

การพยากรณ์โรคเบาหวานด้วยเทคนิคเหมืองข้อมูล

Predicting Diabetes Mellitus by Using Data Mining Techniques

กฤตกนก ศรีพิมพ์สอ^{1*} และ กิตติพล วิแสง²

Kitkanog Sripimsor^{1*} and Kittipol Wisaeng²

สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม^{1,2}

Business Administration and Digital Innovation Mahasarakham Business School Mahasarakham University

E-Mail : 62010982001@msu.ac.th^{1*}, kittipol.w@acc.msu.ac.th²

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาแบบจำลองสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล และ 2) เปรียบเทียบประสิทธิภาพแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน เพื่อสร้างแบบจำลองในการพยากรณ์ผู้ป่วยและนำค่าวัดประสิทธิภาพของการจำแนกประเภทข้อมูลมาเปรียบเทียบโดยค่าความถูกต้อง (Accuracy) ที่ให้ค่ามากที่สุด กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ คือ ผู้ป่วยที่ได้รับการคัดกรองโรคเบาหวาน จำนวน 300 รายการ และกลุ่มผู้ป่วยที่เป็นโรคเบาหวาน จำนวน 100 รายการ รวมทั้งสิ้น 400 รายการ เครื่องมือที่ใช้ในการวิจัย ได้แก่แบบสอบถามและวิเคราะห์ข้อมูลผ่านโปรแกรม Waikato Environment for Knowledge Analysis โดยใช้เทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น เทคนิคนาอิว เบย์ เทคนิคโลจิสติก เทคนิคแรนดอมฟอเรส และเทคนิคต้นไม้ตัดสินใจ สถิติที่ใช้ในการวิจัย ได้แก่ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล และเส้นกราฟ

ผลการวิจัยพบว่าเทคนิคแรนดอมฟอเรส (Random Forest) ให้ค่าความถูกต้องในการทำนายผลเป็นโรคเบาหวานมากที่สุดอยู่ที่ 99.75% มีค่าความแม่นยำ (Precision) ที่ 98.50% ค่าความครบถ้วน (Recall) ได้ 98.50% ค่าวัดประสิทธิภาพโดยรวม (F-Measure) ได้ 98.50% และค่าเส้นกราฟ (ROC) ได้ 95.20% สามารถนำผลที่ได้จากงานวิจัยนี้ไปประยุกต์ใช้ในการประกอบการรักษาผู้ป่วยโรคเบาหวานต่อไปในอนาคต

คำสำคัญ : โรคเบาหวาน, การทำเหมืองข้อมูล, การควบคุมระดับน้ำตาลในเลือด, ผู้สูงอายุ, ความแม่นยำ

ABSTRACT

This research aims to 1) develop a model for diabetic patient prognosis using data mining techniques and 2) compare the performance of the most suitable model for diabetic patient prognosis to create a prognostic model. Furthermore, compare the performance measures of data classification by the Accuracy (Accuracy) that gives the highest value. The sample used in this study was 300 items of patients who received screening for diabetes and 100 items of patients with diabetes, totaling 400 items. The research tools were questionnaires, and data were analyzed through the Waikato Environment for Knowledge Analysis program using multilayer perceptron, Naive Bay, logistics techniques Random Forest Technique, and decision tree technique. The research instruments tools include accuracy, precision, recall, F-measure, and ROC curve.

The results showed that the Random Forest technique gave the highest accuracy in predicting the outcome of diabetes at 99.75% with a precision of 98.50%. The Recall was 98.50%. The overall efficiency measurement (F-Measure) was 98.50%, and the curve value (ROC) was 95.20%. The results of this research could be applied in treating diabetic patients in the future.

Keywords : Diabetes, Data Mining, Blood Sugar Control, Elderly, Accuracy

บทนำ

เบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสำคัญทางด้านสาธารณสุขของโลก มีการประเมินว่ามีผู้ป่วยเบาหวานรวม 425 ล้านรายทั่วโลก แบ่งเป็นกลุ่มผู้สูงอายุ 65 ปีขึ้นไป จำนวน 98 ล้านราย และช่วงอายุ 20-64 ปี จำนวน 327 ล้านราย และคาดการณ์ว่าในปี พ.ศ. 2588 จะมีผู้ป่วยเบาหวานจำนวน 629 ล้านรายทั่วโลก โดยแบ่งเป็นผู้สูงอายุ 65 ปีขึ้นไป จำนวน 191 ล้านราย หรือมีอัตราการเพิ่มอยู่ที่ร้อยละ 94.8 และในช่วงอายุ 20-64 ปี จำนวน 438 ล้านราย หรือมีอัตราการเพิ่มอยู่ที่ร้อยละ 33.9 [1]

โรคเบาหวานเป็นโรคเรื้อรังที่เกิดจากความผิดปกติของร่างกายที่มีการผลิตฮอร์โมนอินซูลินไม่เพียงพอทำให้ระดับน้ำตาลในเลือดสูง ผู้ป่วยเบาหวานมีอัตราการเพิ่มขึ้นอย่างต่อเนื่องทุกปี ผู้ป่วยโรคเบาหวานที่ไม่สามารถควบคุมระดับน้ำตาลในเลือดได้จะทำให้เกิดภาวะแทรกซ้อนของระบบหัวใจและหลอดเลือด ส่งผลต่อการเสียชีวิตและคุณภาพ เกิดภาระค่าใช้จ่ายด้านการรักษาพยาบาลที่เพิ่มขึ้น [2]

สาเหตุที่แท้จริงของโรคที่เกิดขึ้นยังไม่ทราบแน่ชัด แต่จากการศึกษามีปัจจัยหลายอย่างที่เกี่ยวข้องทั้งสาเหตุจากพันธุกรรม สิ่งแวดล้อมและพฤติกรรมสุขภาพ สำหรับด้านพฤติกรรมสุขภาพ พบว่าสาเหตุสำคัญมาจากพฤติกรรมที่ไม่เหมาะสม ดังนั้นการลดพฤติกรรมเสี่ยงและปรับพฤติกรรมในด้านการออกกำลังกาย การรับประทานอาหาร นอกจากจะช่วยป้องกันการเกิดโรคเบาหวานแล้ว ยังช่วยให้ลดภาระค่าใช้จ่าย ลดความทุกข์ทรมานของผู้ป่วยที่เป็นโรคอย่างยาวนาน และส่งเสริมการมีคุณภาพชีวิตที่ดีขึ้นด้วย [3]

จากสถานการณ์โรคเบาหวานในจังหวัดมหาสารคาม ตั้งแต่ปี พ.ศ. 2556 - 2561 มีจำนวนผู้ป่วยโรคเบาหวานเพิ่มมากขึ้น จากปี 2554 มีจำนวนในผู้ป่วยเบาหวาน 9,409 คน เมื่อเวลาผ่านไป 5 ปี พบว่ามีจำนวนผู้ป่วยเบาหวานเพิ่มขึ้นเป็น 13,431 คน ในปี พ.ศ. 2558 [4] สำหรับอำเภอกันทรวิชัย จังหวัดมหาสารคาม มีจำนวน 183 หมู่บ้าน 28,888 ครัวเรือน มีจำนวนประชากร 86,654 คน ปี พ.ศ. 2560 - 2563 อัตราป่วยด้วยโรคเบาหวาน 1,033.9, 2,141.8, 3,548.5, และ 3,751.7 ต่อประชากรแสนคน ตามลำดับ พบอัตราความชุกของโรคเบาหวานในประชากรที่มีอายุตั้งแต่ 35 ปีขึ้นไป ร้อยละ 4.90, 4.96, 5.44 และ 7.30 ตามลำดับ ซึ่งจะพบว่ามีผู้ป่วยเบาหวานรายใหม่เพิ่มขึ้นทุกปี จากการทบทวนวรรณกรรมพบว่า ความรู้เรื่องโรคเบาหวานมีความสัมพันธ์ทางบวกกับพฤติกรรมการจัดการตนเองและสามารถทำนายพฤติกรรมการจัดการตนเองได้ในกลุ่มผู้สูงอายุที่มีภาวะเสี่ยงต่อการเกิดโรคหัวใจและหลอดเลือด การได้รับการสนับสนุนจากบุคคลใกล้ชิด แพทย์ พยาบาล และทีมสุขภาพในกลุ่มผู้สูงอายุโรคเบาหวานชนิดที่ 2 มีความสัมพันธ์กับการปฏิบัติตัวของผู้ป่วยเบาหวาน [5] ในการวิจัยครั้งนี้จะกล่าวถึงผู้ป่วยเบาหวานจากกลุ่มเสี่ยงเบาหวาน อำเภอกันทรวิชัย จังหวัดมหาสารคาม เนื่องจากการตรวจสอบเกณฑ์สถิติผู้ป่วยเบาหวานรายใหม่จากกลุ่มเสี่ยงเบาหวาน มีเกณฑ์ที่เกินมาตรฐาน [6]

ด้วยสาเหตุดังกล่าวข้างต้น ผู้วิจัยจึงให้ความสนใจในการพยากรณ์โรคเบาหวานด้วยเทคนิคเหมืองข้อมูลของผู้มาตรวจร่างกายในโรงพยาบาลกันทรวิชัย ซึ่งการเปรียบเทียบประสิทธิภาพในการพยากรณ์การเป็นโรคเบาหวานขึ้นอยู่กับปัจจัยหลายอย่าง เช่น เพศ อายุ ระดับการศึกษา นับถือศาสนา สถานภาพ ดัชนีมวลกาย ความยาวเส้นรอบเอว ประวัติเป็นความดันเลือดสูง ประวัติการดื่มสุรา ประวัติการสูบบุหรี่ ครอบครัวหรือญาติสนิทมีประวัติการเป็นเบาหวาน ส่วนสูง น้ำหนัก การรับประทานอาหาร การออกกำลังกาย การรับประทานยา ฯลฯ โดยผู้วิจัยสนใจศึกษาการพยากรณ์โรคเบาหวานด้วยเทคนิคเหมืองข้อมูล ของผู้มาตรวจร่างกายที่โรงพยาบาล 5 วิธี คือ วิธีเพอร์เซ็ปตรอนแบบหลายชั้น วิธีนาอิว เบย์ วิธีโลจิสติก วิธีแรนดอมฟอเรท และวิธีต้นไม้ตัดสินใจ เพื่อหาวิธีที่มีประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานที่เหมาะสมต่อไป

1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล
- 1.2 เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 โรคเบาหวาน (Diabetes Mellitus) โรคเบาหวาน (Diabetes Mellitus : DM) คือภาวะที่ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ (Hyperglycemia) ซึ่งเป็นผลมาจากความผิดปกติทางเมตาบอลิซึม เนื่องจากการขาดฮอร์โมนอินซูลินหรือร่างกายไม่ตอบสนองต่อฮอร์โมนอินซูลิน ซึ่งฮอร์โมนอินซูลินเป็นฮอร์โมนสำคัญของร่างกายสร้างและหลังจากเบต้าเซลล์ของตับอ่อน และมีหน้าที่ในการพาน้ำตาลกลูโคสเข้าสู่เนื้อเยื่อต่าง ๆ ของร่างกาย เพื่อเผาผลาญพลังงานในการดำเนินชีวิต ถ้าขาดฮอร์โมนอินซูลินหรือการออกฤทธิ์ของฮอร์โมนอินซูลินบกพร่อง [7]

2.2 เทคนิคเหมืองข้อมูล เทคนิคเหมืองข้อมูลในปัจจุบันมีหลายรูปแบบ ทั้งเทคนิคแบบผู้สอน (Supervised Learning) แบบไม่มีผู้สอน (Unsupervised Learning) แบบการเรียนรู้เกิดมาจากการปฏิสัมพันธ์ (Reinforcement Learning) ซึ่งผู้วิจัยได้ศึกษาเทคนิคที่ใช้ทำเหมืองข้อมูล เพื่อการจำแนกกลุ่มของข้อมูล ได้แก่ วิธีเพอร์เซ็ปตรอนหลายชั้น วิธีนาอิวเบย์ วิธีโลจิสติกส์ วิธีแรนดอมฟอเรส และวิธีต้นไม้ตัดสินใจ เนื่องจากเทคนิคดังกล่าวเป็นเทคนิคที่นิยม และเหมาะกับการจำแนกข้อมูล ที่เป็นหมวดหมู่ข้อมูลต่าง ๆ โดยการเรียนรู้แบบมีผู้สอนจะแตกต่างกับการเรียนรู้แบบไม่มีผู้สอนที่จะไม่ทราบถึงหมวดหมู่ของข้อมูล ตัวอย่างเช่น ในการวิเคราะห์ข้อมูลโรคเบาหวานที่ไม่มีข้อมูลที่บ่งบอกว่าเป็นโรคเบาหวาน ก็จะนำข้อมูลปัจจัยไปวิเคราะห์หาความเสี่ยงว่าเป็นโรคเบาหวานหรือไม่ และข้อมูลที่ผู้วิจัยได้ทำการรวบรวมนั้น มีหมวดหมู่ของข้อมูลที่บ่งบอกว่าเป็นโรคเบาหวานอย่างชัดเจน ผู้วิจัยจึงได้เลือกใช้เทคนิควิธี ดังกล่าวข้างต้น โดยมีรายละเอียดดังนี้

1) เทคนิคเพอร์เซ็ปตรอนหลายชั้น (Multi-layer Perceptron) โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นเป็นโครงข่ายประสาทเทียมแบบ Feed Forward ซึ่งเป็นแบบจำลองทางคณิตศาสตร์ที่ใช้อย่างแพร่หลายและได้รับความสนใจนำมาศึกษาวิจัยอย่างมากในหลากหลายสาขาวิชา พัฒนาขึ้นเพื่อจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ ซึ่งจะจับคู่ระหว่างชุดข้อมูลนำเข้าและชุดของผลลัพธ์ที่เหมาะสม เทคนิคนี้ประกอบด้วย โหนดหลายชั้นในกราฟกำกับโดยแต่ละชั้นจะเชื่อมต่อกับโหนดหนึ่งต่อไปอีกโหนด [8] แม้ว่าโครงสร้างโครงข่ายประสาทเทียมที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัตินั้นการกำหนดชั้นซ่อน 1 ชั้นก็เพียงพอต่อการวิเคราะห์ข้อมูล [9]

2) เทคนิคนาอิวเบย์ (Naive bayes) นาอิวเบย์เป็นวิธีการเรียนรู้ที่ใช้หลักของความน่าจะเป็นตามกฎทฤษฎีของนาอิวเบย์เข้ามาช่วยในการเรียนรู้เพื่อหาสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูลการเรียนรู้แบบนาอิวเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละ สมมติฐาน [10] ถูกลำบากใช้อย่างแพร่หลายในงานจำแนกหมวดหมู่เอกสาร มีประสิทธิภาพการจำแนกที่ดี ซึ่งอาศัยหลักความน่าจะเป็นในการทำนายผลลัพธ์ [11] เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะกับการเรียนรู้ของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน [12]

3) เทคนิคโลจิสติกส์ (Logistic Regression) การถดถอยโลจิสติก (Logistic Regression) ใช้เพื่อหาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยตัวแปรตามมีเพียงสองค่า คือ 0 และ 1 หากตัวแปรอิสระมีค่าน้อย ค่าของตัวแปรตามจะมีค่าเท่ากับ 0 และหากค่าตัวแปรอิสระมีค่ามาก ค่าของตัวแปรตามจะมีค่าเท่ากับ 1 [13] ส่วนตัวแปรอิสระอาจจะเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ จะใช้ประมาณโอกาสที่จะเกิดเหตุการณ์ผลของการวิเคราะห์ขึ้นอยู่กับการปัจจัยที่สำคัญ [14] อาศัยหลักการความน่าจะเป็นของผลลัพธ์การทำนายว่าจะเกิดขึ้นหรือไม่โดยขึ้นอยู่กับการคุณสมบัติของ Logistic Regression [15]

4) เทคนิคแรนดอมฟอเรส (Random Forest) เทคนิคแรนดอมฟอเรส เป็นแบบจำลองอีกการทำนายที่ใช้พื้นฐานจากต้นไม้ตัดสินใจ [16] ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มตัวอย่างข้อมูลและคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจ [17] ถูกสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับ แล้วนำมาสร้างเป็นต้นไม้ [18] ซึ่งจะเรียงลำดับผลลัพธ์ที่สูงที่สุดจะเป็นผลลัพธ์สุดท้ายในเทคนิคป่าสุ่มของเทคนิคการจำแนกประเภทใช้วิธีการทั้งมวล (Ensemble) เพื่อให้ได้ผลลัพธ์ชุดข้อมูลฝึกจะถูกป้อนเพื่อฝึกแบบต้นไม้

การตัดสินใจต่าง ๆ ซึ่งข้อมูลชุดนี้ประกอบด้วยการสังเกตและคุณลักษณะที่จะถูกเลือกแบบสุ่มในระหว่าง การแยกโหนด ระบบป่าฝนอาศัยต้นไม้ตัดสินใจต่าง ๆ แผนผังการตัดสินใจ ทั้งหมดประกอบด้วยโหนดการตัดสินใจ โหนดปลายสุดและโหนดราก โหนดปลายสุดของแผนผังต้นไม้แต่ละต้นเป็นผลลัพธ์สุดท้ายที่สร้างโดยแผนผังการตัดสินใจการเลือกผลลัพธ์สุดท้าย

5) เทคนิคต้นไม้ตัดสินใจ (Decision Trees) เทคนิคการสร้างแบบจำลองที่ใช้ช่วยในการตัดสินใจ โดย แบบจำลองจะอยู่ในรูปแบบโครงสร้างต้นไม้ [19] ต้นไม้ตัดสินใจเป็นแบบจำลองทางคณิตศาสตร์ เพื่อหาทางเลือกที่ ดีที่สุด โดยการนำเอาข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างแบบต้นไม้ [20]

2.3 งานวิจัยที่เกี่ยวข้อง

นพรัตน์ นนทศิริ พิศณุ ชัยจิตวณิชกุล และกริช สมกันธา [21] ศึกษาเรื่อง “การจำแนกข้อมูลเพื่อ วินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิควิธีแบบร่วมกันตัดสินใจและวิธีเลือกคุณลักษณะเด่นไป ข้างหน้า” โดยใช้เทคนิค Decision Tree, Random Forest, NaïveBayes, Support Vector Machine และ K-Nearest Neighbor พบว่าวิธีร่วมกันตัดสินใจให้ผลลัพธ์ที่ดีกว่าการใช้ เทคนิคแบบโมเดลเดี่ยว (Single Model)

เพชรรัตน์ ม่วงน้อย จักรพันธ์ พลาผล และภรณ์ยา ปาลวิสุทธิ [22] ศึกษาเรื่อง “ตัวแบบประเมิน ภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล” โดยใช้เทคนิค Decision Tree พบว่า Random Tree มีประสิทธิภาพสูงสุดโดยมีค่าความถูกต้อง 96.00%

นงเยาว์ ในอรุณ [23] ศึกษาเรื่อง “การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนาย ความเสี่ยงโรคหัวใจและหลอดเลือดโดยใช้อัลกอริทึมเหมืองข้อมูล” โดยใช้เทคนิค Neural Network, Random Forest, NaïveBayes, K-Nearest Neighbors และ Decision Tree พบว่าแบบจำลองโครงข่ายประสาทเทียม พร้อมการเลือกคุณสมบัติของแอตทริบิวต์เป็นแบบจำลองที่เหมาะสมที่สุด

รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย [24] ศึกษาเรื่อง “การจำแนกประเภทผู้ป่วยโรคเบาหวานโดย ใช้เทคนิคเหมืองข้อมูลและการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล” โดยใช้เทคนิค Decision tree, Support Vector Machine, NaïveBayes และ K-Nearest Neighbors พบว่าซัพพอร์ทเวกเตอร์แมชชีนมี ประสิทธิภาพการทำนายสูงสุด คิดเป็น 76.95

ตัวช่วย ดวงแก้ว และสายัญ สายยศ [25] ศึกษาเรื่อง “การวิเคราะห์ปัจจัยที่มีผลต่อการพัฒนาการล่าช้า ตามช่วงอายุในเด็กปฐมวัยด้วยเทคนิคเหมืองข้อมูล” โดยใช้เทคนิค Decision Tree, K-Nearest Neighbors และ NaïveBayes พบว่าเทคนิค Decision Tree มีความถูกต้องแม่นยำมากที่สุดในการทำนาย

ณัฐวดี หงษ์บุญมี และประภาสิริ ตรีพานิชกุล [26] ศึกษาเรื่อง “การเปรียบเทียบประสิทธิภาพ การจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล” โดย ใช้เทคนิค Bayes’ Theorem, Decision Tree และ K-fold cross validation พบว่าการจำแนกข้อมูลโดยใช้ โครงข่ายประสาทเทียมให้ค่าประสิทธิภาพสูงสุดโดยมีค่าความ ถูกต้อง 82.97%

วนิดา พงษ์สงวน ทิพย์า ถินสูงเนิน และมาโนช ถินสูงเนิน [27] ศึกษาเรื่อง “การพัฒนาแบบจำลอง ปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ” โดยใช้เทคนิค Back propagation, Neural Networks, Radial Basis Function Network และ NaïveBayes พบว่าแบบจำลองที่พัฒนาให้ประสิทธิภาพที่มี ค่าความแม่นยำตรง 76.14%

จารี ทองคำ วาทีนิ สุขมาก และกัมพศ สุขมาก [28] ศึกษาเรื่อง “การเปรียบเทียบประสิทธิภาพของ เทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก” โดยใช้เทคนิค พบว่าเทคนิค FP-Growth มีความสามารถในการสร้างกฎความสัมพันธ์ได้มากกว่า เทคนิค Apriori

ณัฐพล แสนคำ ธนากร ปุระรัมย์ และทิพวัลย์ แสนคำ [29] ศึกษาเรื่อง “ระบบสนับสนุนทาง การแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล” โดยใช้เทคนิค Decision tree และ Random Forest พบว่าเทคนิค Random Forest ประสิทธิภาพค่าความถูกต้องสูงที่สุดจากทุกตัวแบบ

เบญจศักดิ์ จงหมื่นไวย [30] ศึกษาเรื่อง “การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้ อัลกอริทึมการจัดกลุ่ม J48 และ NaiveBayes กรณีศึกษาสาธารณสุขโพธิ์กลางนครราชสีมา” โดยใช้เทคนิค Decision Tree (J48) และ NaiveBayes พบว่าตัวแบบที่เหมาะสมสำหรับผู้สูงอายุในกลุ่มนี้คือ Decision Tree จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า การเลือกตัวแบบที่มีประสิทธิภาพนั้นต้องใช้ค่าความถูกต้อง (Accuracy) จากแต่ละตัวแบบผ่านฐานข้อมูล สรุปผลได้ดังตารางที่ 1

ตารางที่ 1 ผลการเปรียบเทียบค่าความถูกต้อง

Author	Data Set	Algorithm	Performance
Xu, He, และ Hu [36]	Data for 520 cases comes from the Sylhet Diabetes Hospital in Sylhet, Bangladesh	Dimensional convolutional neural network	97.02%
Sivaranjani, S Ananya, J Aravinth, และ R Karthika [35]	PIMA Indian diabetes dataset	Random Forest	83.00%
สายชล สันสมบูรณ์ทอง [27]	แบบสอบถาม	Artificial Neural Networks	95.94%
นพรัตน์ นนทศิริ, ราตรี มนต์ศิลา, และ กริช สมกันธา[15]	ข้อมูลเวชระเบียนผู้ป่วยโรคเบาหวานย้อนหลัง จำนวน 1,435 ชุดข้อมูล 16 คุณลักษณะ	Decision Tree	93.73%
ผลการวิจัยนี้	แบบสอบถาม	Random Forest	99.75%

จากตารางที่ 1 ผลการเปรียบเทียบค่าความถูกต้องที่ได้จากฐานข้อมูล พบว่าค่าความถูกต้องที่ได้จาก อัลกอริทึมแต่ละตัวบอกถึงประสิทธิภาพในการที่จะเลือกอัลกอริทึมนั้น เพื่อนำไปใช้ในการพยากรณ์ได้ต่อไป

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 แบบสอบถามข้อมูลทั่วไป จำนวนทั้งสิ้น 13 ข้อ และแบบสอบถามเกี่ยวกับพฤติกรรมความเสี่ยงที่ก่อให้เกิดโรคเบาหวาน จำนวนทั้งสิ้น 32 ข้อ

1.2 โปรแกรม Waikato Environment for Knowledge Analysis

2. กลุ่มเป้าหมาย

2.1 เป็นประชากรผู้ที่เป็นเบาหวาน โดยมีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน ผู้ป่วยที่เข้ารับ การคัดกรองโรคเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม จำนวนทั้งสิ้น 9,841 คน

2.2 เป็นผู้ป่วยที่แพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวานและผู้ป่วยที่เข้ารับการคัดกรองโรคเบาหวาน โรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม จำนวน 400 คน

3. ขั้นตอนการดำเนินการวิจัย

3.1 การเก็บรวบรวมข้อมูลได้จากการสุ่มตัวอย่างผู้ป่วยโรงพยาบาลกันทรวิชัย พบว่ามีจำนวนผู้ป่วยที่ตอบแบบสอบถามออนไลน์ทั้งสิ้น 475 คน และผู้วิจัยได้นำข้อมูลดังกล่าวมาผ่านการวิเคราะห์และตรวจสอบความสมบูรณ์ของข้อมูลพบว่ามีข้อมูลที่ถูกต้องครบถ้วนและสามารถนำไปใช้ในการพยากรณ์โรคเบาหวานได้จริงจำนวนทั้งสิ้น 400 คน

3.2 เก็บรวบรวมข้อมูลโดยแบบสอบถาม ลักษณะของแบบสอบถามมีลักษณะที่เกี่ยวข้องกับ อายุ เพศ ค่าดัชนีมวลกายค่าความดันโลหิตสูง ประวัติเบาหวานในพ่อแม่ พี่น้อง เส้นรอบเอว การสูบบุหรี่ และการดื่มสุรา ในกลุ่มประชากรตัวอย่าง จำนวน 300 คน

3.3 เก็บรวบรวมข้อมูลโดยประวัติผู้ป่วยที่มีลักษณะเกี่ยวข้องกับ อายุ เพศ ค่าดัชนีมวลกายค่าความดันโลหิตสูง และประวัติเบาหวานในพ่อแม่ พี่น้อง เส้นรอบเอว การสูบบุหรี่ การดื่มสุรา และมีแพทย์ผู้เชี่ยวชาญวินิจฉัยว่าเป็นโรคเบาหวาน จำนวน 100 คน

3.4 วัดประสิทธิภาพของเครื่องมือจากข้อมูลที่พบในแบบสอบถามเทียบกับข้อมูลที่แพทย์วินิจฉัยว่าเป็นโรคเบาหวาน

3.5 ผู้วิจัยดำเนินการเข้าเก็บข้อมูลในคลินิกเบาหวานโรงพยาบาลกันทรวิชัย โดยมีกระบวนการในการชี้แจงและขอความยินยอมในกลุ่มอาสาสมัครตามขั้นตอนของการทำวิจัยในคน จำนวนทั้งสิ้น 400 คน โดยใช้วิธีในการสุ่มแบบไม่ใส่กลับคืน

3.6 ระยะเวลาในการดำเนินการระหว่าง สิงหาคม 2563 - ธันวาคม 2564 ตามทะเบียนผู้ป่วยเบาหวานโรงพยาบาลกันทรวิชัย จังหวัดมหาสารคาม

3.7 ตรวจสอบความสมบูรณ์ของแบบสอบถามที่ได้รับการตอบกลับ พบว่ามีข้อมูลที่ไม่ถูกต้องครบถ้วนจำนวน 75 คน ซึ่งผู้วิจัยได้ตัดข้อมูลส่วนนี้ออกตามกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล (Data Cleansing) เพื่อให้การจำแนกข้อมูลถูกต้องมากที่สุด ทำให้เหลือจำนวนข้อมูลที่สามารถนำมาใช้ในการวิเคราะห์ได้จริงจำนวนทั้งสิ้น 400 คน

3.8 นำข้อมูลที่ได้มาทำให้ข้อมูลอยู่ในโครงสร้างหรือรูปแบบบรรทัดฐานเดียวกัน (Normalization) เพื่อลดความซ้ำซ้อนของข้อมูล

3.9 ดำเนินการเก็บรวบรวมข้อมูลจากแบบสอบถามที่ได้รับเพื่อนำมาวิเคราะห์ข้อมูลและแปลผลต่อไป

4. สถิติที่ใช้ในการวิจัย

ผู้วิจัยวิเคราะห์ข้อมูลด้วยการใช้สถิติ ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก ค่า F-measure และ ค่า ROC

ผลการวิจัย

1. ผลการพัฒนาแบบจำลองสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน

โดยใช้เทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น เทคนิคนาอ็ฟ เบย์ เทคนิคโลจิสติก เทคนิคแรนดอมฟอเรท และเทคนิคต้นไม้ตัดสินใจ ในการค้นหาความแม่นยำในการพยากรณ์ ผ่านโปรแกรม Waikato Environment for Knowledge Analysis พบว่า เทคนิคแรนดอมฟอเรส ให้ค่าความถูกต้องมากที่สุดอยู่ที่ 99.75% รองลงมาคือเทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น มีค่า 98.50% เทคนิคต้นไม้ตัดสินใจ มีค่า 94.75% เทคนิคนาอ็ฟ เบย์ มีค่า 78.00% และเทคนิคโลจิสติก มีค่า 77.25% ตามลำดับ สามารถนำผลที่ได้จากงานวิจัยนี้ไปประยุกต์ใช้ในการประกอบการรักษาผู้ป่วยโรคเบาหวานต่อไปในอนาคต

2. ผลการเปรียบเทียบประสิทธิภาพแบบจำลองที่เหมาะสมที่สุดสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวาน ของอัลกอริทึมทั้ง 5 จะสามารถได้โดยการนำข้อมูลชุดเดียวกันมาเข้าตัวแบบทดสอบเพื่อให้สามารถตอบข้อสมมุติฐาน

2.1 วิธีเพอร์เซ็ปตรอนหลายชั้น เมื่อนำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น ทำให้ได้ผลลัพธ์ ดังภาพที่ 1

Correctly Classified Instances	394	98.5	%						
Incorrectly Classified Instances	6	1.5	%						
Kappa statistic	0.9592								
Mean absolute error	0.0213								
Root mean squared error	0.1176								
Relative absolute error	5.6592	%							
Root relative squared error	27.1552	%							
Total Number of Instances	400								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.060	0.980	1.000	0.990	0.960	0.952	0.945	No
	0.940	0.000	1.000	0.940	0.969	0.960	0.952	0.963	Yes
Weighted Avg.	0.985	0.045	0.985	0.985	0.985	0.960	0.952	0.950	

ภาพที่ 1 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น

จากภาพที่ 1 แสดงการวิเคราะห์ข้อมูลด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 98.50% มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 100% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 98.00% มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 94.00% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 100% ค่าช่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 96.90% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 99.00% นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 96.30% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 94.50% แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก

2.2 วิธีนาอีฟ เบย์ เมื่อนำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีวิธีนาอีฟ เบย์ ทำให้ได้ผลลัพธ์ ดังภาพที่ 2

Correctly Classified Instances	312	78	%						
Incorrectly Classified Instances	88	22	%						
Kappa statistic	0.2281								
Mean absolute error	0.3219								
Root mean squared error	0.3967								
Relative absolute error	85.6991	%							
Root relative squared error	91.6083	%							
Total Number of Instances	400								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.973	0.800	0.785	0.973	0.869	0.294	0.755	0.901	No
	0.200	0.027	0.714	0.200	0.313	0.294	0.755	0.530	Yes
Weighted Avg.	0.780	0.607	0.767	0.780	0.730	0.294	0.755	0.808	

ภาพที่ 2 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีนาอีฟ เบย์

จากภาพที่ 2 แสดงการวิเคราะห์ข้อมูลด้วยวิธี Naïve Bayes ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 78.00% มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 71.40% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 75.80% มีค่าพยากรณ์ความระลึกรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 20.00% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 97.30% ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 31.30% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานมีค่าเท่ากับ 86.90% นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 75.50% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 75.50% แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดี

2.3 วิธีโลจิสติกพหุกลุ่ม ทำการนำเข้าข้อมูลจำนวน 400 รายการ ทำการวิเคราะห์ด้วยวิธีโลจิสติกพหุกลุ่ม ทำให้ได้ผลลัพธ์ ดังภาพที่ 3

Correctly Classified Instances	309	77.25	%						
Incorrectly Classified Instances	91	22.75	%						
Kappa statistic	0.2121								
Mean absolute error	0.3245								
Root mean squared error	0.4028								
Relative absolute error	86.3853	%							
Root relative squared error	93.0109	%							
Total Number of Instances	400								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.963	0.800	0.783	0.963	0.864	0.265	0.733	0.882	No
	0.200	0.037	0.645	0.200	0.305	0.265	0.733	0.488	Yes
Weighted Avg.	0.773	0.609	0.749	0.773	0.724	0.265	0.733	0.784	

ภาพที่ 3 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีโลจิสติกพหุกลุ่ม

จากภาพที่ 3 แสดงการวิเคราะห์ข้อมูลด้วยวิธีโลจิสติกพหุกลุ่ม ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 77.25% มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 64.50% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 78.30% มีค่าพยากรณ์ความระลึกรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 20.00% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 96.30% ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 30.05% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานมีค่าเท่ากับ 86.40% นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 73.30% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 73.30% แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดี

2.4 วิธีเทคนิคแรนดอมฟอเรส นำข้อมูลป้อนเข้าจำนวน 400 รายการ และทำการวิเคราะห์ด้วยวิธีเทคนิคแรนดอมฟอเรส ทำให้ได้ผลลัพธ์ ดังภาพที่ 4

Correctly Classified Instances	399	99.75	%						
Incorrectly Classified Instances	1	0.25	%						
Kappa statistic	0.9933								
Mean absolute error	0.1413								
Root mean squared error	0.1645								
Relative absolute error	37.6117	%							
Root relative squared error	37.9941	%							
Total Number of Instances	400								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.010	0.997	1.000	0.998	0.993	1.000	1.000	No
	0.990	0.000	1.000	0.990	0.995	0.993	1.000	1.000	Yes
Weighted Avg.	0.998	0.008	0.998	0.998	0.997	0.993	1.000	1.000	

ภาพที่ 4 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีเทคนิคแรนดอมฟอเรส

จากภาพที่ 4 แสดงการวิเคราะห์ข้อมูลด้วยวิธีเทคนิคแรนดอมฟอเรส ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 99.75% มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 100% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 99.70% มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 99.00% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 100% ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 99.50% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานมีค่าเท่ากับ 99.80% นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 100% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 100% แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก

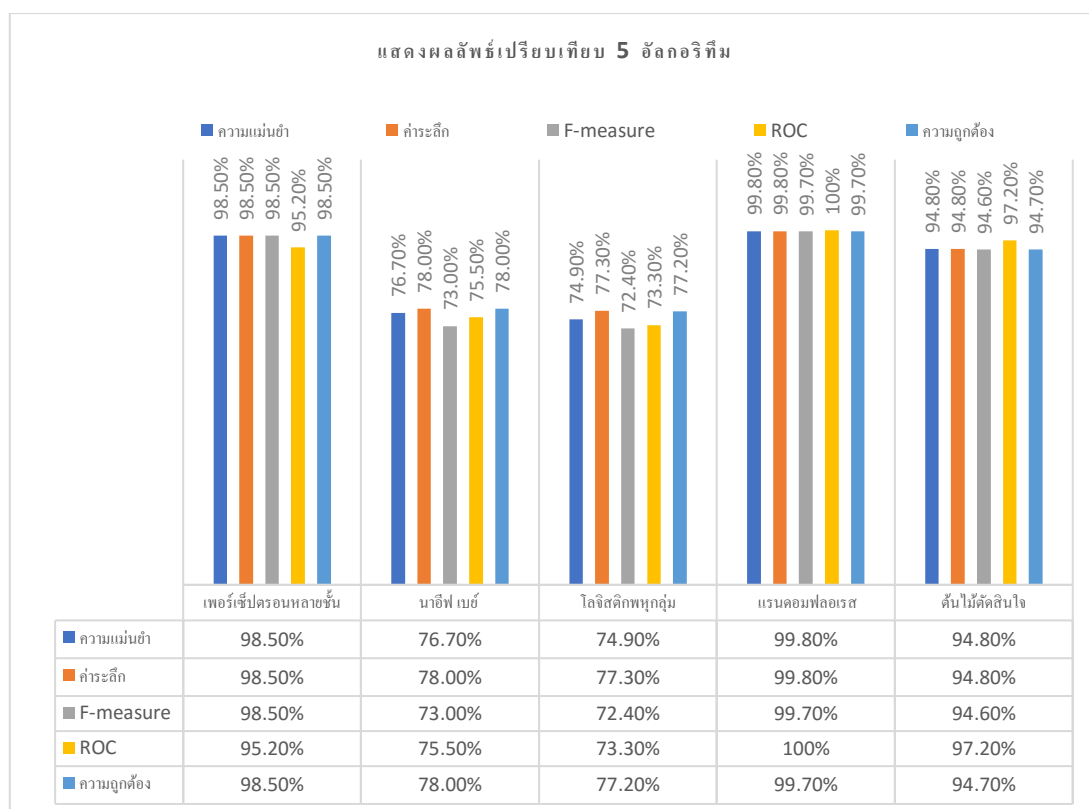
2.5 วิธีต้นไม้ตัดสินใจทำการนำเข้าสู่ข้อมูลจำนวน 400 รายการ ทำการวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ ทำให้ได้ผลลัพธ์ ดังภาพที่ 5

Correctly Classified Instances	379	94.75	%						
Incorrectly Classified Instances	21	5.25	%						
Kappa statistic	0.8537								
Mean absolute error	0.0889								
Root mean squared error	0.2109								
Relative absolute error	23.6764	%							
Root relative squared error	48.6985	%							
Total Number of Instances	400								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.987	0.170	0.946	0.987	0.966	0.857	0.972	0.986	No
	0.830	0.013	0.954	0.830	0.888	0.857	0.972	0.927	Yes
Weighted Avg.	0.948	0.131	0.948	0.948	0.946	0.857	0.972	0.972	

ภาพที่ 5 ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

จากภาพที่ 5 แสดงการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ผลลัพธ์ที่ได้จากอัลกอริทึมให้ค่าความถูกต้องต่อการเกิดโรคเบาหวานเท่ากับ 94.75% มีค่าความแม่นยำจากการพยากรณ์ในรูปแบบที่มีความเสี่ยงต่อการเกิด

โรคเบาหวานเท่ากับ 95.40% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 94.60% มีค่าพยากรณ์ความระลึกในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 83.00% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 98.70% ค่าถ่วงดุลพยากรณ์ออกมาในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 88.80% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงการเกิดโรคเบาหวานมีค่าเท่ากับ 0.966% นอกจากนี้ยังพยากรณ์ค่า ROC ผ่านอัลกอริทึมในรูปแบบที่มีความเสี่ยงต่อการเกิดโรคเบาหวานให้ค่าเท่ากับ 97.20% และพยากรณ์ในรูปแบบที่ไม่มีความเสี่ยงต่อการเกิดโรคเบาหวานเท่ากับ 97.20% แสดงให้เห็นประสิทธิภาพในการทดสอบการพยากรณ์มีความแม่นยำที่ดีมาก ผลความถูกต้องจาก 5 อัลกอริทึม ดังภาพที่ 6



ภาพที่ 6 ผลลัพธ์เปรียบเทียบ 5 อัลกอริทึม

จากภาพที่ 6 แสดงผลลัพธ์ที่ได้จากทั้ง 5 อัลกอริทึม พบว่าวิธีเทคนิคแรนดอมฟลอเรส ให้ค่าความถูกต้องในการพยากรณ์มากที่สุด ที่ 99.75% ค่าความแม่นยำที่ 99.80% ค่าระลึกที่ 99.80% ค่า F-measure ที่ 99.70% และเส้นกราฟ ROC ที่ 100% ดังนั้น จึงสรุปได้ว่าวิธีเทคนิคแรนดอมฟลอเรส มีประสิทธิภาพในการพยากรณ์การเกิดโรคเบาหวานได้ดีที่สุด

อภิปรายผลและสรุปผล

งานวิจัยนี้เป็นการศึกษาเทคนิคการทำเหมืองข้อมูลเพื่อการสร้างแบบพยากรณ์โรคเบาหวาน สร้างแบบจำลองการจำแนกประเภทข้อมูล (Classification) โดยใช้เทคนิคการจำแนกข้อมูลด้วยวิธีเทคนิคเพอร์เซ็ปตรอนแบบหลายชั้น เทคนิคนาอิว เบย์ เทคนิคโลจิสติก เทคนิคแรนดอมฟลอเรส และเทคนิคต้นไม้ตัดสินใจ โดยผู้วิจัยได้ทำการศึกษางานวิจัยที่เกี่ยวข้อง เพื่อค้นหาเทคนิคที่เหมาะสมในการพยากรณ์และจากการศึกษาพบว่า เทคนิคแรน

ดอมฟอเรส (Random forest) ให้ค่าความถูกต้องในการพยากรณ์มากที่สุด ที่ 99.75% ค่าความแม่นยำที่ 99.80% ค่าระลอกที่ 99.80% ค่า F-measure ที่ 99.70% และเส้นกราฟ ROC ที่ 100% เหตุที่เป็นเช่นนี้อาจเนื่องมาจากค่าที่ปรากฏจากการทดสอบ มีประสิทธิภาพที่ดีและน่าเชื่อถือมากในบรรดาเทคนิคที่กล่าวมา ดังนั้นจึงสรุปได้ว่าวิธีเทคนิคแรนดอมฟอเรส (Random forest) มีประสิทธิภาพในการพยากรณ์การเกิดโรคเบาหวานได้ดีที่สุดและมีความเหมาะสมในการนำไปใช้สร้างแบบจำลองพยากรณ์ความเสี่ยงของการเกิดโรคเบาหวาน ซึ่งสอดคล้องกับงานวิจัยเรื่อง การจำแนกผู้ป่วยเบาหวานโดยใช้เทคนิคการโหวตรวม กรณีศึกษา โรงพยาบาลศูนย์อุดรธานี พบว่าเทคนิคแรนดอมฟอเรส (Random forest) ให้ผลลัพธ์การทดสอบประสิทธิภาพที่ดีที่สุด ให้ค่าความถูกต้องที่ 88.03% ให้ค่าความแม่นยำที่ 88.22% ให้ค่าความครบถ้วนที่ 90.36% ให้ค่า F-measure ที่ 89.28% [31]

ข้อเสนอแนะ

ตัวแบบที่ผู้วิจัยได้นำเสนอนั้นเป็นเพียงการใช้เทคนิคหนึ่งของการทำเหมืองข้อมูลจากหลากหลายเทคนิคที่มีอยู่และกลุ่มประชากรตัวอย่างที่เก็บมาเป็นแค่กลุ่มประชากรในพื้นที่ ในอนาคตหากมีการพัฒนางานวิจัยให้มีประสิทธิภาพมากขึ้น อาจต้องเพิ่มปริมาณข้อมูลที่นำมาใช้ในการสร้างตัวแบบหรือขยายการเก็บข้อมูลกลุ่มตัวอย่างประชากรกลุ่มตัวอย่างให้กว้างขึ้น เพื่อให้เกิดความหลากหลายของข้อมูล และเพื่อให้เห็นความแม่นยำของตัวแบบสามารถนำตัวแบบไปใช้ได้อย่างมีประสิทธิภาพ

เอกสารอ้างอิง

- [1] สมาพันธ์เบาหวานนานาชาติ (International Diabetes Federation). 2560. สมาพันธ์เบาหวานนานาชาติ ประเมินทั่วโลกมีผู้ป่วยเบาหวาน 425 ล้านราย Hfocus.Org เจาะลึกระบบสุขภาพ. จาก <https://www.hfocus.org/content/2019/11/18054>
- [2] ฉวีวรรณ ศรีดาวเรือง และจิราพร วรวงศ์. (2565). ความสัมพันธ์ระหว่างความรู้ด้านสุขภาพกับพฤติกรรมสุขภาพตามหลัก 3 อ. 2ส. ของผู้สูงอายุที่ป่วยเป็นโรคเบาหวานชนิดที่ 2. วารสารศูนย์อนามัยที่ 9, 9(2), 456–456.
- [3] จรรยา สุนทรภักดี และชนัญญา เพียงแก้ว. (2561). การพัฒนารูปแบบการส่งเสริมสุขภาพเชิงรุกเพื่อป้องกันการเกิดโรคเบาหวานรายใหม่ใน. วารสารวิชาการสำนักงานสาธารณสุขจังหวัดมหาสารคาม, 2(3), 1-14.
- [4] กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข. (2564). กองโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข. จาก <http://www.thaincd.com/2016/mission/documents.php?tid=32&gid=1-020>
- [5] กมลพร สิริคุตจตุพร, วิราพรรณ วิโรจน์รัตน์ และนารีรัตน์ จิตมนตรี. (2560). ปัจจัยทำนายพฤติกรรมการจัดการตนเองของผู้สูงอายุโรคเบาหวานชนิดที่ 2. วารสารสภาการพยาบาล, 32(1), 81–93. จาก <https://he02.tci-thaijo.org/index.php/TJONC/article/view/82615>
- [6] กระทรวงสาธารณสุข กรมอนามัย. (2561). อัตราผู้ป่วยเบาหวานรายใหม่จากกลุ่มเสี่ยงเบาหวาน เขตฯ 07. จาก http://healthkpi.moph.go.th/kpi2/kpi/index/?id=957&kpi_year=2561&lv=2&z=07
- [7] สุภาภรณ์ ดันดินนทระกุล. (2555). เบาหวาน...ถ้ารู้ทัน...ก็ไร้เสี่ยง. วารสารวิชาการมหาวิทยาลัยการกีฬาแห่งชาติ, 4(2), 125–36. จาก <https://he02.tci-thaijo.org/index.php/TNSUJournal/article/view/254484>
- [8] วิบล ญีก และจารี ทองคำ. (2561). การเปรียบเทียบเทคนิคอนุกรมเวลาเพื่อพยากรณ์ราคาทองและราคาน้ำมัน. วารสาร มทร. อีสาน ฉบับวิทยาศาสตร์และเทคโนโลยี, 11(2), 157–58.
- [9] สายชล สีนสมบูรณ์ทอง. (2563). การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล. วารสารวิทยาศาสตร์และเทคโนโลยี, 28(3), 387–88.
- [10] จุฑาทิพย์ ทัพย์พูล และนิเวศ จิระวิชิตชัย. (2559). การจำแนกจดหมายอิเล็กทรอนิกส์ที่เป็นสแปมโดยใช้เทคนิคเหมืองข้อมูล. วารสาร มทร. วิทยาศาสตร์และเทคโนโลยี, 6(1), 104–5.
- [11] พิพัฒน์ ลิ้มประไพพงษ์ และสมชาย เล็กเจริญ. (2560). ศึกษาปัจจัยที่ส่งผลต่อคะแนนการประเมินคุณภาพภายในระดับคณะของมหาวิทยาลัยรังสิต โดยใช้เทคนิคเหมืองข้อมูล. การประชุมวิชาการระดับชาติ มหาวิทยาลัยรังสิต ประจำปี ๒๕๖ (RSU National Research Conference, 2017), 109–10.

- [12] อนันต์ชัย ขุดิภาสเจริญ และจัญญ์ แสนราช. (2561). การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่ว. *วารสารวิจัย มหาวิทยาลัยขอนแก่น (ฉบับบัณฑิตศึกษา) สาขามนุษยศาสตร์และสังคมศาสตร์*, 6(3), 105–6.
- [13] สุวัชร ศรีเปารยะ และสายชล สันสมบุรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(5), 845–46. จาก <https://li01.tci-thaijo.org/index.php/tstj/article/view/85101>
- [14] กัลยา บุญหล้า และวรวุฒิ มหาโพธิ์. (2565). การจำแนกคุณภาพน้ำในประเทศไทยโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 30(5), 29–30. จาก <https://li01.tci-thaijo.org/index.php/tstj/article/view/255032>
- [15] ปรียานุช ประเสริฐสิริกุล, ศิริสรพร เหล่าหะเกียรติ, เรืองศักดิ์ ตระกูลพุทธิรักษ์ และศศิวิมล สุขพัฒน์. (2565). การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายผลการเรียนของนิสิตที่ใช้งาน ระบบการจัดการเรียนรู้ออนไลน์ด้วยการเรียนรู้ของเครื่อง. *วารสารวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยราชภัฏบุรีรัมย์*, 6(1), 81–82. จาก <https://ph02.tci-thaijo.org/index.php/scibru/article/view/245527>
- [16] ธนัท จระณะสมบุรณ์. (2561). การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร. *บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ*, 4–5.
- [17] อัครพล พลุสวัสดิ์ และจัญญ์ แสนราช. (2562). การศึกษาเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนของนักเรียนโรงเรียนวัดพระขาว (ประชานุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์ แห่งมหาวิทยาลัยราชภัฏเพชรบุรี*, 16(2), 6–7.
- [18] สุภัสสรา สมเจตนา และจารี ทองคำ. (2564). การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ตโฟนของบุตร. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี*, 23(1), 26–27.
- [19] ภูมิพัฒน์ ดวงกลาง และรัชนีา เครือแก้ว. (2562). แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ. *วารสารวิทยาศาสตร์และเทคโนโลยีนายเรืออากาศ*, 15, 3–4.
- [20] อัมมาพร กว่างสวาสดี เพียงฤทัย หนูสวัสดิ์, วราลี คงเหมาะ, ปวีณา ทิพายกุลรักษ์ และบุษกร สังขนันท์. (2562). ระบบการทำนายระดับความเครียดโดยใช้แผนผังการตัดสินใจ. *Rattanakosin Journal of Science and Technology*, 1(2), 13–26. จาก <https://ph02.tci-thaijo.org/index.php/RJST/article/view/239865>
- [21] นรรัตน์ นนทศิริ, พิศณุ ชัยจิตตนิชกุล และกริช สมกันธา. (2565). การจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานโดยใช้เทคนิค วิธีแบบรวมกันตัดสินใจและวิธีเลือกคุณลักษณะเด่นไปข้างหน้า. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏอุดรธานี*, 10(2), 107–22. จาก <https://ph01.tci-thaijo.org/index.php/scudru/article/view/245642>
- [22] เพชรรัตน์ ม่วงน้อย, จักรพันธ์ พลาผล, ภัณญา ปาสวิสุทธิ์ และสาขาวิชาเทคโนโลยี สารสนเทศ. (2564). ตัวแบบประเมินภาวะความเสี่ยงการเป็นโรคซึมเศร้าของนักศึกษาด้วยเทคนิคเหมืองข้อมูล. *วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ*, 7(1).
- [23] นงเยาว์ ในอรุณ. (2564). การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนายความเสี่ยงโรคหัวใจ และหลอดเลือดโดยใช้ อัลกอริทึมเหมืองข้อมูล Efficiency Comparison of Cardiovascular Risk Prediction Models Using Data Mining Algorithms ผู้ช่วยศาสตราจารย์ , สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ และเทคโนโลยี มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา, จังหวัดพระนครศรีอยุธยา 13000. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, 40(2).
- [24] รุ่งโรจน์ บุญมา และนิเวศ จิระวิชิตชัย. (2562). การจำแนกประเภทผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล และการเลือกคุณลักษณะจากความสัมพันธ์ของข้อมูล. *PKRU SciTech Journal*, 3(2), 11–19. จาก <https://ph01.tci-thaijo.org/index.php/pkruscitech/article/view/201722>
- [25] อิทธิพล ดวงแก้ว และสายัญ สายยศ. (2562). การวิเคราะห์ปัจจัยที่ส่งผลต่อการล่าช้าตามอายุของเด็กปฐมวัยด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ (JIST)*, 9(2), 44-55.
- [26] ณัฐดี หงษ์บุญมี และประภาสิริ ตริพาณิขกุล. (2562). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อวิเคราะห์ปัจจัยความเสี่ยงที่ส่งผลต่อการเกิดโรคไฮเปอร์ไทรอยด์ด้วยเทคนิคเหมืองข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยีสารสนเทศ*, 9(1), 41-51.
- [27] วนิตา พงษ์สงวน, ทิพยา ถินสูงเนิน และมานิช ถินสูงเนิน. (2561). การพัฒนาแบบจำลองปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม*, 1(1), 1-8.

- [28] จารี ทองคำ, วาทีนี สุขมาก และภิมพศ สุขมาก. (2561). การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก Performance Comparison of Apriori and FP-Growth Techniques in Generating Association Rules to Prostate Cancer. *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์*, 1(2), 103-111.
- [29] ณัฐพล แสนคำ, ธนากร ปุรารัมย์ และทิพวัลย์ แสนคำ. (2560). ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรัง โดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิชาการ โรงเรียนนายร้อยพระจุลจอมเกล้า* 15(1): 161–70. จาก <https://ph01.tci-thaijo.org/index.php/crma-journal/article/view/243110>
- [30] เบญจกัศ จงหมื่นไวย. (2558). การเปรียบเทียบปัจจัยโรคประจำตัวผู้สูงอายุโดยใช้อัลกอริทึมการจัดกลุ่ม J48 และ NaiveBayes กรณีศึกษาสาธารณสุขโพธิ์กลางนครราชสีมา. *วารสารโครงการงานวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ*, 1(2), 43-51.
- [31] ปพนันศรณ์ สิวสำแดงเดช. (2565). การจำแนกผู้ป่วยเบาหวานโดยใช้เทคนิคการไหวตรวม กรณีศึกษา โรงพยาบาลศูนย์อุดรธานี. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาสารคาม: มหาวิทยาลัยมหาสารคาม.