



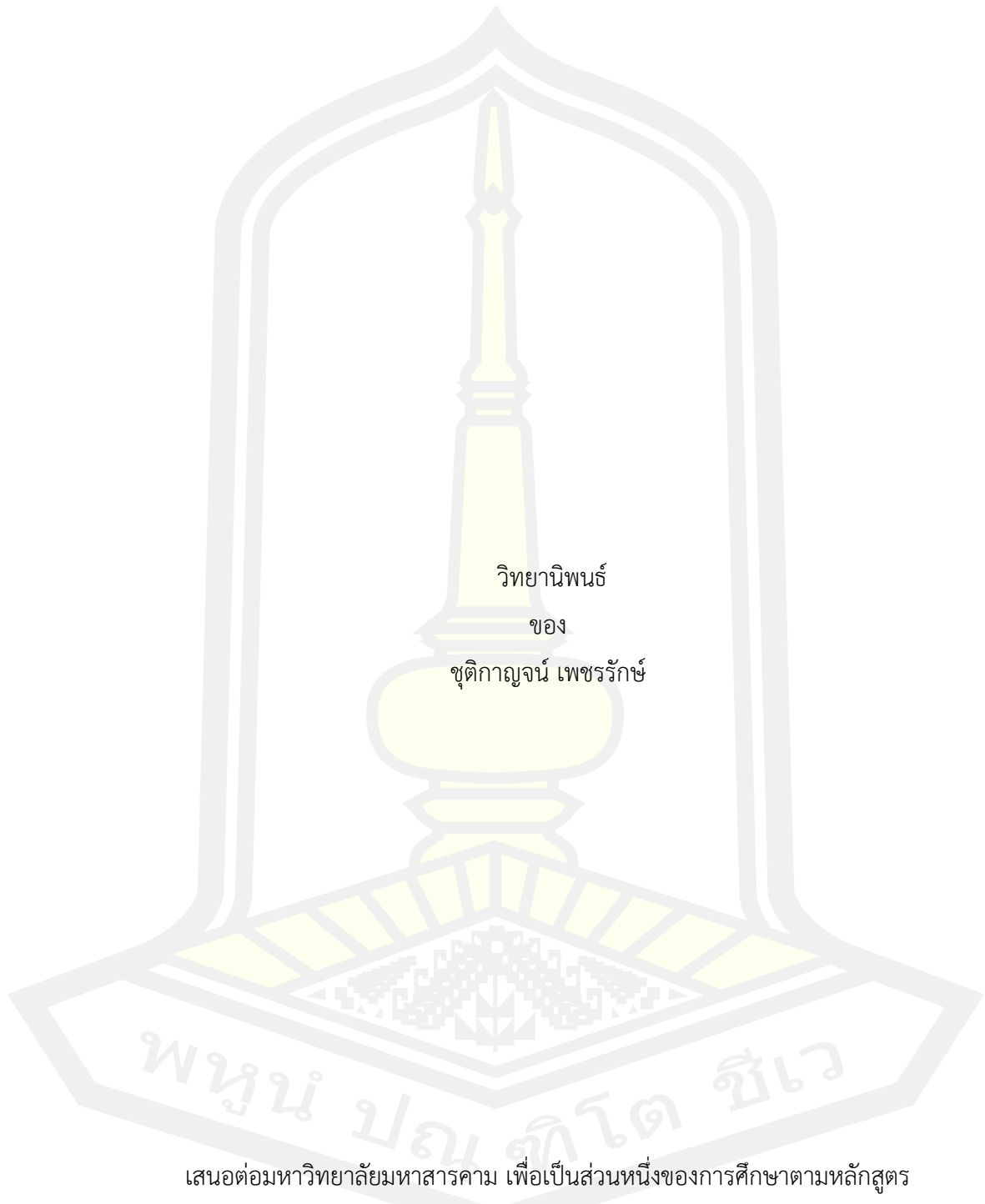
แบบจำลองความรู้สึกของประชาชนต่ออดีตด้วยเทคนิคเหมือนความคิดเห็น

วิทยานิพนธ์
ของ
ชุตติกาญจน์ เพชรรักษ์

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ตุลาคม 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

แบบจำลองความรู้สึกของประชาชนต่ออดีตด้วยเทคนิคเหมืองความคิดเห็น

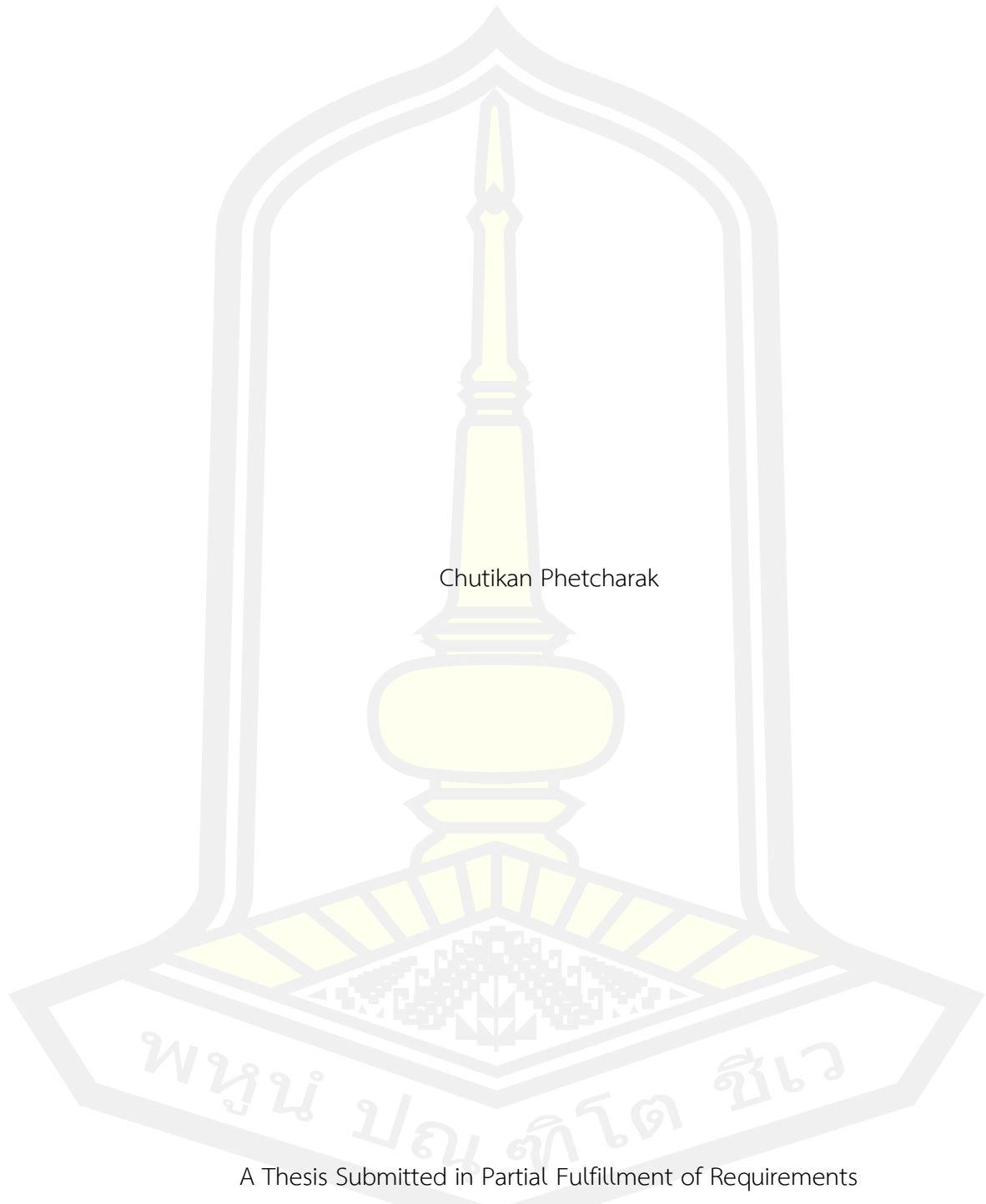


เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ตุลาคม 2565

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Model of People Sentiment towards Autism Using Opinion Mining Techniques



Chutikan Phetcharak

A Thesis Submitted in Partial Fulfillment of Requirements
for Master of Science (Information Technology)

October 2022

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวชุตติกาญจน์ เพชร
รักษ์ แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผศ. ดร. ธาดา จันทะคุณ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. จารีย์ ทองคำ)

.....กรรมการ

(ผศ. ดร. ฉัตรตระกูล สมบัติธีระ)

.....กรรมการ

(ผศ. ดร. สอาทิตย์ แสงประดิษฐ์)

มหาวิทยาลัยขอนแก่นให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

.....
(ผศ. ศศิธร แก้วมัน)

คณบดีคณะวิทยาการสารสนเทศ

.....
(รศ. ดร. กริสน์ ชัยมูล)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง	แบบจำลองความรู้สึกของประชาชนต่อออตติสติกด้วยเทคนิคเหมืองความคิดเห็น		
ผู้วิจัย	ชุตติกาญจน์ เพชรรักรักษ์		
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. จารีย์ ทองคำ		
ปริญญา	วิทยาศาสตรมหาบัณฑิต	สาขาวิชา	เทคโนโลยีสารสนเทศ
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม	ปีที่พิมพ์	2565

บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออตติสติก และเปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนที่มีต่อออตติสติก งานวิจัยนี้ใช้เทคนิคเหมืองความคิดเห็น 5 เทคนิค ได้แก่ เทคนิคนาอิวเบย์ เทคนิคต้นไม้ตัดสินใจ ซี4.5 เทคนิคป่าสุ่ม เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ได้ถูกนำมาสร้างแบบจำลองความคิดเห็นของประชาชนต่อออตติสติก โดยเก็บรวบรวมความคิดเห็นของประชาชนต่อออตติสติกจากสื่อสังคมออนไลน์ในเว็บไซต์พันทิป เฟซบุ๊ก และทวิตเตอร์ จำนวนทั้งหมด 1,766 ความคิดเห็น คุณลักษณะที่ใช้ในงานวิจัยนี้ ได้แก่ คำกริยา คำวิเศษณ์ และคำคุณศัพท์ โดยประเภทของคำดังกล่าวสามารถระบุความรู้สึกได้เป็นอย่างดี หลักการ 10-โฟลตครอสวาเลชันได้ถูกนำมาแบ่งชุดข้อมูลเป็นชุดทดสอบและชุดสอน ค่าความแม่นยำ ค่าความระลึกลับ และค่าความถ่วงดุล ได้ถูกใช้เพื่อการเปรียบเทียบประสิทธิภาพของแบบจำลองโดยใช้จากการศึกษากระบวนการในการทำเหมืองความคิดเห็นพบว่า คำคุณศัพท์ มีผลต่อประสิทธิภาพของแบบจำลองมากที่สุด เมื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความรู้สึกของประชาชนต่อออตติสติกพบว่า แบบจำลองที่สร้างจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เป็นแบบจำลองที่ดีที่สุด โดยให้ค่าเฉลี่ยความแม่นยำ ค่าเฉลี่ยความระลึกลับ และค่าเฉลี่ยความถ่วงดุลสูงถึงร้อยละ 97.71, 98.88 และ 98.29 ตามลำดับ

คำสำคัญ : เหมืองความคิดเห็น, ความรู้สึกของประชาชน, ออตติสติก

TITLE	Model of People Sentiment towards Autism Using Opinion Mining Techniques		
AUTHOR	Chutikan Phetcharak		
ADVISORS	Assistant Professor Jaree Thongkam , Ph.D.		
DEGREE	Master of Science	MAJOR	Information Technology
UNIVERSITY	Maharakham University	YEAR	2022

ABSTRACT

The purpose of this research was to study the opinion mining process for analyzing people's feelings towards autism and to compare models to classify people's feelings towards autism with opinion mining technique. In this research, the Naïve Bayes, C4.5 Decision Tree, Random Forest, K-Nearest neighbor and Support Vector Machine techniques were used to build public opinion on autism models. A total of 1,766 people's opinions on autism were collected from social media sites on Pantip, Facebook and Twitter. Features used in this research are verbs, adverbs and adjectives. Such features can describe feelings very well. The 10-fold cross-validation was employed to divide datasets into test and training sets. Precision, Recall and F-Measure values were utilized to compare the model performance. After studying the process of mining opinions, it was found that the adjective had the greatest effect on the performance of the models. When comparing the model's performance in predicting people's feelings about autism, it was found that, Models created from vector machine support techniques is the best model. in providing precision, recall and F-Measure up to 97.71, 98.88 and 98.29, respectively.

Keyword : Opinion Mining, People Sentiment, Autism

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยดี เนื่องจากได้รับความกรุณาอย่างสูงจาก ผู้ช่วยศาสตราจารย์ ดร.จารีย์ ทองคำ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ที่มอบความรู้และให้คำปรึกษารวมถึง คำแนะนำต่าง ๆ ในการเรียนและการทำวิทยานิพนธ์ และขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ธาดา จันทะคุณ ประธานกรรมการ ผู้ช่วยศาสตราจารย์ ดร.ฉัตรตระกูล สมบัติธีระ และผู้ช่วยศาสตราจารย์ ดร.สาธิต แสงประดิษฐ์ กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ตลอดจนการปรับปรุงแก้ไข ข้อบกพร่องต่าง ๆ จนวิทยานิพนธ์ฉบับนี้เกิดความสมบูรณ์ที่สุด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณอาจารย์สาขาวิชาภาษาไทย คณะมนุษยศาสตร์และสังคมศาสตร์ ที่กรุณาให้ความอนุเคราะห์ช่วยเหลือเรื่องการจัดทำดัชนีคำในวิทยานิพนธ์ฉบับนี้ รวมทั้งให้คำแนะนำต่าง ๆ เกี่ยวกับด้านภาษาไทย ทางผู้วิจัยขอขอบพระคุณในเมตตริจิตที่กรุณาสละเวลาเพื่อช่วยเหลือผู้วิจัย

ผู้วิจัยขอขอบคุณประชาชนชาวไทยผู้มาแสดงความคิดเห็นต่าง ๆ ต่อออกทิสติกบนสื่อสังคมออนไลน์ ทั้งกลุ่มครอบครัวผู้ที่มีความพิการทางออกทิสติกในความดูแล กลุ่มผู้ที่มีความพิการทางออกทิสติก และกลุ่มบุคคลทั่วไป ซึ่งความคิดเห็นดังกล่าวเป็นประโยชน์ต่อการนำมาวิเคราะห์ความรู้สึกรักของประชาชนที่มีต่อออกทิสติกทั้งในเชิงบวกและเชิงลบ และจุดประกายให้เกิดประเด็นการวิจัยในวิทยานิพนธ์ฉบับนี้ขึ้น

ผู้วิจัยมีความคาดหวังว่า วิทยานิพนธ์ฉบับนี้จะมีประโยชน์แก่ผู้ที่สนใจศึกษาเกี่ยวกับการทำเหมืองความคิดเห็น จึงขอมอบความดีทั้งหมดนี้ ให้แก่เหล่าคุณอาจารย์ที่ได้ให้คำแนะนำและแนวทางในการแก้ไขปัญหา รวมทั้งครอบครัวและเพื่อนที่เป็นพลังใจและแรงผลักดันทำให้สามารถผ่านพ้นอุปสรรคไปได้ด้วยดี ทั้งนี้ผู้วิจัยยินดีที่จะรับฟังคำแนะนำจากทุกท่านที่ได้เข้ามาศึกษางานวิทยานิพนธ์ฉบับนี้ เพื่อเป็นประโยชน์ในการพัฒนางานวิจัยต่อไป

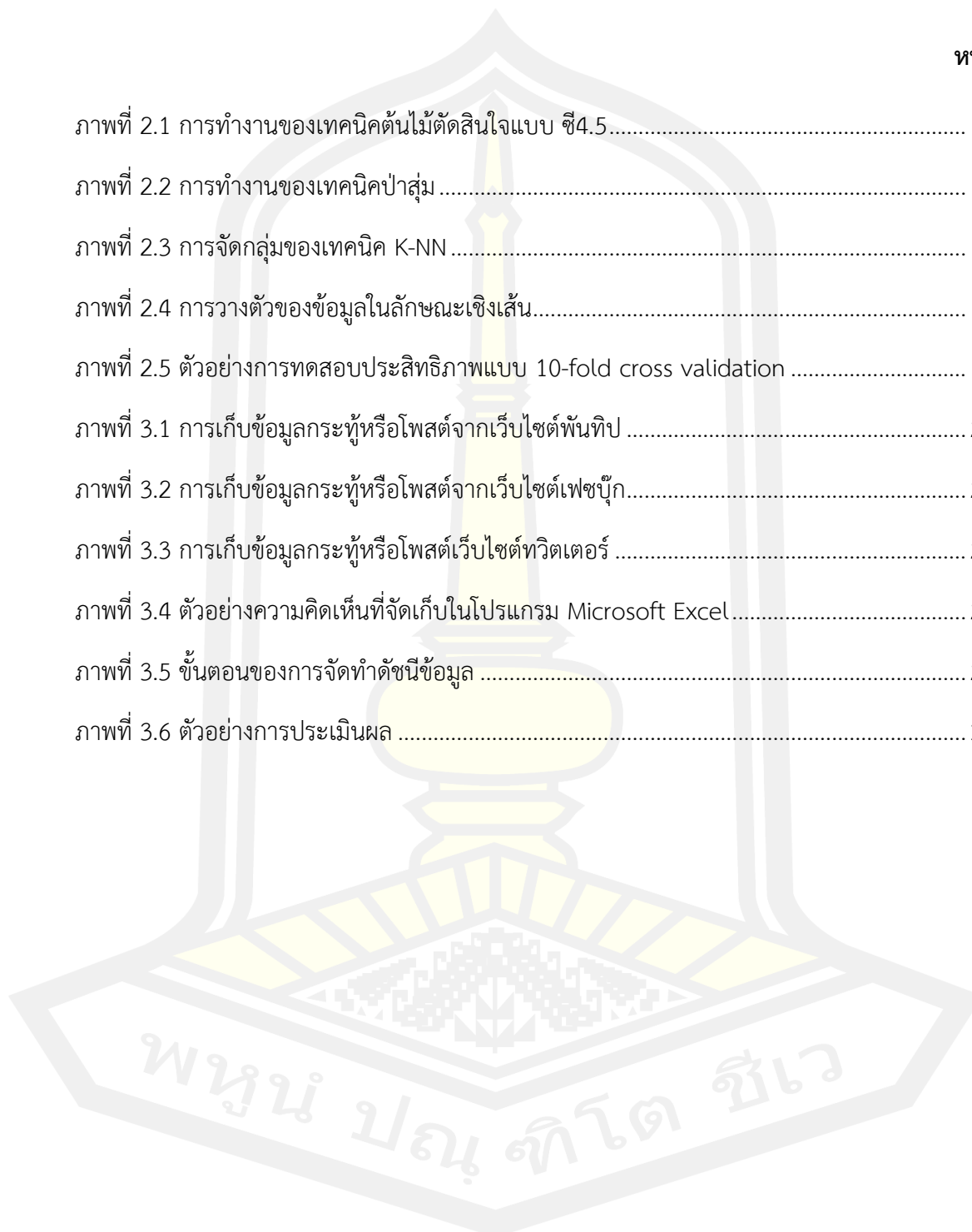
สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญภาพ.....	ฌ
สารบัญตาราง.....	ญ
บทที่ 1 บทนำ.....	1
1.1 หลักการและเหตุผล.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตของการวิจัย.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.5 นิยามศัพท์เฉพาะ.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 ความคิดเห็นของประชาชนต่อออตติค.....	5
2.1.2 ความรู้พื้นฐานเกี่ยวกับออตติค.....	5
2.1.3 การทำเหมืองความคิดเห็น.....	8
2.1.4 เทคนิคที่ใช้ในงานวิจัย.....	8
2.1.5 การวัดประสิทธิภาพของแบบจำลอง.....	17
2.2 งานวิจัยที่เกี่ยวข้อง.....	19
2.2.1 งานวิจัยเกี่ยวกับทัศนคติที่มีต่อออตติค.....	19

2.2.2 งานวิจัยเกี่ยวกับการทำเหมืองความคิดเห็น	20
บทที่ 3 วิธีดำเนินการวิจัย.....	25
3.1 การเตรียมข้อมูล	25
3.2 การจัดทำดัชนีข้อมูล	28
3.2.1 การกลั่นกรองข้อมูล.....	29
3.2.2 กระบวนการตัดคำและกำจัดคำหยุด	29
3.2.3 การทำถ่วงคำ.....	30
3.2.4 การกำหนดคุณลักษณะและการกำหนดคลาส.....	31
3.3 การสร้างแบบจำลอง.....	32
3.4 การประเมินผล	33
บทที่ 4 ผลการวิจัย	36
4.1 ผลการศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของ ประชาชนต่อออตติสติก.....	36
4.2 ผลการเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนต่อออตติสติก ด้วยเทคนิค เหมืองความคิดเห็น.....	38
บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ	43
5.1 สรุป	43
5.2 อภิปรายผล	44
5.3 ข้อเสนอแนะ	45
บรรณานุกรม.....	46
ภาคผนวก.....	52
ประวัติผู้เขียน.....	64

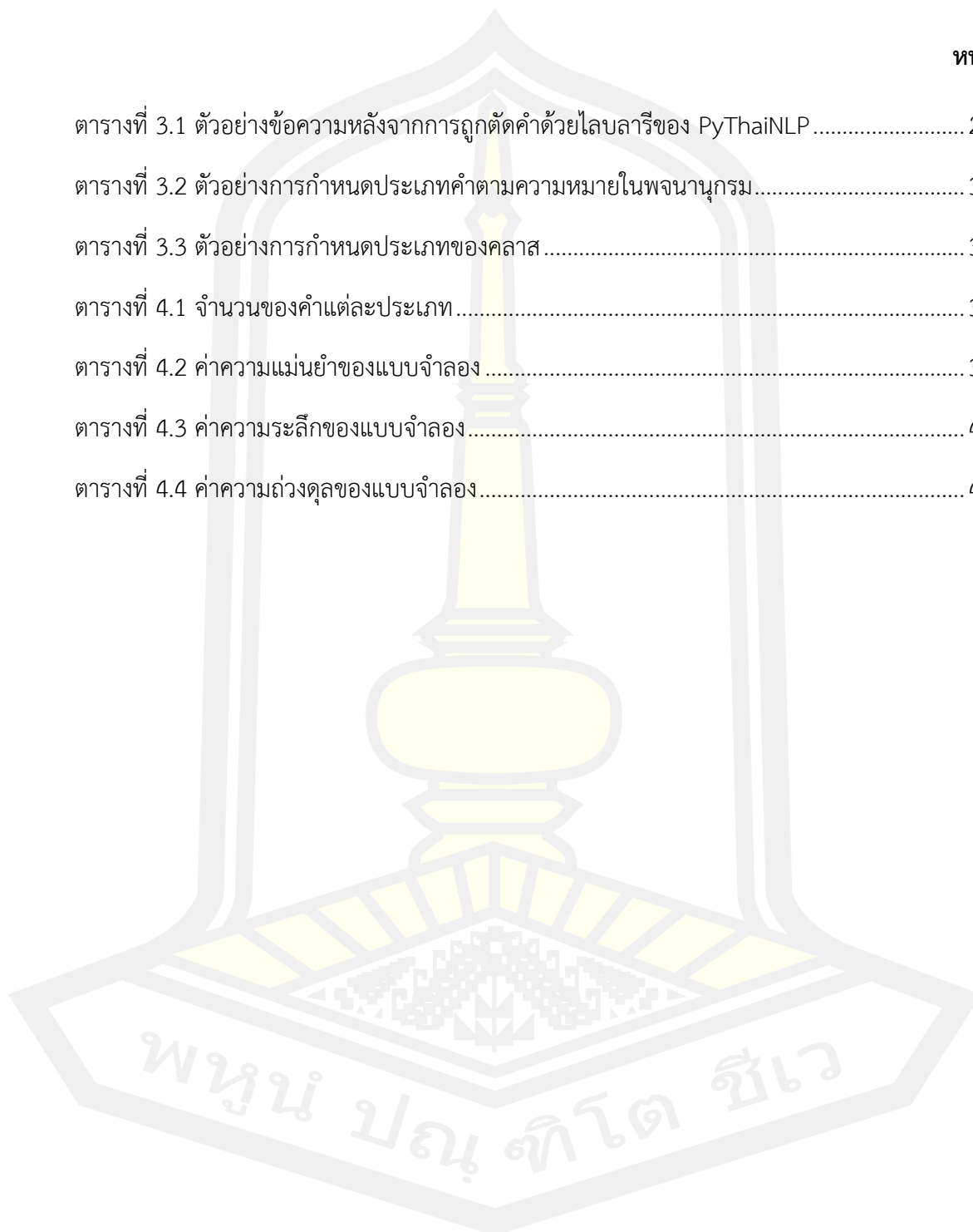
สารบัญภาพ

	หน้า
ภาพที่ 2.1 การทำงานของเทคนิคต้นไม้ตัดสินใจแบบ ซี4.5.....	11
ภาพที่ 2.2 การทำงานของเทคนิคป่าสุ่ม	13
ภาพที่ 2.3 การจัดกลุ่มของเทคนิค K-NN	14
ภาพที่ 2.4 การวางตัวของข้อมูลในลักษณะเชิงเส้น.....	16
ภาพที่ 2.5 ตัวอย่างการทดสอบประสิทธิภาพแบบ 10-fold cross validation	17
ภาพที่ 3.1 การเก็บข้อมูลกระทู้หรือโพสต์จากเว็บไซต์พันทิป	26
ภาพที่ 3.2 การเก็บข้อมูลกระทู้หรือโพสต์จากเว็บไซต์เฟซบุ๊ก.....	26
ภาพที่ 3.3 การเก็บข้อมูลกระทู้หรือโพสต์เว็บไซต์ทวิตเตอร์	27
ภาพที่ 3.4 ตัวอย่างความคิดเห็นที่จัดเก็บในโปรแกรม Microsoft Excel.....	27
ภาพที่ 3.5 ขั้นตอนของการจัดทำดัชนีข้อมูล	28
ภาพที่ 3.6 ตัวอย่างการประเมินผล	34



สารบัญตาราง

	หน้า
ตารางที่ 3.1 ตัวอย่างข้อความหลังจากการถูกตัดคำด้วยไลบรารีของ PyThaiNLP	29
ตารางที่ 3.2 ตัวอย่างการกำหนดประเภทคำตามความหมายในพจนานุกรม	30
ตารางที่ 3.3 ตัวอย่างการกำหนดประเภทของคลาส	32
ตารางที่ 4.1 จำนวนของคำแต่ละประเภท	37
ตารางที่ 4.2 ค่าความแม่นยำของแบบจำลอง	39
ตารางที่ 4.3 ค่าความระลึกลับของแบบจำลอง	40
ตารางที่ 4.4 ค่าความถ่วงดุลของแบบจำลอง	41



บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

ในสังคมไทยปัจจุบันกลุ่มคนพิการที่ได้รับการออกบัตรประจำตัวคนพิการมีจำนวน 2,096,931 คน พบแนวโน้มเพิ่มมากขึ้นเป็น ร้อยละ 3.17 ของประชากรทั้งประเทศ [1] ด้วยคำว่า “พิการ” จึงทำให้ถูกจำกัดความสามารถ ถูกขีดเส้นพื้นที่ ถูกจำกัดการเข้าร่วมสังคม และเกิดปัญหาความไม่เท่าเทียมกันในสังคม เนื่องจากกลุ่มคนเหล่านี้เป็นบุคคลที่มีความบกพร่องทางด้านต่าง ๆ ไม่ว่าจะเป็นความพิการทางอภิสติค ความพิการทางการเห็น ความพิการทางการได้ยินหรือสื่อความหมาย ความพิการทางการเคลื่อนไหวหรือทางร่างกาย ความพิการทางจิตใจหรือพฤติกรรม ความพิการทางสติปัญญา และความพิการทางการเรียนรู้ [2] โดยความบกพร่องดังกล่าวนี้อาจมีมาตั้งแต่กำเนิดหรือเกิดขึ้นภายหลัง [3] ผู้พิการทางอภิสติค คือ บุคคลที่มีข้อจำกัดในการปฏิบัติกิจกรรมในชีวิตประจำวันหรือการเข้าไปมีส่วนร่วมในกิจกรรมทางสังคม ซึ่งเป็นผลมาจากความบกพร่องทางพัฒนาการด้านสังคม ภาษา และการสื่อความหมาย พฤติกรรมและอารมณ์ โดยมีสาเหตุมาจากความผิดปกติของสมองและความผิดปกติที่แสดงก่อนอายุสองปีครึ่ง [2] ซึ่งผู้ที่มีความพิการทางอภิสติคจะมีหลากหลายรูปแบบ โดยจะแสดงอาการผิดปกติออกมา 2 ด้านใหญ่ ๆ ได้แก่ ความบกพร่องในการสื่อสารทางสังคมและการมีปฏิสัมพันธ์ทางสังคม และอีกด้านหนึ่ง คือ มีแบบแผนพฤติกรรม ความสนใจ หรือกิจกรรมที่จำกัดซ้ำ ๆ โดยสามารถสังเกตได้จากพฤติกรรมดังนี้ 1) ความผิดปกติทางสังคมและการมีปฏิสัมพันธ์กับผู้อื่น เช่น เรียกชื่อแล้วไม่หัน ไม่สนใจเวลาพูดด้วย ไม่มองหน้าไม่ค่อยสบตาเวลาพูดคุย ชอบเก็บตัวอยู่ตามลำพัง ชอบเล่นคนเดียว ไม่เข้ามาหาเพื่อให้ออด 2) ความผิดปกติทางภาษาและการสื่อสาร เช่น พูดซ้ำหรือไม่พูดเลย มีภาษาแปลก ๆ ที่คนอื่นไม่เข้าใจ ชอบพูดทวนคำหรือประโยคที่พ่อแม่พูดไป สานต่อบทสนทนาไม่ค่อยได้ ไม่เข้าใจคำพูดซ้ำชั้น อุปมาอุปมัย และ 3) ความผิดปกติของพฤติกรรมหรือความสนใจที่มีลักษณะซ้ำ ๆ จำกัด เช่น ปรับตัวต่อสิ่งใหม่ ๆ ได้ยาก ชอบทำอะไรซ้ำ ๆ หากมีการเปลี่ยนแปลงสิ่งที่เคยทำจะรู้สึกหงุดหงิดและอาละวาด สนใจเรื่องใดเรื่องหนึ่งมากเป็นพิเศษ มีพฤติกรรมกระโดด สะบัดมือซ้ำ ๆ รับประทานอาหารซ้ำ ๆ ซึ่งอาการเหล่านี้มีบางอย่างที่พบได้ในเด็กปกติแต่จะเป็นอยู่ชั่วคราวเท่านั้น ส่วนเด็กออทิสติกจะมีอาการหลายอย่างพร้อมกันและเป็นอยู่นาน [4] โดยความพิการทางอภิสติคที่กล่าวมาข้างต้นนี้มีผลต่อความรู้สึกของคนทั่วไปทั้งในเชิงบวกและเชิงลบ และความรู้สึกมีความสำคัญอย่างยิ่งต่อการพัฒนาในด้านต่าง ๆ ของคนพิการ อาทิ

ด้านการเรียนรู้ ด้านอารมณ์ ด้านสังคม และด้านจิตใจ [5] บางคนมองคนพิการเป็นภาระ ด้วยคุณค่าและชอบเอาความพิการมาล้อเล่น จนทำให้คนพิการใช้ชีวิตและเข้าสังคมยาก ในขณะที่บางคนเปิดใจยอมรับและพร้อมที่จะช่วยเหลือคนพิการ การวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออทิสติกในเชิงบวกและเชิงลบ จะนำไปสู่การสร้างค่านิยมในเชิงบวกของประชาชนในสังคมไทยที่มีต่อผู้ที่มีความพิการทางออทิสติกให้สามารถปรับตัวและมีปฏิสัมพันธ์ระหว่างผู้ที่มีความพิการทางออทิสติกกับผู้อื่น และเป็นประโยชน์สำหรับครอบครัว สถานศึกษาหรือหน่วยงานที่มีผู้ที่มีความพิการทางออทิสติกอยู่ร่วมด้วย เพื่อออกแบบจัดกิจกรรมส่งเสริมและพัฒนาในด้านต่าง ๆ ให้กับผู้ที่มีความพิการทางออทิสติก นอกจากนี้ยังสามารถนำผลการวิจัยไปใช้เป็นแนวทางสำหรับผู้สนใจศึกษาเกี่ยวกับทัศนคติของประชาชนต่อออทิสติกเพื่อนำไปสู่การปรับทัศนคติและการแก้ปัญหาการเข้าถึงสิทธิประโยชน์ต่าง ๆ ของผู้ที่มีความพิการทางออทิสติก อันจะนำไปสู่การยอมรับและการอยู่ร่วมกันในสังคมได้อย่างปกติสุข ซึ่งการนำแบบจำลองมาจำแนกความรู้สึกของประชาชนต่อออทิสติกมาพัฒนาเป็นแอปพลิเคชันจะช่วยให้ผู้สนใจสามารถจำแนกความคิดเห็นเชิงบวก หรือความคิดเห็นเชิงลบต่อออทิสติกได้อย่างรวดเร็วและแม่นยำ

การสร้างแบบจำลองเพื่อการจำแนกด้วยการทำเหมืองความคิดเห็น (Opinion Mining) เป็นการวิเคราะห์จำแนกความคิดเห็น จากความคิดเห็นส่วนใหญ่ที่เกิดขึ้นซึ่งจะเป็นไปในเชิงบวกและเชิงลบ [5] [6] [7] รวมถึงการวิเคราะห์ความรู้สึก โดยเทคนิคที่นิยมนำมาใช้ในการวิเคราะห์ความรู้สึก ได้แก่ เทคนิคป่าสุ่ม (Random Forest: RF) เทคนิคนาอิวเบย์ (Naïve Bayes: NB) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree: DT) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: K-NN) มาใช้ในการวิเคราะห์ความคิดเห็น เช่น พิษณุ พรมลา และ จริญญา แสนราช [8] ได้ศึกษาเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจของผู้เรียนที่มีต่อการจัดการเรียนการสอน ใช้เทคนิคการรวมกลุ่มเพื่อจำแนกความรู้สึกเชิงบวกและเชิงลบ โดยใช้เทคนิค 2 ประเภท คือ เทคนิคการรวมกลุ่มเพื่อจำแนกความรู้สึก ได้แก่ เทคนิค Vote, เทคนิค Bagging, เทคนิค RF และเทคนิคพื้นฐาน ได้แก่ เทคนิค DT, เทคนิค NB และ เทคนิค K-NN พบว่า เทคนิค Vote ให้ประสิทธิภาพมากที่สุด ส่วน สุภัตตรา สมเจตนา และ จารี ทองคำ [9] ได้ศึกษาการเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร โดยใช้เทคนิคริปปเปอร์ (Ripper), เทคนิค DT C4.5, เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค RF แสดงประสิทธิภาพของเทคนิคด้วยค่า Precision, Recall และ F-Measure พบว่า เทคนิค RF เป็นเทคนิคที่ดีที่สุด โดยให้ค่า

Precision 89.62%, Recall 78.38% และ F-Measure 83.55% และ Seed [10] ได้ศึกษาการทำเหมืองความคิดเห็นเกี่ยวกับข้อมูลในทวีตเตอร์ของสายการบินสหรัฐอเมริกา โดยใช้เทคนิค SVM, เทคนิค Logistic Regression, เทคนิค RF, เทคนิค XGBoost, เทคนิค NB และเทคนิค DT พบว่า SVM มีความแม่นยำสูงสุด 83.31%

นอกจากนี้ผู้วิจัยยังได้ศึกษาประเภทของคำที่ส่งผลต่อแบบจำลองเพื่อการจำแนกความคิดเห็นพบว่า ส่วนใหญ่เป็นคำคุณศัพท์ คำวิเศษณ์ และคำกริยา ที่มีผลต่อการจำแนก เช่น ศราวุฒิ เกิดถาวร และจารี ทองคำ [11] ได้ศึกษาผลกระทบของชนิดคำภาษาไทยที่มีต่อประสิทธิภาพของแบบจำลองการจำแนกความคิดเห็นที่ใช้กับความเห็นของคนไทยต่อโรคโควิด 19 จากเว็บไซต์ทวีตเตอร์ และพันทิป ใช้ชุดข้อมูล 3 ชุด ได้แก่ คำวิเศษณ์ คำกริยา และคำวิเศษณ์ร่วมกับคำกริยา ส่วน Tran และคณะ [12] ได้ศึกษาการแยกคำแสดงความคิดเห็นที่มีประสิทธิภาพสำหรับการจำแนกรีวิวอาหารจากเว็บไซต์ Amazon ซึ่งใช้ชุดข้อมูล 2 ชุด ได้แก่ คำคุณศัพท์ และ คำคุณศัพท์กับคำวิเศษณ์ และ Neshan และ Akbari [13] ได้ศึกษาการผสมผสานระหว่างการเรียนรู้ของเครื่องและเทคนิคจากพจนานุกรมสำหรับการวิเคราะห์ความเชื่อมั่น จากเว็บไซต์ Amazon และ IMDB ประเภทคำที่เลือกใช้ ได้แก่ คำคุณศัพท์ คำกริยาวิเศษณ์ และคำกริยา

ดังนั้น ผู้วิจัยจึงมีแนวคิดที่จะพัฒนาแบบจำลองในการทำเหมืองความคิดเห็นของประชาชน ที่มีต่อออกทิสติกโดยนำความคิดเห็นของประชาชนจากโซเชียลมีเดีย ว่ามีความคิดเห็นไปในเชิงบวกหรือเชิงลบ โดยเปรียบเทียบเทคนิคที่ใช้ในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็น ได้แก่ เทคนิคป่าสุ่ม (Random Forest) เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) และเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) แล้ววัดประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออกทิสติก

1.2.2 เพื่อสร้างและเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนที่มีต่อออกทิสติกด้วยเทคนิคเหมืองความคิดเห็น

1.3 ขอบเขตของการวิจัย

1.3.1 เก็บรวบรวมความคิดเห็นของประชาชนต่อออทิสติก จากเว็บไซต์พันทิป เฟซบุ๊ก และ ทวิตเตอร์ จำนวน 1,766 ความคิดเห็น ตั้งแต่เดือนมกราคม 2556 ถึง เดือนกุมภาพันธ์ 2565

1.3.2 การคัดเลือกคุณลักษณะด้วยถ่วงค่า

1.3.3 สร้างแบบจำลองเพื่อจำแนกความคิดเห็นของประชาชนที่มีต่อออทิสติก ได้แก่ เทคนิค ป่าสุ่ม (Random Forest) เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคเพื่อนบ้านใกล้ ที่สุด (K-Nearest Neighbors)

1.3.4 เปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้ทราบกระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของ ประชาชนต่อออทิสติก

1.4.2 ได้แบบจำลองเพื่อจำแนกความรู้สึกของประชาชนต่อออทิสติกที่มีประสิทธิภาพ

1.5 นิยามศัพท์เฉพาะ

1.5.1 แบบจำลอง คือ ต้นแบบที่ใช้ในการจำแนกความคิดเห็นของประชาชนต่อออทิสติก

1.5.2 ออทิสติก หมายถึง บุคคลที่มีข้อจำกัดในการปฏิบัติกิจกรรมในชีวิตประจำวันหรือการเข้าไปมีส่วนร่วมในกิจกรรมทางสังคม ซึ่งเป็นผลมาจากความบกพร่องทางพัฒนาการด้านสังคม ภาษา และการสื่อความหมาย พฤติกรรมและอารมณ์ โดยมีสาเหตุมาจากความผิดปกติของสมองและความผิดปกติที่แสดงก่อนอายุสองปีครึ่ง

1.5.3 ความคิดเห็นของประชาชน คือ ข้อเท็จจริงของการแสดงความคิดเห็นของประชาชนต่อ ออทิสติก

1.5.4 เหมืองความคิดเห็น คือ กระบวนการในการจำแนกความคิดเห็นของประชาชนต่อ ออทิสติกที่เป็นเชิงบวกหรือเชิงลบ

1.5.5 คุณลักษณะ คือ ตัวแปรที่นำมาจากคำที่มีการตัดคำแล้ว

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวกับความคิดเห็นของประชาชนต่อออทิสติก ความรู้พื้นฐานเกี่ยวกับออทิสติก การทำเหมืองความคิดเห็น เทคนิคที่ใช้ในการวิจัย และการวัดประสิทธิภาพของแบบจำลอง รวมทั้งงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ความคิดเห็นของประชาชนต่อออทิสติก

ความคิดเห็นของประชาชนต่อออทิสติก คือ ข้อความแสดงความคิดเห็นของประชาชนที่มีต่อออทิสติก ซึ่งหมายถึงบุคคลที่มีข้อจำกัดในการปฏิบัติกิจกรรมในชีวิตประจำวันหรือการเข้าไปมีส่วนร่วมในกิจกรรมทางสังคม อันเป็นผลมาจากความบกพร่องทางพัฒนาการหลากหลายด้าน เนื่องจากความผิดปกติของสมอง โดยข้อความเหล่านี้ปรากฏอยู่บนสื่อสังคมออนไลน์ที่เปิดเป็นสาธารณะสามารถให้ประชาชนมาแสดงความคิดเห็นได้ มีทั้งการแสดงความคิดเห็นในเชิงบวกและเชิงลบ ซึ่งส่งผลต่อความรู้สึกของบุคคลที่เป็นออทิสติกหรือคนในครอบครัวที่มีบุคคลออทิสติกเป็นสมาชิก หากเป็นความคิดเห็นในเชิงลบอาจจะทำให้บุคคลออทิสติกเกิดการบั่นทอนจิตใจและส่งผลให้มีปัญหาในการอยู่ร่วมกับผู้อื่นในสังคม ในทางตรงกันข้ามความคิดเห็นเชิงบวกอาจจะส่งผลให้บุคคลออทิสติกมีกำลังใจในการใช้ชีวิตไม่รู้สึกว่าตัวเองแตกต่างจากคนอื่น ๆ สามารถที่จะใช้ชีวิตอย่างมีความสุขและกล้าแสดงออกมากยิ่งขึ้น

2.1.2 ความรู้พื้นฐานเกี่ยวกับออทิสติก

คนในสังคมไทยจำนวนไม่น้อยที่ยังไม่เข้าใจในความพิเศษของผู้พิการทางออทิสติก บางคนคิดว่าเด็กพิเศษ หรือ ผู้พิการทางออทิสติกเป็นคนไม่มีความสามารถ แต่ในความเป็นจริงแล้ว กลุ่มคนเหล่านี้เป็นคนที่มีความสามารถมีศักยภาพในเรื่องใดเรื่องหนึ่ง สามารถช่วยเหลือตัวเองได้ โดยสาเหตุของออทิสติก ยังไม่มีข้อสรุปที่ชัดเจนว่าเกิดจากสาเหตุใด แต่มีข้อสันนิษฐานว่ามีปัจจัยหลายประการที่ส่งผลให้การทำงานของระบบประสาทบางส่วนผิดปกติ และบ่งชี้ไปในทางความผิดปกติทางพันธุกรรม กล่าวคือหากมีลูกคนหนึ่งเป็นออทิสติกแล้ว ลูกอีกคนจะมีโอกาสเป็นออทิสติกได้ถึง 3% ปัจจุบันมีการยอมรับถึงสาเหตุสำคัญของโรคที่เกิดจากความบกพร่องของ Central nervous system ซึ่งอาจจะเกิดจากเซลล์สมองที่ผิดปกติและความไม่สมดุลของสารเคมีของระบบประสาท สำหรับความต้องการตามความจำเป็นพื้นฐานของบุคคลออทิสติก ได้แก่ การฟื้นฟูสมรรถภาพ เช่น การแก้ปัญหา

การพูด การปรับพฤติกรรม การจัดกิจกรรมบำบัดทั้งด้านร่างกายและจิตใจ, การบริการการศึกษา โดยพิจารณาจากความต้องการความจำเป็นของแต่ละบุคคลเป็นหลัก โดยมีการใช้เทคนิคการสอน เฉพาะด้านที่เหมาะสม สามารถยืดหยุ่นให้สอดคล้องกับความต้องการและความสามารถของเด็กแต่ละคนได้, การบริการสังคม บุคคลออทิสติกมีขีดจำกัดของความแตกต่างกันในแต่ละด้าน ทำให้มีข้อจำกัด เรื่องโอกาสทางสังคมในการเข้าร่วมกิจกรรมกับคนทั่วไป จึงควรจัดให้บุคคลเหล่านี้มีส่วนร่วมใน กิจกรรมทางสังคมอย่างสม่ำเสมอ และการบริการอาชีพ บุคคลออทิสติกมีความหลากหลาย และมี ข้อจำกัดของแต่ละคนแตกต่างกันออกไป จึงต้องจัดให้เหมาะสมกับศักยภาพและความถนัดของแต่ละ บุคคลเป็นหลัก [14]

การช่วยเหลือออทิสติกมีหลายวิธี เนื่องจากความบกพร่องของโรคนี้อาจมีหลายด้านและไม่มี วิธีที่ดีที่สุดที่จะสามารถช่วยเหลือออทิสติกได้ ดังนั้นคนในครอบครัวจึงต้องให้ความช่วยเหลือ ออทิสติกอย่างรอบด้าน โดยเน้นที่ทักษะทางสังคม ทักษะการสื่อสาร และพฤติกรรมที่เหมาะสม กล่าวคือ ด้านทักษะทางสังคม นับเป็นสิ่งสำคัญที่จะทำให้เด็กออทิสติกอยู่ร่วมกับผู้อื่นและสามารถ พัฒนาตนเองจนใกล้เคียงเด็กคนอื่น ๆ ได้ เช่น ให้เด็กเริ่มหัดเล่นกับคนในครอบครัวก่อน แล้วค่อย ขยับไปเล่นกับเด็กในวัยเดียวกัน หรือ หาเพื่อนมาเล่นกับเด็กออทิสติก โดยมีคนในครอบครัวนั่ง ประคบอยู่ด้วย พร้อมกับยกตัวอย่างให้เด็กออทิสติกเล่นและตอบสนองต่อเพื่อนอย่างเหมาะสม เป็นต้น ด้านทักษะการสื่อสาร จำเป็นต้องเพิ่มทักษะการสื่อสาร ในกรณีที่เด็กออทิสติกยังไม่มีภาษา พูด คนในครอบครัวควรพูดคุยกับเด็ก โดยบอกเป็นคำศัพท์ทั่วไปในชีวิตประจำวัน เช่น นั่ง นอน เดิน แต่ในกรณีที่สามารถพูดสื่อสารได้แล้ว คนในครอบครัวควรสอนการจัดหมวดหมู่ของคำ เพิ่มความยาก เป็นคำที่เกี่ยวข้องรูปร่าง ขนาด และความรู้สึก เป็นต้น สำหรับการฝึกพูดไม่ใช่เพื่อให้เด็กออทิสติกมี คำพูดเท่านั้น แต่เพื่อให้สามารถสื่อสารได้อย่างถูกต้องตามกาลเทศะ ดังนั้น คนในครอบครัวต้องสอน ทักษะทางสังคมควบคู่ไปด้วย และด้านพฤติกรรมที่เหมาะสม ควรเพิ่มพฤติกรรมที่เหมาะสมและลด พฤติกรรมไม่เหมาะสม โดยใช้เทคนิคในการปรับพฤติกรรมเด็กออทิสติกที่ใช้บ่อยและใช้ได้ผล เช่น การจับมือทำเพื่อให้เด็กออทิสติกทำกิจกรรมได้ง่ายขึ้น และการจับให้หยุดเมื่อมีพฤติกรรมที่ไม่ เหมาะสม การให้รางวัลเพื่อให้ทำพฤติกรรมที่เหมาะสมมากขึ้น หรือให้ของที่ชอบ เช่น การให้ขนม ของเล่น เป็นต้น แต่ให้ระวังเด็กออทิสติกจะติดเงื่อนไขในการได้รับรางวัล ครอบครัวควรสลับการให้ รางวัลจากสิ่งของเป็นคำชมหรือทำที่ชื่นชมแทน เช่น ยกนิ้วให้พร้อมแสดงท่าทางชื่นชมไปกับเด็ก เนื่องจากเด็กออทิสติกแต่ละคนมีปัญหาในการปรับตัวต่อสิ่งใหม่ ๆ มีปัญหาติดกับสิ่งเดิม ดังนั้น การปรับพฤติกรรมเป็นเรื่องที่ต้องใช้เวลาและความอดทน [15]

ภาวะออทิสติกเป็นกลุ่มอาการที่ไม่สามารถรักษาให้หายขาดได้ เป้าหมายสำคัญของการรักษา คือ การนำเด็กออกจากโลกของตัวเองสู่สังคมปกติภายนอก และการกระตุ้นพัฒนาการที่บกพร่องให้ใกล้เคียงกับเด็กปกติให้ได้มากที่สุด ดังนั้น การรักษาภาวะออทิสติก จึงต้องใช้หลายวิธีรวมกัน ดังนี้

1. การส่งเสริมพัฒนาการ
2. พฤติกรรมบำบัด
3. การฝึกพูดหรือแก้ไขการพูด
4. การรักษาด้วยกิจกรรมบำบัด ในกรณีที่มีผู้ป่วยมีปัญหาการรับสัมผัสที่ผิดปกติซึ่งส่งผลให้เกิดพฤติกรรมที่ไม่พึงประสงค์
5. การรักษาด้วยยา มีความจำเป็นในกรณีที่ผู้ป่วยมีปัญหาพฤติกรรมรุนแรง และ/หรือมีพฤติกรรมทำร้ายตนเอง อารมณ์รุนแรง

นอกจากนี้หลักการรักษาผู้ป่วยออทิสติกที่สำคัญอีกประการหนึ่ง คือ การส่งเสริมพัฒนาการและการปรับพฤติกรรมให้ผู้ป่วยจะต้องมีความถี่และความเข้มข้นของการฝึกมากพอที่จะส่งเสริมให้เกิดพัฒนาการที่ปกติและพฤติกรรมที่เหมาะสม มีการศึกษาวิจัยในต่างประเทศที่แนะนำให้มีความเข้มข้นของการฝึกอยู่ที่ประมาณ 6-8 ชั่วโมง ในหนึ่งวัน 5-7 วัน ในสัปดาห์ หรืออย่างน้อยต้องมีความถี่ 15 ชั่วโมง ใน 1 สัปดาห์ และฝึกอย่างต่อเนื่องเป็นเวลาอย่างน้อย 2 ปี ควรมีการรักษาอยู่ที่โรงพยาบาลหรือผู้ให้บริการทางสาธารณสุขเท่านั้น และให้พ่อแม่รวมทั้งครอบครัวมีส่วนร่วมในการส่งเสริมพัฒนาการและการปรับพฤติกรรมผู้ป่วยต่อที่บ้านอย่างสม่ำเสมอก็จะช่วยให้ผลการรักษาดีขึ้น [16] ดังตัวอย่างงานวิจัยเกี่ยวกับการบำบัดออทิสติกต่อไปนี้

อัครส ทองเพชร และคณะ [17] ได้ศึกษาการบำบัดทางการพยาบาลระยะแรกต่อความรุนแรงของอาการเด็กออทิสติกวัยก่อนเรียน พบว่า กลุ่มทดลอง หลังได้รับการบำบัดทางการพยาบาลระยะแรก มีความรุนแรงของอาการออทิสติกลดลง มากกว่ากลุ่มควบคุม

นรุตม์ แพงพรมมา, จินตนา ยูนิพันธ์ และ รัชนิกร อุปเสน [18] ได้ศึกษาผลของพฤติกรรมบำบัดโดยครอบครัวมีส่วนร่วมต่อพฤติกรรมก้าวร้าวของเด็กออทิสติก พบว่า พฤติกรรมก้าวร้าวของเด็กออทิสติก หลังได้รับพฤติกรรมบำบัดโดยครอบครัวมีส่วนร่วมลดลงกว่าก่อนได้รับโปรแกรม และพฤติกรรมก้าวร้าวของเด็กออทิสติก หลังการทดลองในกลุ่มที่ได้รับพฤติกรรมบำบัดโดยครอบครัวมีส่วนร่วมลดลงกว่ากลุ่มที่ได้รับการพยาบาลตามปกติ

2.1.3 การทำเหมืองความคิดเห็น

การทำเหมืองความคิดเห็น (Opinion Mining) เป็นกระบวนการในการวิเคราะห์ความคิดเห็นจากข้อความ เพื่อให้ได้มาซึ่งความรู้ที่แฝงอยู่ในความคิดเห็นนั้น ๆ มีการผสมผสานเทคนิคทางภาษา จากข้อความแสดงความคิดเห็นที่เป็นถ้อยคำอันแสดงถึงเหตุผลและความคิดเห็นส่วนตัวของผู้สร้างข้อความ สะท้อนให้เห็นทัศนคติหรือการรับรู้เกี่ยวกับเรื่องราวและเหตุการณ์ของบุคคล จากนั้นพัฒนาให้เป็นระบบแล้วพิจารณาเซตข้อความที่มีความคิดเห็นหรือทัศนคติเกี่ยวกับสิ่งใดสิ่งหนึ่งประกอบอยู่ด้วย โดยการทำเหมืองความคิดเห็นมีวัตถุประสงค์เพื่อดึงคุณลักษณะและส่วนประกอบของสิ่งนั้น ๆ ที่ได้ถูกแสดงความคิดเห็นในรูปแบบของข้อความตัวอักษร เพื่อพิจารณาว่าความคิดเห็นนั้นเป็นความคิดเห็นเชิงบวกหรือเชิงลบ โดยมีขั้นตอนในการทำเหมืองความคิดเห็น 4 ขั้นตอน [9] [19] สำคัญดังนี้

1) การเตรียมข้อมูล (Preprocessing) เป็นกระบวนการระบุถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองความคิดเห็น รวมถึงการนำความคิดเห็นที่ต้องการออกมาจากฐานข้อมูล เพื่อสร้างเป็นชุดข้อมูลในการพิจารณาเบื้องต้นตามขอบเขตที่ต้องการทำการศึกษ ทำให้ข้อมูลความคิดเห็นที่จะนำมาใช้วิเคราะห์มีคุณภาพและมีความถูกต้อง กล่าวคือ เมื่อได้กลุ่มข้อมูลความคิดเห็นที่ต้องศึกษาแล้ว ข้อมูลเหล่านี้จะถูกนำมากรองเพื่อเลือกข้อมูลที่ตรงประเด็นและนำความคิดเห็นที่ไม่ถูกต้องออกหรือเป็นขั้นตอนที่อาจจะต้องแก้ไขก่อนนำไปวิเคราะห์และประมวลผลต่อ

2) การจัดทำดัชนีข้อมูล (Indexing) เป็นกระบวนการในการตัดคำจากชุดข้อมูลที่ได้เตรียมไว้ เพื่อนำมาทำเป็นดัชนีคำและกำหนดความคิดเห็นของข้อความเป็นเชิงบวกหรือเชิงลบ รวมถึงการคัดเลือกคุณลักษณะที่เหมาะสมต่อการจำแนก

3) การสร้างแบบจำลอง เป็นกระบวนการในการค้นหารูปแบบของข้อมูลโดยใช้เทคนิคในการจำแนก (Classification) เพื่อการพยากรณ์

4) การประเมินผล (Interpretation/Evaluation) เป็นขั้นตอนในการวัดประสิทธิภาพของแบบจำลอง เพื่อค้นหาเทคนิคที่มีประสิทธิภาพในการจำแนก

2.1.4 เทคนิคที่ใช้ในงานวิจัย

เทคนิคที่ใช้ในงานวิจัยนี้ เป็นเทคนิคของการทำเหมืองข้อความหรือนำไปสร้างแบบจำลองเพื่อการจำแนกประเภทของคำ โดยใช้โปรแกรม Weka มาเป็นเครื่องมือในการศึกษา จาก

การศึกษางานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อความพบว่า งานวิจัยส่วนใหญ่ใช้เทคนิคการทำเหมืองข้อความ จำนวน 5 เทคนิค ได้แก่ เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิคป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ดังนี้

2.1.4.1 เทคนิคนาอิวเบย์

เทคนิคนาอิวเบย์ (Naïve Bayes: NB) [20] เป็นเทคนิควิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตาม เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ โดยหาว่าสมมติฐานใดน่าจะมีค่าถูกต้องมากที่สุด ซึ่งจะสามารถบ่งบอกได้ถึงความน่าจะเป็นของข้อมูลชุดหนึ่งที่อยู่ในหมวดหมู่ของข้อมูลนั้น สำหรับเทคนิคนี้ใช้ในการจัดการกับปัญหาประเภทจำแนกข้อมูล ผลลัพธ์ของเทคนิคนี้สามารถนำไปทำนายและช่วยอธิบายความหมายของข้อมูล โดยปกติการทำนายผลของ Naïve-Bayes จะถูกต้อง เมื่อตัวแปรอิสระทั้งหมดเป็นอิสระต่อกันโดยไม่ขึ้นกับตัวแปรอิสระตัวใดตัวหนึ่ง แต่เทคนิคนี้มีข้อจำกัดว่า ค่าตัวแปรอิสระหรือตัวแปรตามที่เป็นค่าต่อเนื่องจะถูกแบ่งออกเป็นช่วง ซึ่งการแบ่งช่วงนั้นถ้ามีการแบ่งที่ไม่เหมาะสม ก็จะทำให้ประสิทธิภาพของแบบจำลองที่ได้มีคุณภาพไม่ดีตามไปด้วย [21] ดังสมการที่ 2.1

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad 2.1$$

โดย

X		แทนข้อมูลการแจกแจงความน่าจะเป็น
H	คือ	คลาสของข้อมูล
$P(H)$	คือ	ความน่าจะเป็นก่อนหน้าของ H
$P(X)$	คือ	ความน่าจะเป็นก่อนหน้าของ X
$P(H X)$	คือ	ความน่าจะเป็นของ H เมื่อรู้ว่าเหตุการณ์ X เกิดขึ้น
$P(X H)$	คือ	ความน่าจะเป็นของ X เมื่อรู้ว่าเหตุการณ์ H เกิดขึ้น

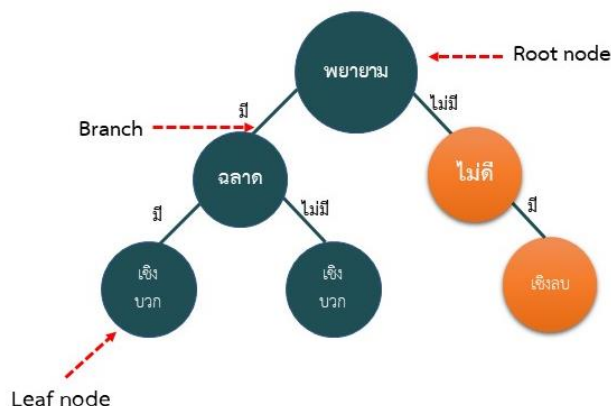
การจำแนกประเภทข้อมูลด้วยเทคนิค NB ถือเป็นวิธีที่มีประสิทธิภาพวิธีหนึ่งที่ใช้ในการจำแนกหมวดหมู่ข้อความ (Text Classification) ได้ดี การทำงานไม่ซับซ้อนเหมือนวิธีการอื่น ๆ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและง่ายต่อการประมวลผล [22]

ธีรยุทธ คุณสุข และ จารีย์ ทองคำ [22] ได้ศึกษากระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นลูกค้า โดยนำเทคนิค 3 เทคนิคมาคัดเลือกคุณลักษณะของข้อความ ได้แก่ Chi-Square Information Gain และ Information Gain Ratio ส่วนในการวัดประสิทธิภาพของการคัดเลือกคุณลักษณะ ใช้เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค DT C4.5 มาวิเคราะห์ข้อมูลที่เก็บรวบรวมจากเว็บไซต์ wongnai.com จำนวนทั้งหมด 4,487 ข้อความ พบว่า เทคนิคการคัดเลือกคุณลักษณะแบบ Information Gain ร่วมกับเทคนิค NB ให้ผลที่ดีที่สุดในการจำแนกความคิดเห็น

2.1.4.2 เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5

เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) [23] เป็นเทคนิคที่นำหลักการของต้นไม้ตัดสินใจมาทำการจำแนกข้อมูลเพื่อการพยากรณ์ ต้นไม้ตัดสินใจประกอบไปด้วย โหนด (Node) ซึ่งทำหน้าที่แสดงคุณลักษณะที่ใช้สำหรับการทดสอบข้อมูล กิ่ง (Branch) เป็นส่วนที่จะแสดงคุณสมบัติในโหนดที่แตกออกมา และใบ (Leaf) จะแสดงกลุ่มหรือคลาสที่กำหนดไว้ โดยการหาความสัมพันธ์ของแต่ละโหนดที่แสดงคุณลักษณะ (Attribute) จะใช้ค่า Information Gain เพื่อหาความสัมพันธ์ของคุณลักษณะในแต่ละโหนด ซึ่งการใช้ค่า Information Gain จะช่วยลดจำนวนครั้งที่ใช้ในการทดสอบ และทำให้ผลที่ได้จากเทคนิคนี้ไม่มีความซับซ้อนมากจนเกินไป [24] เทคนิค DT C4.5 เป็นเทคนิคที่พัฒนามาจากเทคนิค ID3 โดย Quinlan ที่พัฒนาขึ้นมาเพื่อแก้ปัญหาจุดอ่อนของ ID3 ในด้านความอคติ ดังแสดงในภาพ 2.1

พหุบัณฑิต ชีวะ



ภาพที่ 2.1 การทำงานของเทคนิคต้นไม้ตัดสินใจแบบ ซี4.5

จากที่ค่าแบ่งแยกมีการกระจายตัว บ่งบอกถึงการกระจายของข้อมูล ซึ่งวิธีแก้ปัญหาค่าความเอนเอียงจะต้องนำค่าสารสนเทศในการแบ่งแยกหารด้วยค่ามาตรฐานเกน จะได้ค่า Gain Ratio เพื่อเป็นตัวเลือกใช้คุณลักษณะที่จะนำมาเป็นโหนดในลำดับต่อไป ดังสมการที่ 2.2

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad 2.2$$

เมื่อนำข้อมูลมาคำนวณหาค่า Gain Ratio แล้วเลือกค่าสูงสุดมาเป็นโหนดเริ่มต้นและล้างโหนดในระดับต่อไป โดยใช้คุณลักษณะที่มีค่ารองลงมาต่อ ๆ ไป จนกระทั่งข้อมูลมีการกระจายตัวน้อยที่สุดหรือข้อมูลมีค่าเท่ากัน [25] เกณฑ์ที่ช่วยในการตัดสินใจเลือกโหนดเริ่มต้น คือ ทดลองเลือกคุณลักษณะแต่ละตัวมาทำหน้าที่เป็นโหนดเริ่มต้น แล้วหาค่า Gain Ratio ซึ่งเป็นค่าที่บอกว่าคุณลักษณะใดจะทำหน้าที่เป็นโหนดเริ่มต้น และสามารถจำแนกข้อมูลได้มากน้อยเพียงใด หน่วยของการวัดเป็น bits ซึ่งได้มาจากการคำนวณโดยใช้ทฤษฎีสารสนเทศ (Information Theory) โดยที่ค่าสารสนเทศของข้อมูลขึ้นอยู่กับความน่าจะเป็นของข้อมูล และจะเลือกคุณลักษณะที่ให้ค่า Gain Ratio มากที่สุดเป็นโหนดเริ่มต้น [26] ดังสมการที่ 2.3

$$Gain(A) = info(D) - info_A(A) \quad 2.3$$

ในกรณีที่มีข้อมูลในชุดที่นำมาทดสอบมีการกระจายตัวมากหรือไม่เป็นกลุ่มเดียวกัน ทำให้เมื่อมีการจำแนกแล้วเกิดความลำเอียงไปในกลุ่มที่มีข้อมูลอยู่เป็นจำนวนมาก จึงได้เพิ่ม

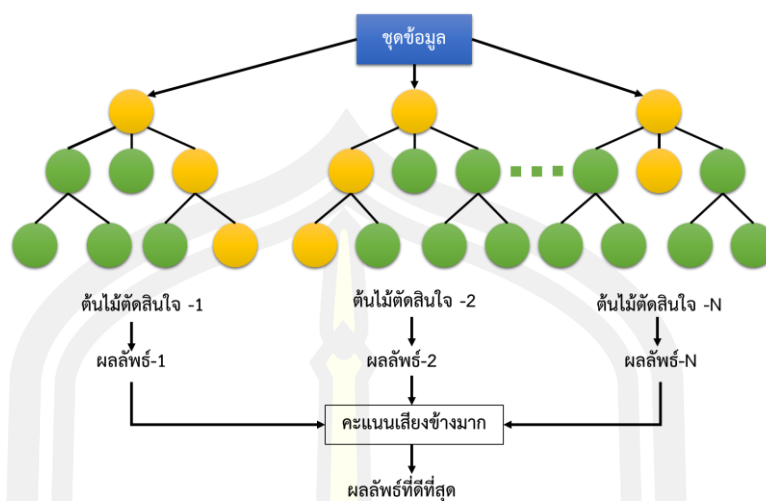
ค่าการแบ่งแยกเพื่อใช้ในการคำนวณและลดปัญหาการเกิดการเอนเอียงในการจำแนก [25]
 ดังสมการที่ 2.4

$$SplitInfo_A(D) = - \sum_{i=1}^n \frac{|t_i|}{|T|} \log_2 \frac{|t_i|}{|T|} \quad 2.4$$

อุกฤษฏ์ ศรีสุข และ จาริ ทองคำ [27] ได้ศึกษาการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรค โดยข้อมูลทั้งหมดถูกรวบรวมมาจากฐานข้อมูล UCI จำนวนทั้งหมด 3 ชุดข้อมูล โดยใช้เทคนิค 5 เทคนิค ได้แก่ เทคนิค DT (C4.5), เทคนิค NB, เทคนิค K-NN, เทคนิค RF และเทคนิค Deep Learning (DL) มาสร้างแบบจำลองเพื่อพยากรณ์การเกิดโรค ได้แก่ โรคมะเร็งเต้านม โรคเบาหวาน และโรคไฮโปไทรอยด์ พบว่า เทคนิค Decision Tree C4.5 เป็นเทคนิคที่ดีที่สุดในการสร้างแบบจำลองในการพยากรณ์โรคไฮโปไทรอยด์

2.1.4.3 เทคนิคป่าสุ่ม

เทคนิคป่าสุ่ม (Random Forest: RF) [28] เป็นเทคนิคที่สุ่มเลือกคุณลักษณะออกมาจากชุดข้อมูลหลาย ๆ ชุด จากนั้นนำชุดของคุณลักษณะเหล่านั้นมาสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจหลาย ๆ ต้น ลักษณะของต้นไม้ที่อยู่ภายในป่าของเทคนิคป่าสุ่มจะถูกควบคุมด้วย 3 ปัจจัย ได้แก่ ต้นไม้แต่ละต้นจะถูกสอน (Train) โดยใช้เซตย่อยจากข้อมูลตัวอย่าง ต่อมาเมื่อต้นไม้โตขึ้นจะสามารถค้นหาโหนด (Node) แต่ละโหนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้โดยใช้การสุ่มเลือกคุณลักษณะจาก N คุณลักษณะ และสุดท้ายจะปล่อยให้ต้นไม้โตขึ้นไปเรื่อย ๆ จนได้ผลลัพธ์ที่ดีที่สุด โดยจะไม่ตัดต้นไม้แต่ละต้นออก หลังจากนั้นทำการให้คะแนนโดยการคัดเลือก (Vote) จากต้นไม้ในป่าสุ่ม หากต้นไม้ต้นใดได้คะแนนสูงสุดก็นำต้นไม้นั้นออกมาสร้างเป็นโมเดล [21] ดังแสดงในภาพ 2.2



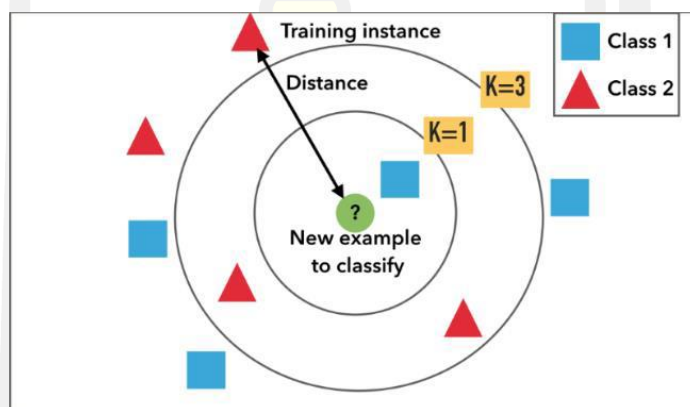
ภาพที่ 2.2 การทำงานของเทคนิคป่าสุ่ม

การใช้ต้นไม้หลายต้นตัดสินใจมีขั้นตอนคือ ให้จำนวนข้อมูลที่ถูกสอน เท่ากับ n จำนวน ส่วนตัวแปร m คือ จำนวนตัวแปรเพื่อใช้ตัดสินใจในโหนดของต้นไม้ แล้วเลือกจำนวนข้อมูลเพื่อมาเป็นชุดฝึก หาค่าความผิดพลาดจากการคาดเดากลุ่ม โดยเริ่มจากการสร้าง bootstrap จากกลุ่มตัวอย่าง n และสร้างต้นไม้จากกลุ่มตัวอย่าง bootstrap จากนั้นเมื่อแต่ละโหนดของต้นไม้มีการเจริญเติบโต จะสุ่มเลือกตัวแปร m เพื่อใช้ในการตัดสินใจในการจำแนกหาค่าความผิดพลาด และค่าที่ดีที่สุดออกมา แล้วจำแนกจุด x เพื่อรวบรวมคะแนนโหวตจากต้นไม้ในป่าสุ่ม จากนั้นใช้การลงคะแนนเสียงส่วนใหญ่ในการตัดสินใจจากแต่ละกลุ่ม อย่างไรก็ตามจะพบว่า เทคนิค RF สามารถให้ความแม่นยำสูง และจัดการข้อมูลได้เป็นจำนวนมาก อีกทั้งยังสามารถตรวจสอบหาค่าความผิดพลาดได้อย่างมีประสิทธิภาพ แต่ก็ยังพบว่ามีข้อเสีย คือ ใช้เวลาในการสร้างโมเดลค่อนข้างนาน และไม่เหมาะที่จะใช้กับข้อมูลที่มีความซับซ้อน [29]

สุภัทสรุ สมนเจตนา และ จาริ ทองคำ [9] ได้ศึกษาเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลอง เพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร โดยใช้เทคนิคคริปเปอร์ (Ripper), เทคนิค DT C4.5 เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค RF แสดงประสิทธิภาพของเทคนิคด้วยค่า Precision, Recall และ F-Measure พบว่า เทคนิค RF เป็นเทคนิคที่ดีที่สุด

2.1.4.4 เทคนิคเพื่อนบ้านใกล้ที่สุด

เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: K-NN) [30] เป็นเทคนิคจำแนกประเภทข้อมูลที่ใช้วิธีการหาระยะห่างระหว่างคุณลักษณะของแต่ละข้อมูล โดยสามารถคำนวณจากสมการระยะทางยูคลิดีเนียน (Euclidean distance) จัดลำดับของระยะห่างจากน้อยไปมากและเลือกชุดข้อมูลที่น้อยที่สุด ตามจำนวน k ที่กำหนดไว้ และกำหนดให้คำตอบของข้อมูลที่ต้องการจำแนกกลุ่มที่มีจำนวนมากที่สุดอยู่ในกลุ่มของชุดข้อมูล k ตัวแรก [31] โดยเริ่มจากขั้นตอนกำหนดจำนวนเพื่อนบ้าน k ซึ่งนิยมกำหนดให้เป็นเลขคี่ จากนั้นคำนวณระยะห่างของข้อมูลที่ต้องการพิจารณากับชุดสอน ซึ่งวิธีนี้จะเหมาะสำหรับข้อมูลแบบตัวเลข แต่ตัวแปรที่เป็นค่าแบบไม่ต่อเนื่องนั้นก็ยังสามารถทำได้แต่จำเป็นต้องมีการจัดการแบบพิเศษเพิ่มขึ้น [32] ดังแสดงในภาพที่ 2.1



ภาพที่ 2.3 การจัดกลุ่มของเทคนิค K-NN

แหล่งที่มา https://www.researchgate.net/figure/Example-of-k-NN-classification-The-test-sample-inside-circle-should-be-classified_fig4_333457161

ในการใช้งานเทคนิค K-NN จะต้องระบุค่าตัวเลขจำนวนเต็มบวกให้กับ k โดยที่ค่านี้จะเป็นตัวบอกจำนวนของกลุ่มเป้าหมายที่จะต้องหาสำหรับการทำนายกรณีใหม่ ได้แก่ 1-NN, 2-NN, 3-NN,...K-NN เช่น 3-NN หมายถึง ขั้นตอนวิธีการนี้จะมีการค้นหา 3 กรณี แล้วพิจารณาว่ากรณีใดมีลักษณะใกล้เคียงกับกรณีใหม่ โดยใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงกันมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจอยู่ใกล้กับข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบที่อยู่ใกล้ที่สุด ซึ่งเป็นวิธีการที่ไม่ซับซ้อนและเข้าใจง่ายที่สุดในการจำแนกประเภทข้อมูล [26]

สูตรระยะทางยูคลิเดียน (Euclidean distance) [33] ดังสมการที่ 2.5

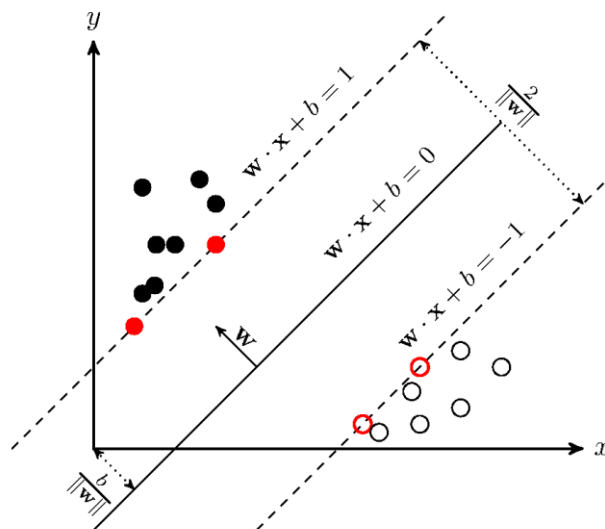
$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad 2.5$$

วสวัตดี อินทร์แปลง และ จาริ ทองคำ [19] ได้วิเคราะห์ความคิดเห็นต่อเกมมือถือพัจจีด้วยเหมืองข้อความ เก็บรวบรวมข้อความที่เป็นความคิดเห็นจาก Google Play จำนวน 3,789 ความคิดเห็น เพื่อใช้แยกคุณลักษณะ โดยเลือกใช้คำวิเศษณ์และคำสแลงบางคำที่มีความหมายอย่างคำวิเศษณ์ เพื่อแยกคุณลักษณะเชิงบวกและเชิงลบ โดยใช้เทคนิค NB, เทคนิค SVM, เทคนิค K-NN, เทคนิค CT C4.5 และเทคนิค RF พบว่า เทคนิค K-NN ให้ผลดีที่สุด

2.1.4.5 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน

เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) [34] เป็นวิธีการพื้นฐานในการเรียนรู้ของเครื่องคอมพิวเตอร์ เทคนิคนี้ใช้แมปฟังก์ชัน (Mapping Function) โดยการแปลงข้อมูลที่มีอยู่ (Input Space) ไปเป็นข้อมูลใหม่ที่เรียกว่า (Feature Space) ในลักษณะเชิงเส้น แล้วจึงนำเอาระนาบการตัดสินใจมาแบ่งกลุ่มข้อมูลออกเป็นสองส่วน และใช้หลักการสร้างเส้นแบ่งกึ่งกลางระหว่างกลุ่มให้มีระยะห่างระหว่างกลุ่มมากที่สุด โดยที่เทคนิค SVM นี้มีประสิทธิภาพสูงในการจำแนกข้อมูล [35] ดังแสดงในภาพ 2.2

พหุ ประทีป ชีวะ



ภาพที่ 2.4 การวางตัวของข้อมูลในลักษณะเชิงเส้น

แหล่งที่มา <http://blog.pengyifan.com/wp-content/uploads/2013/09/svm>

อย่างไรก็ตามปัญหาที่พบในการแบ่งกลุ่มส่วนมากมักพบลักษณะแบบไม่เชิงเส้นตรง ดังนั้น การแปลงข้อมูลจากข้อมูลแบบไม่เชิงเส้นตรงไปสู่ข้อมูลที่เป็นเชิงเส้นตรงจึงต้องใช้ฟังก์ชันเคอร์เนล (Kernel Function) มาช่วยแก้ปัญหาดังกล่าว [36] ในเทคนิค SVM จะเรียกตัวแปรการตัดสินใจว่า คุณสมบัตินี้ และตัวแปรที่เปลี่ยนแปลงจะใช้กำหนดระนาบหลายมิติเรียกว่า คุณลักษณะ (Feature) ส่วนการเลือกที่มีความเหมาะสมที่สุดจะเรียกว่า การคัดเลือกคุณลักษณะ (Feature Selection) จำนวนเซตของคุณลักษณะที่ใช้อธิบายในกรณีหนึ่ง ๆ เรียกว่า เวกเตอร์ ดังนั้นจุดมุ่งหมายของตัวแบบ SVM คือการได้ประโยชน์สูงสุดจากระนาบหลายมิติที่แบ่งแยกกลุ่มของเวกเตอร์ในกรณีนี้ด้วย การคัดเลือกคุณลักษณะเหมาะสมสำหรับข้อมูลที่มีมิติสูง โดยกำหนดให้ $(x_i, y_i), \dots, (x_n, y_n)$ เมื่อ $x \in R^m$, $y \in \{+1, -1\}$ เป็นตัวอย่างที่ใช้สำหรับการสอน n คือ จำนวนข้อมูลตัวอย่าง m คือ จำนวนมิติข้อมูลเข้า และ y คือ ผลลัพธ์ มีค่า $+1$ หรือ -1 [29]

สำหรับปัญหาเชิงเส้น มิติข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยระนาบตัดสินใจ ซึ่งคำนวณได้ ดังสมการที่ 2.6

$$(w \times x) + b = 0 \quad 2.6$$

เมื่อ w คือ ค่าน้ำหนัก และ b คือ ค่า bias สมการที่ใช้สำหรับจำแนกประเภทของข้อมูล ดังสมการที่ 2.7

$$(w \times x) + b > 0$$

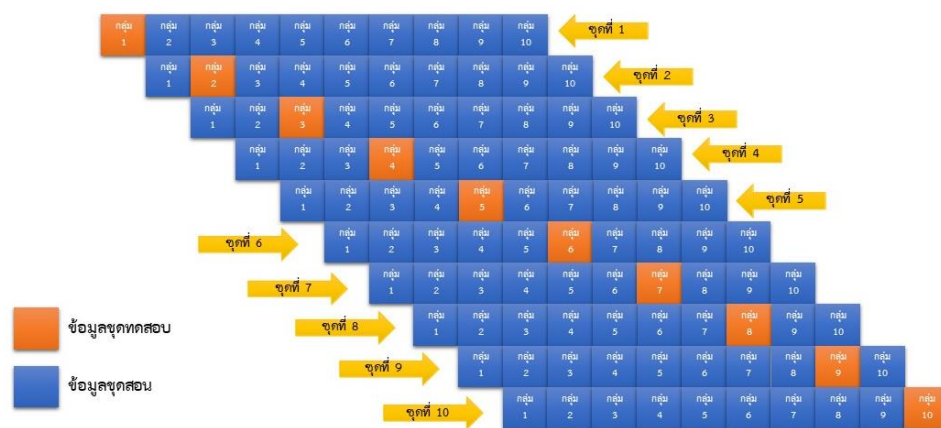
เมื่อ $y_i + 1$ หรือ $y_i - 1$

2.7

ประพัฒน์ พรหมน้ำอ่าง และคณะ [37] ได้ศึกษาการจำแนกกลุ่มข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล เพื่อแบ่งกลุ่มข้อความจากข้อความรีวิว ด้วยเทคนิค 4 เทคนิค ได้แก่ เทคนิค SVM, เทคนิค DT, เทคนิค K-NN และเทคนิค NB พบว่า เทคนิค SVM ได้ค่าความถูกต้องสูงที่สุด

2.1.5 การวัดประสิทธิภาพของแบบจำลอง

การวัดประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation [9] เป็นการแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ ชุดทดสอบ และชุดสอน ซึ่งมีหลักการดังนี้ เริ่มจากเลือกกลุ่มข้อมูลออกเป็น 10 กลุ่มเท่า ๆ กัน จากนั้นจะนำชุดข้อมูลกลุ่มที่ 1 ซึ่งเป็นข้อมูลชุดทดสอบและกำหนดให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน และในการทดลองครั้งที่สองจะใช้ชุดข้อมูลกลุ่มที่ 2 เป็นข้อมูลชุดทดสอบและให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำเช่นเดียวกันจนข้อมูลทุกกลุ่มถูกนำมาเป็นข้อมูลชุดทดสอบทั้งหมดครบตามจำนวนที่แบ่งไว้ โดยผลลัพธ์ที่ได้นั้นจะมากำหนดค่าเฉลี่ยความถูกต้องของการจำแนกข้อมูลในแต่ละรอบ โดยวิธีการทดสอบประสิทธิภาพแบบ 10-fold cross validation มีข้อเสียคือ จะต้องทำการเริ่มทดสอบใหม่โดยจะต้องทำทั้งหมด K รอบ ดังแสดงในภาพที่ 2.3



ภาพที่ 2.5 ตัวอย่างการทดสอบประสิทธิภาพแบบ 10-fold cross validation

จากภาพที่ 2.3 จะเป็นการทดสอบประสิทธิภาพแบบ 10-fold cross validation ซึ่งจะทำให้การแบ่งชุดข้อมูลออกเป็น 10 ชุด โดยในแต่ละรอบจะใช้ชุดข้อมูลเพื่อเป็นชุดข้อมูลทดสอบ 1 ชุด และให้ชุดข้อมูลอื่น ๆ เป็นข้อมูลชุดสอน โดยจะทำการทดสอบทั้งหมด 10 รอบ จากนั้นนำผลที่ได้มาทำการหาค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure)

ค่าความแม่นยำ (Precision) คือ การวัดความสามารถในการที่จะจัดเอกสารที่ไม่เกี่ยวข้องออกไปโดยที่ค่าความแม่นยำตรงนั้นจะเป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้มีการถูกดึงออกมา เพื่อเทียบกับจำนวนเอกสารที่ถูกดึงออกมาทั้งหมด ดังสมการที่ 2.8

$$Precision = \frac{TP}{TP + FP} \quad 2.8$$

ค่าความระลึก (Recall) คือ การวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยค่าความระลึกนั้นจะใช้อัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและได้มีการถูกดึงออกมาเทียบกับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด ดังสมการที่ 2.9

$$Recall = \frac{TP}{TP + FN} \quad 2.9$$

โดยที่

ค่า TP (True Positive) คือ จำนวนของข้อมูลที่แบบจำลองจำแนกความ
คิดเห็นเชิงบวกถูกต้อง

ค่า FP (False Positive) คือ จำนวนของข้อมูลที่แบบจำลองจำแนกความ
คิดเห็นเชิงบวกไม่ถูกต้อง

ค่า FN (False Negative) คือ จำนวนของข้อมูลที่แบบจำลองจำแนกความ
คิดเห็นเชิงลบไม่ถูกต้อง

ค่าความถ่วงดุล (F-Measure) คือ การวัดประสิทธิภาพโดยรวมของทั้งสองค่าระหว่างค่าความแม่นยำและค่าความระลึก ซึ่งนำค่าทั้งสองมาคำนวณร่วมกัน ดังสมการที่ 2.10

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad 2.10$$

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยเกี่ยวกับทัศนคติที่มีต่อออทิสติก

นฤมล ทวีพันธ์ [38] ได้ศึกษาเกี่ยวกับความเครียด ภาระการดูแล และทัศนคติที่มีต่อเด็กออทิสติกของครูที่ดูแลเด็กออทิสติกในโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ ศูนย์วิจัยและพัฒนาการศึกษา เพื่อศึกษาความเครียด ภาระการดูแล และทัศนคติที่มีต่อเด็กออทิสติกของครูที่ดูแลเด็กออทิสติก รวมถึงความสัมพันธ์ระหว่างความเครียดกับภาระการดูแล และทัศนคติที่มีต่อเด็กออทิสติก จากกลุ่มตัวอย่างที่ศึกษาเป็นครูที่ดูแลเด็กออทิสติก จำนวน 160 คน เข้าร่วมตอบแบบสอบถาม 5 ชุด ด้วยตนเอง ได้แก่ แบบสอบถามข้อมูลส่วนบุคคล แบบสอบถามความเครียดของครูที่ดูแลเด็กออทิสติก แบบสอบถามภาระการดูแลของครูที่ดูแลเด็กออทิสติก แบบสอบถามทัศนคติที่มีต่อเด็กออทิสติกของครูที่ดูแลเด็กออทิสติก และแบบประเมินการสนับสนุนทางสังคม จากนั้นนำเสนอความเครียด ภาระการดูแล ทัศนคติที่มีต่อเด็กออทิสติกของครู เป็นค่าเฉลี่ย ค่าสัดส่วน และร้อยละ พบว่า ครูที่ดูแลเด็กออทิสติกส่วนใหญ่มีความเครียดในระดับน้อย มีความรู้สึกไม่เป็นภาระ และมีทัศนคติที่มีต่อเด็กออทิสติกอยู่ในระดับดี

อรพิม วิสารทานนท์ [39] ได้ศึกษาเกี่ยวกับทัศนคติและความต้องการของครูปฐมวัยเขตกรุงเทพมหานครที่มีต่อเด็กปฐมวัยออทิสติก เพื่อศึกษาระดับและเปรียบเทียบทัศนคติ และศึกษาระดับและเปรียบเทียบความต้องการการช่วยเหลือจากภาครัฐของครูปฐมวัยเขตกรุงเทพมหานครที่สอนเด็กปฐมวัยออทิสติก จากกลุ่มตัวอย่างครูปฐมวัยที่สอนเด็กปฐมวัยออทิสติกจำนวน 100 คน เข้าตอบแบบสอบถามทัศนคติของครูปฐมวัยเขตกรุงเทพมหานครที่มีต่อเด็กปฐมวัยออทิสติก พบว่าทัศนคติของครูปฐมวัยในเขตกรุงเทพมหานครที่มีต่อเด็กปฐมวัยออทิสติก โดยภาพรวมและจำแนกรายด้านความรู้และความรู้สึกอยู่ในระดับดี และครูปฐมวัยเขตกรุงเทพมหานครที่สอนเด็กปฐมวัยออทิสติก ต้องการการช่วยเหลือจากภาครัฐอยู่ในระดับมากที่สุด

ขวัญใจ อธิปัญญาวงศ์ และคณะ [40] ศึกษาเกี่ยวกับเจตคติต่อการเป็นเพื่อนกับนักเรียนออทิสติกเรียนร่วมบางเวลาในโรงเรียนระดับประถมศึกษา เพื่อพรรณนาเจตคติต่อการเป็นเพื่อนกับนักเรียนออทิสติกเรียนร่วมบางเวลาในโรงเรียนจิตถรรระดับประถมศึกษา โดยเก็บข้อมูลนักเรียนในชั้นเรียนปกติจำนวน 13 คน และครูประจำชั้น/ครูวิชาปกติ และ ครูการศึกษาพิเศษจำนวน 5 คน ด้วยการสัมภาษณ์ร่วมกับการสังเกตอย่างมีส่วนร่วมนักเรียนในห้องเรียน พบว่า นักเรียนและครูมีเจตคติในด้านความเข้าใจ และความคิดเห็นเกี่ยวกับออทิสติกใกล้เคียงกัน ด้านการประเมินคุณค่าและความรู้สึกพบว่า นักเรียนและครูมีความรู้สึกเชิงบวก ได้แก่ ความเข้าใจเห็นใจ ส่วนความรู้สึกเชิงลบมักเกี่ยวกับความหงุดหงิดของเด็กออทิสติกที่ไม่ตอบสนองหรือกลัวพฤติกรรมที่ควบคุมไม่ได้ โดยสรุปเจตคติต่อการเป็นเพื่อนกับนักเรียนออทิสติกเรียนร่วมบางเวลานั้นมีทั้งเชิงบวกและเชิงลบ

มธุวดี อิงศิริรัตน์ [41] ศึกษาความรู้และทัศนคติของผู้ปกครองเด็กกลุ่มอาการออทิสซึมในโรงพยาบาลอุดรธานี เพื่อสำรวจความรู้และทัศนคติของผู้ปกครองเด็กกลุ่มอาการออทิสซึมที่มารับการรักษาที่โรงพยาบาลอุดรธานี และปัจจัยที่มีผลต่อความรู้และทัศนคติของผู้ปกครอง จากกลุ่มตัวอย่างผู้ปกครองของเด็กกลุ่มอาการออทิสซึมที่มารับการรักษาที่โรงพยาบาลอุดรธานีระหว่าง ธันวาคม 2560 - มีนาคม 2561 จำนวน 40 คน โดยใช้แบบสอบถามความรู้และทัศนคติต่อกลุ่มอาการออทิสซึม พบว่าผู้ปกครองที่ดูแลเด็กกลุ่มอาการออทิสซึม มีระดับความรู้อยู่ในเกณฑ์พอใช้ สำหรับประเด็นที่มีความรู้ต่ำถึงระดับต้องปรับปรุงคือ ด้านสาเหตุและอาการของโรค ส่วนด้านทัศนคติต่อโรคอยู่ในระดับปานกลาง ด้านศักยภาพครอบครัวอยู่ในระดับดี โดยการจัดอบรมให้ความรู้แก่ผู้ปกครองเป็นวิธีการเพิ่มความรู้และปรับทัศนคติให้ผู้ปกครองได้เป็นอย่างดี

2.2.2 งานวิจัยเกี่ยวกับการทำเหมืองความคิดเห็น

พิชญะ พรมลา และ จรรย์ แสนราช [8] ได้ศึกษาเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจของผู้เรียนที่มีต่อการจัดการเรียนการสอน ด้วยกระบวนการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล ใช้เทคนิค 2 ประเภท คือ เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล ได้แก่ Vote, Bagging และ RF และเทคนิควิธีมาตรฐาน ได้แก่ Decision Tree: DT, NB และ K-NN แสดงประสิทธิภาพของเทคนิคด้วยค่า Accuracy, Precision, Recall และ F-Measure โดยพบว่า เทคนิค Vote ให้ประสิทธิภาพมากที่สุด

สุภัทสรุ สมเจตนา และ จารี ทองคำ [9] ได้ศึกษาเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลอง เพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร

โดยเก็บรวบรวมข้อมูลความเป็นความคิดเห็นจากเว็บไซต์พันทิป และเฟซบุ๊ก จำนวน 1,925 ความคิดเห็น และเลือกใช้คำวิเศษณ์ เช่น หงุดหงิด, กี่ข้าง, ก้าวร้าว, กี่ดี, ก็ได้ และกล้า เป็นต้น มาเป็นตัวแปรในการสร้างแบบจำลองโดยใช้เทคนิคริปเปอร์ (Ripper), เทคนิค DT C4.5, เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค RF แสดงประสิทธิภาพของเทคนิคด้วยค่า Precision, Recall และ F-Measure พบว่า เทคนิค RF เป็นเทคนิคที่ดีที่สุด โดยให้ค่า Precision 89.62%, Recall 78.38% และ F-Measure 83.55%

นพมาศ ปักเข็ม และคณะ [42] ได้ศึกษาวิธีการจำแนกประเภทภูมิปัญญาท้องถิ่นแบบอัตโนมัติด้วยวิธีการทางเหมืองข้อมูล เพื่อให้สามารถจำแนกประเภทภูมิปัญญาท้องถิ่นจากข้อมูลแบบพรรณนาโวหารด้วยการประมวลผลของคอมพิวเตอร์ ซึ่งงานวิจัยนี้ไม่ได้จำแนกให้เห็นชนิดของคำอย่างชัดเจน จากนั้นใช้เทคนิค DT, เทคนิค K-NN และเทคนิค NB ทดสอบความแม่นยำด้วยวิธี K-Folds Cross Validation พบว่า เทคนิค NB ให้ค่าความแม่นยำสูงที่สุด

วสวัตดี อินทร์แปลง [43] ได้วิเคราะห์ความคิดเห็นต่อเกมมือถือพีซีด้วยเหมืองข้อความ เพื่อค้นหาเทคนิคการจำแนกที่มีประสิทธิภาพ โดยเก็บรวบรวมข้อความที่เป็นความคิดเห็นจาก Google Play จำนวน 3,789 ความคิดเห็น และเลือกใช้คำวิเศษณ์และคำสแลงบางคำที่มีความหมายเป็นคำวิเศษณ์ เพื่อแยกคุณลักษณะเชิงบวกและเชิงลบ โดยใช้เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค C4.5 ใช้หลักการ 10-fold cross validation แสดงประสิทธิภาพของเทคนิคด้วยค่าความถูกต้อง, ค่าความไว และ ค่าความจำเพาะ พบว่า เทคนิค SVM ให้ผลดีที่สุด ให้ค่าความถูกต้อง 99.65% ค่าความไว 99.30% และค่าความจำเพาะ 100%

สมศักดิ์ ศรีสุวรรณ และ สมัย ศรีสวย [32] ได้วิเคราะห์บทวิจารณ์ออนไลน์ผ่านเครือข่ายเฟซบุ๊กสาธารณะของนักท่องเที่ยวโดยใช้เทคนิคการสกัดคำ จำนวน 13,564 ความคิดเห็น เลือกเอาเฉพาะคำวิเศษณ์ เช่น สวยงาม, ดี, สะอาด และแพง เป็นต้น มาใช้เพื่อแยกความคิดเห็นเชิงบวกและเชิงลบ เปรียบเทียบประสิทธิภาพจากค่าความถูกต้องด้วยเทคนิค NB, เทคนิค K-NN และเทคนิค DT พบว่า เทคนิค NB ให้ผลดีที่สุด โดยให้ค่าความถูกต้อง 87.97% เทคนิค K-NN ให้ค่าความถูกต้อง 83.80% และเทคนิค DT ให้ค่าความถูกต้อง 79.89%

ธีรยุทธ คุณสุข และ จารี ทองคำ [22] ได้ศึกษากระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นลูกค้า โดยเก็บรวบรวมข้อมูลข้อความที่เป็นความคิดเห็นจากเว็บไซต์ wongnai.com จำนวน 4,487 ความคิดเห็น โดยเลือกใช้คำนาม คำกริยา และคำวิเศษณ์

เช่น กลับมา, อีกล้ำ, ปรากฏ, อร่อย และไม่ เป็นต้น มาเป็นตัวแปรในการสร้างแบบจำลอง โดยนำเทคนิค 3 เทคนิค มาคัดเลือกคุณลักษณะของข้อความ ได้แก่ Chi-Square Information Gain และ Information Gain Ratio ส่วนในการวัดประสิทธิภาพของการคัดเลือกคุณลักษณะ ใช้เทคนิค NB, เทคนิค SVM, เทคนิค K-NN และเทคนิค DT C4.5 มาวิเคราะห์ข้อมูลที่เก็บรวบรวมไว้ พบว่า เทคนิคการคัดเลือกคุณลักษณะแบบ Information Gain ร่วมกับเทคนิค NB ให้ผลที่ดีที่สุดในการจำแนกความคิดเห็น

นุชนาฏ ปิ่นเมือง และ จาริ ทองคำ [24] ได้ศึกษาเกี่ยวกับการจำแนกความคิดเห็นของคนไทยเกี่ยวกับสื่อออนไลน์โดยใช้การทำเหมืองข้อความ โดยเก็บรวบรวมข้อมูลความคิดเห็นต่อการใช้บริการพร้อมเพย์บนสื่อสังคมออนไลน์ จำนวน 1,570 ความคิดเห็น และเลือกใช้เฉพาะคำวิเศษณ์ เช่น เมื่อไหร่, อดีตกาล, ยาก และเหมือนเดิม เป็นต้น มาเป็นตัวแปรในการสร้างแบบจำลอง โดยใช้เทคนิค NB, เทคนิค SVM, เทคนิค K-NN, เทคนิค DT และเทคนิค DT C4.5 และใช้ 10-Fold cross validation แบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ และวัดประสิทธิภาพของเทคนิคด้วยค่า Accuracy, Precision และ Recall พบว่า เทคนิค NB ให้ผลดีที่สุด โดยให้ค่าความถูกต้อง 93.88% ค่าความแม่นยำ 94.02% และค่าความระลึก 93.54%

ประพัฒน์ พรหมน้ำอ่าง และคณะ [37] ได้ศึกษาการจำแนกกลุ่มข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล เพื่อแบ่งกลุ่มข้อความจากข้อความรีวิว จำนวน 12,500 ข้อความ ซึ่งงานวิจัยนี้ไม่ได้จำแนกให้เห็นชนิดของคำอย่างชัดเจน จากนั้นใช้เทคนิค 4 เทคนิค ได้แก่ เทคนิค SVM, เทคนิค DT, เทคนิค K-NN และเทคนิค NB พบว่า เทคนิค SVM ได้ค่าความถูกต้องสูงที่สุด

Phung และคณะ [44] ได้ศึกษาเครื่องมือสำหรับการทำเหมืองความคิดเห็นของลูกค้าออนไลน์ โดยรวบรวมรีวิวของนักท่องเที่ยวเกี่ยวกับโรงแรมในเวียดนามบนเว็บไซต์ Agoda.com ที่เป็นภาษาเวียดนาม จำนวน 15,480 ความคิดเห็น ขั้นตอนที่ 1 การรวบรวมและประมวลผลข้อมูล จากเว็บไซต์ Agoda.com โดยใช้โปรแกรมอัตโนมัติ ขั้นตอนที่ 2 การเตรียมชุดข้อมูล หรือการจำแนกประเภทข้อมูล แบ่งชุดข้อมูลที่มีค่าน้อยกว่า 7.0 เป็นลบ และความคิดเห็นที่มากกว่า 7.0 เป็นบวก ขั้นตอนที่ 3 การล้างข้อมูลก่อนประมวลผลชุดข้อมูล หรือการลบคำที่ไม่จำเป็นออก ขั้นตอนที่ 4 การแบ่งส่วนคำ โดยใช้โปรแกรมไพทอนในการตัดคำ ขั้นตอนที่ 5 การแยกคุณลักษณะตามอัลกอริทึม ขั้นตอนที่ 6 การสร้างแบบจำลอง โดยใช้วิธี 2 วิธี คือ วิธี Hold-Out สุ่มแบบข้อมูลออกเป็นชุดสอน 70% และข้อมูลทดสอบ 30% และวิธี K-Fold โดยนำข้อมูลมาวิเคราะห์ด้วยเครื่องมือสำหรับการทำเหมือง

ความคิดเห็น เพื่อค้นหาว่าแบบจำลองใดที่เข้ากันได้กับชุดข้อมูลที่ได้มากที่สุด และใช้แบบจำลองนี้ เพื่อคาดการณ์ความคิดเห็นสำหรับชุดข้อมูลที่รวบรวม โดยใช้เทคนิค NB, SVM, การถดถอยโลจิสติกส์ (Logistic Regression: LR), โครงข่ายประสาทเทียม (Neural Network: NN), DT และ RF ขั้นตอนที่ 7 การวัดประสิทธิภาพ แสดงประสิทธิภาพของเทคนิคด้วยค่า Accuracy, Precision, Recall และ F-Measure พบว่าเทคนิค LR, SVM และ NN มีประสิทธิภาพดีที่สุดในการทำเหมืองความคิดเห็นของลูกค้าออนไลน์ในเวียดนาม

Saad [10] ได้ศึกษาการทำเหมืองความคิดเห็นเกี่ยวกับข้อมูลในทวีตเตอร์ของสายการบินสหรัฐอเมริกา โดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อวิเคราะห์ความคิดเห็นของลูกค้าเกี่ยวกับบริการของสายการบินที่ปรากฏบนทวีตเตอร์เพื่อใช้ปรับปรุงบริการบางอย่างของสายการบิน ผู้วิจัยจัดหมวดหมู่โพสต์บนทวีตเตอร์ เป็นหมวดหมู่เชิงบวก เชิงลบ และเป็นกลาง โดยใช้ข้อมูลจากเว็บไซต์ทวีตเตอร์ที่เกี่ยวกับความคิดเห็นของผู้โดยสาร 6 สายการบินในสหรัฐอเมริกา จำนวน 14,640 รายการ ขั้นตอนที่ 1 ก่อนการประมวลผล นำข้อมูลที่ได้มาตัดคำและเครื่องหมายวรรคตอนที่ไม่จำเป็นออก แล้วแปลงตัวอักษรทั้งหมดให้เป็นตัวพิมพ์เล็กและหาที่มาของคำ ขั้นตอนที่ 2 จำแนกประเภทของคำ ออกเป็นความคิดเห็นเชิงลบ เชิงบวก และความเป็นกลาง ขั้นตอนที่ 3 การตรวจสอบและทดสอบ ข้อมูลแบ่งออกเป็นข้อมูลชุดสอน 70% และข้อมูลทดสอบ 30% จากนั้นใช้เทคนิคการตรวจสอบความถูกต้อง เทคนิคที่ใช้ คือ SVM, LR, RF, XgBoost, NB และ DT แสดงประสิทธิภาพของเทคนิคด้วยค่า Accuracy, Precision, Recall และ F-Measure พบว่า SVM มีความแม่นยำสูงสุด 83.31%

Zvarevashe และ Olugbara [45] ได้ศึกษาวิเคราะห์การทำเหมืองความคิดเห็นของความคิดเห็นที่มีต่อโรงแรม เพื่อให้ได้รับความรู้เชิงลึกเกี่ยวกับความคาดหวังของลูกค้าและสนับสนุนการจัดการความสัมพันธ์กับลูกค้าอย่างมีประสิทธิภาพ โดยขั้นตอนที่ 1 รวบรวมข้อมูลจากการตอบแบบสอบถามของลูกค้าที่มาใช้บริการโรงแรมจากเว็บไซต์เกี่ยวกับโรงแรมจำนวน 259,000 รายการ ขั้นตอนที่ 2 การติดป้ายกำกับความคิดเห็นหรือการแบ่งกลุ่มข้อความโดยใช้โปรแกรม ไพทอนในการตัดคำ แบ่งออกเป็นความคิดเห็นเชิงลบ เชิงบวก และความเป็นกลาง ขั้นตอนที่ 3 แปลงข้อความที่แบ่งกลุ่มตามลักษณะของเวกเตอร์ ขั้นตอนที่ 4 การสร้างแบบจำลอง จากขั้นตอนที่ใช้พบว่าขั้นตอนในการสร้างแบบจำลอง ด้วย NB มีความแม่นยำสูงสุดซึ่งมีถึง 80.9%

จากการวิเคราะห์และสังเคราะห์ พบว่า งานวิจัยโดยส่วนใหญ่จะนำคำวิเศษณ์มาเป็นตัวแปร หรือคุณลักษณะของการสร้างแบบจำลองเพื่อการจำแนก นอกจากนี้ยังพบงานวิจัยที่ใช้คำนามร่วมกับ คำกริยาและคำวิเศษณ์ ตัวอย่างคำวิเศษณ์ เช่น เมื่อไหร่, อร่อย, ดี, แพง, ดีมาก, หงุดหงิด และก็ดี เป็นต้น ตัวอย่างคำนาม เช่น เรา, สาขา และเมนู เป็นต้น และตัวอย่างคำกริยา เช่น มา, ลอง, สั่ง และปรากฏ เป็นต้น โดยผู้วิจัยจะนำเทคนิคและขั้นตอนของการทำเหมืองข้อความ รวมทั้งการเลือก ประเภทของคำจากงานวิจัยดังกล่าวมาข้างต้นมาใช้เป็นแนวทางในการศึกษาครั้งนี้



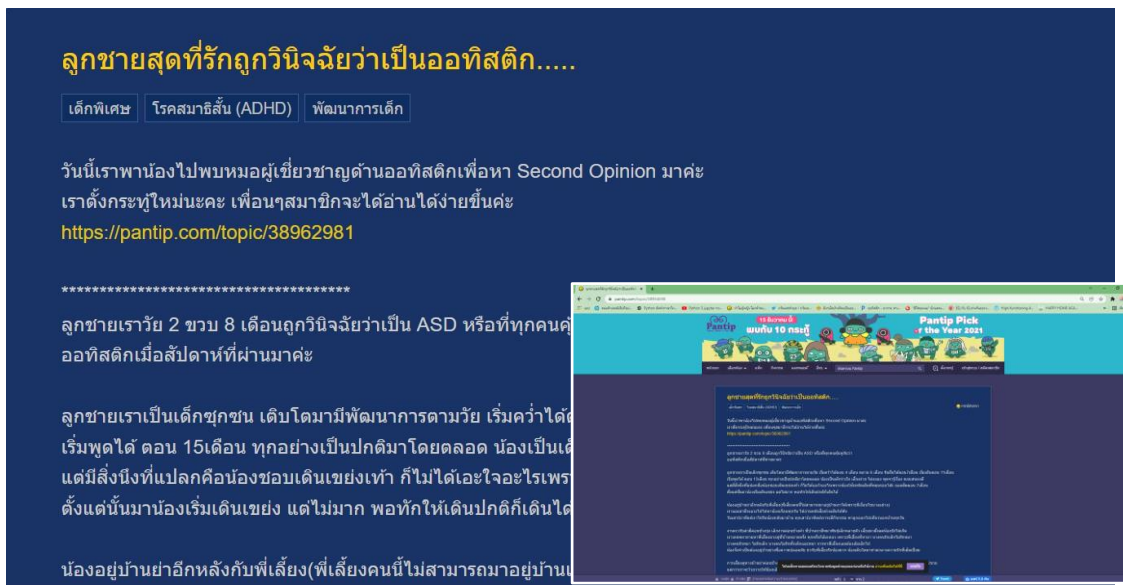
บทที่ 3

วิธีดำเนินการวิจัย

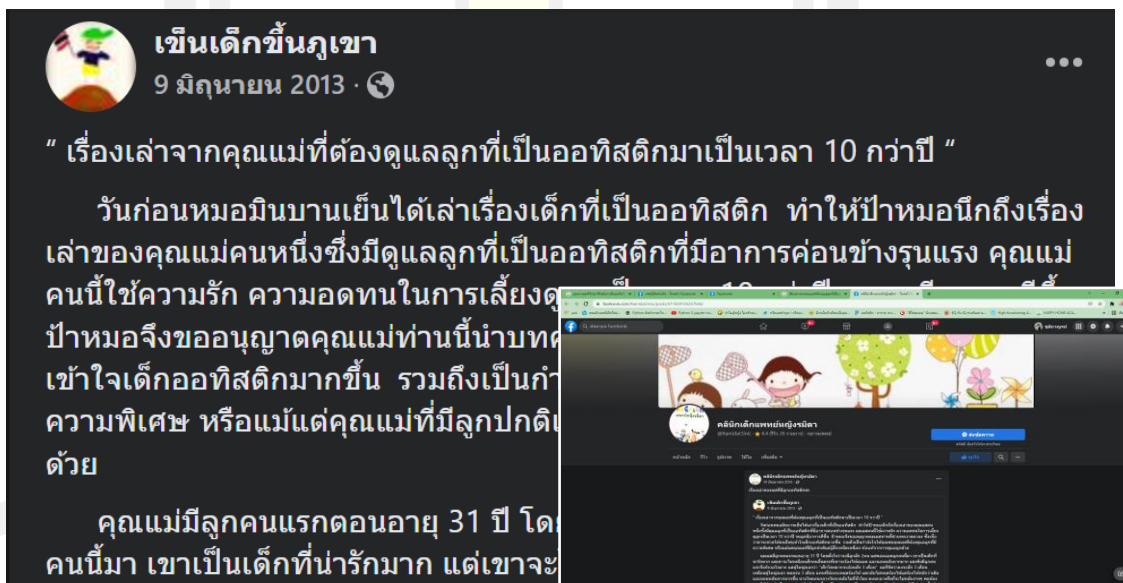
วิธีดำเนินงานวิจัยนี้ประกอบด้วย 4 ขั้นตอน ตามกระบวนการทำเหมืองความคิดเห็น ได้แก่ การเตรียมข้อมูล การจัดทำดัชนีข้อมูล การสร้างแบบจำลอง และการประเมินผล

3.1 การเตรียมข้อมูล

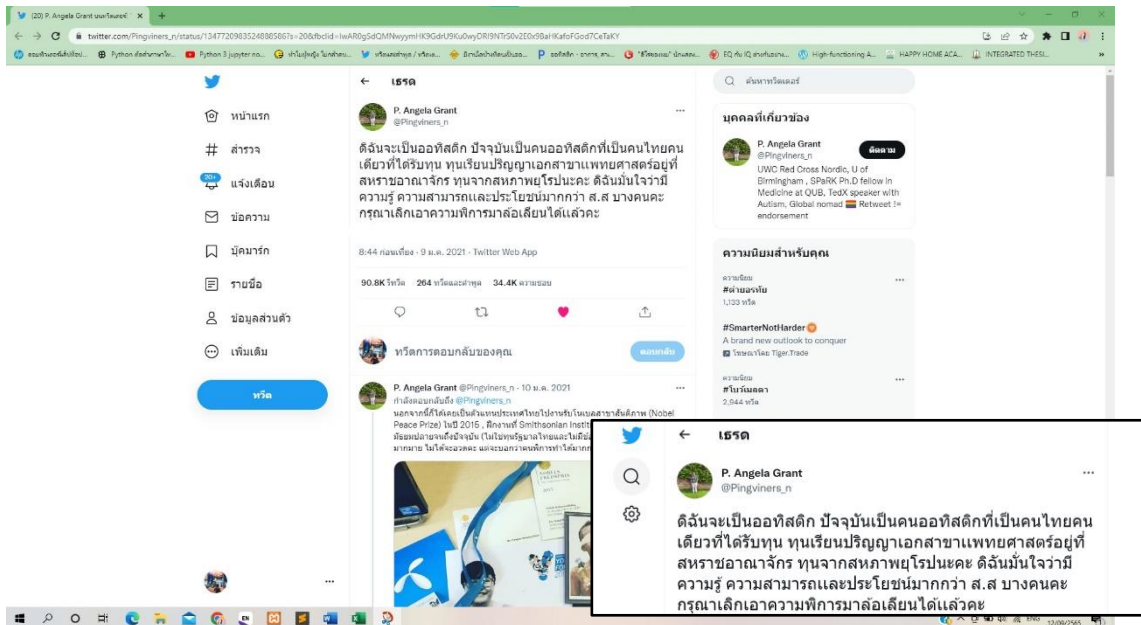
การเตรียมข้อมูลในงานวิจัยนี้ได้ค้นหาข้อความที่เป็นความคิดเห็นของประชาชนต่อออทิสติกและเป็นข้อความภาษาไทย ที่ปรากฏอยู่บนสื่อสังคมออนไลน์ จากเว็บไซต์พันทิป เฟซบุ๊ก และทวิตเตอร์ จำนวน 1,766 ความคิดเห็น ตั้งแต่เดือนมกราคม 2556 ถึง เดือนกุมภาพันธ์ 2565 ซึ่งคัดเลือกกระทู้หรือโพสต์ที่เกี่ยวข้องกับออทิสติกเท่านั้น เช่น ‘เด็กพิเศษมักจะมีความสามารถพิเศษ’, ‘ลูกชายสุดที่รักถูกวินิจฉัยว่าเป็นออทิสติก’, ‘เรื่องเล่าจากคุณแม่ที่ต้องดูแลลูกที่เป็นออทิสติกมาเป็นเวลา 10 กว่าปี’, ‘สำหรับเด็กออทิสติกอย่างผมจะไม่สามารถมีเพื่อนในสังคมเลยใช่หรือไม่’, ‘ชีวิตคนเป็นออทิสติกที่อีกสองเดือนจะอายุ 35’, ‘ให้เด็กกลุ่มออทิสติกแบบผมไปสอนหนังสือเด็กนักเรียนประถม คิดว่าพ่อแม่ที่เป็นผู้ปกครองจะรับได้ไหม’, ‘ไม่สบตา ไม่พาที่ ไม่ชี้นิ้ว’, ‘เด็กที่เป็นออทิสติกอยู่ร่วมกับคนในสังคมไม่ได้จริง ๆ เหรอคะ’, ‘แชร์ความในใจของปัญหาพฤติกรรมเด็กออทิสติก 10 อย่างที่หลายคนอาจยังไม่รู้’, ‘กว่าจะรู้ว่าลูกเป็นออทิสติกก็เกือบสายไปแล้ว’, ‘ลูกชายสุดที่รักถูกวินิจฉัยว่าเป็นออทิสติก 1 ปีเต็มกับความเปลี่ยนแปลง’ เป็นต้น ดังแสดงตัวอย่างในภาพที่ 3.1 ภาพที่ 3.2 และภาพที่ 3.3



ภาพที่ 3.1 การเก็บข้อมูลกระทู้หรือโพสต์จากเว็บไซต์พันทิป



ภาพที่ 3.2 การเก็บข้อมูลกระทู้หรือโพสต์จากเว็บไซต์เฟซบุ๊ก



ภาพที่ 3.3 การเก็บข้อมูลกระทู้หรือโพสต์เว็บไซต์ทวิตเตอร์

จากนั้นทำการดึงข้อมูลความคิดเห็น โดยในแต่ละความคิดเห็นประกอบไปด้วย วันที่โพสต์ หัวข้อโพสต์ วันที่แสดงความคิดเห็น เพศ การมีหรือไม่มีสมาชิกในครอบครัวเป็นออทิสติก และข้อความที่เป็นความคิดเห็นของผู้ใช้งาน ดังแสดงในภาพที่ 3.4

A	B	C	D	E	F
วันที่โพสต์หัวข้อ	หัวข้อโพสต์	วันที่แสดงความคิดเห็น	เพศ	คนในครอบครัวไม่เป็น 0 / เป็น 1	Comment
02/02/2556	คนที่เป็นออทิสติกสามารถมีลูกได้ไหมครั	02/02/2556	ญ	0	ออทิสติกไม่ได้ทำให้เป็นหมัน ถ้าอสุจิและไข่ยังผลิตอยู่ก็สืบพันธุ์ได้นะ ปัจจุบันพบว่ามีความปกติในส่วระบบประสาทบางส่วน ก็น่าจะเป็นที่ถูกรวมส่งต่อได้ละ
	สงสัยมาตั้งแต่เด็กแล้วว่าคนที่เป็นออทิสติก ไม่ว่าจะเป็นตั้งแต่ตอนๆ เป็นต้นไปจนถึงระดับรุนแรง จะสามารถมีลูกได้ไหมครับ ถ้ามีลักษณะการเป็นออทิสติกจะสืบทอดไปยังรุ่นลูกหลานได้ไหมครับ	02/02/2556	ญ	0	คนข้างบ้านจีน แล้วโดนขโมยเงินตั้งครครๆ ลูกออกมาก็เป็นออทิสติกเหมือนกัน แต่เขาเลี้ยงลูกเองได้ พาออกมาเดินเล่นทุกวันเลย สองคนแม่ลูกเราว่าเขาเก่งมากละ
		02/02/2556	N	0	มันเป็นโรคติดต่อทางพันธุกรรมมีผู้ศึกษาเกี่ยวกับโครโมโซมเพศที่สืบหามัน ลูกเกิดมาอาจเป็นโรคได้ อีก อาจส่งต่อสภาพชีวิต เศรษฐกิจของครอบครัวที่อาจแย่งได้ละ
		02/02/2556	N	0	แถวบ้านเราอีก ที่สาวข้างบ้านอายุน่าจะเกือบสามสิบก็เป็นออทิสติก โดนลุงหันขืน ไม่พอคนชี้เหล่าแล้วบ้านก็มาข่มขืนอีก ไม่ได้แจ้งความอะไร แต่บ้านที่เขากันทางบ้านลุงไม่พูดคุยกันอีกเลย
		02/02/2556	ญ	0	สืบทอดได้แต่ไม่ใช่ร้อยเปอร์เซ็นต์ การวิจัยในปัจจุบันยังไม่สามารถระบุได้ว่าที่เปอร์เซ็นต์แต่ก็อย่างว่าขึ้นชื่อว่ามีความเสี่ยง ไม่ว่าจะอะไร ถ้าป้องกันได้ก็ควรป้องกัน
		02/02/2556	ช	0	มีได้ มีโอกาสสืบทอดสูง ทางที่ดีไม่ควรจะมีลูกดีกว่า
		02/02/2556	N	0	มีได้ มีโอกาสสืบทอดสูง ทางที่ดี ไม่ควรจะมีลูกดีกว่า เวลาคุณหมอบอกประวัติตอนฝากครรภ์ เขาจะถามประมาณว่า มีญาติๆ คนใกล้ชิดคนเป็นออทิสติกหรือไม่งั้นมีประวัติครอบครัว
		02/02/2556	ช	0	มีได้ ถ้าไม่แน่ใจว่าเป็นใหม่ลองไปตรวจก่อนดีกว่าเพราะคุณอาจได้รับสิบบทมา หากเป็นก็คิดก่อนจะผสมพันธุ์เพราะส่งต่อไปรุ่นลูกคุณได้ด้วย

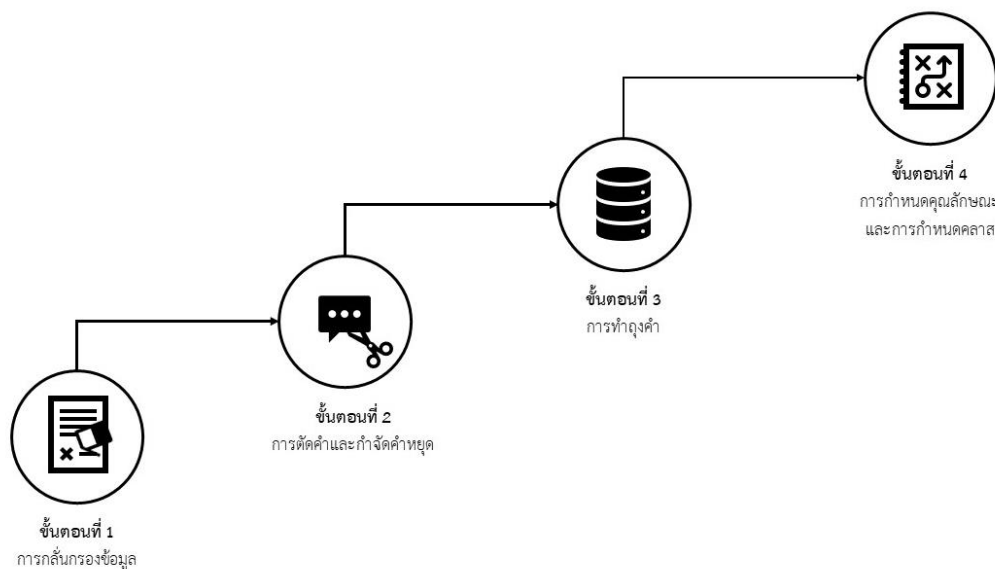
ภาพที่ 3.4 ตัวอย่างความคิดเห็นที่จัดเก็บในโปรแกรม Microsoft Excel

จากข้อมูลความคิดเห็นที่ได้จากเว็บไซต์พันทิป เฟซบุ๊ก และทวิตเตอร์ พบว่ามีความแตกต่างของจำนวนตัวอักษรของข้อความที่เป็นความคิดเห็น กล่าวคือ เว็บไซต์ทวิตเตอร์จะจำกัดตัวอักษรในการแสดงความคิดเห็นในหนึ่งครั้งอยู่ที่ 260 ตัวอักษร ส่วนเว็บไซต์พันทิปและเฟซบุ๊ก สามารถแสดงความคิดเห็นแบบไม่จำกัดตัวอักษร และยังพบข้อคิดเห็นอีกประเด็นหนึ่ง คือ ความนิยมของประชาชนในการใช้โซเชียลมีเดียมีผลต่อจำนวนข้อมูล กล่าวคือ ช่วงระยะเวลาที่ผู้วิจัยเก็บข้อมูล พบว่าประชาชนส่วนใหญ่ให้ความนิยมในการตั้งกระทู้และแสดงความคิดเห็นผ่านเว็บไซต์พันทิปมากกว่าเว็บไซต์เฟซบุ๊กและทวิตเตอร์ จึงทำให้ข้อมูลส่วนใหญ่ที่ได้จะได้จากเว็บไซต์พันทิปมากกว่าเว็บไซต์อื่น

ขั้นตอนการเตรียมข้อมูลเป็นกระบวนการที่ใช้วิเคราะห์ข้อมูลเพื่อเตรียมข้อมูลก่อนเข้าสู่กระบวนการวิเคราะห์ความคิดเห็นของประชาชนต่อออกทิสติก กระบวนการนี้จะทำให้เกิดความมั่นใจในคุณภาพของข้อมูลที่จะนำมาใช้วิเคราะห์ว่ามีความถูกต้อง หรืออยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้

3.2 การจัดทำดัชนีข้อมูล

การจัดทำดัชนีข้อมูลเพื่อแยกประเภทของความคิดเห็นที่เป็นเชิงบวกและความคิดเห็นที่เป็นเชิงลบ ซึ่งมีกระบวนการในการจัดทำดัชนีข้อมูลมี 4 ขั้นตอน ดังแสดงในภาพที่ 3.5



ภาพที่ 3.5 ขั้นตอนของการจัดทำดัชนีข้อมูล

3.2.1 การกลั่นกรองข้อมูล

จากการรวบรวมข้อความที่เป็นความคิดเห็นพบว่า มีความผิดพลาดบางประการที่เกิดจากเจ้าของความคิดเห็นเป็นผู้พิมพ์ผิด เช่น “เปน” ควรพิมพ์ว่า “เป็น” หรือ “ก่อ” ควรพิมพ์ว่า “ก็” รวมถึงการใช้ตัวย่อเพื่อประหยัดเวลาในการพิมพ์ เช่น “อทต” หมายถึง “ออทิสติก” หรือ “ผปก” หมายถึง “ผู้ปกครอง” เป็นต้น ในกระบวนการนี้เป็นการนำข้อความที่เป็นความคิดเห็นมาแก้ไขคำผิดให้ถูกต้อง และกำจัดตัวเลข สัญลักษณ์ ตัวอักษรภาษาอังกฤษ และช่องว่างออก

3.2.2 กระบวนการตัดคำและกำจัดคำหยุด

หลังจากผ่านกระบวนการกลั่นกรองข้อมูลแล้ว นำข้อความที่เป็นความคิดเห็นที่ได้มาตัดคำและกำจัดคำหยุด เป็นการแบ่งตัวคำจากข้อความออกเป็นคำ ๆ ตามรูปแบบของภาษาไทย เพื่อหาขอบเขตของแต่ละหน่วยคำ การตัดคำจะส่งผลให้การจำแนกข้อความมีประสิทธิภาพยิ่งขึ้น จากนั้นกำจัดคำหยุดต่าง ๆ ออกไป เช่น ครับ ค่ะ จ้า เป็นต้น โดยไม่ทำให้ความหมายของข้อความนั้นเปลี่ยนไป โดยใช้โปรแกรมภาษา Python ไลบรารีของ PyThaiNLP ใช้เทคนิคการตัดคำแบบ Multi-Cut เนื่องจากเป็นเทคนิคที่ตัดคำได้เหมาะสม และเพื่อความถูกต้องมากยิ่งขึ้นผู้วิจัยได้ขอคำแนะนำจากผู้เชี่ยวชาญด้านภาษาไทยเข้ามาช่วยพิจารณาในการเลือกเทคนิคการตัดคำที่เหมาะสมและการระบุชนิดของคำ รวมทั้งการแยกประเภทของความคิดเห็นที่เป็นคุณลักษณะเชิงบวกและเชิงลบ ดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อความหลังจากการถูกตัดคำด้วยไลบรารีของ PyThaiNLP

ลำดับ	ข้อความที่ถูกตัดคำ
1	'อาจจะ', 'ดีขึ้น', 'มี', 'การพัฒนา', 'มากขึ้น', 'จน', 'เกือบ', 'เหมือนปกติ', 'แต่', 'คง', 'ไม่', 'หาย'
2	'ลูกชาย', 'เพื่อน', 'จะ', 'ยี่สิบ', 'แล้ว', 'ไม่', 'หาย', 'แต่', 'ที่', 'เคยอ่านหนังสือ', 'เกี่ยวกับ', 'เด็ก', 'ออทิสติก', 'คน', 'หนึ่ง', 'ที่', 'พยายาม', 'หลุด', 'ออกมาจาก', 'โลก', 'ของ', 'ตนเอง', 'จน', 'สามารถ', 'เขียนหนังสือ', 'ได้', 'ก็', 'มี'
3	'คุณหมอบอกว่า', 'เป็น', 'ตลอดชีวิต', 'แต่', 'การสอน', 'ฝึกฝน', 'จะ', 'ช่วย', 'ให้', 'เขา', 'อยู่', 'ใน', 'สังคม', 'ได้', 'อย่าง', 'ไม่', 'แปลกแยก', 'มาก', 'ฟัง', 'แล้ว', 'จิตตก', 'เหมือนกัน'
4	'โธมา', 'จะ', 'เป็น', 'อย่างไร', 'ก็', 'หลาย', 'คน', 'ยังมี', 'ภาวะอารมณ์', 'และ', 'พฤติกรรมที่', 'ไม่', 'เหมาะสม', 'ละ', 'ตัว', 'ผม', 'น้อง', 'เขาว', 'เป็น', 'เด็ก', 'ใน', 'กลุ่ม', 'ออทิสติก', 'อายุ', 'ยี่สิบ', 'สาม', 'ปี', 'ก็', 'ยังมี', 'วุฒิภาวะ', 'เหมือน', 'เด็กประถมอยู่เลย', 'ที่', 'มีปัญหา', 'มาก', 'สุด', 'ก็', 'เป็นเรื่อง', 'ของอารมณ์', 'อารมณ์', 'แปรปรวน', 'ง่าย', 'คิด', 'เยอะ', 'คิดมาก', 'ด้วย', 'ละ', 'พูด'

ลำดับ	ข้อความที่ถูกตัดคำ
	กัน', 'ตามตรง', 'โต', 'ขึ้น', 'มา', 'จะ', 'ดีขึ้น', 'แต่', 'ไม่', 'หาย', 'มี', 'อาการ', 'ไป', 'ตลอดชีวิตละ'
5	'ไอส์ไตน์', 'ก็', 'เป็น', 'ออสติก', 'ผม', 'เลย', 'ไม่', 'แน่ใจ', 'ว่านิยาม', 'คำ', 'ว่า', 'ฉลาด', 'ของ', 'คุณ', 'เจ้าของกระทู้', 'หมายถึง', 'รูปแบบอย่างไร'
6	'คน', 'ข้าง', 'บ้าน', 'เป็น', 'แล้ว', 'โดน', 'ข่มขืน', 'จน', 'ตั้งครรภ์', 'ลูก', 'ออก', 'มา', 'ก็', 'เป็น', 'ออสติก', 'เหมือนกัน', 'แต่', 'เขา', 'เลี้ยง', 'ลูกเอง', 'ได้', 'พา', 'ออก', 'มา', 'เดิน', 'เล่น', 'ทุกวัน', 'เลย', 'สอง', 'คน', 'แม่', 'ลูกเรา', 'ว่า', 'เขา', 'เก่ง', 'มาก'
7	'มัน', 'เป็น', 'โรคติดต่อ', 'ทาง', 'พันธุกรรม', 'รู้สึก', 'จะ', 'เกี่ยวกับ', 'โครโมโซม', 'แห่ง', 'ที่', 'สืบห้า', 'ลูก', 'เกิด', 'มา', 'อาจ', 'เป็น', 'โรค', 'ได้', 'อีก', 'อาจ', 'ส่ง', 'ผล', 'ต่อ', 'สภาพ', 'ชีวิต', 'เศรษฐกิจ', 'ของ', 'ครอบครัว', 'ที่', 'อาจ', 'แย่ง', 'ได้'
8	'มี', 'ได้', 'มี', 'โอกาส', 'สืบทอด', 'สูง', 'ทาง', 'ที่', 'ดี', 'ก็', 'ไม่ควร', 'จะ', 'มี', 'ลูก', 'ดีกว่า'
9	'เหมือน', 'หลาน', 'เลย', 'นอน', 'หนุน', 'ตัก', 'พี่เลี้ยง', 'ดู', 'พัฒนา', 'ได้', 'ทั้งวัน', 'ตอน', 'เด็กเด็ก', 'แต่', 'ปัจจุบันนี้', 'เป็น', 'หนุ่ม', 'แล้ว', 'ดีขึ้น', 'มาก', 'พัฒนาการ', 'ดีขึ้น', 'อย่าง', 'มาก'
10	'คุณแม่', 'ใช้', 'วิธี', 'เล่น', 'กับ', 'ลูก', 'ได้', 'ลูกเรา', 'ก็', 'เป็น', 'แต่', 'พา', 'น้อง', 'หา', 'หมอ', 'ตั้งแต่', 'ขวบ', 'กว่า', 'ทำ', 'กิจกรรมบำบัด', 'ฝึก', 'พูด', 'อาชบาบำบัด', 'คุณหมอ', 'ที่', 'รักษา', 'อยู่', 'ให้', 'เข้า', 'แซมเบอร์', 'ฟรี', 'ด้วย', 'ตอนนี้', 'อายุ', 'หก', 'ขวบ', 'แล้ว', 'ดีขึ้น', 'แล้ว'

3.2.3 การทำถ่วงคำ

การทำถ่วงคำเป็นกระบวนการหลังจากที่ได้ตัดคำและกำจัดคำหยุดแล้ว จะได้คำจำนวนหนึ่งซึ่งไม่ซ้ำกัน จากนั้นนำคำแต่ละคำมากำหนดประเภทของคำตามพจนานุกรม โดยคำที่ได้จะถูกแบ่งออกเป็น 7 ประเภท ได้แก่ คำนาม คำสรรพนาม คำกริยา คำวิเศษณ์ คำบุพบท คำสันธาน และคำอุทาน โดยกระบวนการนี้จะใช้พจนานุกรมอิเล็กทรอนิกส์ ฉบับราชบัณฑิตยสถาน สามารถแสดงตัวอย่างการกำหนดประเภทของคำในตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างการกำหนดประเภทคำตามความหมายในพจนานุกรม

ลำดับ	คำ	ประเภทคำ
1	ลูกชาย	คำนาม
2	ไม่	คำวิเศษณ์
3	ดีขึ้น	คำวิเศษณ์
4	หาย	คำกริยา

ลำดับ	คำ	ประเภทคำ
5	พยายาม	คำกริยา
6	ปัจจุบัน	คำวิเศษณ์
7	เยอะ	คำวิเศษณ์
8	อotisติก	คำนาม
9	ผม	คำนาม
10	ฉลาด	คำวิเศษณ์

จากตารางที่ 3.2 แสดงตัวอย่างการกำหนดประเภทคำตามพจนานุกรม จากนั้นจะเลือกเฉพาะคำวิเศษณ์ คำคุณศัพท์ และคำกริยามาใช้ในการกำหนดคุณลักษณะและการกำหนดคลาส เนื่องจากคำวิเศษณ์ คำคุณศัพท์ และคำกริยาสามารถแสดงอารมณ์ความรู้สึกในเชิงบวกหรือเชิงลบ

3.2.4 การกำหนดคุณลักษณะและการกำหนดคลาส

การกำหนดคุณลักษณะ โดยการนำเอาคำวิเศษณ์ คำคุณศัพท์ และคำกริยามาเป็นคุณลักษณะในการสร้างแบบจำลอง โดยถ้าความคิดเห็นนั้นมีคุณลักษณะจะแทนด้วยจำนวนคุณลักษณะที่ปรากฏ แต่ถ้าความคิดเห็นนั้นไม่ปรากฏว่ามีคุณลักษณะนั้นจะแทนค่าด้วย 0

ส่วนการกำหนดคลาสในงานวิจัยนี้เป็นการนับจำนวนคำที่ปรากฏในแต่ละความคิดเห็นแล้วแทนค่าของจำนวนคำคุณลักษณะที่ไม่ปรากฏเป็น 0 และนับจำนวนคุณลักษณะที่ปรากฏตามจำนวนจริงที่พบในความคิดเห็นนั้น ๆ จากนั้นทำการรวมค่าของคุณลักษณะเชิงบวกเข้าด้วยกัน และรวมค่าคุณลักษณะเชิงลบเข้าด้วยกันตามลำดับ สุดท้ายนำความถี่คุณลักษณะเชิงบวกลบด้วยคุณลักษณะเชิงลบจะได้เป็น 3 กรณี ดังนี้

ถ้าผลที่ได้จากการคำนวณมีค่าเป็นบวก จะแทนค่าของคลาสด้วย P (Positive)

ถ้าผลที่ได้จากการคำนวณมีค่าเป็นลบ จะแทนค่าของคลาสด้วย N (Negative)

แต่ถ้าผลที่ได้จากการคำนวณมีค่าเท่ากับ 0 และข้อความนั้น จะถูกตัดออก ดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 ตัวอย่างการกำหนดประเภทของคลาส

ลำดับ	คุณลักษณะเชิงบวก		คุณลักษณะเชิงลบ		ความถี่ คุณลักษณะ เชิงบวก	ความถี่ คุณลักษณะ เชิงลบ	คลาส
	ดีขึ้น	มาก	ไม่	คิดมาก			
1	1	1	1	0	2	1	P
2	0	0	1	0	0	1	N
3	0	1	1	0	1	1	O
4	0	1	2	1	1	3	N
5	0	0	1	0	0	1	N
6	0	1	0	0	1	0	P
7	0	0	0	0	0	0	O
8	0	0	0	0	0	0	O
9	2	2	0	0	4	0	P
10	1	0	0	0	1	0	P

จากตารางที่ 3.3 การกำหนดประเภทของคลาส จากความคิดเห็นทั้งหมด 1,766 ความคิดเห็น สามารถแบ่งเป็นความคิดเห็นเชิงบวกได้จำนวน 1,588 ความคิดเห็น และความคิดเห็นเชิงลบได้จำนวน 178 ความคิดเห็น ซึ่งผลที่ได้จากการทำดัชนีข้อมูลทำให้ได้ชุดข้อมูล เพื่อนำไปสร้างแบบจำลองเพื่อการจำแนกรู้สึกของประชาชนที่มีต่อออกทิสติก ด้วยเทคนิคเหมือนความคิดเห็นต่อไป

3.3 การสร้างแบบจำลอง

กระบวนการสร้างแบบจำลองความคิดเห็นของประชาชนต่อออกทิสติก ผู้วิจัยได้ใช้เทคนิคเหมือนข้อมูลรวมทั้งหมด 5 เทคนิค โดยทางผู้วิจัยได้ทดลองปรับใช้ค่าพารามิเตอร์อื่นแล้ว ผลที่ได้มีค่าเท่าเดิมแต่ใช้เวลามากขึ้นสามารถสรุปได้ว่า พารามิเตอร์ดังต่อไปนี้เป็นค่าพารามิเตอร์ที่เหมาะสมที่สุดในการสร้างแบบจำลอง ซึ่งมี พารามิเตอร์ ดังนี้

3.3.1 เทคนิคนาอิวเบย์ (Naïve Bayes) มีค่าพารามิเตอร์ ดังนี้

- 1) ค่า numDecimalPlaces กำหนดให้มีค่าเท่ากับ 2

3.3.2 เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) มีค่าพารามิเตอร์ ดังนี้

- 1) ค่า Confidence Factor กำหนดให้มีค่าเท่ากับ 0.25
- 2) ค่า minNumObj กำหนดให้มีค่าเท่ากับ 2
- 3) ค่า numDecimalPlaces กำหนดให้มีค่าเท่ากับ 2
- 4) ค่า numFolds กำหนดให้มีค่าเท่ากับ 3

3.3.3 เทคนิคป่าสุ่ม (Random Forest) มีค่าพารามิเตอร์ ดังนี้

- 1) ค่า maxDepth กำหนดให้มีค่าเท่ากับ 0
- 2) ค่า numDecimalPlaces กำหนดให้มีค่าเท่ากับ 2
- 3) ค่า numExecutionSlots กำหนดให้มีค่าเท่ากับ 1
- 4) ค่า numFeatures กำหนดให้มีค่าเท่ากับ 0
- 5) ค่า numIterations กำหนดให้มีค่าเท่ากับ 100

3.3.4 เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) มีค่าพารามิเตอร์ ดังนี้

- 1) ค่า numDecimalPlaces กำหนดให้มีค่าเท่ากับ 2

3.3.5 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) มีค่าพารามิเตอร์ ดังนี้

- 1) ค่า calibrator ที่จะใช้ กำหนดให้เป็น Logistic
- 2) ค่า kernel กำหนดให้เป็น Polk Kernel

3.4 การประเมินผล

การประเมินผลเป็นขั้นตอนในการวัดประสิทธิภาพของแบบจำลอง เพื่อค้นหาเทคนิคที่มีประสิทธิภาพในการจำแนก ใช้หลักการ 10-Fold Cross Validation เป็นการแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ ชุดทดสอบและชุดสอน โดยเลือกสุ่มข้อมูลออกเป็น 10 ชุดเท่า ๆ กัน จากนั้นในการทดลองครั้งแรกจะใช้ชุดข้อมูลกลุ่มที่ 1 เป็นข้อมูลชุดทดสอบและกำหนดให้ข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน และในการทดลองครั้งที่สองจะใช้ชุดข้อมูลกลุ่มที่ 2 เป็นข้อมูลชุดทดสอบและให้ข้อมูลชุดที่เหลือเป็น

ข้อมูลชุดสอน ทำเช่นเดียวกันจนข้อมูลทุกกลุ่มถูกนำมาเป็นข้อมูลชุดทดสอบทั้งหมดครบตามจำนวนที่แบ่งไว้ ดังภาพที่ 3.5

ข้อมูลชุดที่ 1	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 2	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 3	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 4	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 5	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 6	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 7	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 8	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน	ชุดสอน
ข้อมูลชุดที่ 9	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ	ชุดสอน
ข้อมูลชุดที่ 10	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดสอน	ชุดทดสอบ

ภาพที่ 3.6 ตัวอย่างการประเมินผล

หลังจากนั้นนำผลลัพธ์ที่ได้มาคำนวณหาค่าประสิทธิภาพโดยรวมของแบบจำลอง ซึ่งค่าที่นำมาคำนวณ คือ ค่าจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงบวกถูกต้อง (TP), ค่าจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงบวกไม่ถูกต้อง (FP) และค่าจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงลบไม่ถูกต้อง (FN) มาใช้คำนวณหาค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ส่วนค่าความถ่วงดุล (F-Measure) จะนำค่าที่ได้จากการคำนวณหาค่าความแม่นยำ และค่าความระลึก นำมาคำนวณร่วมกัน

ตัวอย่างการคำนวณหาค่าความแม่นยำ

หลังจากนำชุดข้อมูลที่นำมาวัดประสิทธิภาพของแบบจำลองของเทคนิคซีฟพอร์ตเวคเตอร์แมชชีน ด้วยหลักการ 10-Fold Cross Validation แล้ว จะได้ผลลัพธ์ดังต่อไปนี้

ค่า TP = 1518

ค่า FP = 4

ค่า FN = 8

ค่า TN = 234

จากนั้นนำ

$$\frac{\text{ค่า } TP}{\text{ค่า } TP + \text{ค่า } FP} = \frac{1,518}{1,518 + 4}$$

$$= \frac{1,518}{1,522}$$

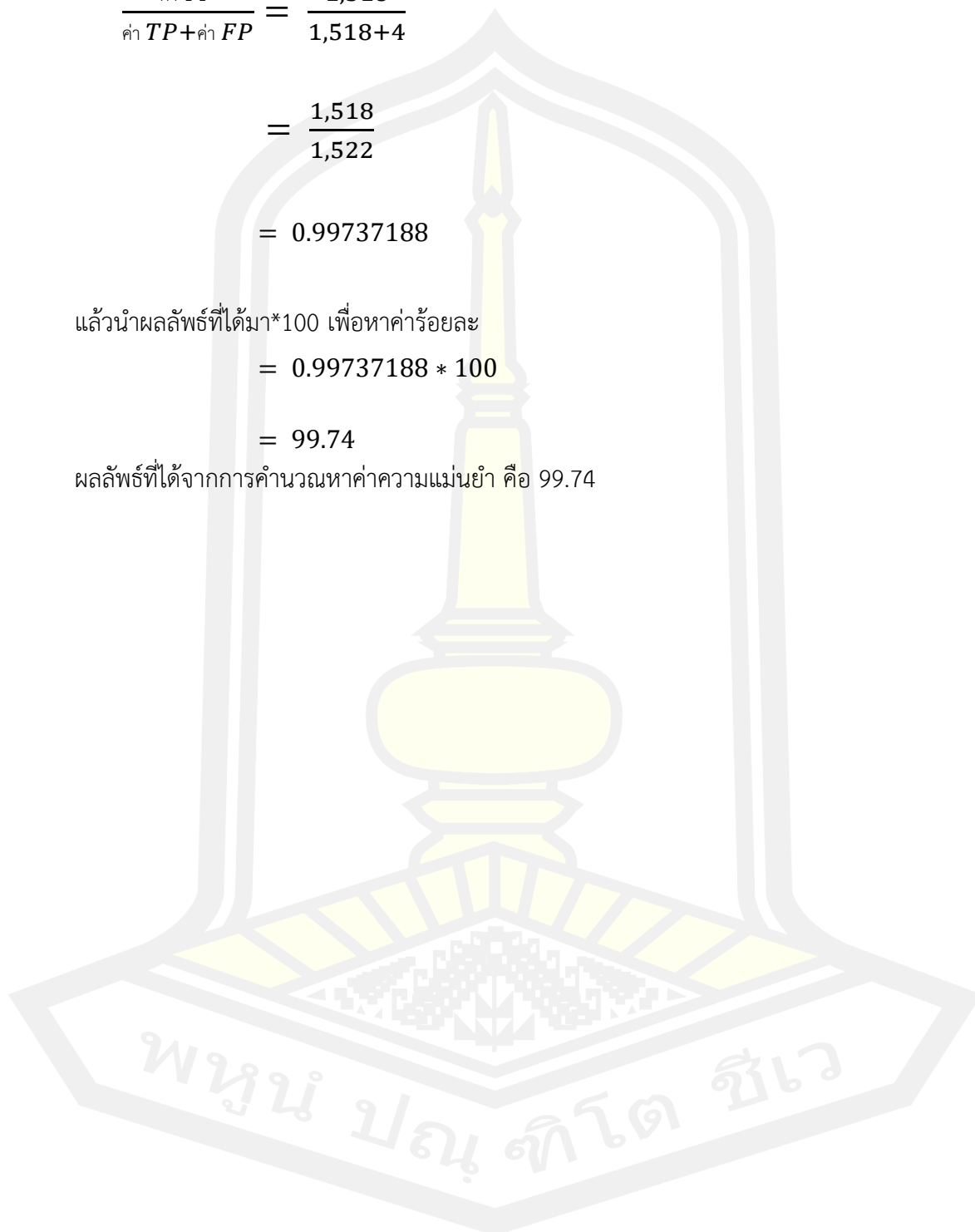
$$= 0.99737188$$

แล้วนำผลลัพธ์ที่ได้มา*100 เพื่อหาค่าร้อยละ

$$= 0.99737188 * 100$$

$$= 99.74$$

ผลลัพธ์ที่ได้จากการคำนวณหาค่าความแม่นยำ คือ 99.74



บทที่ 4

ผลการวิจัย

ในบทนี้ได้กล่าวถึง ผลการศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออตติก และเปรียบเทียบแบบจำลองเพื่อจำแนกรู้สึกของประชาชนที่มีต่อออตติก ผู้วิจัยได้ใช้โปรแกรม Weka เวอร์ชัน 3.9.5 เป็นเครื่องมือที่ใช้ในกระบวนการวิจัยทำเหมืองความคิดเห็น

4.1 ผลการศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนต่อออตติก

ผลการศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนต่อออตติก จากการนำ 4 ขั้นตอนของการทำเหมืองความคิดเห็นมาใช้พบว่า ในขั้นตอนการเตรียมข้อมูลจากเดือนมกราคม 2556 ถึง เดือนกุมภาพันธ์ 2565 ลักษณะของข้อความที่ได้รับรวบรวมมา มีความแตกต่างเรื่องจำนวนคำอย่างเห็นได้ชัด โดยประชาชนที่มาแสดงความคิดเห็น บางคนแสดงความคิดเห็นขนาดยาวเกิน 500 คำ แต่บางคนแสดงความคิดเห็นขนาดสั้นเพียง 1-2 คำ เช่น สู้, น่ารัก, เยี่ยม เป็นต้น ซึ่งความคิดเห็นที่มีจำนวน 1-2 คำนั้น มีผลต่อการนับจำนวนคำ กล่าวคือ เหมือนแปลผลของคำว่า สู้ หรือ น่ารัก ออกมา จะได้ผลเท่ากับ 0 โดยค่า 0 ไม่มีผลต่อการวิเคราะห์

หลังจากเตรียมข้อมูลแล้วนำข้อมูลดังกล่าวมาจัดทำดัชนีข้อมูลพบว่า มีความผิดพลาดบางประการที่เกิดขึ้นจากเจ้าของความคิดเห็นเป็นผู้พิมพ์ผิด เช่น คำว่า ว่า “เปน” “ก้อ” ผู้วิจัยต้องแก้ไขคำดังกล่าวเป็นคำว่า “เป็น” และ “ก็” รวมถึงการใช้ตัวย่อที่ไม่เป็นทางการหรือตัวย่อที่ไม่เป็นที่รู้จักแพร่หลาย เช่น “อทต” หมายถึง ออตติก หรือ “ผปค” หมายถึง ผู้ปกครอง ต้องแก้ไขเป็นคำเต็มว่า “ออตติก” และ “ผู้ปกครอง” เป็นต้น นอกจากนี้ยังพบปัญหาของการตัดคำและกำจัดคำหยุดอีกประการ คือ เมื่อนับจำนวนคำแล้วพบว่า มีคำที่ไม่สามารถระบุเป็นค่าได้ เนื่องจากข้อมูลที่ได้มาจากอินเทอร์เน็ตอาจเกิดความผิดพลาดจากผู้พิมพ์ข้อความแสดงความคิดเห็น เช่น การพิมพ์สระแอ ผู้แสดงความคิดเห็นบางคนใช้สระแอสองครั้งแทนการใช้สระแอ ดังนั้นผู้วิจัยต้องตรวจสอบและแก้ไขคำให้ถูกต้องก่อนที่จะทำถ่วงคำ

จากการทำถ่วงคำเพื่อกำหนดประเภทของคำพบว่า ยังมีการกำหนดประเภทของคำไม่ถูกต้องในบางคำ เช่น คำว่า “แอบแฝง” และ “กังวลใจ” โปรแกรมกำหนดให้เป็นคำนาม แต่ตามหลักของ

ภาษาไทยเป็นคำกริยา เป็นต้น ผู้วิจัยจึงกำหนดประเภทของคำที่ถูกต้องอีกครั้ง กระทั่งได้คำทั้งหมดจำนวน 6,933 คำ แล้วได้ทำการศึกษาคำกริยา (VERB) คำวิเศษณ์ (ADV) และคำคุณศัพท์ (ADJ) เนื่องจากคำทั้ง 3 ประเภท สามารถแสดงอารมณ์ความรู้สึกในเชิงบวกหรือเชิงลบได้ สามารถสรุปจำนวนคำของแต่ละประเภทได้ดังตารางที่ 4.1

ตารางที่ 4.1 จำนวนของคำแต่ละประเภท

ลำดับ	ประเภทคำ	จำนวนคำทั้งหมด	คุณลักษณะเชิงบวก (คำ)	คุณลักษณะเชิงลบ (คำ)
1	VERB	1,881	680	1,201
2	ADJ	162	77	85
3	ADV	301	163	138

จากตารางที่ 4.1 แสดงจำนวนคำในละประเภทของกลุ่มคำพบว่า มีคำที่สามารถนำมาวิเคราะห์จำนวน 2,344 คำ โดยแบ่งเป็น 3 ชุดข้อมูล ได้แก่ ชุดข้อมูลที่มีเพียงคำกริยาจำนวน 1,881 คำ ชุดข้อมูลที่มีเพียงคำคุณศัพท์จำนวน 162 คำ และชุดข้อมูลที่มีเพียงคำวิเศษณ์จำนวน 301 คำ ซึ่งประเภทคำที่พบมากที่สุด ได้แก่ คำกริยา รองลงมาคือ คำวิเศษณ์ และคำคุณศัพท์ ตามลำดับ ด้านคุณลักษณะเชิงบวก พบว่า มีคำกริยามากที่สุดจำนวน 680 คำ รองลงมาคือ คำวิเศษณ์จำนวน 163 คำ และคำคุณศัพท์มีจำนวนน้อยที่สุด 77 คำ และคุณลักษณะเชิงลบ พบว่า มีคำกริยามากที่สุดจำนวน 1,201 คำ รองลงมาคือ คำวิเศษณ์จำนวน 138 คำ และคำคุณศัพท์มีจำนวนน้อยที่สุด 85 คำ

นอกจากนี้ผู้วิจัยได้รวมชุดข้อมูลเข้าด้วยกัน ได้แก่ ชุดข้อมูลที่มีคำวิเศษณ์และคำคุณศัพท์จำนวน 463 คำ และชุดข้อมูลที่มีคำวิเศษณ์รวมกับคำคุณศัพท์และคำกริยาจำนวน 2,344 คำ ทำให้ได้ชุดข้อมูลทั้งหมด 5 ชุดข้อมูล ที่จะนำไปสร้างแบบจำลองการจำแนกความคิดเห็นต่อไป

ในขั้นตอนการสร้างแบบจำลองและการประเมินผลด้วยโปรแกรม Weka จากเทคนิคทั้ง 5 เทคนิค พบข้อผิดพลาดบางประการจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยในครั้งแรกผู้วิจัยได้เลือกใช้การสร้างแบบจำลองด้วยฟังก์ชัน LibSVM ปรากฏว่าได้ผลจากการสร้างแบบจำลองเพียงจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงบวกถูกต้อง (TP) และค่าจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงบวกไม่ถูกต้อง (FP) ส่วนค่าจำนวนของข้อมูลที่แบบจำลองจำแนกความคิดเห็นเชิงลบไม่ถูกต้อง (FN) ได้ค่าเท่ากับ 0 เมื่อนำค่าที่ได้จากการสร้างแบบจำลองของเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ด้วยฟังก์ชัน LibSVM มาคำนวณหาค่าความระลึกได้ผลเป็นค่าเฉลี่ย 100%

ทุกชุดข้อมูล ทำให้ไม่สามารถนำมาวิเคราะห์หาชุดข้อมูลที่เหมาะสมที่สุดได้ ผู้วิจัยจึงได้เปลี่ยนฟังก์ชันในการสร้างแบบจำลองของเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เป็นฟังก์ชัน SMO ทำให้ได้ค่าที่สามารถนำไปประเมินผลและวิเคราะห์ผลต่อได้

4.2 ผลการเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนต่อออตติสติก ด้วยเทคนิคเหมืองความคิดเห็น

ผลการเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนต่อออตติสติก ด้วยเทคนิคเหมืองความคิดเห็น จากชุดข้อมูลทั้งหมด 5 ชุดข้อมูล ได้แก่ ชุดข้อมูลที่มีเพียงคำวิเศษณ์ ชุดข้อมูลที่มีเพียงคำคุณศัพท์ ชุดข้อมูลที่มีเพียงคำกริยา ชุดข้อมูลที่มีคำวิเศษณ์และคำคุณศัพท์ และชุดข้อมูลที่มีคำวิเศษณ์รวมกับคำคุณศัพท์และคำกริยา มาสร้างแบบจำลองเพื่อการจำแนกด้วย 5 เทคนิค ได้แก่ เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิคป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) สามารถแสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุลมาเป็นเกณฑ์ในการเปรียบเทียบแบบจำลอง ดังนี้

4.2.1 ผลการเปรียบเทียบค่าความแม่นยำของแบบจำลอง

จากการเปรียบเทียบค่าความแม่นยำ (Precision) ของแบบจำลองที่สร้างด้วยเทคนิคนาอิวเบย์ (Naïve Bayes), เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิค ป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ในการจำแนกความคิดเห็นของประชาชนต่อออตติสติก ปรากฏผลดังตารางที่ 4.2

พหุ ประสิทธิภาพ

ตารางที่ 4.2 ค่าความแม่นยำของแบบจำลอง

ประเภทของคำ	เทคนิคการจำแนก					Average
	NB	C4.5	RF	K-NN	SVM	
VERB	86.95	88.44	91.90	91.95	95.81	91.01
ADJ	96.45	98.37	99.54	99.61	99.74	98.74
ADV	94.30	97.09	98.30	98.42	99.14	97.45
ADV+ADJ	93.45	96.12	97.40	97.63	98.41	96.60
ADV+ADJ+VERB	90.96	91.21	92.27	93.56	95.45	92.69
Average	92.42	94.25	95.88	96.23	97.71	

ตารางที่ 4.2 แสดงการเปรียบเทียบค่าความแม่นยำ ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ให้ค่าความแม่นยำสูงสุดร้อยละ 99.74 ในชุดข้อมูลที่มีเพียงคำคุณศัพท์ รองลงมาเป็นเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) มีค่าความแม่นยำร้อยละ 99.61 ในชุดข้อมูลที่มีเพียงคำคุณศัพท์ และเทคนิคนาอิวเบย์ (Naïve Bayes) ให้ค่าความแม่นยำต่ำสุดร้อยละ 86.95 ในชุดข้อมูลที่มีเพียงคำกริยา

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของเทคนิค ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นเทคนิคที่มีค่าความแม่นยำเฉลี่ยสูงที่สุดร้อยละ 97.71 รองลงมาเป็นเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) มีค่าความแม่นยำเฉลี่ยร้อยละ 96.23 และเทคนิคนาอิวเบย์ (Naïve Bayes) ให้ค่าความแม่นยำเฉลี่ยต่ำที่สุดร้อยละ 92.42

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของชุดข้อมูล จากชุดข้อมูลที่มีประเภทคำ เป็นคำกริยา คำวิเศษณ์ และคำคุณศัพท์ นอกจากนี้ยังได้นำคำวิเศษณ์รวมกับคำคุณศัพท์ และคำวิเศษณ์รวมกับคำคุณศัพท์และคำกริยา ผลปรากฏว่า ชุดข้อมูลที่มีเพียงคำคุณศัพท์ (ADJ) เป็นชุดข้อมูลที่มีค่าความแม่นยำเฉลี่ยสูงที่สุดร้อยละ 98.74 รองลงมาเป็นชุดข้อมูลที่มีเพียงคำวิเศษณ์ (ADV) มีค่าความแม่นยำเฉลี่ยร้อยละ 97.45 และชุดข้อมูลที่มีเพียงคำกริยา (VERB) ให้ค่าความแม่นยำเฉลี่ยต่ำที่สุดร้อยละ 91.01

4.2.2 ผลการเปรียบเทียบค่าความระลึกของแบบจำลอง

จากการเปรียบเทียบค่าความระลึก (Recall) ของแบบจำลองที่สร้างด้วยเทคนิคนาอิวเบย์ (Naïve Bayes), เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิค ป่าสุ่ม (Random

Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ในการจำแนกความคิดเห็นของประชาชนต่อออตติสติก ปรากฏผลดังตารางที่ 4.3

ตารางที่ 4.3 ค่าความระลึกของแบบจำลอง

ประเภทของคำ	เทคนิคการจำแนก					Average
	NB	C4.5	RF	K-NN	SVM	
VERB	97.00	97.06	100	99.07	97.73	98.17
ADJ	99.61	99.02	99.87	99.87	99.48	99.57
ADV	99.75	98.52	99.82	99.82	99.63	99.51
ADV+ADJ	99.75	98.27	99.88	99.44	99.81	99.43
ADV+ADJ+VERB	98.17	98.61	99.94	98.74	97.73	98.64
Average	98.86	98.30	99.90	99.39	98.88	

ตารางที่ 4.3 แสดงการเปรียบเทียบค่าความระลึก ผลปรากฏว่า เทคนิคป่าสุ่ม (Random Forest) ให้ค่าความระลึกสูงสุดร้อยละ 100 ในชุดข้อมูลที่มีเพียงคำกริยา รองลงมาเป็นเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) มีค่าความระลึกร้อยละ 99.87 ในชุดข้อมูลที่มีเพียงคำคุณศัพท์ และเทคนิคนาอิวเบย์ (Naïve Bayes) ให้ค่าความระลึกต่ำสุดร้อยละ 97.00 ในชุดข้อมูลที่มีเพียงคำกริยา

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของเทคนิค ผลปรากฏว่า เทคนิคป่าสุ่ม (Random Forest) เป็นเทคนิคที่มีค่าความระลึกเฉลี่ยสูงที่สุดร้อยละ 99.90 รองลงมาเป็นเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) มีค่าความระลึกเฉลี่ยร้อยละ 99.39 และเทคนิคต้นไม้ตัดสินใจแบบซี4.5 (Decision tree C4.5) ให้ค่าความระลึกเฉลี่ยต่ำที่สุดร้อยละ 98.30

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของชุดข้อมูล จากชุดข้อมูลที่มีประเภทคำ เป็นคำกริยา คำวิเศษณ์ และคำคุณศัพท์ นอกจากนี้ยังได้นำคำวิเศษณ์รวมกับคำคุณศัพท์ และคำวิเศษณ์รวมกับคำคุณศัพท์และคำกริยา พบว่า ชุดข้อมูลที่มีเพียงคำคุณศัพท์ (ADJ) เป็นชุดข้อมูลที่มีค่าความระลึกเฉลี่ยสูงที่สุดร้อยละ 99.57 รองลงมาเป็นชุดข้อมูลที่มีเพียงคำวิเศษณ์ (ADV) มีค่าความระลึกเฉลี่ยร้อยละ 99.51 และชุดข้อมูลที่มีเพียงคำกริยา (VERB) ให้ค่าความระลึกเฉลี่ยต่ำที่สุดร้อยละ 98.17

4.2.3 ผลการเปรียบเทียบค่าความถ่วงดุลของแบบจำลอง

จากการเปรียบเทียบค่าความถ่วงดุล (F-Measure) ของแบบจำลองที่สร้างด้วยเทคนิคนาอิวเบย์ (Naïve Bayes), เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิค ป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ในการจำแนกความคิดเห็นของประชาชนต่อออตโตสถาน ปรากฏผลดังตารางที่ 4.4

ตารางที่ 4.4 ค่าความถ่วงดุลของแบบจำลอง

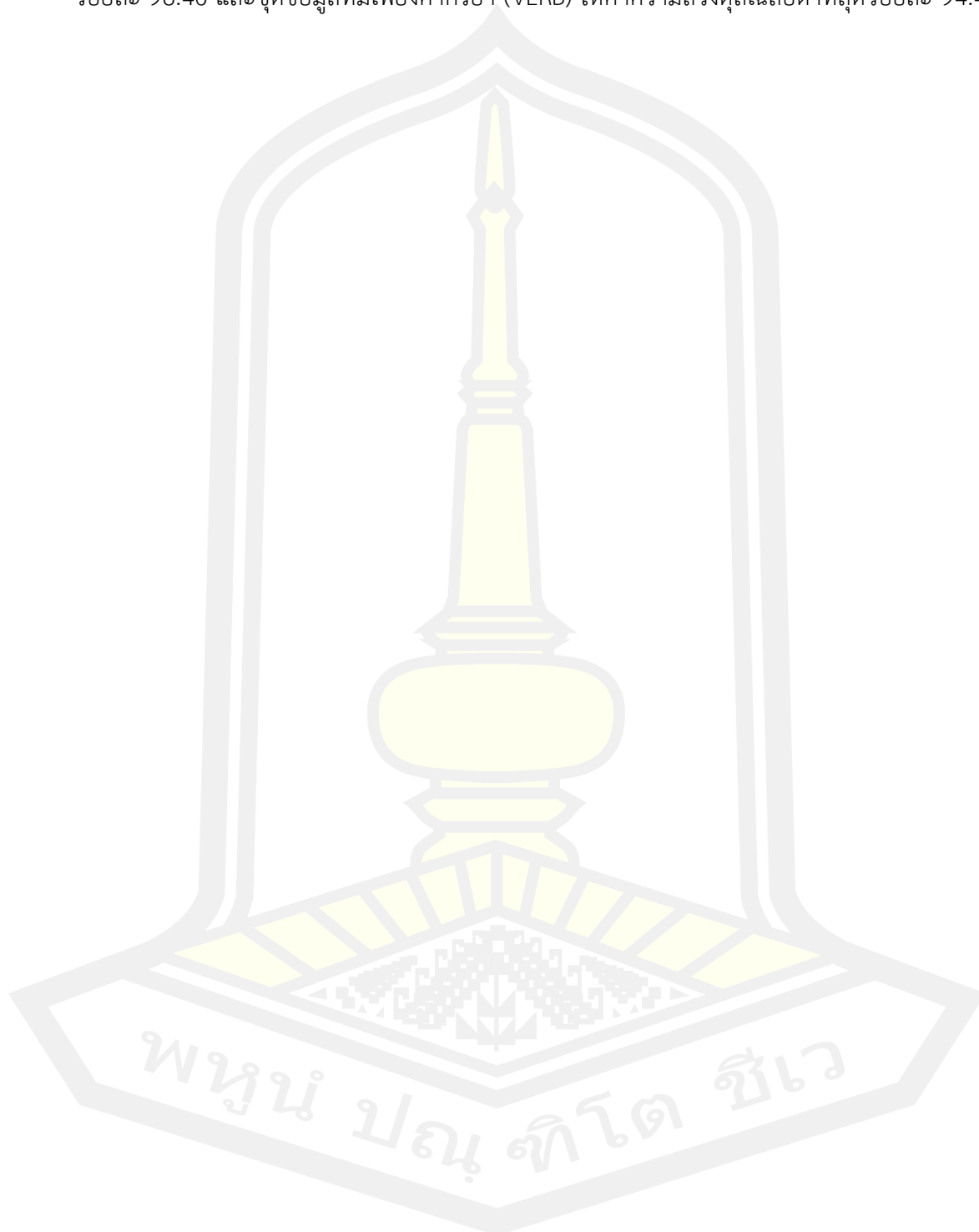
ประเภทของคำ	เทคนิคการจำแนก					Average
	NB	C4.5	RF	K-NN	SVM	
VERB	91.70	92.55	95.78	95.37	96.76	94.43
ADJ	98.00	98.69	99.71	99.74	99.61	99.15
ADV	96.95	97.80	99.05	99.11	99.39	98.46
ADV+ADJ	96.50	97.18	98.62	98.53	99.11	97.99
ADV+ADJ+VERB	94.43	94.77	95.95	96.08	96.58	95.56
Average	95.52	96.20	97.82	97.77	98.29	

ตารางที่ 4.4 แสดงการเปรียบเทียบค่าความถ่วงดุล ผลปรากฏว่า เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ให้ค่าความถ่วงดุลสูงสุดร้อยละ 99.74 ในชุดข้อมูลที่มีเพียงคำคุณศัพท์ รองลงมาเป็นเทคนิคป่าสุ่ม (Random Forest) มีค่าความถ่วงดุลร้อยละ 99.71 ในชุดข้อมูลที่มีเพียงคำคุณศัพท์ และเทคนิคนาอิวเบย์ (Naïve Bayes) ให้ค่าความถ่วงดุลต่ำสุดร้อยละ 91.70 ในชุดข้อมูลที่มีเพียงคำกริยา

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของเทคนิค ผลปรากฏว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นเทคนิคที่มีค่าความถ่วงดุลเฉลี่ยสูงที่สุดร้อยละ 98.29 รองลงมาเป็นเทคนิคป่าสุ่ม (Random Forest) มีค่าความถ่วงดุลเฉลี่ยร้อยละ 97.82 และเทคนิคนาอิวเบย์ (Naïve Bayes) ให้ค่าความถ่วงดุลเฉลี่ยต่ำที่สุดร้อยละ 95.52

เมื่อเปรียบเทียบจากค่าเฉลี่ยภาพรวมของชุดข้อมูล จากชุดข้อมูลที่มีประเภทคำ เป็นคำกริยา คำวิเศษณ์ และคำคุณศัพท์ นอกจากนี้ยังได้นำคำวิเศษณ์รวมกับคำคุณศัพท์ และคำวิเศษณ์รวมกับคำคุณศัพท์และคำกริยา พบว่า ชุดข้อมูลที่มีเพียงคำคุณศัพท์ (ADJ) เป็นชุดข้อมูลที่มีค่าความถ่วงดุล

เฉลี่ยสูงสุดร้อยละ 99.15 รองลงมาเป็นชุดข้อมูลที่มีเพียงคำวิเศษณ์ (ADV) มีค่าความถ่วงดุลเฉลี่ย ร้อยละ 98.46 และชุดข้อมูลที่มีเพียงคำกริยา (VERB) ให้ค่าความถ่วงดุลเฉลี่ยต่ำที่สุดร้อยละ 94.43



บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

จากผลการศึกษากระบวนการในการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออตติสติก และเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนที่มีต่อออตติสติก ในบทนี้ได้กล่าวถึง การสรุป อภิปรายผล และข้อเสนอแนะ

5.1 สรุป

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษากระบวนการทำเหมืองความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อออตติสติก และเพื่อสร้างและเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนที่มีต่อออตติสติก ด้วยเทคนิคเหมืองความคิดเห็น โดยเก็บรวบรวมความคิดเห็นที่เป็นภาษาไทย จากสื่อสังคมออนไลน์ ในเว็บไซต์พันทิป เฟซบุ๊ก และทวิตเตอร์ จำนวนทั้งหมด 1,766 ความคิดเห็น ตั้งแต่เดือนมกราคม 2556 ถึง เดือนกุมภาพันธ์ 2565 โดยแบ่งคำออกเป็น 2 กลุ่ม คือ คำแทนคุณลักษณะเชิงบวกและคำแทนคุณลักษณะเชิงลบ เพื่อนำไปสร้างแบบจำลองจำแนกความคิดเห็น ซึ่งได้นำเทคนิคการวิเคราะห์เหมืองข้อความมาทำการวิเคราะห์ความคิดเห็นโดยทำการเปรียบเทียบด้วยเทคนิคทั้งหมด 5 เทคนิค ได้แก่ เทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจแบบ ซี4.5 (Decision tree C4.5) เทคนิคป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และวัดประสิทธิภาพโดยใช้หลักการ 10-Fold Cross Validation จากผลการศึกษาสามารถสรุปได้ดังนี้

5.1.1 คำที่สามารถนำมาวิเคราะห์ มีจำนวน 2,344 คำ โดยแบ่งเป็น 5 ชุดข้อมูล ได้แก่ ชุดข้อมูลที่มีเพียงคำวิเศษณ์ ชุดข้อมูลที่มีเพียงคำคุณศัพท์ ชุดข้อมูลที่มีเพียงคำกริยา ชุดข้อมูลที่มีคำวิเศษณ์และคำคุณศัพท์ และชุดข้อมูลที่มีคำวิเศษณ์ร่วมกับคำคุณศัพท์และคำกริยา โดยพบว่า ประเภทคำที่ประชาชนส่วนใหญ่ใช้มากที่สุด ได้แก่ คำกริยา รองลงมาคือ คำวิเศษณ์ และประเภทคำที่ใช้น้อยที่สุดคือ คำคุณศัพท์ และจากความคิดเห็นทั้งหมดจำนวน 1,766 ความคิดเห็น สามารถแบ่งเป็น 2 กลุ่มความคิดเห็น ได้แก่ ความคิดเห็นเชิงบวกจำนวน 1,588 ความคิดเห็น และความคิดเห็นเชิงลบจำนวน 178 ความคิดเห็น แสดงให้เห็นว่าประชาชนส่วนใหญ่มีความคิดเห็นเชิงบวกต่อออตติสติกมากกว่าความคิดเห็นเชิงลบ

5.1.2 เทคนิคซ์พอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ให้ผลการทดสอบที่ดีที่สุด โดยให้ค่าประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำเฉลี่ยสูงสุดร้อยละ 97.71 ค่าความระลึกลเฉลี่ยสูงสุดร้อยละ 98.88 และค่าความถ่วงดุลเฉลี่ยสูงสุดร้อยละ 98.29 ส่วนเทคนิคที่ให้ค่าประสิทธิภาพของแบบจำลองต่ำที่สุดคือ เทคนิคเทคนิคนาอิวเบย์ (Naïve Bayes) โดยให้ค่าความแม่นยำเฉลี่ยต่ำที่สุดร้อยละ 92.42 ค่าความระลึกลเฉลี่ยต่ำที่สุดร้อยละ 98.86 และค่าความถ่วงดุลเฉลี่ยต่ำที่สุดร้อยละ 95.52 และผลสรุปการเปรียบเทียบจากชุดข้อมูล สามารถสรุปได้ว่า ชุดข้อมูลที่มีเพียงคำคุณศัพท์ (ADJ) เป็นชุดข้อมูลที่ให้ผลการทดสอบที่ดีที่สุด โดยมีค่าความแม่นยำเฉลี่ยสูงสุดร้อยละ 98.74 ค่าความระลึกลเฉลี่ยสูงสุดร้อยละ 99.57 และค่าความถ่วงดุลเฉลี่ยสูงสุดร้อยละ 99.15 ส่วนชุดข้อมูลที่ให้ค่าการเปรียบเทียบต่ำที่สุดคือ ชุดข้อมูลที่มีเพียงคำกริยา (VERB) โดยมีค่าความแม่นยำเฉลี่ยต่ำที่สุดร้อยละ 91.01 ค่าความระลึกลเฉลี่ยต่ำที่สุดร้อยละ 98.17 และค่าความถ่วงดุลเฉลี่ยต่ำที่สุดร้อยละ 94.43

5.2 อภิปรายผล

จากการศึกษากระบวนการทำความเข้าใจความคิดเห็นในการวิเคราะห์ความรู้สึกของประชาชนที่มีต่อ ออทิสติก และการสร้างและเปรียบเทียบแบบจำลองเพื่อจำแนกความรู้สึกของประชาชนที่มีต่อ ออทิสติก ด้วยเทคนิคเหมือนความคิดเห็น แสดงให้เห็นว่าประเภทคำที่มีผลต่อประสิทธิภาพของ แบบจำลองความรู้สึกของประชาชนต่อออทิสติก คือ คำคุณศัพท์ เป็นประเภทคำที่มีผลต่อการสร้าง แบบจำลองที่ดีที่สุดเมื่อเทียบกับคำวิเศษณ์ คำกริยา คำวิเศษณ์และคำคุณศัพท์ และคำวิเศษณ์ร่วมกับ คำคุณศัพท์และคำกริยา ซึ่งแตกต่างจากงานวิจัยของ ศรารุณี เกิดถาวร และจारी ทองคำ [11] ที่ใช้ ประเภทคำกริยา คำวิเศษณ์ และคำกริยาร่วมกับคำวิเศษณ์ มาสร้างแบบจำลองการจำแนกความคิดเห็นของคนไทยต่อโรคโควิด 19 พบว่า ชุดข้อมูลที่มีเฉพาะคำวิเศษณ์ และชุดข้อมูลที่มีคำวิเศษณ์ และคำกริยาร่วมกัน จะให้ผลดีที่สุดสำหรับการสร้างแบบจำลอง เนื่องจากเมื่อนำข้อมูลมาตัดคำแล้ว ทำให้ได้จำนวนคำน้อย จึงมีความจำเป็นต้องนำคำทั้งสองประเภทมารวมกันเพื่อให้ได้ผลที่ดีที่สุด แต่สำหรับงานวิจัยนี้เมื่อนำข้อมูลมาตัดคำแล้วทำให้ได้คำจำนวนมาก จึงไม่จำเป็นต้องนำประเภทของ คำมารวมกัน นอกจากนี้ยังทำให้เห็นว่าประชาชนชาวไทยโดยส่วนใหญ่มีความรู้สึกต่อออทิสติกในเชิง บวก ดังจะเห็นได้จากผลการทดลองการเปรียบเทียบการจำแนกกลุ่มของเทคนิคทั้ง 5 เทคนิคนี้ ผล การทดลองพบว่า เทคนิคซ์พอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นเทคนิคที่ให้ ประสิทธิภาพดีที่สุดเมื่อเปรียบเทียบกับเทคนิคนาอิวเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจแบบ

ที่ 4.5 (Decision tree C4.5) เทคนิคป่าสุ่ม (Random Forest) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ซึ่งสอดคล้องกับการวิจัยของ วสวัตต์ อินทร์แปลง [43] ที่ได้วิเคราะห์ความคิดเห็นต่อเกมมือถือพับจีด้วยเหมืองข้อความที่พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) มีประสิทธิภาพมากที่สุด จึงทำให้เห็นว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เหมาะที่จะนำมาใช้ในการวิเคราะห์ความคิดเห็นได้ดีที่สุด ส่วนเทคนิคนาอิวเบย์ (Naïve Bayes) ที่ให้ผลของค่าประสิทธิภาพของแบบจำลองน้อยที่สุด ซึ่งอาจจะเกิดจากลักษณะของข้อมูลที่ไม่เหมาะสมต่อการทำงานของโมเดล

5.3 ข้อเสนอแนะ

สำหรับการวิจัยครั้งต่อไป อาจพิจารณาค่าประเภทอื่นที่อาจจะส่งผลต่อประสิทธิภาพของแบบจำลองการจำแนกความรู้สึก เช่น คำนาม เพื่อให้การวิเคราะห์ประโยคจากคำบ่งชี้มีความครอบคลุม และเป็นการเพิ่มประสิทธิภาพในการจำแนกประเภทของคำในประโยคได้ดียิ่งขึ้น นอกจากนี้ การจำแนกกลุ่มของโมเดลทั้ง 5 อาจจะเกิดข้อผิดพลาดขึ้นในกระบวนการตัดคำและกำจัดคำหยุด เนื่องจากข้อมูลที่เป็นข้อความภาษาไทยมีลักษณะของภาษาพูดมากกว่าภาษาเขียนทำให้โครงสร้างประโยคมีความซับซ้อนคล้ายกับภาษาธรรมชาติของมนุษย์ที่คอมพิวเตอร์ไม่สามารถวิเคราะห์ได้หรือวิเคราะห์ผิดพลาด อีกทั้งเรื่องความหมายของคำยังพบว่า คำหนึ่งคำมีหลายความหมาย ทำให้การวิเคราะห์ความหมายในเชิงบวกและเชิงลบทำได้ยาก จึงควรอาศัยผู้เชี่ยวชาญทางด้านภาษาไทยมาช่วยวิเคราะห์คำ

บรรณานุกรม



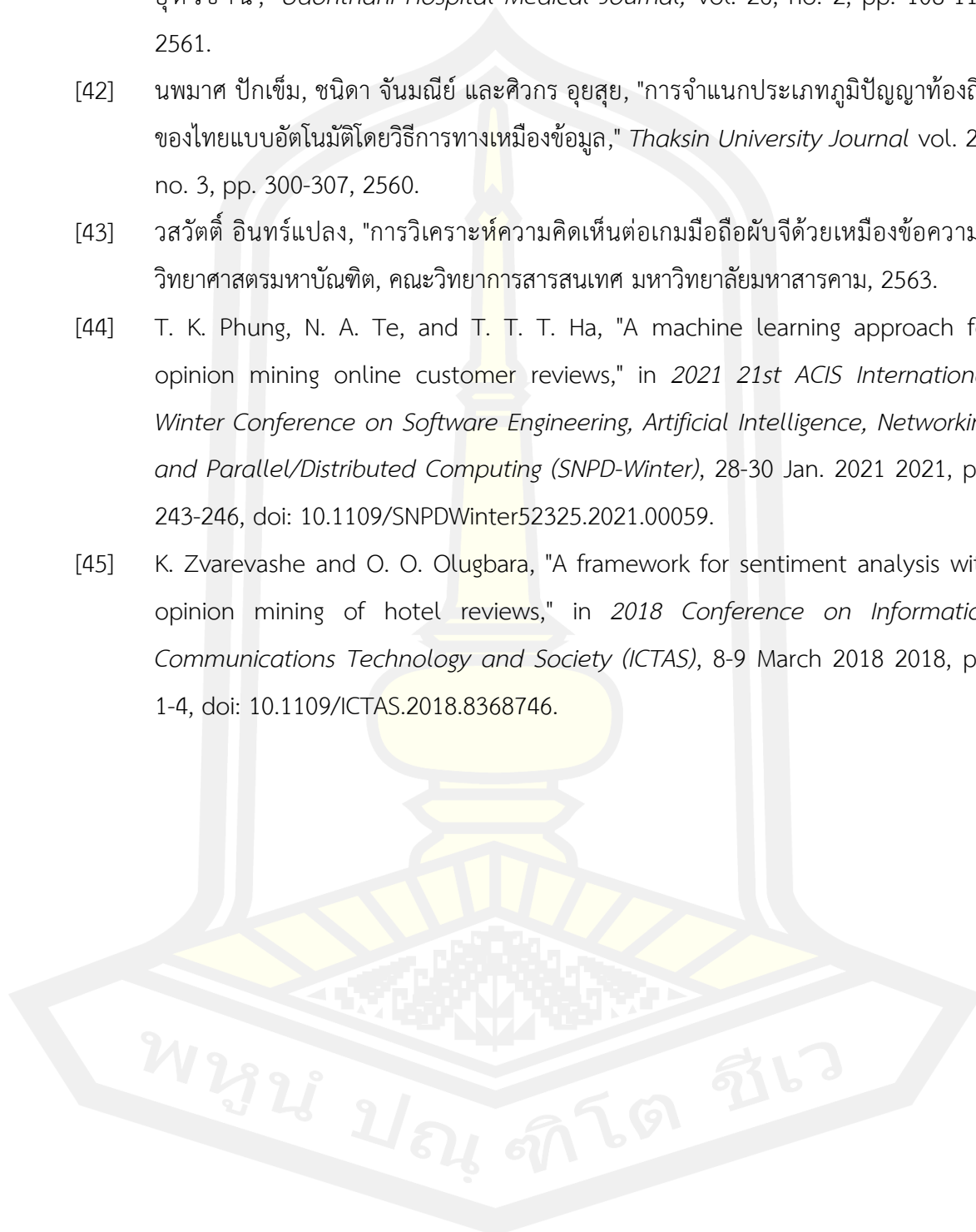
- [1] กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ. "สถานการณ์คนพิการ 30 มิถุนายน 2564 (รายไตรมาส)." https://dep.go.th/images/uploads/files/situation_June64.pdf (accessed 12 กรกฎาคม 2564).
- [2] กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ. "ประกาศกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ เรื่อง ประเภทและหลักเกณฑ์ความพิการ (ฉบับที่ 2) พ.ศ.2555." <http://web1.dep.go.th/sites/default/files/files/services/22.PDF> (accessed 12 กรกฎาคม 2564).
- [3] ราชกิจจานุเบกษา. "พระราชบัญญัติคุ้มครองเด็ก พ.ศ. 2546." <http://www.ratchakitcha.soc.go.th/DATA/PDF/00130679.PDF> (accessed 12 กรกฎาคม 2564).
- [4] วนิตา ชนินทยุทธวงศ์ และคณะ, เด็กออทิสติก คู่มือสำหรับครู, 4 ed. นนทบุรี: โรงพิมพ์ชุมชนสหกรณ์การเกษตรแห่งประเทศไทย จำกัด, 2557.
- [5] สุมาลี ตีจงติง และคณะ, "เจตคติของผู้ปกครองต่อเด็กที่มีความบกพร่องทางการได้ยิน," *Ramathibodi Medical Journal*, vol. 41, no. 1, pp. 74-81, 2561.
- [6] S. de Laat, E. Freriksen, and M. P. J. Vervloed, "Attitudes of children and adolescents toward persons who are deaf, blind, paralyzed or intellectually disabled," *Research in Developmental Disabilities*, vol. 34, no. 2, pp. 855-863, 2013/02/01/ 2013, doi: <https://doi.org/10.1016/j.ridd.2012.11.004>.
- [7] วุฒินันท์ วัชรโกมลพันธุ์ และมณฑิยา รัตนศิริวงศ์วุฒิ, "การพัฒนาแบบจำลองในการจำแนกความพึงพอใจเกี่ยวกับการใช้บริการระบบขนส่งพัสดุในประเทศไทย," *The Sixteenth National Conference on Computing and Information Technology*, pp. 316-321, 2563.
- [8] พิษณุ พรมลา และจรัญ แสนราช, "การเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจเกี่ยวกับการจัดการเรียนการสอนด้วยกระบวนการวิเคราะห์ความรู้สึก โดยใช้เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล," *KKU Research Journal (Graduate Studies)*, vol. 20, no. 4, pp. 140-149, 2563.
- [9] สุภัตตรา สมเจตนา และจารี ทองคำ, "การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร," *Journal of Science and Technology, Ubon Ratchathani University* vol. 23, no. 1, pp. 21-30, 2564.

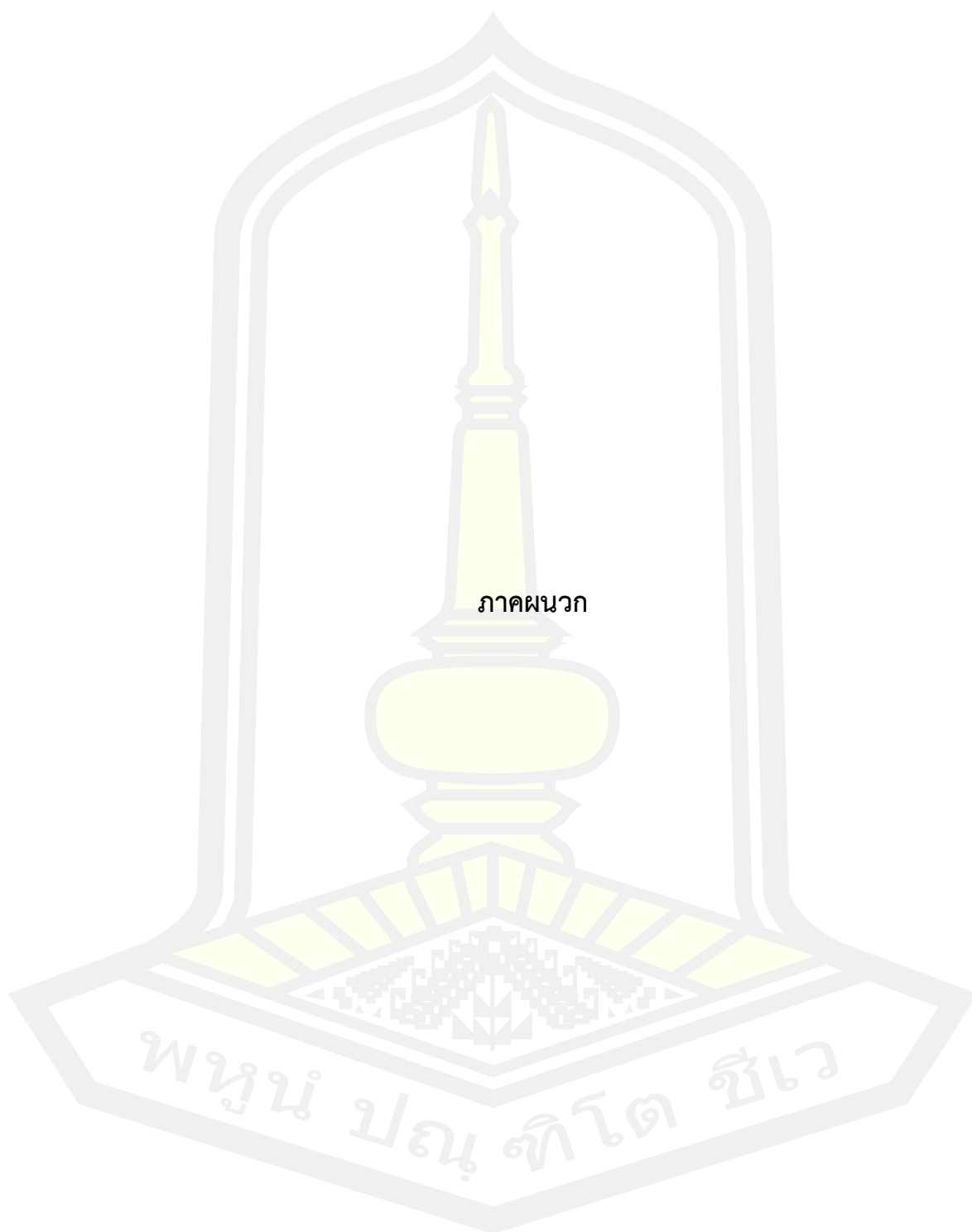
- [10] A. I. Saad, "Opinion Mining on US Airline Twitter Data Using Machine Learning Techniques," in *2020 16th International Computer Engineering Conference (ICENCO)*, 29-30 Dec. 2020, pp. 59-63, doi: 10.1109/ICENCO49778.2020.9357390.
- [11] ศราวุฒิ เกิดถาวร และจारी ทองคำ, "ชนิดของคำภาษาไทยที่มีผลต่อประสิทธิภาพของแบบจำลองการจำแนกความคิดเห็น:กรณีศึกษาของความคิดเห็นของคนไทยต่อโรคโควิด 19," *Journal of Science and Technology, Ubon Ratchathani University*, vol. 24, no. 1, pp. 100-112, 2565.
- [12] P. Tran, N. Trieu, N. Vu, H. Ta Thanh, and H. Huynh, "Effective Opinion Words Extraction for Food Reviews Classification," *International Journal of Advanced Computer Science and Applications*, vol. 11, 01/01 2020, doi: 10.14569/IJACSA.2020.0110755.
- [13] S. A. S. Neshan and R. Akbari, "A Combination of Machine Learning and Lexicon Based Techniques for Sentiment Analysis," in *2020 6th International Conference on Web Research (ICWR)*, 22-23 April 2020, pp. 8-14, doi: 10.1109/ICWR49608.2020.9122298.
- [14] รัตนา เสงส์สวัสดิ์, "ออทิสติก," *Ramkhamhaeng University Journal: The Graduate School* vol. 23, no. 3, pp. 76-88, 2549.
- [15] วนิตา ชนินทยุทรวงศ์ และคณะ, *เด็กออทิสติก คู่มือสำหรับพ่อแม่/ผู้ปกครอง*, 3 ed. นนทบุรี: โรงพิมพ์ชุมชนสหกรณ์การเกษตรแห่งประเทศไทย จำกัด, 2557.
- [16] อนัญญา สิริรัตนันท์ และธีรรัตน์ แทนขำ, *แนวทางการดูแลโรคทางจิตเวชเด็กและวัยรุ่น 4 โรคหลัก (สติปัญญาบกพร่อง ออทิสติก สมาธิสั้น บกพร่องทางการเรียน) สำหรับเครือข่ายบริการสาธารณสุข*. นนทบุรี: โรงพิมพ์ชุมชนสหกรณ์การเกษตรแห่งประเทศไทย จำกัด, 2559.
- [17] อัญชรส ทองเพ็ชร และคณะ, "ผลของการบำบัดทางการพยาบาลระยะแรก ต่อความรุนแรงของอาการเด็กออทิสติกวัยก่อนเรียน," *Thai Journal of Nursing and Midwifery Practice*, vol. 6, no. 1, pp. 19-35, 2562.
- [18] นรุตม์ แพงพรมมา, จินตนา ยูนิพันธุ์ และรัชนิกร อุปเสน, "ผลของพฤติกรรมบำบัดโดยครอบครัวมีส่วนร่วมต่อพฤติกรรมก้าวร้าวของเด็กออทิสติก," *The Journal of Psychiatric Nursing and Mental Health*, vol. 33, no. 2, pp. 91-104, 2559.

- [19] วสวัตต์ อินทร์แปลง และจารี ทองคำ, "การวิเคราะห์ความคิดเห็นต่อเกมมือถือนับพันด้วยเหมืองข้อความ," *Journal of Science and Technology Mahasarakham University*, vol. 39, no. 5, pp. 523-531, 2563.
- [20] George H. John and Pat Langley, " Estimating Continuous Distributions in Bayesian Classifiers.," presented at the Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995.
- [21] ภูริพัทธ์ ทองคำ, "อัลกอริทึมแบบรวมสำหรับการเลือกคุณสมบัติของข้อมูล," *วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)*, คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์, 2559.
- [22] ชีรยุทธ คุณสุข และจารี ทองคำ, "กระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า," *RMUTI Journal Science and Technology* vol. 13, no. 1, pp. 129-143, 2563.
- [23] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA.: Morgan Kaufmann, 1993.
- [24] นุชนาฏ ปิ่นเมือง และจารี ทองคำ, "การจำแนกความคิดเห็นของคนไทยเกี่ยวกับสื่อออนไลน์โดยใช้การทำเหมืองข้อความ," *J Sci Technol MSU*, vol. 37, no. 3, pp. 372-379, 2560.
- [25] สุกัสสรฯ สมเจตนา, "การทำเหมืองความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร," *วิทยาศาสตร์มหาบัณฑิต, คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม*, 2564.
- [26] วิภาวรรณ บัวทอง, การทำเหมืองข้อมูล. ภูเก็ต: คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต, 2559.
- [27] อุกฤษฏ์ ศรีสุข และจารี ทองคำ, "การเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรค," *J Sci Technol MSU*, vol. 40, no. 2, pp. 157-163, 2563.
- [28] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [29] ลลิตี อินทราราม, "การวิเคราะห์รูปแบบการบุกรุกข้อมูลบนเครือข่าย โดยใช้การจำแนกข้อมูลของแบบจำลอง แรนดอมฟอเรส ซัพพอร์ตเวกเตอร์แมชชีน และ แบบผสมผสานกัน," *วิทยาศาสตร์มหาบัณฑิต, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ*, 2554.
- [30] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms.," *Machine Learning*, vol. 6, pp. 37-66, 1991.

- [31] สุพจน์ เสงพระพรหม และสถาพร ประณีธานวิทยา, "การเลือกคุณลักษณะแบบอัตราส่วน สัญญาณต่อสัญญาณรบกวนสำหรับการจำแนกประเภทข้อมูลหลายกลุ่ม," *Journal of Thai Interdisciplinary Research* vol. 11, no. 4, pp. 41-48, 2559.
- [32] สมศักดิ์ ศรีสวการย์ and สมัย ศรีสว, "การวิเคราะห์เหมืองความคิดเห็นโดยใช้เทคนิคการสกัดคำ," *Journal of Applied Information Technology*, vol. 6, no. 2, pp. 95-104, 2563.
- [33] สายชล สีนสมบูรณ์ทอง, การทำเหมือง = *Data Mining*. กรุงเทพมหานคร: จามจุรีโปรดักท์, 2558.
- [34] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, and K.R.K. Murthy, "Improvements to Platt's SMO Algorithm for SVM Classifier Design.," *Neural Computation*, vol. 13, no. 3, pp. 637-649, 2001.
- [35] วชิรวิวัฒน์ จิตต์สกุล และสุนันทา สดสี, "การวิเคราะห์การจำแนกข้อความด้วยการเปรียบเทียบความเสถียรของอัลกอริทึม," *Sripatum Review of Science and Technology* vol. 9, pp. 19-31, 2560.
- [36] ธรณินทร์ สัจจวิริยทรัพย์, "เทคนิคการพยากรณ์แบบผสมของซัพพอร์ตเวกเตอร์แมชชีนและขั้นตอนวิธีเชิงพันธุกรรมสำหรับการพยากรณ์ปริมาณความต้องการพลังงานไฟฟ้าสูงสุดรายปีของประเทศไทย," *The Journal of King Mongkut's University of Technology North Bangkok* vol. 27, no. 3, pp. 453-465, 2560.
- [37] ประพัฒน์ พรหมน้ำอ่าง, วสุวรรธน์ พงศ์ขจร และนิเวศ จิระวิชิตชัย, "การจำแนกกลุ่มข้อความรีวิว โดยใช้เทคนิคเหมืองข้อมูล," *Science and Technology RMUTT Journal*, vol. 6, no. 1, pp. 94-101, 2559.
- [38] นฤมล ทวีพันธ์, "ความเครียด ภาวะการดูแล และทัศนคติที่มีต่อเด็กออทิสติกของครูที่ดูแลเด็กออทิสติกในโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ ศูนย์วิจัยและพัฒนาการศึกษา" *วิทยาศาสตร์มหาบัณฑิต, คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย*, 2555.
- [39] อรพิม วิสารทานนท์, "ทัศนคติและความต้องการของครูปฐมวัยเขตกรุงเทพมหานครที่มีต่อเด็กปฐมวัยออทิสติก," *การศึกษามหาบัณฑิต (การศึกษาปฐมวัย)*, มหาวิทยาลัยศรีนครินทรวิโรฒ, 2552.
- [40] ขวัญใจ อธิปัญญาวงศ์, ศิริวรรณ ศิริบุญ และชัยชนะ นิมนวล, "เจตคติต่อการเป็นเพื่อนกับนักเรียนออทิสติกเรียนร่วมบางเวลาในโรงเรียนระดับประถมศึกษา," *Chulalongkorn Medical Journal*, vol. 53, no. 3, pp. 301-312, 2558.

- [41] มธุวดี อิงศิริโรรัตน์, "ความรู้และทัศนคติของผู้ปกครองเด็กกลุ่มอาการออทิซึมในโรงพยาบาลอุดรธานี," *Udonthani Hospital Medical Journal*, vol. 26, no. 2, pp. 108-118, 2561.
- [42] นพมาศ ปักเข็ม, ชนิตา จันมณี และศิวกร อุษยสุข, "การจำแนกประเภทภูมิปัญญาท้องถิ่นของไทยแบบอัตโนมัติโดยวิธีการทางเหมืองข้อมูล," *Thaksin University Journal* vol. 20, no. 3, pp. 300-307, 2560.
- [43] วสวัตต์ อินทร์แปลง, "การวิเคราะห์ความคิดเห็นต่อเกมมือถือนับจี้ด้วยเหมืองข้อความ," วิทยาศาสตร์มหาบัณฑิต, คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม, 2563.
- [44] T. K. Phung, N. A. Te, and T. T. T. Ha, "A machine learning approach for opinion mining online customer reviews," in *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, 28-30 Jan. 2021 2021, pp. 243-246, doi: 10.1109/SNPDWinter52325.2021.00059.
- [45] K. Zvarevashe and O. O. Olugbara, "A framework for sentiment analysis with opinion mining of hotel reviews," in *2018 Conference on Information Communications Technology and Society (ICTAS)*, 8-9 March 2018 2018, pp. 1-4, doi: 10.1109/ICTAS.2018.8368746.





ภาคผนวก

พหุมนุ ปณฺ ทิโต ชีเว



ภาคผนวก ก
คำคุณลักษณะที่ใช้ในงานวิจัย

พหุบัน ปณฺ ทิตโต สีเว

คำคุณลักษณะที่ประชาชนพูดมากกว่า 10 ครั้ง

คำคุณลักษณะเชิงลบที่ประชาชนพูดถึงมากกว่า 10 ครั้ง มีจำนวน 133 คำ

ลำดับ	คำ
1	โง่
2	ต่ำ
3	แย
4	หนัก
5	เกิน
6	เกือบ
7	ไกล
8	ซ้ำ
9	ซ้ำ
10	ซ้ำซ้ำ
11	น้อย
12	นาน
13	ยาก
14	รุนแรง
15	แรง
16	ไร้สาระ
17	ห่าง
18	เหนียว
19	กลับ
20	กลัว

ลำดับ	คำ
21	กลาย
22	กังวล
23	แก้ง
24	โกรธ
25	ขาด
26	เข้าข่าย
27	ควบคุม
28	คอย
29	คุม
30	เครียด
31	งง
32	งด
33	ชน
34	ต่ำ
35	ดี
36	ดูถูก
37	โดน
38	ตาย
39	ตำหนิ
40	ติด

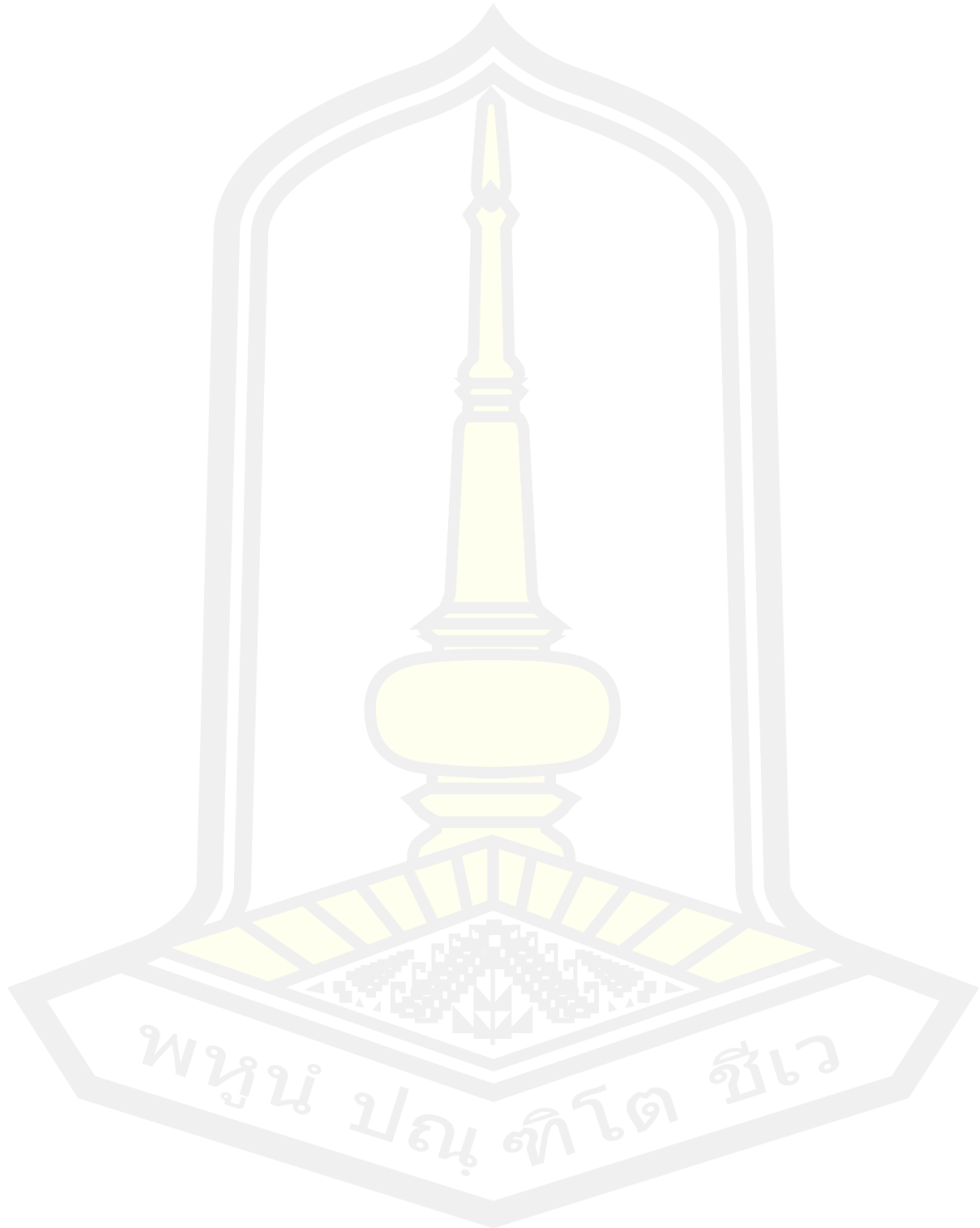
ลำดับ	คำ
41	ดี
42	แตกต่าง
43	แตกต่างกัน
44	ทน
45	ท้อ
46	ทำใจ
47	ทำไม่ได้
48	ทำร้าย
49	ทิ้ง
50	น้ำตาไหล
51	บ่น
52	บังคับ
53	เบื้อ
54	ป่วย
55	แปลก
56	ผิด
57	ผิดปกติ
58	ฝาก
59	พลาด
60	พัง
61	พูดไม่ได้
62	โมโห
63	ไม่

ลำดับ	คำ
64	ไม่กล้า
65	ไม่เก่ง
66	ไม่เข้าใจ
67	ไม่ควร
68	ไม่คอยได้
69	ไม่คิด
70	ไม่เคย
71	ไม่ชอบ
72	ไม่ใช่
73	ไม่ดี
74	ไม่ได้
75	ไม่ได้เลย
76	ไม่ต้อง
77	ไม่ทราบ
78	ไม่ทำ
79	ไม่ning
80	ไม่บอก
81	ไม่ปกติ
82	ไม่เป็น
83	ไม่พอ
84	ไม่พูด
85	ไม่ยอมรับ
86	ไม่รู้

ลำดับ	คำ
87	ไม่รู้จัก
88	ไม่รู้เรื่อง
89	ไม่สนใจ
90	ไม่สบตา
91	ไม่สามารถ
92	ไม่เห็น
93	ไม่หาย
94	ไม่เหมือน
95	ไม่เหมือนกัน
96	ไม่ไหว
97	ไม่ไหว
98	ไม่อยาก
99	ไม่อยากจะ
100	ไม่เอา
101	ยกเว้น
102	ยอม
103	ย่ำ
104	แยก
105	รบกวน
106	รอ
107	ร้องไห้
108	ระราน
109	รังแก

ลำดับ	คำ
110	รำคาญ
111	รีบ
112	ลง
113	ลด
114	ล้อเลียน
115	ลาออก
116	ลำบาก
117	ลื้ม
118	เล็ก
119	วน
120	โววาย
121	สงสัย
122	สงสัย
123	เสีย
124	เสียใจ
125	หงุดหงิด
126	หนี
127	หมด
128	หยุด
129	ห้าม
130	หาย
131	เหนื่อย
132	อาละวาด

133	แบบ
-----	-----



คำคุณลักษณะเชิงบวกที่ประชาชนพูดมากกว่า 10 ครั้ง มีจำนวน 172 คำ

ลำดับ	คำ
1	ใกล้ชิด
2	แข็งแรง
3	คนปกติ
4	ใจเย็น
5	ฉลาด
6	ดีดี
7	เด็กปกติ
8	เด็กพิเศษ
9	เต็ม
10	ทั่วไป
11	ธรรมดา
12	น่ารัก
13	พอ
14	ยิ่งใหญ่
15	สวย
16	สำคัญ
17	สูง
18	ใหม่
19	เก่งกว่า
20	ใกล้
21	ค่อยค่อย

ลำดับ	คำ
22	ง่าย
23	จริง
24	ชัด
25	ชัดเจน
26	ซึ้ง
27	ดี
28	ดีกว่า
29	ดีขึ้น
30	ต่อไป
31	ตามปกติ
32	เต็มที่
33	ทันที
34	ที่สุด
35	น่าสงสาร
36	น่าเห็นใจ
37	แน่
38	แน่นอน
39	ปกติ
40	เพื่อ
41	พร้อม
42	พอดี

ลำดับ	คำ
43	พอสมควร
44	พิเศษ
45	มาก
46	มากกว่า
47	มากขึ้น
48	เร็ว
49	ไว
50	สบาย
51	สบายใจ
52	สม่ำเสมอ
53	สุดยอด
54	เสมอ
55	เหมือนกัน
56	อย่างดี
57	เอาจริง
58	กตัญญู
59	กล้า
60	กอด
61	เก่ง
62	แก้
63	แก้ไข
64	ขอ
65	ขอโทษ

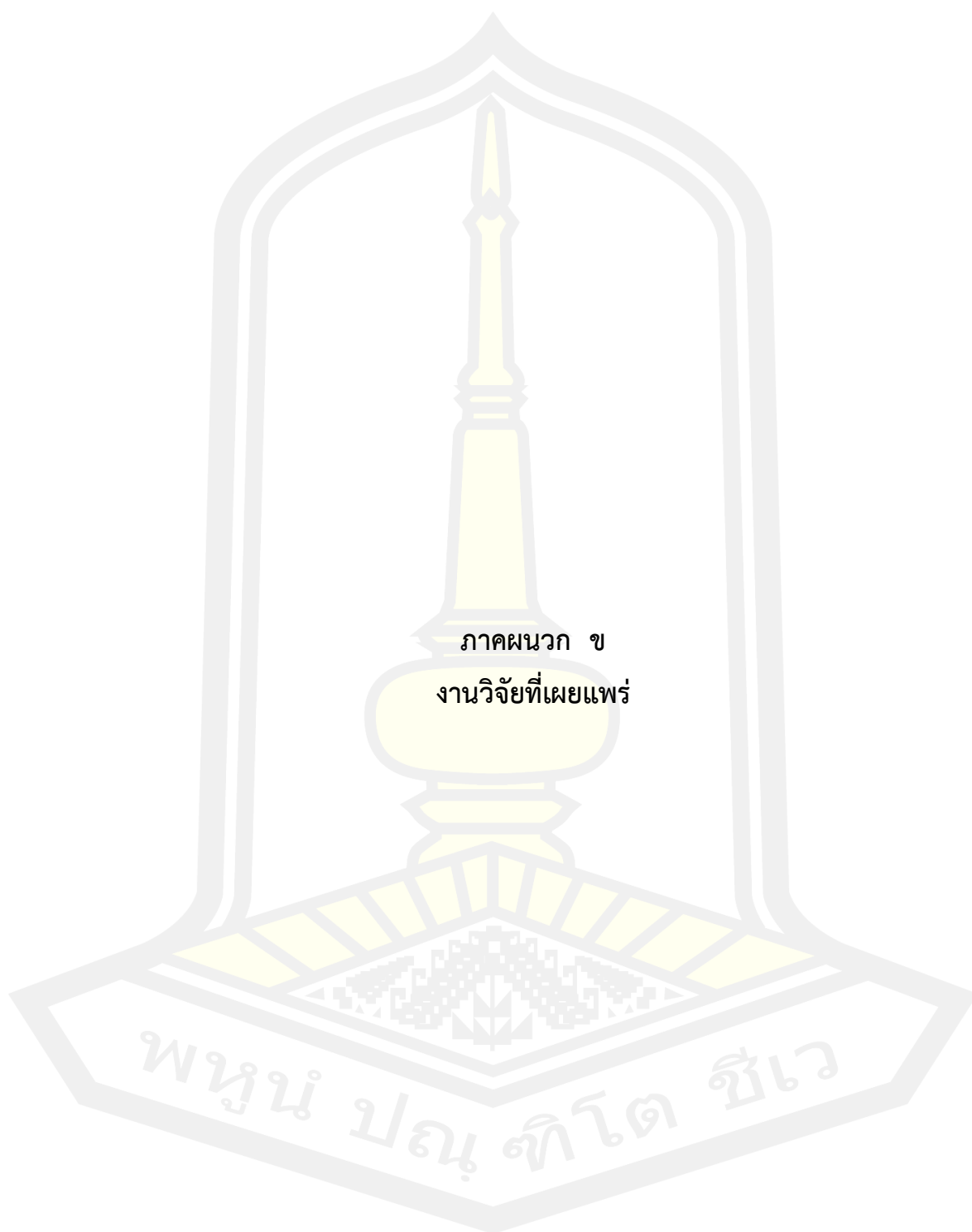
ลำดับ	คำ
66	ขอบคุณ
67	ขออนุญาต
68	เข้มแข็ง
69	เข้าใจ
70	เข้ารอบ
71	เข้าสังคม
72	เขียนได้
73	ควร
74	คิดถึง
75	เคย
76	จบ
77	จำ
78	ชวน
79	ช่วย
80	ช่วยกัน
81	ช่วยเหลือ
82	ชอบ
83	ชื่นชม
84	เชื้อ
85	ใช่
86	ดีใจ
87	ดูแล
88	ได้

ลำดับ	คำ
89	ไต่ยิน
90	ไต่รับ
91	ตรง
92	ตรวจ
93	ต้องการ
94	ต่อสู้อยู่
95	ตั้งใจ
96	ตามใจ
97	เตรียม
98	เติบโต
99	โต
100	ถูก
101	ถูกต้อง
102	แถม
103	ทราบ
104	นับถือ
105	นึกถึง
106	เน้น
107	แนะนำ
108	บวก
109	บำบัด
110	แบ่งปัน
111	ประสบความสำเร็จ

ลำดับ	คำ
112	ปรับ
113	ปรับตัว
114	ปล่อย
115	เป็นปกติ
116	เปลี่ยน
117	เปิด
118	เปิดใจ
119	ผ่าน
120	ฝึก
121	ฝึกฝน
122	พยายาม
123	พัฒนา
124	เพิ่ม
125	ภูมิใจ
126	มั่นใจ
127	มี
128	ไม่ได้ดูถูก
129	ไม่ได้เป็น
130	ไม่นาน
131	ไม่ผิด
132	ไม่ว่า
133	ยอมรับ
134	ยินดี

ลำดับ	คำ
135	ยิ้ม
136	เยี่ยม
137	ร่วม
138	รวย
139	รัก
140	รักษา
141	รับรู้
142	จำเริญ
143	รู้
144	รู้จัก
145	รู้เรื่อง
146	เริ่ม
147	เรียนรู้
148	ส่งเสริม
149	สนใจ
150	สนับสนุน
151	สนุก
152	สบตา
153	สังเกต
154	สามารถ

ลำดับ	คำ
155	สำเร็จ
156	สู้
157	เสร็จ
158	เสริม
159	ใส่ใจ
160	หวัง
161	หัวเราะ
162	เห็นใจ
163	เห็นด้วย
164	เหมาะสม
165	เหมือน
166	ให้
167	ให้ได้
168	อดทน
169	อย่ากังวล
170	เอาใจช่วย
171	เอาใจใส่
172	โอเค



ภาคผนวก ข
งานวิจัยที่เผยแพร่

พหุณฺ์ ปณฺุทฺิโต สิวเว

ประวัติผู้เขียน

ชื่อ	นางสาวชุตติกาญจน์ เพชรรักษ์
วันเกิด	07/05/2536
สถานที่เกิด	โรงพยาบาลค่ายประจักษ์ศิลปาคม
สถานที่อยู่ปัจจุบัน	117 หมู่3 ตำบลโนนสะอาด อำเภอบ้านม่วง จังหวัดสกลนคร 47140
สถานที่ทำงานปัจจุบัน	ศูนย์บริการสนับสนุนนักศึกษาพิการ มหาวิทยาลัยราชภัฏสกลนคร
ประวัติการศึกษา	พ.ศ. 2558 ปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร พ.ศ. 2565 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

พูน ปณ ทัโต ชีเว