

# ระบบคัดกรองข้อความที่เข้าข่ายการละเมิดทางไซเบอร์

ทัชธร สืบอินทร์<sup>1</sup> และ เอกชัย พนมทิพย์<sup>2</sup>

<sup>1</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

<sup>2</sup>คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Emails: 60070138@it.kmitl.ac.th, 60070167@it.kmitl.ac.th

## บทคัดย่อ

ในยุคปัจจุบัน สื่อสังคมออนไลน์นั้นได้รับความนิยมเป็นอย่างมาก แม้สื่อสังคมออนไลน์นั้นจะเป็นช่องทางที่ผู้ใช้สามารถรับรู้ข่าวสาร ความเคลื่อนไหวต่างๆ ของสังคม บุคคลรอบตัว หรือเพื่อนฝูงได้อย่างสะดวกรวดเร็วแล้ว ปัญหาหนึ่งที่เกิดขึ้นตามมาอย่างหลีกเลี่ยงไม่ได้ คือ การละเมิดทางไซเบอร์ หรือการกลั่นแกล้งผู้อื่นผ่านทางสื่อสังคมออนไลน์ ไม่ว่าจะเป็นในรูปแบบของข้อความ รูปภาพ หรือสื่อต่างๆ ที่ทำให้สามารถเห็นได้อย่างสาธารณะ ซึ่งปัญหานี้มักจะส่งผลกระทบต่อสภาพจิตใจของผู้ถูกโจมตีโดยตรง การเฝ้าระวังรักษาสภาพจิตใจของผู้ได้รับผลกระทบนั้น มักเป็นเรื่องที่ซับซ้อนละเอียดอ่อน และใช้เวลา ทางคณะผู้จัดทำเล็งเห็นความสำคัญในการพัฒนาระบบคัดกรองข้อความละเมิดทางไซเบอร์เพื่อเป็นการป้องกัน หรือลดผลกระทบที่อาจจะเกิดขึ้น

ในโครงงานนี้ ผู้จัดทำได้นำเสนอระบบที่สามารถตรวจสอบข้อความจากสื่อสังคมออนไลน์ต่างๆ ของผู้ใช้ด้วยอัลกอริทึมของการเรียนรู้ของเครื่อง (Machine learning) เพื่อระบุว่า ข้อความเหล่านั้นเป็นข้อความที่เข้าข่ายการละเมิดทางไซเบอร์หรือไม่ **คำสำคัญ** – การละเมิดทางไซเบอร์; สื่อสังคมออนไลน์; การเรียนรู้ของเครื่อง; Machine learning

## 1. บทนำ

การละเมิดทางไซเบอร์ (Cyberbullying) ผ่านทางสื่อสังคมออนไลน์ ถือเป็นปัญหาใหญ่ปัญหาหนึ่งในสังคมปัจจุบัน ที่ส่งผลกระทบต่อคนจำนวนมาก จากการสำรวจพบว่า ร้อยละ 75 ของกลุ่มบุคคลที่มีการเข้าถึงอินเทอร์เน็ต ซึ่งมีอายุอยู่ในช่วง 5-28 ปีนั้น ร้อยละ 80 เคยมีประสบการณ์การถูกละเมิดทางไซเบอร์ เนื่องจากสื่อสังคมออนไลน์ เป็นช่องทางที่ทำให้ผู้ใช้สามารถเข้าถึงเนื้อหาต่างๆ ได้อย่างง่ายดาย โดยไม่จำเป็นต้องมีการตรวจสอบใดๆ ทำให้เป็นช่องทางที่สะดวกต่อการสร้างตัวตนในโลกออนไลน์ และกลั่นแกล้งผู้อื่นผ่านทางข้อความ รูปภาพ หรือสื่อต่างๆ ที่ทำให้สามารถเห็นได้อย่างสาธารณะ ก่อให้เกิดความเสียหายทางด้านต่างๆ ต่อผู้ที่ถูกโจมตีทั้งทางตรงและทางอ้อม ทางคณะผู้จัดทำจึงมี

ความสนใจเป็นอย่างมากที่จะพัฒนาระบบคัดกรองข้อความต่างๆ จากสื่อสังคมออนไลน์ เช่น ทวิตเตอร์ (Twitter) โดยอาศัยหลักการการเรียนรู้ของเครื่อง (Machine learning) เพื่อลดผลกระทบที่อาจจะเกิดขึ้นจากการละเมิดทางไซเบอร์ เช่น ความเครียด อาการนอนไม่หลับ อ่อนเพลียเรื้อรัง เจ็บป่วย ทำร้ายตัวเอง จนยกระดับเป็นโรคซึมเศร้า ทำร้ายตนเอง และฆ่าตัวตายได้ในที่สุด

## 2. การทบทวนวรรณกรรมที่เกี่ยวข้อง

### 2.1 การละเมิดทางไซเบอร์ (Cyber Bully)

การละเมิดทางไซเบอร์ (Cyber Bully) หมายถึง การกลั่นแกล้ง การให้ร้าย การด่าว่า การข่มเหง หรือการรังแกผู้อื่นทางสื่อสังคมต่าง ๆ เช่น เฟซบุ๊ก ทวิตเตอร์ [1]

ในปัจจุบัน เนื่องจากการเข้าถึงข้อมูลโดยสื่ออิเล็กทรอนิกส์ถือเป็นเรื่องง่ายด้วยวิวัฒนาการของเทคโนโลยีที่มีการพัฒนาอย่างรวดเร็ว ซึ่งทำให้เกิดปัญหาต่างๆ ตามมา เช่น การรับข้อมูลที่อาจส่งผลถึงสภาวะจิตใจ หรือการนำข้อมูลต่างๆ มาใช้เพื่อวัตถุประสงค์ที่ไม่เหมาะสม หรือปัญหาการระรานทางไซเบอร์นั่นเอง [2]

## 2.2 การเรียนรู้ของเครื่อง (Machine Learning: ML)

การเรียนรู้ของเครื่อง คือ การที่ระบบสามารถมีการเรียนรู้ของข้อมูล โดยปราศจากการเขียนชุดคำสั่งของผู้พัฒนา จนนำไปสู่การให้ค่าผลลัพธ์ที่มีความแม่นยำในระดับที่ยอมรับได้ ผ่านการเรียนรู้และวิเคราะห์โดยตัวระบบเอง [3]

### 2.2.1 การแยกประเภท (Classification)

การแยกประเภทเป็นอัลกอริทึมที่มีจุดประสงค์เพื่อบอกถึงความแตกต่างของชุดข้อมูลยกตัวอย่างเช่น การแยกประเภทรูปภาพแมว โดยซึ่งจะทำการตรวจสอบว่ารูปภาพที่มีผู้ใช้งานเข้าสู่ระบบนั้นเป็นภาพของแมวหรือไม่ใช่ภาพของแมว เป็นต้น โดยในการแยกประเภทของชุดข้อมูลนั้น จำเป็นจะต้องมีข้อมูลเพื่อการเรียนรู้ของเครื่อง เช่นในตัวอย่างนี้คือรูปภาพแมวและรูปภาพของสิ่งที่ไม่ใช่แมวเพื่อให้อัลกอริทึมสามารถเรียนรู้ได้ว่า ลักษณะของแมวเป็นอย่างไร และสามารถใช้ลักษณะที่เรียนรู้ดังกล่าวในการทำนายภาพอื่นๆ ในอนาคตได้

โดยอัลกอริทึมหนึ่งที่นิยมนำมาใช้ในการแยกประเภท คือ เครื่องเวกเตอร์ค้ำยัน (Support Vector Machines : SVM)

เครื่องเวกเตอร์ค้ำยัน เป็นอัลกอริทึมที่ใช้จัดกลุ่มข้อมูลโดยการแบ่งประเภทของข้อมูลออกจากกันโดยจะสร้างเส้นแบ่งแยกข้อมูล (Hyperplane) จะมุ่งเน้นไปที่เส้นแบ่งแยกและกลุ่มข้อมูลให้ดีที่สุด [6] มีฟังก์ชันการตัดสินใจสำหรับการหาเส้นแบ่ง [4] ดังสมการที่ 1

$$f(x) = \sum_{j=1}^n w_j x_j + b \quad (1)$$

จะมีเวกเตอร์ค้ำยันของแต่ละด้านของเส้นแบ่ง มีค่าเป็น 1 (ฝั่งที่เป็นบวก) และ -1 (ฝั่งที่เป็นลบ) มีความชันของฟังก์ชันการตัดสินใจดังสมการที่ 2

$$\frac{\partial}{\partial x} h(x) = \sum_{i=1}^m |w_i| \quad (2)$$

## 2.3 เทคโนโลยีที่เกี่ยวข้อง

การแปลงค่าเป็นตัวเลข คือการนำคำที่สนใจมาผ่านกระบวนการที่สามารถแปลงค่าเป็นตัวเลขที่ใช้แทนของคำนั้นๆ

2.3.1. การแปลงค่าเป็นตัวเลขด้วยกระบวนการ Word2Vec เป็นโมเดลที่ถูกสร้างด้วย Thomas Mikolov ในปี 2013 เพื่อใช้ในแสดง ‘คำ’ ให้อยู่ในค่าที่เป็น เวกเตอร์ [5] โดยจะมีหลักการทำงานดังนี้

- 1) เริ่มต้นด้วยคลังข้อมูลขนาดใหญ่ (Corpus)
- 2) โมเดลจะทำการสุ่มค่า เวกเตอร์ ให้กับแต่ละคำโดยจะมีค่าพารามิเตอร์ชื่อ window of size ให้กำหนดว่าเราจะให้ค่า เวกเตอร์ รอบคำนั้นให้ไปหน้า - หลังกี่คำ โดยค่าที่ได้ออกมาจากการสุ่มคือ ค่าความน่าจะเป็นของการเกิดคำนั้น (Probability) [6]
- 3) หลังจากที่ได้สุ่มค่าเสร็จแล้วโมเดลจะทำการเลื่อนไปสุ่มค่าต่อไปตามรูปภาพที่ 2.2 จนหมดคลังข้อมูลโดยการหาค่าโอกาสการเกิดของคำ จะใช้ฟังก์ชันที่เรียกว่า Likelihood function เพื่อสุ่มค่าโอกาสการเกิดของคำที่จะเกิดในคลังข้อมูลทุกคำ ดังสมการที่ 3

$$\text{Likelihood Fun.} = \prod_{t=1}^T \prod_{\substack{-m \leq j \leq m \\ j \neq 0}} P(w_{t+j} | w_t; \theta) \quad (3)$$

- 4) จากนั้นนำค่าที่ได้มาทำให้มีค่าต่ำที่สุดของแต่ละคำซึ่งจะทำให้เราสามารถทำนาย ‘คำ’ ในบริบทอื่นได้ดียิ่งขึ้น [7] ดังสมการที่ 4

$$J(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{-ms/j \leq m} \log P(w_{t+j}|w_t; \theta) \quad (4)$$

5) หลังจากที่กำหนดค่าโอกาสการเกิดของค่าน้อยที่สุดให้กับค่าเสร็จแล้ว โมเดลจะทำการหาที่ดีที่สุดของ ค่า ให้มีค่าเดียวโดยใช้ฟังก์ชันที่เรียกว่าsoftmax [8] ดังสมการที่ 5

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)} = p_i \quad (5)$$

6) หลังจากที่ได้ค่าของฟังก์ชัน softmax ออกมาเราจะสามารถ นำไปใช้หาค่าความใกล้เคียงกันของค่าเพื่อดูค่าที่มีความหมายใกล้เคียงกันได้

**2.3.2. การแปลงค่าเป็นตัวเลขด้วยกระบวนการ Doc2vec** คือ การทำงานต่อจากการแปลงค่าเป็นตัวเลขด้วยกระบวนการ Word2vec นั่นคือการสนใจโอกาสในการเกิดคำที่อยู่ในประโยคจากการสุ่มตัวเลขให้แก่แต่ละคำ และใช้ค่าเฉลี่ยของการเกิดคำที่ได้มาเป็นตัวแทนของประโยค [7]

ซึ่งค่าตัวเลขที่ได้ในแต่ละคำ เกิดจากการหา ค่าเฉลี่ยจากโอกาสที่มากที่สุดของการเกิดคำนั้นๆ ซึ่งอยู่ในรูปแบบดังสมการที่ 6

$$\frac{1}{T} \sum_{t=k}^{T-k} \log p(W_t|W_{t-k}, \dots, W_{t+k}) \quad (6)$$

**2.2.3 การให้น้ำหนักคำที่ปรากฏอยู่ในเอกสาร (Term Frequency-Inverse Document Frequency : TFIDF)** คือ การใช้ความถี่ (Term Frequencies) แต่ละคำที่พบในข้อมูลทั้งหมดในคลัง มาหาอัตราส่วนของจำนวนคำทั้งหมด เช่น จำนวนคำที่ปรากฏต่อจำนวนคำทั้งหมดในคลัง [9]

**2.2.4 ความคล้ายคลึงด้วยองศา (Cosine distance)** เป็นการกำหนดองศา (cosine) ของมุมระหว่าง เวกเตอร์ 2 จุด ยิ่งจุดของเวกเตอร์ใกล้กันมากเท่าไรค่าความใกล้เคียงก็มีค่าสูงมากเท่านั้น ดังสมการที่

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (7)$$

$$\text{เมื่อ } \vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

### 3. การวิเคราะห์และการออกแบบระบบ

#### 3.1 วิเคราะห์ความต้องการ

โดยเริ่มจากการวิเคราะห์ความต้องการเพื่อแบ่งความต้องการของผู้ใช้ มาเป็นความต้องการของระบบ และทำการออกแบบระบบดังนี้

##### 3.1.1. ความต้องการที่เป็นหน้าที่หลักของระบบ (Functional Requirements)

- ผู้ใช้งานสามารถกรอกข้อมูลประเภทข้อความสำหรับการวิเคราะห์ข้อความที่เข้าข่ายการระรานทางไซเบอร์ได้

- ผู้ใช้งานสามารถตรวจสอบรายละเอียดของผลลัพธ์เป็นรูปแบบของประโยคและผลลัพธ์ของแต่ละคำที่มีความเข้าข่ายการระรานทางไซเบอร์ได้

- ผู้ใช้งานสามารถตรวจสอบประวัติต่างๆของข้อความที่มีผู้ใช้ก่อนหน้านี้มาข้อมูลประเภทข้อความสำหรับการวิเคราะห์ที่เข้าข่ายการระรานทางไซเบอร์ได้

- ผู้ใช้งานสามารถให้คะแนนของการใช้งานของระบบในรูปแบบข้อความและการให้คะแนนได้

- ผู้ใช้งานสามารถตรวจสอบประวัติต่างๆของการให้คะแนนของการใช้งานระบบได้

- ผู้ใช้งานสามารถดูรายละเอียดสำหรับข้อเสนอแนะในการใช้งานระบบคัดกรองข้อความที่เข้าข่ายการระรานทางไซเบอร์

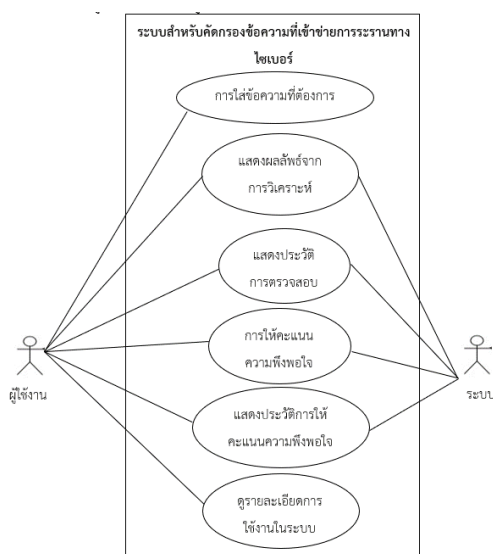
##### 3.1.2. ความต้องการที่ไม่ใช่หน้าที่หลักของระบบ (Non - Functional Requirements)

- ผู้ใช้เกิดความพึงพอใจในการใช้เว็บแอปพลิเคชัน สำหรับคัดกรองข้อความที่เข้าข่ายการระรานทางไซเบอร์

- ส่วนต่อประสานต่างๆทำให้ผู้ใช้งานสามารถใช้งานได้ง่าย สำหรับผู้ที่ไม่มีประสบการณ์หรือไม่ทราบเกี่ยวกับเทคนิคต่างๆในการใช้งาน

- ส่วนของประวัติการค้นหาที่ผู้ใช้งานมีการนำข้อมูลประเภทข้อความมาวิเคราะห์นั้นสามารถนำมาวิเคราะห์และดูแนวโน้มค่าที่มีการใช้งานในช่วงระยะเวลาที่ผ่านมาได้

### 3.2 การออกแบบแผนภาพยูสเคสของระบบ (Use Case Diagram)



รูปภาพที่ 1 แสดงแผนภาพยูสเคสของระบบ

จากรูปที่ 1 แสดงแผนภาพยูสเคสของระบบคัดกรองข้อความที่เข้าช่วยการระรานทางไซเบอร์ ซึ่งมีผู้ใช้งานระบบ 2 กลุ่ม คือ ผู้ใช้งาน และ ระบบ

#### 3.2.1. ยูสเคสระบบสำหรับผู้ใช้งาน

- ยูสเคสการใส่ข้อความที่ต้องการ
- ยูสเคสแสดงผลลัพธ์จากการวิเคราะห์
- ยูสเคสประวัติการตรวจสอบ
- ยูสเคสการให้คะแนนความพึงพอใจ
- ยูสเคสประวัติการให้คะแนนความพึงพอใจ
- ยูสเคสดูรายละเอียดการใช้งานในระบบ

#### 3.2.2. ยูสเคสระบบสำหรับระบบ

- ยูสเคสแสดงผลลัพธ์จากการวิเคราะห์
- ยูสเคสประวัติการตรวจสอบ
- ยูสเคสการให้คะแนนความพึงพอใจ
- ยูสเคสประวัติการให้คะแนนความพึงพอใจ

### 3.3 ฐานข้อมูลของระบบ

ฐานข้อมูลสร้างโดยระบบ Firebase เป็นฐานข้อมูลแบบ NoSQL ประกอบด้วยตาราง 4 ตาราง และตาราง Excel 2 ตารางดังนี้

#### ฐานข้อมูลแบบ NoSQL

1. ตาราง Tweets เก็บข้อมูลที่รวบรวมได้จากแอปพลิเคชันทวิตเตอร์
2. ตาราง Clean Tweets เก็บข้อมูลรายละเอียดข้อความที่นำมาวิเคราะห์จากตาราง Tweets
3. ตาราง Detecting Cyberbullying in Tweets เก็บข้อความที่นำมาพัฒนาการทำงานของเครื่องเวกเตอร์คำย่น
4. ตาราง Labels เก็บข้อความและประเภทข้อความที่ได้จากการทำงานเครื่องเวกเตอร์คำย่น

#### ตารางแบบ excel

1. ตาราง History เก็บข้อมูลการตรวจสอบของผู้ใช้งาน
2. ตาราง Comments เก็บข้อมูลการให้ความคิดเห็นของผู้ใช้

### 4. ระบบต้นแบบ

สามารถแบ่งออกเป็น 2 ส่วน ได้แก่ 1.) ผู้ใช้งานติดต่อกับระบบและ 2.) ส่วนการทำงานของระบบ

#### 4.1 ผู้ใช้งานติดต่อกับระบบ

ส่วนของผู้ใช้งานติดต่อกับระบบ เป็นส่วนที่ผู้ใช้งานกรอกข้อความที่จะตรวจสอบลงบนกล่องข้อความและรับผลที่ได้จากการทำงานของระบบนอกจากนี้ยังสามารถตรวจสอบประวัติการตรวจสอบทั้งหมดที่มีผู้ใช้งานเคยตรวจสอบมา และยังสามารถให้ความคิดเห็นที่มีต่อการใช้งานของระบบได้ทั้งรูปแบบข้อความและรูปแบบการให้คะแนน ซึ่งสามารถตรวจสอบประวัติการให้ความคิดเห็นได้เช่นกัน

#### 4.1.1. หน้าการใส่ข้อความที่ต้องการ

สำหรับการเริ่มต้นใช้งานระบบสำหรับใส่ข้อความที่ต้องการเพื่อตรวจสอบว่าข้อความนั้นเป็นข้อความที่เข้าข่ายการกระรานทางไซเบอร์หรือไม่นั้น ผู้ใช้จะต้องทำการกรอกข้อความลงบนกล่องข้อความสี่เหลี่ยม โดยหากผู้ใช้งานไม่มีการกรอกข้อความในกล่องสี่เหลี่ยม แล้วกดให้ระบบทำงานจะมีการแจ้งเตือนให้ใส่ข้อความก่อน

A light blue rectangular box with a title "Type your sentence:" in red. Below the title is a white text input field. At the bottom center is a red button with the word "Predict" in white.

รูปภาพที่ 4.1 หน้าจอสำหรับการใส่ข้อความที่ต้องการ

#### 4.1.2. หน้าแสดงผลลัพธ์จากการวิเคราะห์

สำหรับการแสดงผลลัพธ์จากการวิเคราะห์นั้น จะแสดงผลข้อความก็ต่อเมื่อผู้ใช้งานกรอกข้อความเข้ามา และทำการกดปุ่ม “Predict” เพื่อให้ระบบทำงาน โดย “Bully” จะใช้แทนข้อความที่เข้าข่ายการกระรานทางไซเบอร์ และ “Not bully” จะใช้แทนข้อความที่ไม่เข้าข่ายการกระรานทางไซเบอร์ ดังรูปภาพที่ 4.2

A light blue rectangular box with a title "Type your sentence:" in red. Below the title is a white text input field. At the bottom center is a red button with the word "Predict" in white. Below the button, the text "bitch : Bully" is displayed in red.

รูปภาพที่ 4.2 หน้าจอสำหรับแสดงผลลัพธ์จากการวิเคราะห์

#### 4.1.3. หน้าประวัติการตรวจสอบ

สำหรับหน้าประวัติการตรวจสอบนั้น จะอยู่ในตำแหน่งด้านขวาของหน้าจอ เป็นการแสดงผลข้อความ

และผลลัพธ์ที่เกิดจากการวิเคราะห์ของระบบต่อข้อความที่กรอกเข้ามาจากผู้ใช้งาน ซึ่งข้อความสีแดงจะเป็นข้อความที่เข้าข่ายการกระรานทางไซเบอร์ และข้อความสีเขียวจะเป็นข้อความที่ไม่เข้าข่ายการกระรานทางไซเบอร์ ดังรูปภาพที่ 4.3 โดยหน้าประวัติการตรวจสอบจะมีการแสดงผลใหม่หลังการทำงานของระบบทุกๆ 30 วินาที

A light blue rectangular box with a title "History" in red. Below the title is a list of text entries. Each entry consists of a text snippet followed by a classification in parentheses. The text snippets are: "222", "1111", "noob1", "bitch 1", "asd", "noob bitch", "very fake", "cdromsd", "123456", "you is a fake friend", "im sadd", and "you are fake". The classifications are: "(not bully)", "(not bully)", "(not bully)", "(bully)", "(not bully)", "(bully)", "(not bully)", "(not bully)", "(not bully)", "(not bully)", "(not bully)", "(bully)", and "(not bully)". The text is color-coded: red for "(bully)" and green for "(not bully)".

รูปภาพที่ 4.3 หน้าจอสำหรับการแสดงประวัติการตรวจสอบ

#### 4.1.4. หน้าการให้คะแนนความพึงพอใจ

สำหรับการให้คะแนนความพึงพอใจของผู้ใช้งาน จะอยู่ด้านล่างของหน้าการใส่ข้อความที่ต้องการ โดยผู้ใช้งานสามารถให้ข้อคิดเห็นต่อการใช้ระบบเป็นข้อความได้ทั้งภาษาไทยและภาษาอังกฤษบนกล่องข้อความภายใต้ข้อความ “Comments” และยังสามารถให้คะแนนความพึงพอใจโดยมีค่าตั้งแต่ 1-5 หลังจากนั้น กดปุ่ม “Submit” เพื่อยืนยันต่อข้อความคิดเห็นและคะแนนความพึงพอใจ หลังจากนั้นระบบจะมีกล่องข้อความแจ้งเตือนขึ้นว่าได้รับข้อความคิดเห็นและคะแนนความพึงพอใจมาแล้ว ผู้ใช้งานจะต้องกดปุ่ม “OK” เพื่อทำการจบการทำงานให้คะแนนความพึงพอใจ ดังรูปภาพที่ 4.4

A light blue rectangular box with a title "Comments" in red. Below the title is a white text input field. To the right of the input field is a vertical list of five radio buttons, each followed by a score: "1 score", "2 score", "3 score", "4 score", and "5 score". At the bottom center is a red button with the word "Submit" in white.

รูปภาพที่ 4.4 หน้าจอสำหรับการให้คะแนนความพึงพอใจ

#### 4.1.5. หน้าแสดงประวัติการให้คะแนนความพึงพอใจ

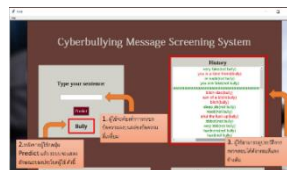
สำหรับการแสดงผลของความคิดเห็นที่ผู้ใช้งานมีต่อระบบนั้น จะอยู่ทางด้านขวาของหน้าการให้คะแนนความพึงพอใจ โดยแสดงข้อความที่ต่อเมื่อมีการกดยืนยันต่อข้อความความคิดเห็นของผู้ใช้งาน และจะทำการแสดงค่าโดยทันทีหลังจากผู้ใช้งานได้ดำเนินการให้คะแนนความคิดเห็นเสร็จสิ้น ซึ่งข้อความความคิดเห็นที่จะแสดงนั้นจะเป็นข้อความความคิดเห็นจากผู้ใช้งานล่าสุด 5 ข้อความ ดังรูปภาพที่ 4.5



รูปภาพที่ 4.5 หน้าจอสำหรับแสดงประวัติการให้คะแนนความพึงพอใจ

#### 4.1.6 หน้าแสดงรายละเอียดการใช้งานในระบบ

สำหรับการแสดงผลของรายละเอียดการใช้งานในระบบนั้น จะอยู่ในส่วนของด้านบนของระบบหากผู้ใช้งานได้ทำการคลิก ปุ่ม 'help' ระบบจะแสดงหน้าต่างแสดงการใช้งานต่างๆที่มีภายในระบบ รวมไปถึงรายละเอียดแสดงหน้าจอต่างๆที่ระบบมีทั้งหมดให้ผู้ใช้งานได้ทำความเข้าใจในการใช้งาน ซึ่งประกอบไปด้วย หน้ารายละเอียดการใช้งาน, หน้าการใส่ข้อความที่ต้องการ, หน้าแสดงผลฟังก์ชันจากกราฟวิเคราะห์, หน้าประวัติการตรวจสอบ, หน้าการให้คะแนนความพึงพอใจ และ หน้าแสดงประวัติการให้คะแนนความพึงพอใจ ดังรูปภาพที่ 4.6



รูปภาพที่ 4.6 หน้าจอแสดงรายละเอียดการใช้งานในระบบ

#### 4.2 ส่วนของการทำงานภายในระบบ



รูปภาพที่ 4.7 แสดงรายละเอียดขั้นตอนการทำงานภายในระบบ

ส่วนของการทำงานภายในระบบ เป็นส่วนที่ระบบใช้ประมวลผลต่างๆ ก่อนที่จะนำผลลัพธ์หรือ output ออกมาให้กับผู้ใช้งานได้รับทราบ โดยการทำงานของระบบมีส่วนการทำงานต่างๆดังนี้

##### 4.2.1. รับข้อความจากผู้ใช้งาน

ระบบสามารถรับข้อความจากผู้ใช้งานซึ่งเป็นข้อความที่อยู่ในรูปแบบของภาษาอังกฤษเท่านั้น

##### 4.2.2 ทำความสะอาดข้อความ

หลังจากที่ระบบได้รับข้อความจากผู้ใช้งาน ข้อความที่ได้จะต้องผ่านกระบวนการทำให้ทุกตัวอักษรเป็นตัวพิมพ์เล็ก ต่อมาทำการลบกริยาต่างๆ (verbs) ที่พบในข้อความออกทั้งหมด หรือเรียกว่า Bag of words

##### 4.2.3. การแปลงคำเป็นตัวเลข

หลังทำความสะอาดข้อความ ระบบจะนำข้อความที่ได้มาแปลงคำเป็นตัวเลข โดยจะใช้กระบวนการให้น้ำหนักคำที่ปรากฏอยู่ในเอกสารมาแปลงคำเป็นตัวเลข และใช้ความคล้ายคลึงด้วยองศา หาค่าที่ใกล้เคียงกันมากที่สุด เพื่อให้คำนั้นอยู่ภายใต้ค่าตั้งแต่ศูนย์ถึงหนึ่ง

#### 4.2.4 การแบ่งประเภทของข้อความ

หลังจากที่ได้ตัวเลขที่มีค่าตั้งแต่ศูนย์ถึงหนึ่งจากการทำงานของทั้งกระบวนการให้น้ำหนักค่าที่ปรากฏอยู่ในเอกสารและความคล้ายคลึงด้วยองศา หาค่าที่ใกล้เคียงกันมากที่สุดแล้ว นำตัวเลขที่ได้มาเข้าสู่การแบ่งประเภทโดยใช้อัลกอริทึมหรือโมเดลจากการเรียนรู้ของเครื่องเวกเตอร์คำย่น โดยสามารถนำมาวิเคราะห์ได้ว่าข้อความที่ระบบปรับมานั้นเป็นข้อความที่เข้าข่ายการระรานทางไซเบอร์หรือไม่

#### 4.2.5 แสดงผลการแบ่งประเภทของข้อความ

การแสดงผลของระบบ จะแสดงผลเป็นข้อความ “Bully” จะใช้แทนข้อความที่เข้าข่ายการระรานทางไซเบอร์และ “Non Bully” จะใช้แทนข้อความที่ไม่เข้าข่ายการระรานทางไซเบอร์

### 5. สรุปและประเมินงานวิจัย

ในโครงงานนี้ ผู้พัฒนาได้ทำการพัฒนาระบบคัดกรองข้อความที่เข้าข่ายการระรานทางไซเบอร์ในหลากหลายรูปแบบ เริ่มต้นจากการพัฒนาโมเดลพื้นฐานในรูปแบบของเครื่องเวกเตอร์คำย่นควบคู่กับ TFI-DF เวกเตอร์ ซึ่งพัฒนามาบนชุดข้อมูลเพื่อการเรียนรู้จากฐานข้อมูล Detecting Cyberbullying in Tweets [10] แล้วทดสอบประสิทธิภาพบนชุดข้อมูลเพื่อการทดสอบจากฐานข้อมูลเดิมพบว่า โมเดลพื้นฐานดังกล่าวมีความแม่นยำอยู่ที่ร้อยละ 90.64

#### 5.1 สรุปภาพรวมการดำเนินงาน

ต่อมาผู้พัฒนาได้ทำการทดลองต่าง ๆ มากมาย เพื่อเพิ่มประสิทธิภาพของโมเดลพื้นฐานนี้โดยอ้างอิงจากสมมุติฐานที่ว่า ถ้าชุดข้อมูลการเรียนรู้มีความหลากหลายและขนาดเพิ่มมากขึ้น ประสิทธิภาพของโมเดลจะเพิ่มขึ้นเช่นเดียวกัน ดังนั้นผู้พัฒนาจึงทำการทดลองดังต่อไปนี้

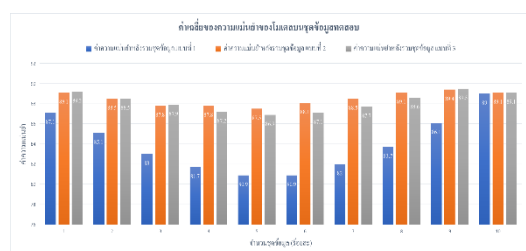
1) สร้างฐานข้อมูล Data from Twitter ซึ่งเป็นข้อมูล ที่ทดสอบแล้วว่ามีความแตกต่างกันกับฐานข้อมูล Detecting Cyberbullying in Tweets [10]

2) การเพิ่มข้อมูลภายในชุดข้อมูลเพื่อการเรียนรู้ อย่างไรก็ตามข้อมูลภายในฐานข้อมูล Data from Twitter ทวิตเตอร์ไม่สามารถนำมาใช้งานได้ทันที เพราะเป็นข้อมูลที่ยังไม่มีการจำแนกประเภทข้อความว่าเป็นข้อความที่เข้าข่ายการระรานทางไซเบอร์หรือไม่เป็นข้อความที่เข้าข่ายการระรานทางไซเบอร์ จึงมีการทดลองเพื่อเพิ่มข้อมูลวิธีการดังต่อไปนี้

- การเพิ่มชุดข้อมูลจากผลลัพธ์ที่ได้จากการทำนายของเครื่องเวกเตอร์คำย่น โดยมีค่าความแม่นยำอยู่ที่ร้อยละ 94.57 เพิ่มขึ้นถึงร้อยละ 3.93

- การเพิ่มชุดข้อมูลจากผลลัพธ์ที่ได้จากค่าความคล้ายกันของข้อความจากค่าความเหมือนโคไซน์ ซึ่งผลของการเพิ่มข้อมูลสามารถเปรียบเทียบความแม่นยำของทั้ง 3 วิธีการประกอบไปด้วย 1) การเพิ่มชุดข้อมูลที่ละส่วน โดยข้อมูลนำเข้ามาจากค่าเฉลี่ยของค่าความคล้ายกันของข้อมูล (กราฟสีฟ้า), 2) การเพิ่มชุดข้อมูลที่ละส่วนโดยข้อมูลนำเข้ามาจากค่าเฉลี่ยของความคล้ายกันของข้อมูลมากที่สุด 100 ประโยคแรก (กราฟสีส้ม) และ 3) การเพิ่มชุดข้อมูลที่ละส่วนโดยข้อมูลนำเข้ามาจากค่าเฉลี่ยของการถ่วงน้ำหนักของค่าความคล้ายกันของข้อมูล (กราฟสีเทา) ดังนี้

**แผนภูมิที่ 5.1 แผนภูมิแสดงค่าเฉลี่ยความแม่นยำ ของการพยายามเพิ่ม ชุดข้อมูลที่ได้จาก**



**cosine similarity ทั้ง 3 วิธีการ**

แผนภูมิที่ 5.1 แสดงค่าความแม่นยำในการเพิ่มชุดข้อมูลที่ได้จากค่าความคล้ายกันของข้อมูล ซึ่งสามารถสรุปได้ว่าการเพิ่มชุดข้อมูลที่ละส่วนโดยข้อมูลนำเข้ามาจากค่าเฉลี่ยของค่าความคล้ายกันของข้อมูลมากที่สุด

ที่สุด 100 ประโยคแรกนั้นมีความแม่นยำที่ดีที่สุดจากทั้ง 3 วิธีการ ซึ่งมีค่าความแม่นยำสูงอยู่ที่ร้อยละ 89.15 33 เป็นเพราะว่าการพยายามเพิ่มชุดข้อมูลในชุดข้อมูลการเรียนรู้จากค่าเฉลี่ยของค่าความคล้ายกันของข้อมูลมากที่สุด 100 ประโยคแรกนั้นจะมีความถูกต้องของข้อความที่ใกล้เคียงกับประโยคที่เข้าข่ายการระรานทางไซเบอร์และมีความคล้ายกันกับประโยคที่ไม่เข้าข่ายการระรานทางไซเบอร์สูง และมีประโยคที่ไม่ส่งผลหรือเป็นประโยคที่มีความผิดพลาดน้อยทำให้มีค่าความแม่นยำมากกว่าการพยายามเพิ่มชุดข้อมูลแบบอื่นๆ อย่างไรก็ตามการเพิ่มข้อมูลโดยอาศัยค่าความคล้ายกันของข้อมูลทั้ง 3 วิธีนั้น ได้ประสิทธิภาพน้อยกว่าการเพิ่มข้อมูลโดยอาศัยการทำนายของโมเดลพื้นฐาน สามารถสรุปได้ว่าการเพิ่มข้อมูลจากการทำนายของโมเดลพื้นฐานเป็นวิธีการที่ดีที่สุด ซึ่งมีค่าความแม่นยำอยู่ที่ร้อยละ 94.57

## 5.2 ปัญหาและอุปสรรคที่พบ

แม้ว่าผู้พัฒนาจะสามารถเก็บฐานข้อมูล Data from Twitter ซึ่งเป็นฐานข้อมูลขนาดใหญ่ได้ก็ตาม การจำแนกประเภทข้อความในฐานข้อมูลดังกล่าวเป็นกระบวนการที่ใช้ทรัพยากร (งบประมาณและเวลา) เป็นอย่างมาก ทำให้ผู้พัฒนาไม่สามารถใช้ประโยชน์จากข้อมูลขนาดใหญ่ได้อย่างเต็มประสิทธิภาพ

## 5.3 แนวทางการพัฒนา

พัฒนาเครื่องมือที่อำนวยความสะดวกในการจำแนกประเภทข้อมูลในฐานข้อมูลขนาดใหญ่ เพื่อให้กระบวนการดังกล่าวสามารถทำได้ง่ายขึ้นหรือใช้ทรัพยากรน้อยลง ซึ่งให้สามารถนำมาประยุกต์ใช้กับฐานข้อมูล Data from Twitter ได้ เพื่อให้ฐานข้อมูลดังกล่าวมีความสมบูรณ์และเป็นประโยชน์มากยิ่งขึ้น

## เอกสารอ้างอิง

[1] MThai.com (2562). รวมศัพท์ใหม่คอมพิวเตอร์และเทคโนโลยีสารสนเทศฉบับราชบัณฑิตยสภา.

สืบค้นเมื่อ 26 กุมภาพันธ์ 2563. เข้าถึงได้จาก : <https://bit.ly/2KY oSpC>

- [2] กรมสุขภาพจิต (2561). 'กรมสุขภาพจิต' เผยกลั่นแกล้ง-รังแก (BULLYING) ในโรงเรียน ไม่ใช่เรื่องล้อเล่น. สืบค้นเมื่อ 26 กุมภาพันธ์ 2563. เข้าถึงได้จาก : <https://www.bangkokbiznews.com/news/detail/859102>
- [3] Nessessence (2561). อะไรคือ การเรียนรู้ของเครื่อง (Machine Learning)? (ฉบับมือใหม่). สืบค้นเมื่อ 26 กุมภาพันธ์ 2563. เข้าถึงได้จาก : <https://bit.ly/3bXev1u>
- [4] phyblas(2561). วิเคราะห์จำแนกประเภทข้อมูลด้วยเครื่องเวกเตอร์ค้ำยัน (SVM). สืบค้นเมื่อ 3 ตุลาคม 2563 เข้าถึงได้จาก :<https://p hyblas.hinaboshi.com/20180709>
- [5] lukkidd, Surasak Lhuengsakul, Kitti Mortin(2561). Word Embedding และ Word2Vec คืออะไร. สืบค้นเมื่อ 26 เมษายน 2563. เข้าถึงได้จาก : <https://bit.ly/2z5 yk7X>
- [6] Mikolov, Tomas. Statistical Language Models based on Neural Networks. PhD thesis, Brno University of Technology, 2012.
- [7] Bengio, Yoshua, Schwenk, Holger, Sen´ecal, Jean-S´ebastien, Morin, Fr´ederic, and Gauvain, Jean-Luc. Neural probabilistic language models. In Innovations in Machine Learning, pp. 137–186. Springer, 2006.
- [8] Mikolov, Tomas, Le, Quoc V., and Sutskever, Ilya. Exploiting similarities among languages for machine translation. CoRR, abs/1309.4168, 2013b.



[9] CHAKRIT(2562). TF-IDF ทำงานอย่างไร?. สืบค้น

เมื่อ 3 ตุลาคม 2563 เข้าถึงได้จาก :

<https://bit.ly/3jqIfbX>

[10] dhavalpotdar(2562). Detecting Cyberbullying

in Tweets. สืบค้นเมื่อ 3 ตุลาคม 2563

เข้าถึงได้จาก : <https://bit.ly/3d3TBpt>