**งานวิจัยที่เกี่ยวข้อง**

สุพาพร บรรดาศักดิ์, เบญญาภา ศรีสว่าง, และสุภาวดี ทองคำ (สุพาพร, เบญญาภา และสุภาวดี, 2559) ศึกษาถึงปัจจัยต่าง ๆ ที่ส่งผลต่อการเป็นโรคข้อเข่าเสื่อม โดยจากการศึกษาพบว่าอัลกอริทึม Naive byes มีประสิทธิภาพค่าความถูกต้อง เท่ากับ 92.1466 %, 99.7382 % อัลกอริทึม Sequential Minimal Optimization (SMO)มีประสิทธิภาพค่าความถูกต้องเท่ากับ 87,4346 %, 99.7382 % อัลกอริทึม Decision Tree (J48) มีประสิทธิภาพค่าความ ถูกต้องเท่ากับ 74.3455 %, 99.4764 % อัลกอริทึม Neural network มี ประสิทธิภาพค่าความถูกต้องเท่ากับ 89,0052 %, 98,9529 % จะเห็นได้ ว่าอัลกอริทึม Naive byes มีประสิทธิภาพค่าความถูกต้องสูงที่สุด และมีการพยากรณ์ที่เที่ยงตรงสามารถนำไปใช้วิเคราะห์ความเสียงคัดกรองผู้ ที่เสี่ยงต่อการเป็นโรคข้อเข่าเสื่อมได้เป็นอย่างดี

นพรัตน์ นนท์ศิริ, ราตรี มนัสศิลา, และกริช สมกันธา (นพรัตน์, ราตรี และกริช, 2564) ได้สร้างแบบจำลองการจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานจาก ผลการเปรียบเทียบพบว่า วิธีต้นไม้ตัดสินใจให้ค่าประสิทธิภาพสูงสุด โดยมีค่าความถูกต้อง 93.73% วิธีนาอีฟเบย์ค่าความถูกต้อง 88.92% วิธีความใกล้เคียงกันที่สุด และวิธีซัพพอร์ตเวกเตอร์แมชชีนค่าความถูกต้อง 86.97% และ 86.13% ตามลำดับ จะพบว่า วิธีต้นไม้ตัดสินใจมีประสิทธิภาพในการสร้าง แบบจำลองมากที่สุดเมื่อเทียบกับวิธีที่ใช้เปรียบเทียบร่วมกัน เนื่องจากเป็นวิธีที่ไม่มีการแจกแจงหรือไม่ใช้พารามิเตอร์ซึ่งไม่ได้ ขึ้นอยู่กับสมมุติฐานการแจกแจงความน่าจะเป็น อีกทั้งสามารถจัดการกับข้อมูลที่มีมิติสูงได้อย่างแม่นยำ เหมาะสมที่จะนำแบบ จำลองไปพัฒนาระบบจำแนกข้อมูลเพื่อวินิจฉัยความเสี่ยงการเป็นโรคเบาหวาน ทางการแพทย์ในการวินิจฉัยความเสี่ยงการเป็นโรคเบาหวานต่อไป เพื่อเป็นแนวทางในการสนับสนุนการตัดสินใจ

นงเยาว์ ในอรุณ (นงเยาว์, 2564) ได้สร้างแบบจำลองการทำนายความเสี่ยงโรคหัวใจและหลอดเลือดโดยใช้อัลกอริทึมเหมืองข้อมูล พบว่าแบบจำลองที่มีประสิทธิภาพการทำนายดีที่สุดคือ แบบจำลองโครงข่ายประสาทเทียมพร้อมการเลือกคุณสมบัติ มีค่าความถูกต้อง 99.29% และต่ำสุดคือ แบบจำลองต้นไม้ตัดสินใจ มีค่าความถูกต้อง 70.39%

สุรวัชร ศรีเปารยะ, สายชล สินสมบูรณ์ทอง (สุรวัชร, สายชล, 2560) ได้เปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม โดยเลือกใช้วิธีความใกล้เคียงกันมากที่สุด เพื่อวัดประสิทธิภาพการจำแนกกลุ่มโดยใช้ข้อมูลผู้ป่วยโรคไตเรื้อรังของโรงพยาบาลอพอลโล ประเทศอินเดีย จากการเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มผู้ป่วยโรคไตเรื้อรัง โดยเปรียบเทียบจากค่าความถูกต้องและคำความ คลาดเคลื่อนกำลังสองเฉลี่ย วิธีการจำแนกกลุ่มที่มีประสิทธิภาพการจำแนกดีที่สุดคือ วิธีต้นไม้ตัดสินใจ ซึ่งให้ค่าความ ถูกต้อง คือ 100 % และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.0059

เบญจภัค จงหมื่นไวย์ (เบญจภัค, 2558) ได้เปรียบเทียบปัจจัยข้อมูลผู้สูงอายุ (เพศ อายุ และโรคประจำตัว) ซึ่งโรคประจำตัว ผู้สูงอายุสามารถจัดกลุ่ม พบว่า แบบจำลอง Decision Tree ร้อยละค่าความถูกต้องของการทานายเท่ากับ 8.56% และ 91.43% เป็นค่าจริง แบบจำลอง Naïve Bayes ซึ่งพบว่า ร้อยละค่าความถูกต้องของการทานายเท่ากับ 10.27 % และ 89.72 % เป็นค่าจริง ดังนั้น การวัดประสิทธิภาพการทางานค่าทานายของผู้สูงอายุพบผู้สูงอายุที่ไม่มีโรคประจาตัวมากกว่าผู้สูงอายุที่มีโรคประจาตัวถึงร้อยละ 70.20 % และตัวแบบที่เหมาะสมสาหรับผู้สูงอายุในกลุ่มนี้คือ Decision Tree

กฤตกนก ศรีพิมพ์สอ และกิตติพล วิแสง (กฤตกนก และกิตติพล, 2566) พัฒนาแบบจำลองสำหรับการพยากรณ์ผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล ผลการวิจัยพบว่าเทคนิคแรนดอมฟอเรส (Random Forest) ให้ค่าความถูกต้องในการทำนายผลเป็นโรค เบาหวานมากที่สุดที่อยู่ที่ 99.75% มีค่าความแม่นยำ (Precision) ที่ 98.50% ค่าความครบถ้วน (Recall) ได้ 98.50% ค่าวัดประสิทธิภาพโดยรวม (F-Measure) ได้ 98.50% และค่าเส้นกราฟ (ROC) ได้ 95.20% สามารถนำผลที่ได้จากงานวิจัยนี้ไปประยุกต์ใช้ในการประกอบการรักษาผู้ป่วยโรคเบาหวานต่อไปในอนาคต

ณัฐพล แสนคำ, ทิพวัลย์ แสนคำ, และธนากร ปุรารัมย์ (ณัฐพล, ทิพวัลย์ และธนากร, 2560) ได้พัฒนาระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคโตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูลเปรียบเทียบประสิทธิภาพของอัลกอริทีมสำหรับทำนายโรคไตเรื้อรัง สรุปได้ว่า เทคนิค Random Forest ที่ใช้ชุดข้อมูลที่มีการเพิ่ม (Oversampling Data) มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงที่สุดจากทุก ตัวแบบที่ค่า 97.29% ค่าความแม่นตรง (Precision) ที่ 95,76% และค่าวัดประสิทธิภาพโดยรวม (F-Measure) เท่ากับ 97,44% และนำเทคนิคเหมืองข้อมูลนี้มาพัฒนาเป็นตัวแบบในการทำนายโรคไต ผลการประเมินประสิทธิภาพในการทำนายโรคไตของระบบพบว่าสามารถทำนายโรคไตของข้อมูลใหม่ ได้ถูกต้อง 95.71% ทั้งนี้ เทคนิคต่าง ๆ และตัวแบบที่ได้พัฒนาขึ้นจะสามารถนำไปต่อยอด เพื่อพัฒนาระบบสนับสนุนทางการแพทย์ที่มีประสิทธิภาพในอนาคต

**ทฤษฎีที่เกี่ยวข้อง**

**Linear Regression** หรือการวิเคราะห์การถดถอยเชิงเส้นเกิดจากการรวมกันของคำ 2 คำได้แก่ Linear ที่แปลว่าเส้นตรง และ Regression ที่แปลว่าการถดถอย ซึ่งถ้าจะให้เข้าใจง่ายก็คือการนำเอาข้อมูลหรือตัวแปรมาหาความสัมพันธ์กันโดยความสัมพันธ์ของข้อมูลจะออกมาในรูปแบบของการเรียงกันเป็นเส้นตรงหรือใกล้เคียง ความสัมพันธ์ที่ได้กล่าวมานั้นคือการหาค่าสหสัมพันธ์หรือ Correlation นั่นเองโดยค่า Correlation หรือ “r” ยิ่งค่ามีความเข้าใกล้ 1 หรือ -1 กำลังจะบ่งบอกว่าค่าความสัมพันธ์ของข้อมูลนั้นยิ่งมีความสัมพันธ์ในรูปแบบเส้นตรง (Linear) แต่ถ้าค่า “r’ เข้าใกล้ 0 จะบ่งบอกว่าความสัมพันธ์ของข้อมูลเหล่านั้นไม่ได้อยู่ในรูปแบบเส้นตรง (Non-linear) Linear Regression สามารถแบ่งออกได้เป็น 2 ประเภทโดยจะแบ่งตามจำนวนตัวแปรอิสระที่จะมเาป็นตัวกำหนดค่าตัวแปรตามหรือค่า Y ได้แก่ Simple Linear Regression เป็นสมการเชิงเส้นที่มีตัวแปรอิสระหรือตัวแปร X มากำหนดตัวแปรตามหรือค่า Y แค่เพียงตัวแปรเดียวโดยสมการจะมีรูปแบบเหมือนกับที่ยกตัวอย่างไปข้างต้น เนื่องจากมีตัวแปรอิสระแค่เพียงตัวเดียวที่ส่งผลกระทบต่อค่า Y ทำให้สมการที่ได้ไม่มีความซับซ้อนในการแปรความหมาย และ Multiple Linear Regression เป็นสมการเชิงเส้นที่มีตัวแปรอิสระหรือตัวแปร X มากกว่า 1 ตัวมาเป็นตัวกำหนดตัวแปรตามหรือค่า Y ความหมายคือมีหลายปัจจัยที่ส่งผลกระทบต่อค่าที่เราให้ความสนใจทำให้การแปลความหมายของสมการมีความซับซ้อนมากยิ่งขึ้น

**เทคนิคต้นไม้แบบสุ่ม (Random Forest)** เป็นอัลกอริทึมประเภทหนึ่งของอัลกอริทึมต้นไม้ตัดสินใจที่มีลักษณะแบบ Unpruned หรือ Regression Trees ซึ่งถูกสร้างขึ้นโดยการสุ่มเลือกตัวอย่างข้อมูล. หลักการของ Random Forest คือการสร้าง model จาก Decision Tree หลายๆ model (ตั้งแต่ 10 model ถึงมากกว่า 1000 model) โดยแต่ละ model จะได้รับ data set ที่ไม่เหมือนกันซึ่งเป็น subset ของ data set ทั้งหมด. เมื่อทำการ prediction, แต่ละ Decision Tree ทำการ prediction แต่ละตัวและคำนวณผล prediction ด้วยการ vote output ที่ถูกเลือกโดย Decision Tree มากที่สุด หรือหาค่าเฉลี่ยจาก output ของแต่ละ Decision Tree (ในกรณี regression)

**เทคนิคต้นไม้ตัดสินใจ (Decision Tree)** เป็นอัลกอริทึมที่ใช้วิธีการแตกแขนงจากโหนดราก (Root Node) เป็นโหนดภายใน (Branch Node) แตกออกไปตามเงื่อนไขหรือข้อมูล จนไปสู่โหนดใบ (Leaf Node) เป็นแบบจำลองที่มีการเชื่อมโยงระหว่างสิ่งที่สนใจกับผลสรุปที่อาจเกิดขึ้นจากค่าของเหตุการณ์ (Jones, 2008) โหนดภายในของต้นไม้ตัดสินใจจะประกอบเป็นคุณลักษณะของข้อมูล ซึ่งเมื่อสอดคล้องกับข้อมูลใด ๆ ก็จะใช้คุณลักษณะนั้นเป็นตัวตัดสินใจว่าข้อมูลจะไปทิศทางใด โหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าของคุณลักษณะในโหนดภายใน และสุดท้ายคือ โหนดใบ เป็นกลุ่มผลลัพธ์ในการจำแนกประเภทข้อมูล ผลลัพธ์ที่ได้สามารถแปลงเป็นกฎ (Rule) ได้ การสร้างจะเริ่มต้นที่โหนดรากเป็นอันดับแรกก่อนจะดำเนินการพิจารณา โหนดใบและกิ่งก้านที่แตกแขนงต่อไป โดยต้องคำนวณหาข้อมูลที่เหมาะสมที่จะเป็น โหนดราก ซึ่งพิจารณาจากค่า Information Gain ที่มากที่สุด ที่ได้จากการคำนวณค่า Entropy เพื่อให้การจำแนกและแยกแยะข้อมูลให้อยู่ในกลุ่มเดียวกันมากที่สุด หลังจากที่ได้โหนดรากแล้วก็จะสร้าง Decision Tree ในลำดับต่อไป จนกทั่งถึงและได้ Decision Tree ที่สมบูรณ์

เทคนิคการเรียนรู้เชิงลึก (Deep Learning) เป็นเทคนิคในกลุ่มโครงข่ายประสาทเทียม **(Artificial Neural Network: ANN)** ที่มีโครงสร้างขนาดใหญ่ประกอบด้วยนิวรอนและชั้นซ่อนจำนวนมาก เป็นอัลกอริทึมที่ถูกสร้างขึ้นเพื่อการเรียนรู้ของเครื่องจักร แต่ระดับ Hidden Layer ของการเรียนรู้เชิงลึกมีมากกว่า ANN ซึ่งแต่ละเลเยอร์จะเปรียบเสมือนประกอบด้วยเซลล์ประสาท (Neural) จำนวนมากที่มีหน้าที่ในการประมวลผล โดยเลเยอร์แรกสุดท้ายจะทำหน้าที่ในการรับข้อมูล (Input Layer) และส่งข้อมูลที่ประมวลผลเสร็จแล้วไปยังเลเยอร์สุดท้าย (Output Layer) การส่งข้อมูลแบบนี้มีข้อดีคือแต่ละเลเยอร์อาจทำให้มีค่าถ่วงน้ำหนัก (Weight) ค่าความเอนเอียงของข้อมูล (Bias) และวิธีการประมวลผลทางคณิตศาสตร์ (Activation Function) เป็นอิสระต่อกันถ้าป้อนข้อมูลเข้าไปให้กับโมเดลมากเท่าไหร่ แต่ละเลเยอร์ก็จะสามารถสกัดคุณลักษณะที่มีความซับซ้อนมากขึ้นทำให้ระบบสามารถตัดสินใจได้ใกล้เคียงกับมนุษย์มากยิ่งขึ้น (สมศักดิ์ ศรีสวการย์ และ สมัย ศรีสวย, 2563)

**เทคนิคนาอีฟเบย์ (Naïve Bayes)** เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes’ Theorem) (รุ่งโรจน์ บุญมา และ นิเวศ จิระวิชิตชัย, 2562) เพื่อหาสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่งๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่งๆ เพื่อหาสมมติฐานที่ดีที่สุด การเรียนรู้แบบเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน ในที่นี่คือคลาสเป้าหมายหรือผลลัพธ์การทำนายโดยการเรียนรู้แบบเบย์เป็นการเรียนรู้เพิ่มได้ เนื่องจากตัวอย่างใหม่ที่ได้มาถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็นทำให้มีการเรียนรู้ที่เปลี่ยนไป วิธีการนี้ตัวแบบจะถูกปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้โดยผนวกกับความรู้เดิมที่มี ซึ่งการทำนายค่าคลาสเป้าหมายของตัวอย่างใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐานจากทฤษฎีของเบย์ เราสามารถคำนวณความน่าจะเป็นของสมมติฐานต่างๆ