

การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการเกิดโรคเบาหวานด้วยเทคนิคเหมืองข้อมูล

เชาว์วรรณ พุทธิรักษา¹ ภวินท์ สีแสง¹ วาสนา จันทิวาสน์¹ สุพจน์ เสงพะพรหม¹ และไกรุ่ง เสงพะพรหม^{1*}

¹สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม

*kairung2011.heng@gmail.com

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบเทคนิคการจำแนกกลุ่มข้อมูลที่เหมาะสมกับการจำแนกข้อมูลการเกิดโรคเบาหวาน ในการวิจัยครั้งนี้ได้ทำการศึกษาคัดเลือกเทคนิคสำหรับใช้ในการวิจัยนี้ทั้งหมด 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และเทคนิคเพื่อนบ้านใกล้เคียงที่สุด เพื่อหาประสิทธิภาพของการจำแนกข้อมูลที่ดีที่สุดในกลุ่มข้อมูลการเกิดโรคเบาหวาน ผลการวิจัยพบว่า เทคนิคที่ให้ผลลัพธ์ที่ดีที่สุดคือ โครงข่ายประสาทเทียม โดยให้ค่าความถูกต้อง เท่ากับ 76.82 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.86 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.80 รองลงมาคือ เทคนิคเพื่อนบ้านใกล้เคียงที่สุด โดยให้ค่าความถูกต้อง เท่ากับ 72.27 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.82 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.77 และสุดท้าย ต้นไม้ตัดสินใจ โดยให้ค่าความถูกต้อง เท่ากับ 71.09 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.89 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.73

คำสำคัญ : การจำแนกข้อมูล ต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้เคียงที่สุด โครงข่ายประสาทเทียม

A Comparison of Data Classification Efficiency for Diabetes Mellitus using Data Mining Techniques

Chaowat Phuttaraksa¹ Pawin Sisang¹ Wassana Juntiwad¹ Supojn Hengpraphorm¹ and
Kairung Hengpraphorm^{1*}

¹ Data Science Faculty of Science and Technology, Nakhon Pathom Rajabhat University

*kairung2011.heng@gmail.com

Abstract

The research aimed to study and compare the techniques classified information to the appropriate classification of diabetes. In this research was to study the technique selected for use in this study, all three techniques include decision trees techniques. Neural networks Technical and closest neighbor to find the composite images of classified information, the best value in the development of diabetes. The results showed that the technique is far better results, Neural networks. The accuracy of 76.82 with a value equal to 0.86 precision by providing accurate value that is equal to 0.80, followed by the nearest neighbor technique. The accuracy of 72.27 by 0.82 by the precision accuracy that is equal to 0.77, and the final decision trees. The accuracy of 71.09 with a value equal to 0.89 precision by providing accurate value that is equal to 0.73.

Keywords: data classification, decision tree, k-NN, artificial neural network

1. บทนำ

โรคเบาหวาน (Diabetes mellitus : DM, Diabetes) เป็นโรคที่มีความผิดปกติเกี่ยวกับการนำน้ำตาลไปใช้ประโยชน์อันเกี่ยวเนื่องกับความบกพร่องของฮอร์โมนอินซูลิน ทำให้ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ หากปล่อยให้ร่างกายอยู่ในสภาวะนี้เป็นเวลานานจะทำให้มีอาการต่าง ๆ เสมอ ก่อให้เกิดอาการและภาวะแทรกซ้อนต่าง ๆ ตามมา ผู้ที่เป็นเบาหวานมักจะมีประวัติคนในครอบครัว (พ่อแม่หรือญาติพี่น้องสายตรง) เป็นโรคนี้ด้วย และมักจะมีภาวะน้ำหนักตัวเกินร่วมด้วย

เบาหวานเป็นโรคที่พบได้สูงในคนทุกอายุและทั้งสองเพศ และพบได้สูงขึ้นตามอายุที่มากขึ้น

โรคเบาหวานมีสาเหตุมาจากการบกพร่องของฮอร์โมนอินซูลิน ผู้ที่เป็นเบาหวานจะพบว่าตับอ่อนผลิตอินซูลินได้น้อยหรือผลิตไม่ได้เลย หรือผลิตได้ปกติ แต่ประสิทธิภาพของอินซูลินลดลง เช่นที่พบในคนอ้วน ซึ่งเรียกว่า “ภาวะดื้อต่ออินซูลิน” (Insulin resistance) เมื่อขาดอินซูลินหรืออินซูลินทำหน้าที่ไม่ได้ น้ำตาลในเลือดจึงเข้าสู่เซลล์ต่าง ๆ ได้น้อยกว่าปกติ จึงทำให้เกิดการคั่งของน้ำตาลในเลือด และน้ำตาลก็จะถูกขับออกมาทางปัสสาวะ นี่จึงเป็นสาเหตุว่าทำไมเราถึงเรียกโรคนี้ว่า “เบาหวาน”

ผู้ป่วยเบาหวานที่เป็นมาก (มีระดับน้ำตาลในเลือดสูงมาก) มักจะมีอาการปัสสาวะบ่อยและมาก เพราะน้ำตาลที่ออกมาทางไตจะดึงเอาน้ำออกมาด้วย จึงทำให้มีปัสสาวะออกมามากกว่าปกติ เมื่อผู้ป่วยถ่ายปัสสาวะมากก็จะทำให้รู้สึกกระหายน้ำจนต้องคอยดื่มน้ำบ่อย ๆ และเนื่องจากร่างกายของผู้ป่วยเบาหวานจะไม่สามารถนำน้ำตาลมาเผาผลาญเป็นพลังงานได้ ร่างกายจึงหันมาเผาผลาญกล้ามเนื้อและไขมันแทน จึงทำให้ร่างกายผอม กล้ามเนื้อฝ่อลีบ ไม่มีไขมันอ่อนเปลี้ยเพลียแรง นอกจากนี้ การมีน้ำตาลในเลือดสูงเป็นเวลานาน ๆ ยังทำให้อวัยวะต่าง ๆ เกิดการเปลี่ยนแปลงผิดปกติและนำมาซึ่งภาวะแทรกซ้อนได้มากมาย

ดังนั้นในงานวิจัยนี้จึงเห็นถึงความสำคัญของการจำแนกกลุ่มผู้ที่มีอาการโรคเบาหวาน เพื่อสำหรับเป็นแนวทางในการสนับสนุนการตัดสินใจทางการแพทย์เกี่ยวกับการวินิจฉัยโรคเบาหวาน โดยในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบเทคนิคการจำแนกกลุ่มข้อมูลที่เหมาะสมกับการจำแนกข้อมูลของโรคเบาหวาน โดยทำการศึกษาทั้ง 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้เคียงที่สุด เพื่อหาเทคนิคที่ดีที่สุด สำหรับการใช้ในการจำแนกกลุ่มอาการของโรคเบาหวาน และนำเสนอเทคนิคที่ดีที่สุดสำหรับการจำแนกกลุ่มข้อมูลการเกิดโรคเบาหวาน

2. วัตถุประสงค์

เพื่อเปรียบเทียบประสิทธิภาพเทคนิคการจำแนกข้อมูลสำหรับโรคเบาหวาน ทั้งหมด 3 วิธี ได้แก่ ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และเทคนิคเพื่อนบ้านใกล้เคียงที่สุด

3. งานวิจัยที่เกี่ยวข้อง

3.1 ทฤษฎีที่เกี่ยวข้อง

เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

เทคนิคต้นไม้ช่วยตัดสินใจ (เดซ ธรรมศิริ, 2554) เป็นเทคนิคที่ได้ผลลัพธ์เป็นแบบจำลองการพยากรณ์ของข้อมูลในรูปแบบของต้นไม้ช่วยตัดสินใจ (Decision Tree) เทคนิคนี้ถือได้ว่าเป็นวิธีการจัดหมวดหมู่หรือแบ่งประเภท (Classification) มี การทำงานเป็นแบบการเรียนรู้โดยมีผู้สอน (Supervised Learning) ซึ่งเป็นเทคนิคของการทำเหมืองข้อมูล (Data Mining) อย่างหนึ่งที่เป็นที่นิยมใช้งานกันเนื่องจากเป็นวิธีที่ง่ายไม่ซับซ้อน

โครงข่ายประสาทเทียม (Artificial Neural Networks - ANN)

โครงข่ายประสาทเทียม (วิทยา พรพัชรพงศ์, 2562) เป็นเทคนิคทางเหมืองข้อมูลที่ให้คอมพิวเตอร์จำลองตัวเองให้ทำงานเลียนแบบการทำงานของเซลล์ประสาทในสมองของมนุษย์ที่สามารถฝึกฝน เรียนรู้ตลอดถึงการนำความรู้หรือทักษะไปประยุกต์ใช้งานได้จริง โดยมีการทำงานในรูปแบบของการประมวลผลพร้อม ๆ กันเป็นจำนวนมาก ซึ่งเทคนิคนี้เป็นวิทยาการแขนงหนึ่งทางปัญญาประดิษฐ์ (Artificial Intelligence)

ในนิวรอล (Neurons) จะประกอบด้วยข้อมูลเข้า (Inputs) และข้อมูลออก (Outputs) แต่ละข้อมูลเข้าจะมีการจำลอง ค่าน้ำหนัก (Weight) ในนิวรอลมีค่าเทรชโฮลด์ (Threshold) ซึ่งเป็นตัวกำหนดค่าน้ำหนักรวมของข้อมูลเข้า ว่าต้องมีค่ามากเท่าไรถึงจะสามารถส่งข้อมูลออกไปยังนิวรอลอื่นได้

เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor: KNN)

เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (นพมาศ ปิกเข็มและคณะ, 2560) เป็นเทคนิคที่ใช้สำหรับการจัดกลุ่มของข้อมูลโดยคำนวณจากระยะห่างของแต่ละคุณลักษณะในข้อมูล (Data) ซึ่งวิธีนี้จะเหมาะสมกับข้อมูลที่เป็นเชิงตัวเลข การจัดข้อมูลที่อยู่ใกล้กัน ให้เป็นกลุ่มเดียวกันจะตรวจสอบจากเงื่อนไขของข้อมูลจะตรวจสอบจากจำนวน K ที่กำหนดไว้แต่เทคนิคนี้จะใช้ระยะเวลาในการประมวลผลที่นาน ถ้าข้อมูลมีปริมาณมากอาจเกิดปัญหาในการคำนวณและใช้ปริมาณทรัพยากรในการประมวลผลสูงมาก เนื่องจากจะใช้เวลาสำหรับการประมวลผลเพิ่มขึ้นตามจำนวนข้อมูลที่เพิ่มขึ้นทั้งหมด

3.2 งานวิจัยที่เกี่ยวข้อง

หทัย แก้ววรรณ และชริยา นนทกาญจน์ (2561) ศึกษาเรื่อง การจำแนกความสูงของกล้วยหอมโดนโครงข่ายประสาทเทียม ความน่าจะเป็นด้วยฟังก์ชันความน่าจะเป็น Generalized Extreme Value โดยสามารถจำแนกระดับความสูงของกล้วยออกเป็น 4 ระดับคือ ดิบ สุกปานกลาง สุก และสุกงอม จากผลการทดลองแสดงให้เห็นว่าค่าพารามิเตอร์ของ GEV ในแต่ละระดับความสูง จะมีความคล้ายคลึงกัน ดังนั้นเมื่อนำมา Train ด้วย Probability Neuron Network: PNN ซึ่งเป็นโมเดลที่มีขนาดเล็กและ มีความซับซ้อนน้อย จากผลการทดลองแสดงให้เห็นว่าผลลัพธ์จากวิธีการที่นำเสนอเป็นที่น่าพึงพอใจ และสามารถพัฒนาต่อไป ในอนาคตเพื่อใช้งานแทนแรงงานมนุษย์ได้

พรเทพ ด้านน้อย และสุวิมล มรรควิบูลย์ชัย (2561) ศึกษาเรื่อง การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการใช้บริการเครือข่ายอินเทอร์เน็ตของสถาบันอุดมศึกษา ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการใช้บริการเครือข่ายอินเทอร์เน็ตของสถาบัน อุดมศึกษา โดยเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลของ 3 เทคนิค คือ เทคนิค Decision Tree เทคนิค K-NN และ เทคนิค Deep Learning โดยการนำข้อมูลเกี่ยวกับข้อมูลการใช้บริการเครือข่ายอินเทอร์เน็ตของสถาบันอุดมศึกษา มีข้อมูลจำนวนที่สามารถใช้ในงานวิจัยได้ 8,530 ชุด ซึ่งทำการแบ่งข้อมูลด้วยวิธี Cross-validation Test โดยการสุ่มข้อมูลเพื่อแบ่ง ข้อมูลออกเป็น 5 ส่วน ในแต่ละส่วนประกอบด้วยข้อมูล จำนวน 1,706 ชุด โดยสร้างโมเดลจากการเลือกข้อมูล 4 ส่วน และ ทดสอบประสิทธิภาพของระบบด้วยข้อมูล 1 ส่วน ผลการวิจัยพบว่า เทคนิคที่ใช้ในการจำแนกข้อมูลการใช้บริการเครือข่าย อินเทอร์เน็ตของสถาบันอุดมศึกษา ที่มีประสิทธิภาพสูงสุด คือ Deep Learning โดยมีค่าความแม่นยำ 88.79% ค่าความ ระลึก 88.55% ค่าความแม่นยำ 89.10% และค่าถ่วงดุล 0.289 ซึ่งเป็นระดับการประเมินที่สามารถยอมรับได้

จักรกฤษณ์ หงส์เวียงจันทร์ นิติมา ลักขณานุรักษ์ และไกรรุ่ง เสงพระพรหม (2562) ศึกษาเรื่อง การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลโรคคอทิสติกด้วยเทคนิคเหมืองข้อมูล การจำแนก ข้อมูลโรคคอทิสติก ในการครั้งนี้ได้ทำการศึกษาคัดเลือกเทคนิคสำหรับใช้ในการวิจัยนี้ทั้งหมด 3 เทคนิค ได้แก่ เทคนิค ต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคเพื่อนบ้านใกล้เคียงที่สุด เพื่อหาประสิทธิภาพของการจำแนกข้อมูล ที่ดีที่สุดในกลุ่มข้อมูลโรคคอทิสติก ผลการวิจัยพบว่า เทคนิคที่ให้ค่าความแม่นยำมากที่สุด ซึ่งให้ประสิทธิภาพที่ดีที่สุด คือ เทคนิคโครงข่ายประสาทเทียม โดยให้ค่าความแม่นยำอยู่ที่ 99.65 %

สุระสิทธิ์ ทรงม้า (2558) ศึกษาเรื่อง การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ รุ่นที่ 3 ที่ใช้เทคนิคการแบ่งข้อมูลที่แตกต่างกัน ผลการวิจัยพบว่า ประสิทธิภาพของการใช้เทคนิคแบ่งข้อมูลด้วยขนาดความกว้างเท่ากันและจำแนกข้อมูลด้วยต้นไม้ตัดสินใจรุ่นที่ 3 มีประสิทธิภาพสูงที่สุด โดยมีอัตราการจำแนกข้อมูลถูกต้อง เท่ากับ 92.30% มีอัตราการจำแนกข้อมูลผิดพลาด เท่ากับ 4.89% และสามารถจำแนกข้อมูลที่ไม่สามารถจัดเข้ากลุ่มได้เท่ากับ 2.81% ผลการทดลองยังแสดงให้เห็นว่าเทคนิคการแบ่งข้อมูลชนิดแบบไม่มีผู้สอนแต่ละชนิดส่งผลต่อประสิทธิภาพ การจำแนกข้อมูลของต้นไม้ตัดสินใจ รุ่นที่ 3 และพบว่าการใช้เทคนิคการแบ่งข้อมูลร่วมกับเทคนิคต้นไม้ตัดสินใจ รุ่นที่ 3 สามารถช่วยกรองข้อมูลที่ไม่อยู่ในกฎที่ สร้างขึ้นได้ซึ่งเป็นสิ่งที่ แตกต่างกับเทคนิคต้นไม้ตัดสินใจ รุ่นที่ C4.5 ฉะนั้น วิธีที่นำเสนอหากเลือกนำไปใช้ได้อย่างเหมาะสมจะเกิดประโยชน์อย่างมาก

เดช ธรรมศิริ และ พยุง มีสีจ (2554) ศึกษาเรื่อง การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม วิธีการจำแนกข้อมูล ด้วยวิธีร่วมกันตัดสินใจแบบการโหวตเสียงข้างมากจาก เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยเลือกตัวแทนที่เหมาะสม ด้วยขั้นตอนวิธีเชิงพันธุกรรมเพื่อให้ความถูกต้องในการจำแนก ข้อมูลที่สูงขึ้น จากการทดสอบความถูกต้องในการจำแนก ข้อมูล Australian Credit, German Credit และ Bankruptcy Data พบว่า เทคนิคการร่วมกันตัดสินใจจากหลายโมเดล ที่ผ่านการเลือกแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม มีผลลัพธ์ที่ดีกว่าการใช้เทคนิคแบบโมเดลเดี่ยว โดยผล การวิจัยพบว่า

สำหรับข้อมูล Australian Credit หลังจากใช้วิธีขั้นตอนเชิงพันธุกรรมช่วยในการเลือกตัวจำแนก ข้อมูลพบว่าการเลือกใช้ตัวจำแนกที่ให้ประสิทธิภาพสูงสุด ได้แก่ การใช้ตัวจำแนกข้อมูล 9 โมเดลร่วมกัน ประกอบด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน 4 โมเดล เทคนิคต้นไม้ตัดสินใจ 3 โมเดล และเทคนิคโครงข่ายประสาทเทียม 2 โมเดล ให้ผลลัพธ์ความถูกต้องสูงสุดที่ 89.01% ข้อมูล German Credit หลังจากใช้วิธีขั้นตอนเชิงพันธุกรรม ช่วยในการเลือกตัวจำแนกข้อมูลพบว่าการเลือกใช้ตัวจำแนกที่ให้ประสิทธิภาพสูงสุดได้แก่ การใช้ตัวจำแนก ข้อมูล 18 โมเดล ซึ่งประกอบด้วยเทคนิคซัพพอร์ตเวกเตอร์ แมชชีน 7 โมเดล ร่วมกัน ประกอบด้วย เทคนิคต้นไม้ตัดสินใจ 6 โมเดล และเทคนิคโครงข่ายประสาทเทียม 5 โมเดล ให้ผลลัพธ์ความถูกต้องสูงสุดที่ 71.50% ข้อมูล Bankruptcy Data หลังจากใช้วิธีขั้นตอนเชิงพันธุกรรม ช่วยในการเลือกตัวจำแนกข้อมูลพบว่าการเลือกใช้ ตัวจำแนกที่ให้ประสิทธิภาพสูงสุดได้แก่ การใช้ตัวจำแนก ข้อมูล 10 โมเดลร่วมกัน ประกอบด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน 4 โมเดล เทคนิคต้นไม้ตัดสินใจ 4 โมเดล และเทคนิคโครงข่ายประสาทเทียม 2 โมเดล ให้ ผลลัพธ์ความถูกต้องสูงสุดที่ 71.50%

เดช ธรรมศิริ และ พยุง มีสัง (2554) ศึกษาเรื่อง การเรียนรู้แบบรวมกลุ่มด้วยโครงข่ายประสาทเทียมเอดาบัพ สำหรับการจำแนกข้อมูล การเพิ่มความแม่นยำสำหรับการจำแนกข้อมูลถือเป็นเรื่องสำคัญ ดังนั้น งานวิจัยครั้งนี้เป็นการนำเสนอวิธีการจำแนกข้อมูลด้วยวิธีร่วมกันตัดสินใจโดยใช้เทคนิคโครงข่ายประสาทเทียมร่วมกับการรวมกลุ่มตัดสินใจโดยใช้เทคนิคเอดาบัพเพื่อให้ความถูกต้องในการจำแนกข้อมูลที่สูงขึ้น ทดสอบบนฐานข้อมูลโรคเบาหวานได้แก่ Diabetes Data จาก UCI การทดสอบความถูกต้องในการจำแนกข้อมูล Diabetes Data พบว่า เทคนิคการร่วมกันตัดสินใจ จากหลายโมเดลที่ผ่านการรวมกลุ่มตัดสินใจด้วยเทคนิคเอดาบัพ มีผลลัพธ์ที่ดีกว่าการใช้เทคนิคแบบโมเดลเดี่ยว โดยผลการวิจัยพบว่าสำหรับข้อมูล Diabetes Data ให้ผลลัพธ์ความถูกต้องสูงสุดที่ 75.02% ในขณะที่โมเดลเดี่ยวให้ความแม่นยำน้อยกว่า

4. ขั้นตอนการดำเนินงาน

4.1 ข้อมูลสำหรับใช้ในการทดลอง

ผู้วิจัยได้ทำการรวบรวมข้อมูลผู้ที่มีอาการเป็นโรคเบาหวานและไม่เป็นโรคเบาหวาน โดยนำข้อมูลดังกล่าวมาจากเว็บไซต์ MLData จากนั้นนำข้อมูลมาทำความสะอาดโดยตัดข้อมูลที่มามีค่าขาดหาย (Missing Values) ออก

4.2 การดำเนินการทดลอง

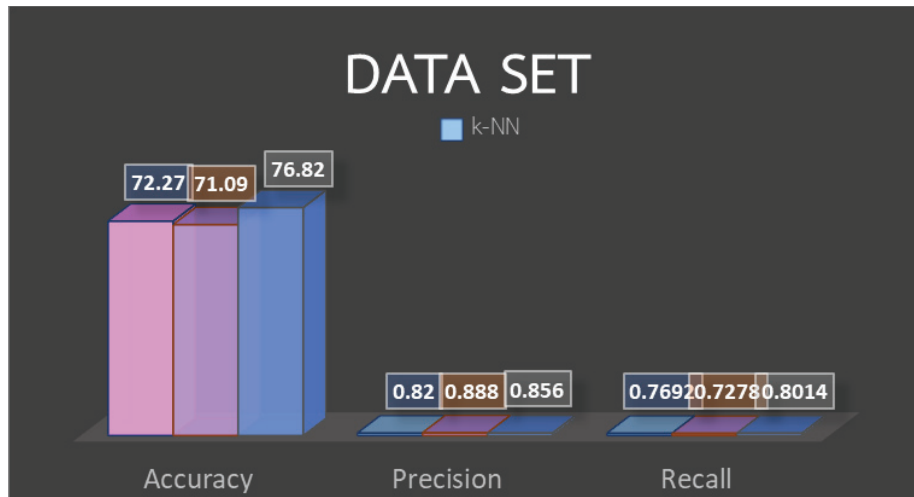
- 1) นำข้อมูลที่ได้จัดเตรียมไว้ในขั้นตอนที่ 4.1 มาทำการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการเกิดโรคเบาหวานด้วย เทคนิคการจำแนกข้อมูล จำนวน 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้าน ใกล้เคียงที่สุด
- 2) นำผลลัพธ์จากการจำแนกกลุ่มที่ได้จากการทดลองมาหาประสิทธิภาพของการจำแนกข้อมูลการเกิดโรคเบาหวาน
- 3) นำเทคนิคการจำแนกกลุ่มข้อมูลการเกิดโรคเบาหวานที่มีประสิทธิภาพมากที่สุดไปพัฒนาระบบการจำแนกกลุ่มข้อมูลโรคเบาหวานต่อไป

5. ผลการดำเนินงาน

ผลการทดลองเปรียบเทียบประสิทธิภาพในการการจำแนกกลุ่มข้อมูลโรคเบาหวาน จำนวน 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้เคียงที่สุด

ตารางที่ 1 ผลการทดลองเมื่อเปรียบเทียบทั้ง 3 เทคนิค

Data Set			
Method	Accuracy	Precision	Recall
k-NN	72.27	0.82	0.7692
Decision Tree	71.09	0.888	0.7278
Neural Net	76.82	0.856	0.8014



ภาพที่ 1 ผลการทดลองเมื่อเปรียบเทียบทั้ง 3 เทคนิค

ผลการทดลองการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลการเกิดโรคเบาหวาน 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้เคียงที่สุด ผลการวิจัยพบว่า เทคนิคที่ให้ผลลัพธ์ที่ดีที่สุดคือโครงข่ายประสาทเทียม โดยให้ค่าความถูกต้อง เท่ากับ 76.82 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.86 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.80 รองลงมาคือ เทคนิคเพื่อนบ้านใกล้เคียงที่สุด โดยให้ค่าความถูกต้อง เท่ากับ 72.27 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.82 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.77 และสุดท้าย ต้นไม้ตัดสินใจ โดยให้ค่าความถูกต้อง เท่ากับ 71.09 โดยให้ค่าความเที่ยงตรง เท่ากับ 0.89 โดยให้ค่าความแม่นยำที่สนใจ เท่ากับ 0.73

6.สรุปผลการดำเนินงานและข้อเสนอแนะ

6.1 สรุปผลการดำเนินงาน

งานวิจัยนี้ได้ทำการศึกษาถึงประสิทธิภาพของวิธีการจำแนกกลุ่มข้อมูลโรคเบาหวานด้วยเทคนิคการจำแนกกลุ่มข้อมูล 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้เคียงที่สุด ผลการวิจัยพบว่า เทคนิคทั้ง 3 เทคนิค เทคนิคที่ให้ค่าความแม่นยำมากที่สุด จะมีประสิทธิภาพดีที่สุด ดังนั้นเทคนิคที่ให้ค่าสามแม่นยำสูงที่สุด คือ เทคนิคโครงข่ายประสาทเทียม ให้ค่าความแม่นยำอยู่ที่ 76.82 %

6.2 ข้อเสนอแนะ

เนื่องจากเทคนิคการจำแนกกลุ่มข้อมูลมีหลากหลายเทคนิคแค่ 3 เทคนิค ที่เลือกมาเปรียบเทียบนั้นอาจจะไม่ใช่วิธีการที่ดีที่สุด ดังนั้น ควรทดลองใช้วิธีการอื่น ๆ ที่ยังไม่ได้ถูกนำมาทำการเปรียบเทียบทำการจำแนกกลุ่มข้อมูลได้หากนำทุกวิธีการมาทำการเปรียบเทียบกันทั้งหมด อาจจะได้วิธีการจำแนกกลุ่มข้อมูลโรคเบาหวานที่ดีที่สุด

7. เอกสารอ้างอิง

- หทัย แก้วภรณ์ และชรียา นนทกาญจน์. (2562).การจำแนกความสุกของกล้วยหอมโดนโครงข่ายประสาทเทียม ความน่าจะเป็นด้วยฟังก์ชันความน่าจะเป็น Generalized Extreme Value. งานประชุมวิชาการระดับชาติ ครั้งที่ 11 มหาวิทยาลัยราชภัฏนครปฐม. หน้า 248– 255.
- พรเทพ ด่านน้อย และสุวิมล มรรควิบูลย์ชัย. (2562).การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลการใช้บริการเครือข่ายอินเทอร์เน็ตของสถาบันอุดมศึกษา. งานประชุมวิชาการระดับชาติ ครั้งที่ 11 มหาวิทยาลัยราชภัฏนครปฐม. หน้า 303– 311.
- จักรกฤษณ์ หงส์เวียงจันทร์, นิติมา ลักษณ์นารักษ์ และไกรรุ่ง เสงพะพรหม.(2562). การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลโรคคอทิสติกด้วยเทคนิคเหมืองข้อมูล.งานประชุมวิชาการระดับชาติ ครั้งที่ 11 มหาวิทยาลัยราชภัฏนครปฐม.หน้า 321– 326.
- สุระสิทธิ์ ทรงม้า.(2558). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจรุ่นที่3 ที่ใช้เทคนิคการแบ่งข้อมูลที่แตกต่างกัน. SDU Research Journal. หน้า 29-45.
- เดช ธรรมศิริ และพยุง มีสัจ. (2554). การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม. วารสารวิชาการพระจอมเกล้าพระนครเหนือ. ปีที่ 21. ฉบับที่ 2, หน้า 293 – 303.
- เดช ธรรมศิริ และพยุง มีสัจ. (2554). การเรียนรู้แบบรวมกลุ่มด้วยโครงข่ายประสาทเทียมเอดาบูท สำหรับการจำแนกข้อมูล, วารสารเทคโนโลยีสารสนเทศ. ปีที่ 7. ฉบับที่ 14, หน้า 7-12.