5.2 ข้อจำกัดงานวิจัย

1. ข้อจำกัดด้านภาษา การศึกษานี้มีการแปลข้อความภาษาไทยเป็นภาษาอังกฤษ ซึ่งอาจทำให้สูญเสียนัยสำคัญบางประการหรือเกิดความคลาดเคลื่อนในการตีความ เนื่องจากแต่ละภาษามีโครงสร้าง ไวยากรณ์ และการใช้คำที่เฉพาะตัว รวมถึงสำนวนหรือคำแสลงที่อาจไม่สามารถแปลได้อย่างตรงตัว

2. ข้อจำกัดด้านแบบจำลองและเครื่องมือ แม้ว่าแบบจำลอง SVM ร่วมกับ TF-IDF จะให้ประสิทธิภาพที่ดี แต่ยังมีข้อจำกัดในการจับความหมายแฝงหรือความหมายเชิงบริบทที่ซับซ้อน เทคนิค LDA ที่ใช้ในการวิเคราะห์หัวข้อก็มีข้อจำกัดในการกำหนดจำนวนหัวข้อที่เหมาะสม และการตีความหัวข้อยังต้องอาศัยการพิจารณาเชิงคุณภาพจากผู้วิจัย

3. ข้อจำกัดด้านขนาดและความหลากหลายของข้อมูล ข้อมูลที่ใช้ในการวิจัยมีจำนวน 1,000 รายการ ซึ่งอาจไม่เพียงพอต่อความคิดเห็นทั้งหมดของนักศึกษานอกจากนี้ข้อมูลยังมาจากระบบประเมินความพึงพอใจออนไลน์เท่านั้น ซึ่งอาจมีอคติจากการเลือกตอบของนักศึกษาที่มีความรู้สึกรุนแรงทั้งในเชิงบวกและเชิงลบ

5.3 ข้อเสนอแนะ

1. การขยายข้อมูลและความหลากหลายของข้อมูล การวิจัยในอนาคตควรเพิ่มขนาดของชุดข้อมูลและรวบรวมข้อมูลจากหลากหลายแหล่ง เช่น แพลตฟอร์มสื่อสังคมออนไลน์ การสัมภาษณ์เชิงลึก หรือแบบสอบถามเฉพาะทาง เพื่อให้ได้ข้อมูลที่ครอบคลุมและหลากหลายมากขึ้น

2. การเพิ่มมิติของการวิเคราะห์ ควรพิจารณาเพิ่มมิติของการวิเคราะห์ เช่น การวิเคราะห์ตามช่วงเวลา (Temporal Analysis) เพื่อติดตามการเปลี่ยนแปลงของความรู้สึกและหัวข้อที่นักศึกษาให้ความสำคัญในแต่ละช่วงเวลา หรือการวิเคราะห์ตามกลุ่มนักศึกษา (เช่น แยกตามคณะ ชั้นปี หรือเพศ) เพื่อให้เข้าใจความต้องการเฉพาะของแต่ละกลุ่ม

3. การวิเคราะห์ข้ามภาษา เนื่องจากข้อมูลเดิมมีทั้งภาษาไทยและภาษาอังกฤษ การวิจัยในอนาคตควรพัฒนาโมเดลที่สามารถวิเคราะห์ข้อความได้ทั้งสองภาษาโดยไม่ต้องผ่านการแปล ซึ่งอาจช่วยลดความผิดพลาดที่เกิดจากการแปลภาษา

4. ควรทดลองใช้โมเดลภาษาแบบ pre-trained ที่ได้รับการฝึกมาเฉพาะสำหรับภาษาไทย เช่น Thai2BERT, DistilBERT หรือ XLM-RoBERTa เพื่อเพิ่มความสามารถในการเข้าใจความรู้สึกเชิงบริบทในภาษาไทยได้ดียิ่งขึ้น

5. ควรมีการปรับเกณฑ์ (threshold) สำหรับการจัดประเภทความรู้สึก หรือพัฒนาโมเดลแบบผสมผสานที่รวมข้อดีของโมเดลเชิงกฎ (rule-based) เช่น VADER และโมเดลเรียนรู้ เช่น SVM หรือ deep learning เพื่อจัดการกับข้อความที่มีความกำกวมได้ดีขึ้น

6. ทดลองโมเดล deep learning อย่าง LSTM และ CNN โดยใช้ embedding เพื่อช่วยให้เข้าใจรูปประโยคและอารมณ์ในระดับลึกมากขึ้น

7. เสนอให้มีการนำผลวิจัยไปพัฒนา dashboard ที่ช่วยให้มหาวิทยาลัยติดตามความคิดเห็นของนักศึกษาแบบต่อเนื่อง และใช้ผลลัพธ์นั้นปรับปรุงบริการแบบเป็นระบบในทุกปีการศึกษา

5.4 บทสรุป

การงานวิจัย "การวิเคราะห์ความรู้สึกของนักศึกษาต่อการให้บริการของมหาวิทยาลัยโดยใช้เทคนิค NLP" แสดงให้เห็นประโยชน์ของการใช้เทคนิคการประมวลผลภาษาธรรมชาติในการวิเคราะห์ทัศนคติของนักศึกษา ทำให้ได้ข้อมูลเชิงลึกสำหรับพัฒนาการบริการของมหาวิทยาลัย

ผลการศึกษาพบว่านักศึกษาส่วนใหญ่มีความรู้สึกเชิงบวกต่อการบริการ 632 รายการ (64.1%) โดยเฉพาะด้านสภาพแวดล้อมทางกายภาพและการบริการมีคะแนนสูงสุด 138 รายการ (74.19%) อย่างไรก็ตาม ยังมีประเด็นต้องปรับปรุง โดยเฉพาะระบบเอกสารและบริการการเงินที่มีความรู้สึกเชิงลบถึง 36 รายการ (46.15%)

จากเทคนิคที่ศึกษา พบว่าแบบจำลอง SVM ร่วมกับ TF-IDF และเทคนิค SMOTE ให้ประสิทธิภาพสูงสุดในการวิเคราะห์ความรู้สึก นอกจากนี้ การวิเคราะห์หัวข้อด้วย LDA ยังช่วยให้สามารถระบุหัวข้อหลักที่นักศึกษาให้ความสำคัญได้ 7 หัวข้อ ซึ่งช่วยให้มหาวิทยาลัยสามารถวางแผนปรับปรุงการให้บริการได้อย่างตรงประเด็น

แม้มีข้อจำกัดบางประการ การศึกษานี้แสดงให้เห็นศักยภาพของเทคโนโลยี NLP ในการวิเคราะห์ข้อมูลเชิงคุณภาพจำนวนมาก ซึ่งสามารถประยุกต์ใช้ในการบริหารสถาบันการศึกษาในอนาคต เช่น การศึกษาความคิดเห็นในสื่อสังคมออนไลน์ การประเมินการเรียนการสอน หรือการติดตามประเด็นที่นักศึกษาให้ความสำคัญ

โดยสรุป การศึกษานี้ไม่เพียงให้ข้อมูลเชิงลึกเกี่ยวกับทัศนคติของนักศึกษา แต่ยังเผยให้เห็นความสำคัญของการใช้เทคโนโลยีปัญญาประดิษฐ์ในการยกระดับคุณภาพการศึกษา ซึ่งก่อประโยชน์ต่อนักศึกษา อาจารย์ บุคลากร และผู้บริหารมหาวิทยาลัย

บรรณานุกรม

Al-Saqqa, S., & Awajan, A. (2019). *The Use of Word2vec Model in Sentiment Analysis: A Survey*. https://doi.org/10.1145/3388218.3388229

Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3).

Almeida, F., & Xexéo, G. (2019). Word embeddings: A survey. *arXiv preprint arXiv:*1901.09069.

Alpaydin, E. (2021). *Machine learning*. MIT press.

Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment analysis: towards a tool for analysing real-time students feedback. 2014 IEEE 26th international conference on tools with artificial intelligence,

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, 63, 743-788.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

Chen, Z., & Liu, B. (2018). *Lifelong machine learning*. Morgan & Claypool Publishers.

Dake, D. K., & Gyimah, E. (2023). Using sentiment analysis to evaluate qualitative students' responses. *Education and Information Technologies*, 28(4), 4629-4647.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453.

Dsouza, D. D., Deepika, D. P. N., Machado, E. J., & Adesh, N. (2019). Sentimental analysis of student feedback using machine learning techniques. *Int. J. Recent Technol. Eng*, 8(14), 986-991.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. kdd,

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61, 863-905.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 463-484.

Goldberg, Y. (2017). *Neural network methods in natural language processing*. Morgan & Claypool Publishers.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1), 5228-5235.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction. In: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. In: Citeseer.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9), 1263-1284.

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in twitter. Proceedings of the first workshop on social media analytics,

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification

evaluations. *International journal of data mining & knowledge management process*, 5(2), 1.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the international AAAI conference on web and social media,

Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences*, 11(9), 3986.

Khaiser, F. K., Saad, A., & Mason, C. (2023). Sentiment analysis of students' feedback on institutional facilities using text-based classification and natural language processing (NLP). *Journal of Language and Communication*, 10(1), 101-111.

Khoshgoftaar, T. M., Golawala, M., & Van Hulse, J. (2007). An empirical study of learning from imbalanced data using random forest. 19th IEEE international conference on tools with artificial intelligence (ICTAI 2007),

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in artificial intelligence*, 5(4), 221-232.

Kumar, A., & Garg, G. (2019). Sentiment analysis of multimodal twitter data. *Multimedia Tools and Applications*, 78, 24103-24119.

Li, Y., & Yang, T. (2018). Word embedding for understanding natural language: a survey. *Guide to big data applications*, 83-104.

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information sciences*, 250, 113-141.

Loria, S. (2018). textblob Documentation. *Release* 0.15, 2(8), 269.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics,

Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*.

MIT press.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:*1301.3781.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. Proceedings of the 2011 conference on empirical methods in natural language processing,

Min, B., Ross, H., Sulem, E., Veyseh, A., Nguyen, T., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2021). *Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey*. https://doi.org/10.48550/arXiv.2111.01243

Naulak, C. (2022). *A comparative study of Naive Bayes Classifiers with improved technique on Text Classification*. https://doi.org/10.36227/techrxiv.19918360.v1

Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, 50(5), 36-43.

Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5, 1-29.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. Proceedings of the eighth ACM international conference on Web search and data mining,

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),

Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep

contextualized word representations. *Psychiatry Research*, 304, 114135.

Shah, G. H., Bhensdadia, C., & Ganatra, A. P. (2012). An empirical evaluation of density-based clustering techniques. *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, 22312307, 216-223.

Solinas, G., Masia, M. D., Maida, G., & Muresu, E. (2012). What really affects student satisfaction? An assessment of quality through a university-wide student survey. *Creative education*, 3(1), 37-40.

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1-34.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune bert for text classification? China national conference on Chinese computational linguistics,

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. Proceedings of the 26th annual international conference on machine learning,

Wen, M., Yang, D., & Rose, C. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us? Educational data mining 2014,

Yao, L., Pengzhou, Z., & Chi, Z. (2019). Research on news keyword extraction technology based on TF-IDF and TextRank. 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS),

Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:*1510.03820.

Zhao, Y., & Karypis, G. (2004). Empirical and theoretical comparisons of selected criterion functions for document clustering. *Machine learning*, 55, 311-331.

ประวัติผู้เขียน