รหัสโครงการ SUT11-1199-65-12-04(NEW)

รายงานการวิจัย

การระบุปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้า

(Identifying Depression Markers in Social Media Content)

ได้รับทุนอุดหนุนการวิจัยจาก

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

รหัสโครงการ SUT11-1199-65-12-04(NEW)

รายงานการวิจัย

การระบุปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้า

(Identifying Depression Markers in Social Media Content)

คณะผู้วิจัย

หัวหน้าโครงการ

อาจารย์ ดร.อรรคพล วงศ์กอบลาภ

สำนักวิชาศาสตร์และศิลป์ดิจิทัล

มหาวิทยาลัยเทคโนโลยีสุรนารี

ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ปีงบประมาณ 2565

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

มิถุนายน 2567

กิตติกรรมประกาศ

การดำเนินโครงการวิจัยเรื่อง “การระบุปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้า” ได้รับเงินอุดหนุนการทำวิจัย ประเภททุนสนับสนุนการสร้างและพัฒนานักวิจัยรุ่นใหม่ ประจำปีงบประมาณ พ.ศ. 2565 จากมหาวิทยาลัยเทคโนโลยีสุรนารี

ผู้วิจัยขอแสดงความขอบพระคุณทุกท่านเป็นอย่างสูงมา ณ.โอกาสนี้

อ.ดร.อรรคพล วงศ์กอบลาภ

มิถุนายน 2567

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทคัดย่อภาษาไทย

ปัญหาด้านโรคที่เกี่ยวกับสุขภาพจิตถือเป็นปัญหาที่สำคัญและควรได้รับการแก้ไขอย่างเร่งด่วน จากสถิติขององค์การอนามัยโลก (World Health Organization) จำนวนของผู้ป่วยที่กำลังเผชิญกับโรคซึมเศร้าคิดเป็นจำนวนประมาณ 5% ของผู้ใหญ่ หรือมีมากกว่า 280 ล้านคนในปี 2566 และการเติบโตที่รวดเร็วของเทคโนโลยีทำให้ผู้คนสามารถเข้าถึงอินเทอร์เน็ตได้ง่ายดายและรวดเร็วขึ้น จึงส่งผลต่อจำนวนผู้ในงานอินเทอร์เน็ตทั่วโลกมีมากกว่า 5.18 พันล้านคน และผู้ใช้งานสื่อสังคมออนไลน์ประมาณ 4.8 พันล้านผู้ใช้งาน ในปี 2566 งานวิจัยชิ้นนี้มีวัตถุประสงค์ในการสกัดปัจจัยที่สามารถระบุผู้ใช้งานบนสื่อสังคมออนไลน์จะป่วยเป็นโรคซึมเศร้าได้ พัฒนาโมเดลการทำนายโอกาสการเป็นโรคซึมเศร้าจากข้อมูลผู้ใช้งานบนสื่อสังคมออนไลน์ และประเมินประสิทธิภาพของโมเดลการทำนายผลจากข้อมูลบนสื่อสังคมออนไลน์ ซึ่งจากผลการทดลองของงานวิจัยชิ้นนี้พบว่าข้อมูลจากสื่อสังคมออนไลน์สามารถใช้ในการหาปัจจัยความแตกต่างระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานทั่วไปได้ โดยปัจจัยที่สามารถใช้ในการแยกความแตกต่างระหว่างผู้ใช้งานทั้งสองกลุ่มคือการใช้คำในข้อความสื่อสังคมออนไลน์ และเวลาที่ใช้งานสื่อสังคมออนไลน์

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทคัดย่อภาษาอังกฤษ

Mental health problems are a serious problem and should be urgently addressed. According to statistics from the World Health Organization (WHO), the number of people suffering from depression is approximately 5% of adults, or more than 280 million people in 2023. The rapid growth of advanced technology has made people access the internet more easily and quickly. As a result, the number of internet users globally is over 5.18 billion and approximately there are 4.8 billion social media users in 2023. The purposes of this study are to extract factors to identify social media users with depression, develop models to predict the chance of depression from users' data on social media, and evaluate the performance of the prediction models. Our study found that user-generated data from social media can be used to find differentiating factors between users suffering from depression and general users. Possible factors used to differentiate between the two groups of users are the use of words in social media messages and time spent on the social media platform.

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ

กิตติกรรมประกาศ ..... *ก*

บทคัดย่อภาษาไทย ..... *ข*

บทคัดย่อภาษาอังกฤษ ..... *ค*

สารบัญ ..... *ง*

สารบัญตาราง ..... *ฉ*

สารบัญภาพ ..... *ช*

บทที่ 1 บทนำ *1*

1. ความเป็นมาและความสำคัญของปัญหา 1
2. วัตถุประสงค์ 2
3. ขอบเขตของการวิจัย 2
4. ประโยชน์ที่คาดว่าจะได้รับ 3

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง *4*

1. การประมวลผลภาษาธรรมชาติ 4
2. การวิเคราะห์ข้อมูลสื่อสังคมออนไลน์ 5

บทที่ 3 วิธีดำเนินการวิจัย *7*

1. การเก็บรวบรวมข้อมูล (Data Collection) 8
2. การระบุประเภทของข้อมูล (Data Annotation) 9
3. การเตรียมข้อมูล (Data Preprocessing) 10
4. การสกัดคุณลักษณะเฉพาะ (Features Extraction) 10
5. การพัฒนาโมเดล (Model Construction) 11
6. การประเมินประสิทธิภาพของโมเดล (Model Evaluation) 11

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4 ผลการวิจัย *12*

1. รายละเอียดข้อมูล 12
2. การประเมินประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่อง 14

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ *21*

1. อภิปรายผล 21
2. ข้อเสนอแนะ 25
3. สรุปผลการวิจัย 26

บรรณานุกรม *28*

ประวัตินักวิจัย *32*

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ 1 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล SVM และสกัดข้อมูลด้วยวิธี Count Vectorizer 15
ตารางที่ 2 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล Logistic Regression และสกัดข้อมูลด้วยวิธี Count Vectorizer 16
ตารางที่ 3 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี Count Vectorizer 16
ตารางที่ 4 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล SVM และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer 17
ตารางที่ 5 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล Logistic Regression และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer 17
ตารางที่ 6 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer 18

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญภาพ

รูปที่ 1 กระบวนการวิเคราะห์หาปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้า 7
รูปที่ 2 หน้าจอแสดงการเข้าถึงข้อมูลบน Twitter 8
รูปที่ 3 คำสั่งการเก็บข้อมูลข้อความจากโปรไฟล์ผู้ใช้งานด้วย Python 9
รูปที่ 4 สัดส่วนประเภทข้อเกี่ยวกับโรคซึมเศร้าและไม่เกี่ยวกับโรคซึมเศร้า 13
รูปที่ 5 จำนวนผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า และจำนวนข้อความของผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า สีเทาอ่อนหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและสีเทาเข้มคือผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า 14
รูปที่ 6 ผลการประเมินโมเดลการทำนายผลด้วย Receiver operating characteristic curves จากโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี Count Vectorizer 19
รูปที่ 7 ผลการประเมินโมเดลการทำนายผลด้วย Receiver operating characteristic curves จากโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer 20
รูปที่ 8 การใช้คำของผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า 22
รูปที่ 9 การใช้คำของผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า 22
รูปที่ 10 ช่วงเวลาในการโพสข้อความบนโปรไฟล์ผู้ใช้งานระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า สีน้ำเงินคือผู้ใช้งานทั่วไป สีส้มหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า 23
รูปที่ 11 วันที่ในการโพสข้อความบนโปรไฟล์ผู้ใช้งานระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า สีน้ำเงินคือผู้ใช้งานทั่วไป สีส้มหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า 24
รูปที่ 12 กระทู้แสดงการพูดคุยเกี่ยวกับการเข้าถึงข้อมูลจาก Twitter 25
รูปที่ 13 รายงานการเปลี่ยนแปลงนโยบาย 26

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 1
บทนำ

งานวิจัยชิ้นนี้ผู้วิจัยสนใจในการวิเคราะห์หาปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้าจากข้อมูลสื่อสังคมออนไลน์ โดยในบทนี้จะอธิบายถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ของงานวิจัยชิ้นนี้ ขอบเขตของการวิจัย รวมถึงประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยชิ้นนี้

1. ความเป็นมาและความสำคัญของปัญหา

ปัญหาด้านโรคที่เกี่ยวกับสุขภาพจิตถือเป็นปัญหาที่สำคัญและเร่งด่วนที่ควรได้รับการแก้ไขอย่างมาก จากสถิติขององค์การอนามัยโลก (World Health Organization) จำนวนของผู้ป่วยที่กำลังเผชิญกับโรคซึมเศร้าคิดเป็นจำนวนประมาณ 5% ของผู้ใหญ่ หรือมีมากกว่า 280 ล้านคนในปี 2566 [1] ผลกระทบของโรคที่เกี่ยวกับสุขภาพจิตไม่เพียงมีผลต่อผู้ป่วยเองเท่านั้น ยังส่งผลต่อคนรอบข้างด้วยอีกด้วย เช่น ครอบครัวและเพื่อน เป็นต้น อีกทั้งยังเป็นปัญหาต่อเศรษฐกิจอีกด้วย มีการคาดการว่ามีการใช้เงินจำนวนมากเพื่อใช้ในการให้บริการและรักษาปัญหาผู้ป่วยด้านสุขภาพจิต รวมถึงผลกระทบที่ผู้ป่วยทางสุขภาพจิตไม่สามารถทำงานได้บรรลุเป้าหมาย รวมเป็นเงินจำนวนมากกว่า 105 พันล้านปอนด์ในอังกฤษ ในปี 2561 [2] และค่าใช้จ่ายเกี่ยวกับการดูแลรักษาปัญหาผู้ป่วยด้านสุขภาพจิตทั่วโลกจะเพิ่มขึ้นเป็น 2.5 ล้านล้าน ดอลลาร์สหรัฐในปี 2573 [3] ปัญหาหลักที่ทำให้ค่าใช้จ่ายมีต้นทุนที่สูงเพราะการรักษาที่ไม่ถูกต้อง [4] ยิ่งไปกว่านั้นการป่วยเป็นโรคทางสุขภาพจิตยังอาจนำไปสู่การฆ่าตัวตายได้อีกด้วย ซึ่งในแต่ละปีจะมีคนจำนวนมากกว่า 7 แสนคนทั่วโลกที่ฆ่าตัวตายสำเร็จ [1], [5]

การเติบโตที่รวดเร็วของเทคโนโลยีทำให้ผู้คนสามารถเข้าถึงอินเทอร์เน็ตได้ง่ายดายและรวดเร็วขึ้น ซึ่งผู้ในงานอินเทอร์เน็ตทั่วโลกมีมากกว่า 5.18 พันล้านคน ในปี 2566 [6] การเพิ่มขึ้นของผู้ใช้งานอินเทอร์เน็ตส่งผลในเกิดเทคโนโลยีใหม่ ๆ ขึ้นมาอย่างมากมาย โดยเฉพาะสื่อสังคมออนไลน์ (Social media) ปัจจุบันสื่อสังคมออนไลน์เข้ามามีบทบาทในชีวิตประจำวันของผู้คนอย่างมาก โดยผู้ใช้งานสามารถที่จะแสดงความคิด ความรู้และอารมณ์ แบ่งปันข้อมูลข่าวสาร รวมถึงสร้างปฏิสัมพันธ์กับผู้ใช้งานคนอื่น ๆ บนสื่อสังคมออนไลน์ได้ ซึ่ง

ผู้ใช้งานสื่อสังคมออนไลน์ ในปี 2566 โดยเฉพาะ Facebook ซึ่งเป็นแพลต์ฟอร์มที่มีผู้ใช้งานมากที่สุดในโลก มีผู้ใช้งานมากกว่า 2.4 พันล้านคน [7] และ Twitter มีผู้ใช้งานมากกว่า 386 ล้านคน [8]

ข้อมูลที่ผู้ใช้จำนวนมากสร้างขึ้นบนสื่อสังคมออนไลน์ ทำให้นักวิจัยจำนวนมากให้ความสนใจที่จะใช้ปัญญาประดิษฐ์ วิทยาศาสตร์ข้อมูลและเทคโนโลยีดิจิทัล เพื่อศึกษาหาความรู้ใหม่จากข้อมูลเหล่านี้ ข้อมูลที่ผู้ใช้ทำการแสดงอารมณ์ ความรู้สึก และกิจกรรมต่าง ๆ บนสื่อสังคมออนไลน์ ได้ถูกนำมาใช้ศึกษา ในหลายๆ สาขาวิชา ทั้งการตลาด การเมือง การเงิน รวมถึงด้านสุขภาพ [9]

ด้วยเหตุผลนี้การนำข้อมูลที่ผู้ใช้งานแสดงไว้บนสื่อสังคมออนไลน์ มาศึกษาถึงความเป็นไปได้ของปัญหาของโรคซึมเศร้าที่กำลังเกิดขึ้นแบบทันท่วงที (Realtime) โดยใช้เครื่องมือทางด้านปัญญาประดิษฐ์ วิทยาศาสตร์ข้อมูลและเทคโนโลยีดิจิทัลซึ่งการใช้เครื่องมือเหล่านี้ทำเราสามารถระบุผู้ใช้งานบนสื่อสังคมออนไลน์ที่มีโอกาสป่วยเป็นโรคซึมเศร้าได้รวดเร็วขึ้นและช่วยให้เค้าได้รับการรักษาได้ทันท่วงที

งานวิจัยชิ้นนี้จึงตั้งเป้าหมายเพื่อพัฒนาโมเดลในการเก็บข้อมูลจากสื่อสังคมออนไลน์อัตโนมัติโดยใช้เครื่องมือปัญญาประดิษฐ์และการเรียนรู้ด้วยเครื่อง จากนั้นจึงสร้างโมเดลการเรียนรู้ด้วยเครื่องเพื่อแบ่งประเภทผู้ใช้งาน

2. วัตถุประสงค์

- สกัดปัจจัยที่สามารถระบุโอกาสที่ผู้ใช้งานบนสื่อสังคมออนไลน์จะป่วยเป็นโรคซึมเศร้าได้
- พัฒนาโมเดลการทำนายโอกาสการเป็นโรคซึมเศร้าจากข้อมูลผู้ใช้งานบนสื่อสังคมออนไลน์
- ประเมินประสิทธิภาพของโมเดลการทำนายผลจากข้อมูลบนสื่อสังคมออนไลน์

3. ขอบเขตของการวิจัย

- ระบุคำค้นเพื่อใช้ในการค้นหากลุ่มเป้าหมายที่มีโอกาสป่วยเป็นโรคซึมเศร้า
- พัฒนาเครื่องมือการเก็บข้อมูลจากสื่อสังคมออนไลน์
- เก็บข้อมูลจากสื่อสังคมออนไลน์ทั้งกลุ่มเป้าหมายและกลุ่มผู้ใช้งานทั่วไป

มหาวิทยาลัยเทคโนโลยีสุรนารี

- สกัดปัจจัยที่สามารถระบุโอกาสที่ผู้ใช้งานบนสื่อสังคมออนไลน์จะป่วยเป็นโรคซึมเศร้าได้
- พัฒนาโมเดลการเรียนรู้ด้วยเครื่องเพื่อใช้ในการทำนายผู้มีโอกาสป่วยเป็นโรคซึมเศร้า

4. ประโยชน์ที่คาดว่าจะได้รับ

ส่วนหนึ่งงานของวิจัยชิ้นนี้ได้ถูกนำไปเผยแพร่เพื่อใช้เป็นองค์ความรู้ในการวิจัยด้านเทคโนโลยีและคอมพิวเตอร์ ในรูปบทความและการนำเสนอผลงานในการประชุมวิชาการ the 21st International Conference on Informatics, Management, and Technology in Healthcare (ICIMTH 2023)

อีกทั้งผลลัพธ์จากงานวิจัยชิ้นนี้สามารถนำไปใช้ในการพัฒนาโมเดลในการทำนายโรคซึมเศร้าจากข้อมูลสื่อสังคมออนไลน์ และสามารถนำไปต่อยอดในการพัฒนาแอพพลิเคชันเพื่อใช้งานกับผู้ใช้จริง ซึ่งจะเป็นนวัตกรรมต่อยอด

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะนำเสนอถึงเทคนิคในปัจจุบันที่ใช้ในการสร้างโมเดลสำหรับการทำนายผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้ากสื่อสังคมออนไลน์ ซึ่งประกอบไปด้วยหัวข้อการประมวลผลภาษาธรรมชาติ และการวิเคราะห์ข้อมูลสื่อสังคมออนไลน์

1. การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ (Natural language processing) เป็นการวิเคราะห์ภาษาที่มนุษย์ใช้ในการติดต่อสื่อสาร ด้วยการใช้เครื่องมือทางเทคโนโลยีและคอมพิวเตอร์เข้ามาช่วยในการทำความเข้าใจภาษาต่างๆ โดยอาจจะอยู่ในรูปแบบของโครงสร้างประโยคที่แน่นอนหรือไม่แน่นอนก็ได้ ไม่ว่าจะเป็นการสื่อสารด้วยเสียงสนทนา รวมถึงการสื่อสารด้วยตัวอักษร การวิเคราะห์ข้อมูลการสื่อสารด้วยตัวอักษรนั้นจะช่วยให้คอมพิวเตอร์สามารถเข้าใจภาษาที่มนุษย์ใช้ในการสื่อสารได้ง่ายและรวดเร็วขึ้น [10] ในช่วงแรกของการวิเคราะห์ข้อมูลการสื่อสารด้วยข้อความนั้น นักวิเคราะห์ต้องทำการอธิบายประกอบ (Annotation) เพื่อช่วยให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ได้มากขึ้น ซึ่งการวิเคราะห์นี้ทำให้เราสามารถมีคลังคำหรือคลังข้อมูล (Corpus) เพื่อใช้ในการวิเคราะห์ข้อมูลการสื่อสารด้วยข้อความอื่นต่อไปได้ รวมถึงการวิเคราะห์ความรู้สึก (Sentiment analysis) อย่างไรก็ตามอย่างการวิเคราะห์ข้อมูลการสื่อสารด้วยข้อความในช่วยแรกนั้นต้องใช้ความพยายามและเวลาของนักวิเคราะห์เป็นอย่างมาก [11]

การศึกษาภาษาธรรมชาติจะใช้เครื่องมือการแบ่งข้อความ (Text segmentation) เพื่อศึกษากระบวนการแบ่งข้อความให้อยู่ในรูปแบบที่เป็นประโยชน์และสามารถสื่อความหมายได้ เช่น การแบ่งด้วยคำ (Words) และประโยค (Sentences) เป็นต้น [10] เครื่องมือการแบ่งข้อความและศึกษาประโยคมีอยู่มากมายในหลายๆ ภาษา รวมถึงภาษาไทยด้วย เช่น PyThaiNLP เป็นต้น [12] เครื่องมือเหล่านี้เป็นเครื่องมือพื้นฐานเพื่อจะช่วยให้นักวิจัยสามารถนำไปต่อยอด เพื่อศึกษาถึงการวิเคราะห์ความรู้สึก หรือการวิเคราะห์หัวข้อของข้อความต่อไป

การศึกษาหัวข้อ (Topic analysis) ของข้อความเราสามารถนำเครื่องมือที่เรียกว่า Latent Dirichlet Allocation (LDA) มาใช้เพื่อวิเคราะห์คำในประโยคแล้วนำมาสรุปถึงหัวข้อหรือกลุ่มคำที่ประโยคเหล่านั้นกำลังสื่อถึงได้ [13] ประโยชน์ของการวิเคราะห์หัวข้อของข้อความทำให้เราสามารถสรุปเนื้อหาหรือเข้าใช้ข้อความต่างๆ ได้สั้นและกระชับขึ้น

2. การวิเคราะห์ข้อมูลสื่อสังคมออนไลน์

เมื่อนักวิจัยเห็นว่าการวิเคราะห์ข้อมูลการสื่อสารด้วยข้อความนั้นมีประโยชน์และประสบความสำเร็จในหลายๆ ด้าน นักวิจัยกลุ่มหนึ่งจึงได้ทดลองนำการวิเคราะห์ข้อมูลการสื่อสารด้วยข้อความมาศึกษาข้อมูลที่เก็บรวบรวมมาจากสื่อสังคมออนไลน์ Coppersmith และคณะได้เสนอวิธีการเก็บรวบรวมข้อมูลจากสื่อสังคมออนไลน์ที่สะดวกและรวดเร็ว โดยใช้วิธีการค้นหาข้อความที่ต้องการบนสื่อสังคมออนไลน์ และนำข้อความที่ค้นหาได้มาทำการจำแนกแล้วนำไปใช้ในการศึกษาต่อไป [14]

อรรคพลและคณะได้ทดลองเก็บรวบรวมข้อมูลจากสื่อสังคมออนไลน์เพื่อใช้ทำการวิเคราะห์ถึงเนื้อหาที่เกี่ยวกับด้านสุขภาพ ซึ่งพบว่าการใช้ข้อมูลบนสื่อสังคมออนไลน์ มาศึกษาด้วยเครื่องมือทางด้านปัญญาประดิษฐ์ วิทยาศาสตร์ข้อมูลและเทคโนโลยีดิจิทัล สามารถเข้าถึงการพูดถึงสุขภาพและการแสดงความเห็นเกี่ยวกับสุขภาพได้อย่างมีประสิทธิภาพ [15]

จากงานวิจัยของกานดาและปราโมทย์ได้ทดลองนำข้อความภาษาไทยในสื่อสังคมออนไลน์มาใช้ในการทำเหมืองความคิดเห็น (opinion mining) เพื่อใช้ในการทำความเข้าใจข้อความการแสดงความคิดเห็นออกเป็นขั้ว ความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง [16]

งานวิจัยของ Amir และคณะได้ทดลองนำข้อมูลที่เก็บรวบรวมจาก Facebook และ Twitter เพื่อทำการวิเคราะห์ทัศนคติของประชาชนในประเทศอังกฤษต่อวัคซีนโรคโควิด ซึ่งพบว่าการนำข้อมูลจากสื่อสังคมออนไลน์ต่างๆ มาวิเคราะห์นั้นสามารถช่วยให้เราเห็นภาพและเข้าใจความเชื่อมั่นต่อวัคซีนโรคโควิดได้อย่างง่ายและรวดเร็ว อีกทั้งยังสามารถนำข้อมูลที่ได้รับมาใช้ในการวางแผนนโนบายทางด้านสาธารณสุขได้อย่างรวดเร็วและมีประสิทธิภาพมากขึ้นอีกด้วย [17]

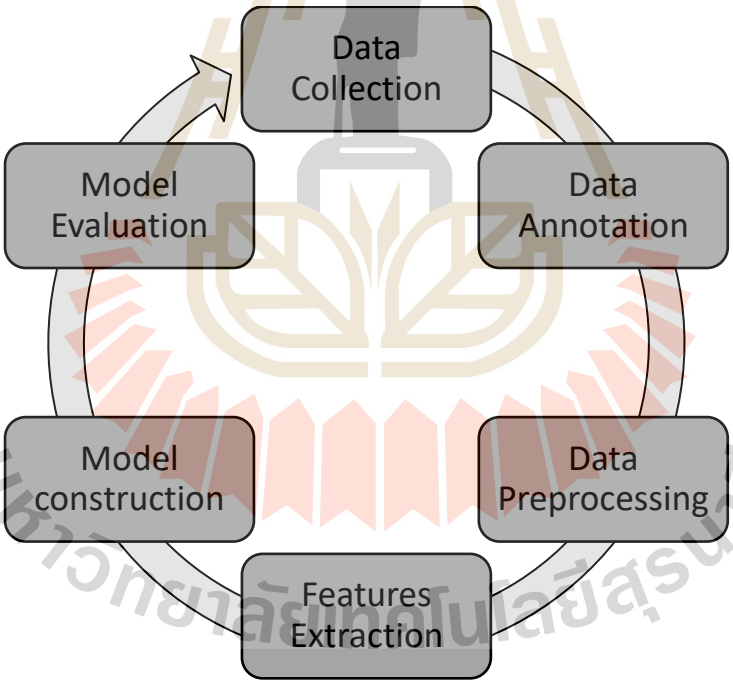
ในปัจจุบันเมื่อเครื่องมือทางด้านปัญญาประดิษฐ์ วิทยาศาสตร์ข้อมูลและเทคโนโลยีดิจิทัลมีความเจริญมากขึ้น นักวิจัยได้ใช้ประโยชน์ของเทคโนโลยีเหล่านี้เพื่อเข้ามาช่วยในการศึกษาและวิเคราะห์ข้อความต่างๆ ในปริมาณมากโดยอัตโนมัติ ซึ่งสามารถช่วยให้คอมพิวเตอร์และนักวิจัยสามารถเข้าใจอารมณ์ ความรู้สึกของผู้ส่งสารได้รวดเร็วและจำนวนมากขึ้น [12], [17], [18] ซึ่งประโยชน์ของเครื่องมือเหล่านี้จะทำให้เราสามารถสรุปถึงเรื่องราวที่ผู้ใช้งานบนสื่อสังคมออนไลน์กำลังพูดคุยหรือแสดงความคิดเห็นกันได้ง่ายและสะดวกขึ้น ทำให้เราสามารถมีข้อมูลมาใช้ในการวางแผนเชิงนโยบายเพื่อแก้ไขปัญหาได้สะดวกและมีประสิทธิภาพมากยิ่งขึ้น [17], [19], [20]

จากศึกษาของผู้วิจัยพบว่ามีความเป็นไปได้ในการใช้ปัญญาประดิษฐ์ วิทยาศาสตร์ข้อมูลและเทคโนโลยีดิจิทัลเพื่อมาศึกษาพฤติกรรมของผู้ใช้งานสื่อสังคมออนไลน์ เพื่อนำมาสร้างชุดความรู้สาธารณะเกี่ยวกับปัญหาสุขภาพจิต และพัฒนาโมเดลการทำนายโอกาสป่วยเป็นโรคซึมเศร้าได้

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 3
วิธีดำเนินการวิจัย

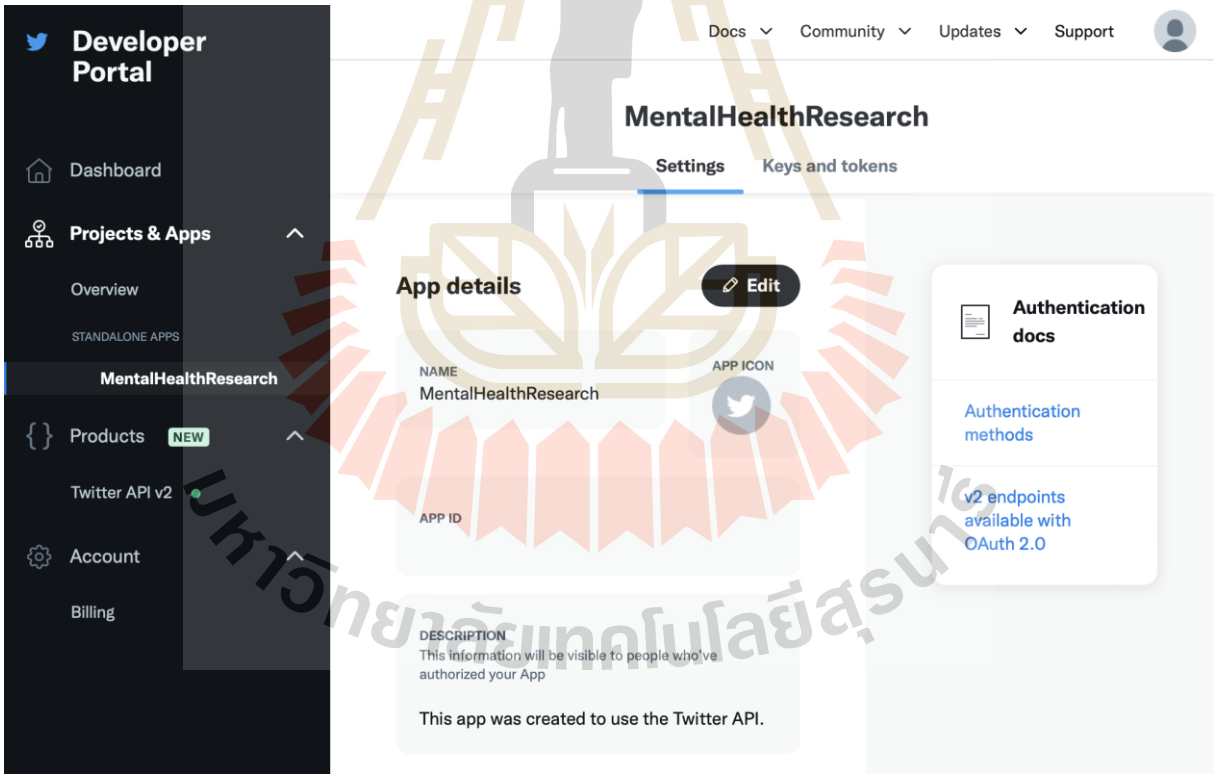
การศึกษาวิจัยครั้งนี้ได้ทำการแบ่งกระบวนการทำวิจัยเพื่อวิเคราะห์หาปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้าออกเป็น 6 ขั้นตอนหลัก ได้แก่ 1.การเก็บรวบรวมข้อมูล 2.การระบุประเภทของผู้ใช้งาน 3.การเตรียมข้อมูล 4.การสกัดลักษณะเฉพาะ (Features) 5.การพัฒนาโมเดล และ 6.การประเมินประสิทธิภาพของโมเดล ดังแสดงในรูปที่ 1 โดยในแต่ละขั้นตอนผู้วิจัยได้ใช้ Python ในการวิเคราะห์และประมวลผลข้อมูล ข้อมูลที่จะเก็บและนำมาใช้ในการวิเคราะห์จะเป็นการเก็บรวบรวมมาจากสื่อสังคมออนไลน์ Twitter โดยข้อมูลที่นำมาจะเป็นข้อมูลภาษาอังกฤษ



รูปที่ 1 กระบวนการวิเคราะห์หาปัจจัยที่ใช้ในการทำนายโอกาสของการเป็นโรคซึมเศร้า

1. **การเก็บรวบรวมข้อมูล (Data Collection)**

งานวิจัยชิ้นนี้เก็บรวบรวมข้อมูลจากแฟลตฟอร์มสื่อสังคมออนไลน์ Twitter ซึ่งแฟลตฟอร์มนี้เป็นแฟลตฟอร์มที่อนุญาติให้ผู้ใช้งานสามารถนำเสนอข้อมูลข้อความ รูปภาพ และวีดีโอได้ ซึ่งผู้ใช้จะใช้แฟลตฟอร์มนี้ในการนำเสนอข้อมูลกิจกรรม อารมณ์ ความรู้สึก รวมถึงความคิดเห็นในเรื่องต่างๆ ได้ ในบางครั้งผู้ใช้งานยังสามารถผู้ถึงเรื่องสุขภาพของตัวผู้ใช้งานเองอีกด้วย การเก็บข้อมูลจากแฟลตฟอร์มนี้จะใช้การดึงข้อมูลผ่านส่วนต่อประสานโปรแกรมประยุกต์ (Application Program Interface – API) ซึ่งการเข้าถึงข้อมูลผู้ใช้งานในแฟลตฟอร์ม Twitter จะต้องทำการสมัครเป็นสมาชิกรวมถึงทำการระบุรายละเอียดโครงการวิจัย และขอบเขตของข้อมูลที่ต้องการเข้าถึง ดังแสดงในรูปที่ 2



รูปที่ 2 หน้าจอแสดงการเข้าถึงข้อมูลบน Twitter

การเขียนคำสั่งเพื่อกำหนดขอบเขตการดึงข้อมูลจะถูกพัฒนาด้วยภาษา Python และใช้ไลบรารี่ Twitter-built v2 ซึ่งการดึงข้อมูลจากแบ่งเป็น 3 ขั้นตอน

1. การค้นหาข้อความที่พูดถึง #depression
2. การดึงข้อมูลข้อความทั้งหมดบนโปรไฟล์ของผู้ใช้งานที่ถูกระบุว่าพูดถึงการเป็นโรคซึมเศร้า
3. การดึงข้อมูลข้อความทั้งหมดบนโปรไฟล์ของผู้ใช้งานทั่วไป

```python
def gettimelines(userid, max_results=100):
  print(userid)
  pagination_token = ""
  timeline = []

  while pagination_token != None:
    try:
      if pagination_token == "":
        tweets = api.get_timelines(user_id=userid, max_results=max_results,
                        tweet_fields=('created_at', 'attachments', 'author_id', 'context_annotations',
                        'conversation_id', 'geo', 'in_reply_to_user_id', 'referenced_tweets',
                        'public_metrics', 'possibly_sensitive', 'source','lang'),
                        )
      else:
        tweets = api.get_timelines(user_id=userid, max_results=max_results,
            tweet_fields=('created_at', 'attachments', 'author_id', 'context_annotations',
            'conversation_id', 'geo', 'in_reply_to_user_id', 'referenced_tweets',
            'public_metrics', 'possibly_sensitive', 'source', 'lang'),
            pagination_token=pagination_token
            )
    except Exception as e:
      print(eval(str(e)))


    timeline += tweets.data
    pagination_token = tweets.meta.next_token

    if not tweets or pagination_token == None:
      break
  time.sleep(10)

  return timeline
```

รูปที่ 3 คำสั่งการเก็บข้อมูลข้อความจากโปรไฟล์ผู้ใช้งานด้วย Python

2. การระบุประเภทของข้อมูล (Data Annotation)

การระบุประเภทของข้อมูลถูกแบ่งเป็น 2 กระบวนการย่อย

1. การระบุประเภทของข้อความ เมื่อทำการค้นหาข้อความที่พูดถึง #depression เรียบร้อยแล้ว จะเข้าสู่กระบวนการระบุประเภทของข้อความ โดยนักวิจัยจะทำการอ่านข้อความที่ละข้อความและพิจารณาถึงความหมายของข้อความ ถ้าข้อความสื่อถึงการพูดเกี่ยวกับการเป็นโรคซึมเศร้าจะถูกระบุเป็น *"ข้อความเกี่ยวกับโรคซึมเศร้า"* ถ้าไม่ได้พูดถึงเรื่องโรคซึมเศร้า เช่น เรื่องภาวะเศรษฐกิจตกต่ำ (Depression) เรื่องตลก และการพูดถึงโรคซึมเศร้าของผู้อื่น เป็นต้น จะถูกระบุเป็น *"ข้อความไม่เกี่ยวกับโรคซึมเศร้า"*
2. การระบุประเภทผู้ใช้งาน ผู้ใช้งานจะถูกระบุเป็น *"ผู้ใช้ประเภทโรคซึมเศร้า"* ต่อเมื่อข้อความที่ถูกค้นหามาถูกระบุเป็นประเภท "ข้อความเกี่ยวกับโรคซึมเศร้า" และสามารถดึงข้อมูลบนโปรไฟล์ของผู้ใช้งานคนนั้นๆ ได้ ส่วน *"ผู้ใช้ประเภทไม่เป็นโรคซึมเศร้า"* จะได้มาจากการสุ่มค้นหาตามช่วงเวลาที่กำหนดและผู้ใช้งานคนนั้นไม่อยู่ในประเภทผู้ใช้งานที่เป็นโรคซึมเศร้า

3. การเตรียมข้อมูล (Data Preprocessing)

เมื่อทำการเก็บข้อมูลจากโปรไฟล์ผู้ใช้งานแต่ละประเภทเรียบร้อยแล้ว ผู้วิจัยได้ใช้เครื่องมือ Regex เพื่อทำการลบข้อมูลที่ไม่จำเป็นออกจากข้อความ เช่น ชื่อของเว็บไซต์หรือลิงค์ต่างๆ (URL) ลบอักขระพิเศษต่างๆ (?.-V@$%) ตัวเลขต่างๆ และการลบคำหยุดต่างๆ (Stop words) เป็นต้น การลบข้อความพวกนี้ออกเพราะเป็นข้อความที่ไม่สื่อความหมาย และทำให้ประหยัดเวลาในการประมวลผลข้อมูล หลังจากทำการลบข้อมูลที่ไม่จำเป็นเรียบร้อยแล้ว จะทำการรวมข้อความทั้งหมดบนโปรไฟล์ของผู้ใช้งานแต่ละคนเข้าด้วยกัน

4. การสกัดคุณลักษณะเฉพาะ (Features Extraction)

เมื่อได้ข้อความที่ทำการลบข้อมูลที่ไม่จำเป็นออกไปเรียบร้อยแล้ว จะเข้าสู่ขั้นตอนการสกัดคุณลักษณะเฉพาะ โดยจะเครื่องมือ Count Vectorizer และ TF-IDF Vectorizer จาก Scikit learn[1] ในการสกัดข้อมูล

[1] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

เครื่องมือเหล่านี้จะทำการแปลงข้อมูลในรูปแบบข้อความเป็นข้อมูลในรูปแบบตัวเลข เพื่อให้อยู่ในรูปแบบที่เหมาะสมกับการประมวลผลด้วยการเรียนรู้ด้วยเครื่องต่อไป

5. การพัฒนาโมเดล (Model Construction)

หลังจากจัดเตรียมข้อมูลที่พร้อมสำหรับการสอนโมเดลการเรียนรู้ด้วยเครื่องเรียบร้อยแล้ว ผู้วิจัยได้ทำการสอนโมเดลการเรียนรู้ด้วยเครื่องโดยเลือกใช้โมเดลดังนี้ Support Vector Machine (SVM), Logistic Regression และ Ada Boost โดยการสอนโมเดลการเรียนรู้ด้วยเครื่องนี้ได้ใช้พารามิเตอร์ค่าเริ่มต้น (Default) ในการกำหนดค่าเพื่อสอนโมเดล รวมทั้งมีการกำหนดค่าสัดส่วนประเภท (Class Weight) เพื่อทำให้โมเดลเรียนรู้ข้อมูลได้อย่างสมดุล

6. การประเมินประสิทธิภาพของโมเดล (Model Evaluation)

เมื่อทำการสอนโมเดลการเรียนรู้ด้วยเครื่องเรียบร้อยแล้ว ผู้วิจัยได้ทำการทดสอบประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่องโดยใช้ค่าค่าถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่าความถ่วงดุล (F-Measure) เพื่อเปรียบเทียบประสิทธิภาพระหว่างโมเดลการเรียนรู้ด้วยเครื่องที่เลือกใช้ โดยผู้วิจัยได้ใช้วิธีการแบ่งข้อมูลออกเป็น Train 70% เพื่อใช้ในการสร้างโมเดล และ Test 30% เพื่อใช้ในการทดสอบโมเดล

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4
ผลการวิจัย

ในบทนี้จะกล่าวถึงกระบวนการเก็บข้อมูลและรายละเอียดข้อมูลที่เก็บรวบรวมมาได้ ผลการวิเคราะห์และรายงานผลการประเมินประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่องเพื่อใช้ในการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า

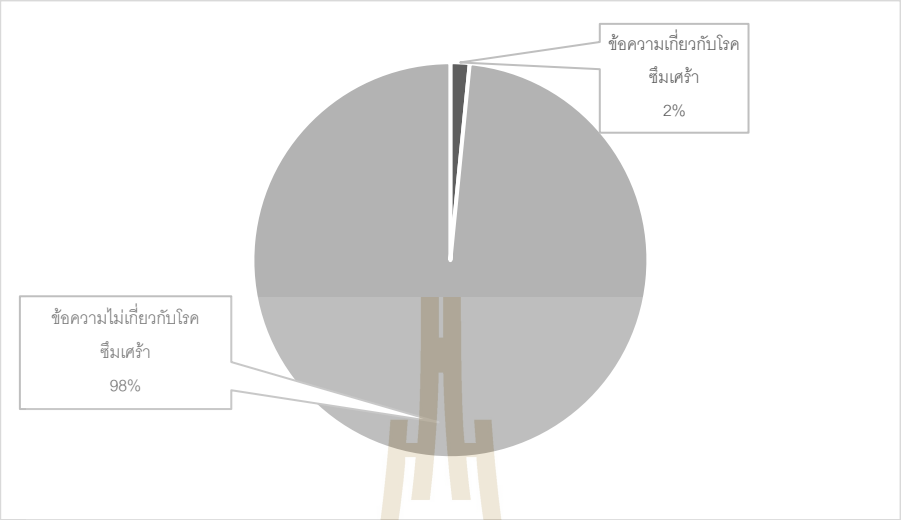
1. รายละเอียดข้อมูล

การเก็บข้อความที่เกี่ยวข้องกับโรคซึมเศร้า

การเก็บข้อมูลด้วยวิธีการค้นหาข้อความที่มีการใช้แฮชแท็ก #depression ระหว่างวันที่ 1 ตุลาคม 2565 ถึง 9 ธันวาคม 2565 ได้ข้อมูลจำนวน 4,673 ข้อความ จากผู้ใช้งาน 2,668 คน เมื่อทำการเก็บรวบรวมข้อมูลเรียบร้อยแล้ว ได้ทำการระบุประเภทข้อความเป็น *“ข้อความเกี่ยวกับโรคซึมเศร้า”* และ *“ข้อความไม่เกี่ยวกับโรคซึมเศร้า”*

หลังจากระบุประเภทข้อความเรียบร้อยแล้ว ได้ข้อความที่เป็นประเภท *“ข้อความเกี่ยวกับโรคซึมเศร้า” ทั้งหมด 73 ข้อความ หรือประมาณ 2% และข้อความประเภท “ข้อความไม่เกี่ยวกับโรคซึมเศร้า” 4,600* ข้อความ (98%) ดังแสดงในรูปที่ 4

มหาวิทยาลัยเทคโนโลยีสุรนารี



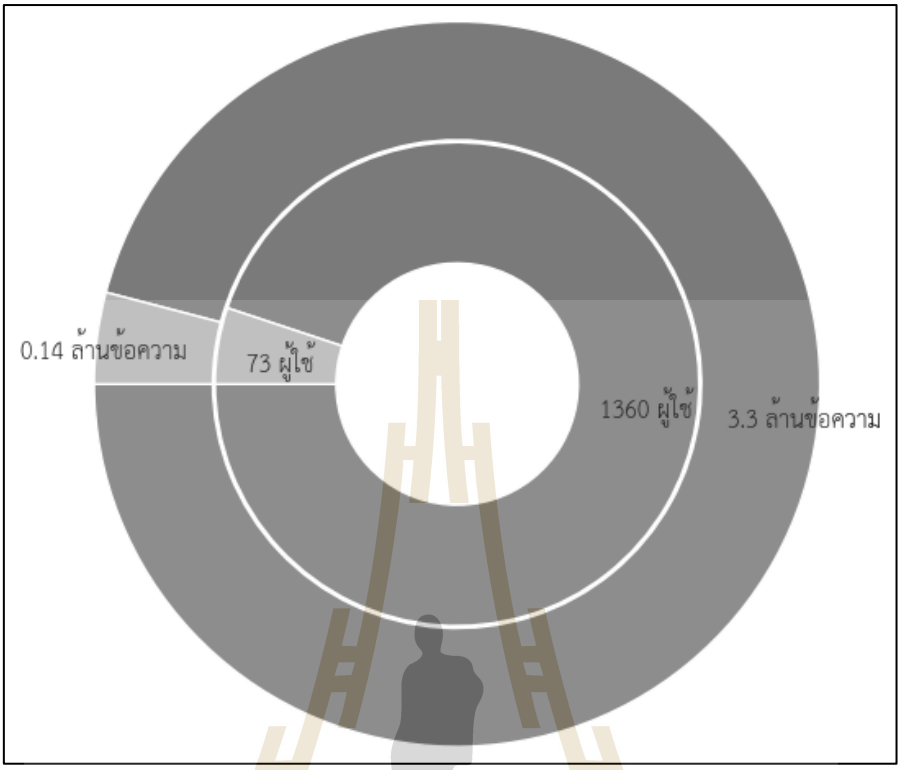
รูปที่ 4 สัดส่วนประเภทข้อเกี่ยวกับโรคซึมเศร้าและไม่เกี่ยวกับโรคซึมเศร้า

การเก็บข้อความจากโปรไฟล์ผู้ใช้งาน

เมื่อได้ข้อความที่พูดถึงการเป็นโรคซึมเศร้าของตัวเองเรียบร้อยแล้ว ผู้วิจัยได้ทำการเก็บข้อความทั้งหมดจากโปรไฟล์ของผู้ที่โพสข้อความที่พูดถึงการเป็นโรคซึมเศร้า จำนวนทั้งหมด 73 โปรไฟล์หรือ 73 ผู้ใช้งาน และทำการระบุผู้ใช้เหล่านี้เป็น *“ผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า”*

จากนั้นผู้วิจัยได้ทำการสุ่มหาผู้ใช้งานที่โพสข้อความระหว่างวันที่ 1 ธันวาคม ถึง 7 ธันวาคม 2565 ซึ่งสามารถได้ผู้ใช้งานทั้งสิ้น 1,360 ผู้ใช้งาน จากนั้นทำการเก็บรวบรวมข้อความทั้งหมดบนโปรไฟล์ของผู้ใช้งานโดยระบุผู้ใช้งานกลุ่มนี้เป็น *“ผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า”* เพื่อนำข้อมูลของผู้ใช้กลุ่มนี้ใช้ในการสร้างโมเดลการเรียนรู้ด้วยเครื่องต่อไป

ข้อมูลในงานวิจัยนี้ประกอบด้วยผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าจำนวน 73 คน และผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าอีกจำนวน 1,360 คน โดยมีข้อความจากผู้ใช้งานเหล่านี้ 1.4 แสนข้อความจากผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า และ 3.3 ล้านข้อความจากผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า (ดังแสดงในรูปที่ 5) ข้อมูลส่วนนี้จะถูกนำไปใช้งานการสร้างโมเดลการเรียนรู้เพื่อทำการทำนายผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าต่อไป



รูปที่ 5 จำนวนผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า และจำนวนข้อความของผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า สีเทาอ่อนหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและสีเทาเข้มคือผู้ใช้งานที่ไม่มีโอกาสเป็นโรคซึมเศร้า

2. การประเมินประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่อง

หลังจากผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลของผู้ใช้งานมีโอกาสเป็นโรคซึมเศร้าจำนวน 73 คน และผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าอีกจำนวน 1,360 คน เรียบร้อยได้ผู้วิจัยได้ทำการสร้างโมเดลการเรียนรู้ด้วยเครื่องเพื่อใช้ในการทำนายผู้ใช้งานมีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า ซึ่งกระบวนการสร้างโมเดลผู้ใช้งานได้ทำการลบชื่อของเว็บไซต์หรือลิงค์ (URL) อักขระพิเศษ (?.-\/@$%) ตัวเลขและคำหยุดต่างๆ ออกจากข้อความแต่ละข้อความของผู้ใช้งาน

เมื่อลบข้อมูลที่ไม่จำเป็นเรียบร้อยแล้ว ผู้วิจัยได้แบ่งข้อมูลออกเป็นสองส่วนได้แก่ Training set 70% เพื่อใช้ข้อมูลในส่วนนี้ในการสร้างโมเดล และ Test set 30% เพื่อใช้ในการทดสอบประสิทธิภาพโมเดล หลังจากนั้นทำการสกัดคุณลักษณะเฉพาะด้วยวิธีการ Count Vectorizer และ TF-IDF Vectorizer เพื่อใช้เป็น

ข้อมูลในการสอนโมเดลการเรียนรู้ด้วยเครื่อง จากนั้นผู้วิจัยได้ทำการสร้างโมเดลการเรียนรู้ด้วยเครื่องโดยเลือกใช้อัลกอริทึม (Algorithm) Support Vector Machine (SVM), Logistic Regression และ AdaBoost ซึ่งได้ผลลัพธ์การประเมินประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่องดังนี้

โมเดลการเรียนรู้ด้วยเครื่องที่สอนด้วยการสกัดคุณลักษณะแบบ Count Vectorizer และ อัลกอริทึม SVM ได้ความถูกต้อง (Accuracy) 61% และได้ค่าความแม่นยำสูงที่สุดที่ 95% ค่าระลึก 61% และค่าความถ่วงดุลที่ 72% ดังแสดงในตารางที่ 1 โดยโมเดล SVM ของ Scikit learn สามารถกำหนดค่าพารามิเตอร์เพื่อถ่วงความสมดุลของข้อมูลได้ ผู้วิจัยจึงทำการปรับค่าพารามิเตอร์ class_weight='balanced' เนื่องจากข้อมูลที่ใช้ในการสร้างโมเดลมีความไม่สมดุล (Imbalanced data) ส่วนพารามิเตอร์อื่นๆ ไม่ได้มีปรับการค่า

Class	Precision	Recall	F1-score
Non-depression	0.99	0.59	0.74
Depression	0.09	0.89	0.16
Weighted average	0.95	0.61	0.72

ตารางที่ 1 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล SVM และสกัดข้อมูลด้วยวิธี Count Vectorizer

ตารางที่ 2 แสดงประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่องที่สอนด้วยการสกัดข้อมูลด้วยวิธี Count Vectorizer และอัลกอริทึม Logistic Regression ซึ่งได้ค่าความถูกต้อง 93% และค่าความแม่นยำ ค่าระลึก และค่าความถ่วงดุลที่ 93% เท่ากัน โดยผู้วิจัยจึงทำการปรับค่าพารามิเตอร์ class_weight='balanced' เพื่อให้โมเดลทำงานได้สมดุลกับข้อมูลมากขึ้น ส่วนพารามิเตอร์อื่นๆ ใช้แบบค่าเริ่มต้น (Defaults)

มหาวิทยาลัยเทคโนโลยีสุรนารี

Class	Precision	Recall	F1-score
Non-depression	0.96	0.96	0.96
Depression	0.15	0.17	0.16
Weighted average	0.93	0.93	0.93

ตารางที่ 2 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล Logistic Regression และสกัดข้อมูลด้วยวิธี Count Vectorizer

ผลประเมินประสิทธิภาพการสร้างโมเดลการเรียนรู้ด้วยเครื่องด้วย AdaBoost และสกัดข้อมูลด้วยวิธี Count Vectorizer ได้ค่าค่าความถูกต้อง 97% ค่าความแม่นยำ 97% ค่าระลึก 97% และค่าความถ่วงดุลที่ 97% ดังแสดงในตารางที่ 3 ได้กำหนดค่า n_estimators=300 ส่วนค่าอื่นๆ ใช้ค่าเริ่มต้น

Class	Precision	Recall	F1-score
Non-depression	0.98	0.99	0.98
Depression	0.62	0.56	0.59
Weighted average	0.97	0.97	0.97

ตารางที่ 3 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี Count Vectorizer

การสร้างโมเดลการทำนายผลผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้าด้วยวิธีการ TF-IDF Vectorizer และ SVM ให้ประสิทธิภาพในการทำนายผลด้วยความถูกต้อง 92% ค่าความแม่นยำ 93% ค่าระลึก 92% และค่าความถ่วงดุลที่ 93% ดังผลลัพธ์ในตารางที่ 4 โดยทำการปรับค่าพารามิเตอร์ class_weight='balanced' และค่าอื่นๆ ใช้ค่าเริ่มต้น

Class	Precision	Recall	F1-score
Non-depression	0.97	0.95	0.96
Depression	0.17	0.22	0.19
Weighted average	0.93	0.92	0.93

ตารางที่ 4 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล SVM และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer

ตารางที่ 5 แสดงผลการสร้างโมเดลการเรียนรู้ด้วยเครื่องเพื่อใช้ในการทำนายผู้ใช้งานมีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าด้วยวิธีการ Logistic Regression และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer ได้ค่าค่าความแม่นยำสูงสุดที่ 95% และได้ค่าความถูกต้องที่ 87% โดยทำการปรับค่าพารามิเตอร์ class_weight='balanced' และค่าอื่นๆ ใช้ค่าเริ่มต้น

Class	Precision	Recall	F1-score
Non-depression	0.98	0.88	0.93
Depression	0.19	0.67	0.30
Weighted average	0.95	0.87	0.90

ตารางที่ 5 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล Logistic Regression และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer

ผู้วิจัยได้ทำการประเมินประสิทธิการทำนายผลโมเดลด้วยการสอนโมเดลด้วย AdaBoost Classifier และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer ได้ค่าความถูกต้องที่ 97% รวมถึงค่าความแม่นยำ ค่าระลึก และค่าความถ่วงดุลที่ 97% เท่ากัน ดังผลลัพธ์ในตารางที่ 6 โดยได้กำหนดค่า n_estimators=300 ส่วนค่าอื่นๆ ใช้ค่าเริ่มต้น

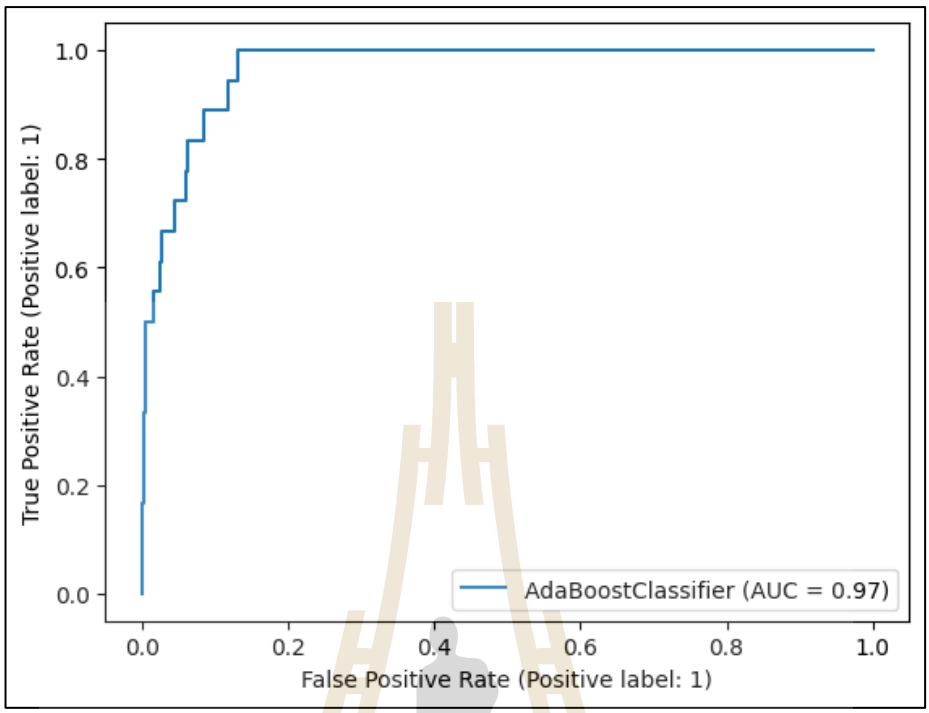
Class	Precision	Recall	F1-score

Non-depression	0.98	0.99	0.98
Depression	0.62	0.56	0.59
Weighted average	0.97	0.97	0.97

ตารางที่ 6 ผลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ด้วยโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer

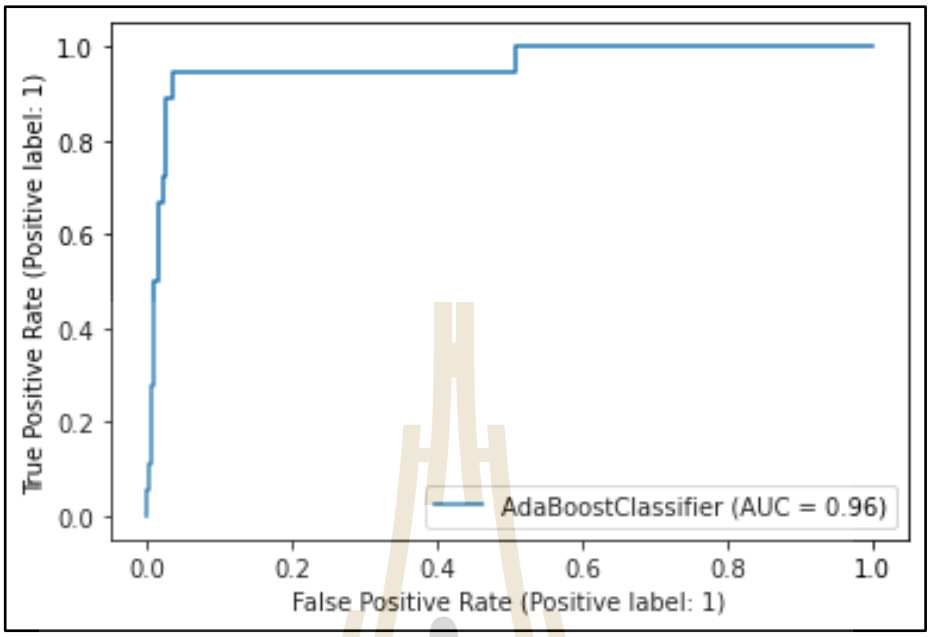
หลักจากผู้วิจัยได้ทำการประเมินประสิทธิภาพโมเดลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้าเรียบร้อยแล้ว ผู้วิจัยได้ทำการเลือกอัลกอริทึมของโมเดลที่มีประสิทธิภาพที่ดีที่สุดจากการสกัดคุณลักษณะมาทำการคำนวณค่า Receiver operating characteristic curves (ROC) ซึ่งโมเดลที่ถูกสอนด้วยอัลกอริทึม AdaBoost Classifier และการสกัดข้อมูลด้วยวิธี Count Vectorizer ได้ค่า ROC เท่ากับ 97% ส่วนการสร้างโมเดลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้าด้วยวิธี AdaBoost Classifier และการสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer สามารถให้ผลลัพธ์ ROC สูงที่สุดเท่ากับ 96% ดังรูปที่ 6 และรูปที่ 7 ตามลำดับ

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 6 ผลการประเมินโมเดลการทำนายผลด้วย Receiver operating characteristic curves จากโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี Count Vectorizer

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 7 ผลการประเมินโมเดลการทำนายผลด้วย Receiver operating characteristic curves จากโมเดล AdaBoost Classifier และสกัดข้อมูลด้วยวิธี TF-IDF Vectorizer

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในบทนี้จะเป็นการสรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะเกี่ยวกับการวิจัยในขั้นตอนต่อไป รวมไปถึงการประยุกต์ใช้ผลการวิจัยที่ได้

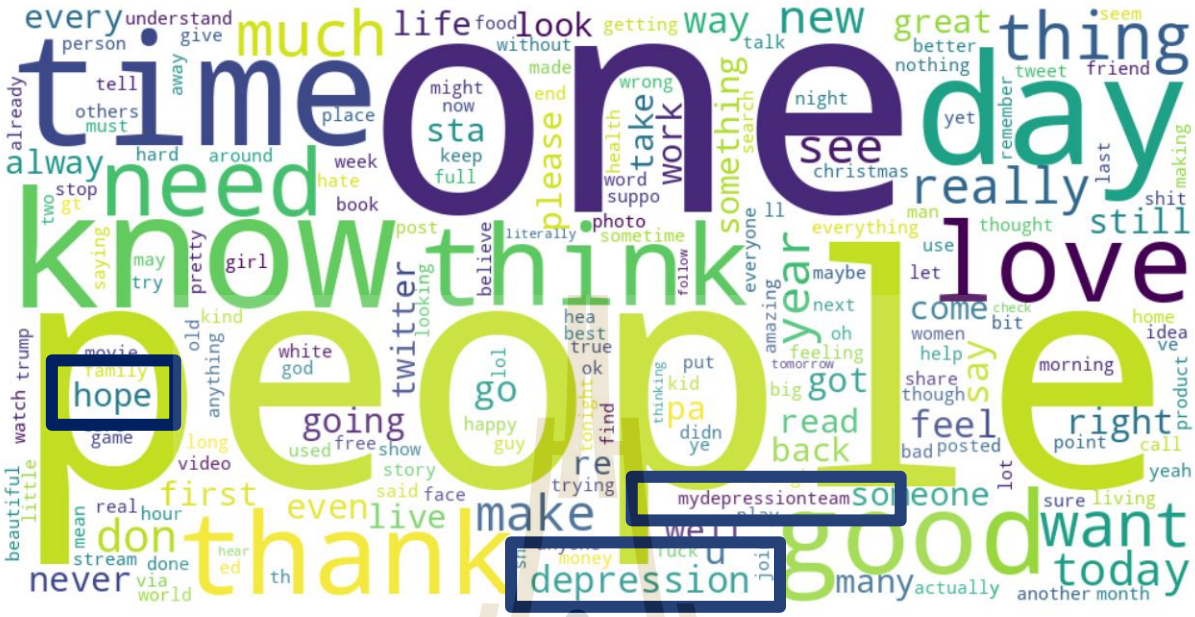
1. อภิปรายผล

งานวิจัยชิ้นนี้นำเสนอการสร้างโมเดลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้า ซึ่งจากผลการประเมินประสิทธิภาพโมเดลการทำนายที่ได้นำเสนอไปในบทที่ 4 ได้แสดงให้เห็นว่าการใช้โมเดลการเรียนรู้ด้วยเครื่องสามารถนำมาใช้ในการสร้างโมเดลการทำนายผู้ใช้งานสื่อสังคมออนไลน์ที่มีโอกาสเป็นโรคซึมเศร้าได้ ซึ่งผลลัพธ์จากการสร้างโมเดลได้แสดงให้เห็นว่า การสร้างโมเดลการเรียนรู้ด้วยเครื่องด้วยอัลกอริทึม AdaBoost Classifier โดยใช้วิธีการสกัดข้อมูลด้วย Count Vectorizer และ TF-IDF ได้ผลลัพธ์ที่คล้ายคลึงกัน ดังแสดงในตารางที่ 3 และ ตารางที่ 6 แต่ผลการประเมินประสิทธิภาพด้วย ROC มีความแตกต่างกันเพียงเล็กน้อย การสร้างโมเดลการเรียนรู้ด้วย AdaBoost Classifier และ Count Vectorizer สามารถทำนายผลได้ดีกว่าการสร้างโมเดลด้วย AdaBoost Classifier และ TF-IDF Vectorizer 1% ดังแสดงในรูปที่ 6 และ รูปที่ 7 จากการศึกษาและวิเคราะห์ผลผู้วิจัยได้คำนึงถึงผลของปัญหาความไม่สมดุลกับระหว่างประเภทข้อมูล (Imbalanced data) ทำให้ผู้วิจัยต้องทำการปรับค่าพารามิเตอร์ของโมเดลที่สร้างขึ้นมา เพื่อแก้ไขปัญหาดังกล่าว และส่งผลให้โมเดลทำนายผลได้มีประสิทธิภาพมากยิ่งขึ้น

การวิเคราะห์ปัจจัยที่สามารถระบุโอกาสที่การเป็นซึมเศร้า

หลังจากที่ผู้วิจัยทำการประเมินประสิทธิภาพโมเดลการทำนายผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าจากข้อมูลสื่อสังคมออนไลน์เรียบร้อยแล้ว ผู้วิจัยได้ทำการศึกษาถึงปัจจัยที่แตกต่างกันระหว่างผู้ใช้งานกลุ่มที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า โดยใช้ Word cloud เป็นเครื่องมือในการวิเคราะห์การใช้คำที่แตกต่างกันระหว่างผู้ใช้งาน 2 กลุ่ม จากรูปที่ 8 แสดงให้เห็นว่าผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าจากใช้คำว่า Depression มากกว่าผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า

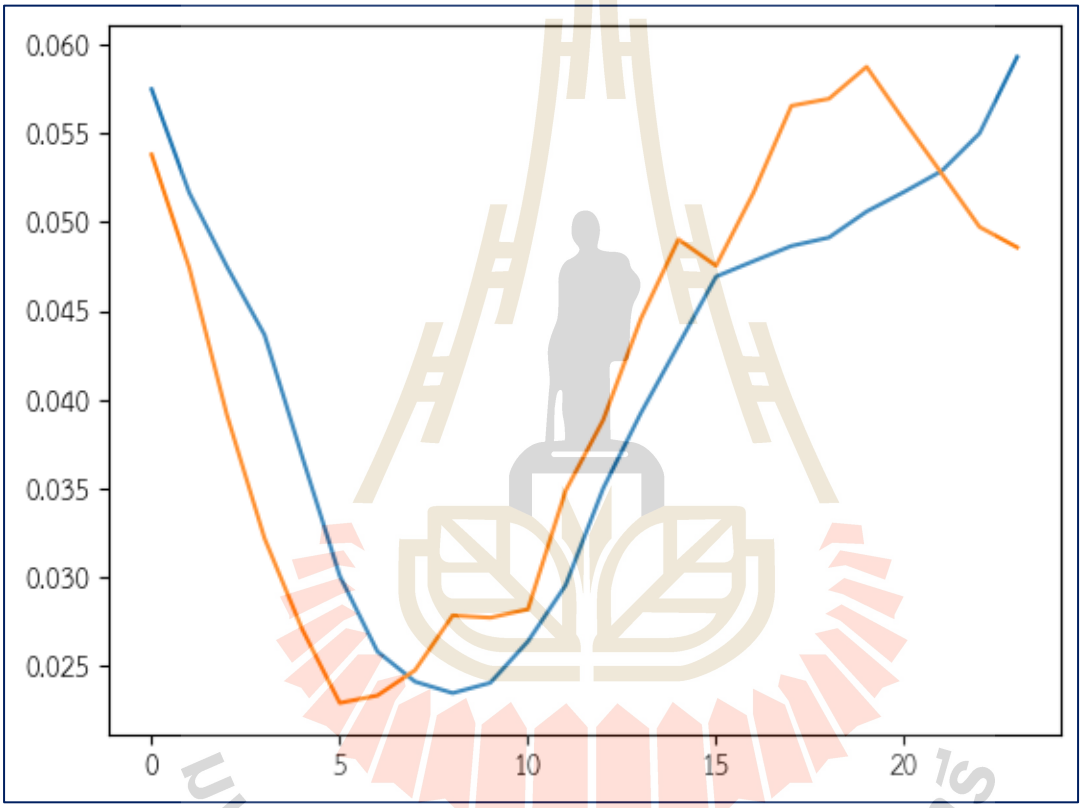
มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 8 การใช้คำของผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า

รูปที่ 9 การใช้คำของผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า

อีกหนึ่งปัจจัยที่น่าสนใจที่สามารถใช้ในการแบ่งกลุ่มผู้ใช้งานออนไลน์ระหว่างผู้ใช้งานกลุ่มที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า คือ การใช้คำว่า “Hope” จากรูปที่ 9 ผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าจะมีการใช้งานคำว่า “Hope” น้อยกว่าผู้ใช้งานอีกกลุ่มหนึ่ง ซึ่งจากงานวิจัย [21] ระบุว่าผู้ที่มีภาวะซึมเศร้ามีความสัมพันธ์กับความหวังในชีวิต ซึ่งมีส่วนสำคัญที่ทำให้ผู้ใช้ที่มีโอกาสเป็นโรคซึมเศร้าใช้คำ “Hope” มากกว่าผู้ใช้งานทั่วไป

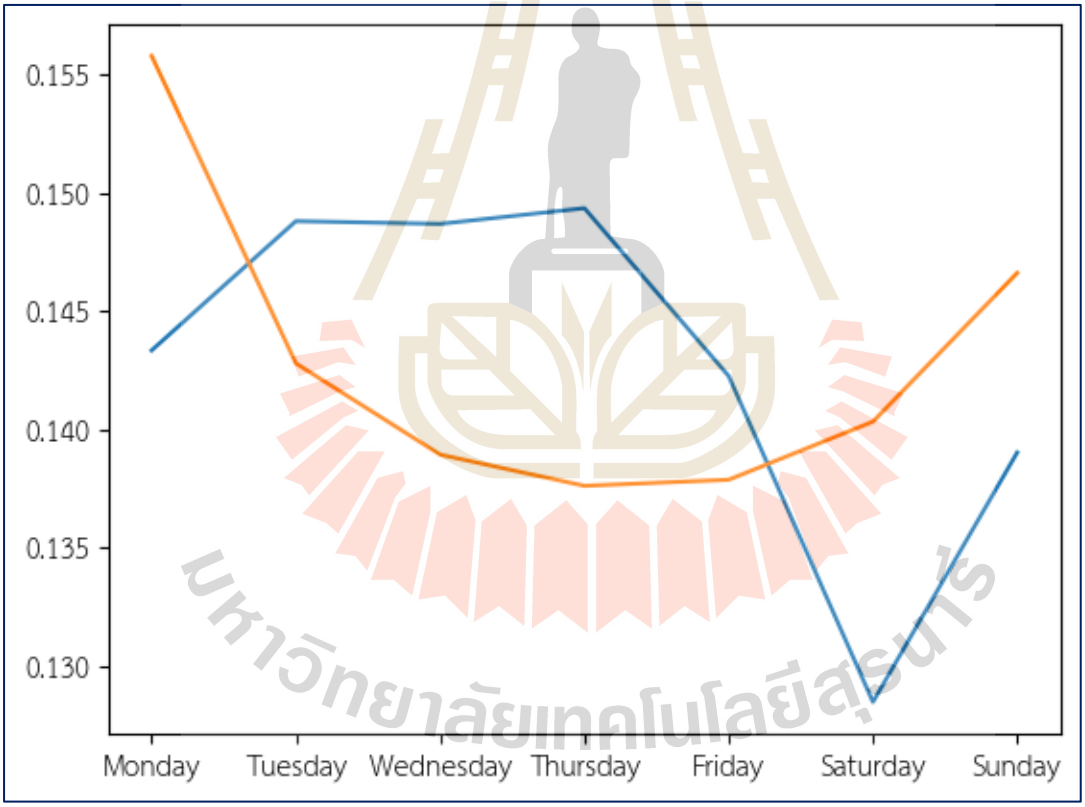


รูปที่ 10 ช่วงเวลาในการโพสข้อความบนโปรไฟล์ผู้ใช้งานระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า สีน้ำเงินคือผู้ใช้งานทั่วไป สีส้มหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า

ปัจจัยต่อมาที่ผู้วิจัยสนใจในการวิเคราะห์ปัจจัยความแตกต่างระหว่างผู้ใช้งานที่เป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าคือ เวลาในการเล่นสื่อสังคมออนไลน์ รูปที่ 10 แสดงให้เห็นว่าผู้ใช้งานทั่วไปมีการเล่นสื่อสังคมออนไลน์มากที่สุดในช่วงเวลา 23.00 น. ในขณะที่ผู้ใช้งานกลุ่มที่มีโอกาสเป็นโรคซึมเศร้าจะใช้งาน

สื่อสังคมออนไลน์มากที่สุดในช่วงเวลา 19.00 น. ส่วนในช่วงเวลาหลังเที่ยงคืนจนถึงเวลา 05.00 น. ผู้ใช้งานทั้งสองกลุ่มจะมีการใช้งานสื่อสังคมออนไลน์น้อยที่สุด ซึ่งสอดคล้องกับงานวิจัย [22]

วันที่ในการโพสข้อความบนโปรไฟล์ผู้ใช้งานระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าเป็นอีกปัจจัยหนึ่งที่สามารถใช้ในการแบ่งแยกผู้ใช้งานทั้งสองกลุ่มออกจากกันได้ รูปที่ 11 แสดงให้เห็นว่าผู้ใช้งานที่เป็นซึมเศร้ามีแนวโน้มที่จะใช้งานสื่อสังคมออนไลน์ในช่วงวันหยุดเสาร์ อาทิตย์มากกว่าผู้ใช้งานทั่วไป ซึ่งข้อค้นพบนี้สอดคล้องกันกับงานวิจัย [23] ในช่วงวันเสาร์อาทิตย์ผู้ใช้งานที่มีโอกาสเป็นซึมเศร้ามีแนวโน้มที่จะหลีกหนีสังคม ทำให้ผู้งานกลุ่มนี้ไม่ออกไปพบปะผู้คน (Social isolation) แต่ใช้เวลาอยู่กับตัวเองแล้วใช้เวลาในการเล่นสื่อสังคมออนไลน์มากกว่าผู้ใช้งานทั่วไป [24]

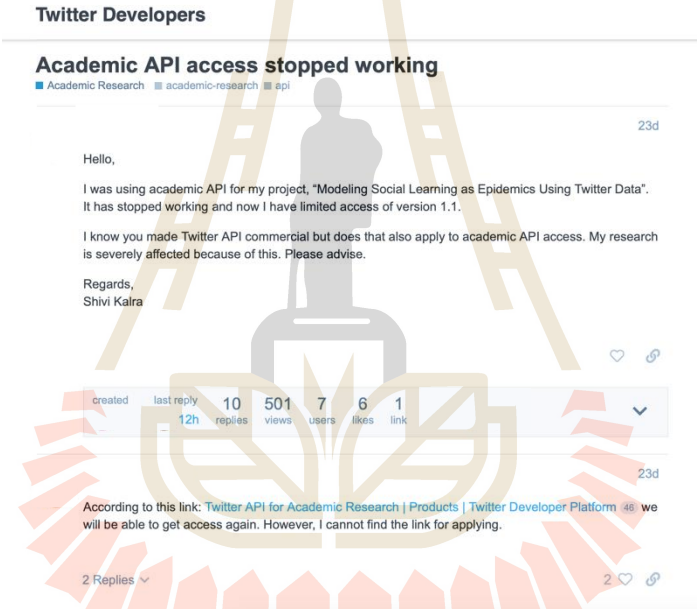


รูปที่ 11 วันที่ในการโพสข้อความบนโปรไฟล์ผู้ใช้งานระหว่างผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้าและผู้ใช้งานที่ไม่เป็นโรคซึมเศร้า สีน้ำเงินคือผู้ใช้งานทั่วไป สีส้มหมายถึงผู้ใช้งานที่มีโอกาสเป็นโรคซึมเศร้า

มหาวิทยาลัยเทคโนโลยีสุรนารี

2. ข้อเสนอแนะ

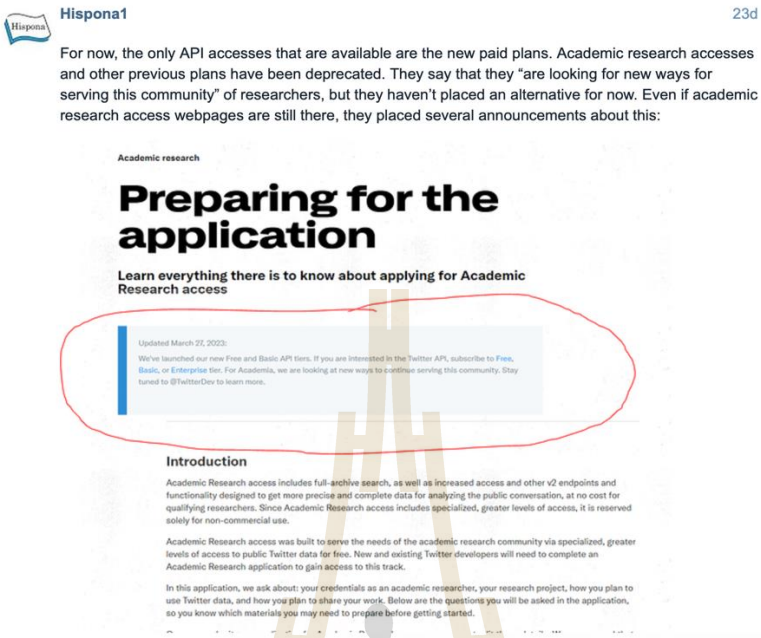
การดำเนินงานวิจัยด้านการวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์จะมีปัญหาหลักด้านการเก็บรวบรวมข้อมูลจากสื่อสังคมออนไลน์ ซึ่งในระหว่างการดำเนินงานวิจัยผู้วิจัยอาจจะพบปัญหา เกี่ยวกับการที่แพลตฟอร์มสื่อสังคมออนไลน์อาจจะมีการเปลี่ยนแปลงนโยบายการเข้าถึงข้อมูลได้ ทำให้ระหว่างการดำเนินการวิจัยอาจจะได้จำนวนข้อมูลไม่ครบถ้วนตามความต้องการได้ ดังแสดงไว้ตามภาพข้างล่าง จากภาพดังกล่าวจะแสดงให้เห็นว่านักวิจัยต่างๆ พบปัญหาด้านการไม่สามารถเข้าถึงข้อมูลสื่อสังคมออนไลน์ได้และอาจจะเปลี่ยนแผนการดำเนินการวิจัย แนวทางแก้ไขโดยทั่วไปคือทำการวิเคราะห์ข้อมูลเฉพาะที่สามารถเก็บรวบรวมได้ก่อนช่วงเวลาการจำกัดการเข้าถึงข้อมูลซึ่งอาจจะส่งผลต่อผลลัพธ์ที่ได้จากงานวิจัยได้



รูปที่ 12 กระทู้แสดงการพูดคุยเกี่ยวกับการเข้าถึงข้อมูลจาก Twitter[2]

มหาวิทยาลัยเทคโนโลยีสุรนารี

[2] https://twittercommunity.com/t/academic-api-access-stopped-working/197861/7



รูปที่ 13 รายงานการเปลี่ยนแปลงนโยบาย

งานวิจัยชิ้นนี้เป็นการแยกประเภทผู้ใช้งานโดยใช้ความเห็นของนักวิจัยซึ่งอาจจะทำให้เกิดความเอนเอียงของข้อมูลได้ (Data biases) ผู้ที่ถูกระบุเป็นประเภทผู้ใช้งานที่มีโอกาสซึมเศร้าเป็นส่วนหนึ่งของผู้ที่ป่วยเป็นโรคซึมเศร้าที่ถูกระบุโดยวิธีการรักษาโดยแพทย์ และผู้ใช้งานสื่องสังคมออนไลน์ส่วนใหญ่เป็นผู้ที่อยู่ในกลุ่มวัยรุ่นและชอบใช้เทคโนโลยีในการติดต่อสื่อสารรวมถึงการแสดงอารมณ์ความรู้สึกผ่านทางโลกออนไลน์ ซึ่งอาจจะมีลักษณะที่แตกต่างกับผู้ป่วยซึมเศร้ากลุ่มอื่นได้ [25]

3. สรุปผลการวิจัย

งานวิจัยชิ้นนี้นำเสนอการสร้างโมเดลการเรียนรู้ด้วยเครื่องเพื่อใช้ในการทำนายผู้ที่มีโอกาสเป็นโรคซึมเศร้าจากข้อมูลสื่อสังคมออนไลน์โดยเฉพาะข้อมูลที่เป็นประเภทข้อความภาษาอังกฤษ ซึ่งกระบวนการสร้างโมเดลประกอบด้วยการเก็บรวบรวมข้อมูล การระบุประเภทผู้ใช้งาน การเตรียมข้อมูล การสกัดคุณลักษณะข้อมูล การพัฒนาโมเดลการเรียนรู้ด้วยเครื่อง และการประเมินประสิทธิภาพโมเดลการเรียนรู้ด้วยเครื่อง ซึ่งจากผลการวิเคราะห์ของงานวิจัยชิ้นนี้แสดงให้เห็นว่าข้อมูลที่ผู้ใช้งานสื่อสังคมออนไลน์สร้างขึ้นมาโดยเฉพาะข้อ

ความสามารถใช้ทำนายการเป็นโรคซึมเศร้าของผู้ใช้งานได้ โดยโมเดลการเรียนรู้ด้วยเครื่องของงานวิจัยชิ้นนี้สามารถทำนายผู้ใช้งานที่เป็นโรคซึมเศร้าได้ความถูกต้องที่ร้อยละ 97 จากผู้ใช้งานที่ถูกระบุว่าเป็นโรคซึมเศร้าจำนวน 73 คน และผู้ใช้งานที่ไม่เป็นโรคซึมเศร้าอีกจำนวน 1,360 คน

จากความสำเร็จดังที่แสดงข้างต้นผู้วิจัยเห็นว่างานวิจัยชิ้นนี้สามารถนำมาประยุกต์ใช้กับการวิเคราะห์ข้อมูลสื่อสังคมออนไลน์ที่ผู้ใช้งานเขียนด้วยภาษาไทยได้ โดยจะเห็นว่าในปัจจุบันมีเครื่องมือการศึกษาการประมวลผลภาษาธรรมชาติด้วยภาษาไทยอยู่มากมายซึ่งเครื่องนี้เหล่านี้เอง สามารถนำมาใช้เวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ในภาษาไทย และใช้สร้างโมเดลการทำนายผลการเป็นโรคซึมเศร้าได้

นอกจากนี้ความเป็นไปได้ดังกล่าวผู้วิจัยเห็นว่าสามารถนำวิธีการนี้ไปสร้างเป็นเครื่องมือในการช่วยเหลือผู้ใช้งานสื่อสังคมออนไลน์ในการทำนายโอกาสของการเป็นโรคซึมเศร้าและสามารถช่วยให้เค้าสามารถได้รับการรักษาที่รวดเร็วขึ้น เพื่อป้องกันปัญหาการเป็นโรคซึมเศร้าในระดับรุนแรงได้

มหาวิทยาลัยเทคโนโลยีสุรนารี

บรรณานุกรม

[1] World Health Organization, “Depression.” Accessed: Mar. 24, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[2] Public Health England, “Health matters: reducing health inequalities in mental illness.” [Online]. Available: https://www.gov.uk/government/publications/health-matters-reducing-health-inequalities-in-mental-illness/health-matters-reducing-health-inequalities-in-mental-illness

[3] D. E. Bloom *et al.*, “The Global Economic Burden of Noncommunicable Diseases,” World Economic Forum. [Online]. Available: http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf

[4] World Health Organisation, “Depression.” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[5] WHO, “Depression, a global public health concern,” WHO Department of Mental Health and Substance Abuse. [Online]. Available: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf

[6] Statista, “Number of internet and social media users worldwide as of April 2023,” https://www.statista.com/statistics/617136/digital-population-worldwide/.

[7] Meta, “Meta Earnings Presentation Q3 2023.” Accessed: Dec. 25, 2023. [Online]. Available: https://s21.q4cdn.com/399680738/files/doc_earnings/2023/q3/presentation/Earnings-Presentation-Q3-2023.pdf

[8] Semrush Inc, “X (Twitter) Statistics: How Many People Use X? (2024),” https://backlinko.com/twitter-users. Accessed: Dec. 25, 2023. [Online]. Available: https://backlinko.com/twitter-users

[9] A. Wongkoblap, M. A. Vadillo, and V. Curcin, “Researching Mental Health Disorders in the Era of Social Media: Systematic Review.,” *J Med Internet Res*, vol. 19, no. 6, p. e228, Jun. 2017, doi: 10.2196/jmir.7215.

[10] L. Deng and Y. Liu, Eds., *Deep Learning in Natural Language Processing*. Singapore: Springer Singapore, 2018. doi: 10.1007/978-981-10-5209-5.

[11] S. Sun, C. Luo, and J. Chen, “A review of natural language processing techniques for opinion mining systems,” *Information Fusion*, vol. 36, pp. 10–25, Jul. 2017, doi: 10.1016/j.inffus.2016.10.004.

[12] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, and S. Nutanong, “WangchanBERTa: Pretraining transformer-based Thai Language Models,” Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.09635

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 1, no. 4–5, pp. 993–1022, 2003.

[14] G. Coppersmith, M. Dredze, and C. Harman, “Quantifying Mental Health Signals in Twitter,” in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 51–60. doi: 10.3115/v1/W14-3207.

[15] A. Wongkoblap, M. A. Vadillo, and V. Curcin, “Deep Learning With Anaphora Resolution for the Detection of Tweeters With Depression: Algorithm Development and Validation Study,” *JMIR Ment Health*, vol. 8, no. 8, p. e19824, Aug. 2021, doi: 10.2196/19824.

[16] ก. แผ่วัฒนากุล and ป. ลือนาม, “การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์,” *Modern Management Journal*, vol. 11, no. 2, pp. 11–20, 2013, [Online]. Available: https://so04.tci-thaijo.org/index.php/stou-sms-pr/article/view/16934

[17] A. Hussain *et al.*, “Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study,” *J Med Internet Res*, vol. 23, no. 4, p. e26627, Apr. 2021, doi: 10.2196/26627.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[19] E. Massaad and P. Cherfan, “Social Media Data Analytics on Telehealth During the COVID-19 Pandemic,” *Cureus*, Apr. 2020, doi: 10.7759/cureus.7838.

[20] V. Grubmüller, K. Götsch, and B. Krieger, “Social media analytics for future oriented policy making,” *European Journal of Futures Research*, vol. 1, no. 1, p. 20, Dec. 2013, doi: 10.1007/s40309-013-0020-7.

[21] K. Kaleta and J. Mróz, “The Relationship between Basic Hope and Depression: Forgiveness as a Mediator,” *Psychiatric Quarterly*, vol. 91, no. 3, pp. 877–886, Sep. 2020, doi: 10.1007/s11126-020-09759-w.

[22] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing Depression from Twitter Activity,” *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15*, pp. 3187–3196, 2015, doi: 10.1145/2702123.2702280.

[23] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, “Early Detection of Depression: Social Network Analysis and Random Forest Techniques,” *J Med Internet Res*, vol. 21, no. 6, p. e12554, Jun. 2019, doi: 10.2196/12554.

[24] L. Ge, C. W. Yap, R. Ong, and B. H. Heng, “Social isolation, loneliness and their relationships with depressive symptoms: A population-based study,” *PLoS One*, vol. 12, no. 8, p. e0182145, Aug. 2017, doi: 10.1371/journal.pone.0182145.

[25] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. B. Silenzio, and M. De Choudhury, “A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: ACM, Jan. 2019, pp. 79–88. doi: 10.1145/3287560.3287587.

มหาวิทยาลัยเทคโนโลยีสุรนารี

ประวัตินักวิจัย

ชื่อ (ภาษาไทย) อาจารย์ ดร.อรรคพล วงศ์กอบลาภ

(ภาษาอังกฤษ) Dr.Akkapon Wongkoblap

ตำแหน่งปัจจุบัน อาจารย์

หน่วยงานที่อยู่ที่ติดต่อได้

สำนักวิชาศาสตร์และศิลป์ดิจิทัล

มหาวิทยาลัยเทคโนโลยีสุรนารี

E-mail: wongkoblap@sut.ac.th

ประวัติการศึกษา

PhD. (2564) Computer Science, King's College London, UK

MSc. (2558) Financial Systems Engineering, University College London, UK

วส.บ. (2553) ระบบสารสนเทศเพื่อการจัดการ มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานทางวิชาการ /ผลงานวิจัย

(1) C. Hmadchayyapum, T. Kopanklang, A. Wongkoblap, P. Iamkhajornchai, T. Angskun, and J. Angskun. The design and development of line chatbots for dental services booking. Suranaree Journal of Social Science, 19(2), 2025 [in press]

(2) S. Kobua, A. Wongkoblap, T. Angskun, and J. Angskun. A model for analyzing the severity level of adverse drug reactions using machine learning. Journal of Science and Technology Mahasarakham University, 44(1), 2025 [in press]

(3) K. Chanvised and A. Wongkoblap. Social media usage reflects depression among adoles- cent users. Journal of Journalism, 17(1):10, 2024

(4) A. Wongkoblap. Automatic profiles collection from twitter users with depressive symp- toms. In Healthcare Transformation with Informatics and Artificial Intelligence, pages 419–422, 2023

(5) Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2022). Social media big data analysis for mental health research. In Mental Health in a Digital World. pp. 109-143. Academic Press.

(6) Ford, K. C., Wongkoblap, A., & Curcin, V. (2019). Public Opinions on Using Social Media Content to Identify Users with Depression and Target Mental Health Care Advertising: Mixed Methods Survey. JMIR Mental Health. Vol. 6. No. 11.

(7) Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2018, June). A multilevel predictive model for detecting social network users with depression. In 2018 IEEE International Conference on Healthcare Informatics (ICHI). pp. 130-135. IEEE.

(8) Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2018). Classifying depressed users with multiple instances learning from social network data. In 2018 IEEE International Conference on Healthcare Informatics (ICHI). pp. 436-436. IEEE.

(9) Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017, August). Detecting and treating mental illness on social networks. In 2017 IEEE International Conference on Healthcare Informatics (ICHI). pp. 330-330. IEEE.

มหาวิทยาลัยเทคโนโลยีสุรนารี