

การสร้างโมเดลทำนายโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล

Building the Predicting Model of the Chance of Repeating the Treatment of Diabetes Patient Using Data Mining

วิธวินท์ แสงมณี¹ วีระวุฒิ รัตนเจริญเลิศ ญัฏฐภาพ โพธิ์รัชต์ และ เพ็ญฤทัย หนูสวัสดิ์

สาขาวิชาเทคโนโลยีสารสนเทศทางธุรกิจ คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ วิทยาเขตวังไกลกังวล

Email : witawin1000000@gmail.com, aitindy14@hotmail.com, Laocyl3er@gmail.com, paingruthai.nus@rmutr.ac.th

บทคัดย่อ

งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการทำนายโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวาน โดยมีวัตถุประสงค์เพื่อ 1.วิเคราะห์ปัจจัยที่ส่งผลต่อโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวาน 2. สร้างโมเดลสำหรับทำนายโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล 3. เปรียบเทียบประสิทธิภาพของโมเดลที่สร้างจากเทคนิค Decision tree ,Naïve Bays และ K-Nearest Neighbors

ผลการสร้างโมเดลในการทำนายโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล คือโมเดลที่สร้างโดยใช้เทคนิค Decision Tree มีประสิทธิภาพการทำนายสูงสุดคิดเป็นร้อยละ 85.5 ซึ่งสามารถนำผลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วนของแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

ABSTRACT

This research applies Data Mining Technique to assess the probability of diabetic patient's readmissions. The objectives are as follows; 1. Analyzing what factors cause the readmissions. 2. Producing model to assess the probability of readmissions by using Data Mining Technique 3. Comparing efficiency of models created from Decision tree, Naïve Bays and K-Nearest Neighbors techniques.

The findings reveal that Decision tree model is found to be the most proficient technique, with the highest percentage of 85.50. As a result, we can apply this model to create a Decision Support System for doctors and patients.

คำสำคัญ— โรคเบาหวาน; การทำเหมืองข้อมูล; Decision tree; Naïve Bays; K-Nearest Neighbors

1. บทนำ

โรคเบาหวาน (Diabetes mellitus (DM)) หรือทั่วไปว่า Diabetes) เป็นกลุ่มโรคเกี่ยวกับการเผาผลาญอาหารซึ่งมีระดับน้ำตาลในเลือดสูงเป็นเวลานาน น้ำตาลในเลือดสูงก่อให้เกิดอาการปัสสาวะบ่อย กระหายน้ำและความหิวเพิ่มขึ้น หากไม่ได้รับการรักษา เบาหวานอาจก่อให้เกิดอาการแทรกซ้อนจำนวนมากภาวะแทรกซ้อนเฉียบพลัน ได้แก่ ภาวะเลือดเป็นกรดจากคีโตนจากเบาหวาน (diabetic ketoacidosis) และโคม่าเนื่องจากออสโมลาร์สูงที่ไม่ได้เกิดจากคีโตน (nonketotic hyperosmolar coma) ภาวะแทรกซ้อนระยะยาวที่ร้ายแรงรวมถึงโรคหัวใจ, โรคหลอดเลือดสมอง, ไตวาย, แผลที่เท้าและความเสียหายต่อตา

เบาหวานเป็นโรคเรื้อรังที่เป็นปัญหาสำคัญทางด้านสาธารณสุขของโลก เป็นภัยคุกคามที่คุกคามอย่างรวดเร็วไปทั่วโลก ส่งผลกระทบต่อการพัฒนาทางเศรษฐกิจอย่างมากจากข้อมูลสมาพันธ์เบาหวานนานาชาติ(international diabetes federation : IDF) ได้รายงานไว้ในปัจจุบันทั่วโลกมีผู้เสียชีวิตด้วยโรคเบาหวาน ๔ ล้านคนต่อปี เฉลี่ย ๘ วินาทีต่อ ๑ คน สำหรับผู้เป็นเบาหวานพบมากกว่า ๓๐๐ ล้านคนและพบว่าคนที่อยู่ในประเทศที่มีรายได้ต่ำและปานกลางมีโอกาสเป็นเบาหวานเร็วกว่าคนที่อยู่ในประเทศที่มีรายได้สูง ๑๐ ปี โดยพบมาก ๒๐ - ขึ้นในวัยทำงานซึ่งปัญหาอย่างหนึ่งที่พบของแพทย์และผู้ป่วยโรคเบาหวานคือแนวทางในการรักษา โดยหลังจากที่ผ่านการรักษาแล้วนั้น ผู้ป่วยจะมีโอกาสกลับมาเป็นโรคเบาหวานอีกหรือไม่ โดยการวิเคราะห์ข้อมูลนี้ต้องอาศัยการวินิจฉัยจากแพทย์เพียงอย่างเดียว ซึ่งบางครั้งทำให้ผู้ป่วยหรือตัวแพทย์เองอาจจะเกิดความล่าช้าในการวินิจฉัยหรือการตรวจสอบอาการที่เกิดขึ้น

จากปัญหาดังกล่าวข้างต้น ทางผู้จัดทำจึงได้มีการนำเทคนิคการทำเหมืองข้อมูล ซึ่งเป็นการค้นหาความรู้จากแหล่งข้อมูลที่มีขนาดใหญ่ เพื่อทำการค้นหากฎความสัมพันธ์การจัดกลุ่มข้อมูล หรือการทำนายพยากรณ์การเกิดเหตุการณ์ต่างๆ ซึ่งในปัจจุบันนั้นหน่วยงานต่างๆ เช่น ธนาคาร

ห้างสรรพสินค้า กลุ่มบริษัทประกันภัย และกลุ่มโรงพยาบาล ต่างๆ ก็ได้มีการนำเทคนิคการเหมืองข้อมูลมาประยุกต์ใช้ในการทำงานด้านต่างๆในส่วนของการการกระทำกับข้อมูลที่มี เพื่อให้เกิดประโยชน์ต่อหน่วยงานหรือองค์กรต่อไป ทางผู้วิจัยจึงได้ทำเทคนิคของการทำเหมืองข้อมูลคือเทคนิค Classification มาใช้สำหรับการทำนายโอกาสการกลับมารักษาทัวของผู้ป่วยโรคเบาหวาน เพื่อวิเคราะห์ว่าผู้ป่วยโรคเบาหวาน จะมีโอกาสการกลับมารักษาทัวซ้ำหรือไม่ นั้นหมายถึงโอกาสของการกลับมาเป็นโรคเบาหวาน โดยใช้ข้อมูลจากการรับยาของผู้ป่วยแต่ละคนหรือปัจจัยอื่นๆที่เกี่ยวข้องมาเป็นปัจจัยสำหรับการวิเคราะห์ ซึ่งจะมีผลดีต่อการทำงานของแพทย์ คือ ช่วยลดระยะเวลาในการประเมินผลการกลับเข้ามารักษา และ ผู้ป่วยสามารถทำได้ด้วยตัวเองโดยการประเมินจากผลน้ำตาลในเลือดและการใช้ยาทำให้เกิดความสะดวกสบายให้แก่ทั้งผู้ป่วยและแพทย์เอง และสามารถนำไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วน of แนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

2. วัตถุประสงค์ของงานวิจัย

- 2.1 เพื่อศึกษาและวิเคราะห์ปัจจัยส่งผลต่อโอกาสการกลับมารักษาทัวของผู้ป่วยโรคเบาหวาน
- 2.2 เพื่อสร้างโมเดลที่ใช้สำหรับทำนายโอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล
- 2.3 เพื่อเปรียบเทียบประสิทธิภาพของโมเดลที่ได้จากการสร้างเทคนิค Decision tree, Naïve Bays และ K-Nearest Neighbors

3. ขอบเขตงานวิจัย

- 3.1 ใช้ข้อมูลกลุ่มตัวอย่างจากจากผู้ป่วยโรคเบาหวานหรือเสี่ยงจะเป็นโรคเบาหวาน โรงพยาบาลของสหรัฐอเมริกาตั้งแต่ปี พ.ศ. 2542-2551เป็นเวลา 10 ปี จำนวน 101,766 แถว จากเว็บไซต์ <https://archive.ics.uci.edu/ml/index.html>
- 3.2 ปัจจัยที่ใช้ในการวิเคราะห์โอกาสการกลับมารักษาทัวซ้ำของผู้ป่วยโรคเบาหวานมีจำนวน 14 แอตทริบิวต์ ได้แก่ รหัสประจำตัวผู้ป่วย (patient_nbr), เพศ (gender), อายุ (age), น้ำหนัก (weight), ประเภทของการเข้ารับการรักษา (admission_type_id), ประเภทของการรับการรักษาต่อ (discharge_disposition_id), ประวัติการรักษาของผู้ป่วย (admission_source_id), ระยะเวลาในการเข้ารับการรักษา (time_in_hospital), จำนวนครั้งในการเข้ารับการรักษาในห้องปฏิบัติการ (number of lab procedures), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยนอกในปีก่อนหน้า (Number of Outpatient visits), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยในในปีก่อนหน้า (Number of Inpatient visits), จำนวน

ในการเข้ารับการรักษาแบบฉุกเฉินของผู้ป่วย (Number of Emergency visit), จำนวนยาที่ได้รับ (Number of medications), จำนวนขั้นตอนของการรักษา (Number of procedures) โอกาสการกลับมารักษาทัวของผู้ป่วย (readmitted)

3.3 ใช้เทคนิค Classification โดยเลือกใช้เทคนิค Decision tree, Naive Bays และ K-Nearest Neighbors และใช้โปรแกรม Rapid miner ในการสร้างโมเดลการทำนาย

3.4 ใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลโดยใช้การแบ่งค่าข้อมูลแบบ Cross-validation Test เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลค่า K= 10 กลุ่ม, k=15 กลุ่ม และ k= 20 กลุ่ม

3.5 หาค่าความถูกต้องของโมเดลที่ได้โดยประเมินจากค่าความถูกต้อง (Accuracy)

4. แนวคิดและทฤษฎีที่เกี่ยวข้อง

4.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) [6] คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมาย ข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูลประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

1. Relational Database เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model

2. Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่เดียวกัน

3. Transactional Database ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปแบบชื้อลูกค้าและรายการสินค้าที่ลูกค้ารายชื้อ

4. Advanced Database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่นๆ เช่น ข้อมูลแบบ Object-Oriented ข้อมูลที่เป็น Text File ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ Web [6]

4.2 การเรียนรู้ต้นไม้ตัดสินใจ (Decision tree)

เทคนิคต้นไม้ตัดสินใจ (Decision Tree) [3] คือ แบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไขแล้วผลลัพธ์” เช่น

“If Income = High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Good”

4.3 ทฤษฎีบทของเบย์ (Naive Bayes)

การเรียนรู้แบบเบย์ เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็นตามกฎของเบย์ (Bayes' Theorem) [10] เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด การเรียนรู้แบบเบย์อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน ในที่นี้คือคลาสเป้าหมายหรือผลลัพธ์การทำนายโดยการเรียนรู้แบบเบย์เป็น (การเรียนรู้เพิ่มเติม เนื่องจาก ตัวอย่างใหม่ที่ได้มาถูกนำมาปรับเปลี่ยนการแจกแจงซึ่งมีผลต่อการเพิ่มหรือ ลดความน่าจะเป็น ทำให้มีการเรียนรู้ที่เปลี่ยนไป วิธีการนี้ตัวแบบจะถูกปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้โดยผนวกกับความรู้เดิมที่มี ซึ่งการทำนายค่าคลาสเป้าหมายของตัวอย่างใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐาน

จากทฤษฎีของเบย์ เราสามารถคำนวณความน่าจะเป็นของสมมติฐานต่าง ๆ โดยใช้สมการ

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$P(h)$ คือ ความน่าจะเป็นก่อนหน้าของสมมติฐาน h

$P(D)$ คือ ความน่าจะเป็นก่อนหน้าของชุดข้อมูล

ตัวอย่าง D

$P(h|D)$ คือ ความน่าจะเป็นของ h ขึ้นต่อ D

$P(D|h)$ คือ ความน่าจะเป็นของ D ขึ้นต่อ h

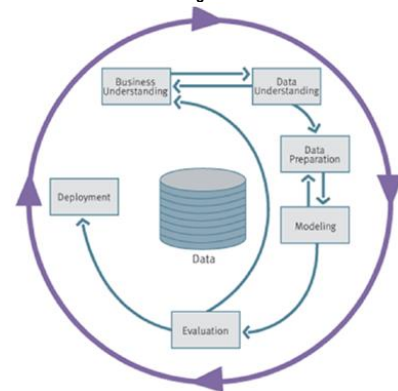
4.4 K-Nearest Neighbors

คือ วิธีการในการจัดแบ่งคลาส เทคนิคนี้จะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน (“K” ใน K-nearest neighbor) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม

ของจำนวนเงื่อนไข หรือกรณีต่างๆสำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

หลักการการทำงานของ K-NN คล้ายๆกับการแบ่งข้อมูลคือทำการวัดระยะห่างระหว่างข้อมูลที่ต้องการทำนายกับข้อมูลที่อยู่ใกล้เคียงเป็นจำนวน K ตัว และคำตอบที่ได้คือคลาสที่พบมากที่สุดของข้อมูลที่เป็นเพื่อนบ้านทั้ง K ตัว ในเทคนิคนี้มักจะใช้วิธีการวัดระยะห่างแบบ Euclidean ซึ่งเกิดจากรากที่สองของผลต่างระหว่างแอตทริบิวต์ต่างๆยกกำลังสอง [11]

4.5 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM



รูปที่ 1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

ที่มา :Eakasit Pacharawongsakda,Ph.D. (2557)

กระบวนการ CRISP-DM ประกอบด้วย 6 ขั้นตอน ได้แก่

1. Business Understanding

เป็นขั้นตอนแรกสุดในกระบวนการ CRISP-DM เป็นการทำความเข้าใจและระบุปัญหาหรือโอกาสเชิงธุรกิจ ระบุ output ที่ต้องการได้จากการวิเคราะห์ข้อมูลด้วย Data mining และวางแผนงาน

2. Data Understanding

ข้อมูลเป็นปัจจัยที่สำคัญที่สุดที่ขาดไม่ได้ในการทำ Data mining ในขั้นตอนนี้เป็นการรวบรวมข้อมูลที่เกี่ยวข้อง ข้อมูลถูกต้อน่าเชื่อถือ ข้อมูลที่ได้มีปริมาณมากพอหรือยัง และข้อมูลที่ได้มีความเหมาะสม มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

3. Data Preparation

ขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลานานที่สุด เนื่องจากโมเดลที่ได้จากการทำดาต้าไมนิงจะให้ผลลัพธ์ที่ถูกต้อนหรือไม่ ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ โดยการเตรียมข้อมูลนั้นสามารถแบ่งออกได้เป็น 3 ขั้นตอนย่อยคือ Data Cleaning Data Selection และ Data Transformation

4. Modeling

เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมนิง เช่น เทคนิค Association เทคนิคการทำ Clustering และเทคนิคการทำ Classification

5. Evaluation

การประเมินหรือวัดประสิทธิภาพของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้น เป็นเพียงการวัดความน่าเชื่อถือของโมเดลเท่านั้น ในขั้นตอนนี้เป็นการประเมินประสิทธิภาพของผลลัพธ์จากโมเดลวิเคราะห์ข้อมูลว่าครอบคลุมและสามารถตอบโจทย์ทางธุรกิจที่ตั้งไว้ในขั้นตอนแรกหรือไม่

6. Deployment

การนำผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมนิงนำไปใช้งานจริง ตัวอย่างเช่น การนำองค์ความรู้ที่ได้ไปใช้ในการจัดโปรโมชันส่งเสริมการขายสินค้า ใช้ในการทำนายแนวโน้มการทุจริตในระบบการเงินของธนาคารหรือตรวจจับความผิดปกติในการซื้อขายหุ้น ในตลาดหลักทรัพย์ เป็นต้น [8]

4.6 งานวิจัยที่เกี่ยวข้อง

(นายปพน,กาญจนา,เบญจภัก,2015) ได้ศึกษาความเสี่ยงของการเกิดภาวะแทรกซ้อนที่เท้าของผู้ป่วยโรคเบาหวาน โดยการพัฒนาและออกแบบให้อยู่ในรูปแบบของเว็บแอปพลิเคชันและใช้การคำนวณสมการคณิตศาสตร์โดยสมการที่ใช้ มีดังนี้สมการเชิงเส้น สมการเอ็กซ์โพเนนเชียล, ลอการิทึม, โพลีโนเมียล, ยกกำลัง เป็นสมการที่ใช้ประเมินผลลัพธ์ซึ่งวิธีการเก็บข้อมูลของปัจจัยที่เกี่ยวข้องกับผู้ป่วยโรคเบาหวานในโรงพยาบาล จากปีค.ศ. 2000 ถึง 2004 โดยใช้เทคนิคของดาต้าไมนิงมาประยุกต์ใช้ เพื่อพัฒนาพร้อมทั้งการคำนวณประสิทธิภาพของเครื่องมือ ซึ่งเทคนิคที่นำมาใช้สามารถจัดกลุ่มตามแบบกฎของเบย์และใช้อัลกอริทึม C4.5มาประยุกต์ใช้ในการทำนายและคาดการณ์ข้อมูล ซึ่งให้ผลการทดลองออกมาคือระบบสนับสนุนการตัดสินใจเพื่อวิเคราะห์ข้อมูลผู้ป่วยโรคเบาหวาน ที่มีความเสี่ยงของการเกิดภาวะแทรกซ้อนที่เท้ามีกระบวนการทั้งหมด 3 วิธีสำหรับผู้ใช้งาน 1. เก็บข้อมูลและตรวจเท้า 2. วิเคราะห์ปัจจัยที่ทำให้เกิดภาวะแทรกซ้อนที่เท้า 3. ประเมินผลการทดลอง ผลลัพธ์ที่ได้จากการประเมินความพึงพอใจโดยผู้เชี่ยวชาญมีค่าเฉลี่ย 3.63 อยู่ในระดับมาก ในส่วนของงานวิจัยในอนาคตสามารถนำไปประยุกต์ใช้ในการทำนิรอลเน็ตเวิร์กได้ [1]

(นางสาวเรวดี,2010) ได้ศึกษาปัจจัยที่มีผลต่อความสำเร็จในการรักษาโรคนิ้วล็อกในแบบต่างๆของคณะแพทยศาสตร์ศิริพยาบาลโดยใช้การใช้เทคนิคดาต้าไมนิงเพื่อนำไปใช้ประกอบการตัดสินใจของแพทย์และบุคคลทั่วไปในการเลือกแนวทางในการรักษาโรคนิ้วล็อกโดยการนำข้อมูลส่วนตัว ข้อมูลการตรวจร่างกายและข้อมูลการรักษาโรคนิ้วล็อกของผู้ป่วยมาทำการวิเคราะห์ด้วยเทคนิคดาต้าไมนิงโดยใช้เทคนิคการแบ่งกลุ่มเทคนิควิธีการในการแบ่งกลุ่ม (Clustering)แบบ K-means Clustering เพื่อแบ่งกลุ่มผู้ป่วยเป็น 3 กลุ่ม ซึ่งให้ผล

สรุปออกมาเป็น 3 กลุ่มคือ กลุ่มที่ 1 มีผู้ป่วยจำนวน 749 คน ซึ่งผู้ป่วยส่วนใหญ่ ไม่เคยได้รับการรักษามาก่อน(80.11%)และมีระดับความเจ็บปวดมากพอสมควร นี้นี้มีการสละแบบเหยียดนิ้วออกเองได้และมีอาการปวดปานกลาง เมื่อกดที่ตำแหน่ง A1 Pulley ที่ปลอกหุ้มเอ็นแล้วได้รับการรักษาโดยการผ่าตัดขยายปลอกหุ้มเอ็น (55.94%) หรือการใช้เข็มเจาะขยายที่ปลอกหุ้มเอ็น(43.93%) กลุ่มที่ 2 มีผู้ป่วยจำนวน 759 คน ซึ่งผู้ป่วยส่วนใหญ่ ไม่เคยได้รับการรักษามาก่อน(91.44%) และมีระดับความเจ็บปวดพอประมาณ (น้อยกว่ากลุ่มที่1) นี้นี้มีการแข็งในทางอ้อมเหยียดก็เมื่อย และมีอาการปวดเล็กน้อย เมื่อกดที่A1 Pulley ที่ปลอกเอ็นแล้วได้รับการรักษาโดยการฉีดยา (84.85%)กลุ่มที่ 3 มีผู้ป่วยจำนวน 448 คน เป็นกลุ่มที่มีจำนวนน้อยที่สุดซึ่งผู้ป่วยส่วนใหญ่เคยได้รับการรักษาด้วยวิธีการฉีดยามาก่อน (97.99%) โดยมีระดับความเจ็บปวดมากพอสมควรมีระดับการล็อกแบบงอนิ้วลงได้แต่เหยียดไม่ออก ต้องใช้มืออีกข้างมาช่วยจับ และมีอาการปวดปานกลางเมื่อกดที่A1 pulley ที่ปลอกเอ็นแล้วส่วนใหญ่ได้รับการรักษาโดยการผ่าตัด (47.99%) เพื่อไม่ให้อาการนิ้วล็อกกลับมาอีกและมีผู้ป่วยบางส่วน เลือกรับการรักษาแบบฉีดยาอีกครั้ง(25.45%) [4]

(ขุภิกกรณ, วิภา,วิทยา,2014)ได้ทำการพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ประวัติผลการตรวจรักษาผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานีตั้งแต่วันที่ 1 มกราคม 2553 ถึงวันที่ 31 ธันวาคม 2556 จำนวน 17,043 คน โดยใช้เทคนิค Data Mining หากกฎความสัมพันธ์(Association Rules) มาวิเคราะห์หาข้อมูลภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานแล้วการพัฒนาแบบจำลองครั้งนี้ต้องการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นเบาหวาน ใช้ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ซึ่งให้ผลออกมาคือผู้ป่วยที่เป็นโรคเบาหวานจะเป็นโรคแทรกซ้อนต่างๆ จากผลวิเคราะห์ข้อมูลภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี พบว่าผู้ป่วยที่เป็น 1) โรคเบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตาแล้วจะเป็นโรคแทรกซ้อน 1.1) โรคจอตาจากโรคเบาหวานชนิดไม่ออกขยาย (เอ็นพีดีอาร์) คิดเป็น 65 % 1.2) โรคต้อกระจกในผู้สูงอายุ คิดเป็น 46 % 1.3)โรคจอตาจากโรคเบาหวานชนิดงอกขยายคิดเป็น 44 % 2)โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตาแล้วจะเป็นโรคแทรกซ้อน 2.1)โรคจอตาจากโรคเบาหวานชนิดไม่ออกขยาย (เอ็นพีดีอาร์) คิดเป็น 77 % 2.2) โรคต้อกระจกในผู้สูงอายุ คิดเป็น 49 % 2.3) โรคจอตาจากโรคเบาหวานชนิดงอกขยายคิดเป็น 36 % 2.4) โรคไตวายเรื้อรังที่มีระดับน้ำตาลเยียด

คิดเป็น 20 % 3) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท นี้แล้วจะเป็นโรคแทรกซ้อน 3.1) โรคต่อกระจอในผู้สูงอายุคิดเป็น 24 % 3.2) โรคหัวใจขาดเลือดเรื้อรังที่มีไตรอะบายลอะเรียคิดเป็น 20 % [2]

5. วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานวิจัยในการสร้างโมเดลทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน อ้างอิงตามวิธีพัฒนาระบบ (Cross Industry Standard Process for Data Mining : CRISP DM)

5.1 การทำความเข้าใจในปัญหา (Business Understanding)

เบาหวานเป็นภาวะเรื้อรังของการมีระดับน้ำตาลในเลือดสูง ซึ่งเกิดการความบกพร่องของการสร้างอินซูลิน หรือการทำงานของอินซูลิน หรือทั้งสองกรณี โรคเบาหวานมักจะเกี่ยวข้องกับโรคแทรกซ้อนที่ร้ายแรงหลายโรคและอาจเป็นสาเหตุให้เสียชีวิตก่อนวัยอันควรปัจจุบันโรคเบาหวานเป็นโรคที่เกิดขึ้นได้กับทุกวัยไม่ว่าจะเป็นวัยเด็ก, วัยทำงาน, วัยชรา เพราะในปัจจุบันอาหารที่เราทานเข้าไปประกอบด้วยน้ำตาลเป็นส่วนใหญ่และการใช้ชีวิตของคนในปัจจุบันก็ไม่ได้ระมัดระวังในการรับประทานอาหาร ซึ่งโรคเบาหวานอาจจะทำให้มีโรคแทรกซ้อนต่างๆเข้ามามากมายที่เป็นอันตรายต่อชีวิตมาก อย่างเช่น โรคความดัน, โรคอ้วน, โรคไขมันอุดตันมีโรคแทรกซ้อนเกี่ยวกับตา, โรคแทรกซ้อนของการตั้งครรภ์ โรคเบาหวานเป็นโรคที่พบได้ในผู้ใหญ่ผู้ป่วยเบาหวานในประเทศไทยมีประมาณ 2.5-6 % ของประชากร จากปัญหาดังกล่าวข้างต้น ทางผู้จัดทำจึงได้มีการนำเทคนิคการทำเหมืองข้อมูล ซึ่งเป็นการค้นหาความรู้จากแหล่งข้อมูลที่มีขนาดใหญ่ เพื่อทำการค้นหาความสัมพันธ์การจัดกลุ่มข้อมูล หรือการทำนายพยากรณ์ เพื่อให้ผู้ทราบได้ว่าตัวของเค้านั้นจำเป็นต้องกลับเข้ามารักษาอีกหรือไม่

5.2 การทำความเข้าใจกับข้อมูล (Data Understanding)

ชุดข้อมูลที่ใช้เป็นกลุ่มตัวอย่างคือข้อมูลของผู้ป่วยโรคเบาหวานหรือเสี่ยงจะเป็นโรคเบาหวาน โรงพยาบาลของสหรัฐอเมริกา ตั้งแต่ปี พ.ศ. 2542-2551 เป็นเวลา 10 ปี จำนวน 101,766 แถว โดยข้อมูลประกอบด้วยคุณลักษณะต่างๆของผู้ป่วย ได้แก่

1. ข้อมูลเบื้องต้นของผู้ป่วย เช่น รหัสประจำตัวผู้ป่วย, อายุ, สัญชาติ, น้ำหนัก และเพศ
2. ข้อมูลการรักษาของผู้ป่วย เช่น ห้องที่รักษา, ประเภทผู้ป่วยนอก, ประเภทผู้ป่วยใน และประเภทการเข้ามารับการรักษา
3. ข้อมูลการให้ยา เช่น insulin, citoglipton, examide, tolazamide และ miglitol

5.3 การเตรียมข้อมูล (Data Preparation)

5.3.1 การคัดเลือก Attribute ที่เกี่ยวข้อง
ทำการตัดแอททริบิวต์ที่ไม่ส่งผลต่อค่าคำตอบจำนวน 39 Attribute และคัดเลือกเฉพาะ Attribute หรือปัจจัยที่ส่งผลต่อการทำนายโอกาสการกลับมารักษาตัวของผู้ป่วยโรคเบาหวาน ได้แก่ รหัสประจำตัวผู้ป่วย (patient_nbr), เพศ (gender), อายุ (age), น้ำหนัก (weight), ประเภทของการเข้ารับการรักษา (admission_type_id), ประเภทของการรับการรักษาต่อ (discharge_disposition_id), ประวัติการรักษาของผู้ป่วย (admission_source_id), ระยะเวลาในการเข้ารับการรักษา (time_in_hospital), จำนวนครั้งในการเข้ารับการรักษาในห้องปฏิบัติการ (number of lab procedures), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยนอกในปีก่อนหน้า (Number of Outpatient visits), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยในปีก่อนหน้า (Number of Inpatient visits), จำนวนในการเข้ารับการรักษาแบบฉุกเฉินของผู้ป่วย (Number of Emergency visit), จำนวนยาที่ได้รับ (Number of medications), จำนวนขั้นตอนของการรักษา (Number of procedures) โอกาสการกลับมารักษาตัวของผู้ป่วย (readmitted)

5.3.2 ลดความซ้ำซ้อนของข้อมูล

5.3.2.1 ทำการแปลงข้อมูลอายุของผู้ป่วยเป็นช่วงๆ ซึ่งแบ่งได้ ช่วงก็คือ 5

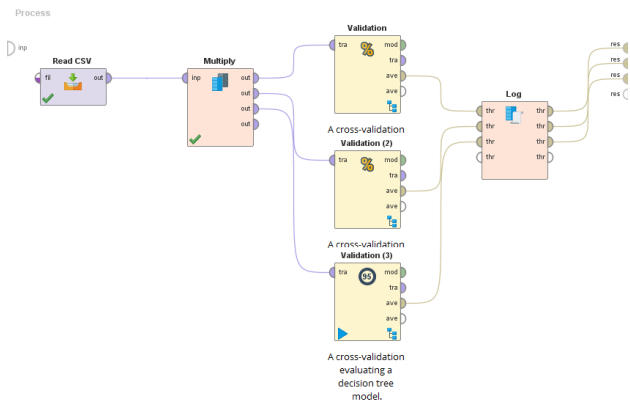
- ช่วงที่ 1 อายุ 0-50 ปี
- ช่วงที่ 2 อายุ 51-60 ปี
- ช่วงที่ 3 อายุ 61-70 ปี
- ช่วงที่ 4 อายุ 71-80 ปี
- ช่วงที่ 5 อายุ 81-100 ปี

5.3.2.2 ทำการแปลงใน Attribute readmitted ซึ่งเป็น Attribute ที่เป็น Label ของโมเดล

- >30 ไปเป็นค่า YES
- <30 ไปเป็นค่า NO

5.4 การสร้างโมเดล (Modeling)

ขั้นตอนนี้จะทำการสร้างโมเดลด้วยเทคนิค Decision tree, Naïve Bays และ K-Nearest Neighbors และใช้โปรแกรม Rapid miner studio 7.2 ในการสร้างโมเดลการทำนาย โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross-validation Test โดยแบ่งข้อมูลค่า K= 10 กลุ่ม, k=15 กลุ่ม และ k= 20 กลุ่ม เพื่อหาค่าความถูกต้องของโมเดลที่มากที่สุด



รูปที่ 2 แสดงกระบวนการสร้างโมเดล

5.5 การประเมินผล (Evaluation)

จากการทำกระบวนการสร้างโมเดลการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานตามเทคนิค Decision tree, Naïve Bays และ K-Nearest Neighbors ทำให้สามารถแบ่งกลุ่มผู้ป่วยที่ต้องกลับเข้ามารักษาทัวซ้ำแล้วผู้ป่วยที่ไม่ต้องกลับเข้ามารักษาทัวซ้ำได้ ซึ่งค่าความถูกต้องของโมเดลที่ดีที่สุดคือเทคนิค Decision tree ซึ่งเป็นค่าที่เชื่อถือได้ ดังนั้นจึงทำให้ผู้ป่วยที่ต้องการทราบถึงการกลับเข้ามารักษาทัวซ้ำซึ่งหมายถึงโอกาสการกลับมาเป็นโรคเบาหวาน ผู้ป่วยสามารถจะทราบถึงข้อมูลหรือพยากรณ์ล่วงหน้าได้ และในทางของแพทย์ยังสามารถทราบถึงโอกาสการกลับมารักษาตัวได้เร็วขึ้นกว่าการที่ต้องไปตรวจผู้ป่วยด้วยตนเองในเวลาสั้นๆ อย่างเช่น 2-3 เดือนตรวจหนึ่งครั้งจึงทำให้สามารถลดระยะเวลาในการรักษาได้

5.6 การนำไปใช้งาน (Deployment)

สามารถนำไปประยุกต์ใช้ในการทางการแพทย์สมัยใหม่ที่มี การนำระบบสารสนเทศเข้ามาช่วยในการสนับสนุนการตัดสินใจ เพื่อลดระยะเวลาในการทำการรักษาของแพทย์และทำนายนการ กลับเข้ามารักษาตัวของผู้ป่วยที่ผู้ป่วยสามารถทำได้ด้วยตนเอง ซึ่งจะช่วยให้ตัวของผู้ป่วยเองสามารถประหยัดเวลาในการที่ต้องมา พบแพทย์บ่อยๆ และสามารถส่งผลต่อการทำนายโอกาสการ กลับมาเป็นโรคเบาหวานของผู้ป่วยได้ด้วย และเพื่อนำโมเดลที่ได้ ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วน ของแนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

6. ผลการดำเนินงานวิจัย

งานวิจัยครั้งนี้ได้ทำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้ในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยทำการสร้างโมเดลจาก 3 เทคนิคของ Classification คือ เทคนิค Decision tree, เทคนิคNaïve Bays และเทคนิค K-Nearest Neighbors โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบ

ประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross-validation เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลค่า K= 10 กลุ่ม, k=15 กลุ่ม และ k= 20 กลุ่ม จากนั้นทำการหาค่าความถูกต้องของโมเดลที่ได้โดยประเมินจากค่าความถูกต้อง (Accuracy)

ตาราง 1. แสดงค่า Accuracyจากเทคนิค Decision tree

	K=10	K=15	K=20
ค่า Accuracy	85.18%	85.28%	85.50%

จากตารางที่ 1 เป็นการแสดงผลค่าความถูกต้องของโมเดลจากการใช้เทคนิค Decision tree โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross-validation เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลระหว่างข้อมูล Training ที่ใช้ในการทดลองและข้อมูลในส่วนของการทดสอบ โดยแบ่งข้อมูลค่า $K = 10$ กลุ่ม, $k = 15$ กลุ่ม และ $k = 20$ กลุ่ม ซึ่งพบว่า การแบ่งข้อมูล training โดยใช้ค่า $k = 20$ ให้ค่าความถูกต้องสูงที่สุด คิดเป็น ร้อยละ 85.50

ตาราง 2. แสดงค่า Accuracy จากเทคนิค Naïve Bays

	K=10	K=15	K=20
ค่า Accuracy	84.17%	84.64%	84.62%

จากตารางที่ 2 เป็นการแสดงผลค่าความถูกต้องของโมเดลจากการใช้เทคนิค Naive Bays โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross-validation เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลระหว่างข้อมูล Training ที่ใช้ในการทดลองและข้อมูลในส่วนของการโดยแบ่งข้อมูลค่า $K=10$ กลุ่ม, $k=15$ กลุ่ม และ $k=20$ กลุ่ม ซึ่งพบว่า การแบ่งข้อมูล training โดยใช้ค่า $k=20$ ให้ค่าความถูกต้องสูงที่สุด คิดเป็น ร้อยละ 84.62

ตาราง 3. แสดงค่า Accuracy จากเทคนิค KNN

	K=10	K=15	K=20
ค่า Accuracy	76.12%	76.13%	76.72%

จากตารางที่ 3 เป็นการแสดงผลค่าความถูกต้องของโมเดลจากการใช้เทคนิค K-Nearest Neighbors โดยใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบประสิทธิภาพโมเดลด้วยวิธีการแบ่งค่าข้อมูลแบบ Cross-validation เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลระหว่างข้อมูล Training ที่ใช้ในการทดลองและข้อมูลในส่วนของการโดยแบ่งข้อมูลค่า K= 10 กลุ่ม, k=15 กลุ่ม และ k= 20 กลุ่ม ซึ่งพบว่า การแบ่งข้อมูล training โดยใช้ค่า k=20

ให้ค่าความถูกต้องสูงสุด คิดเป็น ร้อยละ 76.72

ตาราง 4. แสดงการเปรียบเทียบค่า Accuracy จาก 3 เทคนิค

	DT	NB	KNN
ค่า Accuracy	85.50%	84.62%	76.72%

จากตารางที่ 4 เป็นการแสดงการเปรียบเทียบค่าความถูกต้องที่ได้จากการสร้างโมเดลในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล คือ เทคนิค Decision tree, เทคนิค Naïve Bays และเทคนิค K-Nearest Neighbors ซึ่งจากการดำเนินงานวิจัยนั้น ผลที่ได้คือการสร้างโมเดลในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานจากเทคนิค Decision tree โดยการแบ่งข้อมูลสำหรับการ Training และ testing เป็น 80:20 ให้ค่าความถูกต้องของโมเดลมากที่สุด คิดเป็นร้อยละ 85.50

7. สรุปผลการทดลอง

งานวิจัยนี้เป็นการสร้างโมเดลการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวาน โดยใช้กลุ่มข้อมูลตัวอย่างจากผู้ป่วยที่เป็นโรคเบาหวานในโรงพยาบาล ของสหรัฐอเมริกาตั้งแต่ปี พ.ศ. 2542-2551เป็นเวลา 10 ปี จำนวน 101,766 แถว โดยใช้ขั้นตอนวิธีต่างๆของการทำเหมืองข้อมูลมาเป็นเครื่องมือในการสกัดความรู้ที่ซ่อนอยู่ในข้อมูล เพื่อให้ได้องค์ความรู้ที่ต้องการ โดยมีปัจจัยนำเข้าใช้ในการวิเคราะห์โอกาสการกลับมารักษาซ้ำของผู้ป่วยโรคเบาหวานมีจำนวน มีจำนวน 14 แอตทริบิวต์ ได้แก่ รหัสประจำตัวผู้ป่วย (patient_nbr), เพศ (gender), อายุ (age), น้ำหนัก (weight), ประเภทของการเข้ารับการรักษา (admission_type_id), ประเภทของการรับการรักษาต่อ (discharge_disposition_id), ประวัติการรักษาของผู้ป่วย (admission_source_id), ระยะเวลาในการเข้ารับการรักษา (time_in_hospital), จำนวนครั้งในการเข้ารับการรักษาในห้องปฏิบัติการ (number of lab procedures), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยนอกในปีก่อนหน้า (Number of Outpatient visits), จำนวนครั้งของการเข้ารับการรักษาของผู้ป่วยในในปีก่อนหน้า (Number of Inpatient visits), จำนวนในการเข้ารับการรักษาแบบฉุกเฉินของผู้ป่วย (Number of Emergency visit), จำนวนยาที่ได้รับ (Number of medications), จำนวนขั้นตอนของการรักษา (Number of procedures) โอกาสการกลับมารักษาตัวของผู้ป่วย (readmitted)

ซึ่งทำการนำข้อมูลของผู้ป่วยที่เป็นโรคเบาหวานตามคุณลักษณะที่จัดเก็บมาสร้างโมเดลในการทำนายโดยใช้ เทคนิค Decision tree, Naïve Bays และ K-Nearest Neighbors ผ่าน

กระบวนการขั้นตอนการดำเนินงานวิจัยโดยการประยุกต์ใช้โดยกระบวนการทำเหมืองข้อมูลแบบ CRISP-DM มาเป็นแนวทางในการดำเนินงานวิจัย และใช้วิธีการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพโมเดลโดยใช้การแบ่งค่าข้อมูลแบบ Cross-validation เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล โดยแบ่งข้อมูลระหว่างข้อมูล Training ที่ใช้ในการทดลองและข้อมูลในส่วนของการโดยแบ่งข้อมูลค่า K= 10 กลุ่ม, k=15 กลุ่ม และ k=20 กลุ่ม และหาค่าความถูกต้องของโมเดลที่ได้โดยประเมินจากค่าความถูกต้อง (Accuracy)

ผลการดำเนินงานวิจัยครั้งนี้ โดยการสร้างโมเดลในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานโดยใช้เทคนิค Decision tree, Naïve Bays และ K-Nearest Neighbors ได้ผลการเปรียบเทียบประสิทธิภาพของโมเดลโดยดูจากค่าความถูกต้องสูงสุด ทำให้ได้โมเดลในการทำนายโอกาสการกลับมารักษาตัวซ้ำของผู้ป่วยโรคเบาหวานที่สร้างจากเทคนิค Decision tree จากการแบ่งข้อมูลสำหรับการ Training และ testing เป็น K=20 ให้ค่าความถูกต้องของโมเดลมากที่สุด คิดเป็นร้อยละ 85.50

ตัวอย่างการแปรผลข้อมูล เช่น

- ผู้ป่วยอายุ 0-50 เป็นเพศหญิง ช่วงน้ำหนัก 75 -100 เข้ารับการรักษาแบบฉุกเฉิน มีกระบวนการรักษา คือ ออกจากโรงพยาบาลและไปรักษาโรงพยาบาลใกล้บ้าน และเป็นผู้ป่วยที่เข้ารับการรักษาแบบฉุกเฉิน มีผลการทำนายคือ YES หมายถึง มีโอกาสกลับเข้ามารักษาตัวซ้ำ
- ผู้ป่วยอายุ 81-100 เป็นเพศชาย ช่วงน้ำหนัก 0 -50 เข้ารับการรักษาแบบฉุกเฉิน มีกระบวนการรักษา คือ รับการรักษาตัวที่บ้าน และมีประวัติการรักษาตัวแบบฉุกเฉิน มีผลทำนายออกมาคือ NO หมายถึง ไม่ต้องกลับมารักษาตัวซ้ำอีก

จากผลงานวิจัยสามารถนำความรู้ที่ได้จากการสร้างโมเดลไปเป็นแม่แบบในการพัฒนาระบบสนับสนุนการตัดสินใจเพื่อใช้ในการแพทย์สมัยใหม่ที่มีการนำระบบสารสนเทศเข้ามาช่วยในการสนับสนุนการตัดสินใจ เพื่อลดระยะเวลาในการทำการรักษาของแพทย์และทำนายการกลับเข้ามารักษาตัวของผู้ป่วยโรคเบาหวาน ซึ่งผู้ป่วยสามารถทำได้ด้วยตนเอง และจะทำให้ตัวของผู้ป่วยเองสามารถประหยัดเวลาในการที่ต้องมาพบแพทย์บ่อยๆ และสามารถส่งผลการทำนายโอกาสการกลับมาเป็นโรคเบาหวานของผู้ป่วยได้ด้วย และเพื่อนำโมเดลที่ได้ไปประยุกต์ใช้ในการสร้างระบบสนับสนุนการตัดสินใจในส่วน of แนวทางการรักษาของแพทย์และผู้ป่วยได้ต่อไป

8. ข้อเสนอแนะ

ทางผู้วิจัยได้ดำเนินงานวิจัยครั้งนี้ภายใต้ข้อจำกัดในเรื่องข้อมูล ซึ่งอาจจะส่งผลต่องานวิจัยในแง่ของผลที่เกิดขึ้น ทำให้งานวิจัย อาจจะมีข้อบกพร่องอยู่บ้าง ดังนั้นผู้วิจัยจึงได้เสนอแนะแนวทาง ในการปรับปรุงงานวิจัยในอนาคตต่อไป

8.1 การศึกษางานวิจัยนี้เป็นการนำข้อมูลชุดตัวอย่างของผู้ป่วยที่เป็นโรคเบาหวาน ซึ่งเป็นชุดข้อมูลที่สามารถดาวน์โหลดมาได้จาก แหล่งข้อมูล ทำให้ข้อมูลที่ได้อาจจะไม่ใช่ข้อมูลของผู้ป่วยที่เก็บ ข้อมูลมาจากโรงพยาบาลต่างๆที่อยู่ในประเทศไทย ซึ่งถ้ามีการ พัฒนาต่อยอดต่อไปจากงานวิจัยนี้ ก็สามารถที่จะนำแนวทางที่ได้ ไปประยุกต์ใช้กับข้อมูลของผู้ป่วยที่เป็นโรคเบาหวานในแต่ละ โรงพยาบาลของประเทศไทยได้ และสามารถนำความรู้ที่ได้ไป ประยุกต์ใช้เป็นแม่แบบในการพัฒนาระบบสนับสนุนการตัดสินใจ ทางการแพทย์ที่เกี่ยวกับโรคเบาหวานได้

8.2 การทำเหมืองข้อมูลเป็นเครื่องมือที่ช่วยในการวิเคราะห์หรือ ทำนายโอกาสความน่าจะเป็นที่จะเกิดเหตุการณ์นั้นๆ แต่บางครั้ง ไม่สามารถที่จะเข้าใจธุรกิจหรือปัจจัยแวดล้อมภายนอกต่างๆได้ เท่ากับตัวบุคคล ดังนั้นผู้ใช้งานต้องนำไปวิเคราะห์และตัดสินใจ ร่วมกันระหว่างบุคคลกับระบบด้วยอีกครั้ง

8.3 ในการสร้างโมเดล สิ่งสำคัญคือการเลือกปัจจัยที่เกี่ยวข้อง หรือส่งผลต่อค่าคำตอบ เพราะฉะนั้นจะต้องมีการวิเคราะห์ค่า ทางสถิติเพื่อให้เห็นความสัมพันธ์ระหว่างตัวแปรที่เกิดขึ้นได้อย่าง ชัดเจน

เอกสารอ้างอิง

- [1] ปพน แก้วหิน, กาญจนา เมียกขุนทด และเบญจกัศ จงหมื่น ไวย “ระบบสนับสนุนการตัดสินใจเพื่อวิเคราะห์ข้อมูลผู้ป่วย โรคเบาหวานที่มีความเสี่ยงของการเกิดภาวะแทรกซ้อนที่ เท้า.” มหาวิทยาลัยราชภัฏนครราชสีมา, 2558.
- [2] ขวัญกรณ์ ทราชมอ, วิภาเจริญภัณฑารักษ์ และ วิทยา พร พชรพงศ์, “การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อน ของโรคอื่นในผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมือง ข้อมูลโดยใช้ประวัติผลการตรวจรักษาผู้ป่วยเบาหวาน กรณีสึกษาโรงพยาบาลปทุมธานี.” มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2557.
- [3] พยูน พาณิษฐ์กุล .(2548). การพัฒนาระบบค้ำไม้หนึ่ง โดยใช้ Decision Tree. โครงการพัฒนา ระบบงานปัญญา วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนง วิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหาร ลาดกระบัง.
- [4] เรวดี ศักดิ์ ดุลยธรรม, “การใช้เทคนิคค้ำไม้หนึ่งในการ วิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษา โรคนิ้วล็อก

ในแบบต่างๆ ของคณะแพทยศาสตร์วชิรพยาบาล มหาวิทยาลัยกรุงเทพมหานคร.”, วิทยาลัยราชพฤกษ์ , 2553.

[5] รายงานสถิติสาธารณสุข สำนักนโยบายและกลยุทธ์กระทรวง สาธารณสุข ปีพ .ศ.2558. จำนวนและอัตราผู้เสียชีวิตจาก โรคเบาหวาน, 2558.

[6] อุดลย์ ยิ้มงาม. การทำเหมืองข้อมูล Data Mining[ออนไลน์]. เข้าถึงได้จาก :

http://compcenter.bu.ac.th/index.php?option=com_content&task=view&id=75&Itemid=172. วันที่ค้นข้อมูล): 25 มีนาคม 2560(

[7] Eakasit Pacharawongsakda, Ph.D. An Introduction to Data Mining Techniques Thai Version. พิมพ์ครั้งที่ 2. กรุงเทพฯ : สำนักพิมพ์เอเชีย ดิจิตอลการพิมพ์ จำกัด, 2557.

[8] Eakasit Pacharawongsakda, Ph.D. Prectical Data Mining with RapidMiner Studio 6. พิมพ์ครั้งที่ 34. กรุงเทพฯ : สำนักพิมพ์เอเชีย ดิจิตอลการพิมพ์ จำกัด, 2558.

[9] Chien-Lung Chan (2008), Investigation of diabetic microvascular complications using data mining techniques, Department of Information Management , Yuan Ze Univ., Chungli , Yu- Chen Liu Shih-Hui Luo

[10] Thomas Bayes And Richard Price, An Essay towards solving a ploblem in the Doctrine of chance. By the late Rev.Mr. Bays, communicated by Mr.Price, in a letter to john Canton, A.M.R.S., Philosophical Transactions of the Royal Society of London, 53(0):, 1763

[11] T.cover and P. Hart, Nearest neighbor pattern classification, Information Theory, IEEE Transaction on, 13 (1), 1967