****

**CS 2567/-**

**เอกสารการเสนอหัวข้อสัมมนาทางวิทยาการคอมพิวเตอร์**

จำแนกข้อความการกลั่นแกล้งทางไซเบอร์

Cyberbullying Text Classification

โดย

643020613-5นายธนากร ผาเป้า

643020604-6นายญาณวิทย์ รอบไธสง

อาจารย์ที่ปรึกษาโครงงาน ดร.พงษ์ศธร จันทร์ยอย

ตําแหน่ง อาจารย์

วิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

รายงานเล่มนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314774 โครงงานวิทยาการคอมพิวเตอร์ 1

ภาคการเรียนที่ 1 ปีการศึกษา 2567

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน กรกฎาคม พ.ศ. 2567)

**A black and white logo

Description automatically generated**

**CS 2567/-**

**เอกสารการเสนอหัวข้อสัมมนาทางวิทยาการคอมพิวเตอร์**

จำแนกข้อความการกลั่นแกล้งทางไซเบอร์

Cyberbullying Text Classification

โดย

643020613-5นายธนากร ผาเป้า

643020604-6นายญาณวิทย์ รอบไธสง

อาจารย์ที่ปรึกษาโครงงาน ดร.พงษ์ศธร จันทร์ยอย

ตําแหน่ง อาจารย์

วิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

รายงานเล่มนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314774 โครงงานวิทยาการคอมพิวเตอร์ 1

ภาคการเรียนที่ 1 ปีการศึกษา 2567

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน กรกฎาคม พ.ศ. 2567)

**การเสนอเค้าโครงโครงงานคอมพิวเตอร์**

**สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น**

**ชื่อ** นายธนากร ผาเป้า **รหัสประจำตัว** 643020613-5

Mr.Thanakon phapao

นายญาณวิทย์ รอบไธสง **รหัสประจำตัว** 643020604-6

Mr**.**Yarnnawitrobthaisong

**นักศึกษาระดับปริญญาตรี** หลักสูตรวิทยาศาสตรบัณฑิต วิทยาการคอมพิวเตอร์

**อาจารย์ที่ปรึกษาโครงงาน ดร.พงษ์ศธร จันทร์ยอย**

**Project Advisor** Dr.Pongsathon Janyoi

**1. ชื่อหัวข้อโครงงาน**

ภาษาไทย จำแนกข้อความการกลั่นแกล้งทางไซเบอร์

ภาษาอังกฤษ Cyberbullying Text Classification

## 2. หลักการและเหตุผล

ปัจจุบันการกลั่นแกล้งบนโซเซียลมิเดียด้วยการพิมพ์ข้อความที่มีคำหยาบ การดูถูก เหยียดเชื้อชาติ รูปลักษณ์ร่างกาย สีผิว การศึกษาแม้กระทั่งชนชั้นฐานะ ส่งผลให้ผู้โดนกระทำได้รับผลกระทบทางด้านจิตใจและอาจถึงทางด้านร่างกาย ผู้ที่ถูกกลั่นแกล้งนั้นย่อมเกิดความเครียด วิตกกังวล ซึมเศร้า เก็บตัว โดดเดี่ยว ไปจนถึงมีความคิดที่จะฆ่าตัวตาย นอกจากนี้การกลั่นแกล้งบนสื่อโซเซียลมีเดียยังส่งผลกระทบต่อสังคมโดยรวม ทำให้เกิดความไม่ปลอดภัยและอาจนำไปสู่ความรุนแรงได้

จากปัญหาดังกล่าว ผู้วิจัยจึงได้ศึกษาและจัดทำโมลเดลในการตรวจจับข้อความที่ไม่เหมาะสม จากโพสต์หรือความคิดเห็น (Comment) ต่างๆที่มีแนวโน้มที่ก่อให้เกิดความไม่ปลอดภัย และได้ทำการนำข้อความจากประโยคนั้นๆ จำแนกเป็นหมวดหมู่ เช่นข้อความปกติ ข้อความการเหยียดเพศ ข้อความการเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว ข้อความการเหยียดการศึกษา ชนชั้นฐานะ และข้อความการเหยียดเชื้อชาติ ประโยคข้อความที่มีแนวโน้มจะก่อให้เกิดปัญหาที่จะก่อให้เกิดความรุนแรงบนสื่อโซเซียลมีเดีย ซึ่งโมเดลการคัดแยกประเภทคำการเหยียดจะช่วยให้สามารถคัดแยกคำต่างๆบนสื่อโซเซียลมีเดียได้

โดยโมเดลการจำแนกข้อความการกลั่นแกล้งทางไซเบอร์ จะมุ่งไปที่การจำแนกคำการกลั่นแกล้งในเป็นหมวดหมู่ ประสิทธิภาพและความแม่นยำในการจำแนกข้อความเพื่อให้โมเดลการจำแนกข้อความสามารถใช้งานได้อย่างมีประสิทธิภาพ

### **3. วัตถุประสงค์ของโครงงาน**

เพื่อพัฒนาโมเดลสำหรับการจำแนก คำหยาบ การดูถูปรก เหยียดเชื้อชาติ รูปลักษณ์ร่างกาย สีผิว การศึกษาชนชั้นฐานะ

#### **4.** **ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง**

4.1 ทฤษฏีที่เกี่ยวข้อง

4.1.1 การจำแนกข้อความ (Text classification) เป็นกระบวนการวิเคราะห์ข้อความและกำหนดป้ายกำกับให้กับข้อความนั้นตามเนื้อหาของข้อความ ป้ายกำกับเหล่านี้อาจเป็นประเภทต่างๆ กัน เช่น ประเภทของหัวข้อ ประเภทของอารมณ์ ประเภทของวัตถุประสงค์ การจำแนกประเภทข้อความสามารถทำได้โดยใช้วิธีการต่างๆ กัน โดยทั่วไปแล้ว วิธีการจำแนกประเภทข้อความสามารถแบ่งออกเป็นสองประเภทใหญ่ๆ

วิธีการตามกฎ (Rule-Based-Methods) เป็นวิธีการจำแนกประเภทข้อความตามกฎจะอาศัยกฎทางภาษาศาสตร์ที่จัดทำขึ้นอย่างชัดเจน ระบบจะใช้กฎเหล่านี้เพื่อกำหนดว่าข้อความที่กำหนดควรอยู่ในคลาสใด โดยมองหารูปแบบขององค์ประกอบข้อความที่เกี่ยวข้องกับความหมาย ตัวอย่างเช่น เราสามารถเขียนกฎเพื่อตรวจจับสแปมอีเมลได้ดังนี้

วิธีการเรียนรู้เชิงเครื่องจักร (Machine Learning Methods) วิธีการจำแนกประเภทข้อความโดยอาศัยอัลกอริทึมการเรียนรู้เชิงเครื่องจักรเพื่อเรียนรู้ความสัมพันธ์ระหว่างข้อความและป้ายกำกับ อัลกอริทึมการเรียนรู้เชิงเครื่องจักรที่นิยมใช้สำหรับการจำแนกประเภทข้อความ ได้แก่ อัลกอริทึม logistic regression อัลกอริทึม support vector machine อัลกอริทึม neural network ฯลฯ วิธีการจำแนกประเภทข้อความการเรียนรู้เชิงเครื่องจักรมีความแม่นยำสูงกว่าวิธีการตามกฎ (Rule-Based-Methods) แต่ต้องใช้ข้อมูลตัวอย่างจำนวนมากสำหรับการฝึกโมเดล

A diagram of a classifier

Description automatically generated ตัวอย่างกรณีการใช้งานการจำแนกประเภทข้อความ การตรวจจับสแปม ใช้ในการตรวจจับอีเมลสแปมหรือข้อความสแปมอื่นๆ

**ภาพที่ 1** การจำแนกข้อความ

4.1.2 การเรียนรู้ของเครื่อง (Machine learning) อาศัยการเรียนรู้จากข้อมูลตัวอย่างเพื่อตรวจจับคำหยาบหรือคำที่สื่อถึงคำพูดในเชิงการกลั่นแกล้ง เช่น การใช้โมเดลภาษา (language model) ที่สามารถทำนายหรือสร้างข้อความใหม่ตามที่ถูกต้องและมีความหมาย และใช้หลักการ text classification เพื่อที่ให้ตัวโมเดลนั้นสามารถเข้าใจและตรวจจับคำหยาบคาย

A screen shot of a computer

Description automatically generated

**ภาพที่ 2** ขั้นตอนการแท็กข้อความ

4.1.3 การตัดคำ (Word tokenization) เป็นกระบวนการแบ่งข้อความที่กำหนดออกเป็นหน่วยที่เรียกว่าโทเค็น (Token) โทเค็นอาจเป็นคำ วลี หรือแม้แต่ประโยคทั้งหมดก็ได้ ในกระบวนการสร้างโทเค็น อักขระบางตัว เช่น เครื่องหมายวรรคตอนอาจถูกละทิ้ง

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 3** ตัวอย่างการตัดคำและประโยค

4.1.4 โครงข่ายประสาทเทียมที่เกิดซ้ำ (Recurrent Neural Network) เป็นโมเดลที่นำมาฝึกพิจารณาแต่ละจุดใน Sequence เพื่อช่วยให้การทำนายมีประสิทธิภาพเพิ่มขึ้นโดยการทำงานของ RNN จะทำการรับข้อมูลเข้ามาหรือข้อมูลจากที่ผู้ใช้ทำพิมพ์เข้าไป RNN จะมี Activation function ฟังก์ชั่นการเปิดใช้งาน เช่น ผู้ใช้ต้องการพิมพ์ "Loding" เป็น "Lodink" ระบบจะทำการพิจารณา Activation functions ของอักษรก่อนหน้าจาก "Lodin" ร่วมกันกับ "k" ที่ผู้รูปภาพประกอบด้วย วิชาการพิมพ์, ออกแบบ, ภาพประกอบ

คำอธิบายที่สร้างโดยอัตโนมัติโดยมีความน่าเชื่อถือต่ำใช้ได้พิมพ์เข้าไป และโมเดลจะสามารถทำให้ผลลัพท์ "g" ที่ถูกต้องได้ ตัวอย่างดังรูป

**ภาพที่ 4** ตัวอย่างการทำงานของ4 โครงข่ายประสาทเทียมที่เกิดซ้ำ

4.1.5 การคัดแยกคำตามความสำคัญ (TF-IDF) เป็นวิธีการในการคำนวณความสำคัญของคำศัพท์ในเอกสารหรือเอกสารเอกเปรียบเทียบกันโดยใช้วิธีการนี้เพื่อดูว่าคำนั้นมีความสำคัญแค่ไหนในเอกสารเหล่านั้นๆ โดย TF นั้นจะย่อมาจาก Term Frequency แปลว่าความถี่ของคำ ซึ่ง TF นั้นจะดูว่าคำนั้นปรากฎขึ้นบ่อยเพียงใดในแต่ละประโยค โดยจะใช้ในการหาค่าความถี่ของคำในประโยคด้วยสูตรที่ว่า คำที่ปรากฎ/คำทั้งหมดของทั้งประโยค จะได้ค่าความถี่ของของคำในประโยค ต่อมาคือ IDF ย่อมาจาก Inverse Document Frequency แปลว่าการผกผันในความถี่ ซึ่งจะเป็นการ โดยจะมีสูตรการหาคือ IDF(t) = log [ (1 + n) / (1 + DF(t)) ] + 1 โดย t คือ term หรือคำ 1 คำ n คือจำนวน Document ทั้งหมดที่มีอยู่ DF(t) คือจำนวนประโยคที่พบของคำ t

4.1.6 ถุงคำ (Bag of Words) เป็นเทคนิคที่ใช้ในการแปลงข้อความเป็นเวกเตอร์ (Vector) ของคำที่ปรากฏในข้อความ โดยใช้เพียงความถี่ของคำเท่านั้น โดยใช้ประโยชน์ในการแบ่งประเภทข้อมูลที่เกี่ยวข้องกับข้อความ วิเคราะห์ความถี่ของคำ ทำให้สามารถระบุคำที่มีความสำคัญในบทความ โพสต์บนโซเชียลมีเดีย และความคิดเห็นบนโซเชียลมีเดีย

4.1.7 การปรับใช้โมเดล (Deploy model machine learning) กระบวนการที่นำโมเดล (model) ที่ได้สร้างและฝึก (Trained) ด้วยเทคนิค Machine Learning หรือ Deep Learning ไปใช้งานจริง (putting it into production) ซึ่งสามารถให้บริการหรือใช้งานได้จริงในสภาพแวดล้อมที่สอดคล้องกับการใช้งานจริงของผู้ใช้

กระบวนการการทำงานของการ การปรับใช้โมเดล จะเป็นดังนี้

รูปภาพประกอบด้วย สัญลักษณ์, เครื่องหมาย, ตัวอักษร, ออกแบบ

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 5** ขั้นตอนการปรับแต่งโมเดล

จากภาพที่ 5 จะเห็นว่าจะมีการใส่โมเดลเข้าไปเพื่อปรับแต่ง การปรับแต่งโมเดลจะสามารถปรับแต่งได้จนพอใช้แล้วจึงนำมาใช้

รูปภาพประกอบด้วย ตัวอักษร, เครื่องหมาย, สัญลักษณ์, ออกแบบ

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 6** การบันทึกโมเดลเพื่อนำไปใช้

จากภาพที่ 6 จะเป็นการบันทึกโมเดล ที่เราได้ปรับปรุงเสร็จแล้วนำมาปรับใช้มีการโหลดแล้วนำมาทำนายเพื่อให้ได้ผลลัพธ์ และเมื่อเราได้ทำการทดสอบเสร็จสินแล้วไม่มีปัญหา ขั้นตอนสุดท้ายก็คือ การบันทึก นั้นมาใช้งานตามที่เราต้องการที่จะใช้

4.2 งานวิจัยที่เกี่ยวข้อง

4.2.1 Text Classification Using Machine Learning Techniques ของ M.Ikonomakis , S. Kotsiantis เกี่ยวกับ การจำแนกประเภทข้อความ โดยจะมีเนื้อหาเกี่ยวกับความแม่นยำ ในการวัดประสิทธิภาพการจำแนกข้อความ การใช้ค่าความแม่นยำ (Precision) และ การจำ (Recall) ในกรณีที่ข้อมูลไม่สมดุล การผสมค่าความแม่นยำ และการจำเข้าด้วยกันเพื่อให้ได้ภาพรวมของประสิทธิภาพของตัวจำแนกข้อความ การแทนเอกสารเป็นเวกเตอร์ การทำ การประมวลผลล่วงหน้า (preprocessing) เช่นการลบ การหยุดคำ (stop words) และการใช้ การกั้น (stemming) เพื่อลดขนาดของชุดคุณลักษณะ (feature set)

4.2.2 การพัฒนาโมเดลตรวจจับคําหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิง ของ ณัฐาศิริ เชาว์ประสิทธิ์ และ สมชาย เล็กเจริญ เป็นงานวิจัยที่เกี่ยวข้องกับกระบวนการวิเคราะห์คำหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิง โดยใช้โมเดลในการตรวจจับคำหยาบภาษาไทยด้วยพจนานุกรมคําหยาบที่ผ่านการปรับปรุงใช้เทคนิค TFICF (Term Frequency - Inverse Class Frequency) ใช้ข้อมูลประเภทตัวอักษรบนกระดานสนทนา หรือข้อความแสดงความคิดเห็นตามสื่อบันเทิงออนไลน์ต่างๆ จากเว็บไซต์ ข่าว บันเทิง เกมออนไลน์ และประมูลขายสินค้า เว็บไซต์ดังกล่าวมีการเปิดเสรี รวมถึงการไม่ปิดกั้นคําที่ไม่เหมาะสม หรือคําหยาบ ในการแสดงความคิดของผู้ใช้บริการ ผู้วิจัยได้เก็บตัวอย่างที่ใช้ทดลองได้จํานวนข้อความ 1,214 โพสต์ ( 1 โพสต์ความยาวไม่เกิน 300 ตัวอักษร) ข้อมูลที่นํามาใช้ในการทดสอบนี้เป็นการสุ่มตัวอย่างเพื่อนํามาวิเคราะห์

4.2.3 Tokenization as the initial phase in NLP ของ Jonathan Webster เป็นงานวิจัยที่เกี่ยวข้องกับ Tokenization โดยกล่าวว่า Tokenization เป็นขั้นตอนที่สำคัญในการประมวลภาษาธรรมชาติ (Natural Language Processing) โดยมีปัญหาที่มันซับซ้อน ตัวอย่างเช่น การแบ่งสตริงเป็นหน่วยที่มีความหมาย การจัดกลุ่มคำที่มีความสัมพันธ์กัน Tokenization เป็นการจำแนกข้อความเป็นหน่วยที่มีความหมายและสามารถพัฒนา NLP ให้มีประสิทธิภาพมากขึ้น

4.2.4 Profanity detection in social media text using a hybrid approach of NLP and machine learning ของ Raktim Chatterjee, Sukanya Bhattacharya เกี่ยวกับ การตรวจจับคำหยาบในข้อความบนโซเชียลมีเดียโดยใช้วิธีการผสมผสานระหว่าง NLP และ Machine Learning ซึ่งมีเนื้อหาประกอบด้วยการวิเคราะห์ระบบการตรวจจับคำหยาบที่มีอยู่แล้วและวิเคราะห์ระบบที่นำเสนอใหม่ที่ใช้ Natural Language Processing และ Machine Learning ในการตรวจจับภาษาที่ไม่เหมาะสมบนแพลตฟอร์มโซเชียลมีเดีย รวมถึงการใช้วิธีการแบบ supervised learning problem และการใช้ Generic features เช่น Bag-Of-Words หรือ embeddings รวมถึงการใช้ Lexical resources ในการประมวลผลข้อมูลเพื่อให้ได้ผลลัพธ์ที่แม่นยำมากขึ้นในการจำแนกภาษาที่ไม่เหมาะสม

## 5. วิธีดำเนินการวิจัย

5.1 ศึกษาและวิเคราะห์ปัญหา

ผู้จัดทำได้ศึกษาและวิเคราะห์ปัญหาที่เกิดจากข้อความการกลั่นแกล้งกันในสังคมไซเบอร์ เพื่อที่จะจำแนกประเภทข้อความต่างๆ ออกเป็น ข้อความปกติ ข้อความการเหยียดเพศ ข้อความการเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว ข้อความการเหยียดการศึกษา ชนชั้นฐานะ และข้อความการเหยียดเชื้อชาติ และเพื่อพัฒนาโมเดลการจำแนกข้อความและหมวดหมู่ข้อความ

5.2 ค้นคว้าทฤษฎีเกี่ยวกับวิจัยที่เกี่ยวข้อง

ผู้จัดทำได้ค้นคว้าทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง เพื่อช่วยให้ผู้จัดทำสามารถต่อยอดและพัฒนาโมเดลที่ต้องการได้ เช่น ทฤษฎี การจำแนกข้อความ ในการจำแนกหมวดหมู่ของคำ หรือ การตัดคำ จากประโยคหรือบทความต่างๆงานวิจัยเกี่ยวกับการพัฒนาโมเดลในการตรวจจับคำหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคการวิเคราะห์จำแนกข้อมูล ซึ่งนำข้อมูลจาก NLP for Thai Text Classification thai\_cyberbullying\_lgbt และ PyThaiNLP/wisesight-sentiment มาเพื่อใช้ดำเนินงาน

5.3 ศึกษาข้อมูลและวิธีการใช้เครื่องมือในการดำเนินงาน

ผู้จัดทำได้ทำการศึกษาข้อมูลของเครื่องมือที่ใช้ เช่น วิธีการใช้ Google Colab Python nlktk และ Scikit-learn ในการพัฒนาโมเดล หลักการทำงานของการ ฝึกโมเดล การจำแนกข้อความ การตัดคำ และการปรับแต่งโมเดลเพื่อนำมาใช้งานและพัฒนาต่อยอด

5.4 เขียนเค้าโครงโครงงานและเสนออาจารย์ที่ปรึกษา

ผู้จัดทำได้ทำการเขียนเค้าโครงของโครงงาน เรื่องจำแนกข้อความการกลั่นแกล้งทางไซเบอร์ รายละเอียดและความเป็นไปได้ของโมเดล แล้วจึงนำไปปรึกษาอาจารย์ที่ปรึกษา เพื่อที่จะนำเรื่องดังกล่าวไปจัดทำเป็นโครงงาน

5.5 นำเสนอเค้าโครงงาน

ผู้จัดทำได้นำเสนอเค้าโครงของโครงงานแก่อาจารย์ที่ปรึกษาเพื่อขอคำปรึกษาและขอคำแนะนำในการพัฒนาโมเดล

5.6 วิเคราะห์และออกแบบโมเดล

ผู้จัดทำได้วิเคราะห์ วางแผนในความเป็นไปได้ของทิศทางที่จะพัฒนาโมเดลและทำการออกแบบแนวทางของโมเดลให้ได้ตามความต้องการของผู้จัดทำเพื่อตอบโจทย์ตามวัตถุประสงค์

5.7 สร้างและพัฒนาโมเดล

หลังจากที่ผู้จัดทำได้ทำการออกแบบ วางแนวทางในการพัฒนาโมเดล ผู้จัดทำได้ทำการลงมือสร้างโมเดล โดยการนำข้อมูล ที่ได้มาจาก NLP For thai ซึ่งข้อมูลที่ผู้จัดทำได้นำมาใช้ จะมีสองตัวหลักๆคือ Thai cyberbullying lgbt และ wisesight sentiment นำมาทำการจำแนกข้อความออกเป็นหมวดๆ เช่น คำปกติ คำเหยียดสีผิว ขนขั้น และได้พัฒนาความแม่นยำของโมเดลในการจำแนกหมวดหมู่ของคำให้มีประสิทธิภาพมากขึ้น

5.8 ทดสอบและปรับปรุงแก้ไข

หลังจากที่ผู้จัดทำได้การสร้างโมเดล ผู้จัดทำได้ทำการทดสอบโมเดลเพื่อที่จะหาข้อบกพร่องของโมเดล มาพัฒนาโมเดลให้ได้ตรงกับความต้องการและมีประสิทธิภาพที่ดีให้มากที่สุด

5.9 จัดทำเล่มรายงงาน

ผู้จัดทำได้ทำการจัดทำรูปเล่มรายงาน เพื่อสรุปผลการดำเนินงาน ความคืบหน้าในการดำเนินการ และนำไปเสนอในรายวิชาระเบียบวิธีวิจัยทางวิทยาการคอมพิวเตอร์

5.10 จัดทำเล่มรายงงาน

นำเสนอโมเดลและผลการพัฒนาโมเดล

###### 6. ขอบเขตและข้อจำกัดของการวิจัย

### 6.1 ข้อจำกัดในการใช้งาน

6.1.1 ไม่สามารถจำแนกคำที่ไม่เคยพบมาก่อนได้

6.1.2 การจำแนกมีความซับซ้อนของภาษาและการให้ความหมายเปลี่ยนไปได้

### 6.2 ความสามารถของระบบ

6.2.1 จำแนกประโยคปกติได้

6.2.2 จำแนกประเภทข้อความต่างๆ ออกเป็น ข้อความปกติ ข้อความการเหยียดเพศ ข้อความการเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว ข้อความการเหยียดการศึกษา ชนชั้นฐานะ และข้อความการเหยียดเชื้อชาติ ได้

6.2.3 เข้าใจประโยคและวลีได้ถูกต้อง

## 7. สถานที่ทำวิจัย

สาขาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

## 8. ประโยชน์ที่คาดว่าจะได้รับ

1. ได้รับโมเดลการจำแนกข้อความปกติ ข้อความการเหยียดเพศ ข้อความการเหยียดรูปลักษณ์ร่างกาย หน้าตา สีผิว ข้อความการเหยียดการศึกษา ชนชั้นฐานะ และข้อความการเหยียดเชื้อชาติ

2. ได้เรียนรู้เกี่ยวกับปัญหาการกลั่นแกล้งบนสื่อโซเซียลมีเดีย สาเหตุ ผลกระทบ และแนวทางในการแก้ไขปัญหา

3. ได้เผยแพร่ความรู้เกี่ยวกับปัญหาการกลั่นแกล้งบนสื่อโซเซียลมีเดียสู่สังคม

4. พัฒนาทักษะเกี่ยวกับ NLP และ Machine learning

**9.** **แผนและระยะเวลาดำเนินการ**

ในการดำเนินงานโครงการและพัฒนาโมเดลมีระยะเวลา 5 เดือน โดยเริ่มดำเนินงานในวันที่ 17 มิถุนายน 2567 จนถึงวันที่ 17 ตุลาคม 2567

## 10. งบประมาณ

หมวดวัสดุอุปกรณ์

* ค่าวัสดุสำนักงาน (กระดาษ ปากกา และอื่นๆ )

หมวดค่าใช้สอย

* ค่าถ่ายเอกสาร
* ค่าจัดรูปเล่ม

หมวดค่าใช้จ่ายอื่นๆ

-

**เอกสารอ้างอิง**

[1] ณัฐาศิริ เชาว์ประสิทธิ์, กฤษฎา ศรีแผ้ว. (2563). *วิเคราะห์คําหยาบภาษาไทยบนสื่อออนไลน์ด้วยเทคนิคดาต้าไมน์นิง. มหาวิทยาลัยรัตนบัณฑิต.*

[2] 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), 292-297

[3] CYBER ELITE. (n.d.). *Machine Learning เทคโนโลยีประโยชน์ครอบจักรวาล.* Cyber Trust & Resilience Simplified.

[4] MonkeyLearn. (n.d.). *Go-to Guide for Text Classification with Machine Learning*. <https://monkeylearn.com/blog/text-classification-machine-learning/>

[5] Great Learning Team. (2022). *Tokenising into Words and Sentences | What is Tokenization and it’s Definition?.*

[6] Google Developers. (n.d.). *Text classification*. Google Developers. <https://developers.google.com/machine-learning/guides/text-classification>

[7] Written by Natthanan Bhukan .(2020). *Deploy model machine learning ฉบับมือใหม่.*

[8] Ikonomakis, E. K., Kotsiantis, S., & Tampakas, V. (2005). *Text Classification Using Machine Learning Techniques*. WSEAS Transactions on Computers.

ลงชื่อผู้ทำโครงงาน ................................................

(นายธนากร ผาเป้า)

วันที่ 12 กรกฎาคม 2567

ลงชื่อผู้ทำโครงงาน ................................................

(นายญาณวิทย์ รอบไทสง)

วันที่ 12 กรกฎาคม 2567

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงงาน

………………………………………………………………………………………………………………………………………………………….……..……………………………………………………………………………………………………………………….……..……………………………………………………………………………………………………………………………………………………………………………………………………………………………

(ลงชื่อ) ..................................................

(ดร.พงษ์ศธร จันทร์ยอย)

วันที่ 12 กรกฎาคม 2567