

ปัจจัยแบบจำลองการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวชด้วยเทคนิคเหมืองข้อมูล

Prediction Models of the Sele-Harm for Psychiatric Patients using Data Mining Techniques

โสภิตา สามารต* และจारी ทองคำ

Sopita Samart and Jaree Thongkam

Received: August 7, 2024; Revised: November 5, 2024; Accepted: November 7, 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพแบบจำลองการทำร้ายตนเองของผู้ป่วยจิตเวช ข้อมูลผู้ป่วยได้ถูกรวบรวมจากฐานข้อมูลโรงพยาบาลจิตเวชขอนแก่นราชนครินทร์ จังหวัดขอนแก่น จากเดือนมกราคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2564 จำนวน 4,179 ระเบียบ โดยผู้วิจัยได้นำการคัดเลือกปัจจัยที่เกี่ยวข้องกับการพยากรณ์สำหรับการสร้างแบบจำลองด้วยวิธี Gain Ratio และ Relief จากนั้นได้นำเอาเทคนิคนาอึฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วนมาทำการสร้างแบบจำลองเพื่อการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวช ในการทดลองนี้ 10-Fold Cross Validation ได้ถูกนำมาใช้ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ ส่วนค่าความถูกต้อง ค่าความไว และค่าความจำเพาะ ได้ถูกนำไปใช้ในการเปรียบเทียบประสิทธิภาพแบบจำลอง ผลการทดลองแสดงว่า การคัดเลือกปัจจัยด้วยวิธี Gain Ratio เมื่อนำมาสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ C4.5 หรือเทคนิคสนับสนุนด้วยเครื่องมีค่าความถูกต้องมากที่สุด เหมือนกันที่ร้อยละ 84.66

คำสำคัญ : โรคทางจิตเวช; การคัดเลือกปัจจัย; เหมืองข้อมูล

คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

Faculty of Information Science, Mahasarakham University

* Corresponding Author, Tel. 08 5008 8472, E - mail: 65011282001@msu.ac.th

Abstract

The objective of this research is to create and compare the effectiveness of a self-harm model for psychiatric patients. The data were collected from Khon Kaen Rajanagarindra Psychiatric Hospital database, Khon Kaen province from January 2019 to December 2021. The records were 4,179 records in total. Gain Ratio and Relief methods were used to select relevant factors for building the prediction models. Naive Bay, Decision tree, Machine support, Deep learning, Partial rule techniques were utilized to build self-harm in psychiatric patients prediction models. In this experiment, 10-fold cross validation was used to divide the data into learning and testing datasets. Accuracy, sensitivity, and specificity were employed to compare the effectiveness of the models. The experimental results showed that selection of factors using the Gain Ratio method when used C4.5 decision tree or machine support vector machine techniques has the same highest accuracy at 84.66 %.

Keywords: Psychiatric Disease; Feature Selection; Data Mining

บทนำ

ปัจจุบันการฆ่าตัวตายเป็นปัญหาสำคัญด้านสาธารณสุขของโลก ซึ่งคิด 10 อันดับของสาเหตุการเสียชีวิตของประชากรทั่วโลก โดยเกิดจากพฤติกรรมความรุนแรง เช่น การทำร้ายร่างกาย ความก้าวร้าว การไม่ใช้วิจารณญาณและความเจ็บป่วยทางจิตเวชหรือทางกายที่ส่งผลกระทบต่อสุขภาพจิตของประชาชน (World Health Organization, 2021) ซึ่งกลุ่มที่มีความเสี่ยงสูง คือ กลุ่มที่มีความผิดปกติทางจิต เช่น ผู้ป่วยโรคซึมเศร้า โรคอารมณ์สองขั้ว โรคจิตเภท โรคบุคลิกภาพผิดปกติ กลุ่มที่ใช้สารเสพติด และกลุ่มที่ใช้แอลกอฮอล์ (Khonkaen Rajanagarindra Psychiatric Hospital, National Center for Suicide, 2022) จากการศึกษาพบว่าผู้ที่เคยพยายามฆ่าตัวตาย จะมีโอกาสทำซ้ำและทำสำเร็จในระยะ 1 ปี หลังการทำร้ายตนเองครั้งแรก ในผู้ที่มีความคิดอยากตาย จะมีพฤติกรรมการพยายามฆ่าตัวตายกว่า 20 ครั้ง ถึงจะตายสำเร็จ (World Health Organization, 2019) จึงได้มุ่งเน้นที่จะสร้างแบบจำลองการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวชด้วยเทคนิคเหมืองข้อมูล เพื่อเป็นแนวทางการป้องกันการทำร้ายตนเองในผู้ป่วยจิตเวช ผู้วิจัยจึงมีแนวคิดที่จะศึกษาเกี่ยวกับกลุ่มผู้ป่วยจิตเวชด้วยการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์และการพยากรณ์

เหมืองข้อมูล (Data Mining) คือ กระบวนการในการวิเคราะห์ข้อมูลจากข้อมูลจำนวนมากเพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในข้อมูลนั้น เพื่อให้ได้องค์ความรู้ใหม่ที่สามารถนำไปใช้ประกอบการตัดสินใจ (Rungrattanaubol, 2022) เช่น Boonprasom and Sarach (2019) ได้ทำการศึกษาการวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษาระดับปริญญาตรีจากฐานข้อมูลงานทะเบียนของมหาวิทยาลัยราชภัฏอุบลราชธานี ส่วน Molee et al. (2019) ได้ทำการศึกษาภาวะเสี่ยงโรคของผู้สูงอายุด้วยเทคนิคเหมืองข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจ C4.5, เทคนิคกฎบางส่วน (Partial Rule) และเทคนิคการเหนี่ยวนำ (Induction) ส่วน Boonma and Chirawichitchai (2019) ได้เสนอการสร้างแบบจำลองการจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูลและทำการเปรียบเทียบประสิทธิภาพของแบบจำลองของเทคนิคเหมืองข้อมูลด้วยเทคนิคคอฟฟิเบย์, เทคนิคเคเนียร์เรนเบอร์, เทคนิคต้นไม้ตัดสินใจและเทคนิคสนับสนุนด้วยเครื่อง ส่วน Hongboonmee and Thammakorn (2562) ได้ศึกษาการคัดกรองผู้ที่เสี่ยงต่อภาวะซึมเศร้าผ่านสมาร์ทโฟนโดยใช้เทคนิคเหมืองข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียมและเทคนิคคอฟฟิเบย์ และในด้านธุรกิจ เช่น ระบบแนะนำหนังสือให้กับลูกค้าใน amazon.com ว่า ลูกค้าที่ซื้อหนังสือเล่มหนึ่ง มักจะซื้อหนังสือเล่มใดพร้อมกันด้วยเสมอ ซึ่งสามารถนำไปใช้คาดเดาว่าควรแนะนำหนังสือเล่มใดเพิ่มเติมให้กับลูกค้าที่เพิ่งซื้อหนังสือจากร้านไป

ดังนั้นในงานวิจัยนี้จึงได้สร้างแบบจำลองการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวชด้วยเทคนิคเหมืองข้อมูล ด้วยการคัดเลือกปัจจัยโดยวิธีการ Chi-Square, Gain Ratio และ Relief ร่วมกับเทคนิคนาอ็พเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วน ด้วยหลักการ 10-Fold Cross Validation เพื่อนำมาวิเคราะห์หาประสิทธิภาพของแบบจำลองโดยการค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity)

วิธีดำเนินการวิจัย

ในการวิจัยนี้ผู้วิจัยได้ทำการศึกษาค้นคว้าทฤษฎีและทบทวนงานวิจัยที่เกี่ยวข้อง เพื่อนำมาใช้ในการสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลอง ซึ่งสามารถแบ่งได้ 4 ขั้นตอนประกอบด้วย ขั้นตอนการเตรียมข้อมูล ขั้นตอนก่อนสร้างแบบจำลอง ขั้นตอนการสร้างแบบจำลอง และขั้นตอนการวัดประสิทธิภาพของแบบจำลอง

1. ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูลในงานวิจัยครั้งนี้ได้นำข้อมูลผู้ป่วยจิตเวชจากฐานข้อมูลโรงพยาบาลจิตเวชขอนแก่นราชนครินทร์ จังหวัดขอนแก่น จากวันที่ 1 มกราคม พ.ศ. 2562 ถึงวันที่ 31 ธันวาคม พ.ศ. 2564 จากการศึกษาจากงานวิจัยอื่น ๆ ที่เกี่ยวข้องกับปัจจัยและสาเหตุที่ก่อให้เกิดการทำร้ายตนเอง หรือผู้มีความคิดอยากทำร้ายตนเอง ทำให้ผู้วิจัยมีแนวคิดที่จะศึกษาเกี่ยวกับผู้ป่วยทางด้านจิตเวชที่คาดว่าจะทำร้ายตนเอง โดยอ้างอิงตัวแปรจากงานวิจัยที่เกี่ยวข้อง และจากการสอบถามผู้ที่มีความเชี่ยวชาญในการทำวิจัยเกี่ยวกับผู้ป่วยจิตเวชและเหมืองข้อมูล ซึ่งได้ประกอบด้วยตัวแปรจำนวน 16 ปัจจัยซึ่งสามารถแบ่งเป็นปัจจัยประเภทค่ามาตรานามบัญญัติดังตารางที่ 1 และปัจจัยประเภทค่าตัวเลขดังตารางที่ 2

ตารางที่ 1 ปัจจัยประเภทมาตรานามบัญญัติ

ลำดับที่	ชื่อปัจจัย	ชื่อปัจจัยในระบบ	จำนวนค่าของปัจจัย
1	เพศ	Sex	2
2	สถานภาพ	Status	6
3	บุตรคนที่	Series	11
4	อาชีพ	Occupation	9
5	การศึกษา	Education	10
6	โรคประจำตัว	Diagnosis_cs	31
7	การวินิจฉัยทางสังคม	Social_diag	17
8	การใช้สารเสพติด	Substance	9
9	โรคหลักทางจิตเวช	Main_disease	13
10	โรคร่วมลำดับที่ 1	Second_disease	105
11	โรคร่วมลำดับที่ 2	Third_disease	84
12	คลาส	Class	2

จากตารางที่ 1 แสดงปัจจัยประเภทค่ามาตรานามบัญญัติ (Nominal Scale) ในการจัดข้อมูลเป็นกลุ่มต่างที่แยกออกจากกัน ได้แก่ เพศ สถานภาพ บุตรคนที่ อาชีพ โรคประจำตัว การวินิจฉัยทางสังคม การใช้สารเสพติด โรคหลักทางจิตเวช โรคร่วมลำดับที่ 1 โรคร่วมลำดับที่ 2 และคลาส ซึ่งแสดงจำนวนค่าปัจจัยในการแยกจำนวนปัจจัย

ตารางที่ 2 ปัจจัยประเภทตัวเลข

ลำดับที่	ชื่อปัจจัย	ชื่อปัจจัยในระบบ	ค่าต่ำ	ค่าสูง	ค่ากลาง	S.D.
1	อายุครั้งแรก	Age	14	89	51.5	53.03
2	ความดันบน	SBP	62	200	131	97.58
3	ความดันล่าง	DBP	28	179	103.5	106.77
4	ค่าดัชนีมวลกาย	BMI	13.95	66	19.97	36.80

จากตารางที่ 2 แสดงปัจจัยประเภทค่าตัวเลข (Numeric Value) ที่ใช้แทนจำนวนที่สามารถนำไปคำนวณ ได้แก่ อายุครั้งแรก ความดันบน ค่าความดันล่างและค่าดัชนีมวลกาย เพื่อนำมาใช้ในการหาค่าต่ำ ค่าสูง ค่ากลาง และ ส่วนเบี่ยงเบนมาตรฐาน

2. ขั้นตอนก่อนการสร้างแบบจำลอง

ขั้นตอนก่อนการสร้างแบบจำลองในงานวิจัยนี้ เริ่มจากการทำความสะอาดข้อมูลที่ซ้ำซ้อนและไม่สมบูรณ์ แล้วจึงทำการแปลงรูปแบบข้อมูลให้อยู่ในรูปแบบที่พร้อมจะนำไปวิเคราะห์ รวมถึงการเลือกปัจจัยโดยวิธีการ Chi-Square, Gain Ratio และ Relief มาวิเคราะห์การพยากรณ์การทำร้ายตัวเองของผู้ป่วยจิตเวช ดังนี้

2.1 วิธีการ Chi-Square

วิธีการ Chi-Square เป็นสถิติที่ทดสอบความสัมพันธ์ระหว่างตัวแปรที่ใช้กับข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณที่แบ่งเป็นช่วง และข้อมูลไม่จำเป็นต้องมีการแจกแจงแบบปกติ ตัวเลขที่นำมาวิเคราะห์ในการทดสอบไคสแควร์เป็นความถี่ของแต่ละระดับของตัวแปร ซึ่งความถี่ได้จากตัวอย่างเรียกว่า ความถี่ที่สังเกตได้ (Observed Frequency) แทนด้วยสัญลักษณ์ O และความถี่ตามทฤษฎีเรียกว่า ความถี่คาดหวัง (Expected Frequency) แทนด้วยสัญลักษณ์ E และนำมาเปรียบเทียบความแตกต่าง (Hammongkolpipat, 2017) ดังสมการที่ 1

$$x^2 = \sum \frac{(O-E)^2}{E} \quad (1)$$

เมื่อ x^2 แทน ตัวสถิติทดสอบไคสแควร์
 O แทน ความถี่ที่ได้จากการสังเกต
 E แทน ความถี่ที่คาดหวัง

2.2 วิธีการ Gain Ratio

วิธีการ Gain Ratio เป็นตัวชี้วัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจาก Information Gain เนื่องจากการใช้ Information Gain ในการแบ่งชุดข้อมูลจะมีโอกาสทำให้เกิดความเอนเอียงขึ้น เมื่อปัจจัยที่ทำการพิจารณาได้ค่า gain ที่สูงเป็นจำนวนมาก ทำให้ปัจจัยที่ถูกคัดเลือกไม่ถูกต้อง ตัวอย่างเช่น การพิจารณาปัจจัยที่ทำหน้าที่เป็นตัวระบุเฉพาะ เช่น รหัสผลิตภัณฑ์ การแยกรหัสผลิตภัณฑ์ จะส่งผลให้มีชุดข้อมูลย่อยจำนวนมาก แต่ละชุดข้อมูลย่อยมีเพียง 1 record ซึ่งเมื่อนำค่า Information Gain จะได้ค่าที่สูงจำนวนมากจากความเอนเอียง ทำให้มีการพัฒนาตัวชี้วัดการแบ่งข้อมูลใหม่ที่ชื่อเรียกว่า Gain Ratio โดยการประยุกต์ใช้การทำนอร์มัลไลซ์ค่า Information Gain ด้วยการใช้ค่า Split Information แล้วจะสามารถคำนวณหาค่า Gain Ratio (Xiaowei et al., 2023; Senthil Kumar, 2016) ซึ่งสามารถคำนวณได้ดังสมการที่ 2

$$GainRatio(A) = \frac{Gain(A)}{splitInfo(A)} \quad (2)$$

2.3 วิธีการ Relief

วิธีการ Relief เป็นการกรองข้อมูลเพื่อเลือกปัจจัยที่มีความอ่อนไหวต่อการโต้ตอบของปัจจัย โดยเฉพาะ เดิมถูกออกแบบสำหรับการประยุกต์ใช้กับปัญหาการจำแนกไบนารีที่มีคุณสมบัติไม่ต่อเนื่องหรือเป็นตัวเลข Relief คำนวณคะแนนคุณสมบัติสำหรับแต่ละคุณสมบัติซึ่งสามารถนำไปใช้ในการจัดอันดับและเลือกคุณสมบัติที่ให้คะแนนสูงสุดสำหรับการเลือกคุณสมบัติ หรืออีกทางหนึ่ง คะแนนเหล่านี้จะนำไปใช้เป็นน้ำหนักปัจจัยเพื่อเป็นแนวทางในการสร้างแบบจำลอง (Kenji and Larry, 1992) ซึ่งสามารถคำนวณได้ดังสมการที่ 3

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad (3)$$

เมื่อ	W_i	แทน	น้ำหนักของเวกเตอร์
	X_i	แทน	คุณลักษณะเวกเตอร์
	$nearHit_i$	แทน	ตัวอย่างคลาสเดียวกันที่อยู่ใกล้ที่สุด
	$nearMiss_i$	แทน	ตัวอย่างคลาสที่แตกต่างกันที่ใกล้เคียงที่สุด

3. ขั้นตอนการสร้างแบบจำลอง

ขั้นตอนการสร้างแบบจำลองในงานวิจัยนี้ได้นำเทคนิคที่มีประสิทธิภาพในการต่อการสร้างแบบจำลองในการพยากรณ์ประกอบด้วย 5 เทคนิค

3.1 เทคนิคนาอิวเบย์ เป็นขั้นตอนการประมวลผลในการแก้ปัญหาที่อาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน โดยใช้ขั้นตอนในการจำแนกข้อมูล เพื่อการเรียนรู้ปัญหาที่เกิดขึ้น แล้วนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ โดยใช้หลักการคำนวณหาความน่าจะเป็นในการทำนายผลเป็นอัลกอริทึมในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ และทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้าง เงื่อนไขความจะเป็นสำหรับแต่ละความสัมพันธ์ (Chinprasatsak et al., 2023)

3.2 เทคนิคต้นไม้ตัดสินใจ C4.5 เป็นขั้นตอนการประมวลผลในการแก้ปัญหาที่สามารถนำมาช่วยในการจำแนกข้อมูลและปัจจัยที่ใช้ในการวิเคราะห์ แต่ละตัวมีความสำคัญมากน้อยเหมือนหรือแตกต่างกัน และในขั้นตอนการพยากรณ์ต้นไม้ตัดสินใจจะทำการพยากรณ์ตัวเอง และคำนวณผลการพยากรณ์ผลลัพธ์ที่ถูกเลือกโดยต้นไม้ตัดสินใจมากที่สุด คือ การจำแนกกลุ่ม หรือหาค่าเฉลี่ยจากผลลัพธ์ในกรณีถดถอย (Eibe and Ian, 1998)

3.3 เทคนิคสนับสนุนด้วยเครื่อง เป็นขั้นตอนการประมวลผลในการแก้ปัญหาการจำแนกข้อมูลไปใช้ในการวิเคราะห์ โดยอาศัยหลักการของการหาสมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ ซึ่งเน้นไปยังเส้นแบ่งแยกแยกกลุ่มข้อมูลได้ดีที่สุด ในการจำแนกเชิงเส้นแบบ 2 คลาส โดยใช้ไฮเปอร์เพลนที่เป็นเส้นตรงที่จะสามารถแบ่งแยกข้อมูลออกจากกันได้แต่เส้นตรงที่ถูกพิจารณาให้เป็นเส้นที่ดีที่สุดคือ เส้นตรงที่ลากผ่านข้อมูลที่ใกล้ที่สุดและขนานกับเส้นไฮเปอร์เพลนของข้อมูลแต่ละกลุ่มที่มากที่สุด (Sanguansat, 2019)

3.4 เทคนิคการเรียนรู้เชิงลึก เป็นการเรียนรู้ของพื้นฐานของโครงข่ายประสาทเทียมและการเรียนรู้เชิงปัจจัย ซึ่งจะเป็นการเรียนรู้แบบมีผู้สอน การเรียนรู้แบบกึ่งมีผู้สอน และการเรียนรู้แบบไม่มีผู้สอนที่มีปัจจัยของโครงข่ายแบบหลายชั้น โดยการใช้ข้อมูลที่มีอยู่เพื่อสร้างคุณสมบัติที่ทำให้ได้รับตัวแปรใหม่มาใช้ในการวิเคราะห์ ทำให้เกิดระบบเชิงคาดการณ์ที่มีได้รับการยอมรับที่ดี ปรับแต่งได้ดี มีการปรับปรุงอย่างต่อเนื่องเมื่อมีข้อมูลใหม่เข้ามา และมีประสิทธิภาพ สะดวกและเข้าใจในโครงสร้างที่ชัดเจน (Digital Government Development Agency, 2022)

3.5 เทคนิคกฎบางส่วน (Partial Rule) เป็นเทคนิคที่ใช้แยกและพิชิตข้อมูล ซึ่งในการสร้างกฎการตัดสินใจโดยการนำเอาหลักการของต้นไม้การตัดสินใจ C4.5 และ RIPPER ทั้งสองแบบทำงานในสองขั้นตอน ขั้นแรกจะสร้างชุดกฎเริ่มต้น จากนั้นจึงปรับแต่งโดยใช้ขั้นตอนการเพิ่มประสิทธิภาพที่ค่อนข้างซับซ้อนและจะละทิ้ง (C4.5) หรือปรับ (RIPPER) กฎแต่ละข้อเพื่อให้ทำงานร่วมกันได้ดีขึ้น ในทางตรงกันข้าม ในการพยากรณ์นี้ถึงแม้ว่าเทคนิคนี้จะคล้ายกับต้นไม้ตัดสินใจ C4.5 แต่เทคนิคกฎบางส่วน จะหลีกเลี่ยงการสร้างต้นไม้เต็มรูปแบบทำให้ลดเวลาในการสร้างกฎการตัดสินใจ (Decision Rule) (Eibe and Ian, 1998)

4. การวัดประสิทธิภาพของแบบจำลองการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวช

4.1 ค่าความถูกต้อง (Accuracy) คือ ค่าที่แบบจำลองสามารถจำแนกข้อมูลผู้ป่วยที่เกิดโรค และไม่เกิดโรคได้อย่างถูกต้องต่อข้อมูลทั้งหมด ดังสมการที่ (4)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

4.2 ค่าความไว (Sensitivity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่เกิดโรคจริง ดังสมการที่ (5)

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

4.3 ค่าความจำเพาะ (Specificity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่ไม่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่พยากรณ์ว่าเกิดโรค ดังสมการที่ 6

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

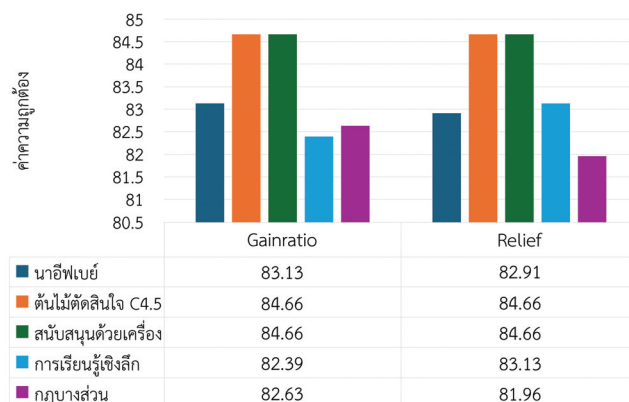
เมื่อ	TP	คือ	จำนวนข้อมูลที่แบบจำลองพยากรณ์การเกิดโรคได้อย่างถูกต้อง
	TN	คือ	จำนวนข้อมูลที่แบบจำลองพยากรณ์การไม่เกิดโรคได้อย่างถูกต้อง
	FP	คือ	จำนวนข้อมูลที่แบบจำลองพยากรณ์การเกิดโรคได้ไม่ถูกต้อง
	FN	คือ	จำนวนข้อมูลที่แบบจำลองพยากรณ์การไม่เกิดโรคได้ไม่ถูกต้อง

ผลการศึกษา

ผลการวิจัยในงานวิจัยนี้ได้ศึกษาเพื่อสร้างและเปรียบเทียบประสิทธิภาพแบบจำลองผู้พยายามทำร้ายตนเองด้วยเทคนิคเหมืองข้อมูล โดยคัดเลือกปัจจัยจากวิธีการ Chi-square, Gain Ratio และ Relief มาใช้ในเทคนิคนาอ์ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วน หลังจากนั้นได้นำวิธี 10-fold cross validation มาใช้ในการแบ่งกลุ่มข้อมูลเป็น 10 ชุดข้อมูลเท่ากัน ซึ่ง 9 ชุดข้อมูลเป็นชุดเรียนรู้ และอีก 1 ชุดเป็นชุดข้อมูลทดสอบ และทำการทดสอบ 10 รอบโดยแต่ละรอบชุดข้อมูลทดสอบจะไม่ซ้ำกัน ทำให้การทดลองไม่อยู่ในชุดข้อมูลเรียนรู้ ในการวัดประสิทธิภาพการพยากรณ์ด้วยค่าความถูกต้อง ค่าความไว และค่าความจำเพาะ จากการทดลองพบว่า การคัดเลือกปัจจัย Chi-square สามารถคัดเลือกปัจจัยได้ทั้งหมด 3 ปัจจัย การคัดเลือกปัจจัยวิธี Gain Ratio คัดเลือกปัจจัยได้ทั้งหมด 12 ปัจจัย และส่วนการคัดเลือกปัจจัยวิธี Relief สามารถคัดเลือกปัจจัยได้ทั้งหมด 15 ปัจจัย น้ำหนักของปัจจัยมีความใกล้เคียงกัน ในการทดลองยังสนใจกลุ่มข้อมูลที่มีจำนวนมาก ผู้วิจัยจึงได้ทำคัดเลือกปัจจัยที่ใช้วิธี Gain Ratio และ Relief ที่มีค่าน้ำหนักจากมากมาหาน้อยจำนวน 12 ปัจจัยในการวัดประสิทธิภาพการพยากรณ์การทำร้ายตนเองของผู้ป่วยจิตเวช

1. ผลของค่าความถูกต้อง

ค่าความถูกต้อง คือ ค่าที่แบบจำลองสามารถจำแนกข้อมูลผู้ป่วยที่เกิดโรค และไม่เกิดโรคได้อย่างถูกต้องต่อข้อมูลทั้งหมด ในการทดลองนี้ได้นำวิธี 10-Fold Cross Validation มาใช้ในการแบ่งกลุ่มข้อมูลเป็น 10 ชุดข้อมูลเท่ากัน ซึ่ง 9 ชุดข้อมูลเป็นชุดเรียนรู้และอีก 1 ชุดเป็นชุดข้อมูลทดสอบ และทำการทดสอบ 10 รอบโดยแต่ละรอบชุดข้อมูลทดสอบจะไม่ซ้ำกันด้วยเทคนิคนาอ์ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วนดังรูปที่ 1

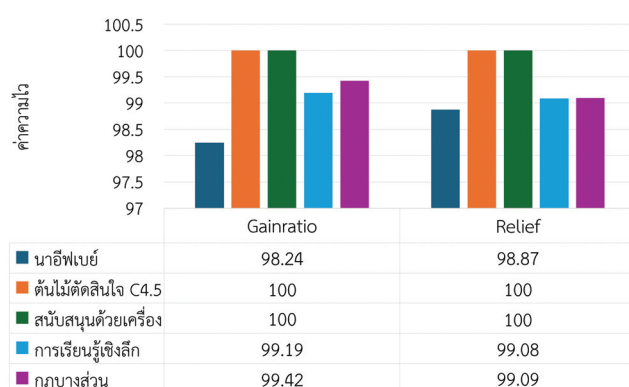


รูปที่ 1 ค่าความถูกต้องของแบบจำลอง

จากรูปที่ 1 แสดงค่าความถูกต้องของแบบจำลองหลังการคัดเลือกปัจจัย โดยการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาใช้ในเทคนิคนาอ็ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วนพบว่า การคัดเลือกปัจจัยด้วยวิธี Gain Ratio เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมีค่าความถูกต้องมีค่าร้อยละ 84.66 รองลงมาเทคนิคนาอ็ฟเบย์มีร้อยละ 83.13 รองลงมาเทคนิคกฎบางส่วนร้อยละ 82.63 และเทคนิคการเรียนรู้เชิงลึกร้อยละ 82.39 ส่วนการคัดเลือกปัจจัยด้วยวิธี Relief เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมีค่าความถูกต้องมากที่สุดมีค่าร้อยละ 84.66 รองลงมาเทคนิคการเรียนรู้เชิงลึกร้อยละ 83.13 รองลงมาเทคนิคนาอ็ฟเบย์ร้อยละ 82.91 และเทคนิคกฎบางส่วนร้อยละ 81.96 เมื่อนำการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาเปรียบเทียบเทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมีค่าความถูกต้องมากที่สุดมีค่าเท่ากับร้อยละ 84.66

2. ผลค่าความไว

ค่าความไว คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่เกิดโรคจริง ในการทดลองนี้ได้นำวิธี 10-Fold Cross Validation มาใช้ในการแบ่งกลุ่มข้อมูลเป็น 10 ชุดข้อมูลเท่ากัน ซึ่ง 9 ชุดข้อมูลเป็นชุดเรียนรู้และอีก 1 ชุดเป็นชุดข้อมูลทดสอบ และทำการทดสอบ 10 รอบโดยแต่ละรอบชุดข้อมูลทดสอบจะไม่ซ้ำกัน ด้วยเทคนิคนาอ็ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วนดังรูปที่ 2



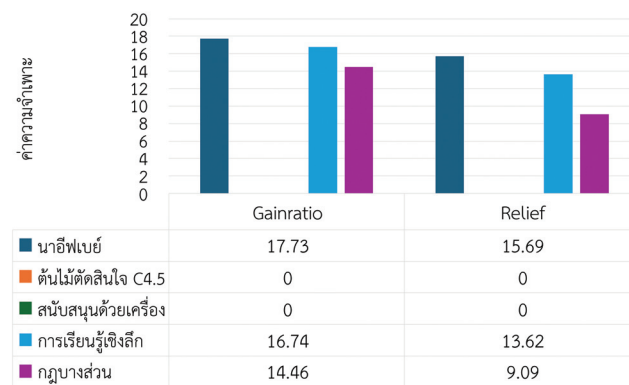
รูปที่ 2 ค่าความไวของแบบจำลอง

จากรูปที่ 2 แสดงค่าความไวของแบบจำลองหลังการคัดเลือกปัจจัย โดยการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาใช้ในเทคนิคนาอ็ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึก

และเทคนิคกฎบางส่วน พบว่าการคัดเลือกปัจจัยด้วยวิธี Gain Ratio เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมือมีความไวมีค่าร้อยละ 100.00 รองลงมาเทคนิคกฎบางส่วนร้อยละ 99.42 รองลงมาเทคนิคนาอึฟเบย์ร้อยละ 99.24 และเทคนิคการเรียนรู้เชิงลึกร้อยละ 99.19 ส่วนการคัดเลือกปัจจัยด้วยวิธี Relief เทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมือที่มีความไวมากที่สุดมีค่าร้อยละ 100.00 รองลงมาเทคนิคนาอึฟเบย์ร้อยละ 99.87 รองลงมาเทคนิคกฎบางส่วนร้อยละ 99.09 และเทคนิคการเรียนรู้เชิงลึกร้อยละ 99.08 เมื่อนำการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาเปรียบเทียบกับเทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมือที่มีความไวมากที่สุดมีค่าเท่ากับร้อยละ 100.00

3. ผลค่าความจำเพาะ

ค่าความจำเพาะ คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่ไม่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่พยากรณ์ว่าเกิดโรคในการทดลองนี้ได้วิธี 10-Fold Cross Validation มาใช้ในการแบ่งกลุ่มข้อมูลเป็น 10 ชุดข้อมูลเท่ากัน ซึ่ง 9 ชุดข้อมูลเป็นชุดเรียนรู้และอีก 1 ชุดเป็นชุดข้อมูลทดสอบ และทำการทดสอบ 10 รอบ โดยแต่ละรอบชุดข้อมูลทดสอบจะไม่ซ้ำกันด้วยเทคนิคนาอึฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่องมือการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วนดังรูปที่ 3



รูปที่ 3 ค่าความจำเพาะของแบบจำลอง

จากรูปที่ 3 แสดงค่าความจำเพาะของแบบจำลองหลังการคัดเลือกปัจจัย โดยการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาใช้ในเทคนิคนาอึฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่องมือ เทคนิคการเรียนรู้เชิงลึก และเทคนิคกฎบางส่วนพบว่า การคัดเลือกปัจจัยด้วยวิธี Gain Ratio เทคนิคนาอึฟเบย์มีค่าความจำเพาะมีค่าเท่ากับร้อยละ 17.73 รองลงมาเทคนิคการเรียนรู้เชิงลึกร้อยละ 16.71 รองลงมาเทคนิคกฎบางส่วนร้อยละ 14.46 รองลงมาเทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมือร้อยละ 0 ส่วนการคัดเลือกปัจจัยด้วยวิธี Relief เทคนิคนาอึฟเบย์มีค่าความจำเพาะมากที่สุดมีค่าเท่ากับร้อยละ 15.69 รองลงมาเทคนิคการเรียนรู้เชิงลึกร้อยละ 13.62 รองลงมาเทคนิคกฎบางส่วนร้อยละ 9.09 รองลงมาเทคนิคต้นไม้ตัดสินใจ C4.5 และเทคนิคสนับสนุนด้วยเครื่องมือร้อยละ 0 เมื่อนำการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และ Relief มาเปรียบเทียบกับเทคนิคนาอึฟเบย์ มีค่าความจำเพาะมากที่สุดมีค่าเท่ากับร้อยละ 17.73

สรุปผลและการอภิปรายผล

ในงานวิจัยนี้ได้ศึกษาปัจจัยการฆ่าตัวตายของผู้ป่วยจิตเวชและเปรียบเทียบประสิทธิภาพแบบจำลองผู้พยายามทำร้ายตนเองจากข้อมูลที่รวบรวมจากฐานข้อมูลโรงพยาบาลจิตเวชขอนแก่นราชนครินทร์ จากวันที่ 1 มกราคม พ.ศ. 2562 ถึงวันที่ 31 ธันวาคม พ.ศ. 2564 จำนวนทั้งหมด 4,179 ระเบียน โดยสมมติฐานว่า การคัดเลือกปัจจัยที่สำคัญจะทำให้มิติของข้อมูลลดลง และน่าจะมีผลการวัดประสิทธิภาพเพิ่มมากขึ้นในการวัดค่าความถูกต้อง ค่าความไว และค่าความจำเพาะ

จึงได้ทำการศึกษาและเปรียบเทียบการนำเทคนิคเหมืองข้อมูลจำนวน 5 เทคนิค ได้แก่ เทคนิคนาอ์ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึกและเทคนิคกฎบางส่วน จากผลการทดลองด้วยวิธีการคัดเลือกปัจจัยจากวิธี Chi-square, Gain Ratio และ Relief พบว่า การคัดเลือกปัจจัย Chi-square สามารถคัดเลือกปัจจัยได้ทั้งหมด 3 ปัจจัย เมื่อนำมาคำนวณในเทคนิคนาอ์ฟเบย์ เทคนิคต้นไม้ตัดสินใจ C4.5 เทคนิคสนับสนุนด้วยเครื่อง เทคนิคการเรียนรู้เชิงลึก และเทคนิคกฎบางส่วน ในเทคนิคต่าง ๆ มาคำนวณหาค่าความถูกต้อง ค่าความไว และค่าความจำเพาะมีค่าเท่ากับ 0 และการคัดเลือกปัจจัยวิธี Gain Ratio คัดเลือกปัจจัยได้ทั้งหมด 12 ปัจจัย ส่วนการคัดเลือกปัจจัยวิธี Relief สามารถคัดเลือกปัจจัยได้ทั้งหมด 15 ปัจจัย น้ำหนักของปัจจัยมีความใกล้เคียงกัน และการทดลองยังสนใจกลุ่มข้อมูลที่มีจำนวนมาก ผู้วิจัยจึงได้ทำคัดเลือกปัจจัยที่ใช้วิธี Gain Ratio และ Relief ที่มีค่าน้ำหนักจากมากมาหาน้อยจำนวน 12 ปัจจัย โดยการวัดประสิทธิภาพการพยากรณ์ด้วยค่าความถูกต้องมากที่สุด มีค่าเท่ากับร้อยละ 84.66 ค่าความไวมากที่สุดมีค่าเท่ากับ 100 และค่าความจำเพาะมีค่ามากที่สุด 17.73

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณโรงพยาบาลจิตเวชขอนแก่นราชนครินทร์ จังหวัดขอนแก่น ที่ให้ข้อมูลการรักษาผู้ป่วยโรคทางจิตเวช มาศึกษาในครั้งนี้ และคณะอาจารย์ เจ้าหน้าที่ภาควิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ในการช่วยผลักดัน ให้คำปรึกษาและข้อเสนอแนะ ให้งานวิจัยนี้ประสบความสำเร็จ

References

- Boonma, R. and Chirawichitchai, N. (2019). Classification of Diabetes Patient by Using Data Mining Techniques and Correlation Based Feature Selection. *PKRU SciTech Journal*, 3(2), 11-19. <https://ph01.tci-thaijo.org/index.php/pkruscitech/article/view/201722/164122>
- Boonprasom, C. and Sarach, C. (2019). The Exploratory Factor Analysis of Undergraduate Students' Dropout at Ubon Ratchathani Rajabhat University. *Technical Education Journal: King Mongkut's University of Technology North Bangkok*, 10(1), 86-97. <https://ojs.kmutnb.ac.th/index.php/jote/article/view/2909/2239>
- Chinprasatsak, K., Insuk, P., Foosarpok, N. and Sonpanow, N. (2023). *The Essence of DEEP LEARNING: AI, illustrated*. Made By AI Company Limited.
- Digital Government Development Agency. (2022). *AI Artificial Intelligence Technology for Government Administration and Services*. Sor. Pichit Printing Company Limited. (in Thai)
- Eibe, F. and Ian, W. (1998). *Generating Accurate Rule Sets Without Global Optimization*. Research Gate.
- Harnmongkolpipat, P. (2017). *Principles of Statistics 1* (7thed). Bangkok: Kasetsart University Press.
- Hongboonmee, N. and Thammakorn, T. (2019). Screening System for Depression on Smartphone using Data Mining Techniques. *Srinakharinwirot University Journal of Sciences and Technology*, 11(21), 100-113. <https://ejournals.swu.ac.th/index.php/SWUJournal/article/view/11425/9654>
- Kenji, K. and Larry, A.R. (1992). *The Feature Selection Problem: Traditional Methods and a New Algorithm*. <https://cdn.aaai.org/AAAI/1992/AAAI92-020.pdf>
- Khonkaen Rajanagarindra Psychiatric Hospital, National Center for Suicide. (2022). *E-Reports*. <https://suicide.dmh.go.th/report/>
- Molee, A., Saokaew, N. and Hemmanee, B. (2019). A Study of Disease Risk of Elderly with Data Mining Technique. *Narkbhutparitat Journal Nakhon Si Thammarat Rajabhat University*, 11(3), 29-34. (in Thai)

- Rungrattanaubol, J. (2022). *Data Mining Techniques*. Naresuan University Publishing House.
- Sanguansat, P. (2019). *Artificial Intelligence with Machine Learning, AI with Machine Learning*. Infopress
- Senthil Kumar, A.V. (2016). *Web Usage Mining Techniques and Applications Across Industries*. IGI Global, United States of America.
- World Health Organization. (2019). *Suicide Worldwide in 2019: Global Health Estimates*. WHO.
- World Health Organization. (2021). *One in 100 Deaths is by Suicide*. WHO. <https://www.who.int/news/item/17-06-2021-one-in-100-deaths-is-by-suicide>
- Xiaowei, H., Gaojie, J. and Wenjie, R. (2023). *Machine Learning Safety*. Singapore.