กฎการจำแนกภาวะไตวายเรื้อรังจากผู้ป่วยโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูล

Classification Rules of Chronic Renal Failure from Diabetic Patients Using Data Mining Techniques

ลียา เสียงเพราะ* และ จารี ทองคำ

Leeya Siangpho* and Jaree Thongkham

ภาควิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มหาสารคาม 44150 ประเทศไทย

Department of Information Technology, Faculty of Informatics, Mahasarakham University,

Maha Sarakham 44150, Thailand

*E-mail: kiizly78@gmail.com

Received: Apr 27, 2026

Revised: Jul 15, 2026

Accepted: Aug 03, 2026

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองของภาวะไตวายเรื้อรังจากผู้ป่วยโรคเบาหวาน ข้อมูลผู้ป่วยถูกรวบรวมจากฐานข้อมูลโรงพยาบาลภูเวียง อำเภอภูเวียง จังหวัดขอนแก่น จากวันที่ 1 มกราคม 2562 ถึง 31 ธันวาคม 2564 จำนวน 3,225 ระเบียน ผู้วิจัยคัดเลือกปัจจัยที่เกี่ยวข้องกับการพยากรณ์สำหรับการสร้างแบบจำลองด้วยวิธี Gain Ratio และ SMOTE จากนั้นนำเทคนิค RIPPER เทคนิค PART Decision List เทคนิคนาอีฟ เบย์ และเทคนิคฟูเรีย มาทำการสร้างแบบจำลองเพื่อพยากรณ์ภาวะไตวายเรื้อรังจากผู้ป่วยโรคเบาหวาน ในการทดลองนี้ 10-Fold Cross Validation ได้ถูกนำมาใช้ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลและประเมินแบบจำลอง เปรียบเทียบประสิทธิภาพด้วยค่าความถูกต้อง ค่าความไว และค่าความจำเพาะ ผลการวิจัยพบว่าเทคนิคฟูเรียที่ใช้ข้อมูลหลังจากการคัดเลือกปัจจัยและปรับสมดุลด้วยวิธีการ SMOTE ให้ประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้องร้อยละ 94.11 ค่าความไวร้อยละ 91.30 และค่าความจำเพาะร้อยละ 96.90 นอกจากนี้ แบบจำลองยังสามารถสร้างกฎการจำแนกภาวะไตวายเรื้อรังได้ โดยกฎที่ค้นพบส่วนใหญ่มีความสัมพันธ์กับโรคร่วม การได้รับยา และปัจจัยอื่น ๆ เช่น ดัชนีมวลกาย ระดับการศึกษา และอาชีพ ซึ่งเป็นปัจจัยสำคัญในการจำแนกผู้ป่วยที่มีความเสี่ยงต่อการเกิดภาวะไตวายเรื้อรังจากโรคเบาหวาน

คำสำคัญ: โรคเบาหวาน การคัดเลือกปัจจัย เหมืองข้อมูล ภาวะไตวายเรื้อรัง ฟูเรีย

Abstract

This research aimed to construct and compare the performance of models for predicting chronic kidney failure in diabetic patients. Patient data was collected from the database of Phu Wiang Hospital, Phu Wiang District, Khon Kaen Province, from January 1, 2019, to December 31, 2021, totaling 3,225 records. The researchers selected predictive factors using the Gain Ratio and SMOTE methods. Then, the RIPPER technique, PART Decision List, Naïve Bayes, and FURIA were used to construct models to predict chronic kidney failure in diabetic patients. In this experiment, 10-Fold Cross Validation was used to cluster the data and evaluate the models. The performance was compared using accuracy, sensitivity, and specificity. The results revealed that FURIA, using data after factor selection and balance adjustment using the SMOTE method, exhibited the best performance with an accuracy of 94.11%, a sensitivity of 91.30%, and a specificity of 96.90%. Moreover, the model was able to generate rules for classifying chronic kidney failure. Most of the discovered rules were related to comorbidities,

medication use, and factors such as body mass index, education level, and occupation, which were important factors in classifying patients at risk of developing chronic kidney failure from diabetes.

Keywords: Diabetes, Factor selection, Data mining, Chronic kidney failure, FURIA

1. บทนำ

โรคเบาหวานเป็นโรคเรื้อรังที่เกิดจากพฤติกรรมและเป็นปัญหาสาธารณสุขที่สำคัญของโลก โดยมีแนวโน้มจำนวนผู้ป่วยเพิ่มขึ้นอย่างต่อเนื่อง ภาวะแทรกซ้อนที่สำคัญคือโรคไตเรื้อรัง (Chronic kidney disease) ซึ่งอาจนำไปสู่ภาวะไตวายระยะสุดท้าย ส่งผลกระทบต่อคุณภาพชีวิตของผู้ป่วยและเพิ่มภาระค่าใช้จ่ายด้านสาธารณสุข นอกจากนี้ การใช้ยารักษาเบาหวานบางชนิดต่อเนื่องเป็นเวลานานอาจส่งผลต่อการทำงานของไต จึงจำเป็นต้องมีการบริหารจัดการยาให้เหมาะสม แม้ว่าปัจจุบันประชาชนจะตื่นตัวในการตรวจคัดกรองโรคไตมากขึ้น แต่ผู้ป่วยจำนวนมากยังไม่ทราบว่าตนเองมีภาวะโรคไตเรื้อรัง ดังนั้น การพัฒนาแบบจำลองเพื่อคัดกรองและทำนายความเสี่ยงของโรคไตเรื้อรังในผู้ป่วยเบาหวานตั้งแต่ระยะเริ่มต้น จึงมีความสำคัญต่อการวางแผนการรักษาและการป้องกันการดำเนินของโรคอย่างมีประสิทธิภาพ

เหมืองข้อมูล (Data mining) เป็นเทคนิคการวิเคราะห์เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล โดยทำการจำแนกประเภท แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น เพื่อจำแนกออกมาเป็นกฎและองค์ความรู้ใหม่ที่สามารถนำไปประกอบการตัดสินใจในด้านต่าง ๆ ได้ งานวิจัยหลายชิ้นได้แสดงให้เห็นถึงประสิทธิภาพของเทคนิคเหมืองข้อมูลในการจำแนกและการพยากรณ์ข้อมูลทางการแพทย์ เช่น วรารินทร์ ปัญญาวงษ์ได้วิจัยปัจจัยที่มีผลต่อการสำเร็จการศึกษาระดับปริญญาตรีของนิสิตคณะสถาปัตยกรรมศาสตร์ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคามโดยการทำเหมืองข้อมูลและสร้างแบบจำลองในการพยากรณ์การ [1] งานวิจัยนี้แสดงให้เห็นว่าเทคนิค RIPPER สามารถสร้างแบบจำลองที่มีประสิทธิภาพ (Precision) สูงสุด 91% ค่าความระลึก (Recall) เท่ากับ 91.10% ต้นแบบกฎการจำลองนี้สามารถนำไปใช้ในการพัฒนาระบบหรือโปรแกรมการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรีได้ จารี ทองคำ และวาทินี สุขมาก ได้พัฒนาแบบจำลองที่มีประสิทธิภาพในการพยากรณ์การรอดชีวิตของผู้ป่วยมะเร็งเต้านม [2] ซึ่งพบว่าเทคนิค PART Decision List มีความถูกต้องสูงสุดที่ 98.89% นิรินทร์ จิวิตัน ได้ทำการวิจัยเกี่ยวกับการเปรียบเทียบประสิทธิภาพในการทำนายความไม่สมดุลของข้อมูลการเสียชีวิตจากอุบัติเหตุบนโครงข่ายถนนจากการคำนวณน้ำหนักของข้อมูลที่มีความเกี่ยวข้องหรือสัมพันธ์ของแอทริบิวต์ (Attribute) ด้วยวิธีการ Gain Ratio และจำลองการพยากรณ์ด้วยวิธีการ K-Fold Cross Validation โดยแบ่งข้อมูลออกเป็น 30 พบว่าเทคนิคที่มีประสิทธิภาพจากจำแนกข้อมูลที่ดีที่สุดคือเทคนิคนาอีฟ เบย์ (Naïve Bayes) มีค่าความถูกต้องที่สุด 72.23% [3] Palacios et al. [4] ได้ทำการวิจัยเกี่ยวกับเด็กที่เป็นโรค Dyslexic ซึ่งพบว่าเทคนิคฟูเรีย (FURIA) มีข้อได้เปรียบจากเทคนิคอื่นในการคำนวณระหว่างขั้นตอนการเรียนรู้ โดยเป็นเทคนิคที่ใช้หลักการสุ่มกลุ่มตัวอย่างในการสร้างกฎความสัมพันธ์ที่มีความแม่นยำสูง ซึ่งแต่ละเทคนิคมีจุดเด่นในการสร้างแบบจำลองที่มีประสิทธิภาพสูงแตกต่างกันไป

จากปัญหาและแนวคิดดังกล่าวผู้วิจัยจึงสนใจที่จะศึกษาวิธีการคัดเลือกปัจจัยด้วยวิธีการ Gain Ratio และ SMOTE สร้างแบบจำลองเพื่อพยากรณ์การจำแนกภาวะไตวายเรื้อรังจากผู้ป่วยโรคเบาหวาน ผู้วิจัยได้นำเทคนิคที่เป็นที่นิยมและมีประสิทธิภาพในการสร้างแบบจำลองจำนวน 4 เทคนิค ได้แก่ เทคนิค RIPPER เทคนิค PART Decision List เทคนิคนาอีฟ เบย์ และเทคนิคฟูเรีย โดยใช้ 10 รอบของ 10-fold Cross Validation ในการแบ่งกลุ่มข้อมูลเป็นชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบเพื่อลดความลำเอียงของผลการทดลอง จากนั้นเปรียบเทียบประสิทธิภาพของแบบจำลองโดยพิจารณาจากค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity) เพื่อที่จะทราบว่าเทคนิคใดมีความสามารถสูงที่สุดในการจำแนกภาวะไตวายเรื้อรังของผู้ป่วยโรคเบาหวาน ซึ่งจะเป็นประโยชน์ต่อการช่วยให้ผู้ป่วยในระยะที่ 1-2 เข้าสู่กระบวนการรักษาได้อย่างทันท่วงที

2. วัสดุ อุปกรณ์และวิธีการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้อง เพื่อใช้เป็นแนวทางในการออกแบบกระบวนการวิเคราะห์ข้อมูลและสร้างแบบจำลอง โดยกระบวนการดำเนินงานวิจัย 4 ขั้นตอน ประกอบด้วย การเตรียมข้อมูล ขั้นตอนก่อนการสร้างแบบจำลอง การสร้างแบบจำลอง และขั้นตอนการวัดประสิทธิภาพของแบบจำลอง

2.1. การเตรียมข้อมูล

การเตรียมข้อมูลในงานวิจัยครั้งนี้ได้นำข้อมูลผู้ป่วยเบาหวานจากฐานข้อมูลโรงพยาบาลภูเวียง อำเภอภูเวียง จังหวัดขอนแก่น จากวันที่ 1 มกราคม 2562 ถึง 31 ธันวาคม 2564 จำนวน 3,225 ระเบียน จำนวน 18 ปัจจัย โดยปัจจัยบางประเภทมีการแปลงเป็นตัวเลข เช่น ได้รับยาแปลงเป็น 1 ไม่ได้รับยาแปลงเป็น 0 จากการศึกษางานวิจัยที่เกี่ยวข้องกับปัจจัยและสาเหตุที่ก่อให้เกิดภาวะไตวายเรื้อรังจากผู้ป่วยโรคเบาหวาน ทำให้ผู้วิจัยมีแนวคิดที่จะศึกษาปัจจัยที่มีแนวโน้มจะเกิดภาวะไตวายเรื้อรัง โดยอ้างอิงจากตัวแปรงานวิจัยที่เกี่ยวข้อง และจากการสอบถามผู้ที่มีความเชี่ยวชาญในการทำวิจัยเกี่ยวกับผู้ป่วยโรคเบาหวานและเหมืองข้อมูล ซึ่งสามารถแบ่งเป็นตัวแปรผู้ป่วยเบาหวานนามบัญญัติดัง Table 1 และตัวแปรผู้ป่วยเบาหวานชนิดตัวเลขดัง Table 2 โดย Table 1 เป็นการนำข้อมูลตัวแปรที่ใช้ในงานวิจัยมาแสดงปัจจัยประเภทนามบัญญัติ (Nominal attributes) ในการจัดข้อมูลเป็นกลุ่มต่าง ๆ ที่แยกออกจากกัน ได้แก่การศึกษา อาชีพ ยารักษาโรค โรคร่วมชนิดที่ 1 โรคร่วมชนิดที่ 2 และคลาส ซึ่งแสดงจำนวนค่าปัจจัยในการแยกจำนวนปัจจัย Table 2 แสดงปัจจัยประเภทค่าตัวเลข (Numeric attributes) ที่ใช้แทนจำนวนที่สามารถนำไปคำนวณได้แก่ ความดันช่วงบน ความดันช่วงล่าง อัตราการหายใจ อัตราชีพจร ดัชนีมวลกาย เพื่อนำมาใช้ในการคำนวณหาค่าต่ำ ค่ากลาง และค่าส่วนเบี่ยงเบนมาตรฐาน เครื่องมือที่นำมาใช้ในการเตรียมข้อมูลได้แก่ โปรแกรม Host XP และใช้โปรแกรม Microsoft Excel ในการจัดรูปแบบข้อมูล เพื่อนำข้อมูลเข้าสู่โปรแกรม Weka ผู้วิจัยได้ใช้เทคนิค Gain Ratio เพื่อคัดเลือกตัวแปรที่มีความสัมพันธ์กับการเกิดไตวายเรื้อรังมากที่สุดก่อนนำไปสร้างแบบจำลองพยากรณ์ส่งผลให้เหลือตัวแปรจำนวน 14 ปัจจัย

Table 1 Nominal attributes

No.	Variables	Details	Number of value
1	Education	การศึกษา	10
2	Occupation	อาชีพ	19
3	Alcohol	พฤติกรรมการดื่มแอลกอฮอล์ของผู้ป่วย	2
4	amLODIPine	ยารักษาโรค	2
5	ENALApril	ยารักษาโรค	2
6	glipiZIDE	ยารักษาโรค	2
7	metFORMIN	ยารักษาโรค	2
8	MIXtard	ยารักษาโรค	2
9	Simvastatin	ยารักษาโรค	2
10	Vitamin	ยารักษาโรค	2
11	Dx0	โรคร่วมชนิดที่ 1	130
12	Dx1	โรคร่วมชนิดที่ 2	118
13	Class	ผลการรักษา (ผลการวินิจฉัยของแพทย์ว่าเป็นหรือไม่เป็นโรคไตวายเรื้อรัง)	2

Table 2 Numeric attributes

No.	Variables	Details	Minimum	Maximum	Mean	SD
1	Bps	ความดันช่วงบน	130	150	139.52	4.85
2	Bpd	ความดันช่วงล่าง	89	100	94.52	2.68
3	Rr	อัตราการหายใจ	18	22	19.58	1.03
4	Pulse	อัตราชีพจร	70	90	78.96	4.72
5	Bmi	ดัชนีมวลกาย	15.11	50.296	26.373	4.28

2.2. การคัดเลือกปัจจัย

ก่อนการสร้างแบบจำลอง ผู้วิจัยได้ดำเนินการตรวจสอบคุณภาพของข้อมูล (Data cleaning) โดยตรวจสอบค่าที่ขาดหาย (Missing values) ตรวจสอบค่าผิดปกติ (Outliers) และตรวจสอบความถูกต้องของข้อมูลทุกตัวแปร หากพบค่าที่ผิดปกติทางการแพทย์ ผู้วิจัยจะตรวจสอบความถูกต้องกับข้อมูลจากเวชระเบียน และตัดข้อมูลดังกล่าวออกจากชุดข้อมูลก่อนนำไปวิเคราะห์ เพื่อป้องกันผลกระทบต่อประสิทธิภาพของแบบจำลอง จากนั้นจึงนำข้อมูลเข้าสู่กระบวนการคัดเลือกปัจจัยด้วยวิธี Gain Ratio และปรับสมดุลข้อมูลด้วยเทคนิค SMOTE โดยกระบวนการคัดเลือกปัจจัยและการปรับสมดุลข้อมูลดำเนินการเฉพาะภายในชุดข้อมูลฝึกสอน (Training set) ของแต่ละรอบในการประเมินแบบ 10-fold Cross Validation เพื่อป้องกันการรั่วไหลของข้อมูล (Data leakage) และลดความเอนเอียงในการประเมินประสิทธิภาพของแบบจำลอง

2.2.1. วิธีการ Gain Ratio

วิธีการ Gain Ratio เป็นตัวชี้วัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจาก Information Gain เนื่องจากการใช้ Information Gain ในการแบ่งชุดข้อมูลจะมีโอกาสทำให้เกิดความเอนเอียงขึ้น [5] เมื่อคุณลักษณะที่ทำการพิจารณาได้ค่า Gain ที่สูงเป็นจำนวนมาก ทำให้คุณลักษณะที่ถูกคัดเลือกไม่ถูกต้อง ตัวอย่างเช่น การพิจารณาปัจจัยที่ทำหน้าที่เป็นตัวระบุเฉพาะ เช่น อัตราส่วนกำไร เมื่อมีการเปลี่ยนแปลงโครงสร้างหุ้นส่วน ซึ่งส่งผลให้มีชุดข้อมูลย่อยจำนวนมาก แต่ละส่วนข้อมูลย่อยมีเพียง 1 Record ซึ่งเหมือนค่า Information Gain จะได้ค่าที่สูงจำนวนมาก จากความเอนเอียง ทำให้มีการพัฒนาตัวชี้วัดการแบ่งข้อมูลใหม่ที่เรียกว่า Gain Ration โดยการประยุกต์ใช้การทำนอร์มัลไลซ์ (Normalize) ค่า Information Gain ด้วยการใช้ค่า “Split Information” ซึ่งสามารถคำนวณได้ดังสมการ 1

$$Splitinfo_A(D) = -\sum_{j=1}^{m} \frac{|D_j|}{|D|} x log_2 \frac{|D_j|}{|D|} \qquad (1)$$

2.2.2. วิธีการ SMOTE

เทคนิคในการสุ่มข้อมูลตัวอย่างของคลาส เป็นเทคนิคทางสถิติที่ใช้ในการจัดการกับปัญหาข้อมูลที่ไม่สมดุลกัน (Imbalance data) [6] ข้อมูลในการทดลองครั้งนี้มีข้อมูลคลาส 1 เพียง 2470 รายการ ส่วนคลาส 0 มี 570 รายการ ซึ่งแตกต่างกันเป็นจำนวนมาก มีผลกระทบต่อประสิทธิภาพในการจำแนกข้อมูล SMOTE จึงมาแก้ไขปัญหาในชุดข้อมูลที่ไม่สมดุลกันเพิ่มจำนวนข้อมูลในคลาสที่มีปริมาณข้อมูลน้อย สามารถคำนวณได้ดังสมการ 2

$$x^i = x + rand(0,1) * |x - x_k| \qquad (2)$$

เมื่อ

x^i	แทน	ข้อมูลที่สังเคราะห์ขึ้นมาใหม่
x	แทน	ข้อมูลเดิมที่สุ่มมา
$rand(0,1)$	แทน	ตัวเลขที่สุ่มระหว่าง 0 ถึง 1
x_k	แทน	ข้อมูลที่เป็นเพื่อนบ้านของ x

2.3. การสร้างแบบจำลอง

การสร้างแบบจำลองในงานวิจัยนี้ได้นำโปรแกรม WEKA มาสร้างแบบจำลอง 4 เทคนิค ประกอบด้วย เทคนิค RIPPER เทคนิค PART Decision List เทคนิคนาอีฟ เบย์ และเทคนิคฟูเรีย ดังนี้

เทคนิค RIPPER เป็นเทคนิคการสร้างแบบจำลองเชิงกฎที่พัฒนาจากอัลกอริทึม IREP เพื่อเพิ่มประสิทธิภาพในการจำแนกข้อมูลและลดข้อผิดพลาดในกระบวนการเรียนรู้แบบมีผู้สอน [7] โดยมีขั้นตอนการสร้างกฎ (Growing) สร้างกฎการตัดสินใจ (Decision rules) และการตัดแต่งกฎ (Pruning phase) ซ้ำหลายรอบเพื่อให้ได้กฎที่ดีที่สุด

เทคนิค PART Decision List เป็นเทคนิคการจำแนกข้อมูลแบบอิงกฎ (Rule-based-classification) [7] พัฒนามาจาก Decision Tree (C4.5) และการสร้างกฎ (Rule learning) ที่มีประสิทธิภาพโดยแต่ละรอบจะเลือกกิ่งของต้นไม้ที่ให้ความแม่นยำสูงมาแปลงเป็นกฎ และนำข้อมูลที่ถูกครอบคลุมโดยชุดกฎนั้นออกจากชุดข้อมูล ทำซ้ำขั้นตอนเดิมจนกว่าจะไม่สามารถสร้างกฎเพิ่มเติมได้

เทคนิคนาอีฟ เบย์ เป็นเทคนิคการจำแนกประเภทที่อิงจากทฤษฎีความน่าจะเป็นโดยมีการตั้งสมมติฐานคุณลักษณะ (Features) ทุกตัวเป็นอิสระต่อกันอย่างสมบูรณ์เมื่อกำหนดคลาส [8] อัลกอริทึมนี้คำนวณความน่าจะเป็นของแต่ละคลาสตามข้อมูลที่ปรากฏ เพื่อทำนายคลาสที่มีความน่าจะเป็นสูงสุด เหมาะสำหรับงานจำแนกข้อมูลที่มีขนาดใหญ่และต้องการความรวดเร็วในการคำนวณ

เทคนิคฟูเรีย เป็นเทคนิคการเรียนรู้กฎที่พัฒนาจากแนวคิด Fuzzy Logic [4] โดยสร้างชุดกฎแบบไม่เรียงลำดับ

(Unordered rules) [9] เพื่อจำแนกข้อมูลได้อย่างยืดหยุ่น เหมาะกับข้อมูลที่คลุมเครือหรือไม่แน่นอน ขั้นตอนการทำงาน คือสร้างต้นไม้ตัดสินใจจากข้อมูล แปลงกิ่งของต้นไม้เป็นกฎแบบ Fuzzy จากนั้นปรับแต่งและตัดแต่งกฎให้เหมาะสม เพื่อใช้ในการจำแนกข้อมูลใหม่อย่างมีประสิทธิภาพ

2.4. การวัดประสิทธิภาพของแบบจำลอง

การวัดประสิทธิภาพของการพยากรณ์การเป็นภาวะไตวายเรื้อรังจากผู้ป่วยเบาหวานโดยใช้หลักการ 10-fold Cross Validation ในการพยากรณ์ข้อมูลออกเป็นสองชุด ประกอบด้วยข้อมูล 9 ชุดเป็นข้อมูลฝึกสอนและชุด 1 ทดสอบ โดยจะทำการทดสอบทั้งหมด 10 รอบ เพื่อหาค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และค่าความจำเพาะ (Specificity)

ค่าความถูกต้อง (Accuracy) คือ ค่าที่แบบจำลองสามารถจำแนกข้อมูลผู้ป่วยที่เกิดโรคและไม่เกิดโรคได้อย่างถูกต้องต่อข้อมูลทั้งหมด ดังสมการที่ 3

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (3)$$

ค่าความไว (Sensitivity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยที่เกิดโรคได้อย่างถูกต้องต่อผู้ป่วยที่เกิดโรคจริง ดังสมการที่ 4

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (4)$$

ค่าความจำเพาะ (Specificity) คือ ค่าที่แบบจำลองสามารถพยากรณ์ข้อมูลผู้ป่วยไม่เกิดโรคได้ อย่างถูกต้องต่อผู้ป่วยที่พยากรณ์ว่าเกิดโรค ดังสมการที่ 5

$$Specificity = \frac{TN}{(TN + FP)} \quad (5)$$

เมื่อ

TP คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์การเกิดโรคไตวายเรื้อรังได้อย่างถูกต้อง

TN คือ จำนวนข้อมูลที่แบบพยากรณ์การไม่เกิดโรคไตวายเรื้อรังได้อย่างถูกต้อง

FP คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์การเกิดโรคไตวายเรื้อรังได้ไม่ถูกต้อง

FN คือ จำนวนข้อมูลที่แบบจำลองพยากรณ์การไม่เกิดโรคไตวายเรื้อรังได้ไม่ถูกต้อง

2.5. จริยธรรมการวิจัย

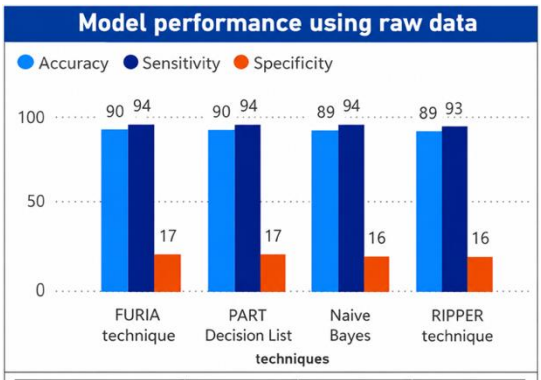
งานวิจัยนี้ผ่านการรับรองจริยธรรมการวิจัยในคน จากมหาวิทยาลัยมหาสารคาม เลขที่การรับรอง 298-357/2566 วันที่รับรอง 7 สิงหาคม 2566

3. ผลการวิจัยและอภิปรายผลการวิจัย

ผลการวิเคราะห์กฎการจำแนกภาวะไตวายเรื้อรังจากผู้ป่วยเบาหวาน ได้วิเคราะห์ปัจจัยของเกิดภาวะไตวายเรื้อรังของผู้ป่วยโรคเบาหวาน จากปัจจัยทั้งหมด 18 ปัจจัย ประกอบด้วย การศึกษา อาชีพยารักษาโรคจำนวน 8 ชนิด โรคร่วมชนิดที่ 1 โรคร่วมชนิดที่ 2 ผลการรักษา (ผลการวินิจฉัยของแพทย์ว่าเป็นหรือไม่เป็นโรคไตวายเรื้อรัง) ความดันช่วงบน ความดันช่วงล่าง อัตราการหายใจ อัตราชีพจร ดัชนีมวลกาย ผู้วิจัยได้ใช้วิธีการ Gain Ratio และ SMOTE ซึ่งเป็นวิธีการที่มีประสิทธิภาพสูงในการวิจัย

3.1. ประสิทธิภาพของแบบจำลองจากข้อมูลดิบ

การสร้างแบบจำลองจากข้อมูลดิบ พบว่าฟูเรียให้ค่าความถูกต้องสูงสุด 90.20% ส่วน RIPPER ให้ค่าความถูกต้องต่ำสุด 88.52% สำหรับนาอีฟ เบย์ ให้ค่าความไวสูงสุด 93.77% และ RIPPER ให้ค่าความไวต่ำสุด 92.59% ในขณะที่ค่าความจำเพาะยังต่ำมาก โดยฟูเรียให้ค่าความจำเพาะสูงสุด 17.37% และนาอีฟ เบย์ ให้ค่าความจำเพาะต่ำสุด 15.90% (Figure 1) สะท้อนว่าแบบจำลองสามารถจับผู้ป่วยได้ดี แต่มีแนวโน้มพลาดทำนายผู้ป่วยกลุ่มเล็ก

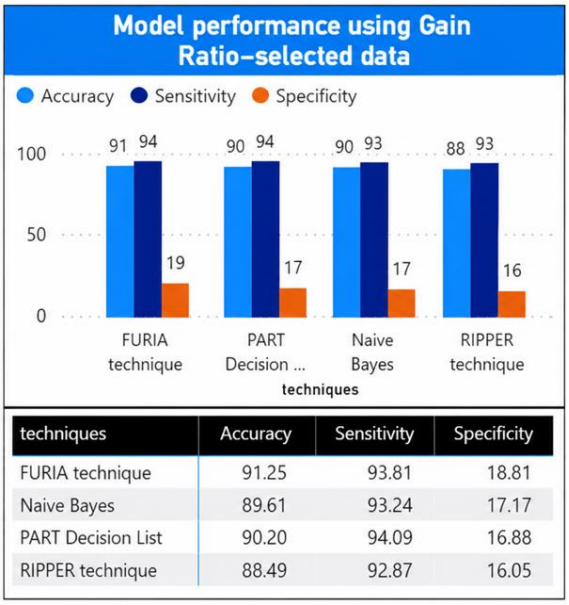


techniques	Accuracy	Sensitivity	Specificity
FURIA technique	90.20	93.72	17.37
Naive Bayes	89.21	93.77	15.90
PART Decision List	89.67	93.56	16.83
RIPPER technique	88.52	92.59	16.47

Figure 1 Model performance using raw data

3.2. ประสิทธิภาพของแบบจำลองจากข้อมูลที่ผ่านกระบวนการคัดเลือกข้อมูล

หลังจากการคัดเลือก 14 ปัจจัยที่มีน้ำหนักสูงสุดโดยใช้วิธี Gain Ratio พบว่าแบบจำลองมีค่าความถูกต้องเพิ่มขึ้นเล็กน้อย ฟูเรียให้ค่าความถูกต้องสูงสุด 91.25% และ PART Decision List ให้ค่าความถูกต้องต่ำสุด 88.49% ส่วน PART Decision List ให้ค่าความไวสูงสุด 94.09% และ RIPPER ให้ค่าความไวต่ำสุด 92.87% ส่วนค่าความจำเพาะ ฟูเรียให้ค่าความจำเพาะสูงสุด 18.81% และ RIPPER ให้ค่าความจำเพาะต่ำสุด 16.05% (Figure 2) การคัดเลือกปัจจัยช่วยลดความเอนเอียงของแบบจำลอง ทำให้สมดุลระหว่างการทำนายผู้ป่วยและผู้ไม่ป่วยดีขึ้น แต่ยังมีข้อจำกัดเรื่องค่าความจำเพาะต่ำ

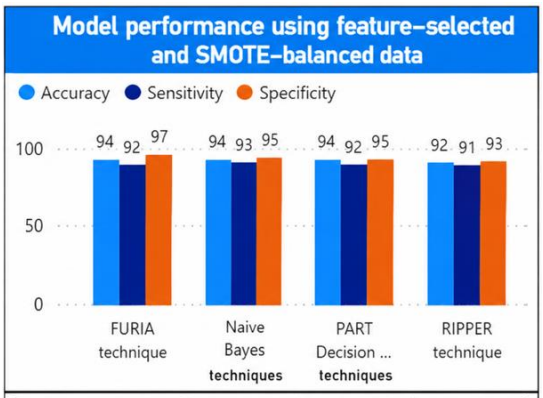


techniques	Accuracy	Sensitivity	Specificity
FURIA technique	91.25	93.81	18.81
Naive Bayes	89.61	93.24	17.17
PART Decision List	90.20	94.09	16.88
RIPPER technique	88.49	92.87	16.05

Figure 2 Model performance using Gain Ratio-selected data

3.3. ประสิทธิภาพของแบบจำลองของข้อมูลที่สมดุล

เพื่อแก้ปัญหาข้อมูลไม่สมดุล ผู้วิจัยได้ใช้เทคนิค SMOTE เพิ่มข้อมูลผู้ป่วยกลุ่มเล็กให้ใกล้เคียงกับข้อมูลกลุ่มใหญ่ ผลการทดลองพบว่าค่าความถูกต้องและความจำเพาะเพิ่มขึ้นอย่างชัดเจน โดยฟูเรียให้ค่าความถูกต้องสูงสุด 94.11% และค่าความจำเพาะสูงสุด 96.74% ส่วน RIPPER ให้ค่าความถูกต้องต่ำสุด 92.36% และค่าความจำเพาะต่ำสุด 93.37% ส่วน นาอีฟ เบย์ ให้ค่าความไวสูงสุด 92.61% และ RIPPER ให้ค่าความไวต่ำสุด 91.38% (Figure 3) การปรับสมดุลข้อมูลช่วยให้แบบจำลองทำนายผู้ป่วยและผู้ไม่ป่วยได้แม่นยำมากขึ้น



techniques	Accuracy	Sensitivity	Specificity
FURIA technique	94.11	91.73	96.74
Naive Bayes	93.82	92.61	95.09
PART Decision List	93.68	92.36	95.07
RIPPER technique	92.36	91.38	93.37

Figure 3 Model performance using feature-selected and SMOTE-balanced data

3.4. กฎการจำแนกภาวะไตวายเรื้อรังที่ได้จากเทคนิคฟูเรีย

เทคนิคฟูเรียสามารถสร้างกฎการจำแนก (Classification rules) ในรูปแบบของ If-Then ซึ่งช่วยให้สามารถอธิบายการตัดสินใจของแบบจำลองได้อย่างชัดเจน จากการทดลองพบว่ามีทั้งหมด 12 กฎ โดยยกตัวอย่างกฎที่สำคัญแสดงดัง Table 3

4. บทสรุป

จากการวิจัยพบว่าเทคนิคฟูเรียร่วมกับเทคนิค SMOTE ให้ประสิทธิภาพดีที่สุด โดยมีค่าความถูกต้อง 94.11% และค่าความจำเพาะ 96.74% สูงกว่าเทคนิคเปรียบเทียบ เนื่องจากเทคนิคฟูเรียสามารถจัดการข้อมูลที่มีความไม่แน่นอนและความสัมพันธ์ที่ซับซ้อนได้อย่างมีประสิทธิภาพ ขณะที่ SMOTE ช่วยแก้ปัญหาความไม่สมดุลของข้อมูล ส่งผลให้แบบจำลองจำแนกผู้ป่วยที่ไม่เป็นโรคได้แม่นยำยิ่งขึ้น แม้ว่าค่าความไวจะลดลงเล็กน้อย แต่ยังคงอยู่ในระดับที่เหมาะสมสำหรับงานด้านการแพทย์ และผลการศึกษาสอดคล้องกับงานวิจัยที่ผ่านมา จึงแสดงให้เห็นว่าเทคนิคฟูเรียร่วมกับเทคนิค SMOTE มีศักยภาพในการพัฒนาระบบสนับสนุนการตัดสินใจทางคลินิกสำหรับการคัดกรองภาวะไตวายเรื้อรังในผู้ป่วยเบาหวาน และงานวิจัยนี้ใช้ข้อมูลจากโรงพยาบาลเพียงแห่งเดียว ทำให้การประยุกต์ใช้กับประชากรอื่นอาจมีข้อจำกัด อีกทั้งยังไม่ได้เปรียบเทียบกับ

Table 3 Classification rules

Rule	If	Then	CF
Rule 1	ถ้าผู้ป่วยมีโรคร่วมชนิดที่ 2 (dx1 = I10)	Class = yes	1.00
Rule 2	ถ้าผู้ป่วยมีโรคร่วมชนิดที่ 1 (dx0 = I10)	Class = yes	1.00
Rule 3	ถ้าผู้ป่วยได้รับยา ENALApril	Class = yes	0.99
Rule 4	ถ้าผู้ป่วยได้รับยา amLODIPine	Class = yes	0.99
Rule 5	ถ้าผู้ป่วยได้รับ Vitamin และ MIXtard พร้อมกัน	Class = yes	0.99

อัลกอริทึมเชิงลึก (Deep learning) ในอนาคตควรนำแบบจำลองไปทดสอบกับโรงพยาบาลหลายแห่ง และพัฒนาเป็นระบบ Clinical decision support system

5. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณโรงพยาบาลภูเวียง จังหวัดขอนแก่น ที่ให้ข้อมูลของผู้ป่วยโรคเบาหวานมาศึกษาวิจัยในครั้งนี้ และคณาจารย์ คณะวิทยาการสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม ในการช่วยผลักดันและให้คำปรึกษาแนะนำต่าง ๆ ให้งานวิจัยครั้งประสบความสำเร็จ

6. References

[1] Panyawong, W. 2020. Utilization of rule-based predicting models Faculty of Architecture, Urban Design and Creative Arts students at Mahasarakham University by data mining. **Journal of Architecture, Design and Construction.** 2(2): 109-119. (*in Thai*)

[2] Thongkham, J. and Sukmak, V. 2021. Predicting breast cancer patient survival. **RMUTI Journal Science and Technology.** 14(1): 44-54. (*in Thai*)

[3] Jiwitan, N. 2022. Comparison of imbalanced data classification efficiency to predict road accident death using data mining techniques. **Information Technology Journal KMUTNB.** 18(1): 75-84. (*in Thai*)

[4] Palacios, A. and et al. 2016. An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia. **Neurocomputing.** 176: 60-71.

[5] Samart, S. and Thongkam, T. 2025. Prediction models of the sefe-harm for psychiatric patients using data mining techniques. **Research on Modern Science and Utilizing Technological Innovation Journal.** 18(1): 110-119. (*in Thai*)

[6] Vasuarayasak, W. and et al. 2025. Enhancing machine learning performance on imbalanced resume data using SMOTE for job candidate classification. **Journal of Computer and Creative Technology.** 13(1): 38-48. (*in Thai*)

[7] Nanaumphai, P. and Thongkam, J. 2019. Intrusion detection using classification techniques in data mining. **Journal of Technology Management Rajabhat Maha Sarakham University.** 6(2): 111-118. (*in Thai*)

[8] Limsripraphan, P. and et al. 2019. Algorithm for mango classification using image processing and Naive Bayes classifier. **Industrial Technology Lampang Rajabhat University Journal.** 12(1): 112-125. (*in Thai*)

[9] Klangnok, P. 2018. **Applying the Ensemble Technique for Improving Rule-based Models Performance.** M.Sc. Thesis, Mahasarakham University. (*in Thai*)