# การทำนายความเสี่ยงในการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุ

# Predicting the Risk for Developing of Essential Hypertension

*ธนัญชัย เพ็งพรหม (Tananchai Pengpom)[1] สรพรรฐ ปัทมามาลย์(Sorapat Patmaman)[2]*
*อิทธิพล ดวงแก้ว(Iddhipol Duangkaew)[3] วาทิตย์ คำพรมมา (Watit Kumpomma)[4]*

*[1234]ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น*
*tanatchai_pp@kkumail.com, sorapat_fresh@kkumail.com, ittipoldk@kkumail.com, watitkp@kkumail.com*

**บทคัดย่อ**

*โรคความดันโลหิตสูง เป็นปัจจัยเสี่ยงทำให้เกิดโรคหัวใจ โรคไต โรคหลอดเลือดหัวใจตีบ โรคอัมพาต โรคหัวใจเป็นโรคที่มีอัตราตายสูงดังนั้นการป้องกันความดันโลหิตสูงสามารถป้องกันอัตราการตายจากโรคหัวใจและโรคอัมพาต*

*เพื่อสามารถป้องกันการสูญเสียหรือเป็นแนวทางในการป้องกันและดูแลควบคุมได้ทันเวลา โดยใช้เทคนิคการทำเหมืองข้อมูลด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) จากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) และ K-nearest neighbor*

*เพื่อวิเคราะห์และทำนายนายผู้ป่วยถึงภาวะความเสี่ยงต่อการเกิดโรคความดันโลหิตสูงและทำการทดสอบโมเดลด้วยเทคนิค 10 Fold Cross Validation*

*พบว่าเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องที่ดีกว่าคือร้อยละ 96.45*

**คำสำคัญ:** โรคความดันโลหิตสูง, การจำแนกประเภทข้อมูล, แผนภาพต้นไม้เพื่อการตัดสินใจ, เพื่อนบ้านที่ใกล้ที่สุด

**Abstract**

*High blood pressure as a risk factor for heart disease, kidney disease, coronary heart disease, stroke, heart disease is a high mortality disease, so prevention of high blood pressure can prevent death from heart disease and paralysis. To prevent loss or preventive measures. And control in time. By using data mining techniques. With Data Classification Techniques Classification from the Decision Tree And K-nearest neighbor to analyze and predict the patient's risk for hypertension and compare the results of the prediction* It was found that the tree tricks were better at 96.45 percent.

**Keyword:** Data Mining, Prediction, Decision Tree, *K-nearest neighbor*

## 1. บทนำ

ในประเทศไทยโรคความดันโลหิตสูงเป็นโรค ที่ติดอันดับ 1 ใน 5 ของโรคชนิดไม่ติดต่อ ส่วนใหญ่ผู้ที่เป็นโรคความดันโลหิตสูงมักจะเป็นชนิดไม่ทราบสาเหตุ (Essential Hypertension) พบมากถึงร้อยละ 95 ในปี 2555 พบว่า อัตราผู้ป่วยในต่อประชากรแสนคนด้วยโรคความดันโรคหิตสูงมีจำนวน 1,009,385 ราย และอัตราการเสียชีวิตของโรคความดันโลหิตสูงในปี 2555 จำนวน 3,684 ราย ในปี 2556

การทำเหมืองข้อมูล (Data mining) คือ กระบวนการสกัดข้อมูลที่มีประโยชน์ เพื่อให้ได้สารสนเทศที่มองเห็นได้ยากเป็นสารสนเทศที่มีเหตุผลและสามารถนำไปใช้ในการช่วยการตัดสินใจได้ ซึ่งการทำเหมืองข้อมูลมีเทคนิคมากมายที่ให้เลือกใช้ แต่ควรเลือกใช้ให้เหมาะสมกับประเภทของงาน ได้แก่ การค้นหาความสัมพันธ์ การจัดกลุ่มข้อมูล การทำนายและการจำแนกประเภทข้อมูล

## 2. วรรณกรรมที่เกี่ยวข้อง

**2.1 โรคความดันโลหิตสูง** (Hypertension)

ความดันโลหิต เป็นแรงดันภายในหลอดเลือดแดง เกิดจากหัวใจบีบและคลายตัวเป็นจังหวะ เพื่อให้เลือดไหลไปเลี้ยงอวัยวะส่วนต่าง ๆ ของร่างกาย โดยค่าความดันโลหิต มี 2 ค่า คือ ค่าบน (ซีสโตลิก) กับค่าล่าง (ไดแอสโตลิก) เช่น 120/80 มม.ปรอท ค่าความดันโลหิต จะเปลี่ยนแปลงตามภาวะร่างกาย อายุ อารมณ์และกิจกรรมต่าง ๆ ความดันโลหิตสูง คือ ภาวะที่ตรวจพบความดันโลหิตอยู่ในระดับตั้งแต่ 140 /90 มม.ปรอท

| ระดับความรุนแรง | ความดันโลหิต | ความดันโลหิ |
|---|---|---|

| | ตัวบน | ต<br>ตัวล่าง |
|---|---|---|
| ความดันโลหิตปกติ | น้อยกว่า 120 และ | น้อยกว่า 80 |
| ระยะก่อนความดันโลหิต | 120-139 | 80 – 89 |
| ความดันโลหิตสูงระยะที่ 1 | และ/หรือ 140- | 90 – 99 |
| ความดันโลหิตสูงระยะที่ 2 | 159 และ/หรือ ≥ | ≥ 100 |
| | 160 และ/หรือ | |

ตารางที่ 1 :
แบ่งระดับความรุนแรงของความดันโลหิตสูงความดันโลหิตสูงมี 2 ชนิด ได้แก่

1. ชนิดไม่พบสาเหตุ พบได้ร้อยละ 50 ซึ่งเป็นความดันโลหิต
ที่ไม่ทราบสาเหตุที่แน่ชัดโดยทั่วไปมักไม่มีอาการผิดปกติ

2. ชนิดที่มีสาเหตุ เป็นความดันโลหิตสูงที่ทราบสาเหตุ
พบได้ ร้อยละ 50 โดยสาเหตุที่พบได้บ่อยที่สุด คือ
ภาวะหยุดหายใจ ขณะนอนหลับ (Obstructive sleep apnea)
พบได้ ประมาณร้อยละ 50

อันตรายและภาวะแทรกซ้อน
หากไม่ควบคุมความดันโลหิตและมีระดับความดันโลหิตสูง
อย่างต่อเนื่อง จะทำให้เกิด
ผนังหลอดเลือดหนาตัวจนหลอดเลือดตีบ
ทำให้หลอดเลือดปริแตก โดยเฉพาะผู้สูงอายุ
เกิดภาวะโรคหัวใจวายเฉียบพลัน ไตวาย หลอดเลือดสมองแตก
โรคอัมพฤกษ์อัมพาต

**2.2 Data Mining**

การทำเหมืองข้อมูล
เป็นกระบวนการสะกัดข้อมูลที่มีประโยชน์
ด้วยขั้นตอนวิธีที่เหมาะสม
เพื่อให้ได้องค์ความรู้ที่สามารถนำไปใช้ประโยชน์เพื่อการอธิบา
ยรูปแบบที่น่าสนใจจากกลุ่มข้อมูลได้ เช่น
แสดงผลในรูปของกฎความสัมพันธ์ (Association Rules) หรือ
การจัดกลุ่มข้อมูล (Clustering) นอกจากนั้น
การทำเหมืองข้อมูล ยังสามารถนำไปใช้ประโยชน์
เพื่อการทำนาย ซึ่งเป็นการนำความรู้ที่เรียนรู้มาจากข้อมูลใน
อดีต มาสร้างรูปแบบจำลอง (Model) แล้วสามารถนำไปใช้
ทำนาย หรือ พยากรณ์ การเกิดเหตุการณ์ที่จะเกิดขึ้นในอนาคต
เทคนิคที่ใช้เรียกว่า เทคนิคการจำแนกประเภท กรณีข้อมูล
เป้าหมายเป็นเชิงกลุ่ม
และเทคนิคการพยากรณ์หรือประมาณค่ากรณีข้อมูลเป้าหมายเป็
นเชิงปริมาณ

**2.3 ต้นไม้ตัดสินใจ (Decision Tree)**

คือการเรียนรู้ของต้นไม้ตัดสินใจ (Decision Tree) [5]
เป็นการเรียนรู้โดยการจำแนกประเภท (Classification)
ข้อมูลออกเป็นกลุ่ม (Class) ต่าง ๆ โดยใช้คุณลักษณะ
(Attribute)
ข้อมูลในการจำแนกประเภทต้นไม้ตัดสินใจที่ได้จากการเรียนรู้
ทำให้ทราบว่าคุณลกษณะใดเป็นตัวกำหนดการจำแนกประเภท
และคุณลักษณะแต่ละตัวมีความสำคัญมากน้อยต่างกันอย่างไร
เพราะฉะนั้น การจำแนกประเภทมีประโยชน์ช่วยให้สามารถ
วิเคราะห์ข้อมูลและตัดสินใจได้ถูกต้องยิ่งขึ้น

**2.4 K-Nearest Neighbor**

คือ วิธีการในการจัดแบ่งคลาส เทคนิคนี้จะตัดสินใจ
ว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง
โดยการตรวจสอบจำนวนบางจำนวน (“K” ใน K-nearest
neighbor)
ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด
โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข
หรือกรณีต่างๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ
ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

**2.5 วิธี Cross-validation Test**

วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัย
เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดลเนื่องจากผล
ที่ได้มีความน่าเชื่อถือ การวัด ประสิทธิภาพด้วยวิธี Cross-
validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน
(มักจะแสดงด้วยค่า k) เช่น 5-fold cross-validation คือ
ทำการแบ่งข้อมูลออกเป็น 5 ส่วน
โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-
validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน
โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน
หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพ
ของโมเดลทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ เช่น
การทดสอบด้วยวิธี 5-fold cross-validation

**2.6 งานวิจัยที่เกี่ยวข้อง**

การทำนายความเสี่ยงในการเกิดภาวะแทรกซ้อนโรคความดันโ
ลหิตสูงของผู้ป่วยโรคเบาหวาน
โดยใช้เทคนิคการทำเหมืองข้อมูล [1]
ด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification)
จากโมเดลแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree)
และการจัดแบ่งคลาส (k-NN)
เพื่อวิเคราะห์และทำนายผู้ป่วยถึงภาวะความเสี่ยงต่อการเกิดโรค
ความดันโลหิตสูงและทำการเปรียบเทียบผลการทำนาย
โดยที่การทำนายด้วยแผนภาพต้นไม้เพื่อการตัดสินใจ

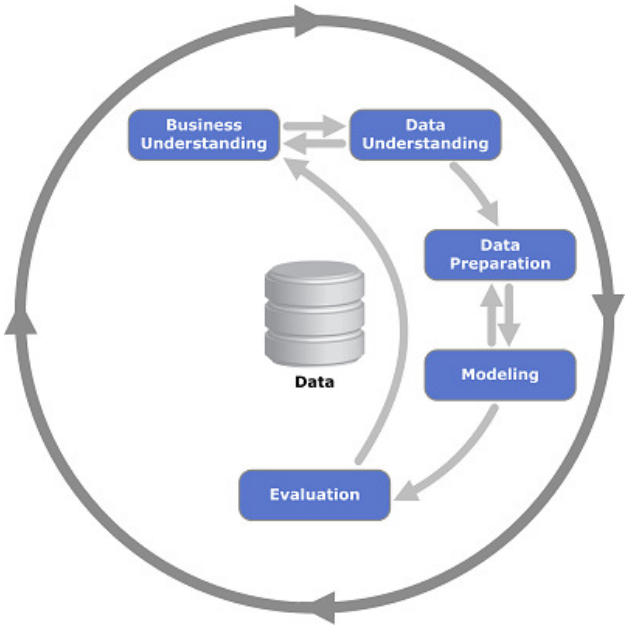
(Decision Tree) มีความแม่นยำมากกว่าการจัดแบ่งคลาส(k-NN) ถึง 2.08 เปอร์เซ็นต์

การศึกษาสาเหตุการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุ กรณีศึกษาโรงพยาบาลภาครัฐแห่งหนึ่งในประเทศไทย [2] ที่เป็นการวิเคราะห์ร่วมกันของ 2 ขั้นตอนวิธีคือการสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) และสร้างแบบจำลองการวิเคราะห์เส้นทาง (Path Analysis) เพื่อให้ได้ตัวแบบที่ใช้ในการหาความสัมพันธ์ของการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุที่มีความแม่นยำมากขึ้น และตัวแบบที่ได้ จะสามารถนำไปใช้สร้างโปรแกรมสนับสนุนทางการแพทย์สะดวกในการวิเคราะห์หาสาเหตุการเกิดโรค

การศึกษาสาเหตุการเกิดโรคความดันโลหิตสูง กรณีศึกษาโรงพยาบาลภาครัฐแห่งหนึ่งในประเทศไทย [3] ด้วยเทคนิคการทำเหมืองข้อมูล (Data Mining) ตามกรอบ CRISPDM ที่เป็นการวิเคราะห์ร่วมกันของ 2 ขั้นตอนวิธี (Algorithm) คือการสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) และการวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression) พบว่ามีโรคร่วมกับความดันโลหิตสูงที่สามารถพยากรณ์ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ ภาวะไขมันในเลือดสูงเกินไม่ระบุรายละเอียด, เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลินไม่มีภาวะแทรกซ้อน, โรคหัวใจจากหลอดเลือดแดงแข็ง, การมีเลนส์เทียมในลูกตา, ผลการตรวจระดับน้ำตาลในเลือดหลังอดอาหารผิดปกติ, ไตวายเรื้อรังไม่ระบุรายละเอียด, ข้อเข่าเสื่อมปฐมภูมิ สองข้าง, การตรวจติดตามผลหลังการรักษาภาวะอื่นด้วยการผ่าตัด, การเจริญเกินของต่อมลูกหมาก,โรคของระบบไหลเวียน และเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลินร่วมกับภาวะแทรกซ้อนทางตา ซึ่งการหาความสัมพันธ์ของการเกิดโรค ทำให้ทราบถึงปัจจัยใดที่ส่งผลให้เกิดโรคความดันโลหิตสูง และเป็นแนวทางในการป้องกันการเกิดโรคความดันโลหิตสูงในอนาคต

## 3. วิธีการดำเนินการวิจัย

งานวิจัยครั้งนี้ ได้นำข้อมูลผู้ป่วยที่เป็นโรคความดันโลหิตสูงฐานข้อมูลที่มีอยู่แล้ว ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary Data) มาศึกษาค้นหาความรู้ความสัมพันธ์ของข้อมูลที่แฝงอยู่ด้วยเทคนิคการทำเหมืองข้อมูลตามกรอบของ CRISP-DM โดยผู้วิจัยได้มีการปรับปรุงกรอบการวิจัยข้างต้นให้เหมาะสมตามวัตถุประสงค์การวิจัย




**ภาพที่ 1 :** ขั้นตอนการดำเนินการวิจัย

3.1 การเข้าใจสภาพของปัญหา

สาเหตุการเกิดโรคความดันโลหิตสูง (Hypertension Disease Understanding) อาจเกิดจากพันธุกรรม สิ่งแวดล้อม เช่น ความเครียด การรับประทานอาหารเค็ม สูบบุหรี่ ไม่ได้ออกกำลังกาย ความดันในกะโหลกศีรษะสูง เป็นต้น ซึ่งแพทย์จะรักษาไปตามอาการ ทางผู้วิจัยจะนำข้อมูลผู้ป่วยมาพยากรณ์โรคความดันโลหิตสูง เพื่อเป็นแนวทางในการหาสาเหตุของการเกิดโรคความดันโลหิตสูง

3.2 การเข้าใจข้อมูลและแหล่งที่มา

ลักษณะข้อมูลโรคความดันโลหิตสูง (Data Hypertension) ข้อมูลที่นำมาใช้ในทำการวิเคราะห์นั้น เป็นข้อมูลถูกเก็บไว้ในฐานข้อมูลส่วนกลาง โดยทางผู้วิจัยได้ทำการคัดเลือกในขั้นตอนที่ 3.3 เพื่อให้ได้ข้อมูลที่เหมาะสมต่อไป

| Data Descriptive |
|---|
| เพศ |
| อายุ |
| อาชีพ |
| ภูมิภาค |
| เชื้อชาติ |
| สัญชาติ |
| ศาสนา |

| |
|---|
| ค่าความดันโลหิตบน BPS |
| ค่าความดันโลหิตล่าง BPD |
| ระยะเวลาที่ป่วย |
| ประเภทของผู้ป่วย |
| การสูบบุหรี่ |
| การดื่มสุรา |
| ประวัติความดันโลหิตในญาติสายตรง |
| ดัชนีมวลกายเส้นรอบเอว |
| โรคที่เกิดร่วมโรคภาวะไขมันในเลือดสูงเกิน |
| โรคที่เกิดร่วมโรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน |
| โรคที่เกิดร่วมโรคหัวใจจากหลอดเลือดแดงแข็ง |
| การมีเลนส์เทียมในลูกตา |
| ผลการตรวจระดับน้ำตาลในเลือดหลังอดอาหารผิดปกติ |
| ต้อกระจกในวัยชราชนิดนิวเคลียส |
| โรคที่เกิดร่วมโรคไตวายเรื้อรัง |
| ข้อเข่าเสื่อมปฐมภูมิสองข้าง |
| การตรวจติดตามผลหลังการรักษาภาวะอื่นด้วยการผ่าตัด |
| การเจริญเกินของต่อลูกหมาก |
| การตรวจทางรังสีวิทยา |
| โรคที่เกิดร่วมโรคของระบบไหลเวียน |
| โรคที่เกิดร่วมโรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลินร่วมกับภาวะแทรกซ้อนทางตา |
| การมีอุปกรณ์ฝังและปลูกถ่ายเพื่อซ่อมแซมหลอดเลือดหัวใจ |
| ภาวะคอเลสเตอรอลในเลือดสูงเกินอย่างเดียว |

ตารางที่ 2 : ลักษณะข้อมูลที่ใช้ในการวิเคราะห์

3.3 การเตรียมข้อมูล

ทำการแปลงค่าข้อมูลที่อยู่ในรูปของอักขระภาษาไทยเป็นภาษาอังกฤษหรือตัวเลข จากข้อมูลผู้ป่วย 2000 คน ชุดที่ 1 เป็นข้อมูลผู้ป่วยความดันโลหิตสูงสำหรับการสร้างโมเดลต้นแบบ จำนวน 1000 คน และข้อมูล ชุดที่ 2 เป็นข้อมูลผู้ป่วยที่ไม่เป็นโรคความดันโลหิตสูง 1000 คน การแปลงข้อมูลสำหรับทดสอบการทำนายของทั้ง 2 โมเดล โดยที่ไม่มีการตั้งสมมุติฐานในการตรวจสอบค่าความดันโลหิต

3.4 การสร้างตัวแบบที่เหมาะสม

จากข้อมูลที่ได้จากการแปลงค่าข้อมูลเรียบร้อยแล้วจากทำการสร้างโมเดลต้นแบบในการทำนายของทั้งโมเดล Decision Tree และ K-nearest neighbor)

3.5 การทดสอบประสิทธิภาพ

นำแบบจำลองการทำนายความเสี่ยงของการเกิดโรคความดันโลหิตสูงไปทดสอบประสิทธิภาพโดยวิธี K-fold Cross validation กำหนดค่า k เท่ากับ 10 และประเมินค่าความถูกต้อง (Accuracy)

| Predict/Actual | HT | Not HT |
|---|---|---|
| Predict HT | TP | FT |
| Predict Not HT | FN | TN |

ตารางที่ 3 : ตาราง Confusion Matrix ของข้อมูลทำนายการเกิดโรคความดันโลหิตสูง

# 4. ผลการดำเนินการวิจัย

| โมเดล | ความถูกต้อง (%) |
|---|---|
| Decision Tree | 96.40 |
| K-nearest neighbor | 92.80 |

ตารางที่ 4 : แสดงการเปรียบเทียบค่าความถูกต้อง

## 5. สรุปผลการดำเนินงานวิจัย

การวิจัยนี้ศึกษาสาเหตุของการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุ พบว่าเทคนิค Decision Tree ให้ผลการทำนายที่ถูกต้องมากกว่าคือ ร้อยละ 96.40 ขณะที่เทคนิค K-nearest neighbor ให้ผลการทำนายถูกต้อง ร้อยละ 92.80 โดยผลที่คาดว่าจะได้รับจากตัวแบบการทำเหมืองข้อมูลสามารถนำไปเป็นประโยชน์ทางด้านการแพทย์ในการหาสาเหตุของการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุ และสามารถนำไปพัฒนาเป็นโปรแกรมในการวิเคราะห์หาสาเหตุของโรคใน อนาคตได้ ส่งผลให้ลดอัตราการเกิดโรคภาวะโรคแทรก ซ้อน ความเสี่ยง และอัตราการตาย เพื่อสร้างคุณภาพชีวิต ที่ดีต่อไป

## เอกสารอ้างอิง


[1] Rungruedee Butdachan, and Todsanai Chumwatana, “Prediction on risk for complications of hypertension in diabetes patient using data mining technique,” Journal of the Thai Medical Informatics Association, Vol 1, pp 1-8, 2017.

[2] ทัศนนท์ ฐานประเสริฐกล และ ดร.กมล เกียรติเรืองกมลา, “The Cause Predicting to Blood Pressure Essential Hypertension A Case Study of Public Hospital in Thailand,” 34th the international graduate conference, pp 1023-1031.

[3] ทัศนนท์ ฐานประเสริฐกลุ และ ดร.กมล เกียรติเรืองกมลา, “The Cause Predicting to Blood Pressure Hypertension A Case Study of Public Hospital in Thailand,” The 8th National Conference on Technical Education, pp 289-297, 2015.

[4] ฉิชา นภาพร จงกะสิกจิ, “ระบบสนับสนุนการตัดสินใจในการวิเคราะห์ความเสี่ยงของการเกิดโรคเรื้อรัง กรณีของโรคเบาหวานและความดันโลหิตสูง,” วารสารวิชาการ คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏลำปาง ปีที่ 9 ฉบับที่ 2 กรกฎาคม 2559 – ธันวาคม 2559, หน้า 11-19.

[5] บรรจบ ดลกุล จารี ทองคา และ วาทินี สุขมาก, “การสร้างแบบจำลองเพื่อพยากรณ์การเกิดแผลที่เท้าของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคเหมืองข้อมูล,” วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม ปีที่ 33 ฉบบัที่ 6 พ.ย.-ธ.ค. 2557 หน้า 703-710.

[6] เอกสิทธิ์ พชัรวงศศ์กัดา .การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิงเบื้องต้น. กรุงเทพฯ: บริษทั เอเชีย ดิจิตอลการพิมพ์จา กัด. 2557.

[7] สำนักนโยบายและแผน. สำนักงานปลัดกระทรวงสาธารณสุข. รายงานสถิติสาธารณสุข พ.ศ. 2555 สืบค้นเมื่อวนัที่ 10 พฤศจิกายน 2557, จาก http://www.bps.moph.go.th/

[8] Patcharaporn Panwong and Natthakan Iam-On, “Predicting Transitional Interval of Kidney Disease Stages 3 to 5 Using Data Mining Method,” Second Asian Conference on Defence Technology (ACDT), 2016.

[9] Sankaranarayanan.S and Pramananda Perumal.T, “Diabetic prognosis through Data Mining Methods and Techniques,” International Conference on Intelligent Computing Applications, pp. 162-166, 2014.

[10] Business Analytics software, CRISP-DM 1.0 Step-by-step Data Mining Guide, Technical report, IBM Corporation 2010.

[11] WEKA: Waikato Environment for Knowledge Analysis, Available at: http://www.cs.waikato.ac.nz/ml/weka/

[12] Rokach L, Maimom O. “Data Mining With DecisionTree Theory and Application,” 2nd ed2014. 11.

[13] Aggarwal C, “k-NN. Data Classification Algorithms and Application 2015,”