

การตรวจจับการกลั่นแกล้งบนอินเทอร์เน็ตของกลุ่มบุคคลที่มีความ หลากหลายทางเพศในภาษาไทย

LGBT Cyberbullying Detection in Thai Language

วีระพงษ์ ทรัพย์สุครี¹, ศุภกิจ จรูญเพ็ญ²

สาขาวิชาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

Emails: Wiraphong.srap@gmail.com, 64011212100@msu.ac.th

บทคัดย่อ

การเข้าถึงสื่อสังคมออนไลน์เป็นเรื่องง่าย และกลายเป็นสิ่งจำเป็นในการดำรงชีวิต ด้านดี ของสื่อสังคมออนไลน์ คือ สามารถใช้ติดต่อสื่อสารถึงกัน ส่วนผลในด้านลบคือ การที่สื่อสังคมออนไลน์ถูกใช้เป็นอีกช่องทางหนึ่งในการคุกคาม กลั่นแกล้ง กัน การกลั่นแกล้งบนโซเชียลมีเดียมีหลากหลายรูปแบบ สามารถสรุปได้ว่า คือการคุกคามหรือทำร้ายผู้อื่น โดยใช้เครื่องมือทางอิเล็กทรอนิกส์ เช่น อีเมล์ ข้อความโต้ตอบแบบทันที ข้อความสั้น ห้องสนทนาออนไลน์ เพื่อให้ผู้อื่นได้รับความเสียหาย เกิดความหวาดกลัว และรู้สึกสั่นหวั่นในชีวิต โดยผู้ที่ตกเป็นเหยื่อยากที่จะ ป้องกันตนเองและถูกกระทำซ้ำไปซ้ำมา งานวิจัยนี้จึงจัดทำขึ้นเพื่อนำระบบประมวลผลภาษาธรรมชาติ ซึ่งเป็นสาขาวิชายอดของภาษาศาสตร์ วิทยาการคอมพิวเตอร์ และปัญญาประดิษฐ์ที่เกี่ยวข้องกับการปฏิสัมพันธ์ระหว่างคอมพิวเตอร์และภาษาของมนุษย์ ซึ่งทำหน้าที่เป็นอีกหนึ่งกลไกสำคัญ ในการนำมาใช้ในการเลือกเนื้อหา ข้อความ การตอบโต้กันบนสื่อสังคมออนไลน์ เพื่อประมวลผลจากและได้อย่างมีประสิทธิภาพเสนอวิธีการในการตรวจจับข้อมูลการกลั่นแกล้งบนโซเชียลมีเดียในข้อความภาษาไทยโดยการนำข้อมูลที่เกิดขึ้นในสังคมออนไลน์บน Twitter มา ทำการประมวลผลผ่านกระบวนการและขั้นตอนในการทำความสะอาดข้อมูล เพื่อฝึกสอนโมเดล 4 อัลกอริทึม คือ K-Nearest Neighbors, Naive Bayes, SVM, CNN และสร้างตัวชี้วัดประสิทธิภาพและความแม่นยำของแบบจำลอง ด้วย Precision, Recall และ f1-score โดย CNN มีความแม่นยำอยู่ที่ร้อยละ 99.35

คำสำคัญ การกลั่นแกล้งบนโซเชียลมีเดีย, ระบบประมวลผลภาษาธรรมชาติ, การวิเคราะห์ภาษาธรรมชาติ, การตรวจจับข้อมูล

1. บทนำ

ทุกวันนี้การสื่อสารเป็นไปอย่างรวดเร็วด้วยเทคโนโลยีและการใช้บริการผ่านอินเทอร์เน็ต หรือ Social Network จึงจำเป็นต้องมีการบริหารข้อมูลขนาดใหญ่ หรือ Big Data ให้เกิดประสิทธิภาพสูงสุดสำหรับองค์กร ปัจจุบันการเผยแพร่สื่อมีทางเลือกมีหลากหลายในการใช้แพลตฟอร์มมากขึ้นกว่าเดิม รวมถึงเราจะเห็นวิธีใหม่ของการทำการตลาด หรือโฆษณาที่ไม่ได้ใช้การผลิตแบบเดิมเพราต้องลงทุนมหาศาล ขณะที่สื่อทางโทรทัศน์ส่วนใหญ่ในปัจจุบันเริ่มมีการใช้ Facebook Live หรือ YouTube แต่ยังมีทางเลือกแพลตฟอร์มอื่น ๆ อีกที่สื่อในประเทศไทยสามารถใช้เป็นช่องทางกระจายข่าวได้อีก เช่น Twitter เป็นต้น สังคมออนไลน์หรือโซเชียลมีเดีย จึงเป็นสถานที่ที่มีข่าว พรีสเซปดอยู่ตามเว็บไซต์มากมาย ทำให้ทั้ง Facebook Twitter YouTube และ Instagram กลายเป็นเครือข่ายบนสังคมออนไลน์ที่ได้รับความนิยมในปัจจุบัน

การเข้าถึงสื่อสังคมออนไลน์เป็นเรื่องง่าย และกลายเป็นสิ่งจำเป็นในการดำรงชีวิต ด้านดีของโซเชียลมีเดีย คือ สามารถใช้ติดต่อสื่อสารถึงกัน ส่วนผลในด้านลบคือ การที่โซเชียลมีเดีย ถูกใช้เป็นอีกช่องทางหนึ่งในการคุกคาม รังแกกัน การรังแกกันบนโลกไซเบอร์มีหลากหลายรูปแบบ การนิยามความหมายจึงมีความแตกต่างกันขึ้นอยู่กับบริบทของสังคมและขอบเขตของการศึกษา ความหมายของการรังแกกันบนโซเชียลมีเดีย (Cyberbullying)

สามารถสรุปใจความได้ว่า การรังแกกันบนโซเชียลมีเดีย คือ การคุยกันหรือทำร้ายผู้อื่น โดยใช้เครื่องมือทางอิเล็กทรอนิกส์ เช่น อีเมล์ ข้อความโต้ตอบแบบทันที ข้อความสั้น ห้องสนทนากลุ่มออนไลน์ เพื่อให้ผู้อื่นได้รับความเสียหาย เกิดความหวาดกลัว และรู้สึกสั่นหวั่นในชีวิต (Strom & Strom, 2006) หรือ การรังแกรูปแบบใหม่ที่แตกต่างจากการรังแกแบบดั้งเดิม ส่วนมากมักเกิดนอกโรงเรียน โดยสอดคล้องกับการศึกษาเกี่ยวกับองค์ประกอบของการรังแกบนโลกไซเบอร์แบบดั้งเดิม ซึ่งมี 3 องค์ประกอบ คือ 1) ผู้กระทำมีความตั้งใจ 2) เป็นการกระทำช้า 3) ผู้กระทำมีอำนาจมากกว่าเหยื่อ (Smith et al., 2008) จากการศึกษา ทั้ง 3 องค์ประกอบ พบร่วมกันว่า การรังแกกันบนโซเชียลมีเดียนั้น หมายถึง พฤติกรรมความก้าวร้าวของบุคคลหรือกลุ่มบุคคล ที่เจตนาใช้เครื่องมืออิเล็กทรอนิกส์ทำร้ายผู้อื่น โดยผู้ที่ตกเป็นเหยื่อยากที่จะป้องกันตนเอง และถูกกระทำช้าไปช้ามา ซึ่งการรังแกกันบนโซเชียลมีเดียนั้นมักจะเกิดขึ้นผ่านทางอินเทอร์เน็ต และเครือข่ายสังคมออนไลน์บนโทรศัพท์เคลื่อนที่เป็นหลัก

2. งานวิจัยและทฤษฎีที่เกี่ยวข้อง

2.1 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

การประมวลผลภาษาธรรมชาติ (NLP) มีต้นกำเนิดย้อนหลังไปถึงช่วงปี 1940 โดยใช้คอมพิวเตอร์พัฒนาเป็นเครื่องแปลภาษา (Machine Translation) ซึ่ง Weaver และ Booth เป็นผู้ริเริ่มโดยใช้คอมพิวเตอร์ในการแปลภาษา ซึ่งต้องการใช้ในการทำลายรหัสของศัตรูในระหว่างสงครามโลกครั้งที่ 2 (Joseph et al., 2016) การประมวลผลภาษาธรรมชาติ เป็นเทคนิคในการคำนวณสำหรับการวิเคราะห์ภาษาตามนุxyzแบบอัตโนมัติ มีวิวัฒนาการมาจากการประมวลผลแบบชุดข้อมูล ซึ่งในเวลาในการคำนวณที่นาน พัฒนาไปจนสู่การวิเคราะห์ข้อมูลนับล้านข้อมูล ที่สามารถประมวลผลได้ในเวลาเป็นหลักวินาที (Cambria & White, 2014) มีพื้นฐานอยู่ในหลากหลายสาขาวิชา ได้แก่ คอมพิวเตอร์ และเทคโนโลยีสารสนเทศ ภาษาศาสตร์ คณิตศาสตร์ วิศวกรรม และอิเล็กทรอนิกส์ ปัญญาประดิษฐ์และจิตวิทยา สามารถการประยุกต์ใช้เพื่อทำความเข้าใจและจัดการข้อมูล หรือคำพูดภาษา ใช้ในการวิเคราะห์และพัฒนาเครื่องมือที่เหมาะสม เพื่อให้ระบบคอมพิวเตอร์เข้าใจและสามารถจัดการภาษาเพื่อทำงานที่ต้องการได้ (Chowdhury, 2003) ในปัจจุบันการพัฒนาระบบจะ

มุ่งเน้นไปที่การแปลภาษาด้วยคอมพิวเตอร์ การดึงข้อมูล การสรุปข้อความ การตอบคำถาม การสร้างแบบจำลองหัวข้อ และการประมวลผลด้วยไวยากรณ์

2.2 ຖរុខ្សីនិងនេវគិតតាមលក្ខមនសំគាលបុគ្គល (Theory of Demographic)

2.2.1 ເພສ (Sex)

ผลการวิจัยเชิงจิตวิทยาแสดงให้เห็นว่า เพศหญิงกับเพศชายมีความแตกต่างกัน ในเรื่องความคิด ทัศนคติ และค่านิยม โดยวัฒนธรรมและสังคมได้กำหนดบทบาทและกิจกรรมของคนสองเพศไว้แตกต่างกัน เพศหญิง เป็นเพศที่มีจิตใจอ่อนไหวหรืออ่อนโยน มีความเป็นแม่บ้านแม่เรือน และถูกซักจุ่งได้ง่ายกว่าเพศชาย เพศหญิงมีแนวโน้มและความต้องการด้านการส่งและรับข่าวสารมากกว่าเพศชาย ในขณะที่เพศชายไม่ได้มีความต้องการที่จะส่งและรับข่าวสารเพียงอย่างเดียว แต่มีความต้องการสร้างความสัมพันธ์อันดีให้เกิดขึ้นจากการรับและส่งข่าวสารนั้นด้วยอีกทั้งเพศชายมักใช้ช่องทางสังคมออนไลน์เพื่อการติดตามข้อมูลข่าวสาร ขณะที่เพศหญิงมักใช้ในการติดต่อสื่อสารกับผู้อื่น โดยเพศชายชอบพูดคุยในเรื่องนามธรรม เช่น การเมือง ศาสนา ขณะที่เพศหญิงก็ชื่นชอบเรื่องนามธรรม แต่ไม่ชอบเปิดเผยความคิดเห็น เพราะเพศหญิงมักถูกซักจุ่งความคิดเห็น ซึ่งจากงานวิจัยพบว่า เพศหญิงมีแนวโน้มที่จะคล้อยตามผู้อื่นหรือถูกซักจุ่งได้ง่ายสามารถรับสารและเลียนแบบพฤติกรรม บุคลิกภาพของผู้อื่นได้ง่ายกว่าเพศชาย ส่วนเพศชายมีความคิดหนักแน่นมากกว่าเพศหญิง แต่เพศหญิงจะมีความรอบคอบ ละเอียดอ่อนด้านความคิดมากกว่าเพศชาย (Patel & Gordon, 1960)

2.2.2 อายุ (Age)

เป็นอีกปัจจัยที่ทำให้คนมีความแตกต่างกัน ทั้งในเรื่องของความคิดและพฤติกรรม คนที่มีอายุน้อยมักจะยึดถืออุดมการณ์ ใจร้อน มีความคิดเห็นแบบเสรีนิยม และมองโลกในแง่ดีมากกว่าคนที่มีอายุมาก ในขณะที่คนที่มีอายุมากจะมีความคิดเชิงอนุรักษ์นิยม ยึดถือการปฏิบัติ มีความระมัดระวัง และมองโลกในแง่ร้ายกว่าคนที่มีอายุน้อย สาเหตุเกิดจากคนที่มีอายุมากมีประสบการณ์ในชีวิตที่เคยผ่านอุปสรรคปัญหาที่ยawnan และมีผลประโยชน์ในสังคมมากกว่า โดยอายุเป็นปัจจัยที่มีส่วนสำคัญในการกำหนดการใช้สื้อและการรับรู้ประโยชน์ของ

สื่อ กลุ่มคนที่มีอายุน้อยจะใช้สื่อเพื่อความบันเทิง แต่กลุ่มที่มีอายุมากจะใช้สื่อ เพราะสนใจที่จะรับรู้เรื่องราวเกี่ยวกับสังคมและความต้องการของตนเองมากกว่า (SAENGROTKITTIKHUN & Choochom, 2021)

2.2.3 ระดับการศึกษา (Education)

คนที่ได้รับการศึกษาในระดับที่แตกต่างกัน ในยุคต่างกัน ในระบบการศึกษาที่ต่างกันในสาขาวิชาที่ต่างกัน ย่อมมีค่านิยม รสนิยม อุดมการณ์ ความรู้สึกนึกคิดและ ความต้องการที่แตกต่าง กัน บุคคลที่มีการศึกษาสูงหรือมีความรู้ดีจะเปรียบมากกว่าใน การรับสารและจะเป็นผู้รับสารที่ดี เพราะคนเหล่านี้มีความรู้ กว้างขวางหลายเรื่อง สามารถเข้าใจสารได้ดี มีความสนใจ ข่าวสารกว้างขวาง จะมีความคิดไตร่ตรอง ไม่เชื่ออะไรง่ายๆ ต้อง มีหลักฐาน หรือเหตุผลสนับสนุนเพียงพอจึงจะตัดสินใจเชื่อ หรือ อีกนัยหนึ่ง ผู้มีการศึกษาสูงมักเชื่อในสื่อสิ่งพิมพ์ เช่น หนังสือพิมพ์มากกว่าสื่อวิทยุ โทรทัศน์และภาพยนตร์ การศึกษา นอกจากจะทำให้คนมีศักยภาพขึ้นแล้ว การศึกษายังทำให้เกิด ความแตกต่างทางทัศนคติ ค่านิยม และคุณธรรม (Mankhatithamt & Thabhiranrak, 2021)

2.2.4 อาชีพ (Occupation)

มืออาชีพในการดำเนินชีวิต เพราะอาชีพเป็นสิ่งที่นำไปสู่ ความต้องการและความจำเป็นในการบริโภคสินค้าและบริการที่ แตกต่างกันออกไป การเลือกใช้สินค้า การแต่งกาย พฤติกรรม การซื้อ เช่น ผู้บริหารระดับสูงจะเลือกใช้สินค้าที่ดูดี เพื่อสร้าง ภาพลักษณ์ที่ดีให้กับตนเอง ในขณะที่พ่อค้า แม่ค้า จะเลือกซื้อ สินค้าที่มีความจำเป็นสำหรับการครองชีพมากกว่าของใช้ สิ่งเบื้องต้น (Sovacool et al., 2018)

2.3 คำจำกัดความของสังคมออนไลน์ (Social Media Definition)

(Kent, 2010) ให้คำจำกัดความกว้างๆ ว่าสังคมออนไลน์ เป็น “การโต้ตอบผ่านช่องทางการสื่อสาร ที่ช่วยให้มีการโต้ตอบและ ข้อเสนอแนะแบบสองทาง” สังคมออนไลน์สมัยใหม่มี ลักษณะเฉพาะ เช่น ศักยภาพสำหรับการโต้ตอบแบบเรียลไทม์ การไม่เปิดเผยตัวตน สามารถแสดงความรู้สึกตรงไปตรงมา ใช้ ระยะเวลาในการตอบสนองสั้นๆ และความสามารถในการ

สื่อสารได้ตลอดเวลา และเนื้อหาที่ผู้ใช้สร้างขึ้นโดยใช้เทคโนโลยี การเผยแพร่ทางอินเทอร์เน็ตซึ่งแตกต่างจากสื่อสิ่งพิมพ์และการ ออกรากурсแบบดั้งเดิม (Terry, 2009)

2.4 แนวคิดของการกลั่นแกล้งบนโซเชียลมีเดีย (Cyberbullying Theory)

คำจำกัดความของการกลั่นแกล้งบนโซเชียลมีเดีย ได้ถูกเปลี่ยน จากการกลั่นแกล้งแบบดั้งเดิมที่เกิดขึ้น กลับกลายมาเป็นการ กระทำผ่านทางโซเชียลมีเดียแทน การกลั่นแกล้งแบบดั้งเดิมและ การกลั่นแกล้งบนโซเชียลมีเดียมีความสัมพันธ์กันและคล้ายคลึง กัน แต่ถึงกระนั้น มีการโต้แย้งว่าการกลั่นแกล้งบนโซเชียลมีเดีย ต้องมีการพิจารณาแยกกัน งานวิจัยหลายชิ้นได้ให้ ความสำคัญของการกลั่นแกล้งบนโซเชียลมีเดียว่า สามารถ ก่อให้เกิดอันตรายเหนือกว่าการกลั่นแกล้งแบบเดิม ๆ ได้ (Englander et al., 2017) การกลั่นแกล้งบนโซเชียลมีเดียจะ เกี่ยวข้องกับการใช้เทคโนโลยีสารสนเทศและการสื่อสาร เช่น อีเมล โทรศัพท์มือถือ การส่งข้อความโต้ตอบแบบทันที เว็บไซต์ ส่วนบุคคล บล็อก เกมออนไลน์ และเว็บไซต์ที่มีการหมิ่นประมาท เพื่อสนับสนุนการจงใจ ทำช้ำ และพฤติกรรมที่เป็นปฏิบัติโดย บุคคลหรือกลุ่มที่มีเจตนาทำร้ายผู้อื่น (Peled, 2019) รวมถึงการ กระทำซึ่งผู้โ杰มตี้ใช้เทคโนโลยีเป็นสื่อกลางในการสื่อสารเพื่อ คุกคาม ทำให้อับอาย ช่มชู่ หรือวิพากษ์วิจารณ์เยื่อของตน การ โจมตี้ในลักษณะต่าง ๆ เช่น การคุกคาม ข้อความที่มีความหมาย แฝงที่หยาบคาย ก้าวร้าว หรือเสื่อมเสีย และความคิดเห็นที่ เผยแพร่บนสื่ออินเทอร์เน็ตฯ เป็นที่แพร่หลายในการโจมตีครั้ง นี้ ในบางกรณี ผู้โ杰มตี้จะอับปอดรูปถ่ายหรือวิดีโอที่เพิ่ง ความรู้สึกไม่สบายและอับอายให้กับ เหยื่อ (Zambrano et al., 2021)

2.5 ปัจจัยที่ก่อให้เกิดการกลั่นแกล้งบนโซเชียลมีเดีย (Factors of Cyberbullying)

2.5.1 การไม่เปิดเผยตัวตน

การกลั่นแกล้งบนโซเชียลมีเดียนี้ สามารถทำได้ผ่าน อุปกรณ์ใดก็ได้ที่ใช้ในการสื่อสารทางอิเล็กทรอนิกส์ ทำให้การ กลั่นแกล้งที่เกิดขึ้นภายใต้โซเชียลมีเดียผู้กลั่นแกล้งสามารถที่จะ ปกปิดตัวตนได้ โดยการสร้างตัวตนปลอมขึ้น (Watts et al., 2017) ก่อให้เกิดการละเมิดความเป็นส่วนตัวของบุคคลมากขึ้น

ซึ่งอาจคงอยู่ได้นานหลายเดือนหรือหลายปี และด้วยความสามารถของการไม่เปิดเผยตัวตนของการกลั่นแกล้งบนโซเชียลมีเดียนี้ จึงทำให้รูปแบบการกลั่นแกล้งน่าดึงดูดยิ่งกว่าการกลั่นแกล้งแบบเดิม ๆ ทำให้สามารถกำหนดเป้าหมายไปยังบุคคลที่ต้องการกลั่นแกล้งได้โดยที่บุคคลนั้นไม่รู้ตัว สามารถทำได้บ่อยครั้งตามต้องการ และลดโอกาสที่จะถูกจับได้ (Wong-Lo et al., 2011)

2.5.2 ความสะดวกในการกระทำ

การใช้งานที่เข้าถึงได้่าย่างผ่านอุปกรณ์ที่เชื่อมเทอร์เน็ตในการเข้าถึง ช่วยให้ผู้กลั่นแกล้งสามารถเข้าถึงเหยื่อได้ตลอด 24 ชั่วโมง 7 วันต่อสัปดาห์ (Watts et al., 2017) ซึ่งแตกต่างจากการกลั่นแกล้งแบบเดิม ๆ ที่ต้องกระทำการเเพชชูหน้า หรือผ่านบุคคลอื่นในการกลั่นแกล้ง ส่งผลให้ในปัจจุบันโซเชียลมีเดียได้กลายเป็นช่องทางยอดนิยมสำหรับการกลั่นแกล้งผ่านทางอินเทอร์เน็ต

2.5.3 การเปลี่ยนแปลงทางอารมณ์ และการระบายความรู้สึก

ด้วยสภาพสังคมที่มีการแข่งขันสูงในหมู่วัยเรียนหรือวัยรุ่น ซึ่งส่งผลต่อการเปลี่ยนแปลงทางอารมณ์ และมักเกิดพฤติกรรมที่ผิดปกติ เช่น ก้าวร้าว ดุดัน ความกลัว โกรธ วิตก กังวล อิจฉาริษยา (Hoff & Mitchell, 2009) ทำให้ต้องการพื้นที่ในการระบายและแสดงออก เพื่อปลดปล่อยความรู้สึกของตัวเอง ไปยังบุคคลอื่น เพื่อให้รู้สึกดีกับตัวเอง หรือ เพราะต้องการให้ออกฝ่ายรู้สึกถึงความทุกข์ยาก ของตน ซึ่งไม่สามารถกระทำได้ในลักษณะเดิมๆ ของความเป็นจริง

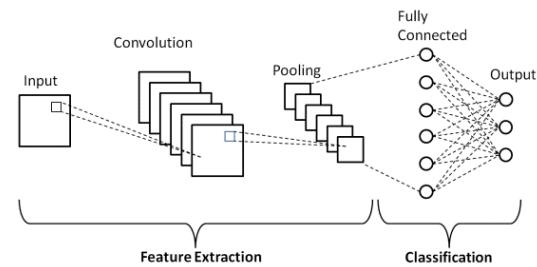
2.5.4 ความอคติที่มีต่อบุคคลอื่น

สาเหตุและปัจจัยที่สำคัญที่สุดที่ก่อให้เกิดการกลั่นแกล้งบนโซเชียลมีเดีย คือ ความเมื่อยล้าต่อตัวบุคคลอื่น ต้องการทำร้ายบุคคลอื่น ด้วยพฤติกรรมทางสังคมที่รุนแรงไม่สามารถแสดงออกได้ซึ่งหน้า (Ovejero et al., 2016) โซเชียลมีเดียจึงเป็นพื้นที่ในการแสดงออกของบุคคลได้อย่างไม่จำกัด

2.6 Convolutional Neural Network (CNN)

การพัฒนาทางด้านปัญญาประดิษฐ์ในระยะหลาຍปีทำให้เทคโนโลยี Convolutional Neural Network (CNN) เป็นที่นิยมอย่างแพร่หลายในการประมวลผลภาพและวิดีโอ โดยเฉพาะในงานวิจัยทางด้านนี้ การนำ CNN มาใช้ได้กับหลายงานต่าง ๆ เริ่มตั้งแต่การจำแนกภาพ การตรวจจับวัตถุ จนถึงการประมวลผลวิดีโอที่มีความซับซ้อน

โครงสร้างของ CNN ประกอบด้วยหลายชั้นที่ทำงานร่วมกัน เพื่อถึงผลลัพธ์ที่แม่นยำ ชั้น Convolution ใช้ในการสกัดลักษณะของภาพ ซึ่งจะมีหน้าที่เคลื่อนที่หน้าต่างที่เรียกว่า



ภาพประกอบที่ 2.1 Convolutional Neural Network

kernel บนภาพเพื่อหาลักษณะเฉพาะ ๆ ของภาพ หลังจากนั้นจะใช้ Pooling Layer เพื่อลดขนาดของข้อมูลและลดการซ้ำซ้อน

2.7 งานวิจัยที่เกี่ยวข้อง

ปัจจุบันมีงานวิจัยที่หลากหลายทำการวิเคราะห์ข้อความกลั่นแกล้งบนอินเทอร์เน็ต การก่อความรุนแรงรูปแบบต่าง ๆ ที่อาศัยเทคโนโลยีสารสนเทศและระบบการติดต่อสื่อสารอิเล็กทรอนิกส์ ทั้งนี้ ไม่ว่าจะในนานาพื้นที่ใหม่สำหรับเปิดเผยเรื่องราวลับ ๆ หรือที่นา อับอายของบุคคลอื่นให้แก่ผู้คนจากทั่วสารทศ หรือในฐานะเครื่องมือส่งผ่านคำชี้มุ่ง คุกคาม หรือ สร้างความ恐怖เครื่องให้แก่เหยื่อ นับเป็นปัญหาร่วมสมัยที่นานาประเทศต่างกันกำลังประสบ เพียงแต่อาจมีระดับความถี่ของการเกิดขึ้น หรือระดับความรุนแรงและความสูญเสียแตกต่างกันไป บ้างเท่านั้น ขึ้นอยู่กับสัดส่วน หรือปริมาณผู้สามารถเข้าถึงเทคโนโลยีได้ วัฒนธรรมความรุนแรง ในสังคม รวมทั้งมาตรการที่ใช้เพื่อแก้ไขปัญหา ซึ่งแน่นอนว่าหมายรวมถึงกฎหมายและการบังคับใช้ ด้วยความรุนแรงในโลกออนไลน์ หลักกรณีส่งผลกระทบกับเหยื่อ

มากเกินกว่าที่ผู้กระทำจะ คาดหมายได้ ตั้งแต่การที่เหยื่อถูกชุดคุ้ยข้อมูลส่วนบุคคลเพิ่มเติม ถูกแอบอ้างเป็นบุคคลอื่น ถูกพักการเรียน ถูกไล่ออกจากงาน ถูกข่มขู่ว่าจะทำร้าย ไปจนถึงการตัดสินใจฆ่าตัวตาย อนึ่ง ปัจจุบัน มีคำศัพท์หลายคำที่ถูกนำมาอธิบายพฤติกรรมความรุนแรงดังกล่าวอาทิ การล่วงละเมิดทางไซเบอร์ (Cyber harassment) การกลั่นแกล้งทางอินเทอร์เน็ต (Cyber bullying) การคุกคามทางไซเบอร์ (Cyber threats) หรือการสะกดรอยตามรังควานทางไซเบอร์ (Cyberstalking) โดยบางคำมีนิยาม ทับซ้อนกันจนอาจทำให้สับสนได้ว่าคำเหล่านี้ มีความหมายเหมือนกัน หรือใช้แทนกันได้หรือไม่ บทความนี้ มุ่งหมายพิจารณาถึงความแตกต่างของคำศัพท์เหล่านั้น รวมทั้ง สำรวจและศึกษาเบื้องต้น

(ณัฐศิริ เชาว์ประสิทธิ์ , 2560) การจำแนกประเภทข้อความภาษาไทยบนสื่อออนไลน์ในเชิงบวกและลบด้วยคลังคำหยาบ การใช้คำหยาบภาษาไทยบนสื่อออนไลน์ ถือได้ว่าเป็นเรื่องละเอียดอ่อนมีผลต่อ อารมณ์ ความรู้สึก ซึ่งทำให้ผู้พบเห็น หรือบุคคลที่ถูกกล่าวหาเสื่อมเสีย ถึงขั้นฟ้องร้องกันมากมาย ถ้ารายแรงมากอาจทำให้เกิดการมาตุรุณ หรือฆ่าตัวตายได้ ซึ่งเป็นปัญหาอย่างมากสำหรับผู้ดูแล เว็บไซต์ งานวิจัยนี้มีวัตถุประสงค์ เพื่อศึกษาการจำแนกประเภทข้อความภาษาไทยในเชิงบวก และลบด้วยคลังคำหยาบ โดยพัฒนาโมเดลสกัดคำหยาบ หรือ ตรวจจับคำหยาบด้วยการปรับปรุงคลัง คำหยาบเพื่อให้การตรวจจับคำมีค่าความแม่นยำสูงที่สุด โดยการหาค่าความถี่ของคำสำคัญด้วยการ ใช้เทคนิค TEICF (Term Frequency Inverse Class Frequency) และศึกษาเปรียบเทียบโมเดลการ สกัดคำหยาบด้วยอัลกอริทึมทั้งหมด 6 อัลกอริมในเทคนิคการจำแนกประเภทข้อมูล ได้แก่ ต้นไม้ ตัดสินใจ (Decision Tree) เคนเนียร์ เรสนेबอร์ (K-Nearest Neighbors) นาอีฟเบย์ (Naive Bayes) การ ถดถอยโลจิสติก (Logistic Regression) ชัพพอร์ตเวกเตอร์ เมมชีน (Support Vector Machines) และ โครงข่ายประสาท เทียม (Neural Network) ในงานวิจัยนี้ ทำการทดลองกับ ข้อความจากกระดาน สนทนาที่ปรากฏในสื่อออนไลน์ ผลการทดลองพบว่า การพัฒนาโมเดลสกัดคำ หรือตรวจจับคำ หยาบโดยการกดคำศัพท์ในคลังคำหยาบสามารถให้ค่าความแม่นยำมากที่สุด และโมเดลการสกัดคำ หยาบด้วยการจำแนกประเภท ข้อความนั้น พบร่วมกับ การถดถอยโลจิสติก (Logistic Regression) ให้ค่า ความถูกต้องมากที่สุด และค่าความคลาดเคลื่อนเฉลี่ยน้อย

ที่สุด สามารถเป็นโมเดลสำหรับการสกัด คำหยาบข้อความภาษาไทยบนสื่อออนไลน์ได้แม่นยำ อีกทั้งมีรูปแบบการวิเคราะห์ ข้อความที่ง่ายต่อ ความเข้าใจมากกว่าเทคนิคอื่นๆ

3. วิธีดำเนินการวิจัย

ในส่วนนี้จะอธิบายกระบวนการที่ใช้ในงานวิจัย ซึ่งประกอบไปด้วยขั้นตอนต่อไปนี้

3.1 การรวบรวมข้อมูลและการเตรียมข้อมูล

งานวิจัยนี้ทำการรวบรวม โดยใช้ข้อมูลจากเว็บไซต์ NLP For Thai ใน โดยมีชื่อว่า [thai_cyberbullying_lgbt](#)

ซึ่งข้อมูลที่รวบรวมถูกดึงข้อมูลมาจากทวิตเตอร์โดยใช้ แพ็กเกจ Python ชื่อ "SNS scrape" (JustAnotherArchivist, n.d.) ข้อความที่ถูกติดป้ายว่าเป็นการกีดกันทางไซเบอร์ถูกเก็บรวบรวมจากคำหยาบภาษาไทยที่เกี่ยวข้องกับกลุ่ม LGBT เช่น "อีตุด", "อีเกย์", "ระเบิดถัง", "ชุดทอง", "อีดิเพช" ตามที่อธิบายในตาราง 1 และประกอบด้วยรายการทั้งหมด 10,072 รายการ ในขณะที่ชุดข้อมูลที่ไม่เกี่ยวข้องกับการกีดกันทางไซเบอร์ถูกดึงข้อมูลจากคำค้นหา "" ซึ่งหมายถึง "happy" ในภาษาไทย โดยทวีทที่ไม่เกี่ยวข้องกับการกีดกันทางไซเบอร์ประกอบด้วยรายการทั้งหมด 10,014 รายการ ชุดข้อมูลทั้งหมดมีทั้งหมด 20,086 รายการ อย่างไรก็ตาม ข้อมูลได้รับการกระบวนการและทำความสะอาดโดยการลบเครื่องหมายวรรคตอนภาษาไทยและคำหยาบ ๆ ในภาษาไทยโดยใช้แพ็กเกจ PythaiNLP และชุดข้อมูลทั้งหมด มีทั้งหมด 20,047 รายการที่ถูกป้ายโดยมี 9,984 ถวะและ 10,063 ถวะที่มีคลาสไม่เกิดกันและคลาสการกีดกันตามลำดับ ในเชิงล้ำดับ, ชุดข้อมูลถูกเข้ารหัสเป็น 1 (การกีดกัน) และ 0 (ไม่การกีดกัน), โดยมีความยาวของข้อความเฉลี่ยของชุดข้อมูลทั้งหมดประมาณ 13 คำ ในเชิงแยกชุดข้อมูล, ชุดข้อมูลได้ถูกแบ่งเป็นอัตราส่วน 70:30 ระหว่างชุดฝึกและชุดทดสอบ โดยมีการแบ่งข้อมูลไว้ฝึกฝนและตรวจสอบดังนี้

ตารางที่ 3.1 การการแบ่งข้อมูลเพื่อฝึกฝนและทดสอบ

	Data
Train	9,811
Test	6,008
Val	4,205

1. การตัดคำ (Word Segmentation)

คือกระบวนการแบ่งข้อความออกเป็นคำ ๆ ตัวอย่างการแยกคำ เช่น "ฉันไปกินข้าว" จะถูกแยกออกเป็น "ฉัน" "ไป" "กิน" "ข้าว" อีกด้วย เช่น "เรื่องอาหารไทย" จะถูกแยกออกเป็น "เรื่อง" "อาหาร" "ไทย" โดยในงานวิจัยนี้ตัดคำภาษาไทยโดยใช้วิธี Maximum Matching ใช้ไลบรารีของ pythainlp โดยนำเข้าโมดูล word_tokenize

2. การกำจัดคำหยุด (Stop Word removal)

การกำจัดคำหยุด (Stop Word removal) คือกระบวนการลบคำที่ไม่จำเป็นออกจากข้อความ คำหยุดเป็นคำที่ 'พบบ่อย เช่น "คือ" "นั้น" "นี่" "ก็" เป็นต้น คำเหล่านี้มักไม่มีความสำคัญต่อความหมายของข้อความมากนัก และไม่ค่อยมีน้ำหนักในการนำมาระหว่างความรู้สึก โดยในงานวิจัยนี้ใช้โมดูล pythainlp พังก์ชั่น thai_stopwords โดยเรียกใช้พังก์ชั่น thai_stopwords

3. แปลงอิโมจิเป็นคำภาษาไทย (thai emoji)

คือการแปลงอิโมจิเป็นภาษาไทย พับบอย เช่น "วันนี้อากาศดีมาก 😊" แปลงออกมาก็จะได้ "วันนี้อากาศดีมาก ยิ่งหวาน" โดยในงานวิจัยนี้ได้ใช้ โมดูล pythainlp โดยเรียกพังก์ชั่น from pythainlp.util import emoji_to_thai

4. ลบตัวเลข (remove number)

การลบตัวเลขออกจากข้อความ เป็นกระบวนการที่สำคัญในการกรองข้อมูลที่มีตัวเลขไม่เกี่ยวข้อง หรือในกรณีที่ต้องการทำนายหรือประมวลผลข้อมูลที่มีเนื้อหาข้อความเท่านั้น ด้วยพังก์ชั่น remove_numbers ที่ได้ถูกนำเสนอ ซึ่งใช้รูปแบบของ Regular Expression เพื่อหาและลบตัวเลขทั้งหมดออกจากข้อความ การลบตัวเลขไม่เพียงแต่ทำให้ข้อความดูสะอาดเรียบร้อย แต่ยังช่วยให้ข้อมูลที่ได้มีข้อมูลตัวเลขที่อาจมีผลกระทบต่อการวิเคราะห์หรือประมวลผล

5. ลบอักษรที่ไม่ใช่ภาษาไทย (remove not thai)

โค้ดที่ให้มีพังก์ชั่นที่ชื่อว่า remove_not_thai ซึ่งมีหน้าที่ลบตัวอักษรที่ไม่ใช่ภาษาไทยออกจากข้อความที่นำเข้า พังก์ชั่นนี้ใช้ PyThaiNLP เพื่อตรวจสอบว่าแต่ละตัวอักษรในข้อความที่นำเข้าเป็นตัวอักษรภาษาไทยหรือไม่ และไม่ใช่ตัวเลขไทยโดยใช้โมดูล pythainlp พังก์ชั่น thai_characters และ thai_digits

6. ลบช่องว่างเพิ่มเติม (remove extra white spaces)

พังก์ชั่นนี้มีประโยชน์ในการทำความสะอาดข้อความโดย

ลบช่องว่างที่เกินมากจากกันในข้อความที่กำหนด

3.2 การสร้างโมเดล

หลังจากการเตรียมข้อมูลเรียบร้อยแล้ว ขั้นตอนถัดไปจะเป็นการสร้างโมเดลเพื่อวิเคราะห์ข้อความว่าเป็นข้อความในเชิงใด ซึ่งในงานวิจัยนี้ทำการสร้างโมเดลจาก 4 อัลกอริทึม คือ K-Nearest Neighbors, Naive Bayes, SVM, CNN โดยอัลกอริทึม KNN กำหนดค่า k = 10 และวัดความใกล้เคียงของข้อมูลที่ต้องการจำแนก (X) กับข้อมูลเรียนรู้ (Y) ส่วน SVM ทำงานโดยพยาามหาเส้นแบ่งที่แยกจุดข้อมูลออกเป็นสองคลาส โดยเส้นแบ่งนี้เรียกว่า เส้นตัดสินใจ (decision boundary) SVM จะพยายามหาเส้น ตัดสินใจที่มีระยะห่างระหว่างจุดข้อมูลมากที่สุด ซึ่งเรียกว่า ระยะขอบ (margin) ซึ่งในงานวิจัยนี้ใช้ linear kernel และ CNN ใช้อัลกอริทึมการอัปเดตค่าน้ำหนักคือ adam พังก์ชั่นความสูญเสียที่ใช้ในการปรับค่าน้ำหนักของโมเดล คือ binary_crossentropy เมماกับจำนวน Class ในการทำนายเพียง 2 Class และจำนวนรอบในการฝึกฝน epochs = 5

3.3 การวัดประสิทธิภาพโมเดล

งานวิจัยนี้ทำการทดสอบประสิทธิภาพของโมเดลโดยแบ่งข้อมูลด้วยเทคนิค Split Test โดยการแบ่งเป็นข้อมูลชุดเรียนรู้ 70% และข้อมูลชุดทดสอบ 30% และ ข้อมูลชุดเรียนรู้ 70% และข้อมูลชุดทดสอบ 30% และใช้ตาราง Confusion matrix เพื่อคำนวณค่าต่างๆ ดังนี้

- ค่าความถูกต้อง (Accuracy) คือ ค่าที่แบบจำลองสามารถถูกต้องได้ในจำนวนของข้อมูลที่ได้รับ โดยวิเคราะห์ขนาดของบริษัท โดยใช้สูตร

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- ค่าความระลึก (precision) คือ การวัดความสามารถในการที่จะจัดเอกสารที่ไม่เกี่ยวข้องออกໄປ โดย precision เป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและถูกต้องของเอกสารที่ถูกต้องทั้งหมดโดยใช้สูตร

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

3. ค่าความแม่นยำ Recall คือ การวัดความสามารถของต้นแบบในการตอบคำตอบที่สนใจและถูกต้องโดย recall เป็นอัตราส่วนของจำนวนกลุ่มที่สนใจและถูกต้อง กับจำนวนกลุ่มที่เกี่ยวข้องทั้งหมดคำนวณได้จากสมการ

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

4. F1-Score เป็นค่าที่ได้จากการเอาค่า precision และ recall มาคำนวณรวมกัน (F1 สร้างขึ้นมาเพื่อเป็น single metric ที่วัดความสามารถของโมเดล ไม่ต้องเลือกระหว่าง precision, recall เพราะเฉลี่ยให้แล้ว)

$$F1 = 2 * \frac{(Precision * recall)}{(Precision + recall)} \quad (4)$$

4. ผลการวิจัย

จากข้อมูล LGBT Cyberbullying Detection in Thai Language จะนำไปสร้างรูปแบบการนำเสนอ Wordcloud จาก โมดูล matplotlib

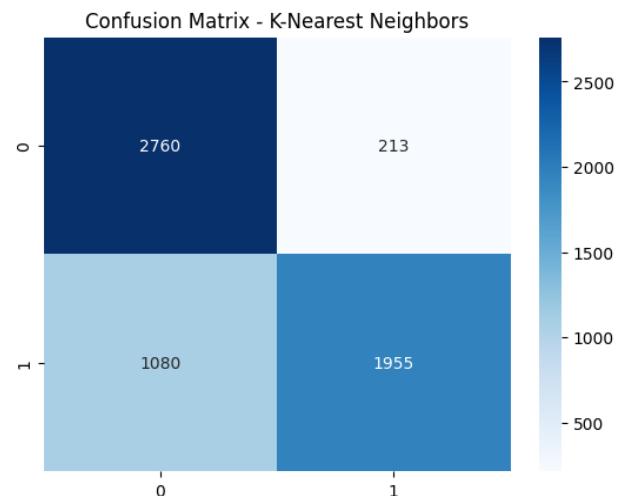


ภาพประกอบที่ 4.1 Word Cloud Positive

จากการวิจัยนี้ใช้ 4 อัลกอริทึม คือ K-Nearest Neighbors, Naive Bayes, SVM , CNN โดยผลของตาราง Confusion Matrix จะมีดังนี้



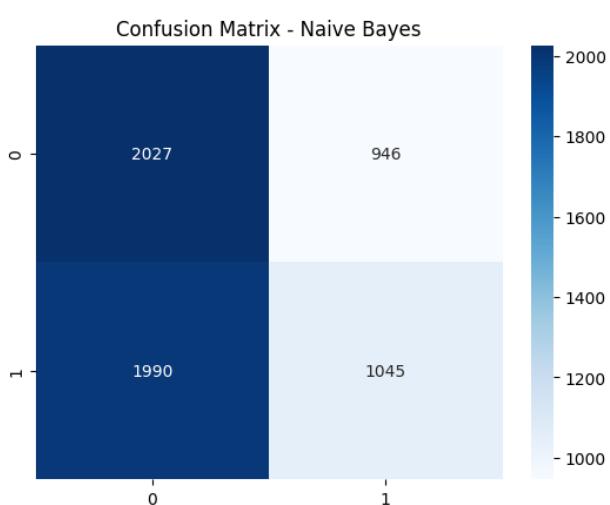
ภาพประกอบที่ 4.2 Word Cloud Negative



ภาพประกอบที่ 4.3 Confusion Matrix KNN

จากการ Confusion Matrix KNN ทำนายผลได้ถูกต้อง 72.7% จาก ($TP + TN / \text{ทั้งหมด}$) และทำนายผลผิดพลาด 27.3% ($FP + FN / \text{ทั้งหมด}$)
โดยมีรายละเอียดดังนี้

- ทำนายผล positive ได้ถูกต้อง 2,760 กรณี (TP)
- ทำนายผล negative ได้ถูกต้อง 1,955 กรณี (TN)
- ทำนายผล positive ผิดพลาด 213 กรณี (FP)
- ทำนายผล negative ผิดพลาด 1,080 กรณี (FN)



ภาพประกอบที่ 4.4 Confusion Matrix Naïve Bayes

($TP + TN / \text{ทั้งหมด}$) และทำนายผลผิดพลาด 29.6%

($FP + FN / \text{ทั้งหมด}$)

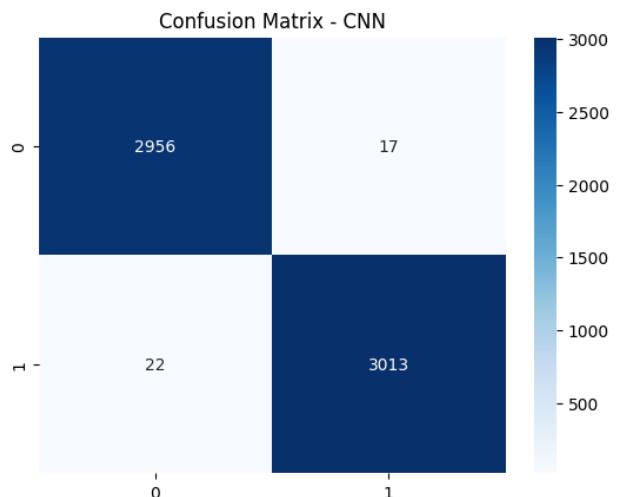
โดยมีรายละเอียดดังนี้

ทำนายผล positive ได้ถูกต้อง 2,649 กรณี (TP)

ทำนายผล negative ได้ถูกต้อง 2,702 กรณี (TN)

ทำนายผล positive ผิดพลาด 234 กรณี (FP)

ทำนายผล negative ผิดพลาด 333 กรณี (FN)



ภาพประกอบที่ 4.6 Confusion Matrix CNN

จากภาพ Confusion Matrix KNN ทำนายผลได้ถูก 57.1%

($TP + TN / \text{ทั้งหมด}$) และทำนายผลผิดพลาด 42.9%

($FP + FN / \text{ทั้งหมด}$)

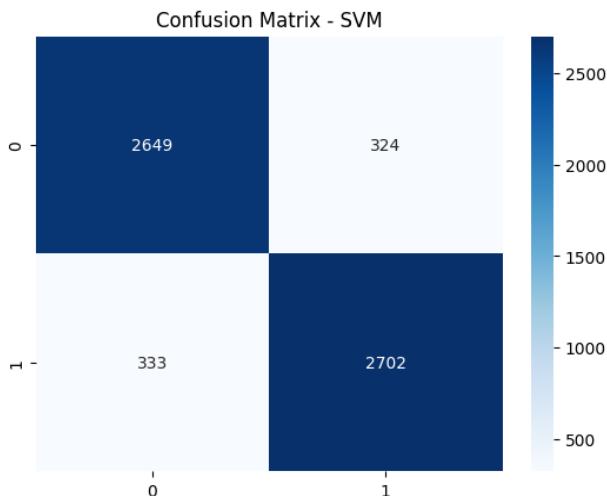
โดยมีรายละเอียดดังนี้

ทำนายผล positive ได้ถูกต้อง 2,027 กรณี (TP)

ทำนายผล negative ได้ถูกต้อง 1,045 กรณี (TN)

ทำนายผล positive ผิดพลาด 946 กรณี (FP)

ทำนายผล negative ผิดพลาด 1,990 กรณี (FN)



ภาพประกอบที่ 4.5 Confusion Matrix SVM

จากภาพ Confusion Matrix KNN ทำนายผลได้ถูก 70.4%

($TP + TN / \text{ทั้งหมด}$) และทำนายผลผิดพลาด 29.6%

($FP + FN / \text{ทั้งหมด}$)

โดยมีรายละเอียดดังนี้

ทำนายผล positive ได้ถูกต้อง 2,956 กรณี (TP)

ทำนายผล negative ได้ถูกต้อง 3,013 กรณี (TN)

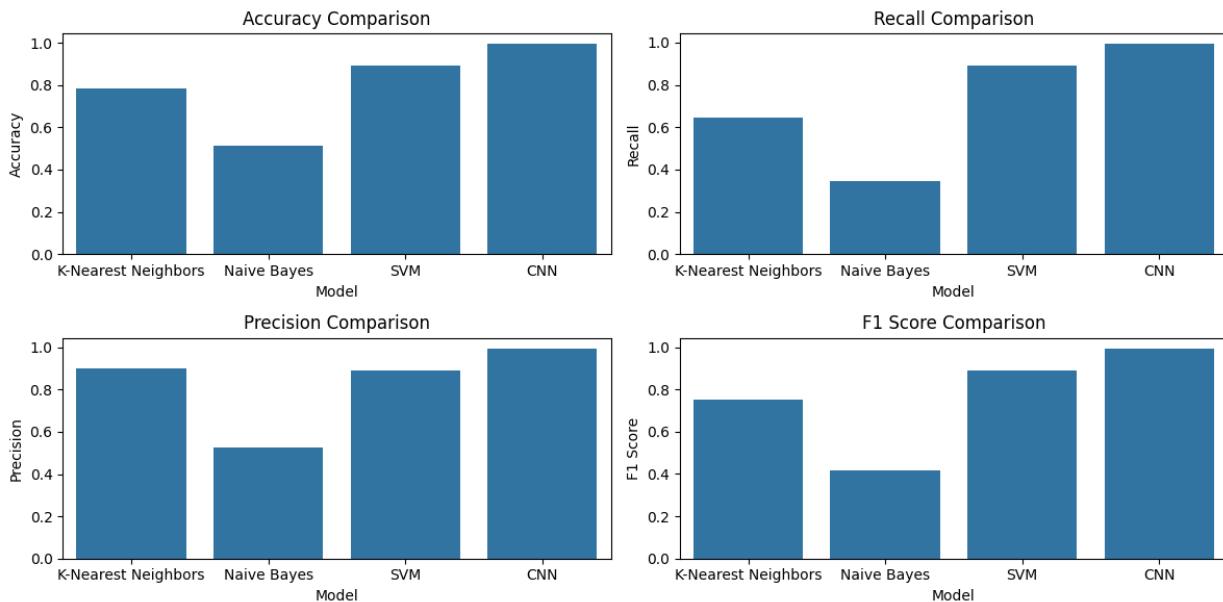
ทำนายผล positive ผิดพลาด 17 กรณี (FP)

ทำนายผล negative ผิดพลาด 22 กรณี (FN)

จากภาพ Confusion Matrix KNN ทำนายผลได้ถูกต้อง 70.4%

สรุปผลวัดประสิทธิภาพ

โดยการวัดประสิทธิภาพด้วย accuracy, recall, precision, f1 score จะได้ผลลัพธ์ดังภาพประกอบที่ 4.6



ภาพประกอบที่ 4.6 สรุปผลการวัดประสิทธิภาพ

ตารางที่ 4.1 ตารางสรุปการวัดประสิทธิภาพ

	Accuracy	Recall	Precision	F1-Score
KNN	78.48	64.42	90.18	75.15
Naive Bayes	51.13	34.43	52.49	41.58
SVM	89.06	89.03	89.29	89.16
CNN	99.35	99.28	99.44	99.36

จากการวัดประสิทธิภาพนี้, เรายกตัวอย่าง CNN (Convolutional Neural Network) ที่มีค่าความถูกต้อง (Accuracy) สูงที่สุดที่ 99.35%, ซึ่งหมายความว่าโมเดลนี้มีความสามารถในการจำแนกข้อมูลอย่างมีประสิทธิภาพมากที่สุดในทั้งสี่โมเดลที่ทดสอบ. การที่มีค่าความถูกต้องสูงมากแสดงถึงความสามารถในการทำนายข้อมูลได้ถูกต้อง

นอกจากนี้, SVM (Support Vector Machine) ก็แสดงให้เห็นถึงประสิทธิภาพที่ดีที่สุดที่เกี่ยวข้องกับค่าความถูกต้องและค่า Precision, Recall, และ F1-Score ที่มีค่าสูงที่สุดหลังจาก CNN ซึ่ง SVM มีค่า Precision, Recall, และ F1-Score ที่ค่อนข้างสูงและใกล้เคียงกัน, แสดงถึงความสามารถสมดุลในการทำนายทั้งทางบวกและลบ

โมเดล KNN (K-Nearest Neighbors) แสดงให้เห็นถึงประสิทธิภาพที่ดีที่สุดในทั้งหมดเมื่อพิจารณาค่า Recall ที่มีค่าสูงที่สุดที่ 64.42%. นั่นหมายความว่า KNN มีความสามารถในการระบุข้อมูลในคลาสบวก (Positive Class) ได้ดีเมื่อเทียบกับโมเดล Naive Bayes ที่มีค่า Recall ต่ำที่สุด

อย่างไรก็ตาม, โมเดล Naive Bayes แสดงให้เห็นถึงประสิทธิภาพที่ต่ำกว่าที่สุดในทั้งสี่โมเดล, ทั้งในแง่ของค่าถูกต้อง, Precision, Recall, และ F1-Score. น่าสังเกตว่า Naive Bayes มีค่า Recall และ F1-Score ที่ต่ำ, แสดงถึงความสามารถในการทำนายคลาสบวกได้ดีเท่ากับโมเดลอื่น ๆ.

ดังนั้น, การเลือกใช้โมเดล CNN หรือ SVM จะขึ้นอยู่กับความต้องการและลักษณะของข้อมูลที่ใช้ในงานนั้น ๆ. ถ้าความถูกต้องและความสามารถในการจำแนกข้อมูลที่สูงสุดเป็นหลัก, โมเดล CNN อาจจะเป็นตัวเลือกที่ดีที่สุด. ในทางตรงกันข้าม, ถ้าความสมดุลในการทำนายทั้งทางบวกและลบเป็นสิ่งที่สำคัญ, SVM ก็อาจจะเป็นตัวเลือกที่ดี

เอกสารอ้างอิง

[1] เช่าวีระสิทธิ์, ณัฐาคิริ (2560) การจำแนกประเภทข้อความภาษาไทยบนสื่อออนไลน์ในเชิงบวกและลบด้วยคลังคำหยาบ. อุทยานวิทยาศาสตร์และเทคโนโลยีราชภัฏรังสิต.

<https://rsuirlibrary.rsu.ac.th/handle/123456789/105>

5

[2] บำรุงศรี วงศ์ (2021) การพัฒนาระบบประเมินผลภาษาธรรมชาติ เพื่อบรรลุการกลั่นแกล้งบนโซเชียลมีเดียในภาษาไทย. จุฬาลงกรณ์มหาวิทยาลัย. [ออนไลน์] จาก

<https://digital.car.chula.ac.th/chulaetd/5181/>

[3] Olweus, D., & Limber, S. P. (2018). Some problems with cyberbullying research. Computers in Human Behavior, 88, 279 - 288.

<https://doi.org/10.1016/j.copsyc.2017.04.012>