



แบบเสนอโครงร่างปริญญานิพนธ์

ชื่อเรื่อง (ภาษาไทย) ระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ
ชื่อเรื่อง (ภาษาอังกฤษ) Depression risk prediction system using text mining techniques.

ผู้เสนอ

นางสาวกรรณิกา ปลั่งกลาง

นายสรวิชัย หนูมาศ

ภาคเรียนที่ 1/2566

ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

แบบเสนอโครงร่างปริญญานิพนธ์
ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม

1. ชื่อปริญญานิพนธ์

ภาษาไทย ระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ
ภาษาอังกฤษ Depression risk prediction system using text mining techniques.

2. ชื่อผู้เสนอ

2.1. นางสาวกรรณิกา ปลั่งกลาง รหัสประจำตัว 6502041510026

2.2. นายสรวิชัย หนูมาศ รหัสประจำตัว 6502041520129

3. หลักการและเหตุผล

ภาวะซึมเศร้า (Depression) เป็นภาวะที่มีความรู้สึกเศร้า หดหู่ ไม่มีความสุข ไม่มีกำลังใจ ในการกระทำการสิ่งใดๆ การแสดงออกด้านอาการและภาวะซึมเศร้านั้น พบว่ามักมีการแสดงออกใน ลักษณะที่สำคัญ คือ การแสดงออกด้านอารมณ์เป็นความรู้สึกสิ้นหวัง เศร้า กระวนกระวาย หงุดหงิด ง่ายไม่เป็นมิตร รวมถึงมีการร้องไห้บ่อยครั้ง มีความเบื่อหน่ายรู้สึกท้อแท้ เป็นต้น [1] ซึ่งโรคซึมเศร้านั้นก่อให้เกิดการสูญเสียด้านสุขภาพอย่างมีนัยสำคัญ และส่งผลกระทบต่อประชากรทั่วโลก องค์การอนามัยโลกได้ประมาณการว่ามีประชากรมากกว่า 280 ล้านคนเป็นโรคซึมเศร้า ซึ่งมากกว่าร้อยละ 4 ของประชากรโลก และมีถึง ร้อยละ 80 ที่ไม่ได้รับการรักษา [2] ในประเทศไทยอัตราการป่วยโรคซึมเศร้า ได้เพิ่มขึ้นอย่างต่อเนื่องในช่วงปี พ.ศ. 2552-2562 โดยในปี พ.ศ. 2552 พบผู้ป่วยโรคซึมเศร้า ในอัตรา 270 รายต่อแสนประชากร ในขณะที่ปี พ.ศ. 2562 พบผู้ป่วยโรคซึมเศร้าในอัตรา 345.1 รายต่อประชากรแสนคน [3] ซึ่งพฤติกรรมการใช้เครือข่ายสังคมออนไลน์ที่มากขึ้นนั้นสามารถส่งผลกระทบต่อประชากรแสนคน [3] ซึ่งพฤติกรรมการใช้เครือข่ายสังคมออนไลน์ที่มากขึ้นนั้นสามารถส่งผลกระทบต่อให้เกิดเป็นปัญหาหลากหลายด้าน ไม่ว่าจะเป็นด้านสุขภาพร่างกาย ด้านสุขภาพจิตที่เกิดจากการเปรียบเทียบทางสังคม การเห็นคุณค่าในตนเองต่อ ปัญหาด้านสุขภาพจิตโรคซึมเศร้า ทำร้ายตนเองและฆ่าตัวตาย เป็นต้น [4] จากข้อมูลดังกล่าวจะเห็นได้ว่าผู้ที่มีความเสี่ยงที่จะเกิดโรคซึมเศร้า มีจำนวนเพิ่มขึ้น และ ส่วนใหญ่จะไม่ได้รับการรักษาหรือไม่ทราบว่าตนเองเกิดภาวะซึมเศร้า ซึ่งนำไปสู่การเกิดโรคซึมเศร้า รวมถึงบุคคลในครอบครัวหรือผู้ที่ใกล้ชิดก็ไม่ทราบเช่นเดียวกัน ทำให้ไม่ได้รับการรักษาที่เหมาะสมและทันทั่วทั้งที่ดังนั้นผู้ที่มีความเสี่ยงที่จะเกิดโรคซึมเศร้าจึงแสดงออกผ่านทางเครือข่าย สังคมออนไลน์ [5]

การวิเคราะห์ด้วยเหมืองข้อความ (Text Mining) เป็นเทคนิคที่ได้รับความนิยมในการ วิเคราะห์ข้อมูลจำนวนมากเนื่องจาก 90% ของปริมาณข้อมูลมหาศาลที่อยู่บนอินเทอร์เน็ตเป็นข้อมูล ที่มีโครงสร้างไม่แน่นอน โดยการทำงานด้วยเหมืองข้อความ เป็นขบวนการ ที่จะช่วยสกัดข้อมูล จากฐานข้อมูลที่มีขนาดใหญ่ เพื่อให้ได้สารสนเทศที่มีประโยชน์โดยข้อมูลที่ ถูกนำมาทำเหมืองข้อความจะ

เป็นข้อมูลที่มีลักษณะเป็นชุดข้อมูล [6] โดยเน้นการทำงานด้านการจัดประเภทเอกสาร (Text Categorization) การจัดกลุ่มเอกสาร (Text Clustering) การสกัดความรู้ (Extraction) การวิเคราะห์ความรู้สึก (Sentiment Analysis) การสรุปเอกสาร (Document Summarization) และการหาความสัมพันธ์ (Entity Relation Modeling) ซึ่งการประมวลผลภาษาธรรมชาติถือเป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) โดยมุ่งศึกษาปัญหาและทำความเข้าใจ กับภาษามนุษย์ โดยใช้องค์ความรู้ด้านภาษา หลักไวยากรณ์ ภาษา และหลักสถิติ [7]

ดังนั้น ผู้วิจัยจึงมีแนวคิดในการพัฒนาระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ โดยที่ใช้ข้อมูลเครือข่ายสังคมออนไลน์เพื่อทำนายภาวะซึมเศร้า โดยใช้เทคโนโลยีดิจิทัลและเครื่องมือการวิเคราะห์ ข้อมูลที่ทันสมัยการวิเคราะห์ความรู้สึก (Sentiment Analysis) และเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งรวมถึงปัจจัยอื่น ๆ ที่เกี่ยวข้อง เพื่อให้ได้รับการวิเคราะห์ที่ครอบคลุมและเป็นประโยชน์ในการทำนายภาวะซึมเศร้า ด้วยการพัฒนาแบบจำลองในการวิเคราะห์ความเสี่ยงของภาวะซึมเศร้าจากข้อมูลเครือข่ายสังคมออนไลน์ มีความหวังว่า จะสามารถสร้างการรับรู้และการดูแลที่เหมาะสมสำหรับภาวะซึมเศร้าในประชาคมไทย โดยการปรับปรุงใน กระบวนการรับรู้และการดูแลภาวะซึมเศร้า มุ่งหวังว่างานวิจัยนี้จะเป็นการสนับสนุนที่เหมาะสมและการสร้าง ความเข้มแข็งสำหรับผู้ที่มีความเสี่ยงต่อภาวะซึมเศร้าในที่สุด

4. วัตถุประสงค์ของโครงการ

4.1 เพื่อพัฒนาระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความและประเมินผ่านหน้าเว็บไซต์

4.2 เพื่อประเมินประสิทธิภาพของวิธีการพยากรณ์ความเสี่ยงภาวะซึมเศร้า ด้วยเทคนิคเหมืองข้อความ

5. ขอบเขตของโครงการ

5.1 รวบรวมข้อมูลเกี่ยวกับภาวะซึมเศร้าบนเครือข่ายสังคมออนไลน์จาก Twitter ในประเทศไทย ประมาณ 20,000 ข้อความโดยใช้เว็บไซต์ Apify.

5.2 เตรียมข้อมูล (Pre Processing) การแก้ไขคำผิด สัญลักษณ์ความรู้สึกและใช้เทคนิคตัดคำ (Tokenization) ในเอกสารภาษาไทย

5.3 ตรวจสอบประโยคที่มีคำที่สามารถบ่งบอกถึงภาวะโรคซึมเศร้าและคำพูดที่ไม่เกี่ยวข้อง

5.4 กำหนดคลาส ให้กับข้อมูลที่เกี่ยวข้องกับภาวะซึมเศร้า ให้ผู้เชี่ยวชาญประเมินและนำผลการประเมินเพื่อแบ่งข้อมูลเป็นชุดการเรียนรู้ (Train), ชุดทดสอบ (Test), และชุดข้อมูลใหม่ (Unseen)

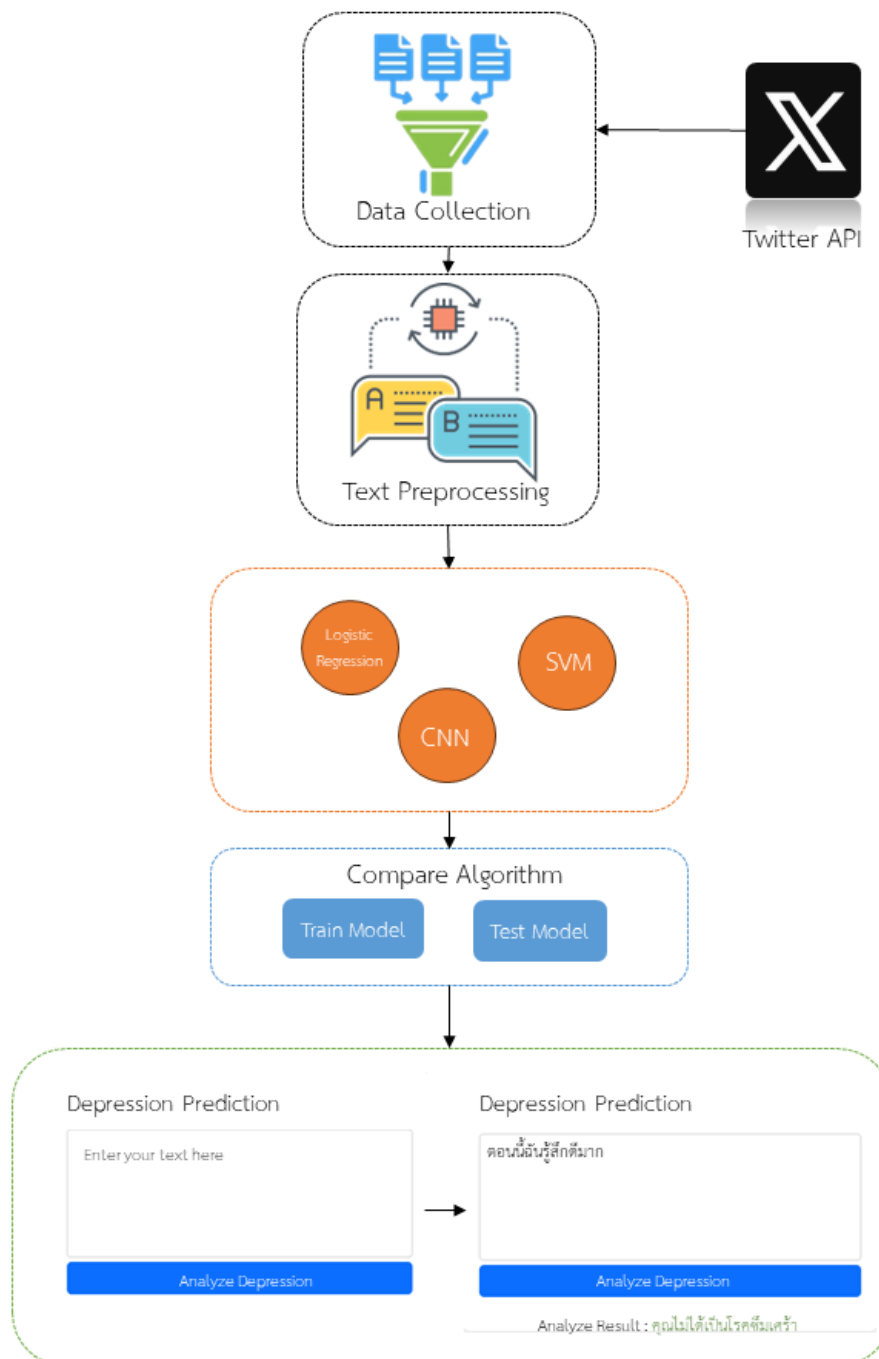
5.5 การทำเหมืองข้อมูลเพื่อจำแนกข้อมูลโดยใช้อัลกอริทึมต่าง ๆ เช่นการถดถอยเชิงเส้น (Logistic Regression), ซัพพอร์ตเวกเตอร์แมชชีน (SVM), และ deep learning model เช่นโครงข่ายประสาทเทียม (Convolutional Neural Network) เพื่อหาค่าความถูกต้องที่เหมาะสมที่สุด

5.6 เปรียบเทียบและประเมินประสิทธิภาพของชุดข้อมูลการเรียนรู้ (Train) และ ชุดทดสอบ (Test)

5.7 พัฒนาระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ

5.8 ประเมินประสิทธิภาพระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ

6. ขั้นตอนการดำเนินงาน



6.1 รวบรวมข้อมูล (Data Collection) ใช้การรวบรวมจาก Twitter API ซึ่งสามารถเก็บได้จากการใช้เว็บไซต์ Apify ที่สามารถผสานรวมกับบริการเว็บมากมายรวมถึง API จำนวนมาก และสามารถสืบค้นจากคำค้น (Keyword) ที่เกี่ยวข้องกับภาวะซึมเศร้า เช่น โรคซึมเศร้า ซึมเศร้า อยากตาย เสียใจ ผิดหวัง ฯลฯ หลังจากนั้นจึงเข้าสู่กระบวนการเตรียมข้อมูล

6.2 เตรียมข้อมูล (Pre Processing) ด้วยวิธี การทำความสะอาดข้อมูล (Data Cleansing) เนื่องจากข้อมูลที่ได้มาเป็น ข้อความ (Text) ที่จำเป็นต้องใช้กระบวนการเพื่อตัดข้อความหรืออักขระที่ไม่เกี่ยวข้องออก และ จัดการข้อมูลให้ เหมาะสมก่อนการทำการวิเคราะห์ความรู้สึกคือการจัดการคำผิดหลังจากทำความสะอาดข้อมูลแล้วจึงนำข้อมูลที่ได้มาผ่านกระบวนการประมวลผลภาษาธรรมชาติโดยใช้ กระบวนการตัดพยางค์ย่อย (Tokenization) เพื่อแยกข้อความในรูปประโยค (Sentence) ออกเป็นคำ (Word) หรือเรียกว่าการตัดคำ (Word Segmentation) การจัดการอีโมจิ (Emoticon) เพื่อ ใช้จำแนกอารมณ์และความรู้สึก โดยใช้การเปรียบเทียบข้อมูลจากฐานข้อมูล Emoji Sentiment Ranking ซึ่งฐานข้อมูล Emoji Sentiment Ranking จะมี ค่าคะแนน ความรู้สึก (Sentiment score) ของแต่ละอีโมจิอารมณ์เพื่อใช้ในการจำแนกอีโมจิ อารมณ์ ด้านบวก และ ด้านลบ

6.2.1 การถดถอยเชิงเส้น (Logistic Regression) การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression) เป็นเทคนิคสถิติที่ใช้ในการพยากรณ์ความน่าจะเป็นที่จะเกิดเหตุการณ์หรือไม่เกิดเหตุการณ์ของเหตุการณ์หนึ่งๆที่สนใจ จะใช้เมื่อตัวแปรตาม (Dependent Variable) เป็นตัวแปรเชิงคุณภาพ ส่วนตัวแปรอิสระ (Independent Variable) หรือตัวแปรพยากรณ์ (Predictor) เป็นตัวแปรที่สนใจหรือเป็นปัจจัยที่สามารถพยากรณ์การเกิดเหตุการณ์หนึ่งที่สนใจหรือเป็นปัจจัยที่มีอิทธิพลกับการเกิดเหตุการณ์หนึ่งที่สนใจได้ ซึ่งเป็นได้ทั้งตัวแปรเชิงปริมาณ และตัวแปรเชิงคุณภาพ สามารถมีได้มากกว่า 1 ตัวแปร

6.1.2 ซัพพอร์ตเวกเตอร์แมชชีน หรือ เอสวีเอ็ม (Support Vector Machine : SVM) จัดเป็นการเรียนรู้ของเครื่องประเภทแบบการเรียนรู้โดยอาศัยตัวอย่างประเภทหนึ่ง ซึ่งมีความสามารถในการจัดหมวดหมู่ และการทำนาย (Regression) โดยเอสวีเอ็มมีพื้นฐานจะมีการคำนวณแบบเชิงเส้น (Linear) ซึ่งจัดอยู่ในประเภทมุ่งหาผลลัพธ์ที่ดีที่สุดของการเรียนรู้ (Discriminative Training) บนการเรียนรู้จากสถิติของข้อมูล ซึ่งทำงานโดยการหาค่าระยะขอบที่มากที่สุด (Maximum Margin) ของระนาบตัดสินใจ (Decision Hyperplane)

6.1.3 การสร้างขั้นแรกคือการพัฒนา deep neural network โดยใช้ภาษาโปรแกรมที่เหมาะสม เช่น Python และส่วนรูปแบบของโครงข่ายนิวรอลเน็ต (Neural Network) ขึ้นมา โครงข่ายนี้ประกอบด้วยชั้นข้อมูล (layers) หลายชั้นที่สามารถเรียนรู้และถอดรหัสข้อมูลจากข้อมูลนำเข้า (input data) และสร้างข้อมูลผลลัพธ์ (output data) ในรูปแบบที่ถูกต้อง โดยอาจจำลองประสิทธิภาพของสมองมนุษย์ในการจดจำและค้นพบลักษณะที่อยู่ในข้อมูล เพื่อให้ระบบสามารถรู้จัก

แพทเทิร์นและคุณสมบัติต่างๆ ในข้อมูลอย่างมีประสิทธิภาพและแม่นยำ หลังจากสร้าง deep neural network แล้วเราสามารถเพิ่มขึ้น CNN (Convolutional Neural Network) เพื่อเพิ่มความสามารถในการสกัดลักษณะเฉพาะต่างๆ (features) ของข้อมูลนำเข้า เช่น เส้นขอบของวัตถุและลักษณะเด่นของรูปภาพ เป็นต้น โดยการใช้ CNN ช่วยให้โครงข่ายสามารถเรียนรู้และจดจำลักษณะเหล่านี้ได้อย่างแม่นยำ การลง CNN เป็นหนึ่งในวิธีที่มีประสิทธิภาพในการปรับปรุงการประมวลผลข้อมูลเชิงภาพเช่นภาพถ่ายหรือภาพวิดีโอ ทำให้โครงข่ายสามารถรู้จักวัตถุและภาพได้อย่างมีประสิทธิภาพและแม่นยำขึ้น

6.3 เปรียบเทียบอัลกอริทึม (Compare Algorithm) ทำการเลือกอัลกอริทึมมาทำการทดสอบกับข้อมูลที่ผ่านมาเตรียมข้อมูลและเลือกอัลกอริทึมที่ดีที่สุด ออกเป็น 2 โมเดลดังนี้

6.3.1 โมเดลการเรียนรู้ (Train Model) ถูกฝึกโดยใช้ข้อมูลที่ได้รับการเตรียม โดยทำการเปรียบเทียบอัลกอริทึม 3 อัลกอโดยเปรียบเทียบค่าความถูกต้องที่ดีที่สุดเพื่อนำไปใช้พัฒนาระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ

6.3.2 ทดสอบโมเดล (Test Model) หลังจากโมเดลการเรียนรู้ ทำการทดสอบโมเดลด้วยชุดข้อมูลชุดทดสอบ เพื่อตรวจสอบค่าความถูกต้องของการเลือกใช้อัลกอริทึมและการประยุกต์ใช้กับข้อมูลที่ไม่ผ่านการเรียนรู้

6.4 พัฒนาระบบการทำนายความเสี่ยงภาวะซึมเศร้าด้วยเทคนิคการทำเหมืองข้อความ เพื่อให้ผู้ใช้สามารถใช้งานได้ง่ายขึ้น โดยระบบสามารถทำนายภาวะซึมเศร้าจากข้อความใหม่และสามารถบอกระดับความรุนแรงของผู้ป่วยหรือคาดการณ์ภาวะซึมเศร้าของตน
ประเมินค่า ความถูกต้อง Precision recall

7. ระยะเวลาการทำโครงการ

ขั้นตอนการดำเนินงาน	พ.ศ.2566							พ.ศ.2567				
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
1.ศึกษาและค้นคว้าข้อมูล												
2.เขียนโปรแกรม												
3.สร้างโมเดล												
4.พัฒนา												
5.ตรวจสอบความถูกต้อง												
6.ทดลองใช้												
7.สรุปผล จัดทำเอกสาร												

8. ประโยชน์ที่คาดว่าจะได้รับ

- 8.1 ระบบที่พัฒนาขึ้นช่วยในการระบุความเสี่ยงในการเกิดภาวะซึมเศร้าของผู้ใช้งานได้
- 8.2 แบบจำลองการทำนายความเสี่ยงภาวะซึมเศร้าสามารถช่วยในการตรวจสอบสุขภาพจิตของผู้ใช้งาน และช่วยในการคัดกรองผู้ที่มีความเสี่ยงในการพบภาวะซึมเศร้าได้

9. เอกสารอ้างอิง

- [1] วราวุฒิ นาคบุญนำ. (2564). การพยากรณ์ภาวะซึมเศร้าด้วยวิธีการเหมืองข้อมูล. วิทยานิพนธ์ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล พุทธศักราช .
- [2] WHO. (2018). Depression. [ออนไลน์]. [สืบค้นวันที่ 16 ตุลาคม 2566].
จาก <http://www.who.int/en/news-room/factsheets/detail/depression>.
- [3] สำนักงานพัฒนาระบบข้อมูลข่าวสารสุขภาพ. (2564). ข้อมูลข่าวสารโรคซึมเศร้า. [ออนไลน์]. [สืบค้นวันที่ 16 ตุลาคม 2566].
จาก.http://social.nesdb.go.th/SocialStat/StatReport_Final.aspx?reportid=367&template=1R2C&yeartype=M&subcatid=15.
- [4] ครองรัตน์ ดุลลาพันธ์.(2561). พฤติกรรมการใช้เครือข่ายสังคมออนไลน์และปัจจัยทางจิตที่พยากรณ์ภาวะโรคซึมเศร้า. วิทยานิพนธ์ศิลปศาสตรมหาบัณฑิต (นิเทศศาสตร์และนวัตกรรม) คณะนิเทศศาสตร์และนวัตกรรมการจัดการสถาบันบัณฑิตพัฒน

บริหารศาสตร์.

- [5] อุไรวรรณ,ปวีณา,อุษามณ,คงเทพ. (2563). การศึกษาวิธีการจำแนกประเภทอาการภาวะซึมเศร้าโดยเทคนิคการทำเหมืองข้อมูล. วารสารวิทยาศาสตร์และเทคโนโลยี นานา เรืออากาศ.สาขาวิชาวิทยาการคอมพิวเตอร์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี.
- [6] วทัญญู ประจติต. (2563).การจำแนกประเภทและวิธี การคัดเลือกคุณลักษณะสำหรับการสนับสนุนการวิเคราะห์ ความรู้สื่อต่อสถานที่ท่องเที่ยวไทยพยากรณ์ภาวะโรคซึมเศร้า. วิทยานิพนธ์ นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยทักษิณ.
- [7] วรณวิภา วงศ์วิไลสกุล. (2556).เหมืองข้อความและการประยุกต์ใช้. วารสารปัญญาภิวัฒน์ ปีที่ 4 ฉบับพิเศษ ประจำเดือนพฤษภาคม. สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์.

10.ความคิดเห็นของที่ปรึกษา

.....

.....

.....

ลงชื่อ



(ผศ. ดร. วาทีนี น้อยเพียร)

ที่ปรึกษาปริญญานิพนธ์

วันที่ เดือน พ.ศ. 2565

.....

.....

.....

ลงชื่อ

()

ที่ปรึกษาปริญญานิพนธ์ร่วม

วันที่ เดือน พ.ศ. 2565

หมายเหตุ โครงการนี้เสนอโดย ☐ ผู้จัดทำโครงการ ☐ อาจารย์ที่ปรึกษา/ที่ปรึกษาร่วม