

## Prediction on risk for complications of hypertension in diabetes patient using data mining technique

**Rungruedee Butdachan, Todsanaï Chumwatana**

College of Information and Communication Technology, Rangsit University, Thailand

---

### Abstract

High blood pressure is especially dangerous for diabetics. Diabetes caused by the body to produce the hormone insulin is not enough for takesugar in the blood to different parts of the body. Resulting in the accumulationof sugar in the arteries, causing organ degeneration which is condition causesymptoms and complications of various organs followed. Researchers are interested to study the complications of high blood pressure that often occurs in diabetes patients. The disease together with high blood pressure has been a factor to kill many people. In this paper, researchers studied diabetes patientswho check-ups regularly and has got measuring blood pressure all the times.To prevent loss or as a guide of prevention, the risk of

complications ofhypertension has been analyzed using data mining techniques. In theexperiment, two classification techniques: decision tree and k-NN were used for analysis and predict a patient's risk to be disease with high blood pressure.The results showed that the accuracy of the decision tree has highly precisemore than k-NN classification technique for 2.08%.

**Keywords:** Diabetes, hypertension, complication diseases, prediction, classification, decision tree, K nearest neighbor clustering.

*Received 31 March 2017; Accepted 31 May 2017*

---

Correspondence: Rungrudee Butdachan, College of Information and Communication Technology, Rangsit University, Thailand Vipahvadee Center, TST Tower, 21 Floor, Vipahvadee Road, Jompol, Jatujak, Bangkok (Tel.: +66-62-651-4268; E-mail address: r.butdachan@gmail.com)

# การกํานายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของ ผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล

รุ่งฤดี บุษดาจันทร์, กณีย์ ชุ่มวัฒนะ

วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต

## บทคัดย่อ

โรคความดันโลหิตสูงเป็นอันตรายอย่างยิ่งสำหรับผู้ป่วยโรคเบาหวาน โรคเบาหวานมักเกิดจากร่างกายผลิตฮอร์โมนอินซูลินไม่เพียงพอในการนำน้ำตาลในเลือดไปเลี้ยงส่วนต่างๆ ของร่างกาย ส่งผลให้เกิดการคั่งของน้ำตาลในเส้นเลือดแดง ทำให้อวัยวะต่างๆ เสื่อม ซึ่งภาวะนี้ทำให้เกิดโรคและอาการแทรกซ้อนต่ออวัยวะต่างๆ ตามมา ผู้วิจัยจึงมีความสนใจที่จะศึกษาภาวะแทรกซ้อนโรคความดันโลหิตสูงที่มักเกิดขึ้นร่วมกันในผู้ป่วยโรคเบาหวานซึ่งโรคความดันโลหิตสูงก็มีผู้เสียชีวิตเป็นจำนวนมากเช่นกัน ผู้วิจัยจึงศึกษาข้อมูลผู้ป่วยโรคเบาหวานที่มีผลการตรวจสุขภาพอย่างสม่ำเสมอ โดยมีการตรวจวัดค่าความดันทุกครั้งมาทำการวิเคราะห์ ความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง เพื่อสามารถป้องกันการสูญเสีย หรือเป็นแนวทางในการป้องกัน และดูแลควบคุมได้ทันเวลา โดยใช้เทคนิคการทำเหมืองข้อมูล ด้วยเทคนิคการจำแนกประเภท

ข้อมูล (Classification) จากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และการจัดแบ่งคลาส (k-NN) เพื่อวิเคราะห์และทำนายผู้ป่วยถึงภาวะความเสี่ยงต่อการเกิดโรคความดันโลหิตสูง และทำการเปรียบเทียบผลการทำนาย โดยที่การทำนายด้วยแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) มีความแม่นยำ มากกว่าการจัดแบ่งคลาส(k-NN) ถึง 2.08 เปอร์เซ็นต์

**คำสำคัญ:** โรคเบาหวาน, โรคความดันโลหิตสูง, ภาวะแทรกซ้อน, การจำแนกประเภทข้อมูล, แผนภาพต้นไม้เพื่อการตัดสินใจ, การจัดแบ่งคลาส

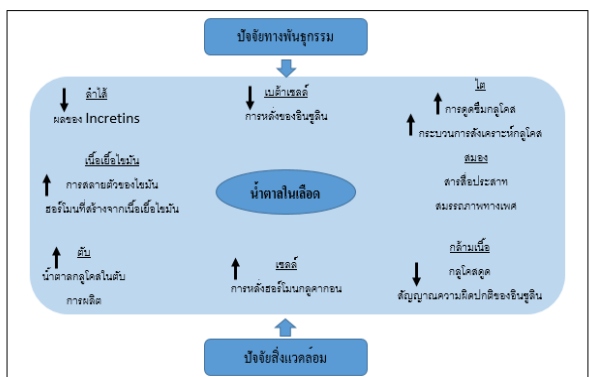
วันที่รับต้นฉบับ 31 มีนาคม 2560; วันที่ตอบรับ 31 พฤษภาคม 2560

## บทนำ

โรคเบาหวาน<sup>1</sup> ถือเป็นโรคเรื้อรังชนิดหนึ่งที่เป็นปัญหาสำคัญทางสาธารณสุขในปัจจุบัน เกิดจากความผิดปกติของร่างกายที่ผลิตฮอร์โมนอินซูลินไม่เพียงพอ ทำให้ระดับน้ำตาลในกระแสเลือดสูง โรคเบาหวานจะแบ่งออกเป็น 2 ชนิด ได้แก่ โรคเบาหวานชนิดที่ 1<sup>2</sup> เกิดจากการที่ตับอ่อนไม่สามารถสร้างฮอร์โมนอินซูลินให้เพียงพอ เนื่องจากเบตาเซลล์ (beta cells) ของตับอ่อนถูกทำลายด้วยระบบภูมิคุ้มกันของร่างกาย และโรคเบาหวานชนิดที่ 2<sup>3</sup> เป็นเบาหวานที่มีส่วนเกี่ยวกับพันธุกรรม และภาวะน้ำหนักตัวมาก ขาดการออกกำลังกาย และสูบบุหรี่ในปี พ.ศ.2558 จะพบว่าผู้ป่วยเป็นโรคเบาหวาน

ผู้นิพนธ์ประสานงาน: รุ่งฤดี บุษดาจันทร์, วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต ศูนย์ศึกษาวិภาวดี อาคารทีเอสที ทาวเวอร์ ชั้น 21 ถนนวิภาวดีรังสิต แขวงจอมพล เขตจตุจักร จังหวัดกรุงเทพฯ (โทร. 062-6514268; E-mail address: r.butdachan@gmail.com)

1 คน จากผู้ใหญ่ 11 คน ซึ่งทุกๆ 6 วินาทีจะมีคนเสียชีวิตจากโรคเบาหวาน<sup>4</sup>



ภาพที่ 1 ปัจจัยที่มีผลต่อการเกิดโรคเบาหวาน

จากข้อมูลข้างต้นจะเห็นได้ว่าเมื่อฮอร์โมนอินซูลินเกิดการ ทำงานผิดปกติ จะส่งผลให้เกิดการคั่งของน้ำตาลในเส้นเลือด ที่นำไปเลี้ยงส่วนต่างๆ ของร่างกาย จึงทำให้เกิดโรคและภาวะแทรกซ้อนตามมาโดยที่ผู้ป่วยโรคเบาหวานไม่รู้ตัว อาทิเช่น โรคความดันโลหิตสูงที่มักเกิดขึ้นร่วมกันในผู้ป่วยโรคเบาหวาน

โรคความดันโลหิตสูง<sup>5</sup> เป็นโรคที่ได้รับการขนานนามว่า “เพชรฆาตเงียบ” เนื่องจากมีความรุนแรงถึงแก่ชีวิตหากไม่ได้รับการวินิจฉัย หรือการรักษาตั้งแต่ต้น โดยทุกๆ คนต้องมีความดันโลหิต เพราะความดันโลหิต จะเป็นแรงผลักดันให้เลือดไปเลี้ยงอวัยวะต่างๆ ของร่างกาย ความดันโลหิตสูงจะทำให้เกิดหลอดเลือดแข็งและตีบ เมื่อหัวใจบีบตัว หัวใจจะบีบเลือดไปยังหลอดเลือดแดง ทำให้เกิดความดันโลหิตสูงซึ่งเกิดจากการบีบตัวของหัวใจ และแรงต้านทานของหลอดเลือด จากการวินิจฉัยทางการแพทย์<sup>6, 7</sup> พบว่าหัวใจคนเราเต้น 60-80 ครั้ง ความดันก็จะเพิ่มขึ้นขณะที่หัวใจบีบตัว และลดลงขณะที่หัวใจคลายตัว ความดันโลหิตของคนเราไม่เท่ากันตลอดเวลาขึ้นกับความเครียด การออกกำลังกาย การนอนหลับ ค่าปกติของคนเราคือ 120/80 มิลลิเมตรปรอท แต่ไม่ควรเกิน 140/90 หากสูงกว่านี้แสดงเป็นโรคความดันโลหิตสูง ในประเทศไทย มีผู้ป่วยเป็นโรคความดันโลหิตสูงเป็นจำนวนมาก และมีอัตราเพิ่มขึ้นถึง 5 เท่าในทุกปี ซึ่งจะมีผู้เสียชีวิตจากโรคความดันโลหิตสูงเป็นจำนวนมาก

จะเห็นได้ว่าโรคเบาหวาน และโรคความดันโลหิตสูงนั้นเป็นอันตรายอย่างยิ่ง หากผู้ป่วยโรคเบาหวานมีภาวะแทรกซ้อนของโรคความดันโลหิตสูงตามมาซึ่งถือว่าเป็นภาวะแทรกซ้อนโรคเรื้อรัง<sup>8, 9</sup> โดยจะนำมาซึ่งการสูญเสีย ทั้งชีวิตและทรัพย์สินเป็นอย่างมาก งานวิจัยนี้ผู้วิจัยจึงต้องการที่วิเคราะห์ความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวาน จากกลุ่มข้อมูลของประชากรผู้ป่วยโรคเบาหวาน ด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) จากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree)<sup>10</sup> และโมเดลการจัดแบ่งคลาส (k-NN)<sup>11</sup> เพื่อวิเคราะห์และทำนายผู้ป่วยโรคเบาหวานถึงภาวะความเสี่ยงต่อการเกิดโรคความดันโลหิตสูงตามมา และทำการเปรียบเทียบผลการทำนาย เพื่อดูความถูกต้องแม่นยำในการทำนายผลของแต่ละโมเดล และเป็นแนวทางในการป้องกันในผู้ป่วยโรคเบาหวานก่อนเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง

## งานวิจัยที่เกี่ยวข้อง

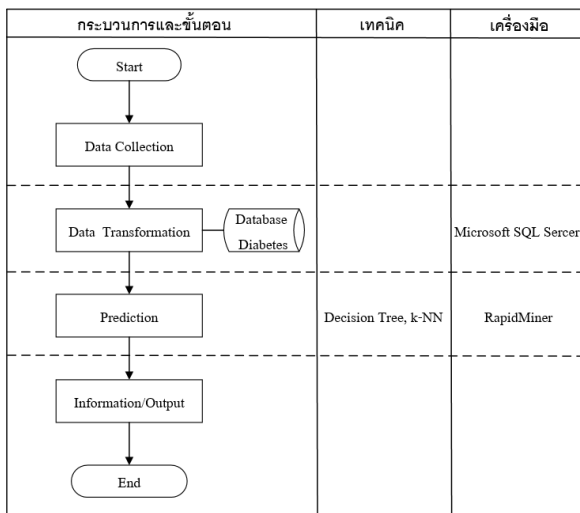
จากการศึกษาข้อมูลโรคเบาหวานและโรคความดันโลหิตสูงแล้ว พบว่ามีผู้ป่วยโรคเบาหวานเป็นจำนวนมากที่เกิดภาวะแทรกซ้อนของโรคความดันโลหิตสูง ผู้วิจัยจึงได้ทำการศึกษาและรวบรวมข้อมูลเกี่ยวกับโรคเบาหวาน และโรคความดันโลหิตสูง

ทั้งจากตำราโรคเบาหวานและโรคความดันโลหิตสูงและจากงานวิจัยด้านการแพทย์ และสิ่งพิมพ์ต่างๆ

จากบทนำข้างต้น ทำให้มีงานวิจัยด้านการแพทย์หลายงาน ที่ทำการวิจัยเกี่ยวกับโรคเบาหวานด้วยเทคนิคการทำเหมืองข้อมูลในรูปแบบต่างๆ ที่น่าสนใจสำหรับงานวิจัยทางการแพทย์และการรักษา เช่นงานวิจัยของ Olaiya Folorunsho<sup>12</sup> ได้ทำการสร้างแบบจำลองเพื่อทำนายผู้ป่วยที่น่าจะเป็นโรคเบาหวาน โดยทำการเปรียบเทียบ 2 โมเดลจากโมเดลโครงข่ายประสาทเทียม (Neural Network) กับแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) ซึ่งแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) มีอัตราความถูกต้องของการทำนายที่ดีกว่า และงานวิจัยของ Cheng-Ding Chang, Chien-Chih Wang, Bernard C. Jiang<sup>13</sup> มีการสร้างแบบจำลองในการคาดการณ์ของโรคความดันโลหิตสูงและโรคไขมันในเลือดที่พบบ่อย ในกรณีวิเคราะห์และคาดการณ์เกิดความดันโลหิตสูงและไขมันในเส้นเลือด ด้วยเทคนิคการทำเหมืองข้อมูลเพื่อดูปัจจัยเสี่ยงที่จะเกิดโรค โดยใช้หลักการของ Multivariate Adaptive Regression Splines (MARS) เพื่อสร้างรูปแบบในการทำนาย อัตราความถูกต้องของการจำแนก 93.07 % ซึ่งมีประสิทธิภาพในการทำนายโรคความดันโลหิตสูงและโรคไขมันในเลือด

## วิธีการดำเนินงานวิจัย

การดำเนินงานวิจัยเริ่มต้นจากการเก็บรวบรวมข้อมูลประวัติการรักษาผู้ป่วยที่เป็นโรคเบาหวาน ปี พ.ศ.2558 จนถึงปีปัจจุบัน พ.ศ.2559 มาทำการจัดเก็บข้อมูลเข้าสู่ฐานข้อมูล Microsoft SQL Server แล้วจึงนำข้อมูลที่ได้นำมาทำการแปลงข้อมูลจากข้อมูลที่เป็นอักขระ ให้เป็นตัวเลข แล้วนำข้อมูลไปทำการสร้างโมเดลขึ้นของทั้ง 2 โมเดล ทั้งโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และโมเดลการจัดแบ่งคลาส (k-NN) เพื่อทำการจัดกลุ่มข้อมูลในการทำนายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวาน ให้เป็นไปตามกระบวนการ และขั้นตอนการวิจัย ดังแสดงในภาพที่ 2



ภาพที่ 2 กระบวนการและขั้นตอนการดำเนินงานวิจัย

จากภาพที่ 2 แสดงกระบวนการเทคนิค และเครื่องมือที่ถูกนำมาใช้ในการทำวิจัยจะประกอบไปด้วยกระบวนการหลักๆ ในแต่ละกระบวนการ ดังนี้

**การเก็บรวบรวมข้อมูล**

การเก็บรวบรวมข้อมูล ในการทำวิจัยจำเป็นต้องทำการศึกษาความเป็นไปได้ของงานวิจัย ในข้อมูลผู้ป่วยโรคเบาหวาน ภาวะแทรกซ้อนต่างๆ ที่เกิดขึ้นตามมา และการวินิจฉัยโรคความดันโลหิตสูง จะพบในผู้ชายมากกว่าผู้หญิง โดยทางการแพทย์จะวินิจฉัยโรคความดันโลหิตสูงได้จากการวัดค่าความดันโลหิต ทั้งค่าความดันบน (BPS) และค่าความดันล่าง (BPD) จากการบีบตัวและคลายตัวของหัวใจ ในการทำนายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวานมีการเก็บรวบรวมข้อมูลผู้ป่วยโรคเบาหวานจากประชากรผู้ป่วยโรคเบาหวานจำนวน 100 คน โดยที่เป็นข้อมูลที่ระบุบอกถึง อายุ เพศ การตรวจคัดกรอง โดยจะมีค่าของระดับน้ำตาลเฉลี่ยสะสมในเลือด (HbA1c) ค่าความดันโลหิตบน (BPS) และค่าความดันโลหิตล่าง (BPD) โดยที่ข้อมูล CID (หมายเลขบัตรประชาชน) และชื่อผู้ป่วยนั้นจะทำการสมมติค่าขึ้นใหม่ เนื่องจากข้อมูลที่ได้ทำการเก็บรวบรวมนั้นเป็นข้อมูลผู้ป่วยจริง ซึ่งไม่สามารถเปิดเผยได้

CID	ชื่อผู้ป่วย	อายุ (ปี)	เพศ	ปีที่เริ่มเป็น	สิทธิการรักษา	Last HbA1c	Last	Last	ปรับปรุงครั้งสุดท้าย
3630100107175	นางศศิธร ออศิ	57	หญิง	2554	บุคคลในครอบครัวของ	7.2	146	83	12/4/2015
3630100105407	นางสาวอวย นิลดี	60	หญิง	2554	สิทธิราชการ/สิทธิ	10	132	74	7/21/2015
3670700923852	นายเสาว รัตติพรณ	79	ชาย	2554	ผู้มีอายุเกิน 60 ปีบริบูรณ์		120	70	7/21/2015
3630200190945	นางเดือน วงษ์ใจ	67	หญิง	2554	บุคคลผู้พิการ		103	46	11/11/2015
3630190016464	นางเรือง แก้วกลิ่น	67	ชาย	2554	ผู้มีอายุเกิน 60 ปีบริบูรณ์	6.7	158	88	12/4/2015

ภาพที่ 3 ตัวอย่างข้อมูลผู้ป่วยโรคเบาหวาน

การจัดเก็บข้อมูลเข้าสู่ฐานข้อมูล โดยการนำข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลมาทำการนำข้อมูลเข้าสู่ฐานข้อมูล SQL Server Management โดยทำการแบ่งการจัดเก็บออกเป็น 5 ชุดข้อมูล ดังนี้ ข้อมูลผู้ป่วย ค่าระดับน้ำตาล อายุ เพศ และค่าระดับความดัน

**การแปลงข้อมูล**

การแปลงข้อมูล โดยการนำข้อมูลที่ได้นำเข้าสู่ฐานข้อมูล แล้วทำการแปลงค่าข้อมูลที่อยู่ในรูปแบบของอักขระในรูปแบบภาษาไทย มาทำให้อยู่ในรูปแบบของภาษาอังกฤษ หรือตัวเลข เช่น ข้อมูลเพศ จะจัดเก็บด้วยข้อมูล หญิง ชาย ทำการแปลงข้อมูลเป็น เพศหญิง = 0 และเพศชาย = 1 โดยทำการแบ่งข้อมูลออกเป็น 2 ชุด จากข้อมูลผู้ป่วย 100 คน ชุดที่ 1 เป็นการแปลงข้อมูลสำหรับการสร้างโมเดลต้นแบบ จำนวน 80 คน และจัดกลุ่มผู้ป่วยด้วยค่าของความดันโลหิตทั้งค่าความดันโลหิตบน (BPS) และค่าความดันโลหิตล่าง (BPD) โดยทำการตั้งสมมุติฐานอ้างอิงจากการวินิจฉัยค่าความดันทางการแพทย์ เพื่อตรวจสอบค่าความดันโลหิต ว่าผู้ป่วยโรคเบาหวานมีความเสี่ยงมากน้อยเพียงไร จากการวัดค่าความดันโลหิตทั้ง 2 ค่า โดยจะสนใจ 3 กลุ่ม ดังนี้ 1) ผู้ป่วยที่มีค่าความดันโลหิตปกติ ค่าความดันโลหิตต่ำกว่า 120/80 มม.ปรอท 2) ผู้ป่วยที่มีค่าความดันโลหิตที่มีความเสี่ยง ค่าความดันโลหิตต่ำกว่า 140/90 มม.ปรอท 3) ผู้ป่วยที่มีค่าความดันโลหิตที่เป็นค่าความดันโลหิตมากกว่า 140/90 มม.ปรอท ขึ้นไป และข้อมูลชุดที่ 2 เป็นข้อมูลผู้ป่วย 20 คน จาก 100 คน ทำการแปลงข้อมูลสำหรับทดสอบการทำนายของทั้ง 2 โมเดล โดยที่ไม่มีการตั้งสมมุติฐานในการตรวจสอบค่าความดันโลหิต

Citizen_ID	Gender	Age	LastBPS_Value	LastBPD_Value
CI001	0	25	125	86
CI002	0	25	132	76
CI003	0	52	144	84
CI004	0	47	138	64
CI005	0	63	165	92
CI006	0	58	139	72
CI007	1	60	155	96
CI008	0	73	135	69
CI009	1	35	124	76
CI010	1	50	129	95
CI011	0	58	134	79
CI012	1	64	208	98
CI013	0	46	124	80
CI014	0	51	112	70

ภาพที่ 4 ตัวอย่างข้อมูลผู้ป่วยโรคเบาหวานที่ได้ทำการแปลงข้อมูลการสร้างโมเดลและการทำนาย

### การสร้างโมเดลและการทำนาย

จากข้อมูลที่ได้จากการแปลงค่าข้อมูลเรียบร้อยแล้วจากข้อมูล ผู้ป่วยโรคเบาหวานจำนวน 100 คน แบ่งเป็น 2 ชุด ดังนี้ 1) ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 80 คน ใช้เป็นข้อมูลในการสร้างโมเดล 2) ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 20 คน ใช้เป็นข้อมูลในการทดสอบโมเดล จากนั้นจึงนำเข้าข้อมูลไปที่เครื่องมือ RapidMiner โดยมีข้อมูลผู้ป่วยจำนวน 80 คน โดยทำการสร้างโมเดลต้นแบบในการทำนายของทั้งโมเดล แผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และโมเดลการจัดแบ่งคลาส (k-NN) ดังภาพที่ 5

Row No.	Citizen_ID	Risk	Gender	Age	LastBPS_Value	LastBPD_Value
1	CI001	Yellow	0	25	125	86
2	CI002	Yellow	0	25	132	76
3	CI003	Red	0	52	144	84
4	CI004	Yellow	0	47	138	64
5	CI005	Red	0	63	165	92

ภาพที่ 5 ข้อมูลในการทดสอบโมเดล

จากข้อมูลผู้ป่วยโรคเบาหวานจำนวน 80 คนที่ใช้ในการสร้างโมเดลต้นแบบนั้นจะมีผู้ป่วยโรคเบาหวานจำนวน 13 คนที่ไม่มีความเสี่ยง (pred. Green) มีผู้ป่วยโรคเบาหวานจำนวน 46 คนที่มีความเสี่ยง (pred. Yellow) และมีผู้ป่วยโรคเบาหวานจำนวน 21 คนที่มีภาวะแทรกซ้อนโรคความดันโลหิตสูง (pred. Red) เมื่อได้โมเดลต้นแบบเรียบร้อยแล้ว จากนั้นนำข้อมูลที่ได้จากการแปลงข้อมูลของผู้ป่วยที่ใช้ในการทดสอบโมเดลจำนวน 20 คนมาทำการทำนายความเสี่ยงที่จะเกิดภาวะแทรกซ้อนของโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวาน และทำการเปรียบเทียบผลความถูกต้องและแม่นยำที่เกิดขึ้นของทั้ง 2 โมเดล ซึ่งมีโมเดลในการทำนายดังนี้

- 1) โมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Model)
- 2) โมเดลการจัดแบ่งคลาส (k-NN Model)

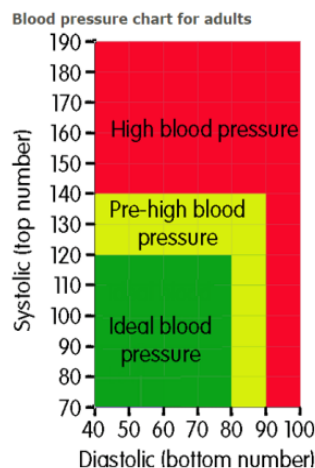
### การแสดงผลและเปรียบเทียบ

จากการสร้างโมเดลทั้ง 2 โมเดลเรียบร้อยแล้ว จากนั้นก็จะทำการเปรียบเทียบผลจากการทำนายที่เกิดขึ้นของทั้ง 2 โมเดล ว่าโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Model) หรือโมเดลการจัดแบ่งคลาส (k-NN Model) มีการทำนายค่าที่ออกมาถูกต้อง และแม่นยำที่สุด

### ผลการวิจัย

จากการเก็บรวบรวมข้อมูลผู้ป่วยโรคเบาหวาน โดยทำการแบ่งข้อมูลผู้ป่วยโรคเบาหวานออกเป็น 2 ชุดให้อยู่ในรูปแบบของไฟล์ Excel ดังนี้ 1) ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 80 คน ใช้เป็นข้อมูลในการสร้างโมเดล 2) ข้อมูลผู้ป่วยโรคเบาหวานจำนวน 20 คน ที่มีค่าการตรวจวัดของค่าความดันโลหิต เพื่อใช้ในการทดสอบแบบจำลองต้นแบบที่สร้างขึ้นแล้วจึงทำการทดสอบ ซึ่งจะประกอบไปด้วย 4 กระบวนการ ดังนี้ 1) การเก็บรวบรวมข้อมูล (Data Preprocessing) 2) การแปลงข้อมูล (Data Transformation) 3) การสร้างโมเดลและการทำนาย (Prediction) โดยทำการสร้างโมเดลต้นแบบ 2 โมเดล คือ 1. แผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และ 2. การจัดแบ่งคลาส(k-NN) 4) การแสดงผลและการเปรียบเทียบ (Information/Output) จากกระบวนการข้างต้นนี้ เพื่อทำนายผู้ป่วยโรคเบาหวานที่มีความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง

ผลที่ได้จากการทดลองของทั้ง 2 โมเดลที่สร้างขึ้น ซึ่งมีค่าความดันโลหิตนั้นจะกำหนดได้ดังภาพที่ 6



ภาพที่ 6 กำหนดช่วงระดับค่าความดันโลหิต

เมื่อทำการทดสอบโมเดลทั้ง 2 โมเดล จากข้อมูลผู้ป่วยโรคเบาหวานที่ใช้ในการทดสอบจำนวน 20 คนนั้น จากการทดสอบข้างต้นพบว่า ผลการทดสอบของทั้งสองโมเดลมีการทำนายความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวานที่ใกล้เคียงกัน แต่กลับให้เปอร์เซ็นต์ในการทำนายที่ต่างกัน

Citizen_ID	LastBPS Value	LastBPD Value	prediction (Risk)	confidence (Yellow)	confidence (Red)	confidence (Green)
CI081	125	86	Yellow	1	0	0
CI082	132	76	Yellow	1	0	0
CI083	144	84	Red	0	1	0
CI084	138	64	Yellow	1	0	0
CI085	165	92	Red	0	1	0
CI086	139	72	Yellow	1	0	0
CI087	155	96	Red	0	1	0
CI088	135	69	Yellow	1	0	0
CI089	124	76	Yellow	1	0	0
CI090	110	63	Green	0	0	1
CI091	134	79	Yellow	1	0	0
CI092	208	98	Red	0	1	0
CI093	124	80	Yellow	1	0	0
CI094	112	70	Green	0	0	1
CI095	137	75	Yellow	1	0	0
CI096	134	61	Yellow	1	0	0
CI097	139	71	Yellow	1	0	0
CI098	129	79	Yellow	1	0	0
CI099	202	93	Red	0	1	0
CI100	112	70	Green	0	0	1

ภาพที่ 7 ตารางแสดงผลการทำนายความเสี่ยงจากโมเดล ทั้ง 2 โมเดล

โดยที่โมเดลที่ 1 แผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) จากข้อมูลชุดทดสอบของผู้ป่วยโรคเบาหวานจำนวน 20 คน พบว่าการทำนายความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวานนั้น ได้มีการทำนายค่าความถูกต้อง (accuracy) เมื่อเทียบกับชุดข้อมูลของโมเดลต้นแบบ โดยเฉลี่ยร้อยละ 98.33 ดังภาพที่ 8

accuracy: 98.33% +/- 5.00% (mikro: 98.33%)				
	true Yellow	true Red	true Green	class precision
pred. Yellow	12	1	0	90.91%
pred. Red	0	4	0	100.00%
pred. Green	0	0	3	100.00%
class recall	100.00%	97.06%	100.00%	

ภาพที่ 8 ตารางแสดงเปอร์เซ็นต์ในการทำนายความเสี่ยงจากโมเดลต้นไม้เพื่อการตัดสินใจ(Decision Tree)

จากผลการทดลองในภาพที่ 8 พบว่าการทำนายความเสี่ยงต่อการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง เมื่อนำข้อมูลสำหรับใช้ในการทดสอบโมเดลจำนวน 20 คน มาทำการทดสอบกับโมเดลต้นแบบที่สร้างขึ้น ผลของการทดสอบจะเห็นได้ว่าผู้ป่วยโรคเบาหวานที่มีความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง (pred. Yellow) มีถึง 12 คน จากจำนวน 20 คน และมีจำนวน 1 คนที่อาจจะมีความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือมีภาวะแทรกซ้อน ทำให้ค่าความแม่นยำ (Precision) เป็นร้อยละ 90.91 และให้

ค่าความระลึกละ 100 เปอร์เซ็น ผู้ป่วยโรคเบาหวานที่มีภาวะแทรกซ้อนของโรคความดันโลหิตสูง (pred. Red) มีจำนวน 4 คน มีค่าความแม่นยำ (Precision) เป็น 100 เปอร์เซ็น และให้ค่าความระลึกละ 97.06 เนื่องจาก มีจำนวน 1 คนมีค่าความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือมีภาวะแทรกซ้อน และผู้ป่วยโรคเบาหวานที่มีค่าความดันปกติ (pred. Green) มีจำนวน 3 คน มีค่าความแม่นยำ (Precision) เป็น 100 เปอร์เซ็น และให้ค่าความระลึกละ 100 เปอร์เซ็น

โมเดลที่ 2 การจัดแบ่งคลาส (k-NN) จากข้อมูลชุดทดสอบของผู้ป่วยโรคเบาหวานจำนวน 20 คน พบว่าการทำนายความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวานนั้น ได้มีการทำนายค่าความถูกต้อง (accuracy) เมื่อเทียบกับชุดข้อมูลของโมเดลต้นแบบโดยเฉลี่ยเป็นร้อยละ 96.25 ดังภาพที่ 9

accuracy: 96.25% +/- 5.73% (mikro: 96.25%)				
	true Yellow	true Red	true Green	class precision
pred. Yellow	10	1	1	95.74%
pred. Red	0	4	0	100.00%
pred. Green	1	0	3	92.31%
class recall	97.83%	95.24%	92.31%	

ภาพที่ 9 ตารางแสดงเปอร์เซ็นต์ในการทำนายความเสี่ยงจากโมเดลการจัดแบ่งคลาส(k-NN)

จากผลการทดลองในภาพที่ 9 พบว่าการทำนายความเสี่ยงต่อการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง เมื่อนำข้อมูลสำหรับใช้ในการทดสอบโมเดลจำนวน 20 คน มาทำการทดสอบกับโมเดลต้นแบบที่สร้างขึ้น ผลของการทดสอบจะเห็นได้ว่าผู้ป่วยโรคเบาหวานที่มีความเสี่ยงที่จะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูง(pred. Yellow) มีถึง 10 คน จากจำนวน 20 คน และมีจำนวน 1 คนที่อาจจะมีความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือมีภาวะแทรกซ้อน และอีก 1 คนอาจจะมีความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือปกติ ทำให้ค่าความแม่นยำ (Precision) เป็นร้อยละ 95.74 และให้ค่าความระลึกละ 97.83 ผู้ป่วยโรคเบาหวานที่มีภาวะแทรกซ้อนของโรคความดันโลหิตสูง (pred. Red) มีจำนวน 4 คน มีค่าความแม่นยำ (Precision) เป็น 100 เปอร์เซ็น และให้ค่าความระลึกละ 95.24 เนื่องจากมีจำนวน 1 คนมีค่าความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือมีภาวะแทรกซ้อน และผู้ป่วยโรคเบาหวานที่มีค่าความดันปกติ (pred. Green) มีจำนวน 3 คน มีค่าความแม่นยำ (Precision) เป็นร้อยละ 92.31

เนื่องจากมี 1 คนอาจจะมีค่าความดันโลหิตที่อยู่ในช่วงคาบเกี่ยวระหว่างมีความเสี่ยงหรือปกติ และให้ค่าความระลึกละ 92.31

จากการวิจัยข้างต้น ด้วยการทดสอบการทำนายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงของผู้ป่วยโรคเบาหวานทั้ง 2 โมเดลนั้น จะเห็นได้ว่าผลที่ได้จากการทำนาย โมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) มีค่าการทำนายที่มีความแม่นยำ และถูกต้องตามที่คาดการณ์ไว้มากกว่าโมเดลการแบ่งคลาส (k-NN) เนื่องจากการทดสอบการทำนายภาวะแทรกซ้อนโรคความดันโลหิตสูงจากข้อมูลผู้ทดสอบจำนวน 20 คนจากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ(Decision Tree) ได้ให้ผลการทำนายความถูกต้องและแม่นยำ ถึง 98.33 เปอร์เซ็นต์ ส่วนการทดสอบการทำนายภาวะแทรกซ้อนโรคความดันโลหิตสูงจากข้อมูลผู้ทดสอบจำนวน 20 คนจากโมเดลการแบ่งคลาส (k-NN) ได้ให้ผลการทำนายความถูกต้องและแม่นยำ ถึงเพียงร้อยละ 96.25 จะเห็นได้ว่าการทำนายจากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) ให้ค่าความแม่นยำ กว่าโมเดลการแบ่งคลาส (k-NN) ถึงร้อยละ 2.08 ดังภาพที่ 10

ผลการทำนายของโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ(Decision Tree)				
accuracy: 98.33% +/- 5.00% (mikro: 98.33%)				
	true Yellow	true Red	true Green	class precision
pred. Yellow	12	1	0	90.91%
pred. Red	0	4	0	100.00%
pred. Green	0	0	3	100.00%
class recall	100.00%	97.06%	100.00%	

ผลการทำนายของโมเดลการแบ่งคลาส(k-NN)				
accuracy: 96.25% +/- 5.73% (mikro: 96.25%)				
	true Yellow	true Red	true Green	class precision
pred. Yellow	10	1	1	95.74%
pred. Red	0	4	0	100.00%
pred. Green	1	0	3	92.31%
class recall	97.83%	95.24%	92.31%	

ภาพที่ 10 ตารางผลการทำนายของโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และโมเดลการจำแนกประเภท (k-NN)

### สรุป

การเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงในผู้ป่วยโรคเบาหวาน เป็นอันตรายอย่างยิ่ง เนื่องจากโรคเบาหวานมักเกิดจากร่างกายผลิตฮอร์โมนอินซูลินไม่เพียงพอในการนำ

น้ำตาลในเลือดไปเลี้ยงส่วนต่างๆ ของร่างกาย ส่งผลให้เกิดการคั่งของน้ำตาลในเส้นเลือดแดง ทำให้อวัยวะต่างๆ เสื่อมซึ่งภาวะนี้ทำให้เกิดโรคและอาการแทรกซ้อนต่ออวัยวะต่างๆ อันเป็นผลให้เกิดการสูญเสียตามมา ดังนั้นเราจึงจะต้องป้องกันตั้งแต่ต้น ก่อนที่ผู้ป่วยโรคเบาหวานจะเกิดภาวะแทรกซ้อนโรคความดันโลหิตสูงขึ้น จากการวิเคราะห์ในการหาแบบจำลองที่เหมาะสมกับข้อมูลตัวอย่างในการวิจัยได้เลือกใช้วิธีการวิเคราะห์ข้อมูลด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) เพื่อสร้างสมมติฐานในการทำนายภาวะแทรกซ้อนโรคความดันโลหิตสูงที่จะเกิดขึ้นกับผู้ป่วยโรคเบาหวาน ด้วยโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และการจัดแบ่งคลาส (k-NN) จากข้อมูลผู้ป่วยโรคเบาหวานที่มีการตรวจคัดกรองอย่างสม่ำเสมอโดยที่มีการเก็บข้อมูลการตรวจวัดค่าความดันมาทำการวิเคราะห์เพื่อทำนายผลที่จะเกิดขึ้นตามมาในอนาคต ว่าผู้ป่วยโรคเบาหวานจะมีภาวะแทรกซ้อนโรคความดันโลหิตสูงหรือไม่ จากการวิเคราะห์ในผลของการทำนายของทั้ง 2 โมเดลจะพบว่าโมเดลในการทำนายด้วยแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) ในการคาดการณ์ผลการวิเคราะห์จะมีประสิทธิภาพดีกว่า ด้วยเปอร์เซ็นต์การทำนายร้อยละ 98.33 เมื่อเทียบกับโมเดลการจัดแบ่งคลาส (k-NN) ที่มีเปอร์เซ็นต์ในการทำนาย 96.25 เปอร์เซ็นต์ ซึ่งโมเดลในการทำนายแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) มีค่าการทำนายมากกว่าร้อยละ 2.08 อย่างไรก็ตามผลการทดลองจะต้องถูกตรวจสอบความถูกต้องในงานวิจัยในอนาคตเพื่อความน่าเชื่อถือ อีกทั้ง การวิเคราะห์การหาแบบจำลองที่เหมาะสมในการทำนายเพื่อให้ผลการทำนายมีประสิทธิภาพมากยิ่งขึ้น ดังนั้นงานวิจัยในอนาคต จึงจำเป็นต้องศึกษาหาโมเดลอื่น เช่น กฎความสัมพันธ์ (Association Rules) ของผู้ป่วยโรคเบาหวานว่ามีปัจจัยใดบ้างที่เป็นสาเหตุทำให้ผู้ป่วยโรคเบาหวานเกิดภาวะแทรกซ้อนของโรคความดันโลหิตสูงขึ้นได้ เพื่อทำการเปรียบเทียบให้งานวิจัยมีความถูกต้องมากยิ่งขึ้น

### เอกสารอ้างอิง

1. สาริส สุนทรโยธิน, ปฎิณัฐ บุรณะทรัพย์ขจร. Diabetes Mellitus ตำราโรคเบาหวาน. โรคเบาหวาน. พิมพ์ครั้งที่ 2 ed. คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย; 2555.
2. Ziegler A-G, Nepom GT. Prediction and pathogenesis in type 1 diabetes. Immunity. 2010;32(4):468-78.
3. Alonso-Magdalena, Paloma, Quesada I, Nadal A. Endocrine disruptors in the etiology of type 2 diabetes mellitus. Nature Reviews Endocrinology. 2011;7(6):346-53.

4. Cavan D, Fernandes JdR, Makaroff L, Ogurtsova K, Webber S. IDF Diabetes Atlas Seventh Edition 2015. 7th ed ed2015.
5. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo Jr JL, et al. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the JNC 7 report. *Jama.* 2003;289(19):2560-71.
6. ศุภชัย ไชยธีระพันธ์. แนวทางรักษาโรคความดันโลหิตสูงในเวชปฏิบัติทั่วไป; 2558.
7. ไวยวรรณ ธนะมัย, สิทธิชัย อาชายินดี, สุวีพร คนละเอียด. ความรู้เบื้องต้นเกี่ยวกับภาวะความดันโลหิตสูง. คู่มือการให้ความรู้เพื่อจัดการภาวะความดันโลหิตสูงด้วยตนเอง. พิมพ์ครั้งที่ 1: สถาบันวิจัยและประเมินเทคโนโลยีทางการแพทย์; 2555.
8. Control D, Group CTR. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl j Med.* 1993;(329):977-86.
9. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *The Lancet.* 1998;352(9131): 837-53.
10. Rokach L, Maimom O. *Data Mining With DecisionTree Theory and Application.* 2<sup>nd</sup> ed2014.
11. Aggarwal C. *k-NN. Data Classification Algorithms and Application*2015.
12. Folorunsho O. Comparative study of different data mining techniques performance in knowledge discovery from medical database. *International Journal.* 2013;3(3).
13. Chang CD, Wang CC, Jiang BC. Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. Taiwan: Ming Chi University of Technology and Yuan Ze University; 2011.