

การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา

โดยใช้เทคนิคเหมืองข้อความ

วรรณรอง ศรีศิริรุ่ง

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

บัณฑิตวิทยาลัย สาขาเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีไทย-ญี่ปุ่น

ปีการศึกษา 2562

THE ANALYSIS OF STUENT EVALUATING TEACHER PERFORMANCE
USING TEXT MINING TECHNIQUE

Waronglong Srisirung

TNI

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology

Graduate School

Thai-Nichi Institute of Technology

Academic Year 2019

หัวข้อวิทยานิพนธ์

การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการ
สอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ

โดย

วรรณกร ศรีศิริรุ่ง

สาขาวิชา

เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม

อาจารย์ที่ปรึกษาวิทยานิพนธ์ (ร่วม)

ดร. ธงชัย แก้วกิริยา

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้บัณฑิตวิทยาลัย
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบริหารธุรกิจ

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. พิชิต สุขเจริญพงษ์)

วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. พยุง มีสัจ)

.....กรรมการ

(รองศาสตราจารย์ ดร. วรากร ศรีเชวงทรัพย์)

.....กรรมการ

(ดร. ณัฐกิตต์ จิตรเอื้อตระกูล)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ (ร่วม)

(ดร. ธงชัย แก้วกิริยา)

วรรณกร ศรีศิริรุ่ง : การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์
โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.
ประจักษ์ เฉิดโฉม อาจารย์ที่ปรึกษาร่วม : ดร. ธงชัย แก้วกิริยา, 73 หน้า.

การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา
โดยใช้เทคนิคเหมืองข้อความ มีวัตถุประสงค์ในการทำวิจัย ดังนี้ 1) เพื่อวิเคราะห์รูปแบบความคิดเห็น
เกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยใช้เทคนิคเหมืองข้อความ 2) เพื่อสังเคราะห์
รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา 3) เพื่อพัฒนาตัวต้นแบบที่
สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา เพื่อระบุปัจจัยที่ส่งผลต่อ
คุณภาพการสอนของอาจารย์ รวมทั้งความคิดเห็นเชิงบวกเชิงลบ และเชิงกลาง จากนั้นได้
ข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์ ในงานวิจัยนี้ดำเนินการโดยรวบรวมข้อมูลจากระบบ
บริการการศึกษาของสถาบันการศึกษาแห่งหนึ่ง จำนวน 3,130 ข้อความ ซึ่งการวิจัยได้ศึกษาการ
วิเคราะห์การจำแนกด้านการประเมิน ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน และ
ด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ นอกจากนี้งานวิจัยนี้มีการ
จำแนกความคิดเห็นแบ่งออกเป็น 3 ด้าน ได้แก่ ความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง จากนั้นทำ
การประเมินประสิทธิภาพการจำแนกข้อความโดยใช้ 3 อัลกอริทึม ได้แก่ Decision Tree , Naïve
Bay และ Random Forest ผลการวิจัยพบว่า การประเมินประสิทธิภาพในการจำแนกโดยใช้
Decision Tree ให้ค่าความถูกต้องในการจำแนกด้านการประเมินมากที่สุด 75.02% และค่าความ
ถูกต้องในการจำแนกความคิดเห็น 75.94% สำหรับประสิทธิภาพในการจำแนกโดยใช้ Naïve Bay มี
ค่าความถูกต้องในการจำแนกด้านการประเมิน 73.42% ซึ่งใกล้เคียงกับ Decision Tree แต่เมื่อ
พิจารณาค่าความถูกต้องในการจำแนกความคิดเห็นมีค่าเป็น 55.40% แต่ไม่สามารถหาค่า
ประสิทธิภาพโดยรวม (F-measure) ในการจำแนกความคิดเห็นเชิงลบได้ ดังนั้นอัลกอริทึมนี้จึงไม่
เหมาะกับการใช้พัฒนาตัวต้นแบบ สำหรับ Random Forest มีค่าความถูกต้องสำหรับการจำแนก
ด้านการประเมิน 66.58% เมื่อพิจารณาค่าความถูกต้องในการจำแนกความคิดเห็นมีค่าเป็น 60.89%
แต่ไม่สามารถหาค่าประสิทธิภาพโดยรวม (F-measure) สำหรับความคิดเห็นเชิงกลางได้ ดังนั้น
อัลกอริทึมนี้จึงไม่เหมาะกับการใช้พัฒนาตัวต้นแบบเช่นกัน

บัณฑิตวิทยาลัย
สาขาวิชา เทคโนโลยีสารสนเทศ
ปีการศึกษา 2562

ลายมือชื่อนักศึกษา
ลายมือชื่ออาจารย์ที่ปรึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษา (ร่วม)

WARONGLONG SRISIRIRUNG : THE ANALYSIS OF STUENT EVALUATING TEACHER PERFORMANCE USING TEXT MINING TECHNIQUE. ADVISOR : ASST. PROF. DR. PRAJAK CHERTCHOM, CO-ADVISOR : DR. THONGCHAI KAEWKIRIYA, 73 PP.

This research's objective is the analysis of student evaluating teacher performance using text mining. The objectives of this research are as follows: 1) to analysis comments' model about student evaluating teacher performance from registrar system using text mining. 2) to synthetic comments' model about student evaluating teacher performance. 3) to develop prototype for analysis teacher performance by setting attribute effect to quality of teaching and then it will suggest the recommendation for teaching development. The 3,130 comments collected by the registrar system of the institute. This research studied the classification analysis of teacher performance that Teacher, Classroom Activity, and Students' Development and the classification analysis of opinion that positive, negative and neutral. Moreover, the research was evaluated by 3 algorithms; Decision Tree, Naïve Bayes and Random Forest. The results show that the decision tree is the most accuracy of attributes classify 75.02% and the most accuracy of comments classify 75.94%. The Naïve Bayes is the accuracy of attributes classify 73.42% that approximate to decision tree's accuracy, while accuracy of comments classify 55.40% but it cannot calculate F-measure for negative comment classify. The Naïve Bay isn't effective for create prototype. The Random Forest is the accuracy of attributes classify 66.58% while accuracy of comments classify 60.89% but it cannot calculate F-measure for neutral comment classify. The Random Forest isn't effective for creating prototype.

Graduate School

Field of Study Information Technology

Academic Year 2019

Student's Signature.....

Advisor's Signature.....

Co-Advisor's Signature.....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์นี้จะสำเร็จลุล่วงไปได้ด้วยดีนั้น ผู้วิจัยขอกราบขอบพระคุณในความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม และ ดร. ธงชัย แก้วกิริยา ซึ่งได้สละเวลาในการให้คำแนะนำ คำปรึกษา การสนับสนุน รวมถึงตรวจสอบข้อบกพร่องทุกขั้นตอนและกำลังใจเต็มเปี่ยมตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้

ทั้งนี้ผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบทุกท่าน ได้แก่ รองศาสตราจารย์ ดร. พยุง มีสัจ รองศาสตราจารย์ ดร. วรากร ศรีเชวงทรัพย์ และ ดร. ณัฐกิตติ์ จิตรเอื้อตระกูล ที่ได้กรุณาให้ข้อเสนอแนะที่เป็นประโยชน์และช่วยตรวจสอบข้อบกพร่องของวิทยานิพนธ์ฉบับนี้ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จและมีความสมบูรณ์

เหนือสิ่งอื่นใดผู้วิจัยขอกราบขอบพระคุณบิดามารดาของผู้วิจัยที่คอยให้กำลังใจ และให้การสนับสนุนในทุกๆ ด้านอย่างที่สุดเสมอมา

ผู้วิจัยหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะเกิดประโยชน์ต่อผู้ที่มีความสนใจ สามารถนำไปต่อยอด องค์ความรู้เพื่อประยุกต์ใช้งานกับระบบการแนะนำการศึกษาเพื่อให้เกิดคุณค่าและสามารถใช้งานได้จริง

วรรณรอง ศรีศิริรุ่ง

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 สภาวะความเป็นมา แนวทางเหตุผล และปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ	3
2 แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	5
2.1 การพัฒนาอาจารย์ด้านการสอน.....	5
2.2 วิธีการประเมินการสอนของอาจารย์โดยนักศึกษา	6
2.3 การทำเหมืองข้อมูลโดยใช้วิธี Cross-Industry Standard Process for Data mining (CRISP-DM)	11
2.4 การทำเหมืองข้อความ.....	13
2.5 งานวิจัยที่เกี่ยวข้องกับการประเมินความคิดเห็นเกี่ยวกับการเรียนการสอน ของอาจารย์.....	26
2.6 งานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อความ	27
3 วิธีการดำเนินงาน	31
3.1 ขั้นตอนที่ 1 การวิเคราะห์ข้อมูลปัญหาของงานวิจัย	34

สารบัญ (ต่อ)

บทที่	หน้า
ภาคผนวก	60
ภาคผนวก ก. โค้ดโปรแกรมพีเอชพี (PHP)	62
ประวัติย่อผู้วิจัย	73



สารบัญตาราง

ตาราง	หน้า
2.1 เปรียบเทียบประสิทธิภาพการค้นคืนข้อมูล.....	24
3.1 ขั้นตอนการวิจัย.....	31
3.2 แผนการดำเนินงานวิจัย.....	33
3.3 ตัวอย่างการทำความสะอาดข้อมูล.....	36
3.4 ข้อมูลความคิดเห็นเพื่อเตรียมนำเข้าสู่การวิเคราะห์.....	37
4.1 สรุปข้อมูลที่ใช้วิเคราะห์.....	39
4.2 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม Decision Tree.....	41
4.3 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม Naive Bayes.....	42
4.4 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม Random Forest	42
4.5 สรุปประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน	43
4.6 แสดงเงื่อนไขจากอัลกอริทึม Decision Tree	44
4.7 ข้อมูลเพื่อใช้ในการวิเคราะห์ประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็น เชิงบวก เชิงลบ และเชิงกลาง.....	44
4.8 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางโดยใช้อัลกอริทึม Decision Tree	46
4.9 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางโดยใช้อัลกอริทึม Naive Bayes	47
4.10 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางโดยใช้อัลกอริทึม Random Forest.....	47
4.11 สรุปประสิทธิภาพการจำแนกความคิดเห็น	48
4.12 แสดงเงื่อนไขการจำแนกความคิดเห็นโดยใช้อัลกอริทึม Decision Tree.....	48
4.13 แสดงน้ำหนักสำหรับการประเมิน	49
4.14 แสดงน้ำหนักสำหรับประเภทความคิดเห็น	50

สารบัญรูป

รูป	หน้า
2.1 CRISP-DM	11
2.2 ขั้นตอนการทำเหมืองข้อความ.....	13
2.3 ต้นไม้ตัดสินใจ (Decision Tree).....	20
2.4 นาอ์ฟเบย์ (Naïve Bay).....	22
2.5 แรนดอมฟอเรส (Random Forest).....	22
2.6 สูตรการประเมินประสิทธิภาพของอัลกอริทึม	25
3.1 กรอบแนวความคิด.....	32
3.2 ความคิดเห็นต่อการสอนของอาจารย์โดยนักศึกษาจากระบบบริหารการศึกษา.....	34
3.3 ประเภทการประเมินการสอนของอาจารย์.....	35
3.4 แผนภาพยูสเคสของระบบจำลองการเสนอแนะเพื่อพัฒนาการสอนของอาจารย์.....	38
4.1 ขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน	40
4.2 การใช้อัลกอริทึม Decision Tree เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามด้านการประเมิน.....	40
4.3 การใช้อัลกอริทึม Naïve Bay เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามด้านการประเมิน.....	40
4.4 การใช้อัลกอริทึม Random Forest เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามด้านการประเมิน.....	41
4.5 ขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง	45
4.6 การใช้อัลกอริทึม Decision Tree เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง.....	45
4.7 การใช้อัลกอริทึม Naïve bay เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง	46
4.8 การใช้อัลกอริทึม Random Forest เพื่อวัดประสิทธิภาพการจำแนกข้อความ ตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง	46
4.9 ตัวต้นแบบสนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา	50
4.10 ตัวอย่างการใช้งานตัวต้นแบบ.....	51

บทที่ 1

บทนำ

1.1 สภาพความเป็นมา แนวทางเหตุผล และปัญหา

สถาบันการศึกษา คือ แหล่งเตรียมความพร้อมในการทำงานให้กับผู้เรียน ดังนั้นการศึกษามีคุณค่าก็ต่อเมื่อผู้เรียนนั้นสามารถนำความรู้ที่ได้จากการเรียนรู้ทฤษฎีต่างๆ ไปประยุกต์ใช้ในชีวิตจริงได้ ดังนั้นการเรียนการสอนแบบให้ผู้เรียนรู้จักคิดวิเคราะห์ สังเคราะห์ จึงควรเริ่มจากการวิธีการสอนของอาจารย์ในสถาบันการศึกษา

อาจารย์เป็นบุคคลที่จะส่งผลกระทบต่อคุณภาพของผู้เรียน อาจารย์ที่มีความเป็นมืออาชีพ มีความสามารถและศักยภาพสูง รวมถึงเป็นผู้ที่มีนวัตกรรมในการเรียนการสอนจะส่งผลให้ผู้เรียนได้รับผลการเรียนรู้ตามที่ต้องการ อาจารย์ยุคใหม่จึงต้องเรียนรู้ตลอดเวลาปรับตัวให้เข้ากับเทคโนโลยี สามารถแนะนำนักศึกษาได้ว่าข้อมูลใดสามารถนำไปใช้กับชีวิตจริงได้หรือวิธีการเลือกเครื่องมือที่เหมาะสม แต่สิ่งสำคัญที่นักศึกษาจะต้องได้รับการปลูกฝังไปพร้อมกับความรู้ทางปัญญา คือ ศีลธรรม คุณธรรม ซึ่งเป็นสิ่งที่อาจารย์ต้องถ่ายทอดให้นักศึกษาเห็นถึงความสำคัญในส่วนนี้ เพื่อดำเนินชีวิตอย่างมีความสุข [1]

คุณลักษณะของการเป็นอาจารย์ที่ดีในแนวคิดของนักวิชาการ [2] จะต้องมีความสำคัญ 10 ประการต่อไปนี้ 1. มุ่งมั่นวิชาการ 2. รักงานสอน 3. อหฺรศิษย์ 4. คิดดี 5. มีคุณธรรม 6. ซื่อสัตย์ 7. อบรมจิตใจ 8. ใฝ่ก้าวหน้า 9. วาจางาม 10. รักความเป็นไทย ซึ่งหลัก 10 ประการนี้ไม่สามารถบ่งชี้คุณลักษณะของการเป็นอาจารย์ที่ดีในความรู้สึกรักของนักศึกษาได้ โดยจากการสำรวจความคิดเห็นคุณสมบัติของ "ครู" ที่สำคัญที่สุดในยุคไทยแลนด์ 4.0 จากประชาชนทั่วประเทศ โดยศูนย์สำรวจความคิดเห็น "นิด้าโพล" สถาบันบัณฑิตพัฒนบริหารศาสตร์ 2561 พบว่า

1. เป็นคนดีมีคุณธรรม (เมตตา กรุณา ซื่อสัตย์ สุจริต)	ผลคะแนน	26.98%
2. เอาใจใส่ในการสอนแบบจริงจัง เป็นแบบอย่างที่ดี	ผลคะแนน	22.82%
3. มีทักษะในการสอน ถ่ายทอดความรู้ได้ดี	ผลคะแนน	21.79%
4. เก่ง มีความรู้ ความสามารถ พัฒนาตัวเอง หาความรู้เสมอ	ผลคะแนน	10.93%
5. เป็นที่ปรึกษา รับฟังปัญหาต่างๆของนักเรียนในเรื่องต่างๆ	ผลคะแนน	9.02%
6. ใจดี ไม่ดุ ไม่ลงโทษเด็ก	ผลคะแนน	4.15%
7. มีระเบียบวินัย	ผลคะแนน	4.15%

จากผลสำรวจจะเห็นได้ว่า นอกจากนักศึกษาจะชอบอาจารย์ที่ใจดีและสอนสนุก "ครู" ที่ในยุคไทยแลนด์ 4.0 ต้องมีคุณสมบัติด้านอื่นๆ ด้วย ได้แก่ เป็นคนดีมีคุณธรรม มีการสอนที่ดี สามารถให้คำปรึกษาที่ดี ซึ่งเป็นสิ่งสำคัญในการพัฒนานักศึกษาให้เติบโตอย่างมีประสิทธิภาพ

ในปัจจุบันสถาบันการศึกษามีการประเมินการสอนของอาจารย์โดยนักศึกษาผ่านระบบ Online ซึ่งข้อมูลจากการประเมินสามารถนำไปใช้สำหรับเป็นข้อมูลในการพัฒนาคุณภาพการเรียนการสอนของอาจารย์ ให้เป็นไปตามวัตถุประสงค์และเป้าหมายของการประกันคุณภาพการศึกษา ลักษณะของข้อมูลความคิดเห็นที่ได้จะประกอบไปด้วย 2 ลักษณะ ได้แก่ ข้อมูลเชิงปริมาณแสดงระดับความพึงพอใจ 1 – 5 และข้อมูลเชิงคุณภาพที่นักศึกษาสามารถแสดงความคิดเห็นเพิ่มเติมได้

จากที่กล่าวมาข้างต้น ความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา สำหรับข้อมูลเชิงคุณภาพจะมีลักษณะเชิงลึกกว่าข้อมูลเชิงปริมาณ นักวิจัยจึงเลือกเทคนิคการทำเหมืองข้อความ (Text Mining) มาใช้ในการวิเคราะห์คุณลักษณะของอาจารย์ เนื่องจากความคิดเห็นเป็นข้อมูลที่มีโครงสร้างที่ไม่แน่นอน ซึ่งเป็นภาษาธรรมชาติที่มนุษย์ใช้ในการสื่อสารและแลกเปลี่ยนประสบการณ์ร่วมกัน ทำให้ยากต่อการนำไปใช้ประโยชน์ เทคนิคการวิเคราะห์เหมืองข้อความและการประมวลผลภาษาธรรมชาติ (Natural Language Processing : NLP) เป็นเทคนิคที่นำมาประยุกต์ใช้ในการสกัดความคิดเห็น จากนั้นนำความคิดเห็นของนักศึกษามาสร้างข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์ต่อไป

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยใช้เทคนิคเหมืองข้อความ

1.2.2 เพื่อสังเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา

1.2.3 เพื่อพัฒนาตัวต้นแบบที่สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา เพื่อระบุปัจจัยที่ส่งผลต่อคุณภาพการสอนของอาจารย์ รวมทั้งความคิดเห็นเชิงบวกเชิงลบ และเชิงกลาง จากนั้นได้ข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์

1.3 ขอบเขตของการวิจัย

1.3.1 ขอบเขตด้านข้อมูล ความคิดเห็นเกี่ยวกับการสอนของอาจารย์สำหรับกลุ่มรายวิชาสำหรับนักศึกษาระดับปริญญาตรี จำนวน 3,130 ข้อความ โดยข้อมูลมาจากระบบบริการการศึกษา (Registrar System) ของสถาบันการศึกษาแห่งหนึ่ง

1.3.2 ขอบเขตระยะเวลา ความคิดเห็นเกี่ยวกับการสอนของอาจารย์สำหรับกลุ่มรายวิชา สำหรับนักศึกษาระดับปริญญาตรี โดยข้อมูลจากระบบบริการการศึกษา (Registrar System) สำหรับช่วงเวลา 1 สิงหาคม 2561 - 31 กรกฎาคม 2562

1.3.3 ขอบเขตด้านเทคนิคและวิธีการ

1.3.3.1 ส่วนของการวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยใช้เทคนิคเหมืองข้อความ เพื่อให้ได้ประสิทธิภาพของอัลกอริทึมที่นำมาใช้ในการจำแนกข้อความด้านการประเมินและจำแนกประเภทความคิดเห็นที่เหมาะสมถูกต้องที่สุดและทำไปเข้าสู่กระบวนการวิเคราะห์ของขั้นตอนการทำเหมืองข้อความต่อไป

1.3.3.2 การประเมินประสิทธิภาพการจำแนกข้อความสำหรับความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา นำอัลกอริทึม 3 ประเภทมาใช้ ได้แก่ อัลกอริทึมต้นสนใจต้นไม้ที่พิจารณาลักษณะของข้อมูลเพื่อเชื่อมโยงข้อมูลเข้าด้วย อัลกอริทึมนาอิวเบย์การที่คาดการณ์ผลลัพธ์โดยทำการวิเคราะห์ระหว่างความสัมพันธ์การสร้างเงื่อนไขของความน่าจะเป็น และอัลกอริทึมแรนดอมฟอร์เรสที่พิจารณาการจำแนกข้อความโดยใช้การโหวตจากกลุ่มต้นไม้ทั้งหมด ในการทดลองวิจัยครั้งนี้จะใช้เครื่องมือ Rapid Miner มาใช้ในการทำเหมืองข้อความ

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ความเข้าใจในแนวคิดกระบวนการจำแนกความคิดเห็นเกี่ยวกับการสอนของอาจารย์โดยนักศึกษา ตั้งแต่เริ่มต้นจนถึงสิ้นสุด ทำให้เกิดกระบวนการเรียนรู้และกระบวนการคิดที่เป็นแบบแผน

1.4.2 รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาทั้งความคิดเห็นเชิงบวกและเชิงลบ

1.4.3 ตัวต้นแบบที่สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา เพื่อระบุปัจจัยที่ส่งผลต่อคุณภาพการสอนของอาจารย์ จากนั้นได้ข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์

1.5 นิยามศัพท์เฉพาะ

1.5.1 ระบบบริการการศึกษา หมายถึง ระบบที่ช่วยให้ข้อมูลกับนักศึกษาเกี่ยวกับการลงทะเบียนเรียน การตรวจสอบผลการเรียน รวมถึงการประเมินการเรียนการสอนของอาจารย์ในรายวิชาที่ลงทะเบียนไว้ สำหรับอาจารย์จะใช้ระบบนี้ในการทำคะแนน ตรวจสอบผลการเรียน และตรวจสอบผลประเมินการสอนของอาจารย์โดยนักศึกษา

1.5.2 **ด้านการประเมิน** หมายถึง หัวข้อสำหรับการประเมินการสอนของอาจารย์สำหรับการวิจัยครั้งนี้ ประกอบด้วย 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน ด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์

1.5.3 **Teaching** หมายถึง หัวข้อการประเมินด้านอาจารย์ผู้สอน

1.5.4 **Activity** หมายถึง หัวข้อการประเมินด้านกิจกรรมการเรียนการสอน

1.5.5 **Development** หมายถึง หัวข้อการประเมินด้านการพัฒนาศักยภาพนักศึกษา
เสริมสร้างคุณลักษณะที่พึงประสงค์



บทที่ 2

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีการศึกษาแนวคิด ทฤษฎี จากหนังสือ บทความวิชาการ และแหล่งสารสนเทศ อินเทอร์เน็ต งานวิจัยที่เกี่ยวข้องต่างๆ ดังหัวข้อต่อไปนี้

- 2.1 การพัฒนาอาจารย์ด้านการสอน
- 2.2 วิธีการประเมินการสอนของอาจารย์โดยนักศึกษา
- 2.3 การทำเหมืองข้อมูลโดยใช้วิธี Cross-Industry Standard Process for Data mining (CRISP-DM)
- 2.4 การทำเหมืองข้อความ
- 2.5 งานวิจัยที่เกี่ยวข้องกับการประเมินความคิดเห็นเกี่ยวกับการเรียนการสอนของอาจารย์
- 2.6 งานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อความ

2.1 การพัฒนาอาจารย์ด้านการสอน

การพัฒนาอาจารย์ในมหาวิทยาลัย จำเป็นต้องมีการคำนึงถึงการพัฒนาหลายด้าน เช่น ด้านวิชาการความเป็นครู ด้านการจัดการเรียนการสอน ด้านวิจัย ด้านสมรรถนะวิชาชีพที่เกี่ยวข้องกับสาขาที่ตนรับผิดชอบ ด้านการบริหารจัดการ และด้านการขัดเกลาทางสังคม [3] ในการพัฒนาอาจารย์ในด้านการสอนนั้น จำเป็นต้องมีการศึกษาปัญหาที่เกิดขึ้น ได้แก่ (1) ด้านสิ่งแวดล้อม เช่น สื่อและอุปกรณ์ในการเรียนการสอน ขนาดห้องเรียนเล็กไปเมื่อเทียบกับจำนวนผู้เรียน (2) ด้านผู้เรียน เช่น พื้นฐานของนักศึกษาแต่ละคนไม่เท่ากัน มีความสนใจในการเรียนไม่เท่ากัน (3) ด้านการเน้นผู้เรียนเป็นสำคัญ เช่น ปัญหาจากตัวอาจารย์ผู้สอนที่ขาดความเข้าใจ และจากตัวนักศึกษาที่เคยชินกันวิธีการสอนแบบเดิม (4) ด้านการปรับปรุงการสอนให้ทันสมัย เช่น การขาดตำรา หนังสือ เอกสาร และวารสารที่ตรงกับศาสตร์จะใช้ และอาจารย์ไม่มีเวลาในการที่จะค้นคว้า จากการพัฒนารูปแบบการจัดการเรียนการสอนที่มีประสิทธิภาพในการจัดการศึกษาระดับปริญญาตรี ของสถาบันอุดมศึกษา สังกัดกรมศิลปากร พบว่า รูปแบบการเรียนการสอนที่มีประสิทธิภาพ ด้านกระบวนการเรียนการสอน ได้แก่ การทำแผนการสอน การจัดทำประมวลการสอน การจัดตารางสอน การจัดอาจารย์เข้าสอน วิธีการสอน และการซ่อมเสริม

กระบวนการพัฒนาอาจารย์มี 4 ขั้นตอน [4] ดังนี้

ขั้นที่ 1 การเตรียมการสำรวจ วิเคราะห์สภาพปัญหา และความต้องการการพัฒนาอาจารย์

ขั้นที่ 2 การสร้างกิจกรรมการพัฒนาอาจารย์ เป็นขั้นตอนที่ต้องดำเนินการร่วมกันระหว่างอาจารย์ และผู้บริหารสถาบันการศึกษา

ขั้นที่ 3 การติดตามผลอย่างต่อเนื่อง

ขั้นที่ 4 การดำเนินการบริหารทรัพยากร ได้แก่ ทรัพยากรบุคคล ทรัพยากรด้านการเงิน ทรัพยากรด้านกายภาพ เช่น สื่อการเรียนการสอน วัสดุอุปกรณ์ และทรัพยากรสารสนเทศ

2.2 วิธีการประเมินการสอนของอาจารย์โดยนักศึกษา

ในหัวข้อนี้มีการอธิบายเกี่ยวกับองค์ประกอบที่ผู้เรียนใช้ในการประเมินการสอนของอาจารย์ ซึ่งการประเมินต้องเป็นระบบ ครอบคลุมทุกมิติ และเป็นไปตามสภาพความจริง รูปแบบการประเมินมีทั้งวิธีการที่แบบง่าย ๆ ไปจนถึงซับซ้อน โดยมีวัตถุประสงค์เพื่อให้มีการกำกับ ติดตาม คุณภาพของการเรียนการสอน การให้รางวัลหรือการยกเลิกการจ้างงาน การวางแผนงบประมาณเพื่อการจัดหาปัจจัยที่ช่วยส่งเสริมการเรียนรู้และการพัฒนาสภาพแวดล้อม รวมถึงผู้สอนจะรับรู้จุดเด่น และจุดที่ควรพัฒนาของตนเอง ข้อมูลที่ได้จากการประเมินที่ดีจะส่งผลให้ผู้ถูกประเมินสามารถตัดสินใจเปลี่ยนแปลงและปรับปรุงการเรียนการสอนให้ดีขึ้น สำหรับการวิจัยครั้งนี้มีวิธีการประเมินมี 2 วิธี ดังนี้

2.2.1 การประเมินโดยการให้คะแนน (Rating Scale)

วิธีการวัดความคิดเห็นของนักศึกษาแบบ Rating Scale สำหรับการวิจัยครั้งนี้ คือ การให้นักศึกษาเลือกแสดงความคิดเห็นจากระดับความคิดเห็น 5 ระดับ ดังนี้

- 1 หมายถึง ไม่เห็นด้วยอย่างยิ่ง
- 2 หมายถึง ไม่เห็นด้วย
- 3 หมายถึง ไม่แน่ใจ/ยังเห็นไม่ชัดเจน
- 4 หมายถึง เห็นด้วย
- 5 หมายถึง เห็นด้วยอย่างยิ่ง

โดยหัวข้อสำหรับการประเมินการสอนของอาจารย์สำหรับการวิจัยครั้งนี้ ประกอบด้วย 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน ด้านการพัฒนา ศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ ซึ่งแต่ละด้านประกอบด้วยหัวข้อย่อย ดังนี้

- ด้านอาจารย์ผู้สอน ประกอบด้วย 5 หัวข้อย่อย ได้แก่

(1) อาจารย์มีบุคลิกดี วางตัวเหมาะสม สามารถเป็นแบบอย่างที่ดีให้กับนักศึกษา (การพูดจา กริยาท่าทาง การแต่งกาย เป็นต้น)

(2) อาจารย์ให้ความสนใจ เอาใจใส่นักศึกษา พร้อมให้ความช่วยเหลือในด้านการเรียนการสอน

(3) อาจารย์ให้ความสนใจและดูแลเอาใจใส่นักศึกษาอย่างทั่วถึง พร้อมให้ความช่วยเหลือในด้านการเรียนการสอน

(4) อาจารย์มีความสามารถในการถ่ายทอดความรู้ อธิบายเนื้อหาได้อย่างชัดเจน เข้าใจง่าย

(5) อาจารย์สามารถให้คำแนะนำหรือตอบข้อซักถามของนักศึกษาได้อย่างชัดเจน และเข้าใจดี

- ด้านกิจกรรมการเรียนการสอน ประกอบด้วย 11 หัวข้อย่อย ได้แก่

(1) อาจารย์ได้ชี้แจงวัตถุประสงค์การเรียนรู้ อธิบายขอบเขตของเนื้อหาวิชา แหล่งค้นคว้าข้อมูล รวมทั้งเกณฑ์การวัดและประเมินผลอย่างชัดเจน

(2) อาจารย์เตรียมเนื้อหาการสอนอย่างเป็นลำดับ มีความต่อเนื่อง น่าสนใจ ทันสมัย และครอบคลุมคำอธิบายรายวิชา

(3) อาจารย์ใช้สื่อการสอนที่เหมาะสม สอดคล้องกับเนื้อหาวิชาและเอื้อต่อการเรียนรู้ของนักศึกษา

(4) อาจารย์ใช้เทคนิค/วิธีการสอนที่หลากหลาย จูงใจให้นักศึกษาสนใจอยากเรียนรู้และติดตามการเรียนรู้

(5) อาจารย์มีการเชื่อมโยงเนื้อหาภาคทฤษฎีกับกรณีศึกษาจากการฝึกปฏิบัติ/การทำงานจริง เพื่อกระตุ้นให้นักศึกษาได้คิด วิเคราะห์ นำความรู้ที่เรียนมาประยุกต์ใช้

(6) อาจารย์เปิดโอกาสให้นักศึกษาได้ซักถาม แสดงความคิดเห็น แลกเปลี่ยนประสบการณ์ หรือมีส่วนร่วมในชั้นเรียน เพื่อสร้างบรรยากาศการเรียนรู้

(7) อาจารย์มอบหมายงานให้นักศึกษาฝึกปฏิบัติอย่างเหมาะสม (ทั้งรูปแบบ ปริมาณงาน และระยะเวลา)

(8) อาจารย์มีการติดตามผลการเรียนรู้ของนักศึกษาอย่างสม่ำเสมอ พร้อมให้ข้อเสนอแนะที่เป็นประโยชน์

(9) อาจารย์ใช้วิธีวัดและประเมินผลการเรียนที่เหมาะสม ครอบคลุมเนื้อหาและวัตถุประสงค์รายวิชา

(10) อาจารย์สอนได้ครอบคลุมเนื้อหาและครบตามแผนการสอนที่กำหนด

(11) เมื่อพบปัญหาในการเรียน นักศึกษาสามารถติดต่อขอคำแนะนำจากอาจารย์ผ่าน ช่องทางต่างๆ ได้ เช่น โทรศัพท์, E-mail, Social Media ต่างๆ

- ด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ ประกอบด้วย 5 หัวข้อย่อย ได้แก่

- (1) นักศึกษาได้รับการเสริมสร้างคุณธรรม จริยธรรม และค่านิยมที่ดีงามในระหว่างการเรียนรู้
- (2) นักศึกษาได้รับการสร้างนิสัยใฝ่เรียน ใฝ่รู้จักวิธีการค้นคว้าและเพิ่มพูนความรู้เพื่อการพัฒนาตนเองอย่างต่อเนื่อง
- (3) นักศึกษาได้รับการส่งเสริมให้รู้จักการทำงานเป็นทีมหรือการทำงานร่วมกับผู้อื่น
- (4) นักศึกษาได้รับการพัฒนาทักษะด้านการสื่อสารและการนำเสนอในรูปแบบต่างๆ
- (5) การเรียนวิชานี้เป็นส่วนหนึ่งที่ทำให้นักศึกษาได้พัฒนาความเป็นมืออาชีพ/ความเชี่ยวชาญในวิชาชีพ

2.2.2 การแสดงความคิดเห็นโดยการบรรยาย

2.2.2.1 ความหมายของความคิดเห็น

พจนานุกรมศัพท์สังคมวิทยาฉบับราชบัณฑิตยสถาน 2542 ได้ให้ความหมาย ความคิดเห็นว่าเป็นข้อพิจารณาว่าเป็นจริงจากการใช้ปัญญาความคิดประกอบถึงแม้จะไม่ได้อาศัยหลักฐานพิสูจน์ ยืนยันได้เสมอไปก็ตาม [5] อธิบายว่า ความคิดเห็นนั้นถูกจัดว่าเป็นส่วนที่มนุษย์ได้แสดงออกมาโดยการพูดหรือเขียน มนุษย์นั้นจะพูดจากใจจริง พูดตามสังคมหรือพูดเพื่อเอาใจผู้ฟังก็ตามแต่เมื่อพูดหรือเขียนไปแล้วก็ทำให้เกิดผลได้ คนส่วนใหญ่มักจะถือว่าสิ่งที่มนุษย์แสดงออกมานั้นเป็นสิ่งที่สะท้อนถึงความในใจด้วยเหตุนี้ จึงเป็นที่นิยมกันมากที่จะสำรวจความคิดเห็นต่อสิ่งหนึ่งหรือเรียกกันว่าการสำรวจประชามติ (Polling) จึงอาจกล่าวได้ว่า การหยั่งประชามติเป็นเครื่องมือสำคัญทางวิชาการที่ศึกษาและสำรวจการแสดงออกทางความคิดเห็นในปัจจุบัน ความคิดเห็นว่าเป็นการแสดงออกทางถ้อยคำ (Verbal Expression) เกี่ยวกับทัศนคติความเชื่อหรือค่านิยมแต่ความคิดเห็นไม่ใช่สิ่งเดียวกับทัศนคติ เพราะในตัวของมันเองไม่จำเป็นต้องประกอบด้วยองค์ประกอบทางอารมณ์หรือพฤติกรรม

สำหรับการวิจัยครั้งนี้ที่มาของความคิดเห็นมาจากระบบบริการการศึกษา (Registrar System) ของสถาบันการศึกษาระดับปริญญาตรีแห่งหนึ่ง โดยเป็นความคิดเห็นเกี่ยวกับรายวิชาแกน ซึ่งผู้บรรยายความคิดเห็นกำลังศึกษาอยู่ในระดับปีที่ 1 – 4 เท่านั้น

2.2.2.2 ปัจจัยที่ทำให้เกิดความคิดเห็น

ความคิดเห็นเป็นสิ่งที่เกิดจากการเรียนรู้มากกว่าเป็นสิ่งที่กำเนิดเอง สิ่งแวดล้อมต่าง ๆ จึงมีอิทธิพลต่อความคิดเห็น ซึ่งได้แก่ ศาสนา ความเชื่อในสังคม ขนบธรรมเนียมประเพณีของสังคม สื่อมวลชนต่าง ๆ ดังนั้น ปัจจัยที่กำหนดความคิดเห็นของบุคคล [6] ได้แก่ 1) การเรียนรู้ซึ่ง ได้แก่ การ

อบรมสั่งสอน อันจะเป็นการสะสมและรวบรวมประสบการณ์เอาไว้เป็นจำนวนมาก เช่น เด็กเกิดในครอบครัวที่นับถือศาสนาพุทธก็จะมีเคลือบฝังในพุทธศาสนา เพราะได้รับอิทธิพลจากการอบรมสั่งสอนประสบการณ์ต่าง ๆ ไว้ 2) ประสบการณ์ส่วนตัวของบุคคลโดยตรง เช่น บุคคลที่เคยรับประทานอาหารทะเลแล้วแพ้ก็จะย่อมมีความคิดเห็นที่ไม่ดีต่ออาหารทะเล 3) เหตุการณ์ประทับใจ ใน 2 ข้อข้างต้นที่กล่าวมานั้น เป็นการสะสมประสบการณ์หลาย ๆ ครั้งและเกิดความคิดเห็น แต่ความคิดเห็นก็สามารถเกิดขึ้นได้หากได้รับเหตุการณ์เพียงครั้งเดียวและรู้สึกประทับใจ ซึ่งอาจจะประทับใจในทางบวกและลบได้ 4) การรับเอาแบบของความคิดเห็นของผู้อื่นมาเป็นของตน โดยยอมรับเอาความคิดเห็นของผู้ที่เหนือกว่ามาปฏิบัติต่อ เช่น รุ่นน้องรับความคิดเห็นบางเรื่องมาจากรุ่นพี่ 5) เกิดจากลักษณะบุคลิกภาพของแต่ละคน เช่น การมองคนในแง่ร้ายก็จะมีแนวโน้มในทางที่มีความคิดเห็นที่ไม่ดีต่อสิ่งต่าง ๆ อยู่เสมอ 6) เกิดจากอิทธิพลจากสื่อมวลชน สื่อมวลชนเป็นแหล่งให้ข้อมูลทีก่อให้เกิดทั้งความเข้าใจ และอารมณ์ซึ่งนำไปสู่การปฏิบัติได้

ความสำคัญของความคิดเห็น การศึกษาเชิงสำรวจความคิดเห็นนั้นเป็นการศึกษาจากองค์ประกอบของความคิดเห็นของบุคคล โดยทั่วไป 3 ประการ [7] คือ บุคคลที่ถูกวัดสิ่งเร้าและการตอบสนอง ซึ่งจะแสดงออกมาในระดับที่สูง ต่ำ มากหรือน้อย โดยทั่วไปแล้ววิธีวัดความคิดเห็นนั้นจะใช้การตอบแบบสอบถามและการสัมภาษณ์ ทั้งนี้เพื่อจะค้นหาความรู้สึกการตัดสินใจจากการประเมินค่าหรือทัศนคติเรื่องใดเรื่องหนึ่ง โดยเฉพาะของบุคคล กลุ่มบุคคลที่มีต่อสิ่งต่าง ๆ ดังนี้ การศึกษาวิจัยด้านความคิดเห็นจึงเป็นประโยชน์ต่อการวางแผนทางกำหนดนโยบาย การพัฒนาหน่วยงานต่าง ๆ ทำให้เกิดการเปลี่ยนแปลงไปในทิศทางที่ดีขึ้นตรงเป้าหมายและยังสามารถตอบสนองความต้องการของหน่วยงานหรือบุคคลในหน่วยงานนั้น โดยเฉพาะอย่างยิ่งหน่วยงานที่มีหน้าที่รับผิดชอบในการปลูกฝังอบรมให้ความรู้ สร้างสรรค์พัฒนาสติปัญญาของบุคคลด้วยแล้ว ยิ่งต้องตระหนักในการสำรวจเก็บข้อมูลความคิดเห็นความรู้สึกเหล่านี้ เพื่อเป็นข้อมูลย้อนกลับมาปรับปรุงเตรียมความพร้อมของหน่วยงานให้ทันสมัยทันต่อสถานการณ์อยู่เสมอ

การวัดความคิดเห็น (Opinion Measurement) ทัศนคติ แรงจูงใจ และค่านิยมได้มีการสร้างแบบทดสอบสำหรับวัดสิ่งต่าง ๆ ดังกล่าว แต่ยังไม่สามารถแยกจากกันได้อย่างเด็ดขาด เพราะมีบางส่วนที่ซ้ำซ้อนกันอยู่ การวัดความคิดเห็นส่วนใหญ่แล้วยังไม่มีการแบ่งแยกออกจากทัศนคติอย่างชัดเจน และมีบ่อยครั้งที่คำทั้งสองใช้สลับกัน อย่างไรก็ตามการสำรวจความคิดเห็นมักจะเป็นการถามสิ่งที่เฉพาะเจาะจง เช่น การสำรวจความคิดเห็นเกี่ยวกับการเลือกตั้งสมาชิกสภาผู้แทนราษฎร การสำรวจความคิดเห็นของประชาชนต่อโครงการสาธารณะต่าง ๆ เป็นต้น ซึ่งผลที่ได้จากการสอบถามความคิดเห็นเหล่านี้จะเป็นตัวชี้วัดความพึงพอใจ ไม่พอใจ เห็นด้วยหรือไม่เห็นด้วยของกลุ่มประชากรดังกล่าว อีกทั้งหน้าที่ของความคิดเห็น โดยทั่วไปหน้าที่ของความคิดเห็นมีดังต่อไปนี้ [8] คือ

- 1) ทำหน้าที่เป็นแรงจูงใจให้บุคคลปรับตัว เมื่อมีความคิดเห็นที่ดีต่อสิ่งใดย่อมเข้าหาสิ่งนั้น และยอม

หลักหนึ่งสิ่งที่มีความคิดเห็นที่ไม่ดี 2) ทำหน้าที่ให้คำนิยมหรือให้ความชื่นชมต่อเนื่องไปถึงสิ่งอื่น ๆ เช่น มีความคิดเห็นว่าการเปลี่ยนแปลงทางสังคมยังต้องใช้วิธีการสันติ ถ้าพบบุคคลที่มีแนวความคิดนี้จะนิยมชมชอบไปด้วย 3) ทำหน้าที่ช่วยให้ตีความหมายของสถานการณ์ต่าง ๆ ได้ เช่น ถ้ามีความคิดเห็นที่ดีต่อพ่อแม่บางครั้งพ่อแม่อาจจะขัดแย้งกับตน ก็จะตีความหมายไปว่าพ่อแม่ทำไปเพราะความหวังดี

ประโยชน์ของความคิดเห็น [9] ประโยชน์ของความคิดเห็นมีดังต่อไปนี้ 1) ช่วยทำให้เข้าใจสิ่งแวดล้อมรอบ ๆ ตัว โดยการจัดรูปหรือการจัดระบบสิ่งของต่าง ๆ ที่อยู่รอบตัว 2) ช่วยให้มี Self-esteem โดยจะช่วยให้บุคคลหลีกเลี่ยงสิ่งที่ไม่ดีหรือปกปิดความจริงบางอย่าง ซึ่งนำความไม่พอใจมาสู่ตัว 3) ช่วยในการปรับตัวให้เข้ากับสิ่งแวดล้อมที่สลับซับซ้อน ซึ่งมีปฏิกิริยาตอบโต้หรือกระทำสิ่งใดสิ่งหนึ่งออกไปนั้น ส่วนมากจะนำความพึงพอใจมาให้ 4) ช่วยให้บุคคลสามารถแสดงออกในด้านค่านิยม ความรู้สึกของตนเอง อันจะนำความพอใจมาสู่บุคคลนั้น ๆ จากการศึกษาข้างต้น ผู้ศึกษาพอสรุปแนวความคิดที่เกี่ยวกับการเกิดความคิดเห็น ประการแรกคือ เกิดจากประสบการณ์ที่บุคคลมีกับสิ่งของ บุคคล หรือสถานการณ์ และประการที่สองเกิดจากค่านิยม และการตัดสินใจค่านิยม ซึ่งแต่ละบุคคลมีค่านิยมและการตัดสินใจค่านิยมที่แตกต่างกันได้ย่อมทำให้มีความคิดเห็นที่แตกต่างกัน

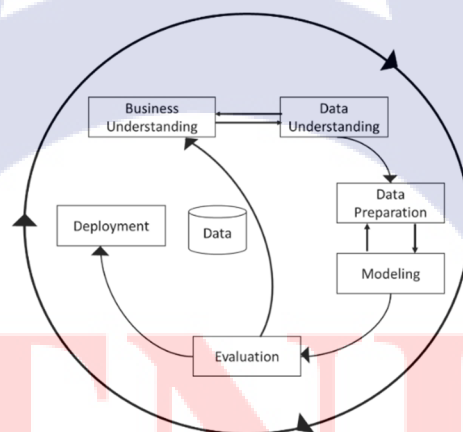
2.2.3 การประยุกต์ใช้การวิเคราะห์ความคิดเห็น

ข้อความบนเครือข่ายสังคมออนไลน์แบ่งเป็น 3 ประเภท ได้แก่ ข้อเท็จจริง ความคิดเห็น และข้อเสนอแนะ โดย (1) ข้อเท็จจริง หมายถึง ข้อความหรือเหตุการณ์ที่เป็นมา หรือที่เป็นอยู่ตามจริง (2) ความคิดเห็นหมายถึงความเห็นหรือข้อวินิจฉัยหรือความเชื่อที่แสดงออกตามความเห็น รู้ คิด และ (3) ข้อเสนอแนะหมายถึงข้อคิดเห็นเชิงแนะนำที่เสนอเพื่อพิจารณา ชี้แจงให้ทำหรือปฏิบัติตาม [10] ตัวอย่าง “เมื่อคืนได้ดูรายการพื้นที่ชีวิต ไม่ชอบพิธีกรเลย อยากให้ปรับปรุงเรื่องการใช้ภาษาหน่อย” จากประโยคตัวอย่าง สามารถจำแนกประเภทของบทวิจารณ์ได้ ดังนี้ ข้อเท็จจริง : “เมื่อคืนได้ดูรายการพื้นที่ชีวิต” การวิเคราะห์เหมือนข้อความ แบ่งข้อมูลออกเป็น 2 รูปแบบคือ (1) ข้อมูลที่เป็นโครงสร้าง (Structured Data) ซึ่งการประมวลผลข้อมูลที่เป็นโครงสร้าง เรียกว่า การวิเคราะห์เหมือนข้อมูล และ (2) ข้อมูลที่ไม่เป็นโครงสร้าง (Unstructured Data) หรือไม่มีโครงสร้างที่ชัดเจน (Implicit Structured Data) ซึ่งส่วนใหญ่มักอยู่ในรูปแบบของข้อความ หรือภาษาธรรมชาติ และเรียกกระบวนการวิเคราะห์ข้อความนี้ว่าการวิเคราะห์เหมือนข้อความ (Text Mining) โดยเป็นการนำความรู้ด้านการวิเคราะห์เหมือนข้อมูล มาประยุกต์ใช้กับหน่วยงาน Cross Industry Standard Process for Data Mining (CRISP DM) ได้นำเสนอกระบวนการวิเคราะห์เหมือนข้อมูล อันประกอบด้วยขั้นตอนดังต่อไปนี้ (1) การทำความเข้าใจกับธุรกิจและระบุปัญหาของงาน (Business Understanding) (2) การรวบรวมความหมายเหมือนกันแต่ใช้คำต่างกันจะแก้ไขให้เป็นคำเดียว เช่น

คำ “คนอ่านข่าว” กับ “ผู้ประกาศข่าว” มีความหมายเหมือนกันจึงแก้ไขเป็น “ผู้ประกาศข่าว” เป็นต้น (3) การแทนข้อความเป็นกระบวนการแทนข้อความให้อยู่ในรูปแบบที่มีโครงสร้างแน่นอน มีเทคนิคที่นิยมใช้คือ วิกเตอร์สเปซโมเดล (Vector Space Model: VSM) เป็นวิธีการหนึ่งในการแทนข้อความให้มีโครงสร้างแบบฟอร์มพีวเจอร์ วิกเตอร์ (Feature Vector) เพื่อให้คอมพิวเตอร์สามารถนำไปประมวลผลต่อได้ โดยมีวิธีการแทนข้อความให้มีโครงสร้าง ดังนี้ การแทนข้อความด้วยถ่วงคำ (Bag-of-Word) เป็นวิธีการที่ง่ายและสะดวกที่สุด โดยแทนข้อความด้วยค่าที่รวบรวมไว้แล้วในคลังคำ หากมีค่าในถ่วงคำเกิดขึ้นในข้อความที่นำมาวิเคราะห์จะแทนค่าด้วย 1 แต่หากในข้อความไม่มีค่าที่กำหนดไว้ในถ่วงคำเกิดขึ้น จะแทนค่าด้วย 0

2.3 การทำเหมืองข้อมูลโดยใช้วิธี Cross-Industry Standard Process for Data mining (CRISP-DM)

CRISP-DM เป็นกระบวนการมาตรฐานขั้นตอนการทำเหมืองข้อมูล ที่ได้รับการพัฒนาในปี ค.ศ.1996 เกิดจากความร่วมมือกัน 3 บริษัท ได้แก่ บริษัท SPSS, บริษัท DaimlerChrysler และ บริษัท NCR ซึ่งประกอบไปด้วย 6 ขั้นตอน ดังนี้



รูปที่ 2.1 CRISP-DM

ขั้นตอนที่ 1 Business Understanding เป็นขั้นตอนแรกสำหรับกระบวนการ CRISP-DM ที่มีวัตถุประสงค์เพื่อทำความเข้าใจปัญหาและสามารถวิเคราะห์ปัญหาทางธุรกิจที่เกิดขึ้น โดยใช้ข้อมูลที่มี รวมถึงการวางแผนระบุสิ่งที่ต้องการหรือเป้าหมายที่ต้องการได้จากการวิเคราะห์ข้อมูล เช่น ทำอย่างไรจึงสามารถเพิ่มยอดขายได้ การแบ่งกลุ่มลูกค้า การทำนายปริมาณน้ำฝน หรือโอกาสในการเจ็บป่วย เป็นต้น

ขั้นตอนที่ 2 Data Understanding ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล จากนั้นตรวจสอบข้อมูลที่ทำกรรวบรวมมาเพื่อให้ข้อมูลมีความถูกต้องสำหรับการนำไปวิเคราะห์ และต้องมีการพิจารณาความจำเป็นในการใช้ข้อมูลทั้งหมด หรือบางส่วนเท่านั้น เพราะจะทำให้ช่วยประหยัดเวลาในการวิเคราะห์ข้อมูลต่อไป เช่น ข้อมูลลูกค้าที่ซื้อสินค้า ข้อมูลการลงทะเบียนเรียน เป็นต้น

ขั้นตอนที่ 3 Data Preparation เป็นขั้นตอนในการแปลงข้อมูลที่ได้เก็บรวบรวมมาให้เป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไป โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง ซึ่งเป็นขั้นตอนที่ใช้เวลานานที่สุด เพื่อให้ข้อมูลมีคุณภาพที่จะนำไปใช้ในการสร้างโมเดลบางได้เป็น 3 ขั้นตอนย่อย ได้แก่

3.1 การคัดเลือกข้อมูล เป็นการกำหนดเป้าหมาย และเลือกเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์

3.2 การกลั่นกรองข้อมูล เป็นการลบข้อมูลที่ซ้ำซ้อน แก้ไขข้อมูลที่ผิดพลาด เช่น ข้อมูลผิดรูปแบบ ข้อมูลสูญหาย

3.3 การแปลงรูปแบบข้อมูล เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้วิเคราะห์ตามอัลกอริทึมของการทำเหมืองข้อมูลที่เลือกใช้

ขั้นตอนที่ 4 Modeling เป็นขั้นตอนจะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล เช่น การจำแนกข้อมูล (Classification) เป็นการสร้างโมเดลจากข้อมูลที่มีอยู่เพื่อทำนายอนาคต การแบ่งกลุ่มข้อมูล (Clustering) เป็นการแบ่งข้อมูลให้เป็นกลุ่มที่มีลักษณะคล้ายกัน การหาความสัมพันธ์ (Association Rules) เป็นการหาความสัมพันธ์ของข้อมูลที่เกิดขึ้นร่วมกัน ซึ่งขั้นตอนนี้จะเลือกนำเทคนิคต่างๆมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด เช่น Decision tree, Naïve Bay, Neural Network, Support Vector Machines เป็นต้น ดังนั้นบางครั้งจะมีการย้อนกลับไปขั้นตอนที่ 3 เพื่อการแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย

ขั้นตอนที่ 5 Evaluation เป็นขั้นตอนที่จะนำผลการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูลนำมาวัดประสิทธิภาพของผลลัพธ์ว่าตรงกับวัตถุประสงค์ที่กำหนดไว้ในขั้นตอนแรก หรือมีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้านี้เพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้มาซึ่งผลลัพธ์ตามที่ต้องการได้

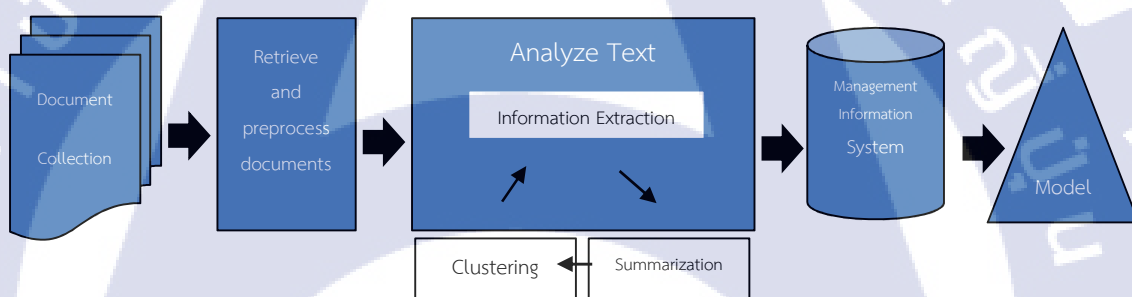
ขั้นตอนที่ 6 Deployment เป็นกระบวนการสกัดผลลัพธ์ที่ได้และนำไปใช้งานกับปัญหาที่กำหนดไว้ตั้งแต่ขั้นตอนแรก

จากขั้นตอนที่กล่าวมาเป็นการทำเหมืองข้อมูล ที่ใช้กระบวนการทางสถิติที่เน้นการทำกับข้อมูลที่มีจำนวนมากในหลากหลายรูปแบบ คัดเลือกเฉพาะข้อมูลที่สำคัญและจำเป็นต้องนำมาใช้งาน จัดทำการกำหนดรูปแบบจัดแบ่งกลุ่มลำดับความสำคัญ จากนั้นจึงจะเริ่มค้นหารูปแบบแนวทางและ

ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยแต่ละขั้นตอนจะอาศัยผลลัพธ์จากอีกขั้นตอนหนึ่งกลายเป็นข้อมูลให้ขั้นตอนต่อไป การทำเหมืองข้อมูลจะช่วยเปลี่ยนข้อมูลดิบให้เป็นสารสนเทศที่มีประโยชน์ การระบุแหล่งข้อมูลที่ต้องการจึงเป็นสิ่งสำคัญต่อผลลัพธ์ที่ได้จากการวิเคราะห์

2.4 การทำเหมืองข้อความ

การวิเคราะห์ข้อมูลเป็นกระบวนการค้นหาและสกัดหารูปแบบจากฐานข้อมูลขนาดใหญ่ เพื่อให้ได้รูปแบบที่สามารถนำมาใช้ประโยชน์ได้ ถ้าข้อมูลที่นำมาใช้ในการสกัดรูปแบบเป็นลักษณะข้อความจะเรียกว่า การทำเหมืองข้อความ (Text Mining) ซึ่งเป็นเทคนิคในการค้นหารูปแบบจากฐานข้อความขนาดใหญ่ โดยใช้กระบวนการค้นหารูปแบบ ความสัมพันธ์ หรือการเปลี่ยนแปลงที่แฝงอยู่ในข้อความ ขั้นตอนในการค้นหารูปแบบ ได้แก่ วิธีทางสถิติ การเรียนรู้ของเครื่อง (Machine Learning) และ การจัดจํารูปแบบ เพื่อให้ได้รูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความ (Text data sets)



รูปที่ 2.2 ขั้นตอนการทำเหมืองข้อความ

สำหรับการทำเหมืองข้อความจะแตกต่างจากการวิเคราะห์เหมืองข้อมูลในส่วนของคุณลักษณะ โครงสร้างของข้อมูล ข้อความไม่มีโครงสร้างที่ชัดเจน แต่ข้อมูลที่มีลักษณะโครงสร้างข้อมูลที่ชัดเจน สามารถจัดทำได้อยู่ในรูปของตารางได้ ดังนั้นข้อความจะนำมาแปลงให้เหมาะสมกับการทำเหมืองข้อมูล โดยการกำหนดหัวข้อคุณลักษณะของข้อความ ซึ่งกระบวนการเตรียมข้อมูลนี้จะทำให้คอมพิวเตอร์สามารถเข้าใจกับคำและประโยคมากขึ้น และถูกนำไปใช้ในการแก้ปัญหาได้หลากหลายวิธี เช่น การสกัดข้อมูล (Information Extraction) การจำแนกเอกสาร (Document Classification) การค้นคืนสารสนเทศ (Information Retrieval) การจัดกลุ่มระบบเอกสาร (Clustering and Organizing Document) การวิเคราะห์คำศัพท์ (Lexical Analysis) เป็นต้น สำหรับการวิเคราะห์ข้อความมีกระบวนการที่แตกต่างจากการวิเคราะห์เหมืองข้อมูลเนื่องจากข้อความมีลักษณะข้อมูลแบบไม่เป็นโครงสร้าง หรือไม่มีทางที่แน่นอน ดังนั้นกระบวนการวิเคราะห์ข้อความจำเป็นต้องมีการ

แปลงรูปแบบข้อความให้เป็นลักษณะข้อมูลที่มีโครงสร้างก่อนการนำไปใช้ในการวิเคราะห์ซึ่งเรียกว่า กระบวนการเตรียมข้อมูล เป็นขั้นตอนที่สำคัญและต้องใช้เวลา [11]

การเตรียมข้อมูลเพื่อใช้ในการค้นหาลักษณะแฝงของข้อมูลได้แบ่งเป็น 3 กระบวนการ [12] ได้แก่ การคัดเลือกข้อมูล (Data selection), การกลั่นกรองข้อมูล (Data cleaning), การสร้างข้อมูลที่เหมาะสม (Data constructing) เช่น การ Derived attributes หรือ การ Generated records เป็นต้น, การรวมข้อมูล (Integrate Data) และสุดท้ายคือการแปลงรูปข้อมูล (Data transformation) ให้อยู่ในรูปแบบที่มีโครงสร้างที่เหมาะสม สำหรับการวิเคราะห์เหมืองข้อความในงานวิจัยนี้ได้จำแนก กระบวนการเตรียมข้อมูลออกเป็น 3 กระบวนการหลัก ได้แก่

สำหรับงานวิจัยนี้ มีการเตรียมข้อมูลเพื่อนำไปวิเคราะห์ตามขั้นตอนดังนี้

1. ทำเอกสารให้เป็นมาตรฐาน (Document Standardization) ที่มีรูปแบบที่ชัดเจน เพื่อสะดวกในการประมวลผล
2. การแยกคำศัพท์ (Word Segmentation) เป็นการแยกข้อความในเอกสารโดยพิจารณาช่องเว้นวรรคระหว่างคำหรือเครื่องหมายวรรคตอน
3. การตัดคำที่ไม่มีนัยสำคัญ (Stop Word Removal) ซึ่งคำเหล่านี้เกิดขึ้นบ่อยแต่ไม่ได้แสดงถึงเนื้อหาหรือไม่มีความหมาย ไม่มีประโยชน์ต่อการจำแนกประเภทเอกสาร เช่น คำสรรพนาม คำสันธาน และคำบุพบท เป็นต้น
4. การระบุหน้าที่ของคำ (Part of Speech Tagging) เพื่อให้ข้อมูลของคำศัพท์มากขึ้น

2.4.1 การทำเหมืองความคิดเห็น (Opinion Mining)

การทำเหมืองความคิดเห็นเป็นเครื่องมือที่ใช้ในการค้นหาความคิดเห็น ความรู้สึก หรือ จุดมุ่งหมายของข้อความ เพื่อที่จะสรุปว่าข้อความนั้นเป็นความคิดเห็นเชิงบวกหรือข้อความนั้นเป็นความคิดเห็นเชิงลบ มีความคิดเห็นเกี่ยวข้องกับอะไร แนวความคิดพื้นฐานของการทำเหมืองความคิดเห็น สามารถแบ่งได้เป็น 2 แบบ คือ แบบใช้ข้อมูลสอน (Supervised Approach) โดยการใช้การเรียนรู้ของเครื่อง (Machine Learning) และแบบไม่ใช้ข้อมูลสอน (Unsupervised Approach) โดยประสิทธิภาพของการทำเหมืองความคิดเห็นทั้ง 2 วิธีจะขึ้นอยู่กับข้อมูลที่สอน การทำเหมืองความคิดเห็นแบบใช้ข้อมูลสอนเป็นการใช้องค์ความรู้เกี่ยวกับการประมวลผลภาษาธรรมชาติ และการประยุกต์ใช้เครื่องมือเกี่ยวกับภาษา เช่น Word Net เพื่อใช้ในการระบุหน้าที่ของคำ รวมถึงการขยายคำศัพท์ ใช้คำที่มีความหมายเหมือนกันคำที่มีความหมายตรงกันข้าม โดยประสิทธิภาพการทำเหมืองความคิดเห็นนั้นจะขึ้นอยู่กับปริมาณคำศัพท์ ที่ครอบคลุม ความเป็นไปได้ของความคิดเห็นที่เกี่ยวข้อง

การวิเคราะห์ข้อความที่มีโครงสร้างไม่แน่นอนจำเป็นต้องมีการแปลงข้อความให้อยู่ในรูปแบบที่มีโครงสร้างก่อน เพื่อให้คอมพิวเตอร์สามารถประมวลผลได้ การแทนข้อความให้อยู่ในรูปเวกเตอร์ เป็นการแทนที่ข้อความให้มีโครงสร้าง สามารถเลือกคุณลักษณะของข้อความมาแทนได้หลายวิธี เช่น คำ วลี หรือหน้าที่ของคำ เป็นต้น โดยปกติแล้วการวิเคราะห์ข้อความจะต้องวิเคราะห์จากคำหลัก เรียกว่าการแทนข้อความด้วยถุงคำ (Bag-of-words) เป็นการแทนคำทุกคำในเอกสารด้วยเวกเตอร์ จากนั้นวิเคราะห์ความถี่ของคำที่เกิดขึ้นสามารถวัดได้ด้วยค่าความถี่คำ-ค่าส่วนกลับความถี่เอกสารที่เกิดคำ (Term Frequency – Inverse Document Frequency: TF-IDF) แนวความคิดของการแทนค่าข้อความด้วยค่า TF-IDF เกิดจากแนวความคิดว่าการแทนข้อความด้วยค่าความถี่อย่างเดียวไม่สามารถจำแนกข้อความได้มีประสิทธิภาพเพียงพอ เนื่องจากค่าความถี่ของคำสูง หมายถึง คำนั้นมีโอกาสเกิดขึ้นในหลายๆเอกสารพร้อมกัน จึงไม่มีประโยชน์ต่อการจำแนกข้อความ ดังนั้น [13] การวัดความถี่ของคำด้วย TF-IDF ที่เพิ่มการคำนึงถึงความถี่ของคำในเอกสารนั้นด้วย โดยคำนำนหนักของคำที่ได้ ดังสมการ นี้

$$w_d = f_{w,d} * \log (|D| / f_{w,d})$$

$f_{w,d}$ คือ ความถี่ของคำ (Term Frequency: TF)

D คือ จำนวนเอกสารทั้งหมดที่นำมาวิเคราะห์

w คือ ค่าความถี่ของเอกสารที่พบคำ

สำหรับการวิเคราะห์ข้อความในระดับประโยคแทนด้วยเอกสารทั้งประโยค ดังนั้นค่า TF คือ ค่าที่เกิดจากการคำนวณค่าความถี่ของคำที่พบในประโยค และ IDF คือ ค่า Log ของสัมประสิทธิ์ของจำนวนประโยคทั้งหมดที่นำมาวิเคราะห์หารด้วยค่าความถี่ของเอกสารที่พบคำ

อย่างไรก็ตาม การแทนข้อความด้วยวิธีการคำนวณค่าจากคำทั้งหมดอาจไม่เหมาะสมเนื่องจากมีคำบางคำที่สามารถเป็นตัวแทนของเอกสารหรือข้อความที่ดี และตัดคำที่ไม่สามารถใช้ตัวแทนเอกสารหรือข้อความที่ดีได้ออก การตัดคำที่ไม่มีความสำคัญ (Stop word removal) คำที่เกิดขึ้นบ่อยในหลายเอกสารแต่ไม่มีประโยชน์ต่อการจำแนกเอกสาร ถ้าตัดคำเหล่านี้ออกไปจะไม่ทำให้ความหมายของเอกสารเปลี่ยนแปลง เช่น คำเชื่อม คำบุพบท คำหยุด เป็นต้น ซึ่งวิธีนี้เป็นอีกวิธีหนึ่งที่ช่วยเพิ่มประสิทธิภาพการจำแนกข้อความและประหยัดพื้นที่ในการจัดเก็บข้อความได้อีกด้วย

2.4.2 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

ภาษาเป็นเครื่องมือที่ใช้ในการสื่อสารทั้งกับมนุษย์ด้วยกันเองหรือแม้กระทั่งสื่อสารกับคอมพิวเตอร์ แต่ภาษาที่ใช้มีรูปแบบที่แตกต่างกันไป ภาษาที่มนุษย์ใช้สื่อสารกับคอมพิวเตอร์ เป็นภาษาที่มีโครงสร้างแน่นอน คอมพิวเตอร์สามารถนำไปประมวลผลได้ทันที เช่น php, java หรือ C++ เป็นต้น แต่สำหรับภาษาที่มนุษย์ใช้สื่อสารกันเองเป็นภาษาที่ไม่มีโครงสร้างหรือรูปแบบเฉพาะตัว และเป็นไปตามธรรมชาติของการเรียนรู้ในสมองมนุษย์แต่ละคนซึ่งมีลักษณะที่แตกต่างกันไป เรียกว่า “ภาษาธรรมชาติ” เป็นภาษาที่ไม่มีโครงสร้างที่แน่นอน การที่คอมพิวเตอร์พยายามทำความเข้าใจกับภาษาธรรมชาติที่มนุษย์ใช้สื่อสารกันนั้น สามารถทำได้ด้วยวิธีการแทนความรู้ การสร้างกฎเกณฑ์ และการประเมินค่าเพื่อหาความหมายของภาษา [14] ดังนั้นการที่คอมพิวเตอร์จะเข้าใจภาษาธรรมชาติได้ดีเพียงไรนั้นขึ้นอยู่กับ 2 แนวทางหนึ่ง คือ พัฒนาการทางด้านปัญญาประดิษฐ์ซึ่งเป็นวิธีการแทนความรู้ (Knowledge representation) และ อีกแนวทางหนึ่งคือการศึกษาและเข้าใจโครงสร้างทางภาษาศาสตร์แบบมีโครงสร้าง [15] ซึ่งทั้งสองกระบวนการดังกล่าวถูกเรียกว่าการประมวลผลภาษาธรรมชาติ โดยระบบประมวลผลภาษาธรรมชาติจะรับข้อมูลอินพุตเป็นข้อความและแทนค่าข้อความด้วยแนวทางการวิเคราะห์ต่าง ๆ เช่น ความรู้ หรือโครงสร้าง เป็นต้น

การประมวลผลภาษาธรรมชาติได้แบ่งระดับขั้นการวิเคราะห์ ดังนี้ [16]

1. การวิเคราะห์ระดับวจีวิภาค (Morphological analysis) เป็นการวิเคราะห์ระดับคำ
2. การวิเคราะห์ระดับวากยสัมพันธ์ (Syntactic analysis) เป็นการวิเคราะห์คำตามหน้าที่ของคำ (Part-of-Speech) เพื่อเป็นข้อมูลพื้นฐานในการตรวจสอบโครงสร้างทางไวยากรณ์เกี่ยวกับการวางตำแหน่งของคำ กลุ่มคำประเภทต่าง ๆ ที่รวมกันเป็นประโยค
3. การวิเคราะห์ระดับความหมาย (Semantic analysis) เป็นการวิเคราะห์เพื่อให้ทราบความหมายของคำแต่ละคำในประโยค
4. วิเคราะห์ระดับวาทนิพจน์ (Discourse integration) เป็นการพิจารณาความหมายของประโยคโดยดูจากประโยคข้างเคียงร่วมด้วย
5. การวิเคราะห์ระดับปฏิบัติ (Pragmatic analysis) เป็นการแปลความหมายของประโยคถึงสิ่งที่ผู้พูดต้องการสื่อความหมายถึง

กระบวนการวิเคราะห์ภาษาจะเริ่มต้นที่ระดับต่ำ สุดก่อนคือการวิเคราะห์ระดับวจีวิภาคหรือคำ ไปจนถึงระดับวากยสัมพันธ์ ที่สามารถอธิบายได้ด้วยโครงสร้างของภาษาที่ประกอบด้วยหน่วยต่าง ๆ ดังต่อไปนี้

1. คำ (Word) คือ หน่วยหนึ่งที่เปล่งเป็นเสียงออกมาจะเป็นอิสระหรือไม่ก็ได้
2. หน่วยคำ (Morpheme) คือ ส่วนประกอบที่มีนัยสำคัญที่แท้จริงทางภาษา คือ คำ และกลุ่มคำ หรือหมายถึงหน่วยคำที่เล็กที่สุดที่มีความหมาย ความแตกต่างระหว่าง หน่วยคำ กับ คำ

คือ หน่วยคำอาจเกิดขึ้นอิสระหรือไม่อิสระก็ได้ แต่คำต้องเป็นหน่วยอิสระเสมอ เช่น นักเรียน “นัก” เป็นหน่วยคำ และ “นักเรียน” จึงถือเป็นคำ

3. วลี (Phrase) คือ กลุ่มของคำหรือคำ ๆ เดียวก็ได้ ที่เป็นส่วนประกอบของประโยค (วลีเป็นส่วนประกอบของประโยค แต่คำไม่ใช่ส่วนประกอบของประโยค) ดังนั้น ไม่ว่าจะเป็นคำๆ เดียว หรือกลุ่มที่ประกอบด้วยหลาย ๆ คำ เมื่อเป็นส่วนประกอบของประโยคเราเรียกเป็น วลี

4. นามวลี (Noun phrase) คือวลีที่ทำหน้าที่เป็นส่วนของประโยคที่เป็นหน่วย ประธาน หน่วยกรรม นามวลี

5. กริยาวลี (Verb phrase) คือวลีที่ทำหน้าที่เป็นภาคแสดงของประโยค

6. ประโยค (Sentence) คือคำหลายคำเรียงกันในการพูดหรือเขียน เพื่อเป็นการแสดงความคิด 1 ความคิดอย่างสมบูรณ์ทางไวยากรณ์ โดยปกติมักประกอบไปด้วยประธานและภาคแสดง

7. ไวยากรณ์ (Grammar) คือ ภาษาที่มีกฎเกณฑ์ และเป็นไปตามระเบียบวิธีการ ประกอบรูปคำให้เป็นประโยค

การวิเคราะห์ในระดับที่สูงขึ้นกว่าระดับวากยสัมพันธ์หรือโครงสร้างของภาษา จำเป็นต้องอาศัยฐานความรู้ที่สูงขึ้น เช่น ฐานความรู้เครือข่ายคำ (WordNet) สำหรับกระบวนการวิเคราะห์ภาษาธรรมชาติในภาษาไทยส่วนใหญ่ เป็นการวิเคราะห์ภาษาในระดับวลีภาคและระดับวากยสัมพันธ์ เท่านั้น เนื่องจากเครือข่ายคำไทย (Thai WordNet) ค่อนข้างมีจำนวนจำกัดและยังอยู่ในขั้นพัฒนา [17]

การวิเคราะห์ภาษาในระดับวลีภาคและวากยสัมพันธ์สำหรับภาษาไทย มีความยุ่งยากกว่าภาษาอังกฤษเล็กน้อย เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีการแบ่งวรรคตอนระหว่างประโยคหรือระหว่างคำที่แน่นอน มักเขียนติดกันไปทั้งย่อหน้า ดังนั้นการวิเคราะห์ในระดับต่ำสุดหรือคำ จึงจำเป็นต้องแบ่งเอกสารตั้งแต่หน่วยใหญ่สุดคือประโยค ให้มีหน่วยเล็กสุดที่มีความหมายก่อน ซึ่งหมายถึงคำ แล้วจึงนำคำแต่ละคำมาประกอบเป็นโครงสร้างประโยค ดังนั้น การวิเคราะห์ภาษาสำหรับภาษาไทยต้องอาศัย 3 กระบวนการพื้นฐานคือ (1) กระบวนการตัดคำ (Tokenization) เพื่อให้มีความสามารถในการวิเคราะห์ระดับวลีภาคได้ (2) กระบวนการกำกับคำตามหน้าที่ของคำ (Part-of-Speech tagging) และ (3) กระบวนการวิเคราะห์โครงสร้างภาษา (Syntactic analysis) เพื่อให้สามารถนำคำแต่ละคำมาประกอบเป็นโครงสร้างประโยคได้

2.4.3 เทคนิคที่ใช้ในการตัดคำ

เทคนิคที่ใช้ในการตัดทำได้ 3 วิธี [18] ได้แก่

1. การใช้กฎไวยากรณ์ทางภาษา (Rule-based) เป็นการสร้างกฎเกณฑ์ให้กับข้อความ มีการกำหนดลักษณะ การผสม สระอักษร การเว้นวรรคหรือการขึ้นย่อหน้า ข้อดี คือ สามารถค้นหาความเกี่ยวข้องของข้อความได้อย่างรวดเร็ว ไม่จำเป็นต้องมีคลังคำศัพท์ที่อยู่ในหน่วยความจำ ข้อเสีย ความยากในการกำหนดหลักไวยากรณ์ให้สมบูรณ์กับข้อความ และไม่สามารถใช้กับคำเฉพาะหรือคำที่มีความหมายคลุมเครือได้

2. การอ้างอิงคำจากพจนานุกรม (Dictionary-based) การตัดคำโดยพจนานุกรมเป็นการตัดคำ โดยใช้สายอักขระ (String) มาเปรียบเทียบกับคำที่มีอยู่ในพจนานุกรม ข้อดี คือ มีความถูกต้องแม่นยำสูงในการค้นหาความสัมพันธ์สามารถแก้ปัญหาเมื่อมีคำศัพท์เพิ่มเติมโดยการปรับปรุงคลังคำศัพท์ให้เป็นปัจจุบันมากขึ้น ข้อเสีย คือ จำเป็นต้องมีหน่วยความจำเป็นปริมาณมากในการจัดทำคลังคำศัพท์ ยากแก่การปรับคลังคำศัพท์สำหรับกรณีที่มีคำศัพท์ใหม่เกิดขึ้น

3. การสร้างโมเดลการเรียนรู้จากฐานข้อความขนาดใหญ่ (Machine Learning or Corpus based) การตัดคำโดยใช้คลังข้อมูลเป็นการตัดคำ โดยนำวิธีการทางสถิติเข้ามาใช้ในการประมวลภาษา โดยใช้คลังข้อมูลทางภาษาเป็นฐานความรู้เกี่ยวกับค่าความถี่ที่ใช้ในการตัดคำ ข้อดี ไม่จำเป็นต้องมีหน่วยความจำสำหรับเก็บคลังคำศัพท์ สามารถแก้ปัญหาคำศัพท์ที่ไม่สามารถหาความสัมพันธ์ได้โดยใช้การเรียนรู้จากฐานข้อความขนาดใหญ่ และสามารถสร้างกฎไวยากรณ์ทางภาษาขึ้นมาได้อัตโนมัติฐานข้อมูลขนาดใหญ่ ข้อเสีย คือ ฐานข้อมูลขนาดใหญ่ต้องมีปริมาณมากพอที่จะ ทำให้เกิดการเรียนรู้โดยให้อัลกอริทึม

2.4.4 วิธีการเรียนรู้ของเครื่อง (Machine Learning Approach)

การจำแนกเอกสารหรือข้อความ เป็นเทคนิคหนึ่งในการเรียนรู้ของเครื่อง ซึ่งงานวิจัยนี้ได้ใช้วิธีการเรียนรู้จากชุดข้อมูลเรียนรู้ เพื่อนำมาสร้างเป็นแบบจำลองแบบอัตโนมัติในการจำแนกข้อความ โดยข้อมูลที่นำมาเรียนรู้นั้นจะต้องมีการกำหนดผลเฉลยไว้แล้ว เพื่อให้คอมพิวเตอร์สามารถเรียนรู้รูปแบบของข้อมูล และสร้างแบบจำลองเพื่อไว้ทำนายหรือจัดกลุ่มของชุดข้อมูลทดสอบได้

เทคนิคการจำแนก (Classification) เป็นเทคนิคในการจำแนกข้อมูลด้วยคุณลักษณะต่างๆ ที่กำหนดไว้แล้ว เทคนิคประเภทนี้เหมาะกับการสร้างแบบจำลองเพื่อการพยากรณ์ข้อมูล (Predictive Modeling) เช่น การแบ่งประเภทลูกค้า เป็นต้น กระบวนการสร้างแบบจำลองนั้นถ้าปริมาณชุดข้อมูลเรียนรู้นั้นมาก จะทำให้แบบจำลองการจำแนกข้อความที่มีประสิทธิภาพสูงขึ้นด้วย แต่ระยะเวลาที่เครื่องใช้ในการสร้างแบบจำลองก็มากตามไปด้วย ในส่วนของงานวิจัยที่ศึกษา ผู้วิจัยจึงขอกว่าถึง 3 อัลกอริทึมสำหรับการจำแนกข้อความ ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) นาอ์ฟเบย์ (Naïve-Bayes) และ แรนดอมฟอเรส (Random Forest)

2.4.4.1 ต้นไม้ตัดสินใจ (Decision Tree)

อัลกอริทึมต้นไม้ตัดสินใจ คือแบบจำลองทางคณิตศาสตร์ เพื่อหาทางเลือกที่ดีที่สุด เป็นเทคนิคหนึ่งสำหรับการจำแนก (Classification) แบบโครงสร้างต้นไม้ ซึ่งเป็นวิธีการแบ่งประเภท หรือแยกหมวดหมู่ข้อมูล ซึ่งเป็นเทคนิคหนึ่งสำหรับการทำเหมืองข้อมูล

ส่วนประกอบของต้นไม้ตัดสินใจเป็นอัลกอริทึมจำแนกข้อมูล ประกอบด้วย (1) โหนด (Node) คือ คุณสมบัติต่างๆเป็นจุดที่แยกข้อมูลว่าจะไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุด เรียกว่า โหนดราก (Root node) (2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติโหนดที่แตกออกมา โดยจำนวนกิ่งจะเท่ากับคุณสมบัติของโหนด (3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล การสร้างต้นไม้ตัดสินใจจะสร้างในลักษณะบนลงล่าง คือ สร้างจากรากแล้วจึงแตกกิ่งไปจนถึงใบ ขั้นตอนการสร้างต้นไม้ตัดสินใจ เริ่มจากการสร้างโหนดรากซึ่งจำแนกโหนดตามกลุ่มข้อมูล หากในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่จะพิจารณาจากค่าความจริงของคุณลักษณะ โดยค่าที่ใช้จะมาจากการคำนวณจากค่า Information Gain ของแต่ละกลุ่มเพื่อที่จะเป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ ที่มีความสามารถในการแบ่งแยกข้อมูลเป็นกลุ่มได้ดีที่สุด ดังนั้นค่าที่มากที่สุดจะถูกเลือกให้เป็นตัวทดสอบใช้ในการตัดสินใจ กิ่งของต้นไม้จะถูกสร้างจากค่าต่างๆที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งตามกิ่งต่างๆที่สร้างขึ้น จากนั้นทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกณฑ์มากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อไป และมีการวนซ้ำ เพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อยๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งเป็นจริง

การคำนวณค่า Information Gain การสร้างต้นไม้ตัดสินใจ ใช้ค่ามาตรฐานอัตราส่วน เกน (Gain Ratio) เพื่อเลือกคุณลักษณะที่จะใช้เป็นรากหรือโหนด ถ้าให้ชุดข้อมูล M ประกอบด้วยค่าที่เป็นไปได้คือ $\{m_1, m_2, \dots, m_n\}$ และให้ความน่าจะเป็นที่จะเกิดค่า m_1 มีค่าเท่ากับ $P(m_1)$ จะได้ว่าค่าเกนสารสนเทศ (Information Gain) ของ M เขียนแทนด้วย $I(M)$ คำนวณได้ดังสมการ [19]

$$I(M) = -\sum_{m=1}^n P(m) \log_2 P(m)$$

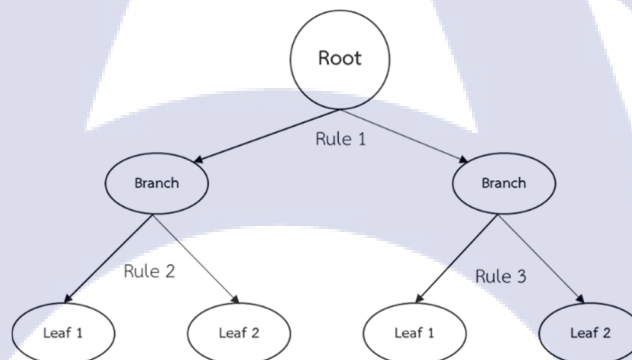
ถ้าให้ข้อมูลสอน คือ T คุณลักษณะทั้งที่เป็นโหนดคือ x และมีค่าทั้งหมดทั้งที่เป็นไปได้ n ค่าโหนดปัจจุบันจะสามารถแบ่งตัวอย่าง T ออกตามกิ่งเป็น $\{t_1, t_2, \dots, t_n\}$ ตามค่าที่เป็นไปได้ของ x ดังนั้นจึงสามารถคำนวณค่าเกนสารสนเทศ หลังจากแบ่งตามคุณลักษณะและค่ามาตรฐานเกน (GAIN) ของคุณลักษณะ x ได้ ดังสมการ

$$Gant(x) = I(T) - I_x(T)$$

จากนั้นคำนวณค่าสารสนเทศของการแบ่งแยก (Split Information) ของคุณลักษณะแต่ละตัว ถ้าให้ T คือ ชุดของตัวอย่าง เมื่อแบ่งตัวอย่างนี้ตามคุณลักษณะ x จะได้ชุดของตัวอย่างย่อยในแต่ละกิ่ง คือ $\{t_1, t_2, \dots, t_n\}$ จำนวน n ชุด ตามค่าที่เป็นไปได้ในคุณสมบัติ x เมื่อคำนวณค่าสารสนเทศของการแบ่งแยกได้ดังสมการ

$$Split\ Information = \sum_{t=1}^T \frac{t_1 l}{Tl} \log_2 \frac{t_1 l}{Tl}$$

คำนวณค่ามาตรฐาน อัตราส่วนเกน (Gain Ratio) ได้จาก $Gain\ Ratio = Gain - Split\ Information$ ทำยสุดจึงเลือกค่า Gain Ratio สูงสุดเป็นคุณลักษณะเริ่มต้น และเลือกคุณสมบัติ ถัดไปตามค่า Gain Ratio น้อยลงตามลำดับ



รูปที่ 2.3 ต้นไม้ตัดสินใจ (Decision Tree)

กระบวนการสร้างต้นไม้ตัดสินใจเริ่มจากคำนวณค่าเกณฑ์การเลือกจากทุกคุณลักษณะในชุดข้อมูล แล้วพิจารณาเลือกคุณลักษณะจากค่าเกณฑ์สารสนเทศสูง (หรือค่าเอ็นโทรพิต่ำ) หรือค่าดัชนี Gini ต่ำก่อน (เพราะถือว่าคุณลักษณะดังกล่าวมีความสามารถในการจำแนกหมวดหมู่สูง) จากนั้นนำคุณลักษณะดังกล่าวมาสร้างเป็นโหนดราก ชุดข้อมูลที่ผ่านโหนดรากจะถูกแบ่งกลุ่มตามค่าในคุณลักษณะที่เป็นไปได้ (ค่าเงื่อนไขการตัดสินใจ) จากนั้นสร้างโหนดกิ่ง โดยคำนวณค่าเกณฑ์การเลือกจากคุณลักษณะที่เหลือ ทำซ้ำกระบวนการเดิมเพื่อสร้างโหนดกิ่งไปจนกว่าชุดข้อมูลที่ถูกแบ่งกลุ่มตามเงื่อนไข จะอยู่ในคลาสเดียวกันทั้งหมด หรือมีค่าเอ็นโทรพิต่ำกับศูนย์ หมายถึงไม่มีการเปลี่ยนแปลงของคำตอบ ทุกข้อมูลให้ค่าความจริงเดียวกัน ตัวอย่าง อัลกอริทึมต้นไม้ตัดสินใจ เช่น

CART, ID3, C4.5 และ CHAID ข้อดีของอัลกอริทึมต้นไม้ตัดสินใจคือ ผู้ใช้สามารถทำความเข้าใจแบบจำลองการตัดสินใจของต้นไม้ได้ แต่ไม่รองรับข้อมูลแบบต่อเนื่อง (Continuous data) หากข้อมูลมีลักษณะดังกล่าว ต้องแบ่งข้อมูลให้เป็นแบบไม่ต่อเนื่องหรือแบบช่วงข้อมูล (Discrete data) ก่อน นอกจากนั้นประสิทธิภาพการจำแนกข้อความอาจดีกว่าวิธีอื่นด้วย จึงมักใช้เป็นพื้นฐานสำหรับเปรียบเทียบผลลัพธ์กับอัลกอริทึมอื่นเท่านั้น

งานวิจัยนี้เลือกใช้อัลกอริทึม Decision Tree แบบ C4.5 เนื่องจาก C4.5 ถูกพัฒนามาจาก ID3 ให้เพิ่มประสิทธิภาพในการคำนวณจัดการได้ดีกับค่าต่อเนื่อง และคุณสมบัติที่เป็นค่าว่าง นอกจากนี้ C4.5 มีความสามารถในการตัดเล็มต้นไม้ตัดสินใจ แก้ปัญหาข้อมูลโอเวอร์ฟิตได้ ดังนั้นงานวิจัยนี้จึงเลือก C4.5 มาใช้ในการประเมินประสิทธิภาพโมเดล

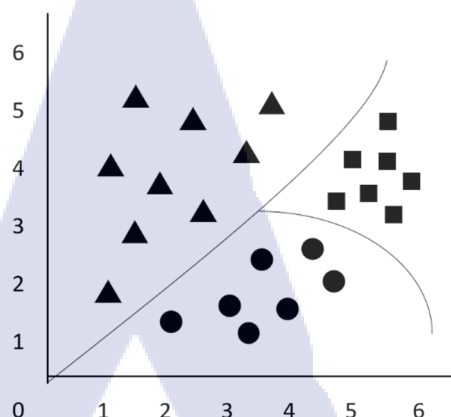
2.4.4.2 นาอิวเบย์ (Naïve-Bayes)

นาอิวเบย์ (Naïve-Bayes) เป็นวิธีการเรียนรู้เพื่อสร้างแบบจำลองการจำแนกเอกสารที่มีพื้นฐานมาจากทฤษฎีความน่าจะเป็นของเบย์ (Bay theorem) แบบจำลองที่ได้จะอยู่ในรูปแบบของความน่าจะเป็น ซึ่งอาศัยค่าความน่าจะเป็นจากชุดข้อมูลเรียนรู้ หรือเรียกว่าความรู้ก่อนหน้า (Prior probability) มาทำนายผลของชุดข้อมูลทดสอบ ด้วยวิธีคำนวณค่าความน่าจะเป็นของชุดข้อมูลที่จะอยู่ในคลาส C คำนวณจนครบทุกคลาส ความน่าจะเป็นของคลาสใดที่มากที่สุดจะถูกเลือกเป็นคำตอบนาอิวเบย์ถูกใช้ในการทำนายผลจัดเป็นเทคนิคในการแก้ปัญหาแบบ Classification ที่สามารถคาดการณ์ผลลัพธ์ได้และสามารถอธิบายได้ด้วย มันจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ การเรียนรู้เบย์อย่างง่ายเป็นวิธีจำแนกประเภทข้อมูลที่มีประสิทธิภาพวิธีหนึ่งที่มีการทำงานที่ไม่ซับซ้อน เหมาะกับกรณีของเซตตัวอย่าง มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน โดยกำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็น [19] ดังสมการ

$$P(a_1, a_2, \dots, a_n | v_j) = \pi P(a_i | v_j)$$

กลุ่ม v_j สำหรับข้อมูลที่มีคุณสมบัติ n ตัว $X = \{a_1, a_2, \dots, a_n\}$ หรือ ใช้สัญลักษณ์ว่า $P(a_1, a_2, \dots, a_n | v_j)$ โดยที่ $\prod$ หมายถึง ผลคูณของค่า $P(a_i | v_j)$ ทั้งหมด $i = 1, 2, 3, \dots, n$ และ $j = 1, 2, 3, \dots, n$ ดังนั้นเราจะได้ค่าวิธีการจำแนกประเภทแบบเบย์อย่างง่าย ดังสมการ

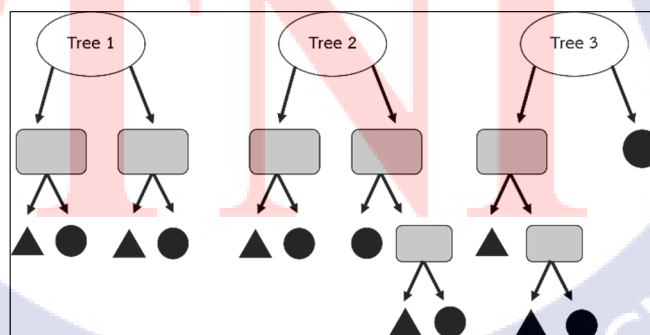
$$v_{NB} = \arg. \max P(v_1) \times \pi P(a_i | v_j)$$



รูปที่ 2.4 นาอีฟเบย์ (Naïve-Bayes)

2.4.4.3 แรนดอมฟอเรส (Random Forest)

Random Forest เป็นอัลกอริทึมที่พัฒนามาจากอัลกอริทึมต้นไม้ตัดสินใจ ลักษณะของต้นไม้ที่เกิดขึ้นจะถูกควบคุมด้วย 3 ปัจจัย ได้แก่ (1) การฝึกสอนโดยใช้เซตข้อมูลตัวอย่าง (2) เมื่อต้นไม้เจริญเติบโตจะสามารถค้นหาโหนดแต่ละโหนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้โดยการสุ่มเลือกคุณสมบัติทั้งหมด (3) ต้นไม้แต่ละต้นไม่มีการตัดออกแต่จะปล่อยให้เจริญเติบโตไปเรื่อยๆ จนได้ผลลัพธ์ที่ดีที่สุดจากการสร้างป่า จากนั้นทำการโหวตโดยต้นไม้ในป่า หากต้นไม้ใดได้รับคะแนนสูงสุดจะถูกนำมาสร้างเป็นโมเดล โดยการสร้างรูปแบบจากอัลกอริทึมต้นไม้ตัดสินใจหลายๆรูปแบบจากข้อมูลชุดใหญ่นำมาทำเป็นข้อมูลชุดย่อย โดยคุณลักษณะของข้อมูลจะมีข้อมูลบางส่วนที่ไม่ถูกเลือกที่เรียกว่า Out-of-Bag (OOB) จากนั้นจะนำข้อมูลไปสุ่มเลือกผลลัพธ์คุณลักษณะข้อมูลโดยการโหวตผลลัพธ์ที่ถูกเลือกทำให้ได้เป็นเงื่อนไขในการแสดงคุณลักษณะที่ต้องการ



รูปที่ 2.5 แรนดอมฟอเรส (Random Forest)

2.4.5 การประเมินประสิทธิภาพการจำแนกประเภทข้อความ (Text Classification Evaluation)

เมื่อผู้วิจัยได้เลือกอัลกอริทึมที่ใช้ในการจำแนกข้อความเพื่อสร้างรูปแบบในการใช้งาน ได้แก่ ต้นไม้ตัดสินใจ นาอ์ฟเบย์ และแรนดอมฟอร์เรส จากนั้นต้องมีการประเมินรูปแบบจากอัลกอริทึม โดยวิธีการดังนี้

2.4.5.1 การวัดประสิทธิภาพในการจำแนกข้อความ

ในการสร้างเงื่อนไขเพื่อจำแนกข้อความ ความสามารถในการจำแนกข้อความเป็นตัวชี้วัดที่สำคัญเพื่อที่จะยืนยันความถูกต้องการจำแนกข้อความที่นำเสนอขึ้นมาใหม่ ได้แก่

1. ค่าความผิดพลาด (Error Rate)
2. ค่าความถูกต้อง (Accuracy)
3. ค่าความแม่นยำ (Precision)
4. ค่าความระลึก (Recall)
5. ค่าประสิทธิภาพโดยรวม (F-measure)

1. ค่าความผิดพลาด (Error Rate) คือ อัตราส่วนของจำนวนข้อมูลที่ไม่สามารถจำแนกได้ตรงกับผู้เชี่ยวชาญ และจำนวนข้อมูลผ่านการจำแนกทั้งหมด ค่าความผิดพลาดจะมีค่าระหว่าง 0-1 หากมีค่าน้อยกว่าแสดงว่ามีการจำแนกที่ดีกว่า ดังสมการนี้

$$\text{ค่าความผิดพลาด (Error Rate)} = \frac{\text{จำนวนข้อมูลที่ไม่สามารถจำแนกได้ตรงกับผู้เชี่ยวชาญ}}{\text{จำนวนข้อมูลผ่านการจำแนกทั้งหมด}}$$

2. ค่าความถูกต้อง (Accuracy) คือ อัตราส่วนของจำนวนข้อมูลที่สามารถจำแนกได้ถูกต้องตรงกับผู้เชี่ยวชาญ และจำนวนข้อมูลผ่านการจำแนกทั้งหมด ค่าความถูกต้องจะแสดงให้ทราบถึงที่น่าจะเป็นที่อัลกอริทึมนั้นจะทำนายถูก ดังนั้นยิ่งค่าความถูกต้องสูงแสดงว่ามีการจำแนกที่ดีกว่า ดังสมการต่อไปนี้

$$\begin{aligned} \text{ค่าความถูกต้อง (Accuracy)} &= \frac{\text{จำนวนข้อมูลที่สามารถจำแนกได้ถูกต้องตรงกับผู้เชี่ยวชาญ}}{\text{จำนวนข้อมูลผ่านการจำแนกทั้งหมด}} \\ &= 1 - \text{Error Rate} \end{aligned}$$

3. ค่าความแม่นยำ (Precision) คือ ค่าความสามารถในการจัดข้อมูลที่ไม่เกี่ยวข้องออกไปเป็นอัตราส่วนของการค้นพบข้อมูลที่ต้องการจากจำนวนเอกสารทั้งหมดที่ทำการค้นคืนมาได้ ถ้าค่าความแม่นยำเท่ากับ 1 หมายถึง ระบบสามารถค้นคืนข้อมูลได้ถูกต้องโดยที่ไม่มีเอกสารที่ไม่เกี่ยวข้องปะปนอยู่

ตารางที่ 2.1 ประเภทการค้นคืนข้อมูล

Category C_i		Expert judgment	
		Yes	No
Classifier judgement	Yes	True Positive	False Positive
	No	False Negative	True Negative

ค่าความถูกต้องเชิงบวก (True Positive) หมายถึง เอกสารที่อยู่กลุ่ม C_i และแบบจำลองทำนายว่าอยู่กลุ่ม C_i

ค่าความผิดพลาดเชิงบวก (False Positive) หมายถึง เอกสารไม่อยู่กลุ่ม C_i แต่แบบจำลองทำนายว่าอยู่กลุ่ม C_i

ค่าความผิดพลาดเชิงลบ (False Negative) หมายถึง เอกสารที่อยู่กลุ่ม C_i แต่แบบจำลองทำนายว่าไม่อยู่กลุ่ม C_i

ค่าความถูกต้องเชิงลบ (True Negative) หมายถึง เอกสารไม่อยู่กลุ่ม C_i และแบบจำลองทำนายว่าไม่อยู่กลุ่ม C_i

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

4. ค่าความระลึก (Recall) คือ การวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยเป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและถูกดึงออกมา กับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

5. ค่าประสิทธิภาพโดยรวม (F-measure) คำนวณจากค่าระลึกและค่าความแม่นยำจะนำมาพิจารณาาร่วมกัน ระบบที่มีประสิทธิภาพดี หมายถึง ค่าระลึกสูงด้วยค่าความแม่นยำที่สูงใกล้เคียงกัน

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

จากสมการประเมินประสิทธิภาพข้างต้นเป็นการประเมินประสิทธิภาพสำหรับชุดข้อมูลที่มี 2 กลุ่ม (Classes) สำหรับชุดข้อมูลที่มีมากกว่า 2 กลุ่ม สามารถประเมินประสิทธิภาพได้ด้วยวิธีการคำนวณค่า Micro averaging และ ค่า Macro averaging ค่า Micro averaging เป็นการวัดประสิทธิภาพด้วยค่าเฉลี่ยแบบให้น้ำหนักทุกเอกสารเท่ากัน (Document-pivoted measurement) โดยจะคำนวณค่าความถูกต้องจากจำนวนเอกสาร ซึ่งจะนำจำนวนเอกสารของแต่ละกลุ่มมารวมกัน เพื่อคำนวณหาค่าความถูกต้องในระดับเอกสาร ค่า Macro averaging เป็นการวัดประสิทธิภาพด้วยค่าเฉลี่ยแบบให้น้ำหนักทุกกลุ่ม (Classes) เท่ากัน (Category-pivoted measurement) โดยจะคำนวณหาค่าความถูกต้องของแต่ละกลุ่มก่อน จากนั้นนำมาเฉลี่ยรวมกันเพื่อให้ได้ค่าความถูกต้องของระบบในระดับกลุ่ม

$$\begin{aligned} Recall_{Micro} &= \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} TP_i + FN_i} \\ Precision_{Micro} &= \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} TP_i + FP_i} \\ Recall_{Macro} &= \frac{\sum_{i=1}^{|C|} Recall_i}{m} \\ Precision_{Macro} &= \frac{\sum_{i=1}^{|C|} Precision_i}{m} \end{aligned}$$

รูปที่ 2.6 สูตรการประเมินประสิทธิภาพของอัลกอริทึม

2.5 งานวิจัยที่เกี่ยวข้องกับการประเมินความคิดเห็นเกี่ยวกับการเรียนการสอนของอาจารย์

ณรงค์ฤทธิ์ ประสานตรี [20] รูปแบบการพัฒนาอาจารย์มหาวิทยาลัยราชภัฏกลุ่มรัตนโกสินทร์ วิจัยเกี่ยวกับการศึกษาสภาพปัญหาในการพัฒนาอาจารย์ สร้างรูปแบบและประเมินรูปแบบการพัฒนาอาจารย์มหาวิทยาลัยราชภัฏกลุ่มรัตนโกสินทร์ โดยใช้แบบสอบถามสภาพปัญหาและความต้องการของอาจารย์มหาวิทยาลัยราชภัฏกลุ่มรัตนโกสินทร์ ประเด็นการสนทนากลุ่ม และแบบประเมินรูปแบบการพัฒนา จากนั้นวิเคราะห์ข้อมูลเชิงปริมาณโดยใช้ค่าสถิติพื้นฐาน ได้แก่ ค่าร้อยละ ค่าคะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ข้อมูลเชิงคุณภาพทำการวิเคราะห์เนื้อหา ผลการวิจัยพบว่า รูปแบบการพัฒนาอาจารย์มหาวิทยาลัยราชภัฏกลุ่มรัตนโกสินทร์ ประกอบด้วยกิจกรรมหลัก -3 มิติ ได้แก่ 1) มิติการสร้างองค์ความรู้และถ่ายทอดความรู้สู่ท้องถิ่นองค์ประกอบที่สำคัญ คือ ด้านการวิจัย และการเขียนตำราและบทความทางวิชาการ 2) มิติการผลิตบัณฑิต มีองค์ประกอบที่สำคัญ คือ ด้านการสอน และด้านการเป็นอาจารย์ที่ปรึกษา 3) มิติการพัฒนา มีองค์ประกอบที่สำคัญ คือ ด้านการบริการวิชาการ และด้านการทำนุบำรุงศิลปและวัฒนธรรม ซึ่งผลการวิจัยนี้ทำให้ทราบถึงความคิดเห็นเกี่ยวกับการด้านการสอนโดยอาจารย์ ซึ่งไม่ใช่ความคิดเห็นที่มาจากนักศึกษาโดยตรงจึงอาจทำให้แนวทางการแก้ปัญหาไม่ตรงกับความต้องการของนักศึกษา

เรืองวิทย์ นนทภา [21] คุณลักษณะของครูที่ผู้เรียนประทับใจ : ต้นแบบของครูดี งานวิจัยนี้ศึกษาเกี่ยวกับเหตุการณ์ สถานภาพ คุณลักษณะของครูที่ผู้เรียนมีความประทับใจ โดยใช้แบบสอบถามชนิดปลายเปิดให้นักศึกษาบรรยายความประทับใจที่มีต่อครู 1 ท่าน จากนั้นวิเคราะห์ข้อมูลด้วยการวิเคราะห์เนื้อหา ผลการวิจัยพบว่า คุณลักษณะของครูในระดับชั้นอนุบาล ประถมศึกษา มัธยมศึกษา และอุดมศึกษาที่ผู้เรียนประทับใจประกอบด้วย 5 ด้าน ได้แก่ คุณลักษณะส่วนตัว การสอน วิชาการ จรรยาบรรณ และมนุษยสัมพันธ์ เหตุการณ์ที่ทำให้ผู้เรียนมีความประทับใจ คือ การช่วยเหลือลูกศิษย์ที่มีปัญหาเป็นที่ปรึกษาได้ทุกเรื่อง มีโอกาสได้ทำกิจกรรมร่วมกับลูกศิษย์ ครูได้แสดงให้เห็นถึงความห่วงใย ให้กำลังใจ ให้คำแนะนำ จนประสบความสำเร็จในการดำเนินกิจกรรมและโครงการ สำหรับคุณลักษณะประทับใจของอาจารย์ในระดับอุดมศึกษา ในด้านการสอน ได้แก่ 1) สอนเข้าใจง่าย อธิบายชัดเจนและใช้เทคโนโลยีเข้ามาประยุกต์ใช้ในการสอน 2) ผลักดันให้นักศึกษามีส่วนร่วมในการทำกิจกรรมหลายๆ อย่าง 3) มีความอดทนในการอธิบายให้นักศึกษาเป็นรายบุคคล เพื่อให้เข้าใจในบทเรียนและมีทักษะในการเรียน ทักษะภาษา 4) สอนสนุก ไม่เครียด ทำให้นักศึกษาล้าชักถาม และ 5) ฝึกให้นักศึกษาคิดวิเคราะห์ ทั้งเรื่องเนื้อหาความรู้ และจริยธรรม ด้านวิชาการ 1) มีความรู้ในเนื้อหา เทคนิคการสอน จิตวิทยา การวัดผล ประเมินผล สามารถประยุกต์ใช้ในการจัดกิจกรรมการเรียนการสอน 2) ใฝ่หาความรู้ สังเกต ปรับปรุงแก้ไขตนเองอยู่เสมอและ 3) ติดตามข่าวสารตลอดเวลา

งานวิจัยกลุ่มนี้แสดงให้เห็นถึงบทบาทของความคิดเห็นต่อการสอนของอาจารย์ว่า คุณลักษณะที่นักศึกษาต้องการมีลักษณะอย่างไร แต่งานวิจัยที่ศึกษานั้นยังคงเป็นรูปแบบการตอบจาก กลุ่มอาจารย์ผู้สอน และมีการสัมภาษณ์กลุ่มผู้เรียนจากนั้นวิเคราะห์บทสัมภาษณ์ที่มีผลออกมาในเชิง บวกเท่านั้น

2.6 งานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อความ

อดินต์ วัฒนบุรณนท์ [22] การจำแนกคุณภาพการใช้งานและปัญหาของเกมแอ็คชั่นจาก บทวิจารณ์ของผู้ใช้ด้วยการทำเหมืองข้อความ และใช้อัลกอริทึมนาอูฟเบย์ในการสกัดข้อความให้ กลายเป็นข้อมูลที่มีโครงสร้าง เพื่อหาความน่าจะเป็นที่ประโยชน์วิจารณ์เกมแอ็คชั่น นั้น เป็นเชิงบวก หรือเชิงลบ จากนั้นจึงประเมินประสิทธิผลของเครื่องมือจากการคำนวณค่าความแม่นยำโดย เปรียบเทียบกับการจำแนกโดยผู้เชี่ยวชาญ ซึ่งผลการทดสอบพบว่า ประสิทธิภาพเท่ากันทั้งหมดในรูป ของความแม่นยำเฉลี่ยของการจำแนกความคิดเห็นเชิงบวกและเชิงลบมีค่าร้อยละ 59.51 และปัจจัยที่ ส่งผลต่อคุณภาพของเกมและปัญหาของเกมมีค่าร้อยละ 64.46 และ 81.01 ตามลำดับ

กานดา แฝ่ววัฒนากุล [23] การวิเคราะห์เหมืองข้อเสนอแนะจากบทวิจารณ์รายการโทรทัศน์ เสนอกระบวนการแก้ปัญหาเกี่ยวกับการจำแนกบทวิจารณ์ในอินเทอร์เน็ตซึ่งมีจำนวนมาก ทั้ง ข้อเท็จจริง ข้อคิดเห็น และข้อเสนอแนะ รวมถึงโครงสร้างประโยคที่ไม่แน่นอนทำให้ยากต่อการ ตีความ ดังนั้นการจำแนกข้อมูลจะช่วยให้การประมวลผลดีขึ้น งานวิจัยนี้จะใช้ 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ นาอูฟเบย์ และซัพพอร์ตเวกเตอร์แมชชีน เพื่อเปรียบเทียบผลลัพธ์ที่ดีที่สุดในการ จำแนกข้อเสนอแนะออกจากบทวิจารณ์ประเภทอื่น โดยข้อเสนอแนะมีการจำแนกได้เป็น 4 ประเภท ได้แก่ ข้อเสนอแนะทางตรง ข้อเสนอแนะเชิงขอร้อง ข้อเสนอแนะเชิงคำถาม และข้อเสนอแนะเชิง เงื่อนไข ผลการทดสอบ พบว่า อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนแบบเคอร์เนลโพลีโนเมียลสามารถ จำแนกข้อเสนอแนะได้ดีที่สุด มีค่าความแม่นยำ 85.75% ค่าความระลึก 93.62% และค่าถ่วงดุล 89.51%

กานดา แฝ่ววัฒนากุล และปราโมทย์ ลือนาม [24] การวิเคราะห์เหมืองความคิดเห็นบน เครือข่ายสังคมออนไลน์ นำเสนอการทำเหมืองความคิดเห็นโดยนำข้อมูลบนเครือข่ายสังคมออนไลน์ เกี่ยวกับคุณลักษณะของผลิตภัณฑ์ด้านต่างๆ และจำแนกความคิดเห็นเป็น 3 ขั้ว ได้แก่ เชิงบวก เชิง ลบ และเชิงกลาง จากนั้นดำเนินการสร้างแบบจำลองจากชุดการเรียนรู้ และมีการประเมิน ประสิทธิภาพของแบบจำลอง และนำไปใช้สร้างระบบการวิเคราะห์ความคิดเห็น

S. Brain et al. [25] วิจัยเกี่ยวกับการใช้เทคนิคเหมืองข้อความ และการวิเคราะห์ความรู้สึก ในการค้นหาการรับรู้เกี่ยวกับร้านอาหารเอเชียโดยใช้ข้อความจากทวิตเตอร์ เพื่อค้นหารูปแบบที่มี นัยสำคัญ คำที่นิยมใช้ และคำที่แสดงถึงอารมณ์ในความคิดเห็น โดยใช้ ScraperWiki และ Microsoft

Excel ในการเก็บรวบรวมความคิดเห็น วิเคราะห์ และจัดการข้อมูล ผลการวิจัยประกอบด้วย 2 ส่วน ได้แก่ ส่วนที่ 1 ผลจากการวัดความถี่ของคำที่เกิดขึ้นในความคิดเห็นเกี่ยวกับร้านอาหาร ส่วนที่ 2 เป็นการเปรียบเทียบค่าความรู้สึกตามเกณฑ์เชิงบวก คะแนน 1 ถึง 4 กลาง 0 และ เชิงลบ -1 ถึง -4 ผลการวิจัยแสดงให้เห็นว่าความคิดเห็นเชิงบวกจะประกอบไปด้วยคำศัพท์ต่างๆที่เกี่ยวกับรสชาติของอาหาร แต่ความคิดเห็นเชิงลบจะประกอบไปด้วยความคิดเห็นต่อสิ่งๆอื่นในร้านอาหารมากกว่า เช่น อุปกรณ์ การบริการ สภาพแวดล้อม

R. Talib et al. [26] Text Mining : Techniques, Applications and Issues นำเสนอการอธิบายและวิเคราะห์ความแตกต่างของเทคนิคการทำเหมืองข้อความ โดยการอธิบายเกี่ยวกับความรู้ที่เกี่ยวข้องในขั้นตอนการทำเหมืองข้อความ เช่น Data Mining , Machine Learning , ความรู้ทางสถิติ , การสืบค้นข้อมูลสารสนเทศ และ ภาษาคอมพิวเตอร์การนำความรู้ที่ได้จากขั้นตอนการทำเหมืองข้อความไปประยุกต์ใช้ในอุตสาหกรรมต่างๆ เช่น การวิเคราะห์ข้อความจากข้อเสนอแนะเกี่ยวกับผลิตภัณฑ์ ระบบการจัดการความสัมพันธ์ของลูกค้า การป้องกันการทุจริต การวิเคราะห์ข้อความจากสื่อโซเชียลมีเดีย การวิเคราะห์ความรู้สึก การสร้างเครื่องมือทำนาย การสร้าง Search Engine เป็นต้น รวมถึงการแนะนำ 5 ขั้นตอนที่จะส่งผลกระทบต่อขั้นตอนการทำเหมืองข้อความ ดังนี้ 1. Information Extraction คือ เทคนิคในการสกัดข้อมูลจากข้อมูลขนาดใหญ่ ให้สามารถระบุ Attribute และ Entity ได้ จากนั้นจึงมีการตรวจสอบและประเมินความถูกต้องโดยใช้การวัดค่า Precision และ Recall 2. Information Retrieval คือ การใช้อัลกอริทึมต่างๆเพื่อหาความสัมพันธ์หรือรูปแบบของกลุ่มข้อความ 3. Natural Language Processing คือ กระบวนการวิเคราะห์ข้อความที่ไม่มีโครงสร้าง เพื่อสร้าง Named Entity Recognition (NER) สำหรับกลุ่มข้อความโดยใช้คำย่อ หรือคำที่มีความหมายเหมือนกัน 4. Clustering คือ การจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมต่างๆ ซึ่งการเลือกอัลกอริทึมจะแตกต่างกันตามลักษณะของข้อมูล ได้แก่ Hierarchical, Distribution, Density, Centroid และ K-means 5. Text Summarization คือ ขั้นตอนการรวบรวมข้อความทั้งหมดให้กระชับขึ้น ขั้นตอนนี้จะทำให้เกิด Lexicon ซึ่งเสมือนคลังคำศัพท์ในการจัดกลุ่มข้อความ

ปราชญธากย์ เหล่าสังข์สุข และคณะ [27] การวิเคราะห์ความคิดเห็นเกี่ยวกับร้านอาหารบนเว็บไซต์รีวิว นำเสนอระบบวิเคราะห์และสรุปความคิดเห็นเกี่ยวกับร้านอาหารจากเว็บไซต์รีวิวโดยอัตโนมัติ โดยใช้เทคนิคการทำเหมืองข้อความ โดยแบ่งรายละเอียดการสรุปออกเป็น 5 ข้อคือ (1) รสชาติอาหาร (2) บริการของร้าน (3) บรรยากาศของร้าน (4) ความสะอาด และ (5) ราคาอาหาร เพื่อหาความหมายเชิงบวกหรือเชิงลบของประโยค ซึ่งจากผลการประเมินความพึงพอใจของผู้ใช้ในด้านความสอดคล้องของผลสรุปจากระบบและผลสรุปจากผู้ใช้พบว่าการสรุปความคิดเห็นด้านบริการและอาหารอยู่ในระดับดี ขณะที่ด้านอื่นๆ อยู่ในระดับพอใช้

พนิดา ทรงรัมย์ [28] การจำแนกความคิดเห็นทางการเมืองบนเครือข่ายสังคมออนไลน์ โดยใช้วิธีการจำแนกแบบความสัมพันธ์ นำเสนอการจำแนกความคิดเห็นที่มีต่อรัฐประหารในประเทศไทย จากข้อความคิดเห็นบนเว็บไซต์เครือข่ายสังคมเฟซบุ๊ก โดยการรวบรวมข้อความคิดเห็นจำนวน 467 ข้อความ แบ่งเป็นเชิงบวก 259 ข้อความ และเชิงลบ 208 ข้อความ จากนั้นกำหนดคุณลักษณะโดยเรียงลำดับ ดังนี้ 1. อิทธิพล 2. โทง 3. ตกต่ำ 4. สงบ 5. ยั่งยืน 6. เจริญ อัลกอริทึมที่ใช้ในการจำแนกคือ Context-bounded analysis (CBA) ซึ่งประกอบไปด้วยการสร้างกฎความสัมพันธ์บนพื้นฐานของอัลกอริทึม Apriori และการสร้างตัวจำแนกจากกฎ ผลการทดสอบระบบ พบว่าความสามารถในการจำแนกความคิดเห็นทางการเมืองได้ถูกต้องสูงสุดถึง 77.75%

สุทธิวรรณ จรรย์พัฒนานุกูล และมาลีรัตน์ โสตานิล [29] ระบบวิเคราะห์ข้อความจากเว็บไซต์ร้านอาหารโดยใช้เหมืองข้อมูลความเห็น นำเสนอการพัฒนาระบบวิเคราะห์ข้อความแสดงความเห็น เพื่อให้ลูกค้าสามารถเลือกใช้บริการร้านอาหารได้ตรงกับความต้องการ โดยวิเคราะห์ข้อความแยกตามทัศนคติต่อคุณลักษณะของร้านอาหาร 5 ด้าน ได้แก่ การบริการ รสชาติ ราคา สิ่งอำนวยความสะดวก และบรรยากาศ จากนั้นนำข้อความแสดงความเห็นมาจำแนกเป็นเชิงบวก และเชิงลบ ซึ่งผลการประเมินประสิทธิภาพของระบบสามารถใช้งานได้เป็นอย่างดี เนื่องจากการทดสอบค่าความแม่นยำ (Precision) ร้อยละ 83.11 ค่าความระลึก (Recall) ร้อยละ 89.87 และค่า F-Measure ร้อยละ 86.36

ทศสิทธิ์ ไพบุลย์นันทพงศ์ และมณเฑียร รัตนศิริวงศ์วุฒิ [30] การจัดหมวดหมู่ข้อมูลจากการทำเหมืองข้อมูลของทวีตเตอร์ กรณีศึกษา #tutorial นำเสนอช่องทางการเก็บข้อมูล และช่องการสืบค้นข้อมูลเกี่ยวกับสื่อการเรียนการสอน (tutorial) ในทวีตเตอร์ โดยการจัดหมวดหมู่ข้อมูลเพื่อสะดวกในการค้นหาและง่ายต่อการใช้งาน ผู้วิจัยพัฒนาระบบโดยการเก็บข้อมูลในทวีตเตอร์เป็น APIs จากนั้นจำแนกข้อมูลเป็น 3 ส่วน ได้แก่ เนื้อหา, URL, Hashtag เพื่อจัดหมวดหมู่ข้อมูล และจัดทำเว็บแอปพลิเคชันภาษา PHP ติดตั้งโปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์ร่วมกับโปรแกรมจัดการฐานข้อมูล MySQL จากนั้นจึงทดสอบระบบโดยผลการประเมินจากผู้เชี่ยวชาญค่าเฉลี่ย 3.65 ส่วนเบี่ยงเบนมาตรฐาน 0.85 สำหรับผลการประเมินจากผู้ใช้งานทั่วไป ค่าเฉลี่ย 3.74 ส่วนเบี่ยงเบนมาตรฐาน 0.82

อติเทพ ไชยสาร และรัฐสิทธิ์ สุขะหุต [31] การประมาณอารมณ์จากความคิดเห็นภาษาไทย โดยใช้เทคนิคการเรียนรู้ของเครื่อง นำเสนอการเปรียบเทียบการประมาณอารมณ์จากความคิดเห็นจาก 3 เทคนิค ได้แก่ นาอึฟเบย์ ซัพพอร์ตเวกเตอร์แมชชีน และต้นไม้ตัดสินใจ โดยการรวบรวมความคิดเห็นจากเว็บไซต์บริการข่าว บริการวาไรตี้บันเทิง และการวิจารณ์สินค้า จำนวน 6,000 ความคิดเห็น จำแนกความคิดเห็นตามกลุ่มอารมณ์ 6 อารมณ์ ได้แก่กลุ่ม Love, Joy, Surprise, Sadness, Fear, และ Anger กลุ่มละ 1,000 ความคิดเห็น การทดสอบการความถูกต้องในการประมาณ

อารมณ์ใช้วิธี k-fold cross validation โดยใช้ค่า $k=10$ ผลการประมาณอารมณ์โดยรวมพบว่า ขั้นตอนวิธีชัพพอร์ตเวกเตอร์แมชชีน สามารถประมาณอารมณ์ได้ถูกต้อง 69.15% ต้นไม้ตัดสินใจได้เท่ากับ 59.93% และ นาอ์ฟเบย์ได้เท่ากับ 57.23%

รวิสุตา เทศเมือง และนิเวศ จิระวิชิตชัย [32] การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธีชัพพอร์ตเวกเตอร์แมชชีน นำเสนอการเปรียบเทียบการสร้างแบบจำลองด้วยอัลกอริทึม 4 วิธี ได้แก่ ชัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ นาอ์ฟเบย์ และ เคเนียร์สเนเบอร์ โดยนำข้อมูลความคิดเห็นการรีวิวสินค้าออนไลน์ ด้านการบริการห้องพักโรงแรม รีสอร์ท จาก Agoda Thailand และ Twitter Thailand ที่จดทะเบียนหลักทรัพย์ จำนวน 2,890 ข้อความ ผู้วิจัยใช้โปรแกรม Import.io เพื่อดึงข้อมูล และใช้โปรแกรม LexTo ในการคัดกรองข้อมูล โดยกำจัดข้อมูลที่ซ้ำกัน จากนั้นใช้โปรแกรม Rapidminer 3 มาใช้ในการสกัดคุณลักษณะของข้อความความคิดเห็นภาษาไทย ซึ่งผลการสร้างแบบจำลองด้วยอัลกอริทึมชัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพในการจำแนกดีที่สุด

งานวิจัยเกี่ยวกับการทำเหมืองข้อความ มีการนำเสนอการทำเหมืองข้อความที่สามารถนำไปใช้ในการวิเคราะห์องค์ประกอบของความคิดเห็นจากการวิเคราะห์คำศัพท์ การจัดหมวดหมู่คำศัพท์จากความคิดเห็น รวมถึงขั้นตอนในการทำเหมืองข้อความข้อความ ซึ่งสามารถนำวิธีการที่ผู้วิจัยใช้นั้นไปใช้ในงานวิจัยเพื่อพัฒนาระบบการพัฒนาอาจารย์ต่อไปได้

จากงานวิจัยทั้งหมดที่ได้ศึกษานั้นพบว่า การวิเคราะห์ความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ จำเป็นต้องใช้คุณลักษณะของอาจารย์ที่ดีและปริมาณความคิดเห็นที่มากพอ จากนั้นจึงสามารถนำมากำหนดการเลือกใช้อัลกอริทึมที่เหมาะสมกับข้อมูล โดยคำนึงถึงวัตถุประสงค์ที่ต้องการและลักษณะของข้อมูลเป็นหลัก งานวิจัยนี้มีการเลือกใช้ข้อมูลจากระบบบริการการศึกษาของสถาบันการศึกษาแห่งหนึ่ง ที่มีจำนวนขนาดข้อมูลเพียงพอ และเป็นข้อมูลที่ไม่ล้ำสมัย

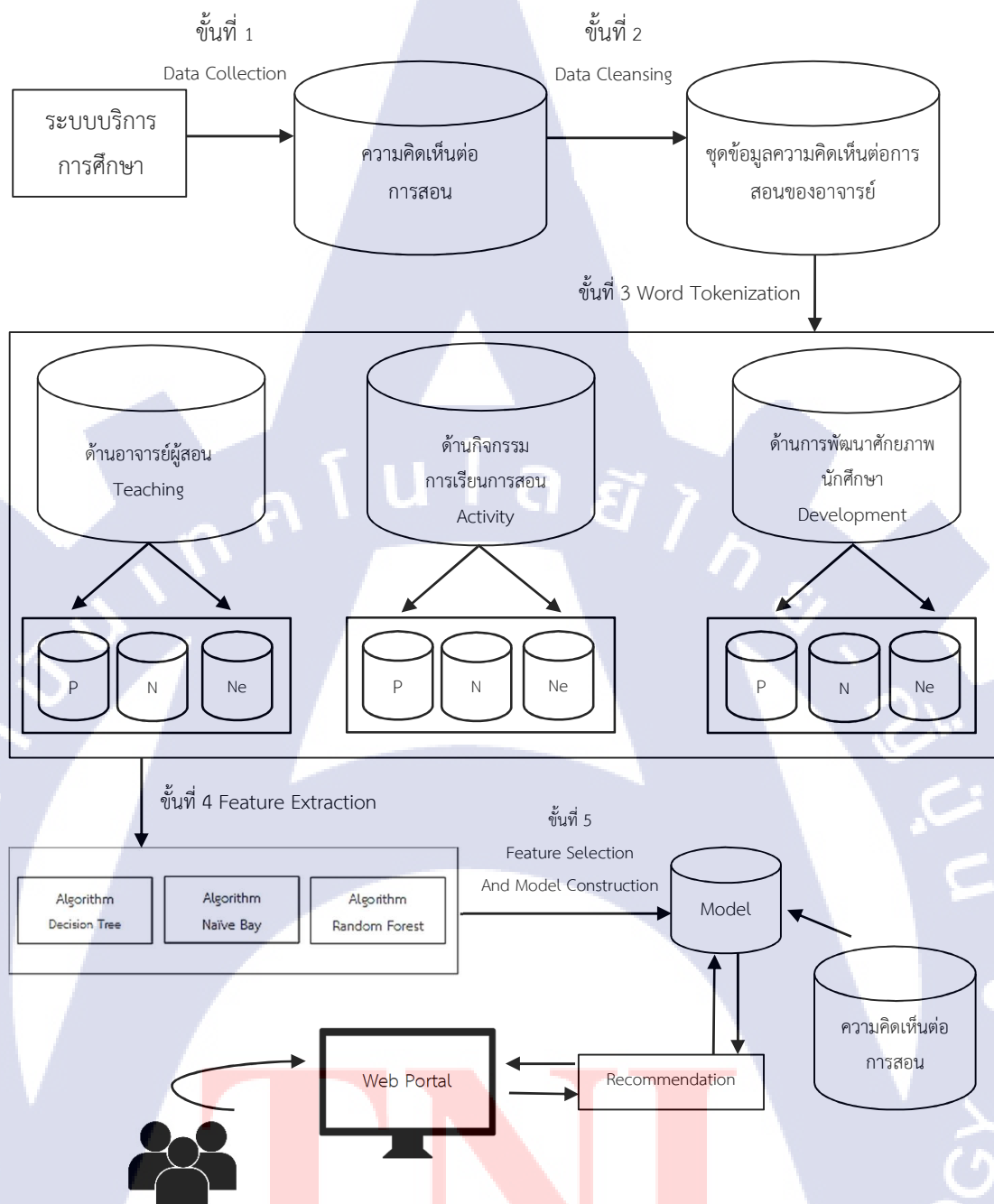
บทที่ 3

วิธีการดำเนินงาน

งานวิจัยเรื่อง การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์
โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ มีขั้นตอนการวิจัยดังนี้

ตารางที่ 3.1 ขั้นตอนการวิจัย

ขั้นตอนหลัก	ขั้นตอนย่อย
ขั้นตอนการนิยามปัญหา	ขั้นที่ 1 การวิเคราะห์ข้อมูลปัญหาของงานวิจัย ขั้นที่ 2 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความคิดเห็น และเทคนิคเหมืองข้อความ
ขั้นตอนการดำเนินงาน	ขั้นที่ 3 ประเมินวิธีการระบุปัจจัยที่ส่งผลต่อการสอนของอาจารย์ ขั้นที่ 4 สร้างคลังคำศัพท์ของปัจจัยที่ส่งผลต่อการสอนของอาจารย์ ขั้นที่ 5 วิเคราะห์และออกแบบโครงสร้างของเครื่องมือที่ใช้ในการจำแนกความคิดเห็น ขั้นที่ 6 การพัฒนาตัวต้นแบบการประเมินการสอนเพื่อพัฒนาการสอนของอาจารย์
ขั้นตอนการสรุปผล	ขั้นที่ 7 จัดทำและนำเสนอบทความวิชาการสำหรับงานวิทยานิพนธ์ ขั้นที่ 8 สรุปผลและจัดทำเอกสารวิทยานิพนธ์



รูปที่ 3.1 กรอบแนวความคิด

ตารางที่ 3.2 แผนการดำเนินการวิจัย (ต่อ)

การดำเนินงาน	ระยะเวลา										
	ปี 2561					ปี 2562					
	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.
- เขียนวิทยานิพนธ์เล่มสมบูรณ์											
- สอบป้องกันวิทยานิพนธ์											

โดยขั้นตอนการดำเนินงานวิจัยในครั้งนี้มีทั้งหมด 5 ขั้นตอน และมีวิธีการแบ่งเป็นรายละเอียดในแต่ละขั้นตอนตารางที่ 3.2 ดังนี้

3.1 ขั้นตอนที่ 1 การวิเคราะห์ปัญหาของงานวิจัย

ผลการประเมินการสอนของอาจารย์ประกอบด้วยข้อมูล 2 ส่วน ได้แก่ ส่วนที่ 1 แบบสอบถามแบบนักศึกษาสามารถเลือกตอบได้ระดับความคิดเห็น 1 ถึง 5 และส่วนที่ 2 คือ ส่วนที่เป็นการแสดงความคิดเห็นแบบบรรยายเป็นตัวหนังสือ ซึ่งในส่วนนี้ผู้วิจัยเห็นว่ายังไม่มีมีการนำความคิดเห็นส่วนนี้ไปใช้ในการปรับปรุงการสอนเท่าที่ควร เนื่องจากอาจารย์บางท่านไม่ให้ความสำคัญกับส่วนนี้และความคิดเห็นมีเป็นจำนวนมากยากแก่การอ่านความคิดเห็นทั้งหมด ผู้วิจัยจึงต้องการนำข้อมูลความคิดเห็นส่วนนี้มาใช้ในการวิเคราะห์การสอนของอาจารย์ให้มีความชัดเจนมากขึ้น

NO	ความคิดเห็น
1	อยากที่แนะนำคืออาจารย์ >< ชอบคุณอาจารย์มากค่ะก็อยากจะทำบทความรู้ให้มากขึ้น ก็มาจัดพิมพ์ออกจะสนุกก็เลย 55555(ตั้งใจในการศึกษาหา? นิดนึง) ให้อาจารย์
2	อยากที่แนะนำคือ
3	อาจารย์สอนดี เข้าใจง่าย ทำให้วิชาที่ยากกลายเป็นวิชาที่สนุก
4	อาจารย์สอนเข้าใจและให้คำแนะนำกับนักศึกษาทุกคน อาจารย์น่ารักมาก ป.ล. แนะนำให้จัดอบรมมากขึ้น
5	อยากที่แนะนำคืออยากให้ อาจารย์ช่วยอาจารย์
6	แนะนำคือ
7	ผมขอแนะนำคืออาจารย์สอนดีมากแต่ไม่ค่อยเข้าใจอะไรจะไรในโลกนี้ อาจารย์ทำให้ผมเข้าใจเป็นอีดีอาจารย์ฝึกคิดในการสอนได้เข้าใจ จากน้องสงครางค์รับ
8	อ.ทำให้หนูรักวิชานี้และอยากไปศึกษาต่อในทีม หนูอยากมาทำโปรเจกต์กับอาจารย์ได้
9	อาจารย์ น่ารัก "
10	"อาจารย์สอนเข้าใจง่ายคือ อาจารย์มีเมตตาและไม่ลืมถามเราว่าไม่มีความสับสนกับวิชาเรียน
11	อาจารย์สอนดีมากแต่เราไม่เข้าใจอาจารย์จะหาอะไรจะเข้าใจ
12	อาจารย์สอนดีมาก และเข้าใจในเนื้อหาเป็นอย่างดี
13	อาจารย์สอนเข้าใจ น่ารัก
14	อาจารย์ สอนเป็นกันเอง สอนเข้าใจง่าย น่ารัก และใส่ใจ นักศึกษามาก
15	ชอบคุณค่ะ ชอบคุณและ ชอบคุณมากค่ะ
16	อาจารย์สอนเข้าใจง่าย เป็นที่สนใจ เนื้อหาและระยะเวลาในการสอนดี
17	อาจารย์น่ารัก สอนเข้าใจง่าย
18	อาจารย์สอนเป็นกันเองดีมากได้ความรู้และประสบการณ์จากอาจารย์ช่วยได้เยอะอีกด้วย
19	เห็นความตั้งใจและความทุ่มเทของอาจารย์และสอคือตั้งใจกับทุกวิชาเรียน
20	อาจารย์สอนดีมากและเข้าใจง่าย
21	อาจารย์ดีมาก อาจารย์สอนไปอย่างช้าๆ หนูทำให้ได้ศึกษาทุกเข้าใจในเนื้อหา ส่วนกับนักศึกษาคนอื่นเค้ายังไม่เข้าใจและยังทำไม่ได้ อาจารย์ก็จะมาสอนส่วนตัวสอนเข้าใจ
22	อาจารย์มีเมตตาและสอนทำให้เข้าใจมากขึ้น อาจารย์เป็นคนที่น่ารักทำให้หนูอยากมาเรียนกับอาจารย์
23	อาจารย์น่ารักมากค่ะ สอนเป็นกันเอง

รูปที่ 3.2 ความคิดเห็นต่อการสอนของอาจารย์โดยนักศึกษาจากระบบบริหารการศึกษ

3.2 ขั้นที่ 2 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความคิดเห็น และเทคนิคเหมืองข้อความ

จากการศึกษาความคิดเห็นความคิดเห็นต่อการสอนของอาจารย์โดยนักศึกษาจากระบบบริหารการศึกษา ซึ่งประกอบด้วย 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างลักษณะที่พึงประสงค์

ผลการประเมินการสอนของอาจารย์โดยนักศึกษา

ปีการศึกษา 2561 ▼

ภาคการศึกษาที่	รหัสวิชา	ชื่อวิชา	ประเภท	Mean	SD	แสดงรายละเอียด
1/2561	BA60203	การบัญชีบริหารเพื่อการจัดการธุรกิจ	ด้านอาจารย์ผู้สอน	4.42	0.69	แสดงรายละเอียด
	BA60203	การบัญชีบริหารเพื่อการจัดการธุรกิจ	ด้านกิจกรรมการเรียนการสอน	4.40	0.70	แสดงรายละเอียด
	BA60203	การบัญชีบริหารเพื่อการจัดการธุรกิจ	ด้านการพัฒนาศักยภาพ นศ. และเสริมสร้างคุณลักษณะที่พึงประสงค์	4.44	0.66	แสดงรายละเอียด
	BA60203			4.41		
	BUS1122	การบัญชีเพื่อการจัดการธุรกิจ	ด้านอาจารย์ผู้สอน	4.00	0.00	แสดงรายละเอียด
	BUS1122	การบัญชีเพื่อการจัดการธุรกิจ	ด้านกิจกรรมการเรียนการสอน	4.00	0.00	แสดงรายละเอียด
	BUS1122	การบัญชีเพื่อการจัดการธุรกิจ	ด้านการพัฒนาศักยภาพ นศ. และเสริมสร้างคุณลักษณะที่พึงประสงค์	4.00	0.00	แสดงรายละเอียด
ผลประเมินโดยรวม (ทุกภาคการศึกษา)				4.41	0.22	-

รูปที่ 3.3 ประเภทการประเมินการสอนของอาจารย์

ความคิดเห็นต่อการสอนของอาจารย์โดยนักศึกษาจากระบบบริหารการศึกษาประกอบด้วย

1. ข้อเท็จจริง หมายถึง ประโยคแสดงสิ่งที่เกิดขึ้นจริง สามารถพิสูจน์ได้
2. ความคิดเห็น หมายถึง ความรู้สึกต่อสิ่งใดสิ่งหนึ่ง ซึ่งมีทั้งข้อแสดงความคิดเห็นที่ชัดเจน คือ ความคิดเห็นเชิงบวก (Positive) เชิงลบ (Negative) สำหรับความรู้สึกที่ไม่ชัดเจนจะแสดงเป็นแบบกลาง (Neutral)
3. ข้อเสนอแนะ หมายถึง การแสดงความคิดเห็นต่อหัวข้อใดหัวข้อหนึ่ง ในกรณีที่ไม่เห็นด้วยกับความคิดเห็นหรือการกระทำของผู้อื่นและนำเสนอความคิดเห็นใหม่ของตนเองเพื่อเป็นแนวทางให้ปฏิบัติ หรือการชี้ให้เห็นข้อบกพร่องพร้อมทั้งเสนอแนวทางแก้ไข โดยข้อเสนอแนะสามารถแบ่งได้ 3 ประเภทย่อยๆ ได้แก่

3.1 ข้อเสนอแนะทางตรง (Explicit suggestion) หมายถึง ข้อเสนอแนะที่มีคำบ่งชี้ข้อเสนอแนะที่ชัดเจน ในการแสดงความคิดเห็นในกรณีที่ไม่เห็นด้วยกับความคิดเห็นหรือการกระทำ

ของผู้อื่น และนำเสนอความคิดเห็นใหม่ของตนเองเพื่อเป็นแนวทางให้ปฏิบัติรูปแบบประโยค ประกอบด้วยคำบ่งชี้ข้อเสนอแนะที่ชัดเจน และมีเจตนาให้ปฏิบัติตามข้อเสนอแนะดังกล่าว

3.2 ข้อเสนอแนะเชิงคำถาม หมายถึง ข้อเสนอแนะที่มีรูปแบบประโยคที่ประกอบด้วย คำบ่งชี้ข้อเสนอแนะในเชิงคำถามอยู่ในประโยค และมีเจตนาการตั้งคำถามถึงสิ่งที่ต้องการให้ปฏิบัติ

3.3 ข้อเสนอแนะเชิงเงื่อนไข หมายถึง ข้อเสนอแนะที่มีรูปแบบประโยคที่ประกอบด้วย คำบ่งชี้ข้อเสนอแนะในเชิงเงื่อนไขอยู่ในประโยค และมีเจตนาในการแสดงข้อเสนอแนะแบบมีเงื่อนไข ในการปฏิบัติ

3.3 ขั้นที่ 3 จัดเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์ปัจจัยที่ส่งผลต่อการสอนของอาจารย์

การทำความสะอาดข้อมูล (Data Cleansing) เพื่อเตรียมข้อมูลไปใช้ในการตัดคำนั้น จำเป็นต้องนำสิ่งที่โปรแกรมไม่สามารถเข้าใจได้ออกจากข้อความนั้นก่อน เนื่องจากจะส่งผลให้การเขียนคำสั่งผิดพลาดไม่สามารถประมวลผลได้ การจัดการแก้ไขข้อความที่ไม่สมบูรณ์หรือผิดพลาดให้ถูกต้อง เช่น สัญลักษณ์ การเว้นวรรค(Space bar) การขึ้นบรรทัดใหม่ คำผิด การขึ้นบรรทัดใหม่ (Enter)โดยใช้ Microsoft Excel จากนั้นดำเนินการวิเคราะห์ความคิดเห็นโดยการจำแนกด้านการ ประเมิน และความคิดเห็นเชิงบวก ลบ กลาง จากนั้นนำข้อมูลทำการตัดคำโดยใช้โปรแกรม LexTo เพื่อเตรียมนำเข้าสู่การวิเคราะห์ตามขั้นตอนการทำเหมืองข้อความต่อไป

ตารางที่ 3.3 ตัวอย่างการทำความสะอาดข้อมูล

ข้อความก่อนแก้ไข	ข้อความที่แก้ไขแล้ว
อาจารย์สอนละเอียดและเข้าใจ	อาจารย์สอนละเอียดและเข้าใจ
อยากให้พูดให้ดังกว่านี้ค่ะ ฟังไม่ค่อยชัด ☹️	อยากให้พูดให้ดังกว่านี้ค่ะ ฟังไม่ค่อยชัด
สร้าง "ความมั่นใจ" เริ่มสอนแรกให้ใจทย่ง่ายๆ ให้ ทำให้ถูก ให้ภูมิใจ	สร้างความมั่นใจเริ่มสอนแรกให้ใจทย่ง่ายๆ ให้ ทำให้ถูกให้ภูมิใจ

ตารางที่ 3.4 ข้อมูลความคิดเห็นเพื่อเตรียมนำเข้าสู่การวิเคราะห์

ความคิดเห็น	ความคิดเห็น (ตัดคำ)	ด้านการ ประเมิน	ด้านความ คิดเห็น
สอนได้เข้าใจสนุกสนานไม่ เครียด มีประสบการณ์ในการ สอนมากๆ	สอนได้ เข้าใจ สนุกสนาน ไม่ เครียด มีประสบการณ์ ในการ สอน มากๆ	Teaching	Positive

ตารางที่ 3.4 ข้อมูลความคิดเห็นเพื่อเตรียมนำเข้าสู่วิเคราะห์ (ต่อ)

ความคิดเห็น	ความคิดเห็น (ตัดคำ)	ด้านการประเมิน	ด้านความคิดเห็น
เข้าใจนักเรียนสอนให้รู้stepใจ เย็นชี้แจงว่าเรื่องที่สอนจะมี ประโยชน์ยังง	เข้าใจ นักเรียน สอน ให้ รู้ step ใจ เย็น ชี้ แจง ว่า เรื่อง ที่ จะ สอน จะ มี ประโยชน์ ยังง	Activity	Positive
ทำให้กล้าแสดงออกกล้าคิด กล้าทำการพูดทำไปใช้ในชีวิต	ทำ ให้ กล้า แสดง ออก กล้า คิด กล้า ทำ การ พูด ทำ ไป ใช้ ใน ชีวิ ต	Development	Positive
บางครั้งอาจารย์ก็สอนเร็วแต่ โดยรวม สอนเข้าใจนะคะแต่ ยังงก็อยากให้ปรับความเร็ว หน่อยคะช้าลงมามีเรียนไม่ ทัน	บาง ครั้ง อา จ การ ย ก็ สอน เร็ว แต่ โดย รวม สอน เข้าใจ นะ คะ แต่ ยังง ก็ อยาก ให้ ปรับ ความ เร็ว หน อย ช้า ลง มา อีก เรียน ไม่ ทัน อา จ การ ย	Activity	Neutral

3.4 ขั้นที่ 4 วิเคราะห์และออกแบบโครงสร้างของเครื่องมือที่ใช้ในการจำแนกความคิดเห็น

วิธีการหาแบบจำลองการจำแนกข้อความที่เหมาะสมที่สุด สามารถทำได้โดยเปรียบเทียบประสิทธิภาพการจำแนกข้อความของ 3 อัลกอริทึม ได้แก่ อัลกอริทึมต้นไม้ตัดสินใจ นาอิวเบย์ และ แรนดอมฟอร์เรส จากนั้นเปรียบเทียบประสิทธิภาพการสกัดข้อเสนอแนะของชุดข้อมูลเรียนรู้ของแต่ละอัลกอริทึม ด้วยค่าความแม่นยำ ค่าความระลึก และค่าถ่วงดุล โดยการทดสอบประสิทธิภาพการจำแนกข้อความจะใช้โปรแกรม RapidMiner จากนั้นหาค่าเฉลี่ยการวัดประสิทธิภาพของแต่ละเทคนิควิธี นำแบบจำลองของแต่ละเทคนิควิธีที่ให้ประสิทธิภาพดีที่สุดมาพัฒนา โดยใช้การวัดประสิทธิภาพโดยการหาค่าความเที่ยง (Precision) ค่าความลึก (Recall) ค่าความถูกต้อง (Accuracy) และค่า F-Measure ดังสมการต่อไปนี้

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

โดยที่ TP คือ จำนวนข้อมูลที่ถูกดึงออกมาอย่างถูกต้อง
 FP คือ จำนวนข้อมูลที่ผิดพลาดที่ถูกดึงออกมา
 TN คือ จำนวนข้อมูลที่ถูกต้องแต่ไม่ถูกดึงออกมา
 FN คือ จำนวนข้อมูลที่ผิดพลาดแต่ไม่ถูกดึงออกมา

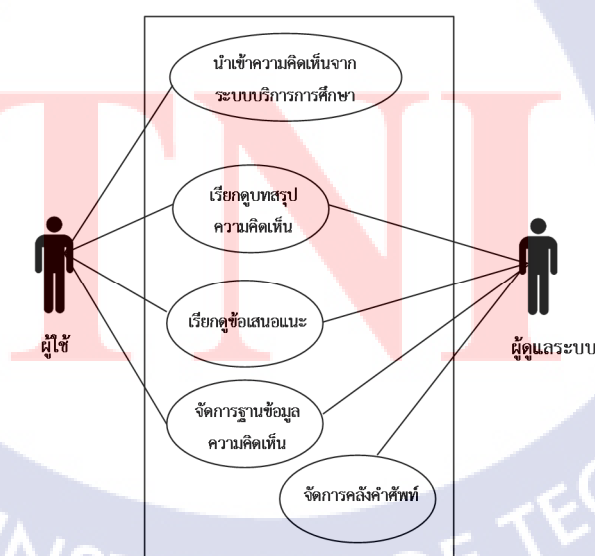
3.5 ขั้นที่ 5 สร้างคลังคำศัพท์ของปัจจัยที่ส่งผลต่อความคิดเห็นการสอนของอาจารย์

สำหรับการสร้างคลังคำศัพท์ของปัจจัยที่ส่งผลต่อความคิดเห็นการสอนของอาจารย์ที่ประกอบด้วย 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างลักษณะที่พึงประสงค์ และประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง สามารถแบ่งได้เป็น 3 ขั้นตอน ดังนี้

1. การสร้างคลังคำศัพท์ของปัจจัยที่ส่งผลต่อความคิดเห็นการสอนของอาจารย์
2. การสร้างคลังคำศัพท์สำหรับคำที่บ่งบอกประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง
3. การสร้างคลังคำศัพท์เกี่ยวกับข้อเสนอแนะ

3.6 ขั้นที่ 6 พัฒนาตัวต้นแบบการประเมินการสอนเพื่อพัฒนาการสอนของอาจารย์

การพัฒนาตัวต้นแบบการประเมินการสอนเพื่อพัฒนาการสอนของอาจารย์ประกอบด้วยส่วนของผู้ใช้งานและผู้ดูแลระบบ



รูปที่ 3.4 แผนภาพยูสเคสของตัวต้นแบบการประเมินการสอนเพื่อพัฒนาการสอน

บทที่ 4

ผลการวิจัย

การศึกษาวิจัยเรื่องการวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยใช้เทคนิคเหมืองข้อความ มีวัตถุประสงค์เพื่อวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยใช้เทคนิคเหมืองข้อความ เพื่อสังเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา และ เพื่อพัฒนาเครื่องมือที่สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา เพื่อระบุปัจจัยที่ส่งผลต่อคุณภาพการสอนของอาจารย์ รวมทั้งความคิดเห็นเชิงบวกและเชิงลบ จากนั้นได้ข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์ สามารถแบ่งผลการดำเนินการและการวิเคราะห์ข้อมูล ดังต่อไปนี้

4.1 การประเมินประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน

4.2 การประเมินประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก
เชิงลบ และเชิงกลาง

4.3 การกำหนดเงื่อนไขตามด้านการประเมินเพื่อการจัดทำตัวต้นแบบ

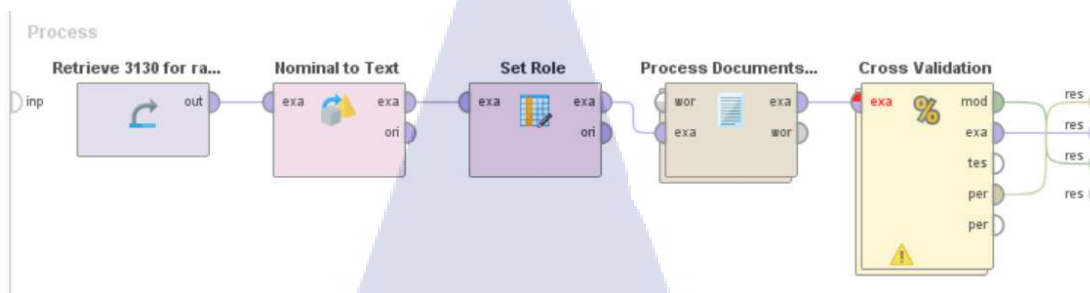
4.4 การกำหนดเงื่อนไขตามประเภทความคิดเห็นเพื่อการจัดทำตัวต้นแบบ

ตารางที่ 4.1 สรุปข้อมูลที่ใช้วิเคราะห์

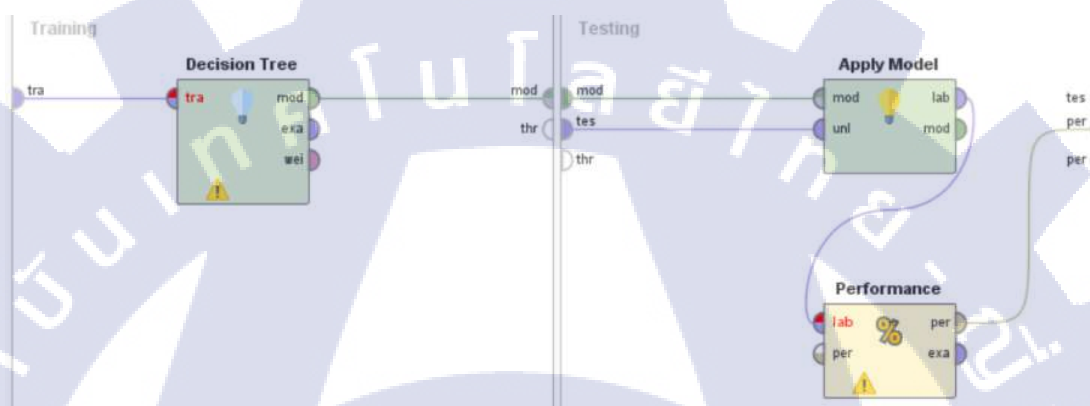
ด้านการประเมิน	จำนวนข้อความ	ประเภทความคิดเห็น	จำนวนข้อความ
Teaching	1,040	Positive	1,737
Activity	1,050	Negative	1,322
Development	1,040	Neutral	71
	3,130		3,130

4.1 การประเมินประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน

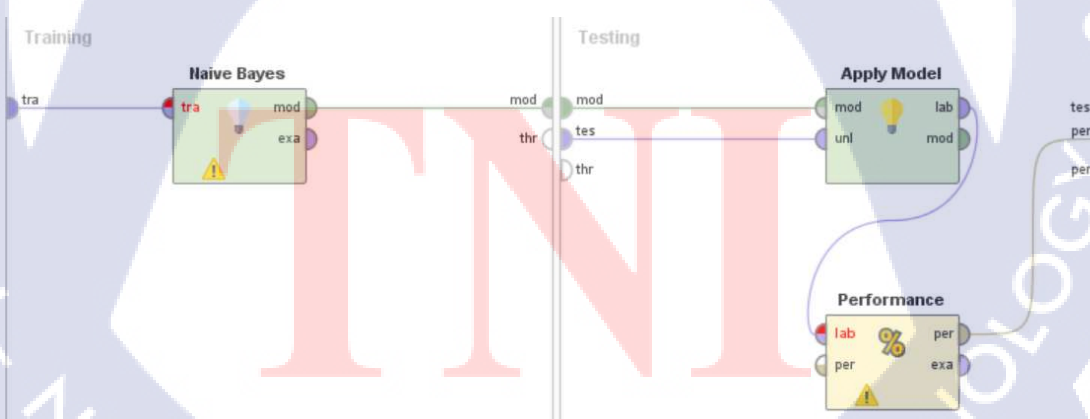
ผู้ทำวิจัยได้ใช้โปรแกรม Rapid miner ในการประเมินประสิทธิภาพการจำแนกข้อความตามด้านการประเมินจากการใช้อัลกอริทึมสำหรับการจำแนก 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ, นาอ็ิบเบย์ และแรนดอมฟอร์เรส ในขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมินจะมีขั้นตอนเหมือนกันโดยเริ่มจากการนำข้อมูลที่เตรียมไว้จำนวน 3,130 ข้อความเข้าสู่ Rapidminer จากนั้นทำการประเมินประสิทธิภาพการจำแนกข้อมูลในส่วนของ Cross Validation ดังนี้



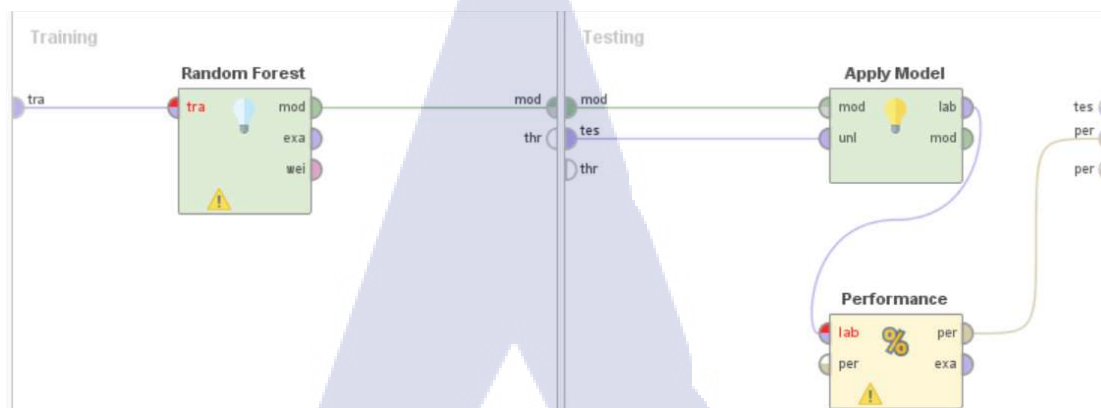
รูปที่ 4.1 ขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน



รูปที่ 4.2 การใช้อัลกอริทึม Decision Tree เพื่อวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน



รูปที่ 4.3 การใช้อัลกอริทึม Naïve Bay เพื่อวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน



รูปที่ 4.4 การใช้อัลกอริทึม Random Forest เพื่อวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน

ตารางที่ 4.2 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม Decision Tree

Decision Tree	true Teaching	true Activity	true Development	class precision	F-Measure
pred. Teaching	544	135	66	73.02%	0.6095
pred. Activity	426	892	112	62.38%	0.7196
pred. Development	70	23	862	90.26%	0.8641
class recall	52.31%	35.51%	82.88%		
accuracy: 75.02%					

จากตารางที่ 4.2 ผลการวัดประสิทธิภาพการจำแนกข้อความโดยใช้อัลกอริทึม Decision Tree ตามด้านการประเมิน พบว่า ค่าความแม่นยำในการทำนาย (Precision) และค่าความระลึก (Recall) ด้าน Development มีค่าสูงสุด 90.26% และ 82.88% ตามลำดับ สำหรับค่าความแม่นยำในการทำนาย (Precision) และค่าความระลึก (Recall) ด้าน Activity มีค่าต่ำสุด 62.38% และ 35.51% ตามลำดับ สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวม ด้าน Development มีค่าสูงสุด 0.8641 ด้าน Teaching มีค่าต่ำสุด 0.6095 โดยค่าความถูกต้องสำหรับอัลกอริทึม Decision Tree ตามด้านการประเมินมีค่า 75.02%

ตารางที่ 4.3 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม

Naive Bayes

Naive Bayes	true Teaching	true Activity	true Development	class precision	F-Measure
pred. Teaching	365	49	0	88.16%	0.5021
pred. Activity	462	944	1	67.09%	0.7638
pred. Development	213	57	1,039	79.37%	0.8845
class recall	35.10%	89.90%	99.90%		
accuracy: 73.42%					

จากตารางที่ 4.3 ผลการวัดประสิทธิภาพการจำแนกข้อความโดยใช้อัลกอริทึม Naive Bayes ตามด้านการประเมิน พบว่า ค่าความแม่นยำในการทำนาย (Precision) ด้าน Teaching มีค่าสูงสุด 88.16% สำหรับค่าความระลึก (Recall) ด้าน Development มีค่าสูงสุด 99.90% สำหรับค่าความแม่นยำในการทำนาย (Precision) ที่ต่ำสุดด้าน Activity 67.09% และค่าความระลึก (Recall) ด้าน Teaching มีค่าต่ำสุด 35.10% สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวม ด้าน Development มีค่าสูงสุด 0.8845 ด้าน Teaching มีค่าต่ำสุด 0.5021 โดยค่าความถูกต้องสำหรับอัลกอริทึม Naive Bayes ตามด้านการประเมินมีค่า 73.42%

ตารางที่ 4.4 แสดงผลประสิทธิภาพการจำแนกข้อความตามด้านการประเมินโดยใช้อัลกอริทึม

Random Forest

Random Forest	true Teaching	true Activity	true Development	class precision	F-Measure
pred. Teaching	739	426	267	51.61%	0.5973
pred. Activity	270	608	36	66.52%	0.6191
pred. Development	31	16	737	94.01%	0.8081
class recall	71.06%	57.90%	70.87%		
accuracy: 66.58%					

จากตารางที่ 4.4 ผลการวัดประสิทธิภาพการจำแนกข้อความโดยใช้อัลกอริทึม Random Forest ตามด้านการประเมิน พบว่า ค่าความแม่นยำในการทำนาย (Precision) ด้าน Development

มีค่าสูงสุด 94.01% และค่าความระลึก (Recall) ด้าน Teaching มีค่าสูงสุด 71.06% สำหรับค่าความแม่นยำในการทำนาย (Precision) ต่ำสุดด้าน Teaching และค่าความระลึก (Recall) ในด้าน Activity มีค่าต่ำสุด 51.61% และค่าความระลึก (Recall) ด้าน Activity ต่ำสุดมีค่า 57.90% สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวม ด้าน Development มีค่าสูงสุด 0.8081 โดยค่าความถูกต้องสำหรับอัลกอริทึม Random Forest ตามด้านการประเมินมีค่า 66.58%

ตารางที่ 4.5 สรุปประสิทธิภาพการจำแนกข้อความตามด้านการประเมิน

Classification Algorithm	Accuracy	F-Measure		
		pred. Teaching	pred. Activity	pred. Development
Decision Tree	75.02%	0.6095	0.7193	0.8641
Naïve Bay	73.42%	0.5020	0.7683	0.8845
Random Forest	66.58%	0.5979	0.6191	0.8081

จากผลการทดสอบพบว่า Decision Tree มีค่าความถูกต้องในการจำแนกข้อมูลตามด้านการประเมินสูงสุด 75.30% และมีค่า F-Measure ด้าน Development และ Teaching สูงสุด ดังนั้นจึงเลือกอัลกอริทึมนี้ในการสร้างเงื่อนไขตามด้านการประเมินสำหรับการเขียน Web portal โดยตัวอย่างเงื่อนไขมีดังนี้

ตารางที่ 4.6 แสดงเงื่อนไขจากอัลกอริทึม Decision Tree

	รูปแบบเงื่อนไข	ด้านการประเมิน
1	อาจารย์ > งาน > ให้ > กลุ่ม	Development
2	อาจารย์ > งาน > เยอะ > สอน	Teaching
3	อาจารย์ > งาน > ให้	Activity
4	อาจารย์ > งาน > ประยุกต์	Teaching
5	อาจารย์ > สอน > ทำให้ > กิจกรรม	Activity
6	อาจารย์ > สอน > ทำให้ > กิจกรรม > ชีวิตประจำวัน	Teaching
7	อาจารย์ > สอน > ข้อมูล	Development
8	อาจารย์ > สอน > ข้อมูล > เล่น	Activity
9	อาจารย์ > วิชา > นำไปใช้	Teaching
10	อาจารย์ > วิชา > รู้ลึก	Activity
11	อาจารย์ > วิชา > รู้ลึก > คุณธรรม	Development
12	อาจารย์ > วิชา > รู้ลึก > เทคนิค	Teaching

ตารางที่ 4.6 แสดงเงื่อนไขจากอัลกอริทึม Decision Tree (ต่อ)

	รูปแบบเงื่อนไข	ด้านการประเมิน
13	อาจารย์ > สอน > เยอะ	Activity
14	อาจารย์ > สอน > เยอะ > ความกล้า	Development
15	อาจารย์ > สอน > การ > ส่ง > ทำให้	Activity
16	อาจารย์ > สอน > การ > ส่ง > กิจกรรม	Activity
17	อาจารย์ > สอน > การ > ส่ง > กิจกรรม > เนื้อหา	Activity
18	อาจารย์ > สอน > การ > ส่ง > กิจกรรม > เนื้อหา > ทำ	Teaching
19	ชี้แจง > บุรณาการ	Development
20	ชี้แจง > บุรณาการ > สื่อการเรียนการสอน	Activity
21	ชี้แจง > บุรณาการ > สื่อการเรียนการสอน > ฝึกงาน	Teaching

จากตารางที่ 4.6 รูปแบบเงื่อนไขที่เกิดขึ้นจากการใช้อัลกอริทึม Decision Tree นำไปใช้ในการพัฒนาตัวต้นแบบต่อไป

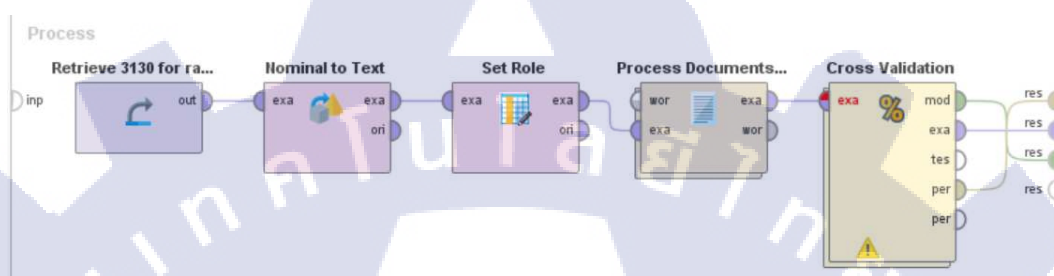
4.2 การประเมินประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และ เชิงกลาง

งานวิจัยนี้จำแนกข้อความตามประเภทความคิดเห็น 3 ลักษณะ ได้แก่ เชิงบวก (Positive) เชิงลบ (Negative) และเชิงกลาง (Neutral) การพิจารณาจำแนกความคิดเห็นจะแยกตามลักษณะของคำกริยา คำวิเศษณ์ คำสันธาน ที่บ่งบอกความรู้สึกในข้อความนั้น ข้อมูลที่เตรียมไว้เพื่อการวิเคราะห์ตามตารางที่ 4.7 ดังนี้

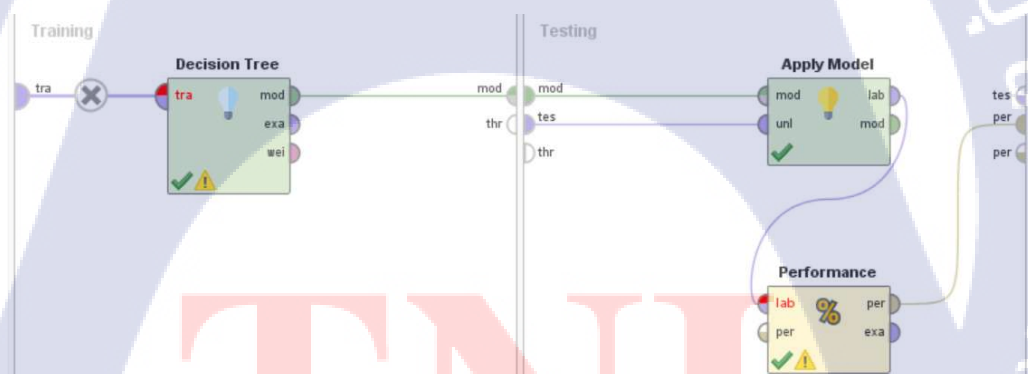
ตารางที่ 4.7 ข้อมูลเพื่อใช้ในการวิเคราะห์ประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง

ความคิดเห็น (ตัดคำ)	Positive/Neutral/Negative
อาจารย์ สอน เข้าใจ ใน เนื้อหา ดีมาก เอาใจใส่ นักศึกษา ทุก คน และ อาจารย์ น่ารัก มากๆ	Positive
พยาม เล่น มุข หรือ สร้าง อะไร ที่ เด็ก ไม่ เรียน แล้ว รู้ สึก เบื่อ หน่าย	Negative
บางครั้ง อาจารย์ ก็ สอน เร็ว แต่ โดยรวม สอน เข้าใจ นะ แต่ ยัง งง ก็ อยาก ให้ ปรับ ความเร็ว หน่อย ซ้ำ ลง มา อีก เรียน ไม่ ทน อาจารย์	Neutral

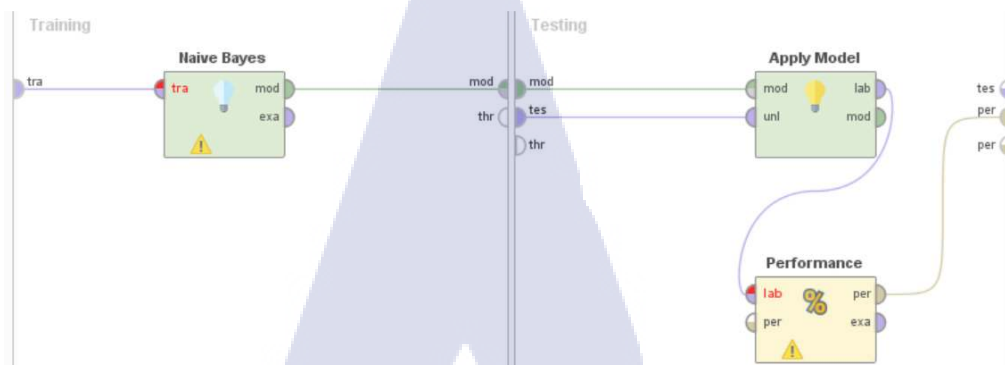
จากนั้นผู้ทำวิจัยได้ใช้โปรแกรม Rapid miner ในการประเมินประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางจากการใช้อัลกอริทึมสำหรับการจำแนก 3 อัลกอริทึม ได้แก่ ต้นไม้ตัดสินใจ, นาอ็ฟเบย์ และแรนดอมฟอร์เรส ในขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามด้านการประเมินจะมีขั้นตอนเหมือนกันโดยเริ่มจากการนำข้อมูลที่เตรียมไว้จำนวน 3,130 ข้อความเข้าสู่ Rapidminer จากนั้นทำการประเมินประสิทธิภาพการจำแนกข้อมูลในส่วนของ Cross Validation ดังนี้



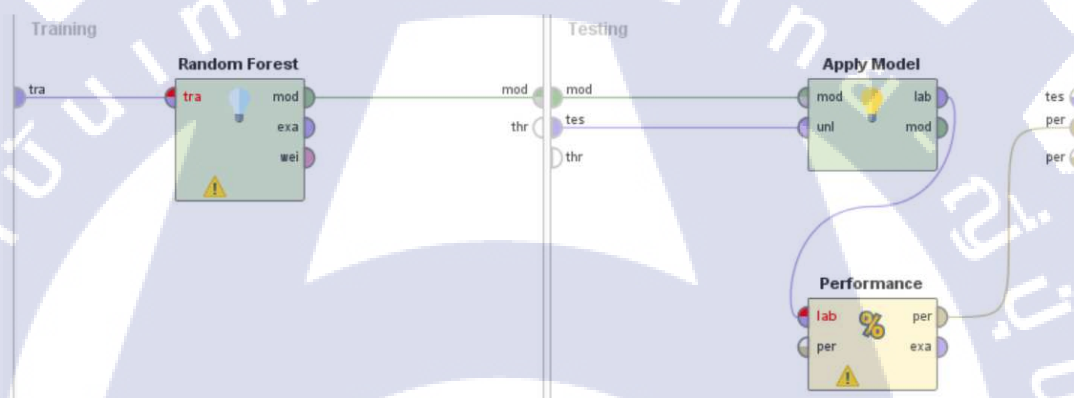
รูปที่ 4.5 ขั้นตอนการวัดประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง



รูปที่ 4.6 การใช้อัลกอริทึม Decision Tree เพื่อวัดประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง



รูปที่ 4.7 การใช้อัลกอริทึม Naïve bay เพื่อวัดประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง



รูปที่ 4.8 การใช้อัลกอริทึม Random Forest เพื่อวัดประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลาง

ตารางที่ 4.8 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางโดยใช้อัลกอริทึม Decision Tree

Decision Tree	true Positive	true Negative	true Neutral	class precision	F-Measure
pred. Positive	1,167	77	26	91.89%	0.7762
pred. Negative	382	1185	20	74.67%	0.8147
pred. Neutral	188	60	25	9.16%	0.1454
class recall	67.18%	89.64%	35.21%		
accuracy: 75.94%					

จากตารางที่ 4.8 ผลการวัดประสิทธิภาพการจำแนกความคิดเห็นตามประเภทความคิดเห็น โดยใช้อัลกอริทึม Decision Tree พบว่า ค่าความแม่นยำในการทำนาย (Precision) ความคิดเห็นเชิงบวก (Positive) มีค่าสูงสุด 91.89% สำหรับค่าความระลึก (Recall) ความคิดเห็นเชิงลบ (Negative) มีค่าสูงสุด 89.64% สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวมด้านการจำแนกความคิดเห็นตามประเภทความคิดเห็นเชิงลบมีค่าสูงสุด 0.8147 โดยค่าความถูกต้องสำหรับอัลกอริทึม Decision Tree มีค่า 75.94%

ตารางที่ 4.9 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และ เชิงกลางโดยใช้อัลกอริทึม Naive Bayes

Naive Bayes	true Positive	true Negative	true Neutral	class precision	F-Measure
pred. Positive	1,733	1,322	70	55.46%	0.7129
pred. Negative	0	0	0	0.00%	-
pred. Neutral	4	0	1	20.00%	0.0263
class recall	99.77%	0.00%	1.41%		
accuracy: 55.40%					

จากตารางที่ 4.9 ผลการวัดประสิทธิภาพการจำแนกความคิดเห็นตามประเภทความคิดเห็น โดยใช้อัลกอริทึม Naive Bayes พบว่า ค่าความแม่นยำในการทำนาย (Precision) ความคิดเห็นเชิงบวก (Positive) มีค่าสูงสุด 55.46% สำหรับค่าความระลึก (Recall) ความคิดเห็นเชิงบวก (Positive) มีค่าสูงสุด 99.77% แต่ไม่สามารถทำนายความคิดเห็นเชิงลบได้ สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวมด้านการจำแนกความคิดเห็นตามประเภทความคิดเห็นเชิงบวกมีค่าสูงสุด 0.7129 โดยค่าความถูกต้องสำหรับอัลกอริทึม Naive Bayes มีค่า 55.40%

ตารางที่ 4.10 แสดงผลประสิทธิภาพการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางโดยใช้อัลกอริทึม Random Forest

Random Forest	true Positive	true Negative	true Neutral	class precision	F-Measure
pred. Positive	1,737	1,153	71	58.66%	0.7394
pred. Negative	0	169	0	100.00%	0.2266
pred. Neutral	0	0	0	0.00%	-
class recall	100.00%	12.78%	0.00%		

จากตารางที่ 4.10 ผลการวัดประสิทธิภาพการจำแนกความคิดเห็นตามประเภทความคิดเห็นโดยใช้อัลกอริทึม Random Forest พบว่า ค่าความแม่นยำในการทำนาย (Precision) ความคิดเห็นเชิงลบ (Negative) มีค่าสูงสุด 100.00% สำหรับค่าความระลึก (Recall) ความคิดเห็นเชิงบวก (Positive) มีค่าสูงสุด 100.00% โดยไม่สามารถทำนายความคิดเห็นเชิงกลางได้เลย สำหรับค่า F-Measure แสดงประสิทธิภาพโดยรวมด้านการจำแนกความคิดเห็นตามประเภทความคิดเห็นเชิงบวก มีค่าสูงสุด 0.7394 โดยค่าความถูกต้องสำหรับอัลกอริทึม Random Forest มีค่า 60.89%

ตารางที่ 4.11 สรุปประสิทธิภาพการจำแนกความคิดเห็น

Classification Algorithm	Accuracy	F-Measure		
		pred. Positive	pred. Negative	pred. Neutral
Decision Tree	75.94%	0.7762	0.8147	0.1454
Naïve Bay	55.40%	0.7129	-	0.0263
Random Forest	60.89%	0.7394	0.2266	-

จากตารางที่ 4.11 แสดงสรุปประสิทธิภาพการจำแนกความคิดเห็น จากผลการทดสอบพบว่า Decision Tree มีค่าความถูกต้องในการจำแนกข้อความตามประเภทความคิดเห็นเชิงบวก เชิงลบ และเชิงกลางสูงสุด 75.94% และสามารถจำแนกได้ครบทุกประเภทความคิดเห็น ดังนั้นจึงเลือกอัลกอริทึมนี้ในการสร้างเงื่อนไขตามประเภทความคิดเห็นสำหรับการพัฒนาตัวต้นแบบ โดยตัวอย่างเงื่อนไขมีดังนี้

ตารางที่ 4.12 แสดงเงื่อนไขการจำแนกความคิดเห็นโดยใช้อัลกอริทึม Decision Tree

	รูปแบบเงื่อนไข	ด้านการประเมิน
1	แต่	Neutral
2	แต่ > ได้ความ	Neutral
3	แต่ > ดี แต่ > รืบ	Neutral
4	แต่ > ดี แต่ > รืบ > เป็นกันเอง	Positive

จากตารางที่ 4.12 รูปแบบเงื่อนไขที่เกิดขึ้นจากการใช้อัลกอริทึม Decision Tree นำไปใช้ในการพัฒนาตัวต้นแบบต่อไป

4.3 การกำหนดเงื่อนไขตามด้านการประเมินเพื่อการจัดทำตัวต้นแบบ

การกำหนดเงื่อนไขตามด้านการประเมินจะพิจารณาจากการให้ค่าน้ำหนักย่อยสำหรับด้านการประเมิน โดยใช้หลักการ (จำนวนคำที่ถูกกล่าวถึง * 100) / จำนวนความคิดเห็นทั้งหมด) การกำหนดค่าน้ำหนักนี้เพื่อเป็นตัวบ่งชี้ว่าแต่ละปัจจัยมีค่าน้ำหนักไปในด้านใด เนื่องจากปัจจัยย่อยที่ประกอบอยู่ในปัจจัยหลักแต่ละด้าน หากไม่ได้กำหนดค่าน้ำหนักของปัจจัยย่อยและให้ผลรวมค่าน้ำหนักในการสนับสนุนการตัดสินใจอาจทำให้การวิเคราะห์ที่ไม่มีความแน่นอน ดังนั้นจึงจำเป็นต้องกำหนดค่าน้ำหนักให้แต่ละปัจจัย ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ ดังตารางที่ 4.13

ตารางที่ 4.13 แสดงน้ำหนักสำหรับด้านการประเมิน

ด้านการประเมิน	คำ	TF-IDF
ด้านอาจารย์ผู้สอน	เป็นกันเอง	0.9726
	เข้าใจง่าย	0.9531
	อธิบาย	0.9229
	อภัยภัย	0.9214
	ใส่ใจ	0.9115
ด้านกิจกรรมการเรียนการสอน	สนุก	0.9290
	สื่อการสอน	0.8403
	เปิดโอกาส	0.7195
ด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้าง คุณลักษณะที่พึงประสงค์	ความคิดสร้างสรรค์	0.6495
	คุณธรรม	0.5425
	นำเสนอ	0.4071

4.4 การกำหนดเงื่อนไขตามประเภทความคิดเห็นเพื่อการจัดทำตัวต้นแบบ

การกำหนดเงื่อนไขตามประเภทความคิดเห็นจะพิจารณาจากการให้ค่าน้ำหนักย่อยสำหรับด้านการประเมิน โดยใช้หลักการ (จำนวนคำที่ถูกกล่าวถึง * 100) / จำนวนความคิดเห็นทั้งหมด) การกำหนดค่าน้ำหนักนี้เพื่อเป็นตัวบ่งชี้ว่าแต่ละปัจจัยมีค่าน้ำหนักไปในด้านใด เนื่องจากปัจจัยย่อยที่ประกอบอยู่ในปัจจัยหลักแต่ละด้าน หากไม่ได้กำหนดค่าน้ำหนักของปัจจัยย่อยและให้ผลรวมค่าน้ำหนักในการสนับสนุนการตัดสินใจอาจทำให้การวิเคราะห์ที่ไม่มีความแน่นอน ดังนั้นจึงจำเป็นต้องกำหนดค่าน้ำหนักให้แต่ละด้าน ได้แก่ เชิงบวก (Positive) เชิงลบ (Negative) เชิงกลาง (Neutral)

ตารางที่ 4.14 แสดงน้ำหนักสำหรับประเภทความคิดเห็น

ด้านการประเมิน	คำ	TF-IDF
เชิงบวก (Positive)	น่ารัก	1
	เข้าใจง่าย	1
	เป็นกันเอง	0.9726
	น่ารัก ใจดี สอน	0.9690
	ดี ให้คำปรึกษา	0.9223
เชิงลบ (Negative)	ไม่เข้าใจ	0.9593
	ไม่ตรงประเด็น	0.6074
	ไม่เหมาะสม	0.5250
เชิงกลาง (Neutral)	บางที่ อาจจะ	0.7904
	ดี แต่	0.7296
	แต่บางที่	0.7023

4.5 การพัฒนาตัวต้นแบบสนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา

ขั้นตอนสุดท้ายการพัฒนาตัวต้นแบบที่สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ ตัวต้นแบบนี้สามารถนำข้อความที่นักศึกษาประเมินเข้าสู่ตัวต้นแบบ จากนั้นจะมีการประมวลผลเพื่อระบุปัจจัยที่ส่งผลกระทบต่อคุณภาพการสอนของอาจารย์และระบุประเภทของความคิดเห็นตามโมเดลที่ได้มาจากการใช้อัลกอริทึม Decision Tree



รูปที่ 4.9 ตัวต้นแบบสนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษา

Feeling Sentiment : อ. สอนเข้าใจ อ. ใจเย็นไม่เข้าใจ อ. อธิบายจนเข้าใจ อ. ใส่ใจเอาใจใส่มีความเป็นห่วง อ. น่ารักเฮฮา ทำให้ ห้องเรียนมีความสนุกสนานไม่น่าเบื่อไม่เครียด ขอบพระคุณอาจารย์ค่ะ ขอบคุณสำหรับการสอนครับ	อ. สอนเข้าใจ อ. ใจเย็นไม่เข้าใจ อ. อธิบายจนเข้าใจ อ. ใส่ใจเอาใจใส่มีความเป็นห่วง อ. น่ารัก เฮฮา ทำให้ห้องเรียนมีความสนุกสนานไม่น่าเบื่อไม่เครียด ขอบพระคุณอาจารย์ค่ะ ขอบคุณ สำหรับการสอนครับ Feeling : ด้านอาจารย์ Sentiment : Positive ประมวลผลใน: 0.0742 วินาที ใจเย็นไม่เข้าใจ อ. อธิบายจนเข้าใจ อ. ใส่ใจเอาใจใส่มีความเป็นห่วง ทำให้ห้องเรียนมีความสนุกสนานไม่น่าเบื่อไม่เครียด ขอบพระคุณอาจารย์ค่ะ ขอบคุณสำหรับการสอนครับ
ตกลง	

รูปที่ 4.10 ตัวอย่างการใช้งานตัวต้นแบบ

ผลจากการวิเคราะห์ความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา สอดคล้องกับวัตถุประสงค์ของงานวิจัยในการวิเคราะห์และสังเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผล ประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ โดยการวิจัยนี้ทำให้ผู้วิจัย เข้าใจถึงกระบวนการในการทำเหมืองข้อความเพื่อให้ได้รูปแบบความคิดเห็นในด้านที่ผู้วิจัยกำหนด ได้แก่ ด้านอาจารย์ผู้สอน ด้านกิจกรรมการเรียนการสอน ด้านการพัฒนาศักยภาพนักศึกษาและ เสริมสร้างคุณลักษณะที่พึงประสงค์ รวมถึงด้านความคิดเห็น ได้แก่ ความคิดเห็นเชิงบวก เชิงลบ และ เชิงกลาง ซึ่งเมื่อผ่านกระบวนการวิจัยตามกรอบแนวคิดการวิจัยแล้ว มีการเลือกอัลกอริทึมที่มี ประสิทธิภาพที่สุดสำหรับขอบเขตการวิจัยนี้นำมาสร้างตัวต้นแบบสนับสนุนการวิเคราะห์ผลประเมิน การสอนของอาจารย์โดยนักศึกษา เพื่อระบุปัจจัยที่ส่งผลต่อคุณภาพการสอนของอาจารย์ ตาม วัตถุประสงค์ของงานวิจัยฉบับนี้ และจากนั้นได้ขอเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์ต่อไป

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

บทที่ 5

บทสรุป อภิปราย และข้อเสนอแนะ

การศึกษาวิจัยเรื่องการวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์ โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ ซึ่งการวิจัยนี้มีวัตถุประสงค์ตามบทที่ 1 ได้นำเสนอข้อสรุปตามลำดับดังนี้

- 5.1 สรุปผลการวิจัย
- 5.2 อภิปรายผลวิจัย
- 5.3 ปัญหาและอุปสรรคในการทำวิจัย
- 5.4 ข้อเสนอแนะงานวิจัย

5.1 สรุปผลการวิจัย

ในการสรุปผลของงานวิจัย ผู้วิจัยได้ทำการสรุปผลตามวัตถุประสงค์ในแต่ละข้อที่ได้กล่าวไว้ข้างต้นในบทที่ 1 ดังต่อไปนี้

การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ โดยมีการรวบรวมข้อมูลจากระบบบริหารการศึกษาของสถาบันการศึกษาแห่งหนึ่งจำนวน 3,130 ข้อความ จากนั้นนำความคิดเห็นมาตัดคำโดยใช้โปรแกรม LexTo และจำแนกด้านการประเมินเป็น 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน (Teaching) ด้านกิจกรรมการเรียนการสอน (Activity) และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ (Development) จากนั้นวิเคราะห์ความคิดเห็นเป็นเชิงบวก (Positive) เชิงลบ (Negative) และเชิงกลาง (Neutral) เมื่อเตรียมข้อมูลเรียบร้อยแล้วข้อมูลเข้าสู่การวิเคราะห์ประสิทธิภาพในการจำแนกข้อความโดยใช้อัลกอริทึมต้นไม้ตัดสินใจ นาอ์ฟเบย์ และแรนดอมฟอร์เรส

สรุปงานวิจัยนี้ ผู้วิจัยมุ่งเน้นเสนอการวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ และประเมินค่าความแม่นยำและค่าความถูกต้องของการจำแนกด้านการประเมินและความคิดเห็น จำแนกโดยใช้ 3 อัลกอริทึม ดังนี้

1.อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree)

1.1 ผลวิเคราะห์การจำแนกด้านการประเมินจากการทำนายทั้งหมด ได้แก่ ด้านอาจารย์ผู้สอน (Teaching) 60.95% ด้านกิจกรรมการเรียนการสอน (Activity) 71.93% และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ (Development) 86.41% ความถูกต้อง (Accuracy) 75.02%

1.2 ผลวิเคราะห์การจำแนกความคิดเห็นจากการทำนายทั้งหมดเชิงบวก (Positive) 77.62% เชิงลบ (Negative) 81.47% และเชิงกลาง (Neutral) 14.54% ความถูกต้อง (Accuracy) 75.94%

2. อัลกอริทึมนาอีฟ เบย์ (Naive Bay)

2.1 ผลวิเคราะห์การจำแนกด้านการประเมินจากการทำนายทั้งหมด ได้แก่ ด้านอาจารย์ผู้สอน (Teaching) 50.21% ด้านกิจกรรมการเรียนการสอน (Activity) 76.83% และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ (Development) 88.45% ความถูกต้อง (Accuracy) 73.42%

2.2 ผลวิเคราะห์การจำแนกความคิดเห็นจากการทำนายทั้งหมดเชิงบวก (Positive) 71.29% และเชิงกลาง (Neutral) 2.63% สำหรับเชิงลบ (Negative) ไม่สามารถประเมินประสิทธิภาพโดยรวมได้ โดยความถูกต้อง (Accuracy) 72.24%

3. อัลกอริทึมแรนดอมฟอเรส (Random Forest)

3.1 ผลวิเคราะห์การจำแนกด้านการประเมินจากการทำนายทั้งหมด ได้แก่ ด้านอาจารย์ผู้สอน (Teaching) 59.79% ด้านกิจกรรมการเรียนการสอน (Activity) 61.91% และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ (Development) 80.81% ความถูกต้อง (Accuracy) 66.58%

3.2 ผลวิเคราะห์การจำแนกความคิดเห็นจากการทำนายทั้งหมดเชิงบวก (Positive) 73.94% และเชิงลบ (Negative) 22.66% สำหรับเชิงกลาง (Neutral) ไม่สามารถประเมินประสิทธิภาพโดยรวมได้ ความถูกต้อง (Accuracy) 60.89%

5.2 อภิปรายผลการวิจัย

จากการทำวิทยานิพนธ์เพื่อการศึกษา เรื่อง การวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ

5.2.1 จากการศึกษางานวิจัย เพื่อวิเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษาโดยใช้เทคนิคเหมืองข้อความ เนื่องจากความคิดเห็นที่นักศึกษาแสดงความคิดเห็นโดยการพิมพ์บรรยายเป็นสิ่งที่สามารถบอกคุณลักษณะบางอย่างของอาจารย์ได้ และสามารถนำความคิดเห็นส่วนนี้ไปปรับปรุงและพัฒนาการเรียนการสอนได้ตามแนวคิดประโยชน์ของความคิดเห็น (ประภาเพ็ญ สุวรรณ, 2530 : 5) ดังต่อไปนี้ 1) ช่วยทำให้เข้าใจสิ่งแวดล้อมรอบ ๆ ตัว โดยการจัดรูปหรือการจัดระบบสิ่งของต่าง ๆ ที่อยู่รอบตัว 2) ช่วยให้มี Self-esteem โดยจะช่วยให้บุคคลหลีกเลี่ยงสิ่งที่ไม่ดีหรือปกปิดความจริงบางอย่าง ซึ่งนำความไม่พอใจมาสู่ตัว 3) ช่วยในการปรับตัวให้เข้ากับสิ่งแวดล้อมที่สลับซับซ้อน ซึ่งมีปฏิกิริยาตอบโต้หรือกระทำสิ่งใดสิ่งหนึ่งออกไปนั้น ส่วนมากจะนำความพึงพอใจมาให้

4) ช่วยให้ผู้บุคคลสามารถแสดงออกในด้านค่านิยม ความรู้สึกของตนเอง อันจะนำความพอใจมาสู่บุคคลนั้นๆ

5.2.2 การสังเคราะห์รูปแบบความคิดเห็นเกี่ยวกับผลประเมินการสอนของอาจารย์โดยนักศึกษา โดยการระบุปัจจัยที่บ่งบอกด้านการประเมิน 3 ด้าน ได้แก่ ด้านอาจารย์ผู้สอน (Teaching) ด้านกิจกรรมการเรียนการสอน (Activity) และด้านการพัฒนาศักยภาพนักศึกษาและเสริมสร้างคุณลักษณะที่พึงประสงค์ (Development) ซึ่งผลการสังเคราะห์รูปแบบทำให้ได้มาซึ่งค่าที่ระบุถึงด้านการประเมินต่างๆ เช่น ด้านอาจารย์ผู้สอนมีคำว่า เข้าใจง่าย สอนสนุก เป็นต้น ด้านกิจกรรมการเรียนการสอนมีคำว่า กิจกรรม สื่อ เป็นต้น ซึ่งสอดคล้องกับงานวิจัย คุณลักษณะของครูที่ผู้เรียนประทับใจ : ต้นแบบของครูดี (เรืองวิทย์ นนทภา) โดยวิธีการวิจัยใช้การตั้งคำถามปลายเปิดให้นักศึกษาบรรยายความประทับใจที่มีต่อครู และการระบุปัจจัยที่แสดงถึงความคิดเห็นเชิงบวก (Positive) เชิงลบ (Negative) และเชิงกลาง (Neutral) ตามงานวิจัยของกานดา แผล้วฒนากุล และปราโมทย์ ลีอนาม การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์ (2556)

5.2.3 ผลการประเมินประสิทธิภาพการวิเคราะห์จำแนกด้านการประเมินและความคิดเห็นที่นำมาใช้ในงานวิจัยครั้งนี้ ได้แก่ การทดสอบแบบ Decision Tree , Naïve Bay และ Random Forest ผลการประเมิน พบว่า การทดสอบแบบ Decision Tree ให้ค่าความถูกต้องในการจำแนกด้านการประเมินมากที่สุด 75.02% และค่าความถูกต้องในการจำแนกความคิดเห็น 75.94% สำหรับผลการประเมินแบบ Naïve Bay มีค่าความถูกต้องในการจำแนกด้านการประเมิน 73.42% ซึ่งใกล้เคียงกับ Decision Tree แต่เมื่อพิจารณาค่าความถูกต้องในการจำแนกความคิดเห็นมีค่าเป็น 55.40% แต่ไม่สามารถหาค่าประสิทธิภาพโดยรวม (F-measure) สำหรับความคิดเห็นเชิงลบได้ ดังนั้นอัลกอริทึมนี้จึงไม่เหมาะกับการทำใช้พัฒนาตัวต้นแบบ สำหรับ Random Forest มีค่าความถูกต้องสำหรับการจำแนกด้านการประเมิน 66.58% เมื่อพิจารณาค่าความถูกต้องในการจำแนกความคิดเห็นมีค่าเป็น 60.89% แต่ไม่สามารถหาค่าประสิทธิภาพโดยรวม (F-measure) สำหรับความคิดเห็นเชิงกลางได้ ดังนั้นอัลกอริทึมนี้จึงไม่เหมาะกับการทำใช้พัฒนาตัวต้นแบบเช่นกัน

5.2.4 กระบวนการวิเคราะห์ความคิดเห็นนำมาซึ่งการพัฒนาตัวต้นแบบที่สนับสนุนการวิเคราะห์ผลประเมินการสอนของอาจารย์โดยนักศึกษาเพื่อให้ได้ข้อเสนอแนะเพื่อการพัฒนาการสอนของอาจารย์

5.3 ปัญหาและอุปสรรคในการทำงานวิจัย

5.3.1 การเก็บข้อมูลที่มีความคลาดเคลื่อนและผิดพลาด เนื่องจากความคิดเห็นที่เราเก็บรวบรวมมานั้น บางความคิดเห็นจะมีการพิมพ์อักษรที่ตกหล่นไปบางความคิดเห็นจะมีการพิมพ์สระและ

วรรณยุกต์เกินมา หรือการใช้คำแสง จึงทำให้การทำความสะอาดข้อมูลใช้เวลานานขึ้น และส่งผลให้การวิเคราะห์มีประสิทธิภาพลดลง

5.3.2 การเลือกอัลกอริทึมที่นำมาใช้ในการประเมินประสิทธิภาพการจำแนกข้อความ ในงานวิจัยอื่นๆ ส่วนใหญ่จะใช้ Support Vector Machine แต่สำหรับข้อมูลชุดนี้ไม่สามารถใช้อัลกอริทึมนี้ได้ ในโปรแกรม Rapid Miner ซึ่งต้องศึกษาเงื่อนไขการเลือกอัลกอริทึมให้เชี่ยวชาญในด้านการใช้งานมากขึ้น

5.4 ข้อเสนอแนะงานวิจัย

5.4.1 ชุดข้อมูลนำมาจากสถาบันการศึกษาแห่งหนึ่งเท่านั้น ดังนั้นเงื่อนไขต่างๆ จะใช้ในการจำแนกได้เฉพาะในสถาบันนี้เท่านั้น

5.4.2 ควรหาชุดข้อมูลที่มาจากหลากหลายแหล่งที่มาให้มากขึ้น เช่น สื่อออนไลน์อื่นๆ เป็นต้น

5.4.3 ควรมีการทดสอบประสิทธิภาพการจำแนกข้อมูลโดยใช้อัลกอริทึมสำหรับการจำแนกอื่นๆ

5.4.4 เนื่องจากข้อมูลความคิดเห็นต่อการเรียนการสอนของอาจารย์โดยนักศึกษาที่นำมาใช้ในการกำหนดด้านการประเมิน และด้านความคิดเห็นมีความไม่สอดคล้องกัน ส่งผลให้การจำแนกข้อมูลที่มีน้อยกว่าไม่สามารถจำแนกได้อย่างมีประสิทธิภาพ ดังนั้นจึงควรมีขั้นตอนการจำทำข้อมูลในสอดคล้องกันก่อนนำข้อมูลไปใช้งานต่อไป

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

บรรณานุกรม

TNI

บรรณานุกรม

- [1] นวพร ชลารักษ์, “บทบาทของครูกับการเรียนการสอนในศตวรรษที่ 21,” *วารสารวิชาการมหาวิทยาลัยฟาร์อีสเทอร์น*, ปีที่ 9, ฉบับที่ 1, หน้า 64-71, พฤษภาคม – กรกฎาคม, 2558.
- [2] ริเรื่องรอง รัตนวิไลสกุล, “การศึกษาคุณลักษณะครูพึงประสงค์ของนักศึกษาและอาจารย์มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี,” *วารสารวิจัยและพัฒนามหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี*, ปีที่ 25, ฉบับที่ 2, หน้า 149-165, เมษายน – มิถุนายน, 2545.
- [3] สำนักงานคณะกรรมการการอุดมศึกษา กระทรวงศึกษาธิการ, *กรอบแผนอุดมศึกษาระยะยาว 15 ปี ฉบับที่ 2 (พ.ศ.2551-2565)*, กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2551.
- [4] ณัฐชนันท์ นุเสน และคณะ, “กระบวนการพัฒนาครู ด้านการวัดและประเมินผลตามสภาพจริง โรงเรียนสังกัดองค์การบริหารส่วนจังหวัด,” *วารสารมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม*, ปีที่ 12, ฉบับที่ 2, หน้า 409-427, กรกฎาคม – ธันวาคม, 2561.
- [5] นพมาศ อีระเวคิน, *จิตวิทยาสังคมกับชีวิต*, กรุงเทพฯ: โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2539.
- [6] ประดิษฐ์ อูปรนัย, *จิตวิทยากับการศึกษา*, กรุงเทพฯ: มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2533.
- [7] B. John, *Research in Education*, New Jersey: Prentice Hall, Inc., 1977.
- [8] กาญจนา คำสุวรรณ และนิตยา เสาวมณี, *จิตวิทยาเบื้องต้น*, กรุงเทพฯ: โรงพิมพ์ทวีการพิมพ์, 2524.
- [9] ประภาเพ็ญ สุวรรณ, *ทัศนคติ: การวัดการเปลี่ยนแปลงและพฤติกรรมอนามัย*, กรุงเทพฯ: โรงพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2530.
- [10] นาวิก นำเสียง, “ข้อมูล สื่อสังคมออนไลน์ที่ผู้บริหารระดับสูงควรรู้,” [Online]. Available: <https://bit.ly/1ahB8HK>. [Accessed: 25 พฤษภาคม 2562]
- [11] R. Feldman and J. Sanger, *The Text Mining Handbook*, England: Cambridge University, 2006.
- [12] E. Brethenoux, “IBM Business Analytics,” IBM [Online], Available: <http://www.ibm.com>. [Accessed: May 25, 2019]
- [13] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Journal of Computer and Communications*, vol.3, no.12, pp. 513-523, December 1988.
- [14] มาสวีร์ มาศดิศรโชติ, “การทำเหมืองความคิดเห็นภาษาไทย,” *วารสารศรีปทุมปริทัศน์*, ปีที่ 6, ฉบับที่ 1, หน้า 62-71, มกราคม – ธันวาคม, 2557.

- [15] ยืน ภู่วรรณ, *การประมวลผลภาษาธรรมชาติ*, กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, 2535.
- [16] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, England: The MIT Press, 1999
- [17] ชนนท์ หลีน้อย, “การสร้างเครือข่ายคำไทยของมโนทัศน์พื้นฐานร่วมของเอนทิตีลำดับที่หนึ่งด้วยวิธีการแปลงสองทางและการใช้พจนานุกรมที่สร้างด้วยวิธีการแตกต่างกัน,” วิทยานิพนธ์ อ.ม. (ภาษาศาสตร์), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2551.
- [18] กานดา รุณนะพงศา และปิยธร อุราธรรมกุล, “การตัดคำภาษาไทยโดยการปรับปรุงกฎและพจนานุกรมแบบใหม่,” รายงานการวิจัย คณะวิศวกรรมศาสตร์, มหาวิทยาลัยขอนแก่น, ขอนแก่น, 2549.
- [19] J.Z. Mohammed and W. Meira, *Data Mining and Analysis Fundamental concepts and Algorithms*, USA: Cambridge University, 2014.
- [20] ณรงค์ฤทธิ์ ประสานตรี, “รูปแบบการพัฒนาอาจารย์มหาวิทยาลัยราชภัฏกลุ่มรัตนโกสินทร์,” *วารสารวิจัยราชภัฏพระนคร สาขามนุษยศาสตร์และสังคมศาสตร์*, ปีที่ 11, ฉบับที่ 1, หน้า 203-216, มกราคม – มิถุนายน, 2559.
- [21] เรืองวิทย์ นนทภา, “คุณลักษณะของครูที่ผู้เรียนประทับใจ: ต้นแบบของครูดี,” *วารสารวิชาการมหาวิทยาลัยฟาร์อีสเทอร์น*, ปีที่ 10, ฉบับที่ 2, หน้า 142-154, เมษายน – มิถุนายน, 2559.
- [22] อติณัฏ วัฒนบุรณนท, “การจำแนกคุณภาพการใช้งานและปัญหาของเกมแอ็คชั่นจากบทวิจารณ์ของผู้ใช้ด้วยการทำเหมืองข้อความ,” วิทยานิพนธ์ วท.ม. (วิศวกรรมซอฟต์แวร์), จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2560.
- [23] กานดา แผลว์พัฒนากุล, “การวิเคราะห์เหมืองข้อเสนอแนะจากบทวิจารณ์รายการโทรทัศน์,” วิทยานิพนธ์ วท.ม. (บริหารเทคโนโลยีสารสนเทศ), สถาบันบัณฑิตพัฒนบริหารศาสตร์, กรุงเทพฯ, 2555.
- [24] กานดา แผลว์พัฒนากุล และปรามิทธิ์ ลีอนาม, “การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์,” *วารสารการจัดการสมัยใหม่*, ปีที่ 11, ฉบับที่ 2, หน้า 25-27, กรกฎาคม – ธันวาคม, 2556.
- [25] S. Brain et al., “Analyzing twitter to explore perceptions of Asian restaurants,” *Journal of Hospitality and Tourism Technology*, vol. 7, no. 4, pp. 405-422, November 2016.

- [26] R. Talib et al., "Text mining: techniques, applications and issues," *International Journal of Advanced Computer Science and Applications*, vol. 7, no.11, pp. 414-418, November 2016.
- [27] ปราชญภาคย์ เหล่าสังข์สุข และคณะ, "การวิเคราะห์ความคิดเห็นเกี่ยวกับร้านอาหารบนเว็บไซต์รีวิว," *วารสารมหาวิทยาลัยทักษิณ*, ปีที่ 20, ฉบับที่ 1, หน้า 39-48, มกราคม - มิถุนายน, 2560.
- [28] พนิดา ทรงรัมย์, "การจำแนกความคิดเห็นทางการเมืองบนเครือข่ายสังคมออนไลน์ โดยใช้วิธีการจำแนกแบบสัมพันธ์," *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี*, ปีที่ 6, ฉบับที่ 2, หน้า 83-93, มกราคม - มิถุนายน, 2559.
- [29] สุทธิวรรณ จรรย์พัฒนานุกูล และมาลีรัตน์ โสदानิล, "ระบบวิเคราะห์ข้อความจากเว็บบอร์ดร้านอาหารโดยใช้เหมืองข้อมูลความเห็น," *วิทยานิพนธ์ วท.ม. (วิศวกรรมคอมพิวเตอร์)*, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ, 2555.
- [30] ทศสิทธิ์ ไพบูลย์นันทพงศ์ และมณเฑียร รัตนศิริวงศ์วุฒิ, "การจัดหมวดหมู่ข้อมูลจากการทำเหมืองข้อมูลของทวิตเตอร์ กรณีศึกษา #Tutorial," *การประชุมทางวิชาการระดับชาติ ด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 8*, NCCIT 2012, 9-10 พฤษภาคม 2555, หน้า 605.
- [31] อติเทพ ไชยสาร และรัฐสิทธิ์ สุขะหุต, "การประมาณอารมณ์จากความคิดเห็นภาษาไทยโดยใช้เทคนิคการเรียนรู้ของเครื่อง," *การประชุมทางวิชาการระดับชาติ ด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 9*, NCCIT 2013, 9-10 พฤษภาคม 2556, หน้า 260.
- [32] รวิสุตา เทศเมือง และนิเวศ จิระวิชิตชัย, "การวิเคราะห์ความคิดเห็นภาษาไทยเกี่ยวกับการรีวิวสินค้าออนไลน์โดยใช้ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมทซิน," *วารสารวิศวกรรมศาสตร์มหาวิทยาลัยสยาม*, ปีที่ 18, ฉบับที่ 1, หน้า 1-11, มกราคม - มิถุนายน, 2560.

ภาคผนวก

TNI

ภาคผนวก ก.

โค้ดโปรแกรมพีแฮชพี (PHP)

TNI

โค้ดการออกแบบ UI หน้าเว็บไซต์

```

<style type="text/css">
    body{
        font-size:16px;
    }
</style>

<div style="float:left; width:400px;">
    <form method="post">
        <b>Feeling Sentiment : </b><br/>
        <textarea name="text_to_segment" cols="40" rows="50"
style="width:300px;height:300px;"><?php echo isset($_POST['text_to_segment'])?
trim($_POST['text_to_segment']):" "></textarea>
        <br/>
        <input type="submit" value="ตกลง" style="width:301px;height:40px;font-
size:18px;background: #eee"/>

    </form>
</div>

<div style="width:1000px;">
    <?php
    if ($_POST) {

        $time_start = microtime(true);

        $text_to_segment = trim($_POST['text_to_segment']);
        echo '<hr/>';
    }

```

```

//echo '<b>ประโยคที่ต้องการตัดคือ: </b>' . $text_to_segment . '<br/>';
include(dirname(__FILE__) . DIRECTORY_SEPARATOR .
'THSplit/segment.php');

$segment = new Segment();
//echo '<hr/>';

$result = $segment->get_segment_array($text_to_segment);
echo implode(' | ', $result);

function convert($size) {
    $unit = array('b', 'kb', 'mb', 'gb', 'tb', 'pb');
    return @round($size / pow(1024, ($i = floor(log($size, 1024)))), 2) . ' '.
$unit[$i];
}

$time_end = microtime(true);
$time = $time_end - $time_start;
        echo '<br/><b>Feeling : </b> ';
        echo '<br/><b>Sentiment : </b> ';
echo '<br/><b>ประมวลผลใน: </b>' .round($time,4).' วินาที';

foreach($result as $row)
{
    if (mb_strlen($row) > 12)
    {
        echo $row.'<br/>';
    }
}
?>

</div>
</body>

```

```
<script type="text/javascript">

    var _gaq = _gaq || [];
    _gaq.push(['_setAccount', 'UA-28746062-1']);
    _gaq.push(['_trackPageview']);

    (function() {
        var ga = document.createElement('script'); ga.type = 'text/javascript';
        ga.async = true;
        ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www')
        + '.google-analytics.com/ga.js';
        var s = document.getElementsByTagName('script')[0];
        s.parentNode.insertBefore(ga, s);
    })();

</script>
</html>
```

TNI

TAHITI - NICHI INSTITUTE OF TECHNOLOGY

โค้ดการทำงานภายในโปรแกรม

```
<?php

class Segment {

    private $_input_string;
    private $_dictionary_array = array();
    private $_thcharacter_obj;
    private $_unicode_obj;
    private $_segmented_result = array();

    function __construct() {
        if (!class_exists('Thchracter')) {
            include(dirname(__FILE__) . DIRECTORY_SEPARATOR . 'thcharacter.php');
        }

        if (!class_exists('Unicode')) {
            include(dirname(__FILE__) . DIRECTORY_SEPARATOR . 'unicode.php');
        }

        $file_handle = fopen(dirname(__FILE__) . DIRECTORY_SEPARATOR .
'dictionary' . DIRECTORY_SEPARATOR . 'dictionary.txt', "rb");
        while (!feof($file_handle)) {
            $line_of_text = fgets($file_handle);
            $this->_dictionary_array[crc32(trim($line_of_text))] = trim($line_of_text);
        }
    }
}
```

```

fclose($file_handle);

// Load Helper Class/
$this->_unicode_obj = new Unicode();
$this->_thcharacter_obj = new Thchracter();
}

private function clear_duplicated($string) {
    $input_string_split = $this->_unicode_obj->uni_strsplit($string);
    $previous_char = "";
    $previous_string = "";
    $dup_list_array = array();
    $dup_list_array_replace = array();
    foreach ($input_string_split as $current_char) {

        if ($previous_char == $current_char) {
            $previous_char = $current_char;
            $previous_string .= $current_char;
        } else {
            if (mb_strlen($previous_string) > 3) {
                $dup_list_array[] = $previous_string;
                $dup_list_array_replace[] = $current_char;
                $string = str_replace($previous_string, $previous_char, $string);
            }
            $previous_char = $current_char;
            $previous_string = $current_char;
        }
    }
    if (mb_strlen($previous_string) > 3) {
        $dup_list_array[] = $previous_string;
        $dup_list_array_replace = $current_char;
    }
}

```



```

        $current_array_result = $this->_segment_by_dictionary_reverse($current_string_reverse_array);
        foreach ($current_array_result as $each_result) {
            if (trim($each_result) != "")
                $this->_segmented_result[] = trim($each_result);
        }
    }

    $tmp_result = array();
    foreach ($this->_segmented_result as $result_row) {
        if (mb_strlen($result_row) > 10) {

            $current_string_array = $this->_unicode_obj->uni_strsplit(trim($result_row));

            $current_array_result = $this->_segment_by_dictionary($current_string_array);

            foreach ($current_array_result as $current_result_row) {
                $tmp_result[] = trim($current_result_row);
            }
        } else {
            $tmp_result[] = $result_row;
        }
    }

    $this->_segmented_result = $tmp_result;
    return $this->_segmented_result;
}

private function _segment_by_dictionary($input_array) {

```

```
$result_array = array();  
$tmp_string = "";  
  
$pointer = 0;  
$length_of_string = count($input_array)-1;  
  
while ($pointer <= $length_of_string) {  
    $tmp_string .= $input_array[$pointer];  
  
    if (isset($this->_dictionary_array[crc32($tmp_string)])) {  
        $dup_array = array();  
        $dup_array[] = array(  
            'title' => $tmp_string,  
            'to_mark' => $pointer + 1,  
        );  
        $count_more = 0;  
        $more_tmp = $tmp_string;  
        //echo $more_tmp.'<br/>';  
  
        for ($i = $pointer + 1; $i <= $length_of_string; $i++) {  
            $more_tmp .= $input_array[$i];  
  
            if (isset($this->_dictionary_array[crc32($more_tmp)])) {  
                $dup_array[] = array(  
                    'title' => $more_tmp,  
                    'to_mark' => $i + 1,  
                );  
                //print_r($dup_array);  
            }  
            $count_more++;  
        }  
    }  
}
```

```
}

if (count($dup_array) > 0) {
    $result_array[] = $dup_array[count($dup_array) - 1]['title'];

    $pointer = $dup_array[count($dup_array) - 1]['to_mark'];

} else {

}

$dup_array = array();
$tmp_string = "";
continue;
}

$pointer++;
}

if ($tmp_string != "") {
    $result_array[] = $tmp_string;
}

if (count($result_array) == 0) {
    return array(implode($input_array));
}

return $result_array;
}

private function _segment_by_dictionary_reverse($input_array) {
```

```
$result_array = array();
$tmp_string = "";

$pointer = 0;
$length_of_string = count($input_array)-1;

while ($pointer <= $length_of_string) {

    $tmp_string = $input_array[$pointer] . $tmp_string;
    if (isset($this->_dictionary_array[crc32($tmp_string)])) {
        $dup_array = array();
        $dup_array[] = array(
            'title' => $tmp_string,
            'to_mark' => $pointer + 1,
        );
        $count_more = 0;
        $more_tmp = $tmp_string;
        //echo $more_tmp.'<br/>';

        for ($i = $pointer + 1; $i <= $length_of_string; $i++) {
            $more_tmp = $input_array[$i] . $more_tmp;

            if (isset($this->_dictionary_array[crc32($more_tmp)])) {
                $dup_array[] = array(
                    'title' => $more_tmp,
                    'to_mark' => $i + 1,
                );
                //print_r($dup_array);
            }
            $count_more++;
        }
    }
}
```

```
}

if (count($dup_array) > 0) {
    $result_array[] = $dup_array[count($dup_array) - 1]['title'];

    $pointer = $dup_array[count($dup_array) - 1]['to_mark'];

    //echo $to_mark;
} else {

}

$dup_array = array();
$tmp_string = "";
continue;
}

$pointer++;
}

if ($tmp_string != "") {
    $result_array[] = $tmp_string;
}

if (count($result_array) == 0) {
    return array(implode(array_reverse($input_array)));
}

return array_reverse($result_array);
}
}
```