

การเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวาน

An Efficiency Comparison of Diabetes Prediction

สายชล สิ้นสมบุญทอง*

ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ถนนฉลองกรุง เขตลาดกระบัง กรุงเทพมหานคร 10520

Saichon Sinsomboonthong *

Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang,

Chalongkrung Road, Ladkrabang, Bangkok 10520

บทคัดย่อ

การศึกษานี้เป็นการเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานของโรงพยาบาลแห่งหนึ่ง วิธีการจำแนกกลุ่มที่นำมาเปรียบเทียบ คือ วิธีความใกล้เคียงกันมากที่สุดโดยใช้อัลกอริทึมชนิด IBk วิธีต้นไม้ตัดสินใจโดยใช้อัลกอริทึมชนิด J48 วิธีโครงข่ายประสาทเทียมโดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม และวิธีนาอิวเบย์ ในการเปรียบเทียบประสิทธิภาพในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 6 วิธี จะใช้ค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ผลการศึกษาพบว่าวิธีโครงข่ายประสาทเทียมมีค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยดีที่สุด คือ 95.94 %, 0.0491 และ 0.0396 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลดีที่สุด

คำสำคัญ : วิธีความใกล้เคียงกันมากที่สุด; วิธีต้นไม้ตัดสินใจ; วิธีโครงข่ายประสาทเทียม; วิธีซัพพอร์ตเวกเตอร์แมชชีน; วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม; วิธีนาอิวเบย์

Abstract

In this study, an efficiency comparison of diabetes prediction were determined and tested against real data obtained from a hospital, and their prediction efficiencies were compared. The tested classification methods were the following: (1) k-nearest neighbor method using IBk algorithm, (2) decision tree method using J48 algorithm, (3) artificial neural network method using multilayer perceptron algorithm, (4) support vector machine method using polynomial kernel, (5) binary logistic regression method, and (6) Naïve Bayes method. The following performance values were employed: accuracy, mean absolute error (MAE), and mean square error (MSE). The important

results are as follows. The artificial neural network method showed the best accuracy, MAE, and MSE at 95.94 %, 0.0491 and 0.0396 respectively. Then the artificial neural network method offered the best efficiencies in prediction method for diabetes.

Keywords: k-nearest neighbor method; decision tree method; artificial neural network method; support vector machine method; binary logistic regression method; Naïve Bayes method

1. บทนำ

Iyer และคณะ [1] ได้กล่าวว่าโรคเบาหวานเป็นโรคที่เกิดขึ้นเมื่อร่างกายผลิตอินซูลินไม่เพียงพอหรือร่างกายไม่สามารถใช้อินซูลินได้อย่างเหมาะสม ทำให้ปริมาณกลูโคสในเลือดสูง เซลล์ในร่างกายย่อยอาหารได้เป็นกลูโคสและกลูโคสเหล่านี้ ถูกขนส่งไปยังเซลล์ต่าง ๆ ในร่างกาย อินซูลินเป็นฮอร์โมนซึ่งผลิตกลูโคสโดยตรง โดยย่อยอาหารในเซลล์ร่างกาย การเปลี่ยนแปลงในการผลิตอินซูลินบางอย่างจะทำให้ระดับน้ำตาลในเลือดเพิ่มขึ้น ส่งผลเสียต่อเนื้อเยื่อและความล้มเหลวของอวัยวะ คนโดยทั่วไปได้รับความทุกข์ทรมานจากโรคเบาหวานเมื่อระดับน้ำตาลในเลือดสูงกว่าปกติ (4.4 ถึง 6.1 mmol/L)

เบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสำคัญทางด้านสาธารณสุขของโลก เป็นภัยคุกคามที่ลุกลามอย่างรวดเร็วไปทั่วโลก ส่งผลกระทบต่อการพัฒนาทางเศรษฐกิจอย่างมาก จากข้อมูลสมาพันธ์เบาหวานนานาชาติ (International Diabetes Federation, IDF) ได้รายงานว่ามีผู้เสียชีวิตด้วยโรคเบาหวาน 4 ล้านคนต่อปี เฉลี่ย 8 วินาทีต่อ 1 คน สำหรับผู้เป็นเบาหวานพบมากกว่า 300 ล้านคน และพบว่าคนที่อยู่ในประเทศที่มีรายได้ต่ำและปานกลางมีโอกาสเป็นเบาหวานเร็วกว่าคนที่อยู่ในประเทศที่มีรายได้สูง 10-20 ปี โดยพบมากขึ้นในวัยทำงาน สำหรับประเทศไทยปี พ.ศ. 2557 มีผู้เสียชีวิตจากโรคเบาหวานทั้งหมด 11,389 คน หรือเฉลี่ยวันละ 32 คน คิดเป็นอัตราตายด้วยโรคเบาหวาน 17.53 ต่อแสน

ประชากร และมีผู้ป่วยด้วยโรคเบาหวานเข้าพักรักษาตัวในโรงพยาบาลสังกัดกระทรวงสาธารณสุข จำนวน 698,720 ครั้ง คิดเป็นอัตราป่วยด้วยโรคเบาหวานเท่ากับ 1,081.25 ต่อแสนประชากร โดย นุชรี และ นิตยา [2]

ปัจจัยเสี่ยงของการเกิดโรคเบาหวาน คือน้ำหนักมากเกินไปและอ้วน ขาดการออกกำลังกาย เคยตรวจพบน้ำตาลในเลือดสูงกว่าปกติ การรับประทานอาหารไม่ถูกสัดส่วน หวาน มัน เค็มมากเกินไป อายุที่เพิ่มขึ้น ความดันโลหิตสูง และไขมันในเลือดสูง มีประวัติญาติสายตรงเป็นโรคเบาหวาน มีประวัติเบาหวานขณะตั้งครรภ์ และพันธุกรรม ในปี พ.ศ. 2555 สำนักกระบาดวิทยาได้รับข้อมูลแฟ้มมาตรฐานจากสำนักระบาดวิทยาและยุทธศาสตร์ สำนักงานปลัดกระทรวง กระทรวงสาธารณสุข ซึ่งรับข้อมูลจากสำนักงานสาธารณสุขจังหวัดทั่วประเทศ จำนวน 76 จังหวัด ยกเว้นกรุงเทพมหานคร พบว่าผู้ป่วยสะสม ณ วันที่ 31 ธันวาคม พ.ศ. 2555 จำนวน 1,799,977 ราย ความชุก 2,800.80 ต่อประชากรแสนคน และระหว่างวันที่ 1 มกราคม พ.ศ. 2555 – 31 ธันวาคม พ.ศ. 2555 มีผู้ป่วยเบาหวานที่ได้รับการขึ้นทะเบียนเป็นผู้ป่วยรายใหม่จำนวน 336,265 ราย อุบัติการณ์ 523.24 ต่อประชากรแสนคน โดยมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่องในช่วงปี พ.ศ. 2551-2555 ผู้ป่วยเบาหวานที่ได้รับการขึ้นทะเบียนเป็นผู้ป่วยรายใหม่เพศชาย 116,715 ราย เพศหญิง 219,550 ราย สัดส่วนเพศชายต่อเพศหญิงเท่ากับ 1:1.9 กลุ่มอายุที่พบสูงสุด คือ กลุ่มอายุมากกว่า

60 ปี รองลงมา คือ กลุ่มอายุ 50-59, 40-49, 15-39 และน้อยกว่า 15 ปี โดยมีอัตราป่วยเท่ากับ 2,128.04, 1,207.35, 463.44, 61.61 และ 8.29 ต่อประชากรแสนคน ตามลำดับ ภาคที่ได้รับรายงานมากที่สุด คือ ภาคกลาง รองลงมา คือ ภาคตะวันออกเฉียงเหนือ ภาคเหนือและภาคใต้ โดยมีอัตราป่วยเท่ากับ 643.16, 603.46, 521.66, 447.33 ต่อประชากรแสนคน ตามลำดับ จังหวัดที่มีอุบัติการณ์ต่อประชากรแสนคน สูงสุด 10 จังหวัด คือ พิจิตร รองลงมา คือ ปราจีนบุรี มหาสารคาม สุราษฎร์ธานี เลย จันทบุรี น่าน ปังกาฬนครปฐม และอ่างทอง ผู้ป่วยส่วนใหญ่เข้ารับการรักษาที่โรงพยาบาลส่งเสริมสุขภาพตำบล ร้อยละ 59.38 รองลงมา คือ โรงพยาบาลชุมชน ร้อยละ 24.94 โรงพยาบาลศูนย์/โรงพยาบาลทั่วไป ร้อยละ 9.94 และอื่น ๆ ร้อยละ 5.74 โดยอมร [3]

ด้วยสาเหตุดังกล่าวข้างต้น ผู้วิจัยจึงให้ความสนใจในการเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานของผู้มาตรวจร่างกายในโรงพยาบาลซึ่งการเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานขึ้นอยู่กับปัจจัยหลายอย่าง เช่น เพศ อายุ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย การออกกำลังกาย ความดันโลหิตสูง ความเครียด การดื่มเครื่องดื่มที่มีแอลกอฮอล์ กรรมพันธุ์ การรับประทานอาหาร ฯลฯ โดยผู้วิจัยสนใจศึกษาการเปรียบเทียบประสิทธิภาพและทำนายผลการเป็นโรคเบาหวานของผู้มาตรวจร่างกายที่โรงพยาบาล 6 วิธี คือ วิธีความใกล้เคียงกันมากที่สุด วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม และวิธีนาอ็ฟเบย์ เพื่อหาวิธีที่มีประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานที่เหมาะสมต่อไป

2. วิธีการวิจัย

2.1 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบฟอร์มใบเข้ามตรวจร่างกายและในการวิเคราะห์ข้อมูลใช้โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) เวอร์ชัน 3.7.4 ซึ่งเป็นโปรแกรมที่สามารถดาวน์โหลดได้จากเว็บไซต์ อยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม WEKA พัฒนามาจากภาษาจาวาทั้งหมด เป็นที่นิยมในการใช้งานด้านการทำเหมืองข้อมูล

2.2 การเก็บรวบรวมข้อมูล การบันทึกและแบ่งข้อมูล

2.2.1 การเก็บรวบรวมข้อมูล เก็บรวบรวมข้อมูลผลการวินิจฉัยของผู้มาตรวจร่างกายในโรงพยาบาลแห่งหนึ่ง ตั้งแต่เดือนมกราคมถึงเดือนเมษายน พ.ศ. 2559 จำนวน 1,233 คน โดยข้อมูลประกอบด้วยคุณลักษณะ (attribute) ต่าง ๆ ดังนี้ เพศ อายุ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย (BMI) การออกกำลังกาย ความดันโลหิตสูง ความเครียด การดื่มเครื่องดื่มที่มีแอลกอฮอล์ กรรมพันธุ์ การรับประทานอาหาร และผลการตรวจวินิจฉัยการเป็นโรคเบาหวาน

2.2.2 การบันทึกข้อมูล บันทึกผลการวินิจฉัยการเป็นโรคเบาหวานของผู้มาตรวจร่างกายจำนวน 1,233 คน ลงในโปรแกรม Microsoft Excel โดยให้แนวคอลัมน์เป็นเพศ อายุ น้ำหนัก ส่วนสูง ค่าดัชนีมวลกาย การออกกำลังกาย ความดันโลหิตสูง ความเครียด การดื่มเครื่องดื่มที่มีแอลกอฮอล์ กรรมพันธุ์ การรับประทานอาหาร และผลการวินิจฉัยการเป็นโรคเบาหวานของผู้มาตรวจร่างกาย ส่วนแนวแถวเป็นลำดับที่ของผู้มาตรวจร่างกายการเป็นโรคเบาหวาน

2.2.3 การแบ่งข้อมูล นำข้อมูลทั้งหมดมาแบ่งออกเป็น 3 ส่วน จากข้อมูลการเป็นโรคเบาหวานของผู้ที่มาตรวจร่างกาย จำนวน 1,233 คน เพื่อใช้

สำหรับการวิเคราะห์ข้อมูล โดยแบ่งข้อมูลออกเป็นสัดส่วนดังนี้ [4]

ส่วนที่ 1 : ข้อมูลชุดฝึกหัด (training data set) เพื่อนำไปสร้างตัวแบบ มีข้อมูล 70 % ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 1 จำนวน 863 คน

ส่วนที่ 2 : ข้อมูลชุดทดสอบประสิทธิภาพ (evaluation data set) เพื่อนำไปทดสอบความถูกต้องของตัวแบบ มีข้อมูล 20 % ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 2 จำนวน 247 คน

ส่วนที่ 3 : ข้อมูลชุดทดสอบ (testing data set) เพื่อนำไปทำนายตัวแบบ มีข้อมูล 10 % ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 3 จำนวน 123 คน

2.3 การวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ ผู้วิจัยได้จัดข้อมูลแต่ละชุดออกเป็น 3 ส่วน โดยส่วนที่ 1 ใช้ข้อมูล 70 % ของข้อมูลทั้งหมดในการสร้างตัวแบบ ส่วนที่ 2 ใช้ข้อมูล 20 % ของข้อมูลทั้งหมดในการทดสอบความถูกต้องของตัวแบบ ส่วนที่ 3 ใช้ข้อมูล 10 % ของข้อมูลทั้งหมดในการทำนายตัวแบบ แปลงไฟล์ข้อมูลให้เป็นนามสกุล *.csv เพื่อใช้วิเคราะห์ประสิทธิภาพการจำแนกกลุ่มข้อมูลในโปรแกรม WEKA ซึ่งเป็นโปรแกรมที่สามารถนำมาทดสอบอัลกอริทึมของวิธีการจำแนกกลุ่มได้เนื่องจากมีอัลกอริทึมที่ได้รับอนุญาตให้เลือกใช้ในโปรแกรมครบตามที่กำหนด ผู้วิจัยได้กำหนดวิธีการจำแนกกลุ่มเพื่อนำมาทดสอบดังนี้

2.3.1 วิธีความใกล้เคียงกันมากที่สุด (K-nearest neighbor) ใช้ัลกอริทึมชนิด IBk เนื่องจากเป็นฟังก์ชันหลักที่สนใจ ซึ่งเป็นพื้นฐานของอัลกอริทึม 8.1 อัลกอริทึม IBk ยังสามารถกำหนดน้ำหนักระยะทางและทางเลือก (option) เพื่อกำหนดค่า k โดยใช้ cross-validation [5]

2.3.2 วิธีต้นไม้ตัดสินใจ (decision tree)

ใช้ัลกอริทึมชนิด J48 ซึ่งพัฒนามาจาก ID3 สามารถใช้ได้กับข้อมูลแบบไม่ต่อเนื่องและแบบต่อเนื่อง ต่างจาก ID3 ที่ใช้ได้เพียงข้อมูลแบบไม่ต่อเนื่องเท่านั้น [6]

2.3.3 วิธีโครงข่ายประสาทเทียม (artificial neural network)

ใช้ัลกอริทึมชนิดเพอร์เซปตรอนหลายชั้น (multilayer perceptron) โดยกำหนดค่าอัตราการเรียนรู้ (learning rate) เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 ค่าโมเมนตัม (momentum) เป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 จำนวนรอบการสอน (training time) 20,000 รอบ การวิจัยครั้งนี้ใช้ัลกอริทึมของวิธีโครงข่ายประสาทชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (hidden layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัติการกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล [7]

2.3.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน (support vector machine)

ใช้ัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล (polynomial Kernel) เนื่องจากงานวิจัยที่อ้างอิงจาก ภรณ์ยา และคณะ [8] ได้ผลว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ัลกอริทึมชนิดโพลิโนเมียลเคอร์เนลดีที่สุด

2.3.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (binary logistic regression)

เป็นการวิเคราะห์การถดถอยแบบหนึ่งโดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือเชิงคุณภาพ หรืออาจมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้ [9]

2.3.6 วิธีนาอิวเบย์ (Naïve Bayes)

คืออัลกอริทึมที่ใช้หลักการของความน่าจะเป็นในการคัดกรอง แต่ละคำตอบ(class)โดยมีคำตอบ 2 คำตอบ[10]

การนำผลการวิเคราะห์ข้อมูลมาประเมินผลเพื่อเปรียบเทียบประสิทธิภาพในการ

ทำนายผลการเป็นโรคเบาหวานของผู้มาตรวจร่างกาย ใช้วิธีการจำแนกกลุ่ม 6 วิธี โดยนำข้อมูลที่ได้จากการวิเคราะห์มาเปรียบเทียบว่าวิธีในการจำแนกกลุ่มข้อมูลวิธีใดที่มีค่าความถูกต้อง (accuracy) มากที่สุด ส่วนค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error, MSE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error, MSE) ที่มีค่าน้อยที่สุด จะทำให้มีประสิทธิภาพในการทำนายผลดีที่สุด โดยที่ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นมาตรวัดการประเมินค่าได้ดี เนื่องจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยประกอบด้วยทั้งความแปรปรวนและความเอนเอียง [11] ส่วนค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ก็มีในงานวิจัยนำมาเปรียบเทียบกันด้วย ดังนั้นในงานวิจัยนี้จึงใช้ทั้งค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

3. ผลการวิจัย

3.1 การทำนายการเป็นโรคเบาหวาน

การทำนายการเป็นโรคเบาหวานจากชุดข้อมูลทดสอบ จำนวน 123 คน โดยวิธีการจำแนกกลุ่มทั้ง 6 วิธี ได้ผลดังนี้

3.1.1 วิธีความใกล้เคียงกันมากที่สุด

ตารางที่ 1 ผลในส่วนของการสรุปผลจากการทำนายการเป็นโรคเบาหวานด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly classified instances	108	87.8049 %
Mean absolute error	0.1228	
Mean squared error	0.1217	

ตารางที่ 1 จากข้อมูลจำนวน 123 คน ทำนายข้อมูลถูกต้องจำนวน 108 คน คิดเป็น 87.8049 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ

0.1228 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควรและมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.1217 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 2 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการเป็นโรคเบาหวานด้วยวิธีความใกล้เคียงกันมากที่สุด

TP rate	FP rate	Precision	Recall	F-measure	Class
0.939	1	0.931	0.939	0.935	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0	0.061	0	0	0	positive

ตารางที่ 2 สำหรับคำตอบไม่เป็นโรคเบาหวาน (class = negative) มีค่าอัตราความถูกต้องเชิงบวก = 0.939 ค่าอัตราความผิดพลาดเชิงบวก = 1 ค่าความแม่นยำ = 0.931 ค่าความระลึก = 0.939 และค่าความถ่วงดุล = 0.935 ส่วนคำตอบเป็นโรคเบาหวาน (class = positive) มีค่าอัตราความถูกต้องเชิงลบ = 0 ค่าอัตราความผิดพลาดเชิงลบ = 0.061 ค่าความแม่นยำ = 0 ค่าความระลึก = 0 และค่าความถ่วงดุล = 0

ตารางที่ 3 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการเป็นโรคเบาหวานด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่แท้จริง	ไม่เป็น	108	7
	เป็น	8	0

ตารางที่ 3 มีข้อมูล 123 คน ตัวแบบ

สามารถทำนายข้อมูลได้ถูกต้อง 108 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าไม่เป็นโรคเบาหวาน 108 คน และจำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 0 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 15 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วไม่เป็นโรคเบาหวาน 7 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วเป็นโรคเบาหวาน 8 คน

3.1.2 วิธีต้นไม้ตัดสินใจ

ตารางที่ 4 ผลในส่วนของการสรุปผลจากการทำนายการเป็นโรคเบาหวานด้วยวิธีต้นไม้ตัดสินใจ

Correctly classified instances	115	93.4959 %
Mean absolute error	0.1137	
Mean squared error	0.0515	

ตารางที่ 4 โดยจากข้อมูลจำนวน 123 คน ทำนายข้อมูลถูกต้องจำนวน 115 คน คิดเป็น 93.4959 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1137 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.0515 ซึ่งมีค่าน้อยมาก แสดงว่าตัวแบบมีความถูกต้องดี

ตารางที่ 5 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการเป็นโรคเบาหวานด้วยวิธีต้นไม้ตัดสินใจ

TP rate	FP rate	Precision	Recall	F-measure	Class
1	1	0.935	1	0.966	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0	0	0	0	0	positive

ตารางที่ 5 สำหรับคำตอบไม่เป็นโรคเบาหวาน (class = negative) มีค่าอัตราความถูกต้องเชิงบวก = 1 ค่าอัตราความผิดพลาดเชิงบวก = 1 ค่าความแม่นยำ = 0.935 ค่าความระลึก = 1 และค่าความถ่วงดุล = 0.966 ส่วนคำตอบเป็นโรคเบาหวาน (class = positive) มีค่าอัตราความถูกต้องเชิงลบ = 0 ค่าอัตราความผิดพลาดเชิงลบ = 0 ค่าความแม่นยำ = 0 ค่าความระลึก = 0 และค่าความถ่วงดุล = 0

ตารางที่ 6 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการเป็นโรคเบาหวานด้วยวิธีต้นไม้ตัดสินใจ

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่แท้จริง	ไม่เป็น	115	0
	เป็น	8	0

ตารางที่ 6 มีข้อมูล 123 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 115 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าไม่เป็นโรคเบาหวาน 115 คน และจำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 0 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 8 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วไม่เป็นโรคเบาหวาน 0 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วเป็นโรคเบาหวาน 8 คน

3.1.3 วิธีโครงข่ายประสาทเทียม

ตารางที่ 7 จากข้อมูลจำนวน 123 คน ทำนายข้อมูลถูกต้องจำนวน 118 คน คิดเป็น 95.935 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0491 ซึ่งมีค่าน้อยมาก แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริง และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.0396 ซึ่งมีค่าน้อยมาก

แสดงว่าตัวแบบมีความถูกต้องดี

ตารางที่ 7 ผลในส่วนของการสรุปผลจากการทำนาย
การเป็นโรคเบาหวานด้วยวิธีโครงข่าย
ประสาทเทียม

Correctly classified instances	118	95.935 %
Mean absolute error	0.0491	
Mean squared error	0.0396	

ตารางที่ 8 ผลในส่วนของความถูกต้องของราย
ละเอียดในแต่ละคำตอบจากการทำนาย
การเป็นโรคเบาหวานด้วยวิธีโครงข่าย
ประสาทเทียม

TP rate	FP rate	Precision	Recall	F-measure	Class
0.983	0.375	0.974	0.983	0.978	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0.625	0.017	0.714	0.625	0.667	positive

ตารางที่ 8 สำหรับคำตอบไม่เป็นโรคเบาหวาน (class = negative) มีค่าอัตราความถูกต้องเชิงบวก = 0.983 ค่าอัตราความผิดพลาดเชิงบวก = 0.375 ค่าความแม่นยำ = 0.974 ค่าความระลึก = 0.983 และค่าความถ่วงดุล = 0.978 ส่วนคำตอบเป็นโรคเบาหวาน (class = positive) มีค่าอัตราความถูกต้องเชิงลบ = 0.625 ค่าอัตราความผิดพลาดเชิงลบ = 0.017 ค่าความแม่นยำ = 0.714 ค่าความระลึก = 0.625 และค่าความถ่วงดุล = 0.667

ตารางที่ 3.9 มีข้อมูล 123 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 118 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าไม่เป็นโรคเบาหวาน 113 คน และจำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 5 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 5 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริง

แล้วไม่เป็นโรคเบาหวาน 2 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วเป็นโรคเบาหวาน 3 คน

ตารางที่ 9 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการเป็นโรคเบาหวานด้วยวิธีโครงข่ายประสาทเทียม

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่แท้จริง	ไม่เป็น	113	2
	เป็น	3	5

3.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน

ตารางที่ 10 ผลในส่วนของการสรุปผลจากการทำนายการเป็นโรคเบาหวานด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly classified instances	115	93.4959 %
Mean absolute error	0.0650	
Mean squared error	0.0650	

ตารางที่ 10 โดยจากข้อมูลจำนวน 123 คน ทำนายข้อมูลถูกต้องจำนวน 115 คน คิดเป็น 93.4959 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0650 ซึ่งมีค่าน้อยมาก แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริง และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.0650 ซึ่งมีค่าน้อยมาก แสดงว่าตัวแบบมีความถูกต้องดี

ตารางที่ 11 สำหรับคำตอบไม่เป็นโรคเบาหวาน (class = negative) มีค่าอัตราความถูกต้องเชิงบวก = 1 ค่าอัตราความผิดพลาดเชิงบวก = 1 ค่าความแม่นยำ = 0.935 ค่าความระลึก = 1 และค่าความถ่วงดุล = 0.966 ส่วนคำตอบเป็นโรคเบาหวาน

(class = positive) มีค่าอัตราความถูกต้องเชิงลบ = 0
ค่าอัตราความผิดพลาดเชิงลบ = 0 ค่าความแม่นยำ
= 0 ค่าความระลึก = 0 และค่าความถ่วงดุล = 0

ตารางที่ 11 ผลในส่วนของความถูกต้องของราย
ละเอียดในแต่ละคำตอบจากการ
ทำนายการเป็นโรคเบาหวานด้วยวิธี
ซัพพอร์ตเวกเตอร์แมชชีน

TP rate	FP rate	Precision	Recall	F-measure	Class
1	1	0.935	1	0.966	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0	0	0	0	0	positive

ตารางที่ 12 ผลในส่วนของเมตริกซ์ความสับสนจาก
การทำนายการเป็นโรคเบาหวานด้วยวิธี
ซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่ แท้จริง	ไม่เป็น	115	0
	เป็น	8	0

ตารางที่ 12 มีข้อมูล 123 คน ตัวแบบ
สามารถทำนายข้อมูลได้ถูกต้อง 115 คน โดยมีจำนวน
ข้อมูลที่จำแนกถูกว่าไม่เป็นโรคเบาหวาน 115 คน และ
จำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 0 คน
ตัวแบบทำนายข้อมูลไม่ถูกต้อง 8 คน โดยมีจำนวน
ข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริง
แล้วไม่เป็นโรคเบาหวาน 0 คน และจำนวนข้อมูลที่
จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้ว
เป็นโรคเบาหวาน 8 คน

3.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

ตารางที่ 13 โดยจากข้อมูลจำนวน 123
คน ทำนายข้อมูลถูกต้องจำนวน 116 คน คิดเป็น

94.3089 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย
(MAE) คือ 0.0910 ซึ่งมีค่าน้อยมาก แสดงว่าค่าที่
ทำนายได้ใกล้เคียงกับค่าที่แท้จริง และมีค่าความ
คลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.0400 ซึ่งมีค่า
น้อยมาก แสดงว่าตัวแบบมีความถูกต้องดี

ตารางที่ 13 ผลในส่วนของการสรุปผลจากการทำนาย
การเป็นโรคเบาหวานด้วยวิธีการถดถอย
โลจิสติกแบบ 2 กลุ่ม

Correctly classified instances	115	94.3089 %
Mean absolute error	0.0910	
Mean squared error	0.0400	

ตารางที่ 14 ผลในส่วนของความถูกต้องของราย
ละเอียดในแต่ละคำตอบจากการ
ทำนายการเป็นโรคเบาหวานด้วยวิธี
การถดถอยโลจิสติกแบบ 2 กลุ่ม

TP rate	FP rate	Precision	Recall	F-measure	Class
0.991	0.750	0.950	0.991	0.970	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0.250	0.009	0.667	0.250	0.364	positive

ตารางที่ 14 สำหรับคำตอบไม่เป็น
โรคเบาหวาน (class = negative) มีค่าอัตราความ
ถูกต้องเชิงบวก = 0.991 ค่าอัตราความผิดพลาดเชิง
บวก = 0.750 ค่าความแม่นยำ = 0.950 ค่าความระลึก
= 0.991 และค่าความถ่วงดุล = 0.970 ส่วนคำตอบ
เป็นโรคเบาหวาน (class = positive) มีค่าอัตราความ
ถูกต้องเชิงลบ = 0.250 ค่าอัตราความผิดพลาดเชิงลบ
= 0.009 ค่าความแม่นยำ = 0.667 ค่าความระลึก
= 0.250 และค่าความถ่วงดุล = 0.364

ตารางที่ 15 มีข้อมูล 123 คน ตัวแบบ
สามารถทำนายข้อมูลได้ถูกต้อง 116 คน โดยมีจำนวน

ข้อมูลที่จำแนกถูกว่า ไม่เป็นโรคเบาหวาน 114 คน และจำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 2 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 7 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วไม่เป็นโรคเบาหวาน 1 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วเป็นโรคเบาหวาน 6 คน

ตารางที่ 15 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการเป็นโรคเบาหวานด้วยวิธีการถดถอยโลจิสติก 2 กลุ่ม

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่แท้จริง	ไม่เป็น	114	1
	เป็น	6	2

3.1.6 วิธีนัฟฟาย

ตารางที่ 16 ผลในส่วนของการสรุปผลจากการทำนายการเป็นโรคเบาหวานด้วยวิธีนัฟฟาย

Correctly classified instances	115	91.0569 %
Mean absolute error	0.1325	
Mean squared error	0.0877	

ตารางที่ 16 โดยจากข้อมูลจำนวน 123 คน ทำนายข้อมูลถูกต้องจำนวน 112 คน คิดเป็น 91.0569 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1325 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ 0.0877 ซึ่งมีค่าน้อยมาก แสดงว่าตัวแบบมีความถูกต้องดี

ตารางที่ 3.17 สำหรับคำตอบไม่เป็นโรคเบาหวาน (class = negative) มีค่าอัตราความ

ถูกต้องเชิงบวก = 0.930 ค่าอัตราความผิดพลาดเชิงบวก = 0.375 ค่าความแม่นยำ = 0.973 ค่าความระลึก = 0.930 และค่าความถ่วงดุล = 0.961 ส่วนคำตอบเป็นโรคเบาหวาน (class = positive) มีค่าอัตราความถูกต้องเชิงลบ = 0.625 ค่าอัตราความผิดพลาดเชิงลบ = 0.070 ค่าความแม่นยำ = 0.385 ค่าความระลึก = 0.625 และค่าความถ่วงดุล = 0.476

ตารางที่ 17 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการเป็นโรคเบาหวานด้วยวิธีนัฟฟาย

TP rate	FP rate	Precision	Recall	F-measure	Class
0.930	0.375	0.973	0.930	0.961	negative
TN rate	FN rate	Precision	Recall	F-measure	Class
0.625	0.070	0.385	0.625	0.476	positive

ตารางที่ 18 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการเป็นโรคเบาหวานด้วยวิธีนัฟฟาย

		ผลการจำแนกการเป็นโรคเบาหวาน	
		ไม่เป็น	เป็น
ผลที่แท้จริง	ไม่เป็น	107	8
	เป็น	3	5

ตารางที่ 18 มีข้อมูล 123 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 112 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าไม่เป็นโรคเบาหวาน 107 คน และจำนวนข้อมูลที่จำแนกถูกว่าเป็นโรคเบาหวาน 5 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 11 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าเป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้วเป็นโรคเบาหวาน 8 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่เป็นโรคเบาหวาน ซึ่งผลที่แท้จริงแล้ว

เป็นโรคเบาหวาน 3 คน

3.2 ผลการเปรียบเทียบประสิทธิภาพในการทำนายผลของวิธีการจำแนกกลุ่ม

การเปรียบเทียบประสิทธิภาพในการทำนายผลการพิจารณาการเป็นโรคเบาหวาน ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีต้นไม้ตัดสินใจ วิธี

โครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม และวิธีนาอ์ฟเบย์ โดยพิจารณาจากค่าความถูกต้อง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ได้ผลดังตารางที่ 19

ตารางที่ 19 ผลการเปรียบเทียบประสิทธิภาพในการทำนายผลของวิธีการจำแนกกลุ่มในการพิจารณาการเป็นโรคเบาหวานทั้ง 6 วิธี

การเปรียบเทียบประสิทธิภาพในการทำนายผล	ค่าความถูกต้อง (accuracy)	ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE)
วิธีความใกล้เคียงกันมากที่สุด	87.80 %	0.1228	0.1212
วิธีต้นไม้ตัดสินใจ	93.50 %	0.1137	0.0515
วิธีโครงข่ายประสาทเทียม	95.94 %	0.0491	0.0396
วิธีซัพพอร์ตเวกเตอร์แมชชีน	93.50 %	0.0650	0.0650
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	94.31 %	0.0910	0.0400
วิธีนาอ์ฟเบย์	91.06 %	0.1325	0.0877

ตารางที่ 19 พบว่าวิธีโครงข่ายประสาทเทียมมีค่าความถูกต้องมากที่สุดคือ 95.94 % มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.0491 และ 0.0396 ตามลำดับ แสดงว่ามีความคลาดเคลื่อนน้อยที่สุด ดังนั้นวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลดีที่สุด

4. สรุปผลการวิจัย อภิปรายผลการวิจัย และข้อเสนอแนะ

4.1 สรุปผลการวิจัย

การเปรียบเทียบประสิทธิภาพการทำนายผลของวิธีการจำแนกกลุ่ม 6 วิธี ในการพิจารณาการเป็นโรคเบาหวานได้ผลสรุปดังนี้

4.1.1 วิธีโครงข่ายประสาทเทียมมีค่าความถูกต้องมากที่สุดคือ 95.94 %

4.1.2 วิธีโครงข่ายประสาทเทียมมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด คือ 0.0491 และ 0.0396 ตามลำดับ

ดังนั้นวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานดีที่สุด

4.2 อภิปรายผลการวิจัย

ผลการศึกษาการเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวาน โดยพิจารณาจากค่าความถูกต้อง พบว่าวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลดีที่สุด ซึ่งให้ผลสอดคล้องกับงานวิจัยใน [12] ที่ได้ศึกษาเรื่องการวิเคราะห์ปัจจัยเสี่ยงของโรคเบาหวาน โดยตัวแบบที่ใช้ศึกษาคือตัวแบบ Back-propagation, Neural networks, Radial basis, Function network และตัวแบบ Naïve Bayes ผลการศึกษาแสดงให้เห็นว่าตัว

แบบ Back-propagation และ Neural Networks ให้ผลลัพธ์ที่มีความถูกต้องมากที่สุด ยังให้ผลสอดคล้องกับงานวิจัยของ ขาญชัย และคณะ [13] ที่ศึกษาการวิเคราะห์ข้อมูลความเสี่ยงของการเป็นโรคหอบหืดด้วยเทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ พบว่าเทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับมีความถูกต้องมากกว่าร้อยละ 90 และสามารถนำไปใช้ประกอบการตัดสินใจของผู้ป่วยเพื่อเข้ารับการรักษาตั้งแต่เริ่มต้นได้นอกจากนี้ยังให้ผลสอดคล้องกับงานวิจัยของ สุมณพิชญ์ และอัจฉรา [14] ที่ศึกษาการวิเคราะห์ความเสี่ยงของการเป็นวัณโรคด้วยเทคนิคโครงข่ายประสาทเทียม พบว่าเทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับเพื่อจำแนกข้อมูลผู้ป่วย ได้แก่ ข้อมูลอาการ ประวัติการเป็นโรคของสมาชิกในครอบครัว และปัจจัยเสี่ยงที่มีผลต่อการเป็นวัณโรค จากผลการวิเคราะห์พบว่าวิธีการนี้มีความถูกต้องแม่นยำและสามารถนำไปใช้ประกอบการตัดสินใจของผู้ป่วยเพื่อเข้ารับการรักษาตั้งแต่ระยะแรกของโรคได้

ส่วนการเปรียบเทียบประสิทธิภาพในการทำนายผลการเป็นโรคเบาหวานโดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ย พบว่าวิธีโครงข่ายประสาทเทียมมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ดังนั้นวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลดีที่สุด ซึ่งสอดคล้องกับข้อมูลโรคไตเรื้อรังโดยใช้วิธีซัพพอร์ต-เวกเตอร์แมชชีนและโครงข่ายประสาทเทียม จากการศึกษาพบว่าวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพในการทำนายผลได้ดีกว่าวิธีซัพพอร์ต-เวกเตอร์แมชชีน [15] แต่ให้ผลตรงข้ามกับงานวิจัยของ ณัฐวุฒิ

และคณะ [16] และ สุรวัชร [17] ที่พบว่าวิธีการจำแนกกลุ่มที่มีประสิทธิภาพในการทำนายผลดีที่สุดโดยเปรียบเทียบจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยคือ วิธีต้นไม้ตัดสินใจ ซึ่งอาจมีความเป็นไปได้เนื่องจากในงานวิจัยนี้ใช้ข้อมูลเพียงเรื่องเดียว คือ ข้อมูลการเป็นโรคเบาหวาน ถ้าใช้ข้อมูลจำนวนหลายเรื่องและตัวแปรอยู่ในกลุ่มเดียวกันอาจให้ผลสอดคล้องกันก็มีความเป็นไปได้

4.3 ข้อเสนอแนะ

4.3.1 ตัวแปรที่นำมาใช้ในการงานวิจัยนี้เป็นเพียงส่วนหนึ่งของการพิจารณาการเป็นโรคเบาหวานเท่านั้น เพื่อให้การทำนายมีประสิทธิภาพมากขึ้น ควรเพิ่มตัวแปรที่เกี่ยวข้องอื่น ๆ อีก

4.3.2 เพื่อให้ผลสรุปครอบคลุมกว้างขวางเพิ่มขึ้น ควรจะศึกษาวิธีอื่น ๆ ที่เป็นเทคนิคการทำเหมืองข้อมูลในการจำแนกกลุ่มเหมือนกัน เช่น วิธีฐานกฎ (Base-rule method) วิธีโครงข่ายความเชื่อของเบย์เซียน (Bayesian belief network method)

4.3.3 เพื่อให้ได้ข้อสรุปของผลการวิเคราะห์ข้อมูลที่มีความสมบูรณ์มากขึ้น เราอาจจะใช้อัลกอริทึมประเภทอื่น ๆ โดยวิธีความใกล้เคียงกันมากที่สุดยังมีอัลกอริทึม KStar และ LWL วิธีต้นไม้ตัดสินใจมีอัลกอริทึม Decision stump, LMT, Random forest, Random tree และ REP tree และ ID3 วิธีโครงข่ายประสาทเทียมสามารถกำหนดค่าอัตราการเรียนรู้และค่าโมเมนตัมที่ละเอียดมากขึ้นกว่าเดิม อาจกำหนดจำนวนชั้นซ่อนมากกว่า 1 ชั้นได้ และอาจเพิ่มจำนวนรอบการสอนให้มากขึ้น เช่น 100,000 รอบ และวิธีซัพพอร์ตเวกเตอร์แมชชีนมีอัลกอริทึม SMO และใช้ฟังก์ชันเคอร์เนล (Kernel function) แบบอื่น ๆ ได้แก่ Normalized poly Kernel, Radial basis function Kernel, Sigmoid Kernel และ Puk

5. รายการอ้างอิง

- [1] Iyer, A., Jeyalatha, S. and Sumbaly. R., 2015, Diagnosis of Diabetes Using Classification Mining Techniques, Int. J. Data Min. Knowl. Manag. Process 5: 1-14.
- [2] นุชรี อาบสุวรรณ และนิตยา พันธุเวชย์, 2558, ประเด็นสาธารณสุขคว้นเบหวนโลก, สำนักโรคไม่ติดต่อ กรมควบคุมโรค กระทรวงสาธารณสุข, กรุงเทพฯ.
- [3] อมรา ทองหงษ์, 2555, โรคเบหวน, สรุปรายงานการเฝ้าระวังโรค, สำนักโรคควบคุมโรค กระทรวงสาธารณสุข, กรุงเทพฯ.
- [4] พยูน พาณิษฐ์กุล, 2548, การพัฒนาระบบดาต้าไมนนิ่งโดยใช้ Decision Tree : โครงการพัฒนาระบบงาน, วิทยานิพนธ์ปริญญาโท, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, กรุงเทพฯ.
- [5] Wu, X. and Kumar, V., 2009, The Top Ten Algorithms in Data Mining, University of Minnesota Department of Computer Science and Engineering, CRC Press, Minneapolis, Boca Raton. 215 p.
- [6] รุจิรา ธรรมสมบัติ, 2554, ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ, รายงานการวิจัย, สาขาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ วิทยาลัยราชพฤกษ์, 68 น.
- [7] Berson, A. and Smith, S.J., 1997, Data Warehousing, Data Mining, and OLAP, McGraw-Hill, New York, 612 p.
- [8] ภรณ์ยา อามฤครัตน์, เดช ธรรมศิริ, วาทีน มัยเพียร, ภัทราวุฒิ แสงศิริ, ณรงค์ โพธิ์ และพยุห มีสัจ, 2553, การเปรียบเทียบประสิทธิภาพการคัดเลือกและจำแนกข้อมูลด้วยวิธีการทางเครือข่ายประสาทเทียม, น. 131-138, การประชุมวิชาการระดับชาติทางด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 5, วิทยาลัยนครราชสีมา, นครราชสีมา.
- [9] กัลยา วานิชย์บัญชา, 2552, การวิเคราะห์ข้อมูลหลายตัวแปร, บริษัท ธรรมสาร จำกัด, กรุงเทพฯ, 589 น.
- [10] Larose, D.T., 2006, Data Mining: Methods and Models, John Wiley and Sons, New Jersey, 322 p.
- [11] Larose, D.T., 2005, Discovering knowledge in Data: An Introduction to Data Mining, John Wiley and Sons, Hoboken, N.J., 222 p.
- [12] กิตติพล วิแสง, สิริภัทร เชี่ยวชาญวัฒนา และคำรณ สุนันติ, 2552, การวิเคราะห์ปัจจัยเสี่ยงของโรคเบหวน, น. 798-805, การประชุมวิชาการแห่งชาติทางด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 5, วิทยาลัยนครราชสีมา, นครราชสีมา.
- [13] ขาญชัย พรหมโคตร, ทิพวรรณ ศรีชนะ และอัจฉรา มหาวีร์วัฒน์, 2554, การวิเคราะห์ข้อมูลความเสี่ยงของการเกิดโรคหอบหืดด้วยเทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ, น. 51, การประชุมวิชาการ มอบ.วิจัย ครั้งที่ 5, มหาวิทยาลัยอุบลราชธานี, อุบลราชธานี.
- [14] สุมณพิชญ์ พลศรี และอัจฉรา มหาวีร์วัฒน์, 2554, การวิเคราะห์ความเสี่ยงของการเป็นวัณโรคด้วยเทคนิคโครงข่ายประสาทเทียม, น. 41, การประชุมวิชาการ มอบ.วิจัย ครั้งที่ 5, มหาวิทยาลัยอุบลราชธานี, อุบลราชธานี.

- [15] Vijayarani, S. and Dhayanand, S., 2015, Kidney disease prediction using SVM and an nalogorithms, Int. J. Cybern. Inform. 4(4): 13-25.
- [16] ณัฐวุฒิ ศิริกุลรุ่งโรจน์, รัฐธชัย บุญวิเศษ, สหัทสนันน์ ไพสานต์ และสุรวัชร ศรีเปารยะ, 2556, การเปรียบเทียบประสิทธิภาพการจำแนกประเภทโดยใช้วิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาทเทียมและวิธีซัพพอร์ตเวกเตอร์แมชชีน, ปัญหาพิเศษปริญญาตรี, สาขาวิชาสถิติ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , กรุงเทพฯ, 132 น.
- [17] สุรวัชร ศรีเปารยะ, 2560, การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย, ว.วิทยาศาสตร์และเทคโนโลยี 25(5): 839-853.