



CS 2567/CS-54

รายงานฉบับสมบูรณ์

จำแนกข้อความการกลั่นแกล้งทางไซเบอร์

Cyberbullying Text Classification

โดย

643020613-5 นายธนากร ผาเป้า

643020604-6 นายญาณวิทย์ รอบไธสง

อาจารย์ที่ปรึกษา : ดร.พงษ์ศธร จันทร์ยอย

รายงานเล่มนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314775 โครงการวิทยาการคอมพิวเตอร์ 2

ภาคการเรียนที่ 2 ปีการศึกษา 2567

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน เมษายน พ.ศ. 2568)



CS 2567/CS-54

รายงานฉบับสมบูรณ์

จำแนกข้อความการกลั่นแกล้งทางไซเบอร์

Cyberbullying Text Classification

โดย

643020613-5 นายธนากร ผาเป้า

643020604-6 นายญาณวิทย์ รอบไธสง

อาจารย์ที่ปรึกษา : ดร.พงษ์ศธร จันทร์ยอย

รายงานเล่มนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314775 โครงการวิทยาการคอมพิวเตอร์ 2

ภาคการเรียนที่ 2 ปีการศึกษา 2567

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน เมษายน พ.ศ. 2568)

ธนากร ผาเป้า และ ญาณวิทย์ รอบโธสง. 2567. **จำแนกข้อความการกลั่นแกล้งทางไซเบอร์.**

โครงการคอมพิวเตอร์ หลักสูตรวิทยาศาสตรบัณฑิตสาขา

วิทยาการคอมพิวเตอร์ สาขาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น.

อาจารย์ที่ปรึกษา: ดร.พงษ์ศธร จันทร์ยอย

บทคัดย่อ

งานวิจัยนี้นำเสนอการพัฒนาโมเดลการจำแนกข้อความสำหรับตรวจจับเนื้อหาการกลั่นแกล้งบนโลกออนไลน์ (Cyberbullying) โดยใช้เทคนิคการเรียนรู้เชิงลึก ทั้งในรูปแบบของเครือข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) โดยใช้ หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory: LSTM) และ หน่วยเวียนกลับแบบมีประตู (Gated Recurrent Unit: GRU) และโมเดลแบบผสมระหว่าง BERT (Bidirectional Encoder Representations from Transformers) กับ GRU ซึ่งช่วยให้สามารถเข้าใจบริบทของข้อความได้ดีซึ่งมากยิ่งขึ้น

ชุดข้อมูลที่ใช้ในการฝึกโมเดลประกอบด้วยจำนวน 2,300 ประโยค รวบรวมและกำกับจากหลายแหล่ง ได้แก่ ชุดข้อมูล *thai_cyberbullying_lgbt*, *PyThaiNLP/wisesight-sentiment* และข้อมูลที่ได้จากแบบสอบถามของผู้ใช้งานโดยแบ่งประเภทของข้อความออกเป็นข้อความทั่วไป ข้อความเหยียดเพศ เหยียดรูปร่างหน้าตา สีผิว การศึกษา ฐานะทางสังคม และเชื้อชาติ

ผลการทดลองแสดงให้เห็นว่า โมเดลผสม BERT-GRU มีประสิทธิภาพสูงที่สุด โดยให้ค่า F1-score เท่ากับ 59% รองลงมาคือ GRU ที่ 56%, RNN ที่ 50% และ LSTM ที่ 43% ตามลำดับ ซึ่งแสดงให้เห็นว่าการผสมการเรียนรู้เชิงบริบทจาก BERT เข้ากับความสามารถในการประมวลผลลำดับของ GRU สามารถยกระดับความแม่นยำในการจำแนกข้อความที่มีลักษณะซับซ้อนได้อย่างชัดเจน

คำสำคัญ: การจำแนกข้อความ, ข้อความบูลลี่, การประมวลภาษาธรรมชาติ

Thanakon Phapao and Yarnnawit Robthaisong. 2024. **Cyberbullying Text Classification**.

Bachelor of Science Project in Computer Science, Department of Computer Science,
College of Computing, Khon Kaen University.

Thesis Advisors: Pongsathon Janyoi, Ph.D.

ABSTRACT

This research presents the development of a text classification model for detecting cyberbullying content using deep learning techniques. The models explored include Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), as well as a hybrid model combining BERT (Bidirectional Encoder Representations from Transformers) with GRU, which enhances the model's ability to understand contextual information in text.

The dataset used for training consists of 2,300 labeled sentences collected from various sources, including the *thai_cyberbullying_lgbt dataset*, *PyThaiNLP/wisesight-sentiment*, and user generated survey responses. The data is categorized into several types of content: general text, gender-based insults, body shaming (including appearance and skin color), educational discrimination, socioeconomic class discrimination, and racial discrimination.

Experimental results show that the BERT-GRU hybrid model achieved the highest performance with an F1-score of 59%, followed by GRU at 56%, RNN at 50%, and LSTM at 43%, respectively. These results demonstrate that combining contextual representations from BERT with the sequence modeling capabilities of GRU significantly improves the accuracy of classifying complex and nuanced bullying content.

Keyword: Text Classification, Bully Text, Natural Language Processing

กิตติกรรมประกาศ

ในการดำเนินโครงการครั้งนี้ ผู้จัดทำได้รับความอนุเคราะห์ และได้รับความช่วยเหลือจากหลายบุคคลด้วยกัน จึงขอขอบคุณไว้ ณ โอกาสนี้ด้วย

ขอขอบพระคุณอาจารย์ทุกท่านที่ให้ความรู้และอบรมสั่งสอน

ขอขอบพระคุณอาจารย์ ดร.พงษ์ศร จันทระยอย ที่เป็นที่ปรึกษาโครงการและได้ให้คำชี้แนะแนวทางในการวิเคราะห์ และพัฒนาโมเดลของโครงการนี้

ขอขอบคุณ คุณพ่อ คุณแม่ ที่คอยเป็นกำลังใจให้ตลอดระหว่างการเรียนและการทำรายงานมาโดยตลอด
ขอบคุณเพื่อนๆ ที่คอยให้คำปรึกษา แนะนำ และคอยให้การช่วยเหลือในทุกด้าน

ผู้จัดทำ

ธนากร ผาเป้า

ญาณวิทย์ รอบไธสง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญรูป	จ
สารบัญตาราง	ฉ
บทที่ 1 บทนำ	1
1. ความเป็นมาและความสำคัญปัญหา	1
2. วัตถุประสงค์	1
3. เป้าหมายและขอบเขต	1
4. ประโยชน์ที่คาดว่าจะได้รับ	1
บทที่ 2 ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง	2
1. ทฤษฎีที่เกี่ยวข้อง	2
2. งานวิจัยที่เกี่ยวข้อง	10
บทที่ 3 วิธีการดำเนินงาน	11
บทที่ 4 ผลการดำเนินการวิจัย	14
บทที่ 5 บทสรุป	28
1. สรุปผล	28
2. ปัญหาและอุปสรรค	28
3. ข้อเสนอแนะ	28
เอกสารอ้างอิง	29
ประวัติผู้เขียน	31

สารบัญรูป

	หน้า
ภาพที่ 1 ตัวอย่างการตัดคำและประโยค	3
ภาพที่ 2 การทำงานของเครือข่ายประสาทเทียมแบบวนซ้ำ	3
ภาพที่ 3 การทำงานของหน่วยความจำระยะสั้นระยะยาว	4
ภาพที่ 4 การทำงานของ GRU	5
ภาพที่ 5 ตัวอย่างการทำงานของ BERT	7
ภาพที่ 6 โครงสร้างของตารางประเมินประสิทธิภาพ	8
ภาพที่ 7 ค่าความแม่นยำของโมเดล RNN	15
ภาพที่ 8 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล RNN	16
ภาพที่ 9 ค่าความแม่นยำในการฝึกโมเดล LSTM	18
ภาพที่ 10 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล LSTM	19
ภาพที่ 11 ค่าความแม่นยำในการฝึกโมเดล GRU	21
ภาพที่ 12 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล GRU	22
ภาพที่ 13 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล BERT	24
ภาพที่ 14 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล BERT	25

สารบัญตาราง

	หน้า
ตารางที่ 1 แผนและระยะเวลาดำเนินการ	13
ตารางที่ 2 ประเภทของข้อความและจำนวนประโยค	14
ตารางที่ 3 การกำหนดป้ายชุดข้อมูล	14
ตารางที่ 4 รายงานผลความแม่นยำแต่ละป้ายกำกับ RNN	17
ตารางที่ 5 รายงานผลความแม่นยำแต่ละป้ายกำกับ LSTM	20
ตารางที่ 6 รายงานผลความแม่นยำแต่ละป้ายกำกับ GRU	23
ตารางที่ 7 รายงานผลความแม่นยำแต่ละป้ายกำกับ BERT	26
ตารางที่ 8 ผลการเปรียบเทียบของทั้ง 4 โมเดล	27

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญปัญหา

ปัจจุบันการกลั่นแกล้งบนโซเชียลมีเดียด้วยการพิมพ์ข้อความที่มีคำหยาบ การดูถูกเหยียดเชื้อชาติ รูปลักษณ์ร่างกาย สีผิว การศึกษาแม้กระทั่งชนชั้นฐานะ ส่งผลให้ผู้โดนกระทำได้รับผลกระทบทางด้านจิตใจและอาจถึงทางด้านร่างกาย ผู้ที่ถูกกลั่นแกล้งนั้นย่อมเกิดความเครียด วิตกกังวล ซึมเศร้า เก็บตัว โดดเดี่ยว ไปจนถึงมีความคิดที่จะฆ่าตัวตาย นอกจากนี้การกลั่นแกล้งบนโซเชียลมีเดียยังส่งผลกระทบต่อสังคมโดยรวม ทำให้เกิดความไม่ปลอดภัยและอาจนำไปสู่ความรุนแรงได้

จากปัญหาดังกล่าว ผู้วิจัยจึงได้ศึกษาและจัดทำโมเดลในการตรวจจับข้อความที่ไม่เหมาะสม จากโพสต์หรือความคิดเห็น (Comment) ต่างๆที่มีแนวโน้มที่จะก่อให้เกิดความไม่ปลอดภัย และได้ทำการนำข้อความจากประโยชน์นั้นๆ จำแนกเป็นหมวดหมู่ เช่นข้อความปกติ ข้อความการเหยียดเพศ ข้อความการเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว ข้อความการเหยียดการศึกษา ชนชั้นฐานะ และข้อความการเหยียดเชื้อชาติ ประโยคข้อความที่มีแนวโน้มจะก่อให้เกิดปัญหาที่จะก่อให้เกิดความรุนแรงบนโซเชียลมีเดีย ซึ่งโมเดลการคัดแยกประเภทคำการเหยียดจะช่วยให้สามารถคัดแยกคำต่างๆบนโซเชียลมีเดียได้

โดยโมเดลการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์ จะมุ่งไปที่การจำแนกคำการกลั่นแกล้งในเป็นหมวดหมู่ ประสิทธิภาพและความแม่นยำในการจำแนกข้อความเพื่อให้โมเดลการจำแนกข้อความสามารถใช้งานได้ อย่างมีประสิทธิภาพ

2. วัตถุประสงค์

พัฒนาโมเดลสำหรับการจำแนกประโยคข้อความที่มีการบูลลี่

3. เป้าหมายและขอบเขต

โมเดลสามารถจำแนกหมวดหมู่ประโยคทั่วไป ประโยคคำบูลลี่ได้เฉพาะภาษาไทย

4. ประโยชน์ที่คาดว่าจะได้รับ

ได้รับโมเดลที่สามารถจำแนกหมวดหมู่ของคำได้

บทที่ 2

ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

1. ทฤษฎีที่เกี่ยวข้อง

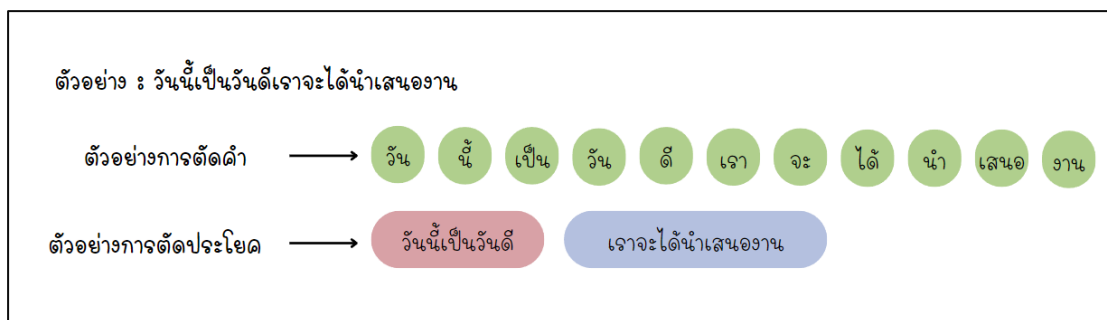
1.1 การจำแนกข้อความ (Text classification) เป็นกระบวนการวิเคราะห์และกำหนดป้ายกำกับให้กับนั้นตามเนื้อหาของ ป้ายกำกับเหล่านี้อาจเป็นประเภทต่างกัน เช่น ประเภทของหัวข้อ ประเภทของอารมณ์ ประเภทของวัตถุประสงค์ การจำแนกประเภทสามารถทำได้โดยใช้วิธีการต่างกัน โดยทั่วไปแล้ว วิธีการจำแนกประเภทสามารถแบ่งออกเป็นสองประเภทใหญ่ๆ

วิธีการตามกฎ (Rule-Based-Methods) เป็นวิธีการจำแนกประเภทตามกฎจะอาศัยกฎทางภาษาศาสตร์ที่จัดทำขึ้นอย่างชัดเจน ระบบจะใช้กฎเหล่านี้เพื่อกำหนดว่าที่กำหนดควรอยู่ในคลาสใด โดยมองหารูปแบบขององค์ประกอบที่เกี่ยวข้องกับความหมาย ตัวอย่างเช่น เราสามารถเขียนกฎเพื่อตรวจจับสแปมอีเมลได้ดังนี้

วิธีการเรียนรู้เชิงเครื่องจักร (Machine Learning Methods) วิธีการจำแนกประเภทโดยอาศัยอัลกอริทึมการเรียนรู้เชิงเครื่องจักรเพื่อเรียนรู้ความสัมพันธ์ระหว่างและป้ายกำกับ อัลกอริทึมการเรียนรู้เชิงเครื่องจักรที่นิยมใช้สำหรับการจำแนกประเภท ได้แก่ อัลกอริทึมการวิเคราะห์การถดถอยโลจิสติก อัลกอริทึม support vector machine อัลกอริทึมเครือข่ายหน่วยประสาท ฯลฯ วิธีการจำแนกประเภทการเรียนรู้เชิงเครื่องจักรมีความแม่นยำสูงกว่าวิธีการตามกฎ (Rule-Based-Methods) แต่ต้องใช้ข้อมูลตัวอย่างจำนวนมากสำหรับการฝึกโมเดล

1.2 การเรียนรู้ของเครื่อง (Machine learning) อาศัยการเรียนรู้จากข้อมูลตัวอย่างเพื่อตรวจจับคำหยาบหรือคำที่สื่อถึงคำพูดในเชิงการกลั่นแกล้ง เช่น การใช้โมเดลภาษา (language model) ที่สามารถทำนายหรือสร้างใหม่ตามที่ถูกต้องและมีความหมาย โดยใช้หลักการ text classification เพื่อที่ให้ตัวโมเดลนั้นสามารถเข้าใจและตรวจจับคำหยาบคาย

1.3 การตัดคำ (Word tokenization) เป็นกระบวนการแบ่งที่กำหนดออกเป็นหน่วยที่เรียกว่าโทเค็น (Token) โทเค็นอาจเป็นคำ วลี หรือแม้แต่ประโยคทั้งหมดก็ได้ ในกระบวนการสร้างโทเค็น อักขระบางตัว เช่น เครื่องหมายวรรคตอนอาจถูกละทิ้ง



ภาพที่ 1 ตัวอย่างการตัดคำและประโยค

จากภาพที่ 2 ตัวอย่างการตัดคำเป็นการแยกแต่ละคำออกจากกัน ซึ่งในภาพแสดงคำที่แยกจากประโยค "วันนี้เป็นวันดีเราจะได้นำเสนองาน" โดยแบ่งออกเป็นคำที่แยกกันอยู่ในวงกลมสีเขียว ได้แก่

วัน → นี → เป็น → วัน → ดี → เรา → จะ → ได้ → นำ → เสนอ → งาน

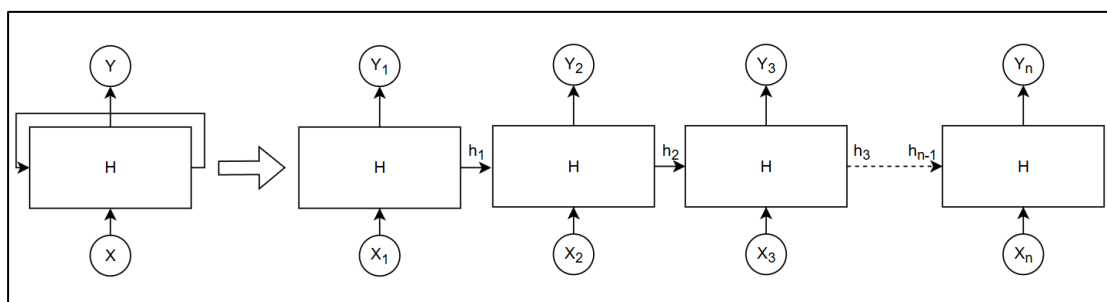
ตัวอย่างการตัดประโยค เป็นการแบ่งประโยคออกเป็นส่วน ๆ ที่มีความหมายชัดเจนและสมบูรณ์ โดยในภาพได้แบ่งเป็น 2 ส่วน ดังนี้:

“วันนี้เป็นวันดี” (วงรีสีชมพู)

“เราจะได้นำเสนองาน” (วงรีสีฟ้า)

1.4 เครือข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) ที่ใช้งานกับข้อมูลที่เป็น ลำดับ (Sequence) เช่น ข้อความ (Sequence of words) ลำดับของคำ เป็นสถาปัตยกรรมเครือข่ายประสาทเทียมแบบวนซ้ำ โครงสร้างของเครือข่ายประสาทเทียมแบบวนซ้ำ จะมีการส่งข้อมูลจากหน่วยประมวลผลไปยังหน่วยต่อไปในลำดับ และมีการย้อนกลับข้อมูลไปยังหน่วยประมวลผลเดิมเพื่อใช้ข้อมูลจากขั้นตอนก่อนหน้าเป็นตัวช่วยในการประมวลผลข้อมูลในปัจจุบัน ส่วนที่สำคัญของ เครือข่ายประสาทเทียมแบบวนซ้ำ จะมี 2 ส่วน ก็คือ

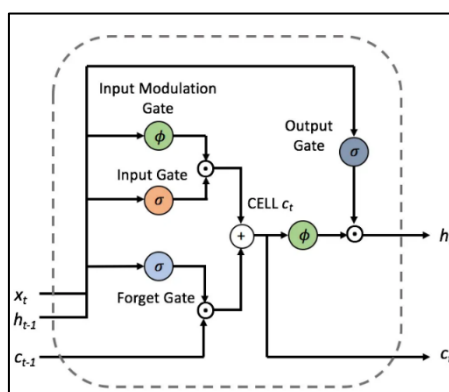
- 1) Hidden state ก่อนหน้า
- 2) การรับเข้าข้อมูล (Input data) ณ ตอนนั้น



ภาพที่ 2 การทำงานของเครือข่ายประสาทเทียมแบบวนซ้ำ

จากภาพที่ 2 อธิบายถึงการวนซ้ำของเครือข่ายประสาทเทียมแบบวนซ้ำ ซึ่งจะแสดงสิ่งที่สำคัญ คือ hidden state ก่อนหน้าการรับเข้าข้อมูลในขณะนั้น โดยการวนซ้ำจะทำให้สามารถนำ hidden state ก่อนหน้านั้นสามารถนำกลับมาใช้ได้หรือการเพิ่ม หน่วยความจำ (Memory) เพื่อเก็บค่า hidden state ที่ได้คำนวณเอาไว้ก่อนหน้า

1.5 หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory) เป็นโครงสร้างที่ต่อยอดมาจากเครือข่ายประสาทเทียมแบบวนซ้ำ โดยที่ หน่วยความจำระยะสั้นระยะยาว นั้นจะมีความละเอียดในการทำงานมากกว่าเครือข่ายประสาทเทียมแบบวนซ้ำ ซึ่งสามารถแก้ปัญหาของเครือข่ายประสาทเทียมแบบวนซ้ำ หรือช่วยให้โมเดลสามารถขยายความจุหน่วยความจำเพื่อรองรับที่มักจะเกิดขึ้นกับประโยคยาวๆ ของข้อมูล หรือช่วยให้โมเดลสามารถขยายความจุหน่วยความจำเพื่อรองรับประโยค จากการเรียนรู้แบบ Long-Term



ภาพที่ 3 การทำงานของหน่วยความจำระยะสั้นระยะยาว (Sirinart,2560)

จากภาพที่ 3 จะเห็นถึงการทำงานของหน่วยความจำระยะสั้นระยะยาว โดยจะมีการกำหนดค่า อ่าน (read) เขียน (write) และ (forget) เพื่อคอยสั่งการทำงาน จะทำให้โมเดลหน่วยความจำระยะสั้นระยะยาวสามารถรู้ได้ว่า ควรจะ อ่าน เขียน หรือ ลืม ในขั้นตอนที่เหมาะสม โดยแต่ละตัวแปรจะมีหน้าที่ดังนี้

- 1) Cell state เป็น state ของ Memory cell ใน หน่วยความจำระยะสั้นระยะยาว
- 2) Gate เป็นตัวควบคุมการไหลของตัวข้อมูล
- 3) Forget ทำหน้าที่ล้าง Cell state เดิมออกไป เพื่อเตรียมตัวรับข้อมูลใหม่ โดยจะตัดสินใจที่

Forget state ว่าควรล้างข้อมูลข้อมูล หรือไม่ล้าง Forget state จะต้องดูข้อมูลที่รับเข้า และ Hidden state ประกอบการตัดสินใจ โดยใช้ ฟังก์ชันซิกมอยด์ (Sigmoid Function) เป็นตัวตัดสินใจ ดังสมการต่อไปนี้

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (1)$$

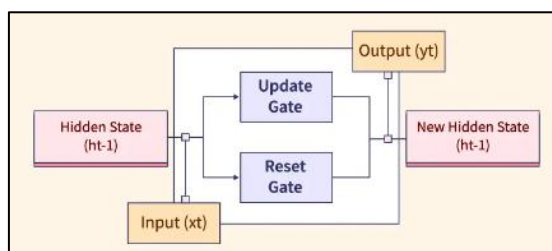
4) เขียน เป็นตัวแปลที่จะตัดสินใจกับ การรับเข้าข้อมูล ว่าควรมีการอัปเดตค่าที่เข้ามากับค่าเดิมหรือไม่ โดยจะยังคงใช้ ฟังก์ชันซิกมอยด์เป็นตัวตัดสินใจว่าควรอัปเดตหรือไม่อัปเดต และต้องคำนวณค่า การรับเข้าข้อมูล ที่เข้ามากับ hidden state ก่อนหน้าดังสมการต่อไปนี้

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (2)$$

5) อ่าน เป็นการดึงค่าจาก cell state หรือ hidden state เพื่อนำไปใช้ในการคำนวณผลลัพธ์ หรือ hidden state สำหรับลำดับถัดไป โดยมีสมการดังนี้

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (3)$$

1.6 หน่วยเวียนกลับแบบมีประตู (Gated Recurrent Unit) เป็นโมเดลที่พัฒนาเพื่อให้โครงสร้างง่ายขึ้นเมื่อเทียบกับหน่วยความจำระยะสั้นระยะยาว โดยรวมเอาคุณสมบัติหลักของ input gate และ forget gate เข้าด้วยกันเป็น update gate และ reset gate ทำให้โมเดลนี้มีประสิทธิภาพใกล้เคียงกับหน่วยความจำระยะสั้นระยะยาว แต่ใช้ทรัพยากรในการประมวลผลน้อยกว่า โดยกระบวนการทำงานของแต่ละตัวแปลจะเป็นดังนี้



ภาพที่ 4 การทำงานของ GRU (Towards Data Science,2018)

จากภาพที่ 4 สามารถอธิบายการทำงานและการหาค่าจากการทำงานของโมเดลได้ดังนี้

1) Update gate จะทำหน้าที่คล้ายกับ forget gate และ input gate ใน หน่วยความจำระยะสั้นระยะยาว คือคำนวณว่าข้อมูลใหม่ที่เข้ามาควรถูกบันทึกไว้ในหน่วยความจำมากน้อยเพียงใด และข้อมูลเก่าที่อยู่ในหน่วยความจำควรถูกลืมหรือปรับปรุงไปมากน้อยเพียงใด จะมีสมการดังนี้

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (4)$$

2) Reset Gate กำหนดว่าข้อมูลเก่าที่มีอยู่นั้นควรถูกลืมมากน้อยเพียงใด และควรให้ข้อมูลใหม่เข้ามา มีผลต่อการคำนวณมากน้อยเพียงใด จะมีสมการดังนี้

$$z_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (5)$$

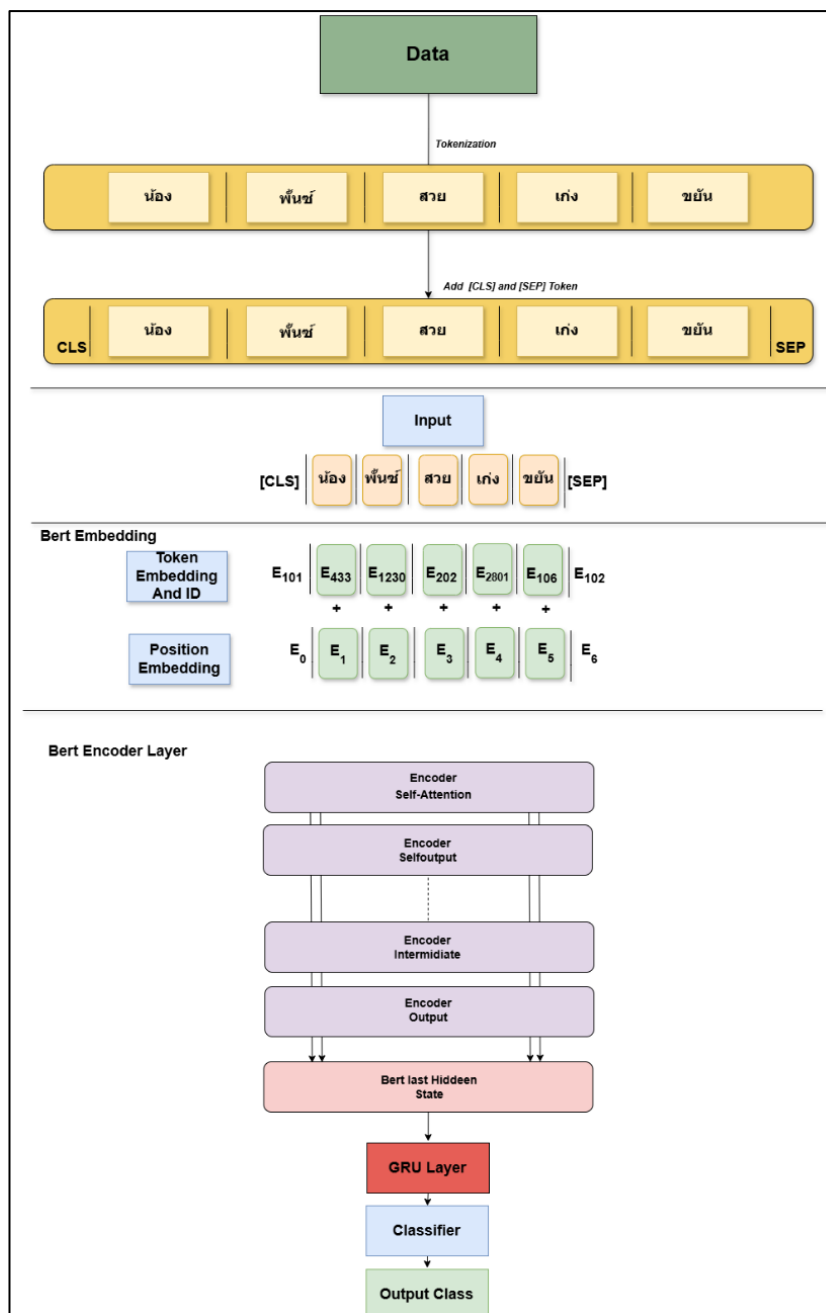
3) การอัปเดตสถานะหน่วยความจำ สถานะหน่วยความจำใหม่ของ GRU คำนวณโดยใช้ข้อมูลจาก update gate และ reset gate ดังนี้

$$h_t = \sigma(1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (6)$$

โดยจะมี \tilde{h}_t เป็นสถานะใหม่ที่ถูกรับปรุงจากข้อมูลปัจจุบัน x_t ซึ่งคำนวณโดยใช้ reset gate

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t]) \quad (7)$$

1.7 การแสดงตัวเข้ารหัสแบบสองทิศทางจากเครื่องแปลง (Bidirectional Encoder Representations from Transformers) เป็น AI Deep Learning แบบเครื่องแปลงสำหรับงาน การประมวลภาษาธรรมชาติ การฝึกฝนของ ตัวเข้ารหัสแบบสองทิศทางจากเครื่องแปลงจะต้องทำการฝึก 2 รอบ ไม่เหมือนกับ หน่วยความจำระยะสั้นระยะยาว หรือ เครือข่ายประสาทเทียมแบบวนซ้ำ โดยตัวโมเดล ตัวเข้ารหัสแบบสองทิศทางจากเครื่องแปลง จะแบ่ง Pre-training ก่อนเพื่อเป็นการทำความเข้าใจภาษาที่จะใช้ก่อน หลังจากที่ได้เข้าใจภาษาของโมเดลแล้ว จะทำการฝึกฝนอีกรอบ เป็นการฝึกครั้งที่ 2 (Fine Tune) คือการนำข้อมูลที่ใช้ในการฝึกแบบที่มีการแยกประเภท เช่น ข้อความทั่วไป ข้อความคำหยาบ ให้กับตัวโมเดล ตัวเข้ารหัสแบบสองทิศทางจากเครื่องแปลง ทำการเรียนรู้และสามารถแยกแยะได้



ภาพที่ 5 ตัวอย่างการทำงานของ BERT+GRU

จากภาพที่ 5 เป็นการนำโมเดล BERT - GRU จะสามารถอธิบายถึงตัวอย่างการทำงานของโมเดลได้ดังนี้

1) การเตรียมข้อมูล (Data Preparation) ข้อมูลที่รับเข้ามาเป็นข้อความ เช่น "น้อง พันธ์ สวย เก่ง ขยัน" ข้อความจะถูก Tokenization แยกออกเป็นคำ ๆ (Tokens) ได้เป็น น้อง พันธ์ สวย เก่ง ขยัน และทำการเพิ่มโทเคน CLS จุดเริ่มต้น และ SEP จุดสิ้นสุด เพื่อบ่งบอกขอบเขตของข้อมูล เช่น CLS น้อง พันธ์ สวย เก่ง ขยัน SEP

2) การสร้างค่าเวกเตอร์ (BERT Embedding) ข้อความที่ผ่านการ Tokenization จะถูกแปลงเป็นค่าเวกเตอร์

3) การประมวลผลด้วย BERT Encoder Layer โมเดล BERT ประกอบด้วยหลาย ๆ Encoder Layer แต่ละ Encoder Layer จะทำการเรียนรู้บริบทของคำในประโยคทั้งด้านซ้ายและขวาพร้อมกัน ใช้ Attention Mechanism ในการหา Feature ที่สำคัญ

4) การประมวลผลด้วย GRU Layer เอาเวกเตอร์ที่ได้จาก BERT Encoder Layer มาเป็นอินพุตให้กับ GRU Layer GRU ทำหน้าที่จับลำดับของคำในบริบทที่ซับซ้อนเพิ่มเติมจากที่ BERT ประมวลผล ส่งผลลัพธ์เข้าสู่ การจำแนกประเภทของคำ (Classifier Layer) ชั้นสุดท้าย เพื่อให้โมเดลตัดสินใจ

5) ผลลัพธ์ (Output) สุดท้าย โมเดลจะให้ค่าเวกเตอร์ที่ผ่านการประมวลผลทั้งหมดไปที่ การจำแนกประเภทของคำ ส่งผลลัพธ์ออกมาเป็น คลาสที่คาดการณ์ได้ เช่น คำทั่วไปทางการชม คำหยาบ เป็นต้น

1.8 ตารางประเมินประสิทธิภาพ (Confusion Matrix) ใช้ประเมินประสิทธิภาพของความถูกต้องของโมเดล เพื่อหาค่าการทำนายที่ถูกต้องของโมเดล (Accuracy) ผลของการทำนายจะประกอบไปด้วย 4 วิธี ดังนี้

- 1) True Positive (TP) สิ่งที่ทำนาย ตรงกับสิ่งที่เกิดขึ้นจริง ในกรณี ทำนายว่าจริง และสิ่งที่เกิดขึ้น ก็คือ จริง
- 2) True Negative (TN) สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณี ทำนายว่า ไม่จริง และสิ่งที่เกิดขึ้น ก็คือ ไม่จริง
- 3) False Positive (FP) สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือทำนายว่า จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง
- 4) False Negative (FN) สิ่งที่ทำนายไม่ตรงกับที่ที่เกิดขึ้นจริง คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้น คือ จริง

	Predicted Positive (1)	Predicted Negative (0)	
Actual Positive (1)	True Positive	False Negative	Sensitivity = $\frac{TP}{TP + FN}$
Actual Negative (0)	False Positive	True Negative	Specificity = $\frac{TN}{FP + TN}$
	Precision = $\frac{TP}{TP + FP}$		

ภาพที่ 6 โครงสร้างของตารางประเมินประสิทธิภาพ (Pagongatchalee,2021)

จากภาพที่ 6 จะแสดงค่าทั้ง 4 ค่า สามารถนำมาคำนวณเพื่อหาค่าความแม่นยำได้ดังนี้

(1) ความแม่นยำ (Accuracy)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

จากสมการแสดงถึงความสามารถของโมเดลในการทำนายได้ถูกต้องทั้งหมด

(2) ความแม่นยำของการทำนายบวก (Precision)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

จากสมการแสดงถึงสัดส่วนของการทำนายบวกที่ถูกต้อง

(3) ค่าความไว (Recall หรือ Sensitivity)

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

จากสมการแสดงถึงความสามารถของโมเดลในการทำนายค่าบวกจากทั้งหมดที่เป็นบวกจริง

(4) F1 Score เป็นการเอา Recall กับ Precision มาหาค่า Harmonic mean

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

จากสมการเป็นค่าเฉลี่ยถ่วงน้ำหนักของ Precision และ Recall ซึ่งมีค่าสูงสุดเมื่อ Precision และ Recall มีค่าเท่ากัน

(5) ค่าลบจากทั้งหมด (Specificity)

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (12)$$

จากสมการแสดงถึงความสามารถของโมเดลในการทำนายค่าลบจากทั้งหมดที่เป็นลบจริง

2. งานวิจัยที่เกี่ยวข้อง

2.1 Text Classification Using Machine Learning Techniques ได้อธิบายเกี่ยวกับการจำแนกประเภท โดยจะมีเนื้อหาเกี่ยวกับความแม่นยำ ในการวัดประสิทธิภาพการจำแนก การใช้ค่าความแม่นยำ (Precision) และการจำ (Recall) ในกรณีที่ข้อมูลไม่สมดุล การผสมค่าความแม่นยำ และการจำเข้าด้วยกันเพื่อให้ได้ภาพรวมของประสิทธิภาพของตัวจำแนก การแทนเอกสารเป็นเวกเตอร์ การทำ การประมวลผลล่วงหน้า (Preprocessing) เช่น การลบ การหยุดคำ (Stop words) และการใช้ การกั้น (Stemming) เพื่อลดขนาดของชุดคุณลักษณะ (Feature set) [12]

2.2 การพัฒนาโมเดลตรวจจับคำหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิ่ง ได้อธิบายเกี่ยวกับงานวิจัยที่เกี่ยวข้องกับกระบวนการวิเคราะห์คำหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิ่ง โดยใช้โมเดลในการตรวจจับคำหยาบภาษาไทยด้วยพจนานุกรมคำหยาบที่ผ่านการปรับปรุงใช้เทคนิค TFICF (Term Frequency - Inverse Class Frequency) ใช้ข้อมูลประเภทตัวอักษรบนกระดานสนทนา หรือแสดงความคิดเห็นตามสื่อออนไลน์ต่างๆ จากเว็บไซต์ ข่าว บันเทิง เกมออนไลน์ และประมูลขายสินค้า เว็บไซต์ดังกล่าวมีการเปิดเสรี รวมถึงการไม่ปิดกั้นคำที่ไม่เหมาะสม หรือคำหยาบ ในการแสดงความคิดเห็นของผู้ใช้บริการ ผู้วิจัยได้เก็บตัวอย่างที่ใช้ทดลองได้จำนวน 1,214 โพสต์ (1 โพสต์ความยาวไม่เกิน 300 ตัวอักษร) ข้อมูลที่นำมาใช้ในการทดสอบนี้เป็นการสุ่มตัวอย่างเพื่อนำมาวิเคราะห์ [1]

2.3 การจำแนกประเภทข้อความภาษาไทยบนสื่อออนไลน์ในเชิงบวกและลบด้วยคลังคำหยาบ ของ ณฐาศิริ เชาว์ประสิทธิ์ มีจุดประสงค์หลักเพื่อพัฒนาโมเดลการจำแนกประเภทข้อความภาษาไทยบนสื่อออนไลน์ในเชิงบวกและลบด้วยการใช้คลังคำหยาบ ซึ่งได้ทำการศึกษาและพัฒนาโมเดลโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) จำนวน 6 แบบ ได้แก่ Decision Tree, Naïve Bayes, K-Nearest Neighbors (K-NN), Neural Network, Logistic Regression, และ Support Vector Machines (SVM) [2]

2.4 Profanity detection in social media text using a hybrid approach of NLP and machine learning ของ Raktim Chatterjee, Sukanya Bhattacharya เกี่ยวกับ การตรวจจับคำหยาบในโซเชียลมีเดีย โดยใช้วิธีการผสมผสานระหว่าง การประมวลภาษาธรรมชาติ และ วิธีการเรียนรู้เชิงเครื่องจักร ซึ่งมีเนื้อหาประกอบด้วย การวิเคราะห์ระบบการตรวจจับคำหยาบที่มีอยู่แล้วและวิเคราะห์ระบบที่นำเสนอใหม่ที่ใช้ การประมวลภาษาธรรมชาติ และ วิธีการเรียนรู้เชิงเครื่องจักร ในการตรวจจับภาษาที่ไม่เหมาะสมบนแพลตฟอร์มโซเชียลมีเดีย รวมถึงการใช้วิธีการแบบ Supervised Learning Problem และการใช้ Generic features เช่น Bag-Of-Words หรือ Embeddings รวมถึงการใช้ Lexical resources ในการประมวลผลข้อมูลเพื่อให้ได้ผลลัพธ์ที่แม่นยำมากขึ้นในการจำแนกภาษาที่ไม่เหมาะสม [13]

บทที่ 3

วิธีการดำเนินงาน

3.1 วิธีการดำเนินงาน

วิธีการดำเนินงานของการสร้างและพัฒนาโมเดลจำแนกข้อความการกลั่นแกล้งทางไซเบอร์ผู้วิจัยได้ทำการแบ่งวิธีการดำเนินงานออกเป็นดังนี้

3.1.1 ศึกษาข้อมูลและวิธีการใช้เครื่องมือในการดำเนินงาน

- 1) Colab
- 2) Pythainlp
- 3) Text classification
- 4) Machine learning
- 5) Word tokenization
- 6) Recurrent Neural Network (RNN)
- 7) Long Short-Term Memory (LSTM)
- 8) Gated Recurrent Unit (GRU)
- 9) Bidirectional Encoder Representations from Transformers (BERT)
- 10) TF-IDF
- 11) Bag of Words
- 12) Confusion Matrix
- 13) Deploy model machine learning

3.1.2 ขั้นตอนจัดเก็บและรวบรวมข้อมูล

โดยได้จัดเก็บและรวบรวมข้อมูลจาก NLP for Thai Text Classification thai_cyberbullying_lgbt PyThaiNLP/wisesight-sentiment และข้อมูลจากแบบฟอร์มที่ผู้จัดทำวิจัยได้สร้างขึ้นเพื่อหาข้อมูลข้อความของแต่ละหมวดหมู่เพิ่ม โดยจะนำจำนวนชุดข้อมูลทั้งหมดแบ่งเป็น 3 ชุด

- 1) ชุดฝึก 70%
- 2) ชุดทดสอบ 20%
- 3) ชุดตรวจสอบ 10%

3.1.3 การจัดเตรียมข้อมูล

ขั้นตอนการจัดเตรียมข้อมูลในการสร้างและพัฒนาโมเดลการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์มีขั้นตอนดังต่อไปนี้

- 1) แบ่งแยกข้อมูลจากที่รวบรวมมาทั้งหมดเป็น 6 หมวดหมู่
- 2) ทำความสะอาดข้อมูลคือการลบอักขระพิเศษ อีโมจิ ช่องว่างและหว่างข้อความออก
- 3) ตัดคำ ทำการตัดคำออกเป็นหน่วยย่อยๆ เพื่อให้สามารถวิเคราะห์ข้อมูลได้ง่ายขึ้น
- 4) กำหนดป้ายให้กับชุดข้อมูล

3.1.4 โมเดลการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์

โมเดลของการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์ที่ได้เลือกนำมาฝึกฝนมีด้วยกันทั้งหมด 4 ตัว คือ

- 1) RNN
- 2) LSTM
- 3) GRU
- 4) BERT+GRU

3.1.5 การตั้งค่าของทั้ง 4 โมเดล (Model Setting)

- 1) กำหนดขนาดของเวกเตอร์ (Hidden State) เป็น 256
- 2) ขนาดของผลลัพธ์ที่ออกจากโมเดล (Output DIM) เป็น 6
- 3) กำหนดจำนวนชั้น (N Layer) เป็น 2
- 4) กำหนดให้โมเดลเป็น Bidirectional
- 5) กำหนดอัตราการ Dropout เป็น 0.25
- 6) จำนวนรอบในการฝึกเป็น 200 รอบ

3.1.6 กำหนดอัตราการเรียนรู้ของโมเดล (Learning Rate)

การกำหนดอัตราการเรียนรู้ของโมเดลจะใช้ค่า Learning Rate ที่ผู้พัฒนาโมเดลได้ทำการทดสอบแล้วเห็นผลว่าเหมาะสมสำหรับชุดข้อมูลปัจจุบันที่ผู้พัฒนารอบรวมมาได้ โดยจะได้ค่าของแต่ละโมเดล ดังนี้

- 1) โมเดล RNN เป็น 0.000005
- 2) โมเดล LSTM เป็น 0.000005
- 3) โมเดล GRU เป็น 0.000005
- 4) โมเดล BERT เป็น 0.000005

3.1.7 เปรียบเทียบโมเดล

กระบวนการเปรียบเทียบโมเดลคือการนำโมเดลทั้ง 4 ตัวที่ได้เลือกมาฝึกฝนและพัฒนา นำมาเปรียบเทียบผลลัพธ์ในการจำแนกข้อความ เพื่อที่จะได้ผลสรุปค่าที่ดีที่สุดในการทำงานของโมเดลทั้งหมด เช่น ผลลัพธ์ของ LSTM และ GRU ที่ได้ผลลัพธ์ต่างกัน

หลังจากที่เปรียบเทียบโมเดลทั้ง 4 ตัว (RNN LSTM GRU และ BERT+GRU) ก่อนหน้านั้นแล้ว จะนำโมเดลตัวที่ดีที่สุดมาทำการเปรียบเทียบกับโมเดล BERT เพื่อดำเนินการตามแผนการทำงานต่อไป

3.1.8 สรุปผล

สรุปผลจากกระบวนการทำงาน การเก็บข้อมูล การจัดเตรียมข้อมูล การจัดทำพัฒนาโมเดลประสิทธิภาพของโมเดล และผลจากการเปรียบเทียบโมเดล

3.2 แผนและระยะการดำเนินงาน

ตารางที่ 1 แผนและระยะการดำเนินงาน

การดำเนินการ	ระยะเวลาการดำเนินการ									
	ภาคเรียนที่ 1					ภาคเรียนที่ 2				
	7	8	9	10	11	12	1	2	3	4
ศึกษาข้อมูลและวิธีใช้เครื่องมือในการดำเนินงาน										
ขั้นตอนจัดเก็บและรวบรวมข้อมูล										
การจัดเตรียมข้อมูล										
นำเสนอความคืบหน้าครั้งที่ 1 (กลางภาค)										
โมเดลการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์										
เปรียบเทียบโมเดล										
สรุปผลการดำเนินงาน										
นำเสนอความคืบหน้าครั้งที่ 2 (ปลายภาค)										
นำเสนอความคืบหน้าครั้งที่ 3 (กลางภาค)										
นำเสนอโครงงานฉบับสมบูรณ์										

3.3 ค่าใช้จ่าย

- ค่าถ่ายเอกสารรูปเล่ม 500 บาท
- Colab Pro เดือนละ 343 บาท จำนวน 4 เดือน
- Colab Pro เพิ่มเติม 343 บาท
- รวม 2,215 บาท

บทที่ 4

ผลการดำเนินการวิจัย

4.1 ผลที่ได้จากการรวบรวมข้อมูล

จะได้จำนวนทั้งหมด 2,300 ประโยค โดยได้จัดเก็บและรวบรวมข้อมูลจาก NLP for Thai Text Classification thai_cyberbullying_lgbt PyThaiNLP/wisesight-sentiment และข้อมูลจากแบบฟอร์ม และทำการแบ่งประเภทของประโยคเป็นทั้งหมด 6 ประเภท

ตารางที่ 2 ประเภทของข้อความและจำนวนประโยค โดยจะแบ่งเป็นทั้งหมด 6 ประเภทข้อความ

ประเภทของข้อความ	จำนวนประโยค
ข้อความทั่วไป	585
ข้อความคำทั่วไปการชม	537
ข้อความคำหยาบปกติ	503
ข้อความเหยียดเพศ	245
ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว	230
ข้อความการเหยียดการศึกษา ขนชั้นฐานะ การเหยียดเชื้อชาติ	200
รวม	2300

จากตารางที่ 2 จะแสดงให้เห็นผลที่ได้จากการแบ่งประเภทของข้อมูลทั้งหมด 6 ประเภทได้แก่ ข้อความทั่วไป ข้อความคำทั่วไปการชม ข้อความคำหยาบปกติ ข้อความเหยียดเพศ ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว และ ข้อความการเหยียดการศึกษา ขนชั้นฐานะ การเหยียดเชื้อชาติ

4.2 ผลที่ได้จากการกำหนดป้าย

จะได้ป้ายกำกับของข้อมูลทั้ง 6 ป้ายกำกับจากชุดข้อมูลทั้งหมด 6 ประเภท

ตารางที่ 3 การกำหนดป้ายชุดข้อมูล

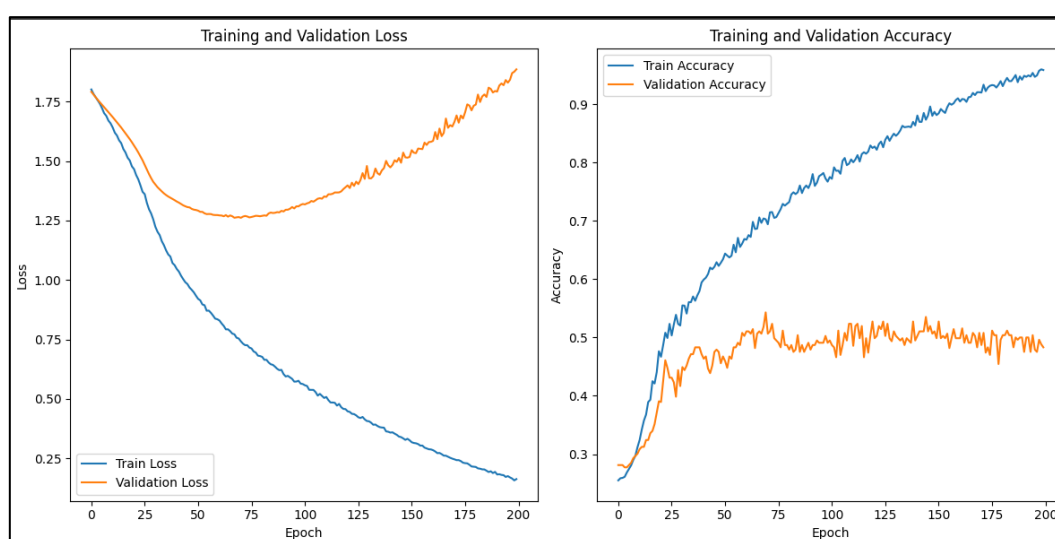
ชุดข้อมูล	ประเภทของข้อความ	ป้ายกำกับ
1	ข้อความคำทั่วไปการชม	Positive
2	ข้อความทั่วไป	Neutral
3	ข้อความคำหยาบปกติ	Negative
4	ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว	Body shaming
5	ข้อความการเหยียดการศึกษา ชนชั้นฐานะ การเหยียดเชื้อชาติ	Classism
6	ข้อความเหยียดเพศ	Sexism

จากตารางที่ 3 จะแสดงให้เห็นถึงการกำหนดป้ายของแต่ละชุดข้อมูล โดยจะได้ทั้งหมด 6 ป้ายกำกับดังนี้ ชุดที่ 1 แทนด้วย Positive ชุดที่ 2 แทนด้วย Neutral ชุดที่ 3 แทนด้วย Negative ชุดที่ 4 แทนด้วย Body shaming ชุดที่ 5 แทนด้วย Classism และชุดที่ 6 แทนด้วย Sexism

4.3 ผลที่ได้จากการฝึกฝนโมเดล RNN

ผลที่ได้จากการฝึกฝนโมเดล RNN จะได้ดังนี้

- 1) ค่าความแม่นยำในการฝึกโมเดล RNN



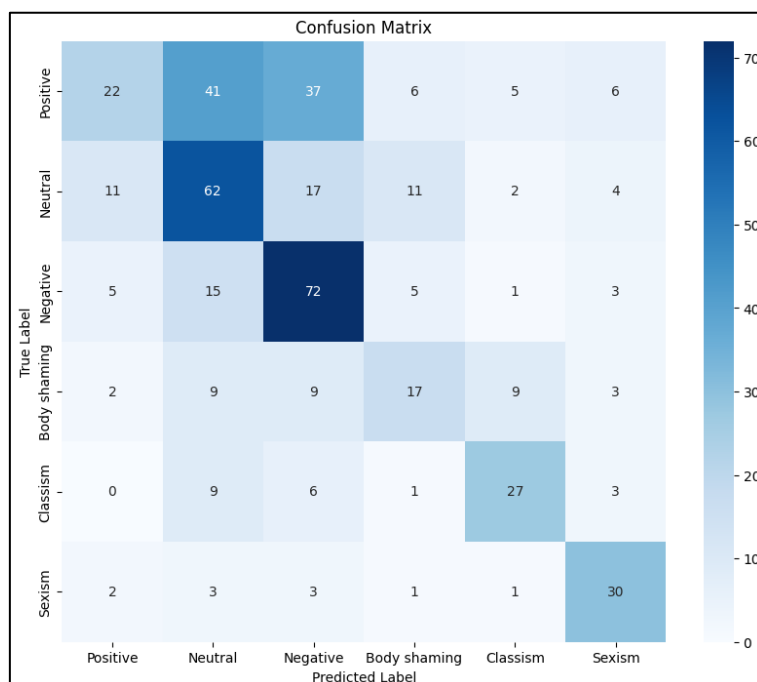
ภาพที่ 7 ค่าความแม่นยำของโมเดล RNN

จากภาพที่ 7 ทางซ้ายมือ (Loss over Epochs) เส้นสีฟ้าจะแทนค่าความสูญเสียของการฝึกฝน (Train Loss) ลดลงตั้งแต่เริ่ม จนเข้าใกล้ 0 หมายความว่าโมเดลนั้นเรียนรู้ได้ดีขึ้นเรื่อยๆ และเส้นสีส้มแทนค่าการสูญเสียการตรวจสอบ (Validation Loss) ตอนเริ่มจะลดลงเล็กน้อยในตอนเริ่มและเพิ่มขึ้นเรื่อยๆหลังจากนั้น โดยจะมีค่าตั้งแต่ 1.75 ถึง 2.00 โดยประมาณ

จากภาพทางขวามือ (Accuracy over Epochs) เส้นสีฟ้าจะแทนค่าแม่นยำในการฝึกฝน (Train Accuracy) จะเห็นว่าค่านั้นเพิ่มขึ้นอย่างรวดเร็วจาก 0.20 จนเกือบถึง 1.00 ซึ่งหมายความว่าโมเดลนั้นสามารถจดจำข้อมูลในการฝึกฝนได้ดี แต่เส้นสีส้มที่แทนค่าความถูกต้องในการตรวจสอบ (Validation Accuracy) มีการเรียนรู้ได้ดีในตอนเริ่มและเรียนรู้ได้ดีที่สุดในรอบที่ 60 หลังจากการฝึกโมเดล 200 รอบ

สรุปค่าความแม่นยำของโมเดล RNN ได้ว่า มีแนวโน้มที่จะเกิดการ overfitting เพราะ Train Loss ลดลงมาก ในขณะที่ Validation Loss เริ่มเพิ่มขึ้น หรือไม่ไปในทิศทางเดียวกัน และค่า Train Accuracy สูงขึ้นเรื่อยๆ แต่ค่า Validation Accuracy ค่อนข้างนิ่ง

2) ค่าการประเมินประสิทธิภาพของการทำนาย (Confusion Matrix) ของโมเดล RNN



ภาพที่ 8 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล RNN

จากรูปที่ 8 แสดงค่า Confusion Matrix ของโมเดล RNN พบว่าหมวดหมู่ที่ทำนายได้ถูกต้องมากที่สุดคือ Negative และ Neutral โดยมีการทำนายถูกต้อง 72 ครั้งจากตัวอย่างทั้งหมด 107 ตัวอย่าง และ 62 ตัวอย่างตามลำดับ แต่เมื่อพิจารณาตามค่า F1-Score ของคลาส Positive จะอยู่เพียง 50% แต่คลาสที่มีประสิทธิภาพในการทำนายมากที่สุดคือคลาส Sexism ซึ่งอยู่ที่ 67% และต่ำที่สุดคือคลาส Positive อยู่ที่ 28%

เมื่อพิจารณาการทำนายผิดของโมเดลในแต่ละคลาส พบว่าคลาสที่มีการทำนายผิดมากที่สุดตามค่า Recall คือคลาส Positive โดยมีค่าอยู่เพียง 19% แสดงให้เห็นว่าโมเดลนั้นยังไม่สามารถระบุข้อความที่เป็นเชิงบวกได้ดีมากเท่าที่ควร ในขณะที่ทำนายได้ผิดน้อยที่สุดคือคลาส Sexism, Negative มีค่า Recall ถึง 75%, 71%

3) ค่า F1-Score ของโมเดล RNN

ตารางที่ 4 รายงานผลความแม่นยำแต่ละป้ายกำกับ RNN

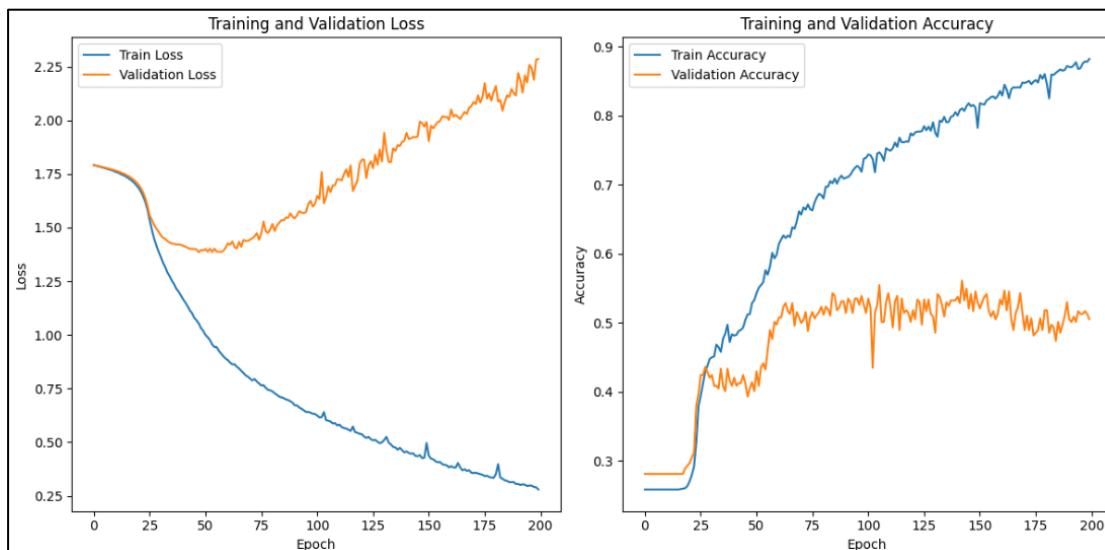
ชุดข้อมูล	ประเภทของข้อความ	ป้ายกำกับ	Precision Score	Recall	F1-Score	Support
1	ข้อความคำทั่วไปการชม	0	0.52	0.19	0.28	117
2	ข้อความทั่วไป	1	0.45	0.58	0.50	107
3	ข้อความคำหยาบ	2	0.50	0.71	0.59	101
4	ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว	3	0.41	0.35	0.38	49
5	ข้อความการเหยียดการศึกษา ชั้นชั้น ฐานะ การเหยียดเชื้อชาติ	4	0.60	0.59	0.59	46
6	ข้อความเหยียดเพศ	5	0.61	0.75	0.67	40
Macro Average			0.52	0.53	0.50	460
Weighted Average			0.50	0.50	0.47	460

จากตารางที่ 4 จะแสดงถึงค่าความแม่นยำของแต่ละหมวดหมู่ โดยหมวดหมู่ที่ได้ค่า F1-Score ที่ดีที่สุดคือข้อความเหยียดเพศ เป็น 67% และน้อยที่สุดคือ ข้อความคำทั่วไปการชม เป็น 28% และค่าเฉลี่ยของค่า F1-Score ทั้งหมดคือ 47% โดยใช้การคำนวณแบบ Weighted Average

4.4 ผลที่ได้จากการฝึกฝนโมเดล LSTM

ผลที่ได้จากการฝึกฝนโมเดล LSTM จะได้ดังนี้

1) ค่าความแม่นยำในการฝึกโมเดล LSTM



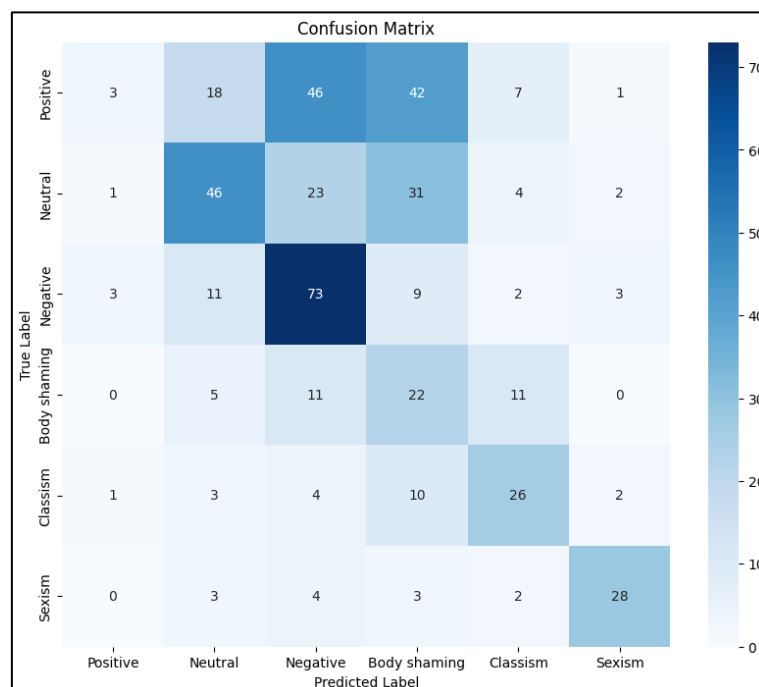
ภาพที่ 9 ค่าความแม่นยำในการฝึกโมเดล LSTM

จากภาพที่ 9 ทางซ้ายมือ (Loss over Epochs) เส้นสีฟ้าจะแทนค่าความสูญเสียของการฝึกฝน (Train Loss) ลดลงตั้งแต่เริ่ม จนเข้าใกล้ 0 หมายความว่าโมเดลนั้นเรียนรู้ได้ดีขึ้นเรื่อยๆ และเส้นสีส้มแทนค่าการสูญเสียการตรวจสอบ (Validation Loss) ตอนเริ่มจะลดลงในตอนเริ่มและเพิ่มขึ้นเรื่อยๆหลังจากการฝึกที่ 75 โดยจะมีค่าตั้งแต่ 1.10 ถึง 2.30 โดยประมาณ

จากภาพทางขวามือ (Accuracy over Epochs) เส้นสีฟ้าจะแทนค่าแม่นยำในการฝึกฝน (Train Accuracy) จะเห็นว่าค่านี้นั้นเพิ่มขึ้นอย่างรวดเร็วจาก 0.20 จนเกือบถึง 1.00 ซึ่งหมายความว่าโมเดลนั้นสามารถจดจำข้อมูลในการฝึกฝนได้ดี แต่เส้นสีส้มที่แทนค่าความถูกต้องในการตรวจสอบ (Validation Accuracy) มีการเรียนรู้ได้ดีในตอนเริ่มและเรียนรู้ได้ดีที่สุดในรอบที่ 100 ในจำนวนรอบทั้งหมดในการฝึกโมเดล

สรุปค่าความแม่นยำของโมเดล LSTM ได้ว่า มีแนวโน้มที่จะเกิดการ overfitting เพราะ Train Loss ลดลงมากในขณะที่ Validation Loss เริ่มเพิ่มขึ้น หรือไม่ไปในทิศทางเดียวกัน และค่า Train Accuracy สูงขึ้นเรื่อยๆ แต่ค่า Validation Accuracy ค่อนข้างนิ่ง

2) ค่าการประเมินประสิทธิภาพของการทำนาย (Confusion Matrix) ของโมเดล LSTM



ภาพที่ 10 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล LSTM

จากรูปที่ 10 แสดงค่า Confusion Matrix ของโมเดล LSTM พบว่าหมวดหมู่ที่ทำนายถูกต้องเยอะที่สุดคือ Negative คือ 73 ครั้งจากตัวอย่างข้อมูล 101 ตัวอย่าง แต่เมื่อพิจารณาตามค่า F1-Score ของคลาส Sexism จะอยู่ที่ 74% เป็นคลาสที่มีประสิทธิภาพในการทำนายมาก และต่ำที่สุดคือคลาส Positive อยู่ที่ 5%

เมื่อพิจารณาการทำนายผิดของโมเดลในแต่ละคลาส พบว่าคลาสที่มีการทำนายผิดมากที่สุดตามค่า Recall คือคลาส Positive โดยมีค่าอยู่ที่ 3% ในขณะที่คลาสที่ทำนายได้ผิดน้อยที่สุดคือคลาส Negative และ Sexism มีค่า Recall ที่ 72% 70%

3) ค่า F1-Score ของโมเดล LSTM

ตารางที่ 5 รายงานผลความแม่นยำแต่ละป้ายกำกับ LSTM

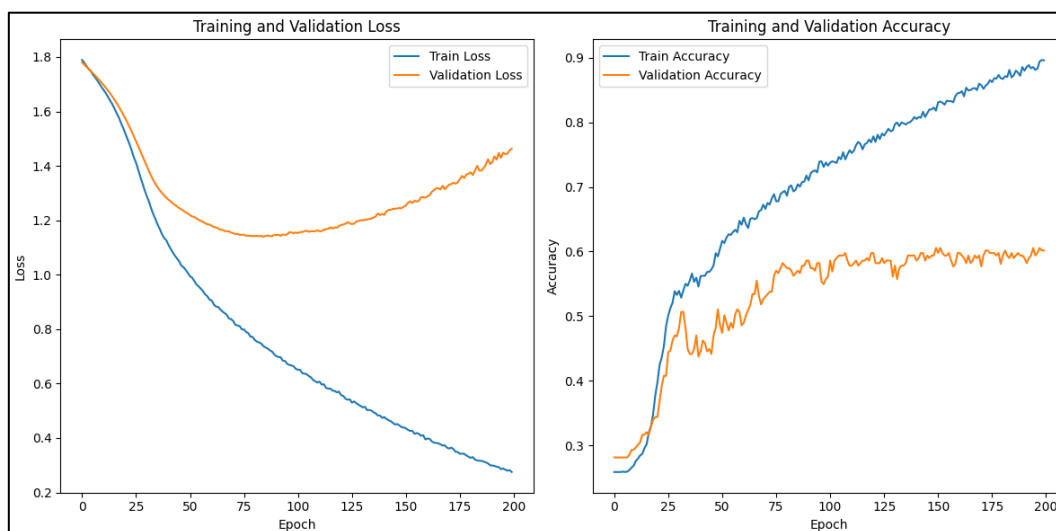
ชุดข้อมูล	ประเภทของข้อความ	ป้ายกำกับ	Precision Score	Recall	F1-Score	Support
1	ข้อความคำทั่วไปการชม	0	0.38	0.03	0.05	117
2	ข้อความทั่วไป	1	0.53	0.43	0.48	107
3	ข้อความคำหยาบ	2	0.45	0.72	0.56	101
4	ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว	3	0.19	0.45	0.27	49
5	ข้อความการเหยียดการศึกษา ชนชั้นฐานะการเหยียดเชื้อชาติ	4	0.50	0.57	0.53	46
6	ข้อความเหยียดเพศ	5	0.78	0.70	0.74	40
Macro Average			0.47	0.48	0.44	460
Weighted Average			0.46	0.43	0.39	460

จากตารางที่ 5 จะแสดงถึงค่าความแม่นยำของแต่ละหมวดหมู่ โดยหมวดหมู่ที่ได้ค่า F1-Score ที่ดีที่สุดคือ ข้อความเหยียดเพศ เป็น 74% และน้อยที่สุดคือ ข้อความคำทั่วไปการชม เป็น 3% และค่าเฉลี่ยของค่า F1-Score ทั้งหมดคือ 39% โดยใช้การคำนวณแบบ Weighted Average

4.5 ผลที่ได้จากการฝึกฝนโมเดล GRU

ผลที่ได้จากการฝึกฝนโมเดล GRU จะได้ดังนี้

1) ค่าความแม่นยำในการฝึกโมเดล GRU



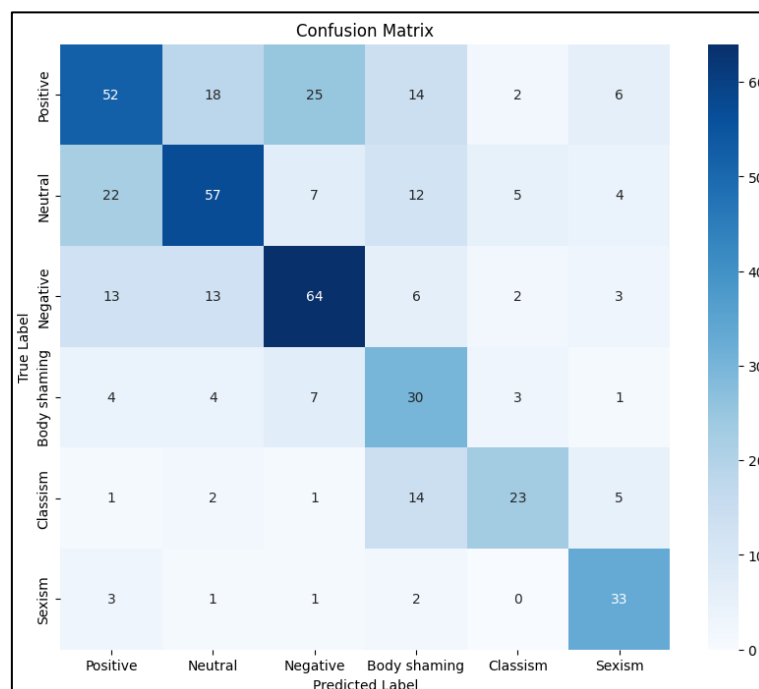
ภาพที่ 11 ค่าความแม่นยำในการฝึกโมเดล GRU

จากภาพที่ 11 ทางซ้ายมือ (Loss over Epochs) เส้นสีฟ้าจะแทนค่าความสูญเสียของการฝึกฝน (Train Loss) ลดลงตั้งแต่เริ่ม จนเข้าใกล้ 0 หมายความว่าโมเดลนั้นเรียนรู้ได้ดีขึ้นเรื่อยๆ แต่เมื่อถึงการฝึกที่ 75 โมเดลมีการเกิด Overfitting ทำให้เกิดความไม่เสถียรในการฝึกโมเดล และเส้นสีส้มแทนค่าการสูญเสียการตรวจสอบ (Validation Loss) ตอนเริ่มจะลดลงในตอนเริ่มและเพิ่มขึ้นเรื่อยๆหลังจากการฝึกที่ 75 โดยจะมีค่าตั้งแต่ 1.10 ถึง 1.50

จากภาพทางขวามือ (Accuracy over Epochs) เส้นสีฟ้าจะแทนค่าความแม่นยำในการฝึกฝน (Train Accuracy) จะเห็นว่าค่ามันเพิ่มขึ้นจาก 0.20 จนเกือบถึง 1.00 ซึ่งหมายความว่าโมเดลนั้นสามารถจดจำข้อมูลในการฝึกฝนได้ดี แต่เส้นสีส้มที่แทนค่าความถูกต้องในการตรวจสอบ (Validation Accuracy) มีการเรียนรู้ได้ดีในตอนเริ่มและเรียนรู้ได้ดีที่สุดในรอบที่ 150 หลังจากนั้นค่าความถูกต้องในการตรวจสอบ (Validation Accuracy) มีค่าที่ลดลงมาต่ำที่สุดที่ 0.5 และเพิ่มขึ้นไปที่ 0.6 และไม่มีการเพิ่มขึ้นหลังจากนั้น

สรุปค่าความแม่นยำของโมเดล GRU ได้ว่า มีแนวโน้มที่จะเกิดการ overfitting เพราะ Train Loss ลดลงมากในขณะที่ Validation Loss เริ่มเพิ่มขึ้น หรือไม่ไปในทิศทางเดียวกัน และค่า Train Accuracy สูงขึ้นเรื่อยๆ แต่ค่า Validation Accuracy มีค่าที่ไม่เสถียร

2) ค่าการประเมินประสิทธิภาพของการทำนาย (Confusion Matrix) ของโมเดล GRU



ภาพที่ 12 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล GRU

จากรูปที่ 12 แสดงค่า Confusion Matrix ของโมเดล GRU พบว่าหมวดหมู่ที่ทำนายถูกต้องเยอะที่สุดคือ Negative คือ 64 ครั้งจากตัวอย่างข้อมูล 101 ตัวอย่าง แต่เมื่อพิจารณาตามค่า F1-Score ของคลาส Negative จะอยู่ที่ 62% ในขณะที่ประสิทธิภาพในการทำนายมากที่สุดคือคลาส Sexism ซึ่งอยู่ที่ 81% และต่ำที่สุดคือคลาส Body shaming อยู่ที่ 47%

เมื่อพิจารณาการทำนายผิดของโมเดลในแต่ละคลาส พบว่าคลาสที่มีการทำนายผิดมากที่สุดตามค่า Recall คือคลาส Positive โดยมีค่าอยู่ที่ 44% แสดงให้เห็นว่าโมเดลคลาสนี้ได้ดี ในขณะที่คลาสที่ทำนายได้ผิดน้อยที่สุดคือ Sexism โดยมีค่า Recall ที่ 72%

3) ค่า F1-Score ของโมเดล GRU

ตารางที่ 6 รายงานผลความแม่นยำแต่ละป้ายกำกับ GRU

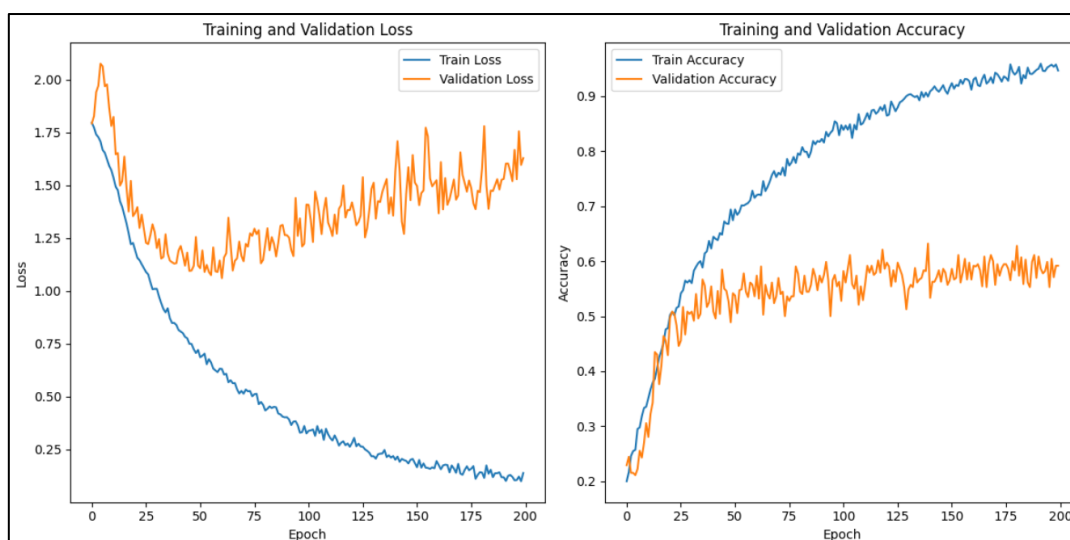
ชุดข้อมูล	ประเภทของข้อความ	ป้ายกำกับ	Precision Score	Recall	F1-Score	Support
1	ข้อความคำทั่วไปการชม	Positive	0.55	0.44	0.49	117
2	ข้อความทั่วไป	Neutral	0.60	0.53	0.56	107
3	ข้อความคำหยาบ	Negative	0.61	0.63	0.62	101
4	ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีดผิว	Body shaming	0.38	0.61	0.47	49
5	ข้อความการเหยียดการศึกษา ชนชั้นฐานะ การเหยียดเชื้อชาติ	Classism	0.66	0.50	0.57	46
6	ข้อความเหยียดเพศ	Sexism	0.63	0.82	0.72	40
Macro average			0.58	0.58	0.57	0.59
Weighted average			0.56	0.55	0.57	0.56

จากตารางที่ 6 จะแสดงถึงค่าความแม่นยำของแต่ละหมวดหมู่ โดยหมวดหมู่ที่ได้ค่า F1-Score ที่ดีที่สุดคือ ข้อความเหยียดเพศ เป็น 72% และน้อยที่สุดคือ ข้อความคำทั่วไปการชม เป็น 44% และค่าเฉลี่ยของค่า F1-Score ทั้งหมดคือ 57% โดยใช้การคำนวณแบบ Weighted Average

4.6 ผลที่ได้จากการฝึกฝนโมเดล BERT+GRU

ผลที่ได้จากการฝึกฝนโมเดล BERT+GRU จะได้ดังนี้

1) ค่าความแม่นยำในการฝึกโมเดล BERT+GRU



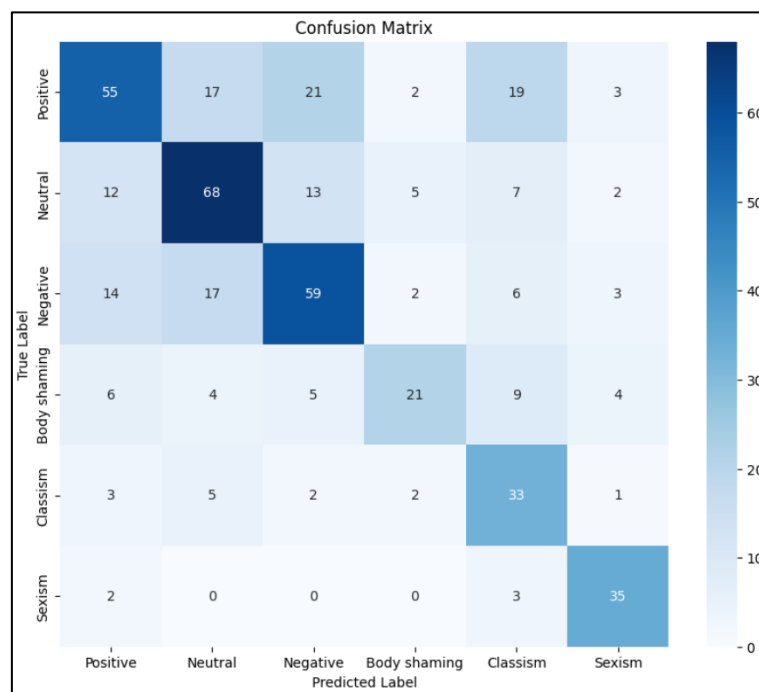
ภาพที่ 13 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล BERT+GRU

จากภาพที่ 14 ทางซ้ายมือ (Loss over Epochs) เส้นสีฟ้าจะแทนค่าความสูญเสียของการฝึกฝน (Train Loss) ลดลงตั้งแต่เริ่ม จนเข้าใกล้ 0 เมื่อถึงรอบที่ 80 หมายความว่าโมเดลนั้นเรียนรู้ได้ดีขึ้นเรื่อยๆ แต่เมื่อถึงการฝึก รอบที่ 60 โมเดลมีการเกิด Overfitting ทำให้เกิดความไม่เสถียรในการฝึกโมเดล และเส้นสีส้มแทนค่าการสูญเสีย การตรวจสอบ (Validation Loss) ตอนเริ่มจะลดลงในตอนเริ่มและเพิ่มขึ้นเรื่อยๆหลังจากการฝึกที่รอบที่ 60 โดยจะมี ค่าตั้งแต่ 1.10 ถึง 1.75

จากภาพทางขวามือ (Accuracy over Epochs) เส้นสีฟ้าจะแทนค่าแม่นยำในการฝึกฝน (Train Accuracy) จะเห็นว่าค่านี้นั้นเพิ่มขึ้นจาก 0.20 จนเกือบถึง 1.00 ภายในการฝึกโมเดล 100 รอบ ซึ่งหมายความว่า โมเดลนั้นสามารถจดจำข้อมูลในการฝึกฝนได้ดี แต่เส้นสีส้มที่แทนค่าความถูกต้องในการตรวจสอบ (Validation Accuracy) มีการเรียนรู้ได้ดีในตอนเริ่มและเรียนรู้ได้ดีที่สุดในรอบที่ 130 หลังจากนั้นค่าความถูกต้องในการ ตรวจสอบ (Validation Accuracy) มีค่าที่ลดลงมาต่ำที่สุดที่ 0.5 และเพิ่มขึ้นไปที่ 0.6 และไม่มีการเพิ่มขึ้นหลังจาก นั้น

สรุปค่าความแม่นยำของโมเดล BERT ได้ว่า มีแนวโน้มที่จะเกิดการ overfitting เพราะ Train Loss ลดลง มากในขณะที่ Validation Loss เริ่มเพิ่มขึ้น หรือไม่ไปในทิศทางเดียวกัน และค่า Train Accuracy สูงขึ้นเรื่อย ๆ แต่ ค่า Validation Accuracy มีค่าที่ไม่เสถียร

2) ค่าการประเมินประสิทธิภาพของการทำนาย (Confusion Matrix) ของโมเดล BERT+GRU



ภาพที่ 14 ค่าการประเมินประสิทธิภาพของการทำนายของโมเดล BERT+GRU

จากรูปที่ 15 แสดงค่า Confusion Matrix ของโมเดล BERT พบว่าหมวดหมู่ที่ทำนายถูกต้องเยอะที่สุดคือ Neutral คือ 68 ครั้งจากตัวอย่างข้อมูล 107 ตัวอย่าง แต่เมื่อพิจารณาตามค่า F1-Score ของคลาส Neutral จะอยู่ที่ 60% ในขณะที่คลาสที่มีประสิทธิภาพในการทำนายมากที่สุดคือคลาส Sexism ซึ่งอยู่ที่ 85% และต่ำที่สุดคือคลาส Positive อยู่ที่ 49%

เมื่อพิจารณาการทำนายผิดของโมเดลในแต่ละคลาส พบว่าคลาสที่มีการทำนายผิดมากที่สุดตามค่า Recall คือคลาส Positive โดยมีค่าอยู่ที่ 38% แสดงให้เห็นว่าโมเดลไม่สามารถทำนายคลาสนี้ได้ดี ในขณะที่คลาสที่ทำนายได้ผิดน้อยที่สุดคือ Sexism โดยมีค่า Recall ที่ 88%

ตารางที่ 7 รายงานผลความแม่นยำแต่ละป้ายกำกับ BERT+GRU

ชุดข้อมูล	ประเภทของข้อความ	ป้ายกำกับ	Precision Score	Recall	F1-Score	Support
1	ข้อความคำทั่วไปการชม	Positive	0.66	0.38	0.49	117
2	ข้อความทั่วไป	Neutral	0.56	0.64	0.60	107
3	ข้อความคำหยาบ	Negative	0.53	0.77	0.63	101
4	ข้อความเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว	Body shaming	0.56	0.51	0.53	49
5	ข้อความการเหยียดการศึกษา ชนชั้นฐานะ การเหยียดเชื้อชาติ	Classism	0.63	0.52	0.57	46
6	ข้อความเหยียดเพศ	Sexism	0.83	0.88	0.85	40
Macro average			0.63	0.62	0.61	460
Weighted average			0.61	0.60	0.59	460

จากตารางที่ 7 จะแสดงถึงค่าความแม่นยำของแต่ละหมวดหมู่ โดยหมวดหมู่ที่ได้ค่า F1-Score ที่ดีที่สุดคือ ข้อความเหยียดเพศ เป็น 85% และน้อยที่สุดคือ ข้อความคำทั่วไปการชม เป็น 49% และค่าเฉลี่ยของค่า F1-Score ทั้งหมดคือ 59% โดยใช้การคำนวณแบบ Weighted Average

4.7 ผลการเปรียบเทียบของทั้ง 4 โมเดล

จากการฝึกโมเดลทั้ง 4 โมเดล ได้แก่ RNN LSTM GRU และ BERT โดยใช้ค่าเฉลี่ย F1-Score เป็นตัววัดผลในการประเมินประสิทธิภาพการทำงานของแต่ละโมเดล สามารถสรุปผลได้ดังนี้

ตารางที่ 8 ผลการเปรียบเทียบของทั้ง 4 โมเดล

ป้ายกำกับ	Precision				Recall				F1-Score			
	RNN	LSTM	GRU	BERT	RNN	LSTM	GRU	BERT	RNN	LSTM	GRU	BERT
Positive	0.52	0.38	0.55	0.66	0.19	0.03	0.44	0.38	0.28	0.05	0.49	0.49
Neutral	0.45	0.53	0.60	0.56	0.58	0.43	0.53	0.64	0.50	0.48	0.56	0.60
Negative	0.50	0.45	0.61	0.53	0.71	0.72	0.63	0.77	0.59	0.56	0.62	0.63
Body shaming	0.41	0.19	0.38	0.56	0.35	0.45	0.61	0.51	0.38	0.27	0.47	0.53
Classism	0.60	0.50	0.66	0.63	0.59	0.57	0.50	0.52	0.59	0.53	0.57	0.57
Sexism	0.61	0.78	0.63	0.83	0.75	0.70	0.82	0.88	0.67	0.74	0.72	0.85
Macro Average	0.52	0.47	0.58	0.63	0.53	0.48	0.58	0.62	0.50	0.44	0.57	0.61
Weighted Average	0.50	0.46	0.56	0.61	0.50	0.43	0.55	0.60	0.47	0.39	0.57	0.59

จากตารางที่ 7 เป็นค่า Precision Recall F1-Score ค่าที่ได้มากที่สุดและค่าที่น้อยได้ดังนี้

1) โมเดล BERT เป็นโมเดลที่มีประสิทธิภาพดีที่สุด คือ ให้ค่า Precision Recall และ F1-Score สูงที่สุดในหลายหมวดหมู่ มี F1-Score สูงสุด ในหมวดหมู่ Sexism 0.85 Neutral 0.60 Negative 0.63 ค่าของ Macro Average 0.61 มี Recall สูงที่สุดในหลายหมวดหมู่ คือ Negative 0.77 Neutral 0.64 และ Sexism 0.88 มีค่า Precision สูง ในบางประเภท คือ Positive 0.66 Body shaming 0.56 และ Sexism 0.83

2) โมเดล GRU เป็นโมเดลที่มีประสิทธิภาพรองจาก BERT มี Precision สูงที่สุดในหลายหมวดหมู่ คือ Neutral (0.60), Negative 0.61 และ Classism 0.66 มีค่า Recall อยู่ในระดับปานกลาง แต่มีค่าที่โดดเด่นในบางประเภท เช่น Body Shaming 0.61 F1-Score ใกล้เคียงกับ BERT ในบางหมวดหมู่ เช่น Classism 0.57 และ Macro Average 0.57

3) โมเดล RNN มี Recall ที่ค่อนข้างสูงในบางหมวดหมู่ เช่น Negative 0.71 และ Sexism 0.75 แต่ Precision และ F1-Score ไม่สูงมาก เมื่อเทียบกับ BERT และ GRU

4) โมเดล LSTM เป็นโมเดลที่มีผลลัพธ์แย่ที่สุดโดยรวม ให้ค่า Precision Recall และ F1-Score ต่ำที่สุดในหลายหมวดหมู่ ตัวอย่างเช่น F1-Score ของ Positive 0.05 และ Body Shaming 0.27 จะให้ได้ชี้ชัดว่ามีค่าที่ต่ำมาก

บทที่ 5

บทสรุป

1. สรุปผล

ผู้จัดทำได้ทำการพัฒนาโมเดลโดยใช้หลักการของการจำแนกข้อความ ในการจำแนกหมวดหมู่ข้อความ โดยได้รับชุดข้อมูลมาจาก NLP for Thai Text Classification thai_cyberbullying_lgbt PyThaiNLP/wiseweight-sentiment เพจข่าวออนไลน์ต่างๆบน Facebook TikTok X และจากการสร้างแบบสอบถามการใช้คำหยาบคายบนสื่อโซเชียลมีเดีย นำชุดข้อมูลที่ได้รับมาทำการแยกหมวดหมู่โดยแบ่งเป็นทั้งหมด 6 หมวดหมู่จาก 2300 ประโยค จากนั้นเตรียมข้อมูลสำหรับการฝึกโมเดล โดยได้นำโมเดลมาทดลองทั้งหมด 4 โมเดล ได้แก่ Recurrent Neural Network (RNN) Long Short-Term Memory (LSTM) Gated Recurrent Unit (GRU) และ Bidirectional Encoder Representations from Transformers ทดสอบด้วยชุดข้อมูลโดยแบ่งเป็นชุดเรียนรู้ 70% ชุดทดสอบ 20% ชุดตรวจสอบ 10% และฝึกฝนทั้งหมด 200 รอบ จากผลการประเมินประสิทธิภาพของโมเดล (Model Evaluation Metrics) โมเดลที่มีประสิทธิภาพมากที่สุดคือ Bidirectional Encoder Representations from Transformers มีค่า F1-Score อยู่ที่ 0.59 ต่อมาเป็น Gated Recurrent Unit (GRU) มีค่า F1-Score (Weighted Average) เป็น 0.57 ซึ่งดีกว่า Long Short-Term Memory (LSTM) และ Recurrent Neural Network (RNN) ที่มีค่าเท่ากับ 0.47 และ 0.39 ตามลำดับ ดังนั้นจะสรุปได้ว่า Bidirectional Encoder Representations from Transformers เป็นโมเดลที่มีค่าที่ดีที่สุดในปัจจุบันที่ผู้พัฒนาโมเดลได้ทำการทดสอบและพัฒนาโมเดลมาจนถึงตอนนี้

2. ปัญหาและอุปสรรค

1. ความกำกวมทางภาษาหรือมีความคลุมเครือของข้อความ
2. ภาษาที่สามารถเกิดขึ้นได้ทุกวันอาจทำให้โมเดลมีความล้าหลังในทางด้านข้อมูล
3. ป้ายกำกับ (Label) มีไม่มากพอทำให้การฝึกโมเดลไม่มีประสิทธิภาพ
4. โมเดลยังไม่สามารถระบุว่าเป็นข้อความการกลั่นแกล้งแบบใดได้
5. ชุดข้อมูลยังไม่มีคุณสมบัติ

3. ข้อเสนอแนะ

1. เพิ่มป้ายกำกับ (Label) ให้กับชุดข้อมูลก่อนนำไปฝึกโมเดล
2. ทำโมเดลให้สามารถระบุข้อความว่าเป็นการกลั่นแกล้งแบบใด
3. เพิ่มชุดข้อความที่มีความหลากหลายและครอบคลุม
4. ลดจำนวนรอบในการฝึกโมเดล

เอกสารอ้างอิง


1. ณัฐาศิริ เชาว์ประสิทธิ์, กฤษฎา ศรีแก้ว. (2563). การพัฒนาโมเดลตรวจจับคำหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคด้าโมเนนิง. มหาวิทยาลัยรัตนบัณฑิต.
2. ณัฐาศิริ เชาว์ประสิทธิ์. (2560). การจำแนกประเภทข้อความภาษาไทยบนสื่อออนไลน์ในเชิงบวกและลบด้วยคลังคำหยาบ. ปทุมธานี: มหาวิทยาลัยรังสิต.
3. MonkeyLearn. (n.d.). Go-to Guide for Text Classification with Machine Learning.
4. Amazon Web Services. (n.d.). What is a recurrent neural network (RNN)?. AWS.
5. GoogleDevelopers. (n.d.). Text classification algorithms. Google Developers.
6. CYBER ELITE. (n.d.). Machine Learning เทคโนโลยีประโยชน์ครอบคลุมจักรวาล. Cyber Trust & Resilience Simplified.
7. Great Learning Team. (2022). Tokenising into Words and Sentences | What is Tokenization and it's Definition?.
8. Kanoktipsatharporn, S. (2018, November 23). Natural Language Processing (NLP) คืออะไร รวมคำศัพท์เกี่ยวกับ Natural Language Processing (NLP) – NLP ep.1. BUA Labs.
9. Softnix. (2019, May 28). TF-IDF ทำงานยังไง? [How TF-IDF works?].
10. Kanoktipsatharporn, S. (2020, November 17). การทำความเข้าใจเทคโนโลยี Chatbot ทำงานอย่างไร BUA Labs.
11. Written by Natthanan Bhukan. (2020). Deploy model machine learning ฉบับมือใหม่.
12. Ikonomakis, E. K., Kotsiantis, S., & Tampakas, V. (2005). Text Classification Using Machine Learning Techniques. WSEAS Transactions on Computers.
13. Chatterjee, R., Bhattacharya, S., & Kabi, S. (2021). Profanity detection in social media text using a hybrid approach of NLP and machine learning. International Journal of Advance Research, Ideas and Innovations in Technology, 7(1), 292-294.
14. Pagongatchalee, N. (2021, August 16). Confusion matrix: เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนายใน machine learning. Medium.
15. Sinart. (2560). Long short-term memory (LSTM). Medium.
16. Towards Data Science. (2018, March 24). Understanding GRU networks. Towards Data Science.


เอกสารอ้างอิง (ต่อ)

17. Das, S., Datta, S., & Obaid, (2023). **Applying interpretable machine learning to classify traffic injury types using BERT**. Natural Language Processing Journal. Retrieved from sciencedirect

ประวัติผู้เขียน

นายธนากร ผาเป้า เกิดเมื่อวันที่ 28 มกราคม พ.ศ. 2545 และ นายญาณวิทย์ รอบไธสง เกิดเมื่อวันที่ 12 พฤศจิกายน พ.ศ. 2545 นักศึกษาชั้นปีที่ 4 หลักสูตร วิทยาการคอมพิวเตอร์ สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น เมื่อปีการศึกษา 2564 เข้าศึกษาในระดับบัณฑิตศึกษา สาขาวิชา วิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

ลงชื่อผู้ทำโครงการ 
(นายธนากร ผาเป้า)

ลงชื่อผู้ทำโครงการ 
(นายญาณวิทย์ รอบไธสง)

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงการ

.....
.....
.....

(ลงชื่อ)
(ดร.พงษ์ศธร จันทร์ยอย)