

ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษา ด้วยเทคนิคเหมืองข้อมูล

Model for Assessing the Risk of Depression of Students with Data Mining Techniques.

เพชรรัตน์ ม่วงน้อย^{1*}, จักรพันธ์ พลาผล² และ ภรณ์ยา ปาลวิสุทธิ์³

Petcharat Maungnoi^{1*}, Jakkapan Palaphol² and Paranya Palwisut³

สาขาวิชาเทคโนโลยีสารสนเทศ^{1,2} คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม

สาขาวิชาวิทยาการข้อมูล³ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม

Department of Information Technology^{1,2}, Faculty of Science and Technology, Nakhon Pathom Rajabhat University

Department of Data Science³, Faculty of Science and Technology, Nakhon Pathom Rajabhat University

E-Mail: 604230015@webmail.npru.ac.th^{1*}, 604230006@webmail.npru.ac.th², paranya@npru.ac.th³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาและพัฒนาตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคการเหมืองข้อมูล กลุ่มตัวอย่างที่ใช้ในการวิจัย ได้แก่ นักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏนครปฐม จำนวน 600 คน โดยสุ่มตัวอย่างจากทุกคณะ เครื่องมือที่ใช้ในการวิจัย ได้แก่ 1) แบบสอบถามข้อมูลทั่วไป และ 2) แบบประเมินแบบทดสอบภาวะซึมเศร้า PHQ-9 เทคนิคการวิเคราะห์ข้อมูลที่ใช้ในการวิจัย คือ เทคนิคเหมืองข้อมูล จำนวน 5 อัลกอริทึม ได้แก่ Random, Tree, LMT, PART J48, and JRIP และทำการเปรียบเทียบประสิทธิภาพด้วยการทดสอบ Cross-validation

ผลการวิจัยพบว่า Random Tree มีประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 96.00% รองลงมาคือ อัลกอริทึม LMT มีค่า 95.90% อัลกอริทึม PART มีค่า 95.10% อัลกอริทึม J48 มีค่า 94.80% และอัลกอริทึม JRIP มีค่า 94.00% ตามลำดับ

คำสำคัญ: โรคซึมเศร้า, เหมืองข้อมูล, ต้นไม้ตัดสินใจ

ABSTRACT

The purposes of this research are 1) to study and develop a risk techniques model for depression among students that may cause self-harm using data mining and the sample used in the research is 600 undergraduate students, Nakhon Pathom Rajabhat University randomly sampling from all faculties. The research instruments are 1) general information questionnaires and 2) PHQ-9 depression test assessment. The data analysis techniques used in this research consist of five data mining techniques: Random Tree, LMT, PART J48, and JRIP, the efficiency is compared by a Cross-validation test.

The results show that the data using the Random Tree is the most efficient with an accuracy of 96.00%, followed by the LMT with an accuracy of 95.90%, PART with an accuracy of 95.10%, J48 with an accuracy of 94.80%, and JRIP with an accuracy of 94.00% respectively.

Keywords: Depression, Data Mining, Decision Trees

บทนำ

ภาวะซึมเศร้า (Depression) หรือที่รู้จักกันว่าโรคซึมเศร้า เป็นภาวะความผิดปกติทางอารมณ์ที่ผู้ป่วยอาจรู้สึกเศร้าสิ้นหวัง ช่วยเหลือตัวเองไม่ได้ หรือรู้สึกท้อแท้ด้อยค่า หดหู่ วิตกกังวล มีความรู้สึกท้อแท้ไร้ค่า โดดเดี่ยว หงุดหงิดง่าย แม้ความรู้สึกและอารมณ์เหล่านี้จะเกิดขึ้นได้กับทุกคนเป็นครั้งคราว [1] แต่อาการของภาวะซึมเศร้านั้นมีความรุนแรงและยาวนานกว่ามาก จนถึงขั้นส่งผลกระทบต่อการใช้ชีวิตในด้านต่าง ๆ โรคซึมเศร้าเกิดขึ้นได้กับคนทุกเพศทุกวัย แต่โดยมากมักเริ่มตั้งแต่ช่วงอายุ 20-30 ปี โรคความผิดปกติทางอารมณ์ส่วนใหญ่จะเริ่มพัฒนามาจากช่วงวัยรุ่นที่มีความเครียดและความวิตกกังวลสูง ยิ่งในช่วงวัยรุ่นประสบกับความกังวลมากเท่าไรก็ยิ่งมีความเสี่ยงต่อการเกิดโรคซึมเศร้าในวัยผู้ใหญ่มากขึ้นเท่านั้นด้วย [2]

โรคซึมเศร้าอาจเกิดได้จากหลายสาเหตุและปัจจัยประกอบกันจนพัฒนาไปเป็นภาวะซึมเศร้า ได้แก่ การทำงานของสมองบางส่วนที่ผิดปกติ และมีระดับของสารเคมีไม่สมดุลกัน ลักษณะนิสัยเดิมของผู้ป่วยที่เอื้อต่อการเผชิญภาวะซึมเศร้าเหตุการณ์ตึงเครียดในชีวิตและการกระทบกระเทือนทางจิตใจอย่างรุนแรงรวมถึงอาการป่วยจากโรคและการใช้ยารักษาโรคที่ตามมาด้วยภาวะแทรกซ้อน หรือผลข้างเคียงในลักษณะอาการซึมเศร้า จากสถิติทั่วโลกพบว่า มีผู้ป่วยโรคซึมเศร้า ประมาณ 350 ล้านคน มีความชุกราว 2-10% และเป็นในเพศหญิงมากกว่าเพศชาย ส่วนในไทยมีจำนวนผู้ป่วยซึมเศร้าเพศหญิงเป็นอันดับที่ 3 และเพศชายเป็นอันดับที่ 8 ในปัจจุบันพบได้มากในหมู่ วัยรุ่นและวัยทำงาน โดยมักเริ่มตั้งแต่อายุ 20-30 ปี โดยสถิติการฆ่าตัวตายในประเทศไทยปี 2562 เฉลี่ย 14.4% คนต่อคนไทย 1 แสนคน มากสุดอันดับ 32 ของโลก ซึ่งปัจจัยที่มีความสัมพันธ์กับการฆ่าตัวตาย ในปี 2561-2562 พบเป็นปัญหาเรื่องความสัมพันธ์ 53.04%, ปัญหาจากสุรา 29%, โรคทางกาย 25.7%, ปัญหาด้านเศรษฐกิจ 19%, โรคทางจิต 12% และโรคซึมเศร้า 7.8% และพบว่า ผู้ที่พยายามฆ่าตัวตายจะไม่กลับมาทำร้ายตัวเองซ้ำใน 1 ปี ถึง 94.27% [3]

ดังนั้นงานวิจัยนี้จะศึกษาและจัดทำตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาที่ส่งผลต่อการทำร้ายตัวเอง ด้วยเทคนิคเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision tree) เพื่อให้ได้เครื่องมือช่วยในการทำนายความเสี่ยงในการเป็นโรคซึมเศร้าและแหล่งเป็นความรู้แก่ผู้สนใจในเรื่องนี้

1. วัตถุประสงค์การวิจัย

1. เพื่อศึกษาและพัฒนาตัวแบบประเมินความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันนั้นการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล [4]

2.2 CRISP-DM (Cross-Industry Standard Process for Data Mining) เป็นกระบวนการ มาตรฐานในการวิเคราะห์ข้อมูลด้านเทคนิคเหมืองข้อมูล พัฒนาโดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอน ดังนี้ [5]

2.2.1 การทำความเข้าใจในเรื่องปัญหาที่จะทำ (Business Understanding) เป็นขั้นตอนแรกสุดในกระบวนการ CRISP-DM เป็นการทำความเข้าใจระบุปัญหาหรือโอกาสเชิงธุรกิจ ระบุ output ที่ต้องการได้จากการวิเคราะห์ข้อมูลด้วยเหมืองข้อมูลและวางแผนงาน

2.2.2 การทำความเข้าใจกับข้อมูลที่จัดเก็บ (Data Understanding) ข้อมูลเป็นปัจจัยที่สำคัญที่สุดที่ขาดไม่ได้ในการทำเหมืองข้อมูล ในขั้นตอนนี้เป็นการรวบรวมข้อมูลที่เกี่ยวข้องข้อมูลถูกต้องน่าเชื่อถือ ข้อมูลที่ได้มีปริมาณมากพอหรือยังและข้อมูลที่ได้มีความเหมาะสม มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

2.2.3 การเตรียมข้อมูล (Data Preparation) ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลานานที่สุดเนื่องจากโมเดลที่ได้จากการทำเหมืองข้อมูล จะให้ผลลัพธ์ที่ถูกต้องหรือไม่ขึ้นขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้โดยการเตรียมข้อมูลนั้น สามารถแบ่งออกได้เป็น 3 ขั้นตอนย่อย คือ Data Cleaning, Data Selection และ Data Transformation

2.2.4 การสร้างโมเดล (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล เช่น เทคนิค Association เทคนิคการทำ Clustering และเทคนิคการทำ Classification

2.2.5 การประเมินผล (Evaluation) การประเมินหรือวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้น เป็นเพียงการวัดความน่าเชื่อถือของโมเดลเท่านั้น ในขั้นตอนนี้เป็นการประเมินประสิทธิภาพของผลลัพธ์จากโมเดลวิเคราะห์ข้อมูลว่าครอบคลุมและสามารถตอบโจทย์ทางธุรกิจที่ตั้งไว้ในขั้นตอนแรกหรือไม่

2.2.6 การนำไปใช้ (Deployment) การนำผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลนำไปใช้งานจริง

2.3 กฎการจำแนก (Classification Rules) เป็นกระบวนการในการจัดแบ่งข้อมูล ขึ้นอยู่กับลักษณะของวัตถุประสงค์นั้น ๆ ส่วนการจำแนกในด้านการทำเหมืองข้อมูล เป็นการวิเคราะห์เซตของกลุ่มข้อมูล (Data object) ที่ยังไม่จัดแบ่งประเภท เพื่อสร้างโมเดลหรือฟังก์ชันออกเป็นชุดข้อมูล (Class) ซึ่งลักษณะของคลาสถูกอธิบายโดยกลุ่มของคุณสมบัติ (Attribute) และกลุ่มของข้อมูล (Training data set) ที่ใช้ในการสร้างโมเดลกฎการจำแนก [4]

2.4 ต้นไม้ตัดสินใจ (Decision Tree) เป็นวิธีหนึ่งที่สำคัญในการจำแนกกฎ โดยจะมีลักษณะเป็นการทำงานเหมือนโครงสร้างต้นไม้ ที่แต่ละโหนด (Node) แสดงคุณลักษณะ Attribute ที่ใช้ทดสอบข้อมูลแต่ละกิ่งแสดงผลในการทดสอบและลิฟโหนด (Leaf Node) แสดงกลุ่มหรือคลาส (Class) ที่กำหนดไว้ ซึ่งต้นไม้ตัดสินใจนี้ง่ายต่อการเข้าใจและการปรับเปลี่ยนเป็นกฎการจำแนก Classification Rules อัลกอริทึมพื้นฐานของการสร้างต้นไม้ตัดสินใจ คือ อัลกอริทึมละโมภ (Greedy Algorithm) โดยจะสร้างต้นไม้จากบนลงล่างแบบวนซ้ำ (Recursive) ด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย (Divide-and-Conquer) ซึ่งรูปแบบของต้นไม้จะประกอบด้วยโหนดแรกสุดที่เรียกว่า Root Node จาก Root Node ก็จะแตกออกเป็นโหนดลูก และที่โหนดลูกก็จะมีลูกของตัวเอง ซึ่งโหนดในระดับสุดท้ายจะเรียกว่า Leaf Node [5]

Gini Index ค่าที่บ่งบอกว่าคุณลักษณะหรือปัจจัยใดควรนำมาใช้เป็นคุณลักษณะในการแบ่งกลุ่มของอัลกอริทึม

(1)

$$\text{Gini}(ti) = 1 - \sum Ni=1 [p(ti)]^2$$

Entropy ค่าคาดคะเนของข้อมูลเป็นค่าที่แยกโดยใช้ลักษณะประจำของอัลกอริทึม IDF3

(2)

$$\text{Entropy}(ti) = 1 - \sum Ni=1 [p(ti)] \log_2 p(ti)$$

โดยที่

ti คือ คุณลักษณะที่นำมาวัดค่า Entropy

P (ti) คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม i กับจำนวนสมาชิก ทั้งหมดของกลุ่มตัวอย่างซึ่งแต่ละอัลกอริทึมจะให้ผลของโครงสร้างต้นไม้ตัดสินใจที่แตกต่างกันไป

- 1) True Positive (TP) เป็นจำนวนข้อมูลที่มีประเภที่เป็นบวกและถูกตัวแบบจำแนกประเภทว่าเป็นบวก
- 2) False Positive (FP) เป็นจำนวนข้อมูลที่มีประเภที่เป็นลบ แต่ถูกตัวแบบจำแนกประเภทว่าเป็นบวก
- 3) True Negative (TN) เป็นจำนวนข้อมูลที่มีประเภที่เป็นลบ และถูกตัวแบบจำแนกประเภทว่าเป็นลบ

4) False Negative (FN) เป็นจำนวนข้อมูลที่มีประเภทเป็นบวก แต่ถูกตัวแบบจำแนกประเภทว่าเป็นลบ

ตารางที่ 1 จำนวนของข้อมูลในแต่ละประเภท

Actual Class	Predicted Class		
		Class = Yes	Class = No
	Class = Yes	A (TP)	B (FN)
	Class = No	C (FP)	D (TN)

2.5 ตัวชี้วัดที่สามารถคำนวณได้จาก Confusion Matrix [6]

1. ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส

(3)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

2. ค่าความระลึก (Recall) เป็นการวัดความระลึกของโมเดล โดยพิจารณาแยกทีละคลาส

(4)

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

3. ค่าความถ่วงดุล (F-measure) เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส

(5)

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. ค่าความถูกต้อง (Accuracy) เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส คือ จำนวน True Positive ของทุกคลาสรวมกัน

(6)

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

การวัดประสิทธิภาพของโมเดล การแบ่งข้อมูลเพื่อนำไปทดสอบประสิทธิภาพของโมเดลมี 3 วิธีการ ดังนี้

วิธี Self-Consistency Test หรือเรียกว่า Use Training set วิธี Split Test และวิธี Cross-Validation Test โดยในงานวิจัยนี้เลือกวิธี Cross-Validation Test เป็นวิธีที่นิยมใช้ในการทดสอบประสิทธิภาพของโมเดล เนื่องจากผลที่ได้มีความน่าเชื่อถือออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลส่วนหนึ่งจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดลโดยทวนไปจนครบจำนวนที่แบ่งไว้

2.6 Synthetic Minority Over-sampling Technique: SMOTE เป็นเทคนิคที่ใช้ในการแก้ปัญหาที่ต้องการจำแนกข้อมูลไม่สมดุล ซึ่งข้อมูลมีจำนวนตัวอย่างแตกต่างกันมากในแต่ละคลาส เมื่อทำการจำแนกประเภทจะทำให้การเรียนรู้แต่ข้อมูลกลุ่มที่มาก ผลที่ได้ก็จะจำแนกไปในข้อมูลกลุ่มมากวิธี SMOTE เป็นวิธีการเพิ่มจำนวนข้อมูลประเภทที่มีข้อมูลน้อย ให้เพิ่มปริมาณข้อมูลใกล้เคียงกับประเภทที่มีมากที่สุดโดยสุ่มค่าขึ้นมาจากค่าและหาค่า

ระยะห่างระหว่างค่าที่เลือกกับทุก ๆ ค่าเลือกค่าที่ใกล้เคียงที่สุด เช่น กำหนดไว้ 5 ค่าสุ่มค่าจากที่เลือก 1 ใน 5 ค่าอยู่ระหว่างค่าที่เลือกตอนแรกและค่าที่สุ่มมาตอนหลัง [7]

2.7 อัลกอริทึมที่ใช้ในงานวิจัย การเปรียบเทียบหาเทคนิคและตัวแบบที่สามารถพยากรณ์ได้แม่นยำและมีประสิทธิภาพสูงที่สุดโดยวัดจากค่าความถูกต้องของแต่ละตัวแบบ ซึ่งเทคนิคต้นไม้ตัดสินใจที่นำมาใช้เปรียบเทียบประสิทธิภาพในงานวิจัย การเลือกอัลกอริทึมที่ผู้วิจัยได้ทำการศึกษาพบว่าเป็นที่นิยมใช้อย่างแพร่หลาย แล้วเหมาะกับข้อมูลที่มีเป็นแบบ ordinal หรือ nominal ชุดข้อมูลในงานวิจัยเป็นประเภทนี้ และข้อมูลที่เราเอามาใช้พิจารณาจากการที่ review paper หรือข้อมูลที่เป็นข้อมูลพื้นฐานของนักศึกษา โดยอัลกอริทึมทั้ง 5 มีรายละเอียดดังนี้ [8]

2.7.1 อัลกอริทึม JRIP คือ กระบวนการเรียนรู้กฎเชิงประพจน์ โดยมีการตัดแต่งเพิ่มส่วนแบบทำซ้ำเพื่อสร้างการลดข้อผิดพลาด

2.7.2 อัลกอริทึม LMT คือ การจำแนกรูปแบบที่มีการเชื่อมโยงและผสมผสานการถดถอยโลจิสติกส์ (LR) และการตัดสินใจการเรียนรู้ในการตัดสินใจของต้นไม้

2.7.3 อัลกอริทึม PART เป็นกระบวนการแยกและพิชิตข้อมูล สร้างแผนผังการตัดสินใจการวนซ้ำในแต่ละครั้ง บางส่วนเพื่อให้ได้ Leaf Node ที่ดีที่สุด

2.7.4 อัลกอริทึม RT หรือ Random Tree มีการพิจารณาแอตทริบิวต์ที่สุ่มเลือกในแต่ละโหนด ไม่ทำการตัดแต่งกิ่ง นอกจากนี้ยังมีตัวเลือกในการอนุญาตการประมาณค่าความน่าจะเป็นของคลาส (หรือค่าเฉลี่ยเป้าหมายในการถดถอย) ตามชุดการระงับ (backfitting) ตัวเลือกที่ถูกต้อง

2.7.5 อัลกอริทึม J48 หรืออัลกอริทึมของ C4.5 เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจจากกลุ่มของข้อมูลฝึกสอน โดยใช้ความถูกต้องของแต่ละคุณลักษณะของข้อมูล เพื่อใช้ในการตัดสินใจแบ่งกลุ่มข้อมูลกลุ่มย่อย ๆ โดยพิจารณาจากค่าความแตกต่างใน Entropy ผลลัพธ์จากการเลือกคุณลักษณะสำหรับแบ่งกลุ่มข้อมูลด้วยค่า Normalized information gain ที่สูงที่สุดนั้นคือ การสร้างการตัดสินใจ

2.8 งานวิจัยที่เกี่ยวข้อง

ปณิธา บุญพาณิชย์ รัศมีน กัลยาศิริ, และณภัทรวรรต บัวทอง [9] ศึกษาเรื่องภาวะซึมเศร้าและความเหนื่อยล้าของพนักงานต้อนรับบนเครื่องบินในประเทศไทย ผลที่ได้จากการวิจัยครั้งนี้พบว่า กลุ่มตัวอย่างมีภาวะสุขภาพจิตด้านซึมเศร้าอยู่ในระดับเล็กน้อย (ร้อยละ 16.30%) มีความเหนื่อยล้าอยู่ในระดับปานกลาง (ร้อยละ 60.80%)

ชนิษฐา ชัดิยะ [10] ศึกษาเรื่องพยากรณ์ความเสี่ยงของโรคซึมเศร้า โดยใช้เทคนิคเหมืองข้อมูล ผลที่ได้จากการวิจัยครั้งนี้พบว่า ตัวแบบพยากรณ์การสร้างโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ จากการที่ได้นำเสนอการนำเทคนิคตัวแบบพยากรณ์ด้วยเทคนิค J48 จะเป็นอัลกอริทึมที่เหมาะสมมาใช้ในการงานวิจัยการพัฒนาตัวแบบพยากรณ์ความเสี่ยงต่อการเป็นโรคซึมเศร้า โดยที่มีค่าความถูกต้องในการจำแนกข้อมูล 84.00% ซึ่งจะเหมาะสมกับงานวิจัยนี้

สุทธนันท์ ชุนแจ่ม, โสภณ แสงอ่อน, และทัศนาว ทวีคุณ [3] ศึกษาเรื่องการสำรวจงานวิจัยเกี่ยวกับภาวะซึมเศร้าในประเทศไทย ผลที่ได้จากการวิจัยครั้งนี้พบว่า งานวิจัยเกี่ยวกับภาวะซึมเศร้าในผู้ใหญ่และผู้สูงอายุในประเทศไทย เป็นการศึกษาในผู้ป่วยทางกายเป็นส่วนใหญ่ ดังนั้นจึงควรสนับสนุนให้มีการศึกษาในผู้ป่วยทางจิตหรือในกลุ่มที่เสี่ยงต่อการเกิดภาวะซึมเศร้า เช่น กลุ่มผู้ดูแลกลุ่มคนทำงานให้เพิ่มมากขึ้นและสนับสนุนให้มีการวิจัยเกี่ยวกับภาวะซึมเศร้าในด้านการส่งเสริมป้องกันและฟื้นฟูให้มากขึ้น

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 แบบสอบถามข้อมูลทั่วไปและแบบทดสอบภาวะซึมเศร้า PHQ-9

2. กลุ่มเป้าหมาย

2.1 นักศึกษามหาวิทยาลัยราชภัฏนครปฐมระดับปริญญาตรี ที่กำลังศึกษาอยู่ในปีการศึกษา 2563 โดยใช้วิธีสุ่มกลุ่มแบบการเลือกแบบเจาะจง (Purposive Sampling) จากทุกคณะ จำนวน 600 คน จากจำนวนนักศึกษาทั้งหมด 18,682 คน

3. ขั้นตอนการดำเนินการวิจัย

3.1 การศึกษาเบื้องต้นการทำความเข้าใจในเรื่องปัญหาที่จะทำ ผู้วิจัยศึกษาและทบทวนผลงานวิจัยที่เกี่ยวข้องกับงานวิจัยข้างต้น เพื่อเป็นแนวทางในการดำเนินงานวิจัย ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ โดยปัจจัยที่เกี่ยวข้องที่นำมาวิเคราะห์ในงานวิจัยนี้คือ ลักษณะการใช้ชีวิตประจำวัน สถานะทางสังคมที่พบเจอ และช่วงวัยอายุของผู้ที่เป็นโรคซึมเศร้า ซึ่งในปัจจุบันพบได้มากในหมู่วัยรุ่นและวัยทำงานโดยมักเริ่มตั้งแต่อายุ 20-30 ปี หรือวัยศึกษาระดับอุดมศึกษา และศึกษาเครื่องมือหรือซอฟต์แวร์ที่นำมาใช้ในการพัฒนาการสร้างโมเดลและทำการทดสอบอัลกอริทึม โดยซอฟต์แวร์ที่เลือกใช้คือ โปรแกรม Weka [6] เป็นโปรแกรมที่สามารถนำมาทดสอบอัลกอริทึมของตนไม่ตัดสินใจได้และสามารถแสดงค่าความถูกต้อง

3.2 การทำความเข้าใจกับข้อมูลที่จัดเก็บ การศึกษาข้อมูลผู้วิจัยได้ดำเนินการศึกษาข้อมูลจริง รวมถึงการเก็บข้อมูลต่าง ๆ ที่มีความสัมพันธ์เกี่ยวข้องกับงานวิจัยในครั้งนี้ เพื่อใช้ในการสร้างโมเดลและทดสอบอัลกอริทึมมีดังต่อไปนี้

3.2.1 การกำหนดกลุ่มตัวอย่าง คือ นักศึกษาทั้งชายและหญิงที่กำลังศึกษาอยู่ในมหาวิทยาลัยราชภัฏนครปฐม โดยสุ่มกลุ่มตัวอย่างที่จะเก็บข้อมูลจากแต่ละคณะ จำนวน 5 คณะ

3.2.2 การเก็บรวบรวมข้อมูล ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถามจาก Google form โดยการแจกแบบสอบถามให้กลุ่มตัวอย่างทำ จำนวน 600 คน

3.2.3 นำข้อมูลที่ได้จากแบบสอบถามที่พัฒนาโดย คณะแพทยศาสตร์ โรงพยาบาลรามธิบดี [11] จำนวน 600 ชุด นำมาจัดเก็บลงในฐานข้อมูลเพื่อทำการเตรียมนำเข้าโปรแกรม Weka ดำเนินการสร้างโมเดลและทดสอบหาอัลกอริทึมที่เหมาะสมมากที่สุด

3.3 การเตรียมข้อมูลสำหรับทำเหมืองข้อมูล จากที่ได้ศึกษาและเก็บรวบรวมข้อมูลจากขั้นตอนก่อนหน้านี้ ในขั้นตอนนี้จะต้องมีการจัดเตรียมข้อมูลและการคัดเลือกข้อมูลที่จะนำมาใช้งาน เนื่องจากข้อมูลที่ได้มานั้นยังไม่สามารถนำมาใช้งานได้ ต้องผ่านกระบวนการทดสอบและวิเคราะห์ด้วยเครื่องมือของการทำเหมืองข้อมูล ให้เป็นไปตามมาตรฐานเดียวกัน และคัดเลือกข้อมูลที่มีความถูกต้องสมบูรณ์ คัดกรองเอาข้อมูลที่ไม่เกี่ยวข้องกับการนำไปวิเคราะห์ออกไป ข้อมูลที่ได้จากแบบสอบถาม จากคน 600 คน แบ่งเป็น 16 คุณลักษณะ ซึ่งคุณลักษณะที่ 16 เป็นคุณลักษณะที่เป็นกลุ่มของข้อมูลที่ใช้ในการจำแนกข้อมูล กลุ่มที่แบ่งได้จากประเมินจากแบบสอบถาม โดยแบ่งออกเป็น 5 กลุ่มได้แก่

กลุ่มที่ 1 แทนค่าด้วย A คือ กลุ่มคนที่อยู่ในระดับปกติหรือไม่มีความเสี่ยงจำนวน 191 คน

กลุ่มที่ 2 แทนค่าด้วย B คือ กลุ่มคนที่อยู่ในระดับมีความเสี่ยงเล็กน้อย จำนวน 167 คน

กลุ่มที่ 3 แทนค่าด้วย C คือ กลุ่มคนที่อยู่ในระดับมีความเสี่ยงปานกลาง จำนวน 175 คน

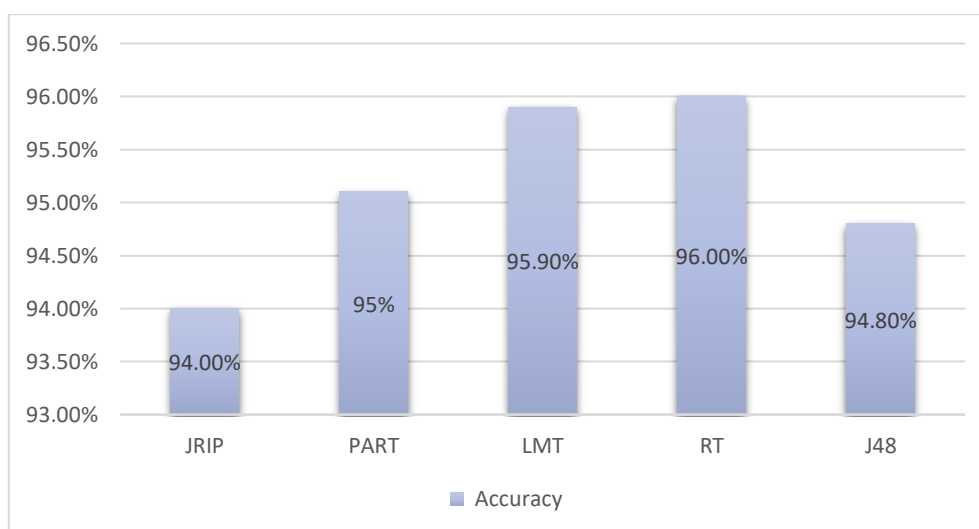
กลุ่มที่ 4 แทนค่าด้วย D คือ กลุ่มคนที่อยู่ในระดับมีความเสี่ยงรุนแรงค่อนข้างมาก จำนวน 49 คน

กลุ่มที่ 5 แทนค่าด้วย E คือ กลุ่มคนที่อยู่ในระดับมีความเสี่ยงรุนแรงมาก จำนวน 18 คน

3.4 การตรวจสอบประสิทธิภาพของตัวแบบ โดยในงานวิจัยนี้เลือกใช้วิธี Cross validation Test ในการทำงานวิจัยเพื่อใช้ในการแบ่งข้อมูลโมเดลเนื่องจากผลที่ได้มีความน่าเชื่อถือ จากนั้นทำการหาค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล เพื่อหาตัวแบบที่มีประสิทธิภาพดีที่สุดของอัลกอริทึม JRIP, PART, LMT, RT, J48 และ RF มาใช้เปรียบเทียบมีรายละเอียดแสดงดังตารางที่ 2 และภาพที่ 1

ตารางที่ 2 ผลค่าเฉลี่ยผลการจำแนกประเภทข้อมูลของอัลกอริทึมแต่ละชนิด

อัลกอริทึม CLASSIFY	ค่าความถูกต้อง ACCURACY	ค่าความแม่นยำ PRECISION	ค่าความระลึก RECALL	ค่าความถ่วงดุล F-MEASURE
JRIP	94.00%	94.00%	94.00%	94.00%
PART	95.10%	95.10%	95.10%	95.10%
LMT	95.90%	95.90%	95.90%	95.90%
RT	96.00%	96.00%	96.00%	96.00%
J48	94.80%	94.80%	94.80%	94.80%



ภาพที่ 1 การเปรียบเทียบข้อมูลในรูปแบบกราฟ

จากภาพที่ 1 ตัวแบบการจำแนกข้อมูลที่ได้จากการจำแนกประเภทข้อมูลด้วยอัลกอริทึม Random Tree มีเปอร์เซ็นต์ความถูกต้องมากที่สุด 96.00% จึงเป็นตัวแบบที่เหมาะสมต่อการนำไปประยุกต์ใช้สำหรับการพัฒนาตัวแบบประเมินความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาโดยตัวอย่างกฎที่สกัดได้จากอัลกอริทึม Random Tree ดังภาพที่ 2

If Mobile Phone Usage Time = 3-4hour and Number of Siblings = 2 People and Status = Single and Class Year = One and Pet = Yes and then class = A

If Mobile Phone Usage Time = More10andSex = F and Meeting Friends = Some days and Decreased Interest in Activities = No and The Marital Status of Parents = Together and Faculty = Dept and Age = 17-19Year and then class = B

ภาพที่ 2 ตัวอย่างกฎที่สกัดได้จากอัลกอริทึม Random Tree

3.5 การนำตัวแบบไปใช้งาน เป็นการนำตัวแบบที่สร้างได้เมื่อทำการทดสอบประสิทธิภาพไปใช้ทำนายหรือหาคำตอบให้กับข้อมูลใหม่ซึ่งยังไม่รู้คลาสคำตอบ ในงานวิจัยนี้ยังสามารถนำตัวแบบที่ได้จากต้นไม้ตัดสินใจมาประยุกต์ใช้งานต่อยอดในรูปแบบของการสร้างเว็บไซต์ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคมะเร็งลำไส้ของนักศึกษาด้วยเทคนิคเหมืองข้อมูลในอนาคตได้

4. สถิติที่ใช้ในการวิจัย

ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน นำผลที่ได้เทียบกับเกณฑ์การประเมิน ดังนี้

ค่าเฉลี่ยเท่ากับ 4.51 – 5.00 หมายถึง ระดับมากที่สุด

ค่าเฉลี่ยเท่ากับ 3.51 – 4.50 หมายถึง ระดับมาก

ค่าเฉลี่ยเท่ากับ 2.51 – 3.50 หมายถึง ระดับปานกลาง

ค่าเฉลี่ยเท่ากับ 1.51 – 2.50 หมายถึง ระดับน้อย

ค่าเฉลี่ยเท่ากับ 1.01 – 1.50 หมายถึง ระดับน้อยที่สุด

ผลการวิจัย

1. ผลการพัฒนาตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคมะเร็งลำไส้ของนักศึกษาด้วยเทคนิคเหมืองข้อมูล

ผู้วิจัยได้ดำเนินการพัฒนาตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคมะเร็งลำไส้ของนักศึกษาที่ส่งผลต่อการทำร้ายตัวเอง ตามขั้นตอนการวิจัยและเลือกตัวแบบจากอัลกอริทึม Random Tree ซึ่งเป็นตัวแบบที่มีประสิทธิภาพดีที่สุดมาใช้ในการพัฒนาระบบในส่วนของการทำนายความเสี่ยง โดยภาพรวมของการทำงานมีดังภาพที่ 3



ภาพที่ 3 ภาพรวมการทำงานของตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษา

อภิปรายผลการวิจัย

การพัฒนาตัวแบบประเมินความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษา โดยใช้เทคนิคการทำเหมืองข้อมูล พบว่า ในการเลือกเทคนิคมาทดลองเปรียบเทียบ ได้แก่ 1) อัลกอริทึม JRIP, อัลกอริทึม LMT, อัลกอริทึม PART, อัลกอริทึม Random Tree, และ อัลกอริทึม J48 พบว่า ตัวแบบที่ได้จากอัลกอริทึม Random Tree มีประสิทธิภาพสูงที่สุด โดยที่ค่าความ ถูกต้องเท่ากับ 96.00% ค่าความระลึกเท่ากับ 96.00% ค่าความแม่นยำเท่ากับ 96.00% และค่า ความเหวี่ยงเท่ากับ 96.00% ซึ่ง Random Tree เป็นกระบวนการของต้นไม้แบบสุ่มโดยมีหลักการสร้างต้นไม้จากการสุ่มต้นไม้หลาย ๆ แบบในแต่ละโหนดแล้วเลือกมาประมวลผล โดยไม่ใช้การการตัดแต่ง และเมื่อพิจารณาจำนวนกฎที่สกัดได้จาก Random Tree พบว่า มีจำนวนกฎมากกว่าเทคนิคอื่นค่อนข้างมาก

ข้อเสนอแนะ

จากผลการวิจัยพบว่า ในการสร้างตัวประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาเพื่อลดหรือช่วยป้องกันไม่ให้เกิดภาวะซึมเศร้าของนักศึกษา ส่วนมากลักษณะของผู้ที่มีความเสี่ยงการเป็นโรคซึมเศร้าจะเป็นเพศหญิง มีโอกาสเป็นมากกว่าผู้ชาย อยู่ในช่วงอายุ 20-22 ปีหรือในช่วงเข้าศึกษาในมหาวิทยาลัย ระดับปีที่ 3 และ 4 ซึ่งช่วงอายุนี้อาจมีความเครียดจากการเรียนหรือการกดดันจากทางบ้านหรือปัญหาต่าง ๆ รวมไปถึงการใช้เวลาอยู่กับการเล่นโทรศัพท์มือถือมากเกินไป พักผ่อนน้อย มีนิสัยเก็บตัว มีเพื่อนน้อย ไม่มีการพบปะสังสรรค์กับผู้อื่น มีภาระหนี้สิน ซึ่งจากที่กล่าวมาคือ ลักษณะของผู้ที่มีความเสี่ยงของผู้เป็นโรคซึมเศร้าที่ได้จากการวิจัย ทางผู้วิจัยเห็นว่า ควรได้รับ การพัฒนา ดังต่อไปนี้

1. ศึกษาและประยุกต์การใช้อัลกอริทึมอื่น ๆ ที่เป็นการแบ่งกฎการจำแนกเหมือนกัน เพื่อช่วยในการวิเคราะห์การตัดสินใจ
2. ข้อมูลที่เก็บรวบรวมมาใช้ในการวิเคราะห์ที่มีจำนวนน้อยเกินไป ดังนั้น ในอนาคตควรมีการรวบรวมข้อมูลเพื่อทำการวิเคราะห์อีกครั้ง เพื่อให้ได้ผลของการวิเคราะห์ข้อมูลมีความสมบูรณ์มากขึ้น

เอกสารอ้างอิง

- [1] POB PAD. (ม.ป.ป.). *โรคซึมเศร้า*. <https://www.pobpad.com/โรคซึมเศร้า>
- [2] Weka. (2561). *โรคซึมเศร้า*. <https://weka.sourceforge.io/doc.dev/allclasses-noframe.html>
- [3] สุทธนันท์ ชุนแจ่ม, โสภณ แสงอ่อน, และ ทศนา ทวีคุณ. (2554). การสำรวจงานวิจัยที่เกี่ยวกับภาวะซึมเศร้าในประเทศไทย. *Rama Nurs J*, 17(3), 412-429. file:///C:/Users/Admin/Downloads/9046-Article%20Text-18465-1-10-20130607.pdf
- [4] ศจี วาณิช. (2558). Data Mining (เหมืองข้อมูล). <http://sajeegm301.blogspot.com/2015/11/data-mining.html>
- [5] จิราภา เลหาหะวรรณนท์, รชต ลิ้มสุทธิวันภูมิ, และ บัณฑิต ฐานะโสภณ. (2558). การใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือก. *วารสารเทคโนโลยีสารสนเทศลาดกระบัง*, 4(2), 1-9. file:///C:/Users/Admin/Downloads/65-361-1-PB.pdf
- [6] ดนัย คำยา. (2560, 9 เมษายน). *การใช้งานโปรแกรม Weka ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล*. <https://www.glurgeek.com/education/howto-weka/>
- [7] กมลรัตน์ สมใจ. (2563, 6-7 กุมภาพันธ์). การเปรียบเทียบประสิทธิภาพโมเดลการตัดสินใจเลือกกลุ่มวิชาของนักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์. ใน *วิทยาศาสตร์เทคโนโลยี และนวัตกรรมเพื่อสังคม*. [Symposium]. งานประชุมวิชาการระดับชาติด้านวิทยาศาสตร์และเทคโนโลยีเครือข่ายสถาบันอุดมศึกษาภาคใต้ ครั้งที่ 5 มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย นครศรีธรรมราช. <http://dspace.bru.ac.th/xmlui/bitstream/handle/123456789/6006/Template%20NSCIC%202020E0.pdf?sequence=1&isAllowed=y&fbclid=IwAR1tSgZbpipO5-XLXoS-GjYJGiZecPLWu7bMSnldj1735B9czOPTxttYNk>
- [8] ภรณ์ยา ปาลวิสุทธิ. (2559). การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดเชื้อเอนเทอโรเน็ต. *วารสารเทคโนโลยีสารสนเทศ*, 12,(1), 54-63. file:///C:/Users/Admin/Downloads/72840-Article%20Text-172635-1-10-20161222.pdf
- [9] ปณิดา บุญพาณิชย์, รัชมน กัลยาศิริ, และ ณัฏฐวรรต บัวทอง. (2561). ภาวะซึมเศร้าและความเหนื่อยล้าของพนักงานต้อนรับบนเครื่องบินในประเทศไทย. *จุฬาลงกรณ์เวชสาร*, 62(6), 1023-1035.
- [10] ขนิษฐา ขัตติยะ. (2561). *พยากรณ์ความเสี่ยงของโรคซึมเศร้า โดยใช้เทคนิคเหมืองข้อมูล*. นครปฐม: มหาวิทยาลัยราชภัฏนครปฐม.
- [11] คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล. (ม.ป.ป.). *แบบทดสอบภาวะซึมเศร้า PHQ-9*. https://med.mahidol.ac.th/th/depression_risk