

การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวาน
โดยใช้เทคนิคการทำเหมืองข้อมูล

นายชฎิภรณ์ ทรายหมอ



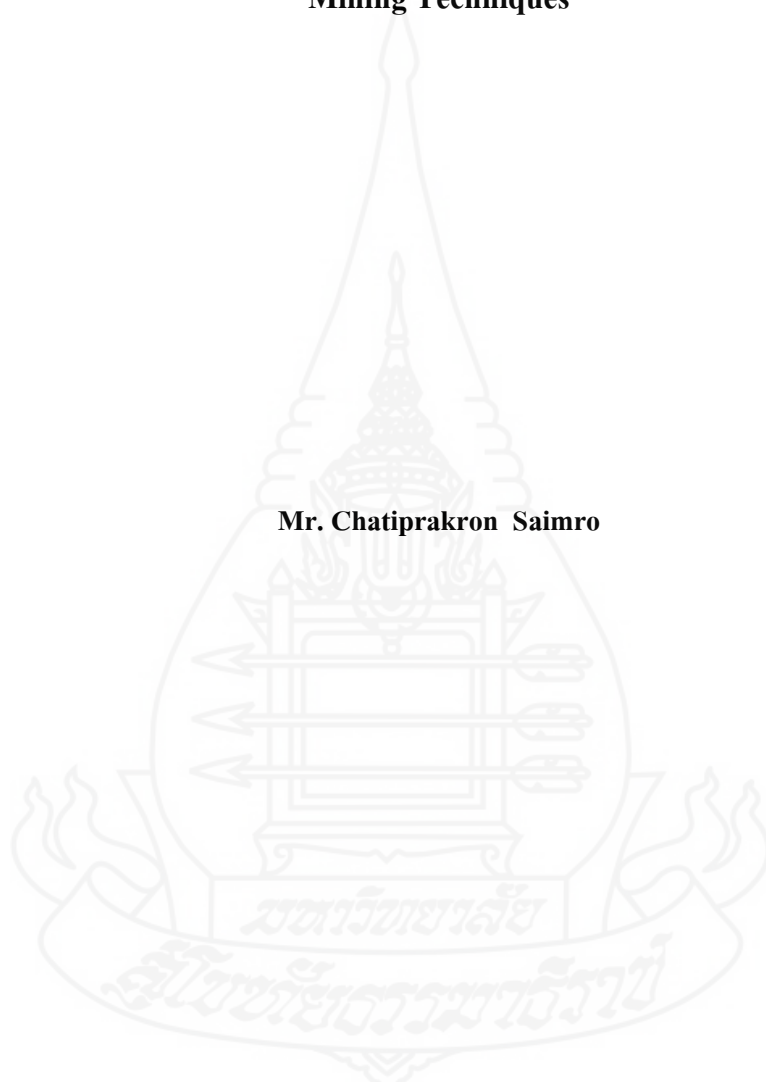
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมชิราช

พ.ศ. 2556

Complications of Diabetic Patients Using Data

Mining Techniques

Mr. Chatiprakron Saimro



A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Information and Communication Technology

School of Science and Technology
Sukhothai Thammathirat Open University

2013

หัวข้อวิทยานิพนธ์ การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล

ชื่อและนามสกุล นายชฎิกกรณ์ ทรายหมอ

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

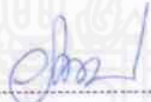
สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ดร. วิภา เจริญกัณธารักษ์
2. อาจารย์วิทยา พรห้ชรพงศ์


วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 31 กรกฎาคม 2557

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(อาจารย์ ดร. ดวงดาว วิชากุล)


..... กรรมการ
(รองศาสตราจารย์ ดร. วิภา เจริญกัณธารักษ์)


..... กรรมการ
(อาจารย์วิทยา พรห้ชรพงศ์)


..... ประธานกรรมการบัณฑิตศึกษา
(ศาสตราจารย์ ดร. เทีรวรรณ ศรีพหล)

ชื่อวิทยานิพนธ์ การพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล

ผู้วิจัย นายชฎิกกรณ์ ทราชหมอ รหัสประจำตัว 2549600142

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)

อาจารย์ที่ปรึกษา (1) รองศาสตราจารย์ ดร. วิภา เจริญกัญชาวิทย์ (2) อาจารย์วิทยา พรพิชรพงศ์ ปีการศึกษา 2556

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน 2) นำแบบจำลองที่ได้ไปทดลองใช้กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวานของโรงพยาบาลปทุมธานี

วิธีการดำเนินการวิจัย 1) ทำความเข้าใจปัญหาที่ดำเนินงาน การตรวจรักษาของแพทย์ต่อผู้ป่วย และโครงสร้างข้อมูลของการจัดเก็บประวัติการรักษา เพื่อสร้างแบบจำลองให้ตรงตามความเป็นจริง 2) จัดเตรียมข้อมูลให้มีความถูกต้อง โดยทำความเข้าใจปัญหาที่ดำเนินงาน ทั้งนี้ข้อมูลที่ใช้ในการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน จะนำมาจากข้อมูลผลตรวจการวินิจฉัยโรคเบาหวานตั้งแต่ที่ผู้ป่วยเป็นโรคเบาหวาน ของโรงพยาบาลปทุมธานี 3) สร้างแบบจำลองตัวแปรที่สำคัญได้แก่ จำนวนครั้งการมาตรวจรักษาด้วยแพทย์ จำนวนโรคที่ทำการลงทะเบียนผู้ป่วยเบาหวานเท่านั้น เครื่องมือที่ใช้ในการวิจัยประกอบด้วย โปรแกรม WEKA เพื่อพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยเบาหวาน MicroSoft Visual Foxpro เพื่อจัดเก็บข้อมูลผลการตรวจวินิจฉัยโรค Microsoft Excel 2010 สำหรับสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยเบาหวาน 4) เปรียบเทียบผลการพยากรณ์โดยวิธีการหาความสัมพันธ์ของข้อมูลระหว่าง Apriori Algorithm และ FPGrowth Algorithm 5) ทดสอบกับผลการพยากรณ์สถานการณ์จริง นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ทดลองใช้ตั้งแต่ 1 เมษายน พ.ศ. 2557 ถึง 15 กรกฎาคม พ.ศ. 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

ผลการพยากรณ์ความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี พบว่าผู้ป่วยที่เป็นโรค 1) เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E103) แล้วจะเป็นโรคแทรกซ้อนดังนี้ 1.1) โรคจอตาจากโรคเบาหวานชนิดไม่ออกซายหรือเอ็นพีดีอาร์ (รหัสโรค H3600) คิดเป็น 65 % 1.2) โรคต่อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 46 % 1.3) โรคจอตาจากโรคเบาหวานชนิดออกซายหรือพีดีอาร์ (รหัสโรค H3602) คิดเป็น 44 % 2) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E113) แล้วจะเป็นโรคแทรกซ้อน 2.1) โรคจอตาจากโรคเบาหวานชนิดไม่ออกซาย (รหัสโรค H3600) คิดเป็น 77 % 2.2) โรคต่อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 49 % 2.3) โรคจอตาจากโรคเบาหวานชนิดออกซาย (รหัสโรค H3602) คิดเป็น 36 % 2.4) โรคไตวายเรื้อรังที่มีได้ระดับระยะเฉียบ (รหัสโรค N189) คิดเป็น 20 % 3) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท (รหัสโรค E114) แล้วจะเป็นโรคแทรกซ้อน 3.1) โรคต่อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 24 % 3.2) โรคหัวใจขาดเลือดเรื้อรังที่มีได้ ระบุระยะเฉียบ (รหัสโรค I251) คิดเป็น 20 % ข้อเสนอแนะงานวิจัยควรรวบรวมข้อมูลประวัติการรักษาด้านอื่น ข้อมูลประชากรพื้นที่อื่นไม่เฉพาะเจาะจง

จึงมาทำการวิเคราะห์ข้อมูลด้วย ควรเลือกใช้เทคนิคการสร้างเหมืองข้อมูลรูปแบบอื่น ในการค้นหาความสัมพันธ์ของข้อมูล เพื่อเพิ่มประสิทธิภาพในการวัดความถูกต้องของการพยากรณ์ต่อไป

คำสำคัญ แบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น โรคเบาหวาน เหมืองข้อมูล กฎความสัมพันธ์

อัลกอริทึม Apriori อัลกอริทึม FPGrowth

Thesis title: Complications of Diabetic Patients Using Data Mining Techniques
Researcher: Mr. Chatiprakron Saimro; **ID:** 2549600142;
Degree: Master of Science (Information and Communication Technology);
Thesis advisors: (1) Vipa Jaroenpuntaruk, Associate Professor;
 (2) Wittaya Pornpatcharapong; **Academic year:** 2013

Abstract

The research aimed at 1) developing predictive models for complications of diseases in diabetic patients, and 2) applying the model for field trial Diababetes Clinic, Medicine Department, Pathum Thani Hospital.

This research was conducted in 5 main steps as follows: 1) Study problems of working procedure, doctors' diagnoses, and data structure of patient profile in order to obtain the accurate model; 2) Prepare complete data for further process. The trained data employed in the model development was from the diagnoses' results of diabetic patients in Pathum Thani Hospital. 3) Generate the model using the important variables in the research such as the number of times that the patients visited the doctors and the number of diseases that were recorded in the patients' profiles. 4) Compare the predicted models by using association between Apriori Algorithm and FPGrowth Algorithm. The instruments used were consisted of a) WEKA program for data analysis of complications of diseases in diabetic patients, b) Microsoft Visual Foxpro for recording diabetic diagnosis results, and c) Microsoft Execl 2010 for creating the model for complications of diseases in diabetic patients. 5) The predictive models were tested in real situations. The information system was developed to present the prediction for complications of diseases in diabetic patients at Medication Unit, Diabetic Patients Department of Pathum Thani Hospital from April 1, 2014 to July 15, 2014.

The findings revealed that the patients with insulin-dependent diabetes mellitus type 1 at with ophthalmic complications (E103) were 1.1) 65% likely to have non-proliferative diabetic retinopathy, NPDR, (H3600), 1.2) 46% likely to have senile nuclear cataract (H251), and 1.3) 44% likely to have proliferative diabetic retinopathy, PDR, (H3602). The research also found that the patients with non-insulin-dependent diabetes mellitus type 2 at with ophthalmic complications (E113) were 2.1) 77% likely to have non-proliferative diabetic retinopathy (H3600) 2.2) 49% likely to have senile nuclear cataract (H251) 2.3) 36% likely to have proliferative diabetic retinopathy (H3602), and 2.4) 20% likely to have chronic renal failure, unspecified (N189). Moreover, the findings showed that the patients with non-insulin-dependent diabetes mellitus type 3 with neurological complications (E114) were 3.1) 24% likely to have senile nuclear cataract (H251), and 3.2) 20% likely to have chronic ischemic heart disease (I251) unspecified. Recommendation for further research should cover into other medical history information. Data for the additional models should be randomized with nonspecific selected demographic areas. Moreover, new algorithm may be applied to discover unexpected models.

Keywords: Predictive Model Complication of Diseases, Diabetic, Data Mining, Association Rule, Apriori Algorithm, FPGrowth Algorithm

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยได้รับความร่วมอนุเคราะห์จากท่านผู้ทรงคุณวุฒิหลายท่าน โดยเฉพาะอาจารย์ รศ.ดร.วิภาเจริญภัณฑารักษ์ อาจารย์ที่ปรึกษาหลัก และอาจารย์วิทยาพรพัชรพงศ์ อาจารย์ที่ปรึกษาร่วมของวิทยานิพนธ์ฉบับนี้ ซึ่งได้สละเวลาในการให้คำปรึกษาแนวทางการตลอดจนแนะนำเกี่ยวกับข้อมูลและเนื้อหาที่จะนำมาปรับปรุงแก้ไขข้อบกพร่องต่างๆ เพื่อให้เป็นวิทยานิพนธ์ที่สมบูรณ์ฉบับหนึ่ง ผู้วิจัยกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ นายแพทย์ทรงพล ชวาลตันพิพัทธ์ผู้อำนวยการโรงพยาบาลปทุมธานี และนายแพทย์มณฑิเตอร์ เพ็งสมบัติรองผู้อำนวยการกลุ่มภารกิจด้านพัฒนาระบบบริการและสนับสนุนบริการสุขภาพโรงพยาบาลปทุมธานีและทันตแพทย์คนุยศ ศรีศัมภวงค์ หัวหน้ากลุ่มงานสารสนเทศ-เวชระเบียน / ศูนย์คอมพิวเตอร์ โรงพยาบาลปทุมธานีและนางปีตมา วงศ์นคร พยาบาลวิชาชีพชำนาญการ หัวหน้ากลุ่มงานการพยาบาลผู้ป่วยนอก/ศูนย์เปล โรงพยาบาลปทุมธานี ที่ให้ความอนุเคราะห์ข้อมูลและสถานที่เพื่อนำระบบเข้ามาทำการทดสอบการใช้งานและขอขอบคุณเจ้าหน้าที่ศูนย์สารสนเทศ โรงพยาบาลปทุมธานี ที่อำนวยความสะดวกในการให้ข้อมูลผลการรักษาผู้ป่วยเบาหวานตลอดจนเอกสารที่เกี่ยวข้องรวมทั้งให้ความช่วยเหลือแนะนำในส่วนต่างๆของระบบได้เป็นอย่างดีและขอขอบคุณกลุ่มงานห้องตรวจอายุรกรรม โรงพยาบาลปทุมธานี ที่เสียสละเวลาให้ความร่วมมือในการให้ข้อควรปฏิบัติ/คำแนะนำ ของผู้ป่วยเบาหวาน และทดสอบการทำงาน ของระบบรวมทั้งผู้ที่มีส่วนร่วมทุกท่านที่ได้เอื้อนามไว้ ณ ที่นี้

นอกจากนี้ขอขอบคุณคุณอาจารย์สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อนักศึกษาและผู้ที่มีส่วนเกี่ยวข้องในการทำวิทยานิพนธ์ครั้งนี้ทุกท่านที่ได้กรุณาให้การสนับสนุนช่วยเหลือและให้กำลังใจตลอดมาจนกระทั่งการทำวิจัยวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ด้วยดี

ชฎิกกรณ์ ทราชมอ

กรกฎาคม 2557

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ญ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์การวิจัย.....	2
ขอบเขตของการวิจัย.....	3
วิธีการดำเนินการวิจัย.....	3
กรอบแนวคิดการวิจัย.....	4
นิยามศัพท์เฉพาะ.....	5
ประโยชน์ที่คาดว่าจะได้รับ.....	6
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง.....	7
ความรู้พื้นฐาน โรคเบาหวานและโรคแทรกซ้อน (Diabetes Mellitus& Complications).....	7
ทฤษฎีคลังข้อมูล (Data warehouse).....	9
ทฤษฎีเหมืองข้อมูล (Data Mininig).....	16
หลักการของคริปส์-ดีเอ็ม.....	42
โปรแกรมเวก้า (WEKA).....	47
งานวิจัยที่เกี่ยวข้อง.....	49
บทที่ 3 วิธีดำเนินการวิจัย.....	54
ประชากรและกลุ่มตัวอย่าง.....	54
เครื่องมือที่ใช้ในการดำเนินงานวิจัย.....	55
ขั้นตอนการดำเนินงาน.....	55

สารบัญ (ต่อ)

	หน้า
บทที่ 4 ผลการวิจัย.....	76
ผลการวิเคราะห์ข้อมูล.....	76
การพัฒนาระบบโปรแกรมนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้อง จากการเป็นโรคเบาหวานเพื่อนำแบบจำลองที่ได้ไปทดลองใช้.....	83
บทที่ 5 สรุปการวิจัย.....	89
สรุปการวิจัย.....	89
อภิปรายผล.....	91
ข้อเสนอแนะ.....	92
บรรณานุกรม.....	94
ภาคผนวก.....	98
ก ตัวอย่างการใช้โปรแกรมเวก้า.....	99
ข เอกสาร อนุญาตให้เก็บข้อมูลเพื่อการศึกษาวิจัย.....	108
ค เอกสารรับรองทดลองใช้ “ระบบนำเสนอผลการพยากรณ์โอกาส การเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน”.....	111
ประวัติผู้วิจัย.....	113



สารบัญตาราง

	หน้า
ตารางที่ 2.1 ตัวอย่างข้อมูลรายการการซื้อสินค้า.....	28
ตารางที่ 2.2 กลุ่มข้อมูลที่เกิดขึ้นบ่อยที่ได้จากฐานข้อมูลตัวอย่าง.....	28
ตารางที่ 2.3 ตัวอย่างกฎความสัมพันธ์ที่สร้างได้จากกลุ่มข้อมูลที่เกิดขึ้นบ่อย.....	29
ตารางที่ 2.4 ฐานข้อมูลการซื้อสินค้า.....	37
ตารางที่ 2.5 ค่า Conditional pattern base และค่า Conditional FP-Tree.....	41
ตารางที่ 2.6 ผลลัพธ์ของการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยของขั้นตอนวิธี FP- Growth.....	42
ตารางที่ 3.1 โครงสร้างข้อมูลเพิ่มประวัติผู้ป่วย.....	56
ตารางที่ 3.2 โครงสร้างข้อมูลเพิ่มผลการตรวจรักษาวินิจฉัยโรคของผู้ป่วยเบาหวานผู้ป่วยนอก.....	57
ตารางที่ 3.3 โครงสร้างข้อมูลเพิ่มผลการตรวจรักษาวินิจฉัยโรคของผู้ป่วยเบาหวานผู้ป่วยใน.....	57
ตารางที่ 3.4 โครงสร้างข้อมูล รหัสโรค.....	58
ตารางที่ 3.5 โครงสร้างข้อมูล รหัสกลุ่มโรค.....	58
ตารางที่ 3.6 โครงสร้างฐานเก็บผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน.....	62
ตารางที่ 4.1 รหัสโรคเบาหวานทั้งหมด.....	78
ตารางที่ 4.2 ผลการตรวจวินิจฉัยโรคเบาหวาน ในผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี.....	80
ตารางที่ 4.3 โรคที่ผู้ป่วยมีโอกาสที่จะเป็นหลังจากเป็น โรคเบาหวาน.....	83
ตารางที่ 4.4 โครงสร้างข้อมูลรหัสโรค.....	85
ตารางที่ 4.5 โครงสร้างข้อมูลเพิ่มประวัติผู้ป่วย.....	86
ตารางที่ 4.6 โครงสร้างข้อมูลรหัสกลุ่มโรค.....	86

สารบัญภาพ

	หน้า
ภาพที่ 1.1 กรอบแนวคิดการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากที่ผู้ป่วยเป็นโรคเบาหวาน.....	4
ภาพที่ 2.1 วิวัฒนาการของสารสนเทศในคลังข้อมูล.....	10
ภาพที่ 2.2 กระบวนการในการพัฒนาคลังข้อมูล.....	11
ภาพที่ 2.3 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล.....	17
ภาพที่ 2.4 สถาปัตยกรรมระบบเหมืองข้อมูล.....	18
ภาพที่ 2.5 การวิเคราะห์พฤติกรรมการซื้อขาย.....	21
ภาพที่ 2.6 แสดงตัวอย่างการค้นหาความสัมพันธ์.....	22
ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภท.....	24
ภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่ม.....	25
ภาพที่ 2.9 กระบวนการค้นหาความสัมพันธ์.....	26
ภาพที่ 2.10 ขั้นตอนการทำงานของ Apriori Algorithm.....	29
ภาพที่ 2.11 การอ่านข้อมูลจากฐานข้อมูลเพื่อค้นหา L_1 ของ Apriori Algorithm.....	30
ภาพที่ 2.12 การสร้างกลุ่มข้อมูลที่ซ้ำ C_2 ของ Apriori Algorithm.....	31
ภาพที่ 2.13 การนับค่าความถี่ของข้อมูลที่ซ้ำ C_2 ของ Apriori Algorithm.....	31
ภาพที่ 2.14 ผลลัพธ์การค้นหาข้อมูล L_2 ของ Apriori Algorithm.....	32
ภาพที่ 2.15 การสร้างกลุ่มข้อมูลที่ซ้ำ C_3 ของ Apriori Algorithm.....	32
ภาพที่ 2.16 การนับค่าความถี่ของข้อมูลที่ซ้ำ C_3 ของ Apriori Algorithm.....	33
ภาพที่ 2.17 ผลลัพธ์การค้นหาข้อมูล L_3 ของ Apriori Algorithm.....	33
ภาพที่ 2.18 การสร้างกลุ่มข้อมูลที่ซ้ำ C_4 ของ Apriori Algorithm.....	34
ภาพที่ 2.19 การนับค่าความถี่ของข้อมูลที่ซ้ำ C_4 ของ Apriori Algorithm.....	34
ภาพที่ 2.20 ขั้นตอนการทำงานของขั้นตอนวิธี FP-Growth.....	36
ภาพที่ 2.21 การอ่านข้อมูลจากฐานข้อมูลเพื่อค้นหา L_1 ของขั้นตอนวิธี FP-Growth.....	37
ภาพที่ 2.22 การสร้างตาราง Header ของขั้นตอนวิธี FP-Growth.....	38
ภาพที่ 2.23 การอ่านรายการข้อมูลแรกจากฐานข้อมูลของขั้นตอนวิธี FP-Growth.....	39
ภาพที่ 2.24 การอ่านรายการข้อมูลที่สองจากฐานข้อมูลของขั้นตอนวิธี FP-Growth.....	39
ภาพที่ 2.25 การอ่านรายการข้อมูลที่สามจากฐานข้อมูลของขั้นตอนวิธี FP-Growth.....	40

สารบัญภาพ (ต่อ)

	หน้า
ภาพที่ 2.26 การอ่านรายการข้อมูลที่สืบจากฐานข้อมูลของขั้นตอนวิธี FP-Growth.....	40
ภาพที่ 2.27 การอ่านรายการข้อมูลที่สืบจากฐานข้อมูลของขั้นตอนวิธี FP-Growth.....	40
ภาพที่ 2.28 การสร้างตาราง Header ของขั้นตอนวิธี FP-Growth.....	41
ภาพที่ 2.29 การแบ่งระดับงานของคริสป์-ดีเอ็ม.....	43
ภาพที่ 2.30 แบบจำลอง CRISP-DM.....	45
ภาพที่ 2.31 โลโก้ของซอฟต์แวร์ WEKA เป็นรูปนกเฉพาะถิ่นของประเทศนิวซีแลนด์.....	47
ภาพที่ 2.32 หน้าการหลักของซอฟต์แวร์ WEKA.....	48
ภาพที่ 3.1 ข้อมูลทะเบียนประวัติผู้ป่วย.....	59
ภาพที่ 3.2 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก.....	59
ภาพที่ 3.3 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน.....	60
ภาพที่ 3.4 ข้อมูลรหัสโรค ICD10.....	61
ภาพที่ 3.5 ข้อมูลกลุ่มโรค.....	61
ภาพที่ 3.6 การนำข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน.....	62
ภาพที่ 3.7 ซอร์ซโค้ด (Source Code) การนำข้อมูลผลตรวจผู้ป่วยเบาหวานผู้ป่วยนอก.....	63
ภาพที่ 3.8 ซอร์ซโค้ด (Source Code) การนำข้อมูลผลตรวจผู้ป่วยเบาหวานผู้ป่วยใน.....	63
ภาพที่ 3.9 ผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานทั้งหมด (Diabetes_Opd_DX).....	64
ภาพที่ 3.10 ข้อมูลออกมาเป็นรูปแบบ File Microsoft Execl 2010.....	65
ภาพที่ 3.11 การนำเข้าข้อมูลเข้าโปรแกรม WEKA เพื่อทำการวิเคราะห์ข้อมูล.....	66
ภาพที่ 3.12 การกำหนดค่า Min Support, Confidence , num Rules.....	67
ภาพที่ 3.13 ผลการพยากรณ์ด้วย Apriori Algorithm.....	68
ภาพที่ 3.14 ผลการพยากรณ์ด้วย FPGrowth Algorithm.....	68
ภาพที่ 3.15 การบันทึกส่งข้อมูลออกในรูปแบบ ไฟล์ TEXT FILE.....	69
ภาพที่ 3.16 ข้อมูลในรูปแบบ ไฟล์ TEXT FILE.....	69
ภาพที่ 3.17 หน้าจอหลักระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน.....	71

สารบัญภาพ (ต่อ)

	หน้า
ภาพที่ 3.18 หน้าจอรระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจาก การเป็นโรคเบาหวาน.....	73
ภาพที่ 3.19 พิมพ์ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็น โรคเบาหวาน.....	74
ภาพที่ 3.20 พิมพ์ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็น โรคเบาหวาน.....	75
ภาพที่ 4.1 ข้อมูลผลการวินิจฉัยโรคของผู้ป่วยเบาหวาน.....	77
ภาพที่ 4.2 กำหนดค่าสนับสนุน กำหนดค่าความเชื่อถือ.....	81
ภาพที่ 4.3 การบันทึกข้อมูลออกในรูปแบบ ไฟล์ Text File.....	84
ภาพที่ 4.4 ข้อมูลในรูปแบบ ไฟล์ Text File.....	85
ภาพที่ 4.5 ตารางเก็บผลการวิเคราะห์ข้อมูลรูปแบบฐานข้อมูล.....	88



บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

โรคเบาหวาน (Diabetes Mellitus: DM) หมายถึง ภาวะที่ร่างกายมีระดับน้ำตาลในเลือดสูงกว่าปกติ (Hyperglycemia) โดยตรวจพบระดับน้ำตาลในเลือดก่อนอาหารเช้าสูงกว่า 140 มิลลิกรัมต่อเลือด 100 มิลลิกรัม อย่างน้อย 2 ครั้ง หรือการตรวจพบระดับน้ำตาลในเลือดเวลาใดก็ตามสูงกว่า 200 มิลลิกรัมต่อเลือด 100 มิลลิกรัม ร่วมกับมีอาการของโรคเบาหวาน

เป็นโรคที่พบได้ในผู้ใหญ่ คือประมาณ 2.5-6 % ของประชากร (ถ้าคิดจากคนไทย 60 ล้านคน ก็จะมีคนเป็นโรคเบาหวานประมาณ 3 ล้านคน) โรคเบาหวานเป็นสาเหตุสำคัญอันดับแรกที่ทำให้เกิดโรคไตวายเรื้อรัง นอกจากนั้นยังเป็นปัจจัยสำคัญ ทำให้เกิดโรคของหลอดเลือดตีบตันได้ในทุกส่วนของร่างกาย เช่น โรคอัมพฤกษ์อัมพาตจากเส้นโลหิตสมองตีบ โรคกล้ามเนื้อหัวใจขาดเลือดจากเส้นเลือดหัวใจตีบ โรคเส้นเลือดที่ไปเลี้ยงปลายมือปลายเท้าอุดตัน และยังเป็นสาเหตุของตาบอดจากทั้งเบาหวานขึ้นตา และต้อกระจกได้อีกด้วย

โรคเบาหวานไม่ใช่โรคที่เกิดขึ้นรวดเร็วอาการอาจไม่ปรากฏเป็นเดือนหรือเป็นปีทั้งๆที่น้ำตาลในเลือดสูงและมีน้ำตาล ในปัสสาวะ หรือมีความผิดปกติเพียงเล็กน้อย เช่น ปัสสาวะบ่อยหรือหิวบ่อย แต่ไม่ได้นึกถึงโรคเบาหวาน จึงพบว่ากว่าจะทราบก็ต่อเมื่อผู้ป่วยมีอาการมากแล้ว หรือตรวจพบโดยบังเอิญเมื่อเข้าโรงพยาบาลด้วยโรคอื่น หรือการตรวจสุขภาพประจำปี อาการที่สำคัญของโรคเบาหวาน ปัสสาวะบ่อยและมาก กระหายน้ำ หิวบ่อย กินจุ แต่ผอมลง อ่อนเพลีย น้ำหนักลดลงมาก เป็นแผลฝี ได้ง่าย แต่รักษาหายยาก ค้นตามผิวหนัง โดยเฉพาะบริเวณอวัยวะสืบพันธุ์ ช่องคลอด ปวดเจ็บตามกล้ามเนื้อ ชาตามมือ เท้า หมดความรู้สึกทางเพศ ตาพร่ามัว ต้องเปลี่ยนแว่นตาบ่อยๆ คลอดบุตรที่มีน้ำหนักเกิน 4 กิโลกรัม โรงพยาบาลปทุมธานี เปิดให้บริการประชาชนครั้งแรก เมื่อปี พ.ศ. 2496 จนถึงปัจจุบัน มีเตียงรองรับผู้ป่วย 377 เตียง วิทยาลัยนักรังสีวิทยาโรงพยาบาลปทุมธานี เป็นโรงพยาบาลตติยภูมิที่มีคุณภาพอย่างยั่งยืน เป้าประสงค์ ประชาชนได้รับบริการสุขภาพที่มีคุณภาพ พันธกิจให้บริการสุขภาพระดับตติยภูมิ แบบองค์รวมอย่างมีคุณภาพ พัฒนาให้เป็นองค์กรแห่งการเรียนรู้ ทำการวิจัยด้านสาธารณสุข ร่วมผลิตและฝึกอบรมบุคลากรสาธารณสุข สร้างและสนับสนุนเครือข่ายบริการสุขภาพทุกระดับ โดยประชาชนและทุกภาคส่วนมีส่วนร่วมในการดูแล

สุขภาพ โรงพยาบาลปทุมธานีไม่มีเขตรับผิดชอบโดยตรง จำนวนผู้ป่วยที่ขึ้นทะเบียนประวัติการรักษาประมาณ 328,812 คน การรักษาผู้ป่วยเบาหวานที่มารับบริการตรวจรักษาที่โรงพยาบาลปทุมธานีพบว่าผู้ป่วยเบาหวานไม่สามารถควบคุมปริมาณน้ำตาลในเลือดได้ สาเหตุมาจากการไม่ออกกำลังกาย ควบคุมปริมาณอาหาร ไม่รับประทานยาตามแพทย์สั่ง การปรับลดยาทานเอง ไม่มาพบแพทย์ตามนัด มีภาวะเครียด การขาดความรู้ของผู้ดูแลผู้ป่วยเบาหวานซึ่งมักจะพบมากในกลุ่มผู้สูงอายุหรือกลุ่มผู้ที่ทำงานเป็นต้น ผู้ป่วยเบาหวานจะมีภาวะแทรกซ้อนของโรคอื่น ดังนั้นผู้วิจัยจึงได้ศึกษาย้อนหลังจากเอกสารที่เกี่ยวข้อง เกี่ยวกับผลการวินิจฉัยโรคในผู้ป่วยเบาหวาน ที่มารับบริการที่กลุ่มงานอายุรกรรมห้องตรวจเบาหวาน โรงพยาบาลปทุมธานี สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานแล้ว โดยใช้ข้อมูลผลการวินิจฉัยโรคเบาหวานของผู้ป่วยเบาหวานเท่านั้น ปัจจุบันโรงพยาบาลปทุมธานีมีระบบสารสนเทศในการให้บริการและจัดเก็บข้อมูล ระบบโปรแกรมที่ใช้อยู่ชื่อ Ez-Hosp พัฒนาโดยแอปพลิเคชัน MicroSoft Visual Foxpro เริ่มใช้ตั้งแต่ปี 2547 จนถึงปัจจุบัน ซึ่งมีการจัดเก็บข้อมูลทะเบียนประวัติผู้ป่วย ผลการตรวจวินิจฉัยโรคของผู้ป่วยที่มารับบริการ จัดเก็บในรูปแบบฐานข้อมูล ซึ่งข้อมูลที่จัดเก็บดังกล่าวสามารถนำมาใช้ประโยชน์โดยการสร้างเป็นคลังข้อมูล และใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) หาคความสัมพันธ์ (Association) มาวิเคราะห์หาข้อมูลภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานแล้ว จากแบบจำลองสามารถนำมาทดลองใช้โดยให้ความรู้กับตัวผู้ป่วยเบาหวาน การให้คำแนะนำของแพทย์ เป็นแนวทางให้เจ้าหน้าที่/แพทย์ ในการกำหนดมาตรการป้องกันเฝ้าระวังแก้ไข เพื่อลดอัตราการเกิดภาวะแทรกซ้อนของโรคอื่น และสามารถนำไปใช้ประโยชน์ต่อไปได้

2. วัตถุประสงค์การวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อ

2.1 สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน ด้วยใช้ข้อมูลผลการตรวจวินิจฉัยโรค โรงพยาบาลปทุมธานี โดยใช้เทคนิคการทำเหมืองข้อมูล

2.2 นำแบบจำลองที่ได้ไปทดลองใช้ กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

3. ขอบเขตของการวิจัย

สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวานแล้ว ใช้ประวัติผลการตรวจรักษาผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี ตั้งแต่วันที่ 1 มกราคม 2553 ถึงวันที่ 31 ธันวาคม 2556 จำนวน 17,043 คน โดยใช้ข้อมูลผลการตรวจวินิจฉัยโรคผู้ป่วยเบาหวาน ตัดข้อมูลรายการโรคที่น้อยกว่า 50 ครั้ง และตัดข้อมูลผู้ป่วยเบาหวานที่มาตรวจรักษาไม่ถึง 10 ครั้งออก มาใช้ในการวิเคราะห์ข้อมูลตามแบบจำลอง คริสป์-ดีเอ็ม (CRISP-DM) ในการทำเหมืองข้อมูล เทคนิคการทำเหมืองข้อมูลที่ใช้ในงานวิจัยนี้ เป็นการหาความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm เพื่อวิเคราะห์ข้อมูลภาวะแทรกซ้อนของโรคอื่น เป็นการพยากรณ์โรคอื่นในอนาคต หลังจากเป็นโรคเบาหวานแล้ว

4. วิธีการดำเนินการวิจัย

งานวิจัยนี้เป็นการวิจัยโดยการพัฒนาเหมืองข้อมูลพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยใช้เทคนิค Data Mining การพัฒนาแบบจำลองครั้งนี้ต้องการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นเบาหวาน ใช้ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน จึงกำหนดขั้นตอนการดำเนินการดังต่อไปนี้

4.1 ออกแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โรงพยาบาลปทุมธานี โดยใช้เทคนิคการทำเหมืองข้อมูล

4.2 พัฒนาระบบ โดยแบ่งออกเป็น 2 ส่วนดังนี้

4.2.1 *พัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น* หลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยวิธีการหาความสัมพันธ์ของข้อมูล (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm เครื่องมือที่ใช้ประกอบด้วยโปรแกรม WEKA ใช้ในการวิเคราะห์ข้อมูล MicroSoft Visual Foxpro เพื่อจัดเก็บข้อมูลผลการตรวจวินิจฉัยโรค Microsoft office 2010 เพื่อสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากที่ผู้ป่วยเป็นโรคเบาหวาน

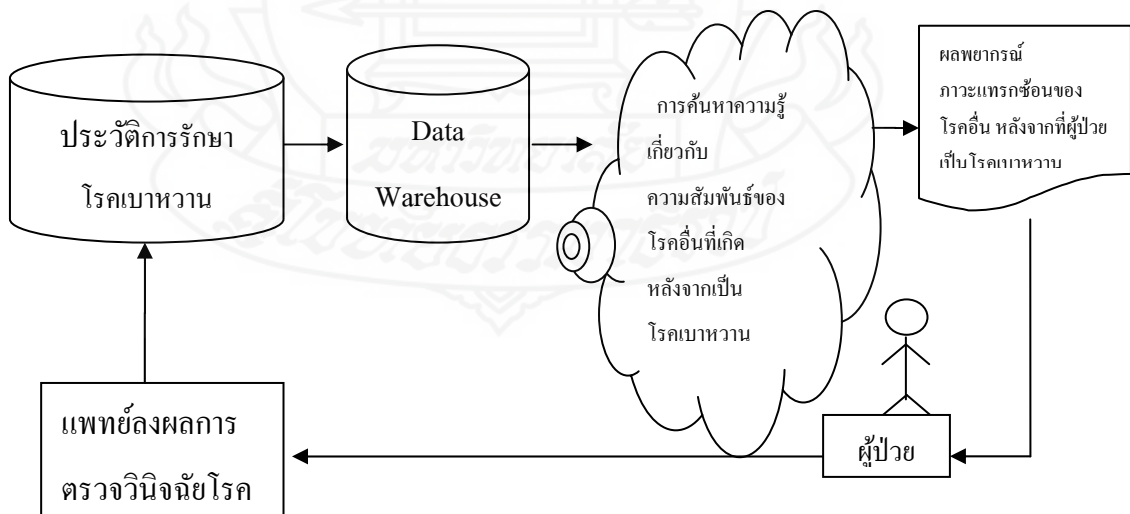
4.2.2 *พัฒนาระบบสารสนเทศ* ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

4.3 นำผลการวิเคราะห์ข้อมูล ความสัมพันธ์ที่ได้จากกระบวนการทำเหมืองข้อมูล มาทำการเปรียบเทียบผลพยากรณ์จากการหาความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm

4.4 นำแบบจำลองที่ได้ไปทดสอบกับสถานการณ์จริง โดยพัฒนาระบบสารสนเทศระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี ข้อมูลที่นำเสนอประกอบด้วยประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์

5. กรอบแนวคิดการวิจัย

กรอบแนวคิดงานวิจัยประกอบด้วย การใช้ Data Mining เพื่อค้นหาภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานประกอบไปด้วย ผู้ป่วยที่มีผลการตรวจวินิจฉัยโรคเป็นโรคเบาหวานทั้งหมดเป็นตัวแปรต้น และข้อมูลผลการตรวจวินิจฉัยโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวานแล้ว มาพัฒนาเป็นคลังข้อมูล (Data Warehouse) เพื่อหาความรู้ที่ต้องการทราบเกี่ยวกับภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูลมาเป็นเครื่องมือค้นหาความรู้ดังนี้



ภาพที่ 1.1 กรอบแนวคิดการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน

6. นิยามศัพท์เฉพาะ

6.1 ผู้ป่วยเบาหวาน หมายถึง ผู้ป่วยที่เป็นโรคเบาหวานด้วยรหัสโรค E10-E14 ตามข้อมูลรหัสโรค ICD101 (ICD101: ICD-10-TM International Statistical Classification of Diseases and Related Health Problems 10th Revision, Thai Modification, 2012)

6.2 การทำเหมืองข้อมูล (Data Mining) หมายถึง กระบวนการวิเคราะห์ข้อมูลที่มีปริมาณมหาศาลเพื่อค้นหาความรู้ที่ไม่เคยทราบมาก่อน โดยอัลกอริทึมที่พัฒนามาจากศาสตร์ด้านสถิติ และด้านปัญญาประดิษฐ์

6.3 การลงการตรวจผลวินิจฉัยโรค หมายถึง การลงผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานที่มารับบริการในแต่ละครั้ง โดยแพทย์จะเป็นผู้ทำการลงผลการตรวจวินิจฉัยโรค ด้วยระบบคอมพิวเตอร์ ของโรงพยาบาลปทุมธานี

6.4 ห้องตรวจ หมายถึง ห้องตรวจโรคอายุรกรรม แผนกรักษาโรคเบาหวาน ที่มีแพทย์เฉพาะทางด้านอายุรกรรม ทำการตรวจรักษาผู้ป่วยเบาหวาน ในโรงพยาบาลปทุมธานี

6.5 ภาวะแทรกซ้อน หมายถึง โรคที่เกิดหลังจากโรคเบาหวานสามารถเกิดได้ทุกระบบ ได้แก่ โรคหัวใจและหลอดเลือด โรคตา โรคไต โรคแทรกซ้อนที่ขา โรคแทรกซ้อนทางระบบประสาท ความดันโลหิตสูง โรคทางช่องปาก โรคผิวหนัง

6.6 Apriori Algorithm หมายถึง อัลกอริทึมที่ใช้ในการค้นหาภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน เป็นขั้นตอนวิธีในการค้นหาข้อมูลที่ปรากฏร่วมกันเพื่อนำไปสร้างกฎความสัมพันธ์

6.7 FPGrowth Algorithm หมายถึง อัลกอริทึมที่ใช้ในการค้นหาภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน เป็นขั้นตอนวิธีในการค้นหาข้อมูลที่ปรากฏร่วมกัน เพื่อนำไปสร้างกฎความสัมพันธ์

6.8 ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน หมายถึง ระบบสารสนเทศที่พัฒนาเพื่อการทดสอบกับสถานการณ์จริง นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

7. ประโยชน์ที่คาดว่าจะได้รับ

7.1 ได้แบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน โรงพยาบาลปทุมธานี

7.2 นำแบบจำลองผลพยากรณ์ภาวะแทรกซ้อนของโรคอื่น ไปใช้กับการให้บริการผู้ป่วยเบาหวาน กลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

7.3 สามารถนำมาอธิบายให้ความรู้กับตัวผู้ป่วยเบาหวานหรือผู้ดูแลผู้ป่วยเบาหวาน ในการดูแลรักษา และให้ปฏิบัติตามคำแนะนำของแพทย์

7.4 เป็นแนวทางให้เจ้าหน้าที่/แพทย์ ทำการรักษาดูแลผู้ป่วยเบาหวานในการกำหนดมาตรการป้องกัน ฝ้าระวังแก้ไข เพื่อลดอัตราการเกิดภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยเบาหวาน

7.5 ลดอัตราเสี่ยงต่อการเป็น อัมพฤกษ์ อัมพาต ของผู้ป่วยเบาหวาน

7.6 ลดอัตราเสี่ยงต่อการเสียชีวิตของผู้ป่วยเบาหวาน

7.7 ลดค่าใช้จ่ายงบประมาณในการดูแลผู้ป่วยเบาหวาน



บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาค้นคว้าข้อมูลเกี่ยวกับทฤษฎีและงานวิจัยที่เกี่ยวข้องจากเอกสารบทความวิชาการและแหล่งสารสนเทศอินเทอร์เน็ต ในหัวข้อดังต่อไปนี้

1. ความรู้พื้นฐานโรคเบาหวานและโรคแทรกซ้อน (Diabetes Mellitus & Complications)
2. ทฤษฎีคลังข้อมูล (Data Warehouse)
3. ทฤษฎีเหมืองข้อมูล (Data Mining)
4. หลักการของคริปส์-ดีเอ็ม
5. โปรแกรมเวก้า (WEKA)
6. งานวิจัยที่เกี่ยวข้อง

1. ความรู้พื้นฐานโรคเบาหวานและโรคแทรกซ้อน (Diabetes Mellitus & Complications)

โรคเบาหวาน คือ ภาวะที่ร่างกายเกิดความผิดปกติ เนื่องจากมีระดับน้ำตาลในเลือดสูง เกิดจากร่างกายไม่สามารถนำน้ำตาลซึ่งได้จากอาหารที่รับประทานเข้าไป ไปใช้เป็นพลังงานในเซลล์ได้ตามปกติร่างกายของคนเราจำเป็นต้องใช้พลังงานในการดำรงชีวิต พลังงานเหล่านี้ได้จากอาหารต่างๆ ที่รับประทานเข้าไป โดยเฉพาะอาหารประเภทแป้งซึ่งถูกย่อยสลายกลายเป็นน้ำตาลกลูโคสในกระเพาะอาหารและถูกดูดซึมเข้าไปในกระแสเลือดเพื่อส่งผ่านไปเลี้ยงเนื้อเยื่อต่างๆ ของร่างกาย แต่การที่ร่างกายจะนำน้ำตาลไปใช้เป็นพลังงานได้นั้นมีความจำเป็นต้องอาศัยองค์ประกอบหลักๆ 2 ชนิด ได้แก่ ฮอร์โมนจากตับอ่อนชื่อ อินซูลินซึ่งมีหน้าที่เป็นตัวพาน้ำตาลกลูโคสเข้าสู่เซลล์ของเนื้อเยื่อต่างๆ และ สารที่ทำหน้าที่กระตุ้นการทำงานของฮอร์โมนอินซูลิน สาร GTF (Glucose Tolerance Factor) หากในร่างกายของเรามีปริมาณสาร GTF น้อยกว่าปกติจะทำให้ประสิทธิภาพการทำงานของอินซูลินลดน้อยลงไปด้วย ดังนั้นหากกลไกทั้งสองอย่างหรืออย่างใดอย่างหนึ่งทำงานบกพร่อง จะทำให้น้ำตาลไม่สามารถเข้าสู่เซลล์ได้ตามปกติ และทำให้น้ำตาลเหลือตกค้างอยู่ในกระแสเลือดมากกว่าปกติ ส่วนใหญ่จะรู้เมื่อมีอาการจากการที่มีน้ำตาลในเลือดขึ้นสูง ได้แก่ อาการคอแห้ง หิวน้ำบ่อย ปัสสาวะบ่อย และปัสสาวะเป็นจำนวนมาก หิวบ่อย ทานจุ แต่น้ำหนักลดลงเรื่อยๆอ่อนเพลีย ไม่มีแรง เป็นต้น บางรายอาจไม่มีอาการดังกล่าวหรือมีอาการไม่มาก ไม่ชัดเจน

แต่จะพบแพทย์ด้วยอาการของภาวะแทรกซ้อนจากโรคเบาหวาน เช่น ตามัว มองเห็นไม่ชัดเนื่องจากมีต่อกระจกหรือจอประสาทตาเสื่อม เท้าชา ไม่รู้สึก หรือมีอาการปวดแสบร้อนที่เท้า เป็นแผลที่เท้าเรื้อรังไม่หาย หรือนิวเท้าดำเนื่องจากขาดเลือดไปเลี้ยง หรือมีอาการของโรคไตวาย เช่น บวม ชีด ปัสสาวะเป็นฟอง เป็นต้น อาการของโรคเบาหวานและภาวะแทรกซ้อนต่างๆ เหล่านี้จะปรากฏเมื่อเป็นโรคเบาหวานมานานหรือมีระดับน้ำตาลในเลือดขึ้นสูงมาก คนที่เริ่มเป็นโรคเบาหวานใหม่ๆ มักจะไม่มีอาการ แต่ไม่ควรรอจนเป็นมากแล้วค่อยมาตรวจและเริ่มการรักษา เพราะภาวะแทรกซ้อนจากโรคเบาหวานส่วนใหญ่ถ้าเป็นมากแล้วจะรักษาไม่หาย ดังนั้นคนที่สงสัยว่าจะเป็นโรคเบาหวานหรือมีความเสี่ยงต่อโรคเบาหวานควรได้รับการตรวจว่าเป็นโรคเบาหวานหรือไม่เสียแต่เนิ่นๆ

ผู้ที่มีปัจจัยเสี่ยงของการเกิดโรคเบาหวานประกอบด้วย บุคคลที่อ้วนหรือลงพุง เคยเป็นโรคเบาหวานตอนตั้งครรภ์ ได้รับยาบางชนิดเช่น สเตียรอยด์ เคยเป็นโรคของตับอ่อนหรือแม่กระทั่งผู้สูงอายุก็มีโอกาสเป็นโรคเบาหวานได้ และปัจจัยที่เกิดจากความผิดปกติภายในร่างกายเช่น บุคคลที่มีประวัติครอบครัวเป็นโรคเบาหวาน เช่นบิดา มารดา หรือพี่น้องเป็นโรคเบาหวาน เป็นต้น การเกิดภาวะแทรกซ้อนในผู้ป่วยเบาหวาน เกิดขึ้นจากระดับน้ำตาลในเลือดที่สูง เป็นระยะเวลานาน จะเป็นพิษต่อร่างกาย โดยเฉพาะอย่างยิ่งต่อระบบหลอดเลือดซึ่งเป็นเสมือนท่อส่งน้ำเลี้ยงของร่างกาย ผู้ป่วยเบาหวานที่ควบคุมไม่ดีจะมีความผิดปกติของระดับไขมันในเลือดด้วย โดยจะมีไขมันในเลือดสูงกว่าปกติ และมีความดันเลือดสูงกว่าปกติ ทั้งระดับน้ำตาลที่สูง ไขมันในเลือดที่สูง และความดันโลหิตที่สูงจะมีผลต่อผนังหลอดเลือด เกิดการเสื่อมสภาพ มีการอักเสบ และมีการสะสมของไขมันที่ผนังหลอดเลือดทำให้ตีบ แคบลง หรืออาจตันไปในที่สุด เลือดผ่านไปยังอวัยวะต่างๆ ไม่ได้ เกิดภาวะขาดเลือด ขาดสารอาหาร ขาดออกซิเจน ทำให้อวัยวะนั้นๆ เสียหาย เช่นถ้า เกิดกับหลอดเลือดหัวใจ เกิดเป็นโรคหลอดเลือดหัวใจตีบ กล้ามเนื้อหัวใจตาย ถ้าเกิดกับเส้นเลือดสมอง ก็จะมีโอกาสเป็นอัมพฤกษ์ อัมพาต ถ้าเกิดกับเส้นประสาท ทำให้เส้นประสาทเสื่อม โดยเฉพาะที่บริเวณเท้า ถ้าเกิดกับจอประสาทตาทำให้จอประสาทตาเสื่อม มีเลือดออก จอประสาทตาหลุดลอก ทำให้ตาบอด ถ้าเกิดกับเส้นเลือดที่ไต ทำให้ไตขาดเลือด ไตเสื่อม เป็นโรคไตวายในที่สุด

ที่มา: <http://www.vcharkarn.com/varticle/36790#chapter1> July 25, 2014

2. ทฤษฎีคลังข้อมูล (Data warehouse)

คลังข้อมูล เป็นระบบข้อมูลขนาดใหญ่ที่เอื้ออำนวยให้ผู้ใช้งานสามารถใช้ข้อมูลได้อย่างมีประสิทธิภาพ เป็น ที่เก็บรวบรวมข้อมูลจากหลายแหล่งที่มา หลายช่วงเวลา มีคุณลักษณะของข้อมูลแตกต่างกัน นำมาไว้รวม ณ ที่เดียวกันใช้เพื่อการวิเคราะห์ (กิตติพงษ์ กลมกล่อม, 2552: 5)

คลังข้อมูล (Data warehouse) เป็นการเก็บข้อมูลในเชิงหัวข้อ (Subject - oriented) ซึ่งทำการรวบรวมข้อมูลมาจากแหล่งต่างๆ (Integrated) โดยจะเก็บข้อมูลเป็นระยะเวลานาน (Time - variant) และ ไม่มีการเปลี่ยนแปลงของชุดข้อมูลได้โดยง่าย (Non - volatile) ซึ่งเป็นข้อมูลสำหรับที่ใช้ช่วยการสนับสนุนในการดำเนินการตัดสินใจเพื่อการบริหาร (Roiger and Geatz, 2003: 148)

2.1 วิวัฒนาการของคลังข้อมูล

จากความเป็นมาของคลังข้อมูล แสดงให้เห็นว่าคลังข้อมูลมีวิวัฒนาการมาจากฐานข้อมูลปฏิบัติการ (สุวรรณณี อัสวกุลชัย, 2555: 4) ในที่นี้จะกล่าวถึงการกำเนิดของฐานข้อมูล ดังนี้

ปี พ.ศ. 2503 เริ่มมีการพัฒนาฐานข้อมูล มีการค้นคว้าและพัฒนาระบบฐานข้อมูลมาเรื่อยๆ เช่น ฐานข้อมูลแบบลำดับขั้น (Hierarchical database) ลักษณะคล้ายต้นไม้ที่กล่าวห้วงฐานข้อมูลแบบเครือข่าย (network database) และฐานข้อมูลเชิงสัมพันธ์ (relational database) ซึ่งมีเครื่องมือในการออกแบบฐานข้อมูลที่เรียกว่า อีอาร์ไดอะแกรม (Entity Relationship - ER - diagram) เป็นต้น มีเทคนิคการใช้ดัชนี (index) การบริหารข้อมูล และภาษาในการสืบค้นข้อมูล (query language)

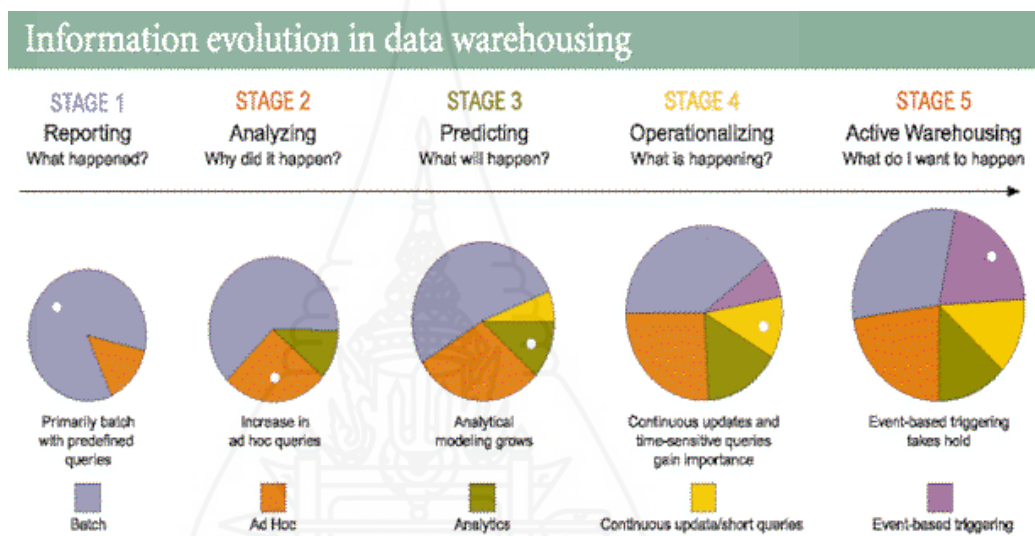
ปี พ.ศ. 2513 คลังข้อมูลก่อกำเนิดที่สถาบันเทคโนโลยีแมสซาชูเซตส์หรือเอ็มไอที (Massachusetts Institute of Technology – MIT) โดยนักวิจัยพยายามแยกความแตกต่างระหว่างฐานข้อมูลปฏิบัติการกับกระบวนการวิเคราะห์ข้อมูล จึงได้มีการพัฒนาแนวทางของโครงสร้างโดยแบ่งออกเป็นชั้นๆ ให้จัดเก็บข้อมูลแยกต่างหากจากส่วนวิเคราะห์ ผลการศึกษาในเบื้องต้นมีอุปสรรคเนื่องจากความยุ่งยากในการประมวลผลที่จะต้องไปเรียกข้อมูลมาจากที่จัดเก็บหลายๆ ฐานข้อมูล จึงไม่ประสบความสำเร็จ

ปี พ.ศ. 2523 ฐานข้อมูลมีการพัฒนาให้มีศักยภาพมากขึ้น ได้มีการนำระบบจัดการฐานข้อมูลเชิงสัมพันธ์ (Relational Database Management System – RDBMS) เป็นแบบจำลองข้อมูลชนิดขั้นสูง (advanced data models) โดยพัฒนามาจากฐานข้อมูลเชิงสัมพันธ์ และได้นำหลักการที่นักวิจัยจากเอ็มไอทีคิดค้นในตอนต้นมาออกแบบส่วนของการเก็บข้อมูล (data capture) การเข้าถึงข้อมูล (data access) การกำหนดไดเรกทอรี (directory) และการให้บริการแก่ผู้ใช้ (user services) ซึ่งส่วน

การนำเข้าข้อมูล นับเป็นส่วนที่เกี่ยวข้องกับฐานข้อมูลปฏิบัติการ ส่วนการเข้าถึงข้อมูล เป็นส่วน
ของกระบวนการวิเคราะห์ข้อมูล ซึ่งโครงสร้างนี้เริ่มเป็นการพัฒนาลงข้อมูลขึ้นมาอีกระดับหนึ่ง

ปี พ.ศ. 2531 ถึงพ.ศ. 2534 ได้มีการนำเทคโนโลยีการทำคลังข้อมูล (Data warehousing) การทำเหมืองข้อมูล (Data mining) ฐานข้อมูลหลายมิติ (multidimensional databases) และฐานข้อมูลบนเว็บ (web databases) มาใช้ในธุรกิจ

ทั้งนี้ ถ้าพิจารณาเฉพาะส่วนของสารสนเทศในคลังข้อมูล พบว่ามีวิวัฒนาการใน
อีกรูปแบบหนึ่ง



ภาพที่ 2.1 วิวัฒนาการของสารสนเทศในคลังข้อมูล

ที่มา: <http://dssresources.com/papers/features/brobst&rarey01062003.html> Retrieved July 25, 2014

ยุคที่ 1 สารสนเทศในคลังข้อมูลที่รวบรวมไว้ ถูกนำไปใช้เพื่อแสดงรายงานเป็นระยะๆ
ว่าเกิดอะไรขึ้น เช่น ในช่วงไตรมาสแรกของปี ได้รายได้เท่าไร เป็นต้น

ยุคที่ 2 เริ่มมีการวิเคราะห์สารสนเทศในคลังข้อมูลแบบทันทีทันใด สามารถตอบ
คำถามว่าทำไมถึงเกิดเหตุการณ์นั้นขึ้น โดยการใช้โอแลป (Online Analytical Processing – OLAP)
มาช่วยวิเคราะห์

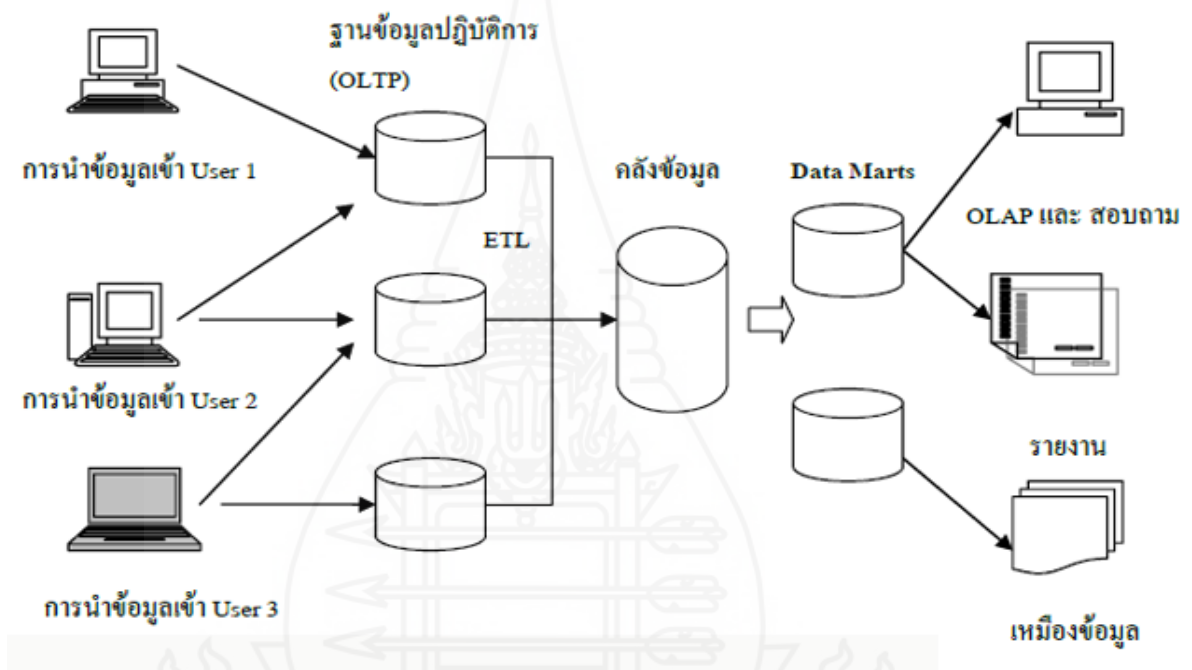
ยุคที่ 3 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์เพื่อการพยากรณ์โดยใช้หลักการของ
เหมืองข้อมูลมาช่วยวิเคราะห์ เช่น ความต้องการสินค้าของลูกค้า เป็นต้น

ยุคที่ 4 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์มากขึ้นอย่างต่อเนื่อง เพื่อสามารถ
ตอบได้ว่าเกิดอะไรขึ้นในอาทิตย์หน้า

ยุคที่ 5 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์เพื่อกำหนดสิ่งที่จะเกิดขึ้นในอนาคต โดยผ่านระบบที่มีการติดต่อสื่อสารที่ได้ต่อบกัน (Interactive) โดยตรง

2.2 กระบวนการในการพัฒนาคลังข้อมูล

เนื่องจากข้อมูลพื้นฐานของฐานข้อมูลในคลังข้อมูลประกอบด้วยข้อมูลจำนวนมากจะต้องมีการออกแบบคลังข้อมูลเพื่อรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล (สุวรรณณี อัสวกุลชัย, 2555: 20)



ภาพที่ 2.2 กระบวนการในการพัฒนาคลังข้อมูล

2.3 สถาปัตยกรรมคลังข้อมูล (Data Warehouse Architecture - DWA)

DWA เป็นโครงสร้างมาตรฐานที่ใช้อธิบายเพื่อให้เข้าใจแนวคิด และกระบวนการของคลังข้อมูลนั้นๆ ซึ่งโดยทั่วไปแล้วคลังข้อมูลแต่ละระบบอาจจะมีรูปแบบที่ไม่เหมือนกันได้ เพื่อให้เหมาะสมกับองค์กรนั้นๆ (เลิศ เลิศศิริ โสภณ, 2541: 95) ทั้งนี้ส่วนประกอบต่างๆ ภายใน DWA ที่สำคัญ ได้แก่

2.3.1 Operational database หรือ external database layer ทำหน้าที่จัดการกับข้อมูลในระบบงานปฏิบัติการหรือแหล่งข้อมูลภายนอกองค์กร

2.3.2 Information access layer เป็นส่วนที่ผู้ใช้ปลายทางติดต่อผ่านโดยตรง ประกอบด้วยฮาร์ดแวร์และซอฟต์แวร์ ที่ใช้ในการแสดงผลเพื่อการวิเคราะห์ โดยมีเครื่องมือช่วย

เป็นตัวกลางที่ผู้ใช้ใช้ติดต่อกับคลังข้อมูล โดยในปัจจุบันเครื่องมือที่ได้รับความนิยมเพิ่มขึ้นอย่างรวดเร็วนั้นคือ Online Analytical Processing Tool หรือ OLAP tool ซึ่งเป็นเครื่องมือที่มีความสามารถในการวิเคราะห์ที่ซับซ้อน และแสดงข้อมูลในรูปแบบหลายมิติ

2.3.3 Data access layer เป็นส่วนต่อประสานระหว่าง Information access layer กับ operational layer

2.3.4 Data director (metadata) layer เพื่อให้เข้าถึงข้อมูลได้ง่ายขึ้น และเป็นการเพิ่มความเร็วในการเรียกและดึงข้อมูลของคลังข้อมูล

2.3.5 Process management layer ทำหน้าที่จัดการกระบวนการทำงานทั้งหมด

2.3.6 Application messaging layer เป็นมิดเดิลแวร์ทำหน้าที่ในการส่งข้อมูลภายในองค์กรผ่านทางเครือข่าย

2.3.7 Data warehouse (physical) layer เป็นแหล่งเก็บข้อมูลของทั้ง Information data และ External data ในรูปแบบที่ง่ายแก่การเข้าถึงและยืดหยุ่นได้

2.3.8 Data staging layer เป็นกระบวนการแก้ไข และดึงข้อมูลจาก External database

2.4 เทคนิคในการสร้างคลังข้อมูล

การเคลื่อนที่ของข้อมูลในคลังข้อมูล ข้อมูลที่จะจัดเก็บภายในคลังข้อมูลมีการเคลื่อนที่ของข้อมูล (Information flow) 5 ประเภท (www.gotoknow.org/posts/452189) ดังนี้

2.4.1 in flow คือการนำข้อมูลจากฐานข้อมูลอื่นเข้าสู่คลังข้อมูลทั้งฐานข้อมูลภายในและภายนอกองค์กร โดยในขั้นนี้อาจมีการเปลี่ยนแปลงโครงสร้างข้อมูล การทำ demoralize การลบหรือเพิ่มฟิลด์เพื่อให้ข้อมูลทั้งหมดอยู่ในเนื้อหาที่สนใจเดียวกัน ในขั้นตอนนี้อาจใช้เครื่องมือที่เรียกว่า Data Warehouse Tool

2.4.2 up flow เมื่อข้อมูลที่เราต้องการอยู่ในคลังข้อมูลแล้วในบางครั้งอาจต้องมีการเพิ่มคุณค่าให้กับข้อมูลด้วยเพื่อให้ข้อมูลอยู่ในรูปแบบที่เป็นประโยชน์มากที่สุดต่อการนำเครื่องมือมาใช้ ซึ่งได้แก่การจัดกลุ่มข้อมูลหาค่าทางสถิติที่ซับซ้อน จัดข้อมูลให้อยู่ในรูปแบบหรือเทมเพลตมาตรฐาน

2.4.3 down flow เป็นขั้นตอนของการปรับปรุงเปลี่ยนแปลงข้อมูลที่เก่า และไม่อยู่ในเนื้อหาที่องค์กรสนใจออกไปจากคลังข้อมูลขององค์กร

2.4.4 out flow เป็นขั้นที่ผู้ใช้เรียกใช้ข้อมูลในคลังข้อมูลผ่านเครื่องมือต่างๆ โดยการเรียกใช้อาจมีเพียงขอเรียกเป็นครั้งคราวเป็นประจำทุกวัน/เดือน หรือแม้กระทั่งต้องการแบบทันที

2.4.5 Meta flow ข้อมูลที่จัดเก็บในคลังข้อมูลจะถูกทำข้อมูลไว้อีกชุดหนึ่ง เป็นแหล่งที่มาของข้อมูลนั้น หรือแม้กระทั่งที่อยู่ของข้อมูลนั้นในคลังข้อมูลและข้อมูลอื่นที่เกี่ยวข้อง

2.5 วิธีการออกแบบคลังข้อมูล

วิธีการนี้ถูกเสนอโดย Kimball ในปี 1996 เรียกว่าระเบียบวิธี 9 ขั้น หรือ Nine - Step Methodology (Connolly, 2002: 922) โดยวิธีการนี้เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่าดาต้ามาร์ท (Data mart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จแล้ว จึงนำมารวมกันเป็นคลังข้อมูลขององค์กรในขั้นสุดท้าย ซึ่งขั้นตอนทั้ง 9 ขั้นตอนมีรายละเอียดดังนี้

2.5.1 กำหนดดาต้ามาร์ท คือการเลือกที่จะสร้างดาต้ามาร์ทของระบบงานใดบ้าง และระบบงานใดเป็นระบบงานแรก โดยองค์กรจะต้องสร้าง E - R model ที่รวมระบบงานทุกระบบขององค์กรไว้ แสดงการเชื่อมโยงของแต่ละระบบงานอย่างชัดเจน และสิ่งที่จะต้องคำนึงถึงในการเลือกระบบงานที่จะเป็นดาต้ามาร์ทแรกนั้น มี 3 ปัจจัยที่เกี่ยวข้อง ได้แก่ จะต้องสามารถพัฒนาออกมาได้ทันตามเวลาที่ต้องการ โดยอยู่ในงบประมาณที่กำหนดไว้และต้องตอบปัญหาทางธุรกิจให้แก่องค์กรได้ ดังนั้นดาต้ามาร์ทแรกควรจะเป็นของระบบงานที่นำรายได้เข้ามาสู่องค์กร เช่นระบบงานขาย เป็นต้น

2.5.2 กำหนด Fact Table ของดาต้ามาร์ท คือการกำหนดเนื้อหาหลักที่ควรจะเป็นของดาต้ามาร์ท โดยการเลือกเอนทิตีหลักและกระบวนการที่เกี่ยวข้องกับเอนทิตีนั้นๆ ออกมาจาก E - R model ขององค์กร นั้นหมายถึงจะทำให้เราทราบถึง Dimension Table ที่ควรจะมีด้วย

2.5.3 กำหนดแอตทริบิวต์ที่จำเป็นของแต่ละ Dimension Table คือการกำหนดแอตทริบิวต์ที่บอกหรืออธิบายรายละเอียดของ dimension ได้ ทั้งนี้แอตทริบิวต์ที่เป็น primary key ควรเป็นค่าที่คำนวณได้ กรณีที่มีดาต้ามาร์ทมากกว่าหนึ่งดาต้ามาร์ทมี dimension เหมือนกัน นั้นหมายถึงว่าแอตทริบิวต์ใน dimension นั้นจะต้องเหมือนกันทุกประการ แต่นั่นก็ไม่อาจจะแก้ไขปัญหาการจัดเก็บข้อมูลซ้ำซ้อน อันจะนำมาสู่ความแตกต่างกันของข้อมูลชุดเดียวกัน ปัญหานี้จึงเป็นการดีที่จะมีการใช้ Dimension Table ร่วมกันในแต่ละ Fact Table ที่จำเป็นต้องมี dimension ดังกล่าว โดยเรียก Dimension Table ลักษณะแบบนี้ว่า conformed และเรียก Fact Table ว่า fact constellation เราสามารถกำหนดข้อดีของการใช้ Dimension Table ร่วมกันได้ดังนี้

- 1) แน่ใจได้ว่าในแต่ละรายงานจะออกมาสอดคล้องกัน
- 2) สามารถสร้างดาต้ามาร์ทในเวลาต่างๆ กันได้
- 3) สามารถเข้าถึงดาต้ามาร์ทโดยผู้พัฒนากลุ่มอื่นๆ
- 4) สามารถรวบรวมดาต้ามาร์ทหลายๆ อันเข้าด้วยกัน

5) สามารถออกแบบคลังข้อมูลร่วมกันได้

2.5.4 กำหนดแอตทริบิวต์ที่จำเป็นใน Fact Table โดยแอตทริบิวต์หลักใน Fact Table จะมาจาก primary key ในแต่ละ Dimension Table นอกจากนี้แล้ว ยังสามารถมีแอตทริบิวต์ที่จำเป็นอื่นๆ ประกอบอยู่ด้วย เช่นแอตทริบิวต์ที่ได้จากการคำนวณค่าเบื้องต้นที่จำเป็นสำหรับการคงอยู่ของแอตทริบิวต์อื่นใน Fact Table เรียกอีกอย่างหนึ่งว่า Measure การกำหนดแอตทริบิวต์นี้ไม่ควรจะเลือกแอตทริบิวต์ที่คำนวณค่าไม่ได้ เช่นเป็นตัวหนังสือหรือไม่ใช่ตัวเลข เป็นต้น และไม่ควรเลือกแอตทริบิวต์ที่ไม่เกี่ยวข้องกับเนื้อหาของ Fact Table ที่เราสนใจด้วย

2.5.5 จัดเก็บค่าการคำนวณเบื้องต้นใน Fact Table คือการจัดเก็บค่าที่ได้จากการคำนวณให้เป็นแอตทริบิวต์หนึ่งใน Fact Table ถึงแม้ว่าจะสามารถหาค่าได้จากแอตทริบิวต์อื่นๆ ก็ตาม ทั้งนี้เพื่อให้การสอบถามมีประสิทธิภาพมากขึ้น สามารถทำงานด้วยความเร็วที่เพิ่มขึ้น เนื่องจากไม่ต้องคำนวณค่าใหม่ทั้งหมด ถึงแม้ว่าจะเกิดความซ้ำซ้อนของข้อมูลในการจัดเก็บบ้างก็ตาม

2.5.6 เขียนคำอธิบายของ Dimension Table ทั้งนี้ก็เพื่อให้ผู้ใช้สามารถใช้งานดาต้ามาร์ทได้อย่างมีประสิทธิภาพเพราะเกิดความเข้าใจอย่างดีในส่วนต่างๆ

2.5.7 กำหนดระยะเวลาในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจจะเป็นการจัดเก็บเพียงช่วงระยะเวลา 1 - 2 ปี หรือนานกว่านั้น ขึ้นอยู่กับความต้องการขององค์กร เนื่องจากองค์กรแต่ละประเภทมีความต้องการในการจัดเก็บข้อมูลต่างช่วงเวลากัน ทั้งนี้ขึ้นอยู่กับความจำเป็นหรือข้อกำหนดในการดำเนินธุรกิจมีข้อสังเกตอยู่ 2 ประการที่น่าสนใจและสำคัญสำหรับการออกแบบแอตทริบิวต์ในเรื่องของการจัดเก็บข้อมูล ดังนี้

- 1) ข้อมูลที่ถูกจัดเก็บไว้นานเกินไปมักเกิดปัญหาการอ่าน หรือแปลข้อมูลนั้นๆ จากเพิ่มหรือลบค่า
- 2) เมื่อมีการนำรูปแบบเก่าของ Dimension Table มาใช้อาจเกิดปัญหาการเปลี่ยนแปลงของ dimension อย่างซ้ำๆ ได้

2.5.8 การติดตามปัญหาการเปลี่ยนแปลงของ dimension คือ การเปลี่ยนแอตทริบิวต์ของ Dimension Table เก่ามาใช้แล้วส่งผลกระทบต่อข้อมูลปัจจุบันของ Dimension Table โดยสามารถแบ่งประเภทของปัญหาที่เกิดขึ้นได้เป็น 3 ประเภท ดังนี้

- 1) เกิดการเขียนทับข้อมูลใหม่โดยข้อมูลเก่า
- 2) เกิดเรคอร์ดใหม่ๆ ขึ้นใน Dimension
- 3) เกิดเรคอร์ดที่มีทั้งค่าเก่าและใหม่ปนกันไป

2.5.9 กำหนดคิวิรีเป็นการออกแบบด้านกายภาพ เพื่อให้ผู้ใช้เกิดความสะดวกในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพเมื่อดำเนินการทั้ง 9 ขั้นตอนสำหรับแต่ละดาต้ามาร์ทเสร็จแล้ว จึงจะนำทั้งหมดมารวมกันเป็นภาพของคลังข้อมูลขององค์กรต่อไป

2.6 การแปลงข้อมูลเข้าสู่ดาต้ามาร์ท

เมื่อเราออกแบบฐานข้อมูลสำหรับแต่ละดาต้ามาร์ทเสร็จแล้ว ขั้นตอนต่อไปที่สำคัญยิ่งก็คือการนำข้อมูลจากแหล่งข้อมูลไปแปลงให้อยู่ในแพลตฟอร์มของฐานข้อมูลที่ได้ออกแบบไว้ นั่นก็คือการแปลงข้อมูล หรือ Extraction Transformation and Loading (ETL) นั่นเอง โดยที่คุณภาพของการแปลงข้อมูลเป็นสิ่งที่สำคัญมากสำหรับการสร้างคลังข้อมูล ความซับซ้อนของการแปลงข้อมูลและโครงสร้างของข้อมูลจะแตกต่างกันไปตามคลังข้อมูลของแต่ละองค์กรต้องการ โดยที่การแปลงข้อมูล

หมายรวมถึงตั้งแต่การวิเคราะห์แหล่งข้อมูล กำหนดการส่งข้อมูลรวบรวมหรือสร้างข้อมูลภายนอก วางแผนและสร้างรูทีนของการแปลงข้อมูล และตรวจสอบความถูกต้องของข้อมูลที่ได้ (นงลักษณ์ พลอยปลื้ม, 2540: 126) สามารถสรุปเป็นขั้นตอนได้ดังนี้

2.6.1 วิเคราะห์แหล่งข้อมูล เช่น ปริมาณของข้อมูล จำนวนและชนิดของการเข้าถึงแหล่งข้อมูล แพลตฟอร์มและภาษาโปรแกรมที่ใช้ เป็นต้น

2.6.2 ย้ายข้อมูลที่ต้องการจากระบบเดิม มาไว้ในบริเวณที่ใช้ปรับแต่งข้อมูลหรือเรียกบริเวณนี้ว่า staging area เพื่อนำมาเลือกเฉพาะส่วนที่ต้องการแปลงข้อมูลและตรวจสอบความถูกต้อง หรือการทำความสะอาดข้อมูล

2.6.3 กำหนด primary key ของ Fact Table และ Dimension Table และกำหนด foreign key ระหว่าง Fact Table กับ Dimension Table

2.6.4 ย้ายข้อมูลที่ทำความสะอาดแล้ว จาก staging area ลงสู่เซิร์ฟเวอร์ของดาต้ามาร์ท

2.6.5 สร้าง metadata ของแต่ละดาต้ามาร์ท โดยเก็บรายละเอียดของข้อมูลการอัปเดตและส่งออกไว้ในดาต้ามาร์ท

2.6.6 ตรวจสอบความถูกต้องของข้อมูล ซึ่งจะต้องกระทำตลอดทั้งกระบวนการแปลงข้อมูลจะทำได้ดังนี้

1) ตรวจสอบผลรวมทั้งหมดของจำนวนข้อมูลที่ดึงมาจากแหล่งข้อมูลที่เพิ่มเข้าไป

2) ตรวจสอบแก้ไขข้อมูลในระบบเดิมของแหล่งข้อมูล ซึ่งควรจะเก็บข้อมูลในการตรวจแก้ไขไว้ใน Metadata ของการแปลงข้อมูลด้วย

3) ตรวจสอบค่าของข้อมูลให้ถูกต้องในกระบวนการรวบรวมข้อมูล

4) ตรวจสอบผลรวมของข้อมูลหลังจากการย้ายข้อมูลลงสู่ดาต้ามาร์ทแล้ว

3. ทฤษฎีเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล คือ การค้นหาความรู้ในคลังข้อมูล (อังกฤษ: Knowledge Discovery in Databases - KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (Pattern) จากข้อมูลจำนวนมากโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และการรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

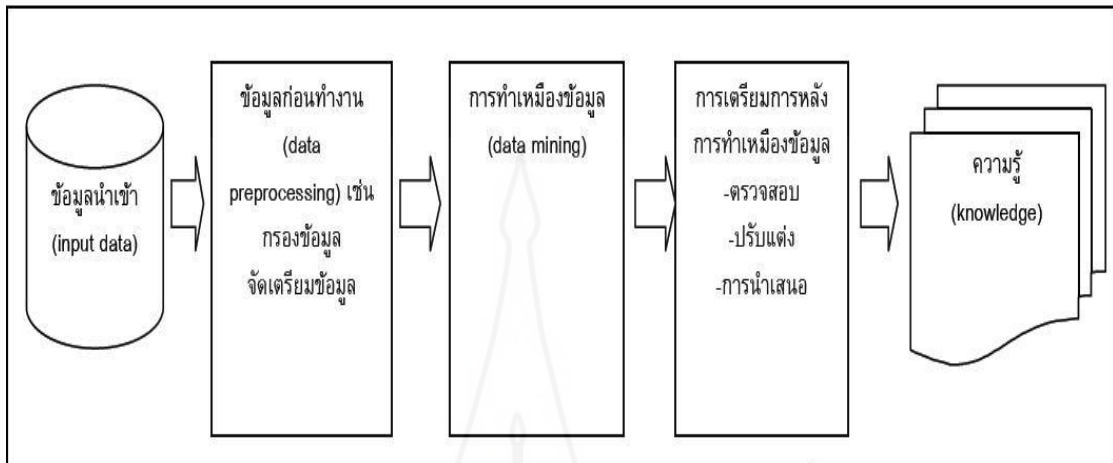
http://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_Related_Health_Problems July 24,2014

เหมืองข้อมูลเป็นเทคโนโลยีที่ช่วยให้การวิเคราะห์ข้อมูลทำได้โดยอัตโนมัติและมีประสิทธิภาพสูง ซึ่งส่งผลให้มีการนำเหมืองข้อมูลไปใช้อย่างแพร่หลายในทุกวงการ โดยเฉพาะกรณีที่ข้อมูลมีขนาดใหญ่มาก โดยระบบคอมพิวเตอร์จะทำหน้าที่ค้นหาแนวโน้มและลักษณะที่น่าสนใจต่างๆ ที่ปรากฏในข้อมูล ตลอดจนวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลนั้น ซึ่งจะช่วยให้สามารถพยากรณ์แนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้นในอนาคตได้ รวมถึงเข้าใจความสัมพันธ์ที่เชื่อมโยงข้อมูลแต่ละกลุ่มเข้าด้วยกัน (Nicholson, 2003)

การทำเหมืองข้อมูล เป็นกระบวนการในการค้นหาความสัมพันธ์ใหม่ระหว่างข้อมูลด้วยกันรวมทั้งแพทเทิร์นและแนวโน้ม โดยการกลั่นกรองข้อมูลปริมาณมหาศาลที่จัดเก็บอยู่ ด้วยการใช้เทคโนโลยีการรู้จำแพทเทิร์น ความรู้ทางสถิติและคณิตศาสตร์มาประยุกต์เข้าด้วยกัน (Gartner Group www.gartner.com อ้างใน Larose (2005) “Discovering Knowledge in Data: An Introduction to Data Mining”)

จากความหมายของการทำเหมืองข้อมูลที่กล่าวข้างต้นได้ว่า การทำเหมืองข้อมูล หมายถึง กระบวนการค้นหาความรู้ (Knowledge) ซึ่งเป็นรูปแบบและความสัมพันธ์ที่น่าสนใจที่มีอยู่ในข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ซับซ้อน โดยมี อัลกอริทึม ให้เลือกหลากหลายวิธี เป็นขั้นตอนวิธีการ ค้นหา รูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำแพทเทิร์น การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ ซึ่งจะช่วยให้สามารถพยากรณ์แนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้น ในอนาคตได้ รวมถึงเข้าใจความสัมพันธ์ที่เชื่อมโยงข้อมูลแต่ละกลุ่มเข้า

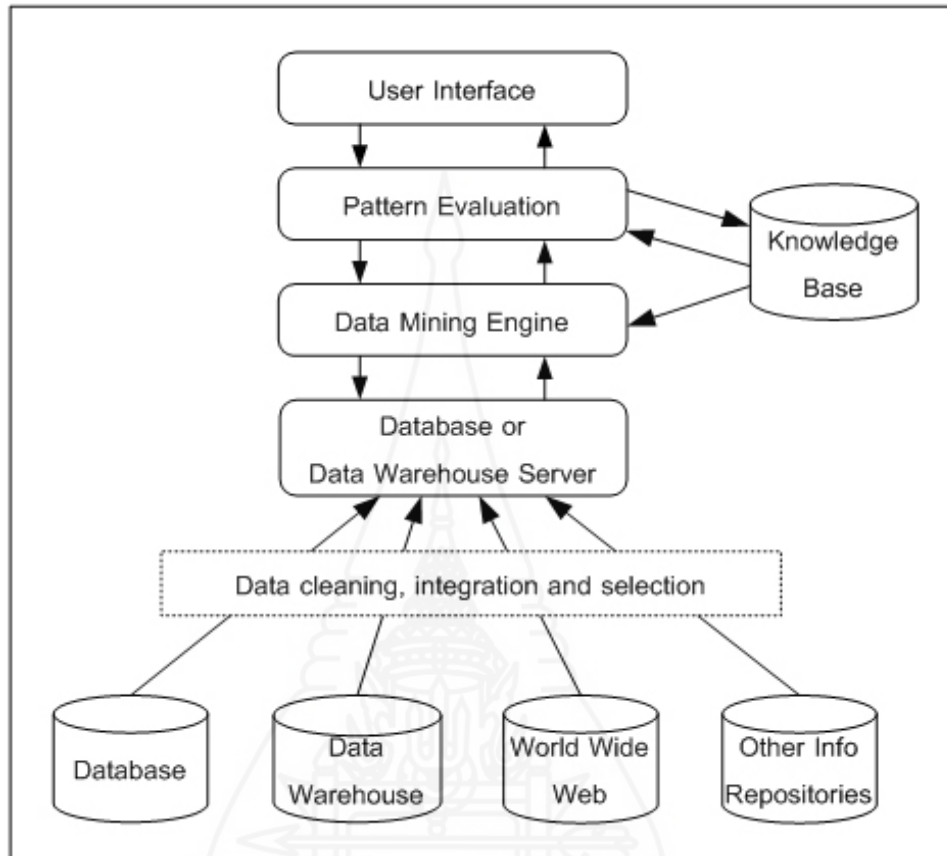
ด้วยกัน การค้นหาความสัมพันธ์ (Association) การจัดหมวดหมู่ (Classification) การรวมกลุ่ม (Clustering) การทำนายล่วงหน้า (Prediction) โดยมีขั้นตอนในการทำเหมืองข้อมูล ดังภาพที่ 2.3



ภาพที่ 2.3 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล



3.1 สถาปัตยกรรมระบบเหมืองข้อมูล



ภาพที่ 2.4 สถาปัตยกรรมระบบเหมืองข้อมูล

ที่มา : <http://siripornk.blogspot.com/2010/08/data-mining.html> June 24, 2014

3.1.1 แหล่งข้อมูลที่ใช้ทำเหมืองข้อมูล (Data Sources) เป็นแหล่งข้อมูลที่มีการจัดเก็บไว้อยู่ในฐานข้อมูล โดยอาจอยู่ในรูปของภาพยนตร์ เสียง รูปภาพ แผนภาพ ฐานข้อมูล เป็นต้น การทำเหมืองข้อมูลต้องการค้นหาความรู้ที่ซ่อนเร้นอยู่ ซึ่งมีขั้นตอนการดำเนินการเพื่อให้ข้อมูลอยู่ในสภาพที่พร้อมใช้งาน การคัดเลือกข้อมูลตามความต้องการใช้ในการทำเหมืองข้อมูล การทำความสะอาดข้อมูล (Data Cleaning) เป็นการจัดการข้อมูลที่ไม่สมบูรณ์ ไม่ถูกต้อง รายละเอียดขาดหายไป เนื้อหาข้อมูลขัดแย้งกันเอง

1) การบูรณาการข้อมูล (Data Integration) การนำข้อมูลจากแหล่งต่างๆ ที่สัมพันธ์กันมาไว้ในที่เดียวกัน จำเป็นต้องปรับเปลี่ยนรูปแบบให้เป็นรูปแบบเดียวกัน

2) การคัดเลือกข้อมูล (Data Selection/Extraction) เป็นการนำข้อมูลที่ต้องการมาใช้งานมาจัดเก็บตามความต้องการใช้ในการทำเหมืองข้อมูล

3.1.2 คลังข้อมูล (Data Warehouse Server) เป็นระบบแม่ข่ายที่ใช้จัดเก็บข้อมูลเพื่อรองรับข้อมูลทั้งหมดในการทำเหมืองข้อมูล

3.1.3 ฐานความรู้ (Knowledge Base) หมายถึง ความรู้ความเข้าใจในสิ่งหรือเรื่องที่ต้องการวิเคราะห์หรือค้นหา ความรู้ความเข้าใจเหล่านี้เกิดจากการเรียนรู้และประสบการณ์ในเรื่องดังกล่าว มีความเข้าใจในธรรมชาติของข้อมูลที่มีอยู่และความรู้ในการเลือกวิธีการทำเหมืองข้อมูลได้อย่างเหมาะสมกับลักษณะงานหรือข้อมูลที่ต้องการดำเนินการ

3.1.4 กลไกการทำเหมืองข้อมูล (Data Mining Engine) เป็นวิธีการทำเหมืองข้อมูล ได้แก่ การกำหนดคุณสมบัติข้อมูล การวิเคราะห์ความสัมพันธ์ของข้อมูล การจำแนกหมวดหมู่ การจัดกลุ่ม การค้นหาความผิดปกติของข้อมูล และเครื่องมือที่ใช้ในการทำเหมืองข้อมูลแต่ละประเภท เช่น วิธีการทางสถิติ

3.1.5 ส่วนประเมินรูปแบบ (Pattern Evaluation Module) ในการทำเหมืองข้อมูล สิ่งที่สำคัญคือ การสร้างแบบจำลองหรือตัวแบบ (Model) เพื่อแสดงถึงสมมติฐานได้อย่างชัดเจนให้สามารถค้นพบความรู้จากแบบจำลองที่สร้างขึ้นมาได้ ทั้งนี้แบบจำลองที่สร้างขึ้นควรต้องมีความเหมาะสมในการพยากรณ์ได้ และเครื่องมือหรือกลไกเหล่านี้ ยังใช้ตรวจสอบและประเมินแบบจำลองที่จัดทำขึ้นมาด้วย เพื่อให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือ เหมาะสม และถูกต้องมากที่สุด

3.1.6 ส่วนที่ใช้ติดต่อกับผู้ใช้งาน (User Interface) แบ่งเป็น

- 1) ส่วนรองรับการป้อนคำสั่งเข้าของผู้ใช้งาน เช่น การค้นหา การใส่เงื่อนไขการทำเหมืองข้อมูล
- 2) ส่วนนำเสนอผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลในรูปแบบต่างๆ เช่น ตาราง แผนภูมิ และรูปแบบความสัมพันธ์ของข้อมูลที่ค้นหา

3.2 ประเภทการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นการค้นหารูปแบบที่เกิดขึ้นกับข้อมูลที่สนใจ ซึ่งแบ่งตามลักษณะงานที่ดำเนินการ (Data mining Operation/Functionality) เป็นประเภทต่างๆ ได้ดังนี้ (วิชาเจริญทัศน์วารักษ์, 2555)

3.2.1 การค้นหาคุณลักษณะ/รายละเอียดของข้อมูล เป็นการค้นหาคุณลักษณะเฉพาะของข้อมูล (Class/Concept description) ที่เราให้ความสนใจหรือต้องการดำเนินการ เพื่อเข้าใจในรายละเอียดที่เกี่ยวข้องกับข้อมูลเหล่านั้น การทำเหมืองข้อมูลประเภทนี้ แบ่งได้ 2 ลักษณะ คือ

- 1) การกำหนดคุณลักษณะเฉพาะข้อมูล (Data Characterization) เป็นการค้นหาคุณลักษณะเฉพาะของข้อมูล เพื่ออธิบายรายละเอียดเจาะลึกของข้อมูลนั้น ทำให้เข้าใจภาพรวมของข้อมูลในเรื่องนั้น ตัวอย่างเช่น การค้นหาคุณลักษณะ/ประวัติของลูกค้า (Customer Profile) ที่ซื้อ

สินค้ากับบริษัท “STOU Electronics” มากกว่า \$1,000 ต่อปี โดยใช้คำสั่งเอสคิวแอล (SQL) เพื่อคัดกรองข้อมูลที่ต้องการในเบื้องต้น พบว่า ลูกค้ามีอายุระหว่าง 40-50 ปี มีงานทำ และมีประวัติ/เครดิต (Credit Rating) อยู่ในระดับดี เป็นต้น

2) การกำหนดคุณลักษณะแตกต่าง (Data Discrimination) เป็นการค้นหาและศึกษารายละเอียดของข้อมูลที่เราสนใจเปรียบเทียบกับข้อมูลเดิมที่มีคุณสมบัติที่แตกต่างหรือตรงกันข้าม ซึ่งการเปรียบเทียบข้อมูล มักแบ่งเป็น 2 กลุ่มคือ กลุ่มหลัก และกลุ่มที่แตกต่าง ทั้งนี้ผู้ทำเหมืองข้อมูลต้องเป็นผู้กำหนดเงื่อนไขเอง เช่น การศึกษาเปรียบเทียบรายละเอียดข้อมูลของกลุ่มลูกค้าที่ซื้อสินค้ากับบริษัท STOU Electronics มากกว่า 3 ครั้งต่อเดือน เทียบกับกลุ่มลูกค้าที่ซื้อสินค้าน้อยกว่า 3 ครั้งต่อปี เป็นต้น ผลที่ได้จากการค้นหารายละเอียดคือประวัติการซื้อสินค้าของลูกค้า เช่น กลุ่มที่ซื้อสินค้าประจำ จะเป็นผู้ที่มีอายุระหว่าง 20-40 ปี จบการศึกษาระดับปริญญาตรี ขณะที่กลุ่มที่แตกต่าง จะมีประวัติของลูกค้าที่มีอายุต่ำกว่า 20 หรือ มากกว่า 60 ปี และจบการศึกษาในระดับที่ต่ำกว่าปริญญาตรี เป็นต้น ทั้งนี้อาจจะใช้ข้อมูลที่จัดเก็บในคลังข้อมูลมาดำเนินการโดยใช้คำสั่งเอสคิวแอล (SQL) ในการค้นหาตามเงื่อนไขที่ต้องการได้ เครื่องมือหรือวิธีการในการทำเหมืองข้อมูลแบบนี้ ใช้วิธีการทางสถิติ เช่น ค่าเฉลี่ย (Mean) ค่ามัธยฐาน (Median) ค่าฐานนิยม (Mode) การหาค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) การเขียนกราฟต่างๆ การหาความสัมพันธ์ระหว่างตัวแปรที่สนใจ และการเปรียบเทียบข้อมูลด้วยตารางหลายมิติ (Multidimensional Data Cube) เป็นต้น การนำเสนอผลลัพธ์ที่ได้ อาจจะเป็นแผนภูมิแบบต่างๆ เช่น แผนภูมิแท่ง เส้น และวงกลม เป็นต้น

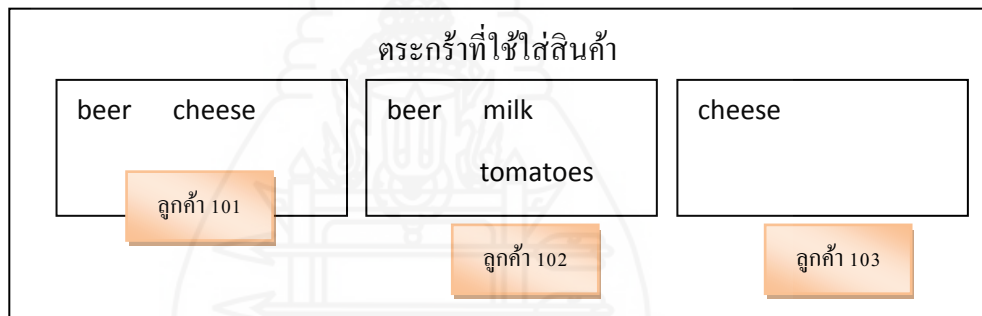
3.2.2 การค้นหาความสัมพันธ์ เป็นการทำเหมืองข้อมูลแบบการค้นหาความสัมพันธ์ (Frequent Pattern And Association) ประกอบด้วย การค้นหารูปแบบที่เกิดควบคู่กัน (Frequent Pattern) และสร้างกฎความสัมพันธ์ที่เกิดขึ้น (Association Rule) มีรายละเอียดดังนี้

1) การค้นหารูปแบบของข้อมูลที่เกิดขึ้นควบคู่กันในรายการเดียวกันเสมอ (Frequent Item Set) การทำเหมืองข้อมูลประเภทนี้นิยมใช้ในการวางแผนการตลาด การวางแผนการส่งซื้อสินค้าว่าควรส่งซื้อสินค้าใดร่วมกันบ้าง การจัดชั้นวางสินค้าอย่างเหมาะสม การจัดรายการส่งเสริมการขายสินค้า จึงมักนิยมเรียกการทำเหมืองข้อมูลประเภทนี้ว่า การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เป็นการวิเคราะห์หรือค้นหาความสัมพันธ์ข้อมูลที่เกิดควบคู่กันในแต่ละรายการ (Transaction) เช่น เมื่อซื้อนม มักจะซื้อ ขนมปัง ไข่ ร่วมกันด้วยเสมอ เป็นต้น

2) การค้นหาความสัมพันธ์ (Association) หลังจากการค้นหารูปแบบที่เกิดขึ้นควบคู่กัน (Frequent Item Set) ได้แล้ว จะสร้างกฎความสัมพันธ์ของข้อมูล (Association Rule) โดยการกำหนดค่า “Support” ซึ่งเป็นจำนวนรายการข้อมูลที่น่ามาดำเนินการ และค่า “Confidence” เป็น

ค่าความน่าจะเป็นที่ใช้ในการวิเคราะห์ความสัมพันธ์ เช่น STOU Electronics ต้องการค้นหารายการข้อมูล (Transaction) ในการซื้อสินค้าของลูกค้า ตามกฎความสัมพันธ์ต่อไปนี้ “Buys (X, ‘Computer’) => Buys (X, ‘Software’) (Support=1%, Confidence=50%)” หมายความว่า 1% ของจำนวนทรานแซกชันที่นำมาดำเนินการค้นหา (Support=1%) พบว่าโอกาสหรือความน่าจะเป็นของลูกค้าที่ซื้อ “Computer” แล้วจะซื้อ “Software” มี 50 % (Confidence=50%) ผลที่ได้ทำให้ผู้วิเคราะห์มีความเข้าใจความเชื่อมโยงสัมพันธ์ของข้อมูลได้ดียิ่งขึ้น

ตัวอย่างของการค้นหากฎความสัมพันธ์ คือการวิเคราะห์พฤติกรรมของผู้บริโภค โดยการหาความสัมพันธ์ระหว่างสินค้าที่ต่างชนิดกัน แต่ลูกค้าจะเลือกซื้อไปพร้อมกัน ดังภาพที่ 2.5 โดยการค้นหาความสัมพันธ์นี้จะช่วยให้ทราบว่าสินค้าใดถูกลูกค้าเลือกซื้อไปคู่กันมากที่สุด และนำกฎความสัมพันธ์ที่ค้นพบนั้นมาใช้ในการปรับปรุงกลยุทธ์การขายสินค้าหรือใช้ประกอบการพิจารณาในการจัดวางสินค้า ทำให้เกิดความสะดวกในการเลือกซื้อสินค้า และเป็นการเพิ่มยอดขายให้กับร้านค้า



ภาพที่ 2.5 การวิเคราะห์พฤติกรรมกรรมการซื้อสินค้า

กฎความสัมพันธ์สามารถเขียนให้อยู่ในรูปไอเท็มเซตที่เป็นเหตุ ไปสู่อิเท็มเซตที่เป็นผล เช่น ลูกค้าที่ซื้อ beer ส่วนใหญ่จะซื้อ milk ด้วย ก็สามารถเขียนความสัมพันธ์ได้เป็น {beer} → {milk} เป็นต้น

การอธิบายวิธีการค้นหาความสัมพันธ์จะต้องใช้นิยามดังต่อไปนี้

- 1) ไอเท็มเซต (Item Sets : I) คือเซตที่มีไอเท็มทั้งหมดเป็นสมาชิก ซึ่งไอเท็มในที่นี้อาจเป็นชื่อของหน่วยใดๆ ที่นำมาใช้ในการเรียนรู้
- 2) ทรานแซกชัน (Transaction : T) เป็นเซตย่อยของไอเท็ม โดยที่ $T \subseteq I$
- 3) เซ็ตของข้อมูล (Data: D) คือเซตที่มีทรานแซกชันทุกตัวเป็นสมาชิก โดยที่ $D = \{t_1, t_2, \dots, t_3\}$
- 4) การกำหนดค่าสนับสนุน (Support) s_{min}' โดยที่ $0 \leq s_{min}' \leq 1$

- 5) การกำหนดค่าความเชื่อมั่น (Confidence) c_{min}' โดยที่ $0 \leq c_{min} \leq 1$
- 6) ทรานแซกชัน T บรรจุเซตย่อยของไอเท็ม X ก็ต่อเมื่อ $X \subseteq t$
- 7) ค่าสนับสนุนของไอเท็มเซต X คือ $s(X) = \frac{|\{t \in D \mid X \subseteq t\}|}{|D|}$
- 8) กฎที่ได้อยู่ในรูปแบบ $L \rightarrow R$ เมื่อ $L \subset I, R \subset I$ และ $L \cap R = \emptyset$
- 9) ค่าความเชื่อมั่นของกฎ $L \rightarrow R$ คือ $c(L, R) = s(L \cup R) / s(L)$
- 10) ค่าสนับสนุนของกฎ $L \rightarrow R$ คือ $s(L, R) = s(L \cup R)$
- 11) ทุกกฎความสัมพันธ์ $L \rightarrow R$ เมื่อ $L \subset I, R \subset I$ และ $L \cap R = \emptyset$ แล้วจะต้องประกอบไปด้วยค่าสนับสนุนและค่าความเชื่อมั่น ซึ่งนิยามได้ดังนี้

(1) กฎความสัมพันธ์ $L \rightarrow R$ มีค่าสนับสนุนเท่ากับ s ก็ต่อเมื่อ s% ของทรานแซกชันใน D บรรจุ $L \cup R$

(2) กฎความสัมพันธ์ $L \rightarrow R$ มีค่าความเชื่อมั่นเท่ากับ c ก็ต่อเมื่อ c% ของทรานแซกชันใน D บรรจุ $L \cup R$ แต่ละกฎที่ถูกค้นพบเรียกว่ากฎความสัมพันธ์ โดยการค้นหาความสัมพันธ์คือการค้นหาความสัมพันธ์ทั้งหมดในทรานแซกชันทุกตัว โดยกฎความสัมพันธ์ที่หาได้ทั้งหมดจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่กำหนดไว้ ($s(L \cup R) \geq s_{min}$) และมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่กำหนดไว้ ($c(L, R) \geq c_{min}$) โดยกฎความสัมพันธ์ที่มีค่าทั้งสองสูงถึงเกณฑ์ที่กำหนดจะถูกเรียกว่า “แข็งแกร่ง” (Strong) ค่าสนับสนุนเรียกว่า “การปรากฏบ่อย”

Database D		จะได้กฎความสัมพันธ์ดังนี้	
tid	Item		
101	beer		$\{beer\} \rightarrow \{milk, tomatoes\}$
101	cheese	$s_{min} = 0.2$	ค่าสนับสนุน = $s(\{beer, milk, tomatoes\})$
102	bee □	$c_{min} = 0.5$	= 1/3
102	milk		= 33.33%
102	tomatoes	ค่าความเชื่อมั่น	= 1/2
103	cheese		= 50%

ภาพที่ 2.6 แสดงตัวอย่างการค้นหาความสัมพันธ์

ที่มา : Han *et al.*, (2006)

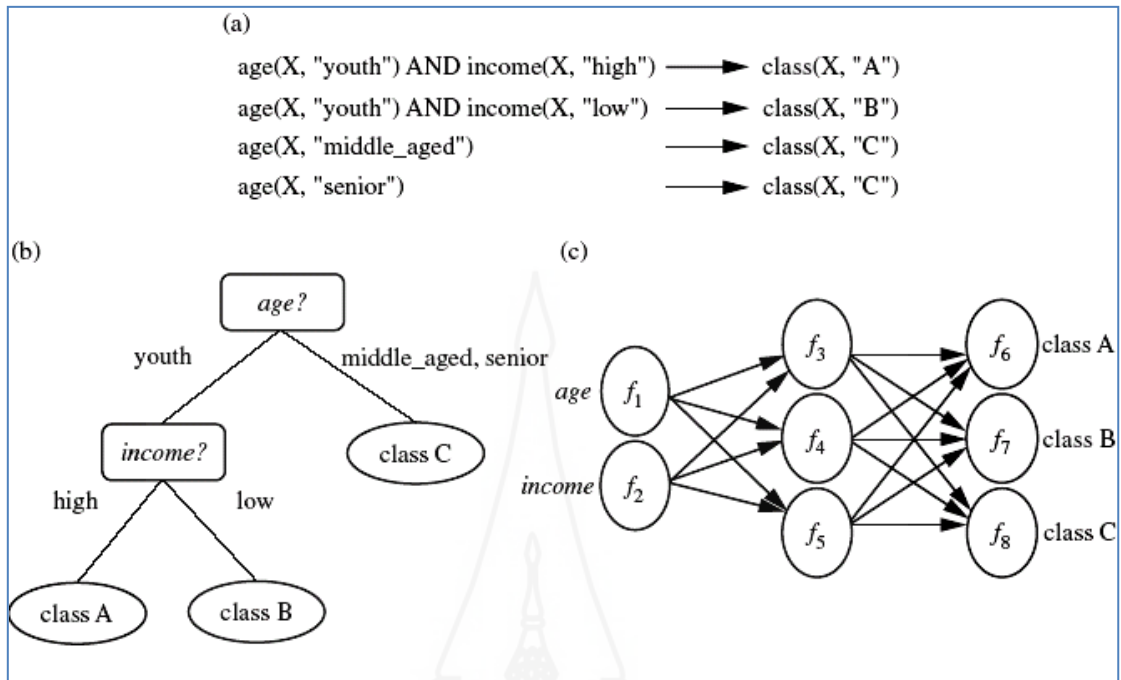
เครื่องมือในการทำเหมืองข้อมูลประเภทนี้มักจะเป็นใช้เทคนิคหรืออัลกอริทึมมาช่วยรวบรวมและวิเคราะห์ความถี่ของการเกิดข้อมูล เช่น การหาจำนวนความถี่การเกิดรายการ (Frequent item set) กฎความสัมพันธ์ (Association Rule) ด้วยอัลกอริทึมอะพริออริ (Apriori) อัลกอริทึมเอฟพี-โกรท (FP-Growth) เป็นต้น

3.2.3 การจำแนกประเภทและการพยากรณ์ การจำแนกประเภทและการพยากรณ์ (Classification and Prediction) แบ่งเป็น

1) การจำแนกประเภท (Classification) เป็นการทำความเข้าใจข้อมูลเพื่อจำแนกว่าข้อมูลที่ต้องการวิเคราะห์ว่าอยู่ในประเภท (Class) ใด โดยอาศัยแบบจำลองหรือตัวแบบที่สร้างขึ้นมา การจำแนกประเภท (Classification) จะใช้ข้อมูลที่มีอยู่จำนวนหนึ่ง หรือชุดข้อมูลเพื่อสร้างแบบจำลอง (Training Data Set) หลังจากที่ได้แบบจำลองแล้วก็จะต้องทำการทดสอบแบบจำลองดังกล่าวด้วยข้อมูลอีกกลุ่ม (Testing Data Set) เพื่อตรวจสอบความเที่ยงหรือความน่าเชื่อถือของแบบจำลองนั้น การทำความเข้าใจข้อมูลโดยวิธี จำแนกประเภท (Classification) จะใช้สำหรับข้อมูลที่มีค่าไม่ต่อเนื่อง (Discrete data) เช่น ข้อมูลที่ระบุเป็นผลการตัดสินใจว่า ใช่ หรือ ไม่ใช่ เสี่ยง หรือ ไม่เสี่ยง

ตัวอย่างเช่น ฝ่ายสินเชื่อของธนาคารต้องการวิเคราะห์ความเสี่ยงในการให้สินเชื่อลูกค้าของธนาคาร ซึ่งมีแบบจำลองที่สร้างขึ้นเพื่อกำหนดหรือใช้พยากรณ์ประเภท (Class) โดยมีเงื่อนไข ดังนี้ หากเงินเดือนตั้งแต่ 20,000 ขึ้นไป และมีงานประจำ มีประเภท (Class) ความเสี่ยงการให้สินเชื่อ คือ ไม่เสี่ยง แต่หากเงินเดือนต่ำกว่า 20,000 และงานที่ทำเป็นงานชั่วคราว มีประเภท (Class) ความเสี่ยงการให้สินเชื่อ คือ เสี่ยง ดังนั้นหากต้องการพยากรณ์ ความเสี่ยงในการให้สินเชื่อแก่นาย A ซึ่งมีเงินเดือน 20,000 และมีอาชีพรับราชการ การพยากรณ์ประเภท (Class) ความเสี่ยงของการให้สินเชื่อนาย A คือ ไม่เสี่ยง

เครื่องมือในการทำเหมืองข้อมูลแบบจำแนกประเภท (Classification) ได้แก่ การใช้อัลกอริทึมสำหรับสร้างต้นไม้การตัดสินใจ (Decision tree) ทฤษฎีเบย์ (Bayes Thorem) นาอิว์เบย์ (Naïve Bayes) โครงข่ายความเชื่อแบบเบย์ (Bayes Belief network) การวิเคราะห์แบบ Rule Base ในรูปแบบ IF_THEN Rule โครงข่ายประสาทประติษฐ์หรือนิวรัลเน็ตเวิร์ก (Neural Networks) เอสวีเอ็ม (SVM -Support Vector Machine) กฎความสัมพันธ์ (Association Rule) การหาคุณสมบัติที่มีความคล้ายคลึงหรือใกล้เคียงกันโดยวิธี K-nearest Neighbor การหาเหตุและผลเพื่อตอบสนองมติฐานโดยใช้วิธี Case Base Reasoning และอัลกอริทึมที่ใช้หลักการเชิงพันธุกรรม (Genetic Alogorithm) ภาพที่ 2.7 เป็นตัวอย่างเครื่องมือที่ใช้ในการจำแนกประเภท (Classification) 2.7 (a) IF-THEN Rule 2.7 (b) ต้นไม้การตัดสินใจ (decision tree) 2.7 (c) โครงข่ายประสาทประติษฐ์ (Neural Networks)



ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภท

ที่มา : Han *et al.*,(2006)

2) การพยากรณ์ (Prediction) มีหลักการเช่นเดียวกับการจำแนกประเภท แต่จะใช้กับข้อมูลที่มีค่าต่อเนื่อง (Continuous data) หรือค่าข้อมูลที่มีค่าเป็นลำดับการเกิดของข้อมูล (Ordered Data) นั้น จำแนกอยู่ในประเภท (Class) ใด เช่น ข้อมูลอายุ เงินเดือน ความสูง เวลา รายได้จากการขายสินค้า เป็นต้น การทำเหมืองข้อมูลประเภทนี้เพื่อใช้พยากรณ์หรือประมาณการผลลัพธ์จากกลุ่มข้อมูลที่มีอยู่ในขณะนั้น เช่น การพยากรณ์ยอดขายสินค้าจากข้อมูลปัจจุบัน เครื่องมือในการทำเหมืองข้อมูลแบบพยากรณ์ (Prediction) ได้แก่ การวิเคราะห์ความถดถอย (Regression Analysis) การวิเคราะห์และตรวจสอบด้วยการใช้ไคสแควร์ (Chi Square Automatic Interaction Detection – CHAID)

3.2.4 การจัดกลุ่ม การทำเหมืองข้อมูลประเภทจัดกลุ่ม (Clustering) ไม่ได้กำหนดหรือแบ่งกลุ่มไว้ล่วงหน้า เป็นเทคนิคจัดกลุ่มสมาชิกที่มีความคล้ายคลึงกันสูงสุด (Nearest Neighbor) เพื่อจำแนกข้อมูลแต่ละหน่วยในชุดข้อมูล โดยวัดค่าความห่างจากจุดศูนย์กลาง (Distance Measurement) การวัดค่าเบี่ยงเบน (Deviation Measurement) เป็นการรวมหน่วยที่คล้ายกันมากที่สุดเข้าเป็นกลุ่มเดียวกัน ซึ่งบางครั้งเรียกเทคนิคนี้ว่า k-nearest neighbor ตัวอย่างเช่น การทำเหมืองข้อมูลเพื่อวิเคราะห์ตำแหน่งที่อยู่ของลูกค้าบริษัท STOU Electronics เพื่อจัดกลุ่มลูกค้าว่าอาศัยใน

ในเขตพื้นที่ใดบ้าง ดังภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่มได้ จะเห็นว่า การจัดกลุ่มเหล่านี้ไม่ได้มีการกำหนดว่าจะต้องเป็นกลุ่มใดล่วงหน้า ในการเกิดกลุ่มเหล่านี้เป็น เพราะข้อมูลในแต่ละกลุ่มมีคุณลักษณะเป็นไปในทางเดียวกันจึงรวมตัวกันเป็นกลุ่มๆ ได้ ซึ่งต่างจากการจำแนกประเภท (Classification) ที่มีการกำหนดกลุ่มหรือประเภท (Class) ไว้ล่วงหน้า เครื่องมือในการทำเหมืองข้อมูลแบบการจัดกลุ่ม (Clustering) เช่น การใช้การวิเคราะห์ค่าเฉลี่ยแบบเค (k-mean analysis)



ภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่ม
ที่มา: Han,Kamber (2006)

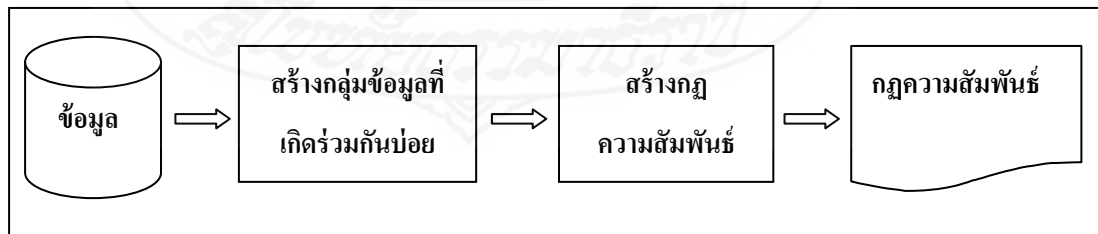
3.2.5 การค้นหาหน่วยข้อมูลที่ผิดปกติ เป็นการทำเหมืองข้อมูลเพื่อค้นหาหน่วยข้อมูลที่แตกต่างหรือผิดปกติ (Outlier Analysis) จากคุณสมบัติข้อมูลทั่วไปที่จัดเก็บในฐานข้อมูลอย่างมาก สาเหตุของการเกิดหน่วยเช่นนี้อาจจะมาจากความผิดพลาดในการบันทึกข้อมูล หรือความผิดพลาดอื่น (Noise) หรือเป็นลักษณะผิดปกติจริง (Exception) การค้นพบหน่วยผิดปกติและจัดการกับข้อมูลดังกล่าวอย่างเหมาะสมในการทำเหมืองข้อมูลจะทำให้ข้อความรู้ที่ได้มีคุณค่ามากขึ้น เช่น ติดตามการฉ้อโกงในการใช้บัตรเครดิต ที่มีรายการและยอดการใช้จ่ายที่พุ่งสูงขึ้นอย่างรวดเร็ว เมื่อเทียบกับพฤติกรรมการใช้เวลาปกติทั่วไป เครื่องมือที่ใช้ในการทำเหมืองข้อมูลประเภทนี้ มักจะเป็นการใช้สถิติเพื่อวิเคราะห์ความน่าจะเป็น การวัดค่าความห่างจากจุดศูนย์กลาง (Distance Measurement) และการวัดค่าเบี่ยงเบน (Deviation Measurement)

3.2.6 การวิเคราะห์การเปลี่ยนแปลงที่อิงตามกาลเวลา เป็นการทำเหมืองข้อมูลเพื่อค้นหารูปแบบข้อมูลที่มีการเปลี่ยนแปลงตามเวลา (Evolution Analysis) มักจะดำเนินการผสมผสาน

วิธีต่างๆ ที่กล่าวมาข้างต้นเพื่อค้นหารูปแบบเพื่อใช้ประโยชน์จากข้อมูลที่มีอยู่ เช่น การวิเคราะห์การเปลี่ยนแปลงของราคาหุ้นในแต่ละประเภทเพื่อการลงทุนที่เหมาะสม เป็นต้นเครื่องมือที่นิยมใช้ เช่น การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

3.3 ทฤษฎีเทคนิคของการหาความสัมพันธ์ (Association)

การค้นหากฎความสัมพันธ์ของข้อมูลในฐานข้อมูล (Rygielski et al., 2002; Agrawal et al., 1993; Berberidis et al., 2005) ได้มีการพัฒนาขึ้นครั้งแรกโดยนักวิจัยจากศูนย์วิจัย IBM (International Business Machines Corporation) ประเทศสหรัฐอเมริกา มีจุดประสงค์เพื่อการค้นหาความสัมพันธ์ที่น่าสนใจซึ่งซ่อนอยู่ในข้อมูล โดยการนำไปใช้ที่เห็นได้ชัดเจนสำหรับการค้นหาความสัมพันธ์ คือ การค้นหากฎความสัมพันธ์ของการซื้อสินค้าของลูกค้าว่าจะซื้อสินค้าใดบ้างร่วมกันในรถเข็นในซูเปอร์มาร์เก็ต (Market Basket Analysis) เพื่อทำความเข้าใจถึงพฤติกรรมการซื้อสินค้าของลูกค้า ยกตัวอย่างเช่น เมื่อลูกค้าซื้อนมแล้วจะซื้อขนมปังด้วย เป็นต้น การค้นหากฎความสัมพันธ์นั้นมีขั้นตอนการทำงานที่สำคัญที่สุด คือ การค้นหากฎข้อมูลที่ปรากฏร่วมกันบ่อย ซึ่งจะใช้เวลาในการทำงานนานที่สุด ดังนั้นในการเลือกขั้นตอนวิธีจึงต้องเลือกขั้นตอนวิธีที่เหมาะสมกับลักษณะของข้อมูลในฐานข้อมูล เพื่อลดระยะเวลาและเนื้อที่หน่วยความจำในการทำงาน โดยทั่วไปการค้นหากฎความสัมพันธ์สามารถแสดงดังภาพที่ 2.9 แบ่งการทำงานออกเป็น 2 ขั้นตอน คือ ขั้นตอนการค้นหากฎข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด โดยหาได้จากกลุ่มข้อมูลในฐานข้อมูลที่เกิดขึ้นร่วมกันในแต่ละรายการข้อมูล โดยกลุ่มข้อมูลเหล่านั้นจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนดจึงจะถือว่าเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย และขั้นตอนการนำกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยมาสร้างกฎความสัมพันธ์ โดยกฎความสัมพันธ์จะเป็นที่ยอมรับได้หากตรวจสอบค่าความเชื่อมั่นของกฎนั้นแล้วมีค่ามากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่ผู้ใช้กำหนด



ภาพที่ 2.9 กระบวนการค้นหากฎความสัมพันธ์

3.3.1 นิยามกฎความสัมพันธ์ กำหนดให้ $I = \{i_1, i_2, i_3, \dots, i_m\}$ เป็นเซตของข้อมูลเรียกว่าสินค้า (Item) กำหนดให้ T เป็นชุดข้อมูล (Transaction) โดยที่ T เป็นกลุ่มข้อมูล (Itemset)

ซึ่ง $T \subseteq I$ จะเรียกชุดข้อมูล T ว่าครอบคลุมกลุ่มข้อมูล X ก็ต่อเมื่อ $X \subseteq T$ กฎความสัมพันธ์สรุปได้ในรูป $X \Rightarrow Y$ โดยที่ $X \subset I, Y \subset I$ และ $X \cap Y = \emptyset$ กฎความสัมพันธ์ $X \Rightarrow Y$ เกิดขึ้นในฐานข้อมูล D ด้วยค่าความมั่นใจ c เมื่อ c เปรอ์เซ็นต์ของชุดข้อมูลที่มีกลุ่มข้อมูล X ในฐานข้อมูล D มีกลุ่มข้อมูล Y ปรากฏอยู่ด้วย กฎความสัมพันธ์ $X \Rightarrow Y$ จะเป็นกฎที่น่าสนใจก็ต่อเมื่อ กฎนั้นมีค่าความมั่นใจ c ไม่ต่ำไปกว่าค่าความมั่นใจขั้นต่ำ (Minimum Confidence) กฎความสัมพันธ์ $X \Rightarrow Y$ มีค่าสนับสนุนเท่ากับ s เมื่อ s เปรอ์เซ็นต์ของชุดข้อมูลในฐานข้อมูล D มี $X \cup Y$ ปรากฏอยู่กลุ่มข้อมูล X จะเป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อยก็ต่อเมื่อ ค่าสนับสนุน (Support) ของกลุ่มข้อมูลนั้นมีค่าไม่ต่ำไปกว่าค่าสนับสนุนขั้นต่ำ (Minimum Support) กลุ่มข้อมูลนอกจากที่กล่าวนี้เป็นกลุ่มข้อมูลนอกเหนือความสนใจ

3.3.2 การสืบค้นกฎความสัมพันธ์ ประกอบด้วยการทำงาน 2 ขั้นตอนหลักคือ

1) การสืบค้นกลุ่มข้อมูลที่เกิดขึ้นบ่อย (Frequent Itemsets) การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยเป็นขั้นตอนที่สำคัญสำหรับการค้นหาความสัมพันธ์ เนื่องจากต้องค้นหากลุ่มข้อมูลที่อาจเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยเพื่อนำ ไปสร้างกฎความสัมพันธ์เป็นจำนวนมาก ซึ่งขั้นตอนนี้จะใช้เวลานานและเนื้อที่ในหน่วยความจำจำนวนมาก จึงทำให้ต้องมีการเลือกขั้นตอนวิธีที่เหมาะสมกับลักษณะข้อมูลในฐานข้อมูล ทรัพยากรของเครื่องคอมพิวเตอร์ที่มีอยู่ หรือผลลัพธ์ของข้อมูลที่ต้องการจากการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย

2) การนำเอากลุ่มข้อมูลที่เกิดขึ้นบ่อย ที่ได้จากขั้นตอนที่ 1 มาสร้างเป็นกฎกลุ่มข้อมูลใดๆ จะถือว่าเป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อยก็ต่อเมื่อ กลุ่มข้อมูลนั้นมีค่าสนับสนุนที่ได้จากจำนวนการปรากฏของกลุ่มข้อมูลนั้นในฐานข้อมูลไม่น้อยไปกว่า ค่าสนับสนุนขั้นต่ำ ซึ่งเป็นค่าที่ผู้กำหนดขึ้นเพื่อเป็นเกณฑ์ในการตัดสินใจว่า กลุ่มข้อมูลใดเป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อย เช่น ต้องมีข้อมูลผู้ซื้อนมกับขนมปังพร้อมกันไปจากร้านค้าเกิดขึ้นในฐานข้อมูลไม่ต่ำกว่า 600 ราย จากลูกค้าทั้งหมด 1000 ราย จึงจะถือว่านมกับขนมปังเป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อย

กฎของความสัมพันธ์ที่ได้จากการสืบค้น จะเป็นกฎที่เชื่อถือได้ก็ต่อเมื่อกฎที่สร้างขึ้นมานั้นมีค่าความเชื่อมั่นไม่น้อยกว่าค่าความเชื่อมั่นขั้นต่ำ โดยค่าความเชื่อมั่นของกฎ $X \Rightarrow Y$ จะมีค่าเท่ากับ ค่าสนับสนุนของ XY หารด้วยค่าสนับสนุนของ X เช่น ค่าความเชื่อมั่นของกฎ ถ้าลูกค้าซื้อขนมปังไปแล้วจะซื้อนมไปพร้อมกันด้วย มีค่าเท่ากับจำนวนลูกค้าที่ซื้อนมพร้อมกันกับขนมปังหารด้วยจำนวนลูกค้าที่ซื้อนมไปอย่างเดียวเป็นต้น

ตัวอย่างการสืบค้นกฎความสัมพันธ์จากฐานข้อมูลรายการการซื้อสินค้าของร้านค้าร้านหนึ่งปรากฏตามตารางที่ 2.1 โดย รายการที่ คือ รหัสประจำรายการการซื้อสินค้า และข้อมูลคือรายการของสินค้าที่ถูกซื้อ ไปพร้อมกัน ในรายการการซื้อสินค้านั้น กำหนดให้ค่าสนับสนุนขั้นต่ำ

ที่ใช้ในการสืบค้นกลุ่มข้อมูลที่เกิดขึ้นบ่อยมีค่าเป็น 60 % จากการทำงานของ การสืบค้นกฎ ความสัมพันธ์ในขั้นตอนที่ 1 จะได้กลุ่มข้อมูลที่เกิดขึ้นบ่อยในฐานข้อมูลดังแสดงในตารางที่ 2.2 โดยกลุ่มข้อมูล คือ กลุ่มข้อมูลที่เกิดขึ้นบ่อย และค่าสนับสนุน คือ จำนวนครั้งที่กลุ่มข้อมูลนั้น ปรากฏในฐานข้อมูลและจากการทำงานของ การทำเหมืองข้อมูลในขั้นตอนที่ 2 โดยกำหนดให้กฎที่ สร้างขึ้นอยู่ในรูปของกฎความสัมพันธ์ โดยมีค่าความมั่นใจขั้นต่ำ เป็น 80 % จะได้ตัวอย่างของกฎ ความสัมพันธ์พร้อมทั้งค่าความมั่นใจดังแสดงในตารางที่ 2.3 โดยกฎ คือ กฎที่สร้างได้จากกลุ่ม ข้อมูลที่เกิดขึ้นบ่อย และค่าความมั่นใจ คือ ค่าความมั่นใจของกฎนั้น

ตารางที่ 2.1 ตัวอย่างข้อมูลรายการการซื้อสินค้า

รายการที่	ข้อมูล
1	A C D
2	B C E
3	A B C E
4	B E
5	A B C E

ตารางที่ 2.2 กลุ่มข้อมูลที่เกิดขึ้นบ่อยที่ได้จากฐานข้อมูลตัวอย่าง

กลุ่ม	ค
{AC}	3 (60%)
{BC}	3 (60%)
{BE}	4 (80%)
{CE}	3 (60%)
{BCE}	3 (60%)

ตารางที่ 2.3 ตัวอย่างกฎความสัมพันธ์ที่สร้างได้จากกลุ่มข้อมูลที่เกิดขึ้นบ่อย

กฎ	ค่าความ
$\{A\} \Rightarrow \{C\}$	4/4 (100%)
$\{B\} \Rightarrow \{C\}$	3/4 (75%)
$\{B\} \Rightarrow \{E\}$	4/4 (100%)
$\{C\} \Rightarrow \{E\}$	3/4 (75%)
$\{BC\} \Rightarrow \{E\}$	3/3 (100%)

จากตัวอย่างกฎความสัมพันธ์ที่ได้จากตัวอย่างงานนี้จะแปลความหมายได้ดังนี้ $\{C\} \Rightarrow \{E\}$ หมายความว่า รายการการซื้อสินค้าในฐานข้อมูลตัวอย่างที่มี $\{C\}$ ปรากฏอยู่มีความเป็นไปได้ 75% ที่จะมี $\{E\}$ ปรากฏอยู่ด้วย

3.3.3 หลักการทำงานของอัลกอริทึม Apriori ขั้นตอนวิธี Apriori (R. Agrawal and R. Srikant, 1993) เป็นขั้นตอนวิธีพื้นฐานในการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยและสร้างกฎความสัมพันธ์ เป็นขั้นตอนวิธีที่ได้รับการยอมรับและได้รับความนิยมเป็นอย่างมาก อีกทั้งขั้นตอนวิธี Apriori ยังเป็นขั้นตอนวิธีที่มีอิทธิพลต่อการศึกษาและพัฒนาขั้นตอนวิธีอื่นๆ มีขั้นตอนการทำงานดังภาพที่ 2.10

```

1  $L_1 = \{\text{large 1-itemsets}\};$ 
2 For ( $k=2; L_{k-1} \neq \emptyset; k++$ ) do begin
3  $C_k = \text{apriori-gen}(L_{k-1});$  // New candidates
4 Forall transactions  $t \in D$  do begin
5  $C_t = \text{subset}(C_k, t);$  // Candidates contained in  $t$ 
6 Forall candidates  $c \in C_t$  do
7  $c.\text{count}++;$ 
8 End
9  $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\};$ 
10 End
11  $\text{Answer} = \bigcup_k L_k;$ 

```

ภาพที่ 2.10 ขั้นตอนการทำงานของ Apriori Algorithm (Agrawal and Srikant, 1993)

จากภาพที่ 2.10 โดยการทำงานของขั้นตอนวิธี Apriori สามารถอธิบายได้ดังนี้

1) อ่านชั้นข้อมูลจากฐานข้อมูลครั้งแรกเพื่อนับค่าความถี่ของแต่ละชั้นข้อมูลที่ปรากฏทั้งหมดในฐานข้อมูล

2) ตรวจสอบค่าความถี่ของแต่ละชั้นข้อมูล เพื่อกำหนดค่าสนับสนุนโดยหากชั้นข้อมูลนั้นๆ มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำก็จะถือว่าเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยที่มีขนาดของชั้นข้อมูล 1 ชั้นข้อมูล (L_1 : Frequent 1-Itemsets)

3) นำ L_1 ที่ได้มาสร้างกลุ่มข้อมูลที่ซึ่งที่มีขนาดของชั้นข้อมูล 2 ชั้นข้อมูล (C_2 : Candidate 2-Itemsets)

4) อ่านชั้นข้อมูลจากฐานข้อมูลอีกครั้งเพื่อนับค่าความถี่ของ C_2 และตัด C_2 ที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนขั้นต่ำ โดยหาก C_2 มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำก็จะกลายเป็น L_2

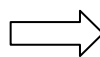
5) ทำในหัวข้อที่ 3) และ 4) ซ้ำจนกว่าไม่สามารถสร้าง C_k จาก L_{k-1} ได้ เมื่อ k คือขนาดของชั้นข้อมูล จึงสิ้นสุดการสร้างกลุ่มข้อมูลที่ซึ่งและจบการทำงานของขั้นตอนวิธี Apriori ทำให้ได้กลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด

(1) ตัวอย่างการทำงานของขั้นตอนวิธี Apriori เพื่อให้เข้าใจถึงกระบวนการทำงานของขั้นตอนวิธี Apriori ได้อย่างชัดเจนขึ้น จึงยกตัวอย่างการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย พร้อมทั้งกำหนดค่าสนับสนุนขั้นต่ำเป็น 0.4 (มีค่าความถี่อย่างน้อยเท่ากับ 2 รายการข้อมูลปรากฏในฐานข้อมูล) โดยตัวอย่างการทำงานของขั้นตอนวิธี Apriori สามารถอธิบายได้ดังนี้

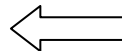
รหัสรายการ (TID)	ชั้นข้อมูล (Items)
A001	A C D
A002	B C E
A003	A B C E
A004	B E
A005	A B C E

Database

Item	Count
A	3
B	4
C	4
E	4



Item	Count	Support
A	3	0.6
B	4	0.8
C	4	0.8
D	1	0.2
E	4	0.8

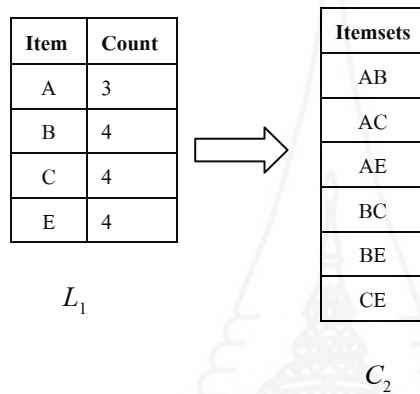


C_1

L_1

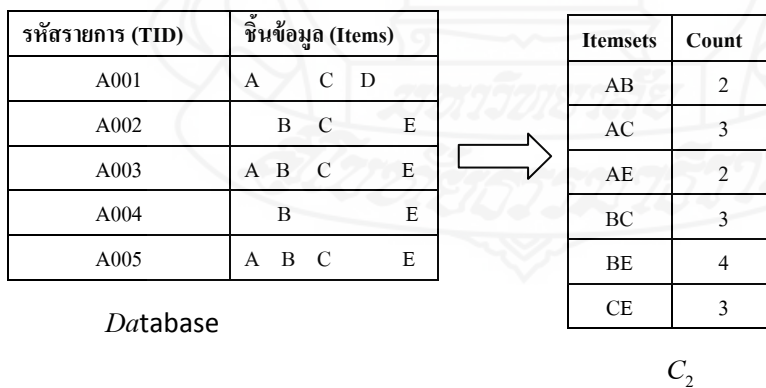
ภาพที่ 2.11 การอ่านข้อมูลจากฐานข้อมูลเพื่อค้นหา L_1 ของ Apriori Algorithm

(2) นำ L_1 ที่ได้ไปสร้างกลุ่มข้อมูลที่ซ้ำ C_2 โดยวิธีการสร้าง C_2 นั้นจะจับคู่ร่วมกัน (Join) ของชั้นข้อมูลทุกตัวที่เป็น L_1 ทุกกรณีที่เป็นไปได้ที่ประกอบด้วยชั้นข้อมูล 2 ชั้นข้อมูล ดังภาพที่ 2.12



ภาพที่ 2.12 การสร้างกลุ่มข้อมูลที่ซ้ำ C_2 ของ Apriori Algorithm

(3) หลังจากสร้างกลุ่มข้อมูลที่ซ้ำ C_2 แล้วนั้น ขั้นตอนต่อไป ก็อ่านชั้นข้อมูลจากฐานข้อมูลเพื่อนับค่าความถี่ของ C_2 แสดงดังภาพที่ 2.13



ภาพที่ 2.13 การนับค่าความถี่ของข้อมูลที่ซ้ำ C_2 ของ Apriori Algorithm

(4) นำค่าความถี่ของ C_2 ที่ได้ไปคำนวณหาค่าสนับสนุนเพื่อใช้ในการตัดสินใจให้เป็นกลุ่มข้อมูล L_2 ผลลัพธ์ที่ได้แสดงดังภาพที่ 2.14

Itemsets	Count	Support
AB	2	0.4
AC	3	0.6
AE	2	0.4
BC	3	0.6
BE	4	0.8
CE	3	0.6

C_2

→

Itemsets	Count	Support
AB	2	0.4
AC	3	0.6
AE	2	0.4
BC	3	0.6
BE	4	0.8
CE	3	0.6

L_2

ภาพที่ 2.14 ผลลัพธ์การค้นหากลุ่มข้อมูล L_2 ของ Apriori Algorithm

(5) นำกลุ่มข้อมูล L_2 ที่ได้ไปสร้างกลุ่มข้อมูลทำชิง C_3 โดยวิธีการสร้างจะจับคู่ร่วมกันของกลุ่มข้อมูลทุกตัวที่เป็น L_2 ที่มีชั้นข้อมูลตัวแรกเหมือนกัน ซึ่งผลลัพธ์ของการจับคู่จะประกอบด้วยชั้นข้อมูล 3 ชั้นข้อมูล ดังภาพที่ 2.15

Itemsets	Count	Support
AB	2	0.4
AC	3	0.6
AE	2	0.4
BC	3	0.6
BE	4	0.8
CE	3	0.6

L_2

→

Itemsets
ABC
ABE
ACE
BCE

C_3

ภาพที่ 2.15 การสร้างกลุ่มข้อมูลทำชิง C_3 ของ Apriori Algorithm

(6) หลังจากสร้างกลุ่มข้อมูลที่ซิง C_3 แล้วนั้น ขั้นตอนต่อไปคือ อ่านชั้นข้อมูลจากฐานข้อมูลเพื่อนับค่าความถี่ของ C_3 แสดงดังภาพที่ 2.16

รหัสรายการ (TID)	ชั้นข้อมูล (Items)
A001	A C D
A002	B C E
A003	A B C E
A004	B E
A005	A B C E

Database

→

Itemsets	Count
ABC	2
ABE	2
ACE	2
BCE	3

C_3

ภาพที่ 2.16 การนับค่าความถี่ของข้อมูลที่ซิง C_3 ของ Apriori Algorithm

(7) กำหนดหาค่าสนับสนุนจากค่าความถี่ที่ได้ พร้อมทั้งตัดกลุ่มข้อมูลที่ซิง C_3 ที่ไม่ผ่านค่าสนับสนุนขั้นต่ำ ซึ่งผลลัพธ์ที่ได้จะเรียกว่า L_3 แสดงดังภาพที่ 2.17

Itemsets	Count	Support
ABC	2	0.4
ABE	2	0.4
ACE	2	0.4
BCE	3	0.6

C_3

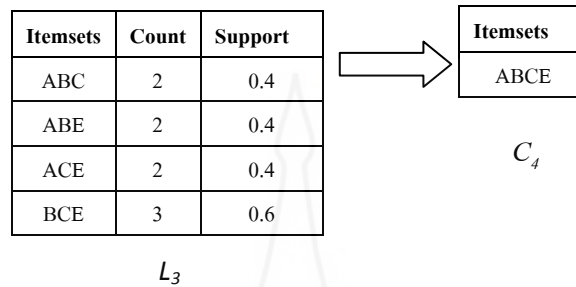
→

Itemsets	Count	Support
ABC	2	0.4
ABE	2	0.4
ACE	2	0.4
BCE	3	0.6

L_3

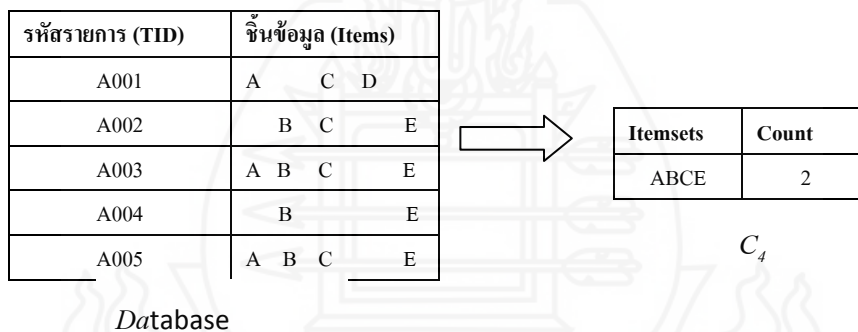
ภาพที่ 2.17 ผลลัพธ์การค้นหากลุ่มข้อมูล L_3 ของ Apriori Algorithm

(8) นำกลุ่มข้อมูล L_3 ที่ได้ไปสร้างกลุ่มข้อมูลทำชิง C_4 โดยวิธีการสร้างจะจับคู่ร่วมกันของกลุ่มข้อมูลทุกตัวที่เป็น L_3 ที่มีชั้นข้อมูลตัวแรกเหมือนกัน ซึ่งผลลัพธ์ของการจับคู่จะประกอบด้วยชั้นข้อมูล 4 ชั้นข้อมูล ดังภาพที่ 2.18



ภาพที่ 2.18 การสร้างกลุ่มข้อมูลทำชิง C_4 ของ Apriori Algorithm

(9) หลังจากสร้างกลุ่มข้อมูลทำชิง C_4 แล้วนั้น ขั้นตอนต่อไปคือ อ่านชั้นข้อมูลจากฐานข้อมูลเพื่อนับค่าความถี่ของ C_4 แสดงดังภาพที่ 2.19



ภาพที่ 2.19 การนับค่าความถี่ของข้อมูลทำชิง C_4 ของ Apriori Algorithm

จากภาพประกอบที่ 2.11 -2.19 มีการทำงานดังนี้คือ เริ่มจากพิจารณาหา 1-frequent itemset ของรายการดำเนินการซึ่งมีรายการทั้งหมด 5 รายการ พิจารณา item ทั้งหมดในที่นี้ประกอบด้วย A,B, C,D และ E และนับความถี่ของ item ที่ปรากฏในรายการ จะได้ว่า A มีปรากฏในรายการทั้งหมด 4 ครั้งหรือ 4 รายการนั่นเอง จากนั้นทำการนับความถี่ของ item ทุกตัวผลที่ได้แสดงดังตาราง C_1 ซึ่งเรียก itemset ที่เกิดว่า 1-Candidate itemset หลังจากนั้นเปรียบเทียบความถี่ที่ได้กับค่าสนับสนุนต่ำสุดที่กำหนดซึ่งตัวอย่างกำหนดให้เท่ากับ 40% ดังนั้น 40% ของการดำเนินการ 5 รายการเท่ากับ 1 เพราะฉะนั้น itemset ที่ในการดำเนินการต้องปรากฏมากกว่า 1 ครั้งจึงจะพิจารณาว่าเป็น 1-frequentitemset ดังแสดงในตาราง L_1 ในการสร้าง 2-Candidate itemset เป็นการเชื่อม 1-

frequent itemset กับ 1-frequent itemset โดยเทียบสมาชิกตัวหน้าของแต่ละ itemset หากมีสมาชิกเหมือนกันนำ itemset มาเชื่อมต่อกันผลการทำงานดังตาราง C_2 หลังจากนั้นนับค่าความถี่ของการเกิดขึ้นต้องนับรายการที่มี Candidate itemset ปรากฏพร้อมกันในรายการเช่น itemset {AB} ปรากฏในรายการทั้งหมด 2 รายการเป็นต้น แล้วทำการกรอง 2- Candidate itemset ที่มีค่าความถี่ต่ำกว่าสนับสนุนต่ำสุดผลลัพธ์ ที่ได้แสดงดังตาราง L_2 การทำงานรอบถัดไปสร้าง 3-Candidate itemset นับค่าความถี่และคัด Candidate itemset ที่ไม่ผ่านเกณฑ์ผลที่ได้ดังตาราง C_2 และ L_3 ตามลำดับ ทำการสร้าง Candidate itemset ในลำดับถัดไปคือ 4-Candidate itemset นับความถี่และกรอง Candidate itemset ที่ไม่ผ่านค่าที่กำหนดเพื่อสร้าง 4-frequent itemset ซึ่งจะพบว่าในขั้นนี้ 4-frequent itemset ไม่สามารถสร้าง 5-Candidate itemset ได้อีกแล้วเพราะฉะนั้นหยุดการทำงาน ผลลัพธ์สุดท้ายที่ได้คือ frequentitemset ที่มีสมาชิกคือ {A}, {B}, {C}, {E}, {AB}, {AC}, {AE}, {BC}, {BE}, {CE}, {ABC}, {ABE}, {ACE}, {BCE} และ {ABCE}

3.3.4 หลักการทำงานของขั้นตอนวิธี FP-Growth Algorithm Jiawei Han และคณะ (Charu C. Aggarwal, Jiawei Han, 2014) ได้มีการพัฒนาขั้นตอนวิธีใหม่ขึ้นมาเพื่อลดจำนวนของการอ่านข้อมูลจากฐานข้อมูล พร้อมทั้งนำเสนอโครงสร้างข้อมูลแบบใหม่ขึ้นมาที่มีชื่อว่า FP-Tree โดยใช้ชื่อว่าขั้นตอนวิธีแบบ FP-Growth เป็นขั้นตอนวิธีที่ช่วยให้มีการอ่านข้อมูลข้อมูลเพียง 2 ครั้งและไม่มีการสร้างกลุ่มข้อมูลทำซ้ำ เพื่อลดระยะเวลาในการประมวลผลให้สามารถทำงานได้เร็วขึ้น หลักการทำงานของขั้นตอนวิธี FP-Growth แสดงดังภาพที่ 2.20 ซึ่งขั้นตอนวิธี FP-Growth เป็นขั้นตอนวิธีที่มีลักษณะการค้นหากลุ่มข้อมูลที่ปรากฏบ่อยรูปแบบ (Pattern Growth) โดยการทำงานของขั้นตอนวิธี FP-Growth อธิบายหลักการทำงานได้ดังนี้

1) อ่านข้อมูลจากฐานข้อมูลครั้งแรกเพื่อนับค่าความถี่ของแต่ละชั้นข้อมูลแล้วนำชั้นข้อมูลที่ไม่น้อยกว่าค่าสนับสนุนขั้นต่ำ (L_1) มาเรียงลำดับตามค่าความถี่ของแต่ละชั้นข้อมูลจากมากไปหาน้อยแล้วนำมาสร้างตาราง Header (Header Table)

2) อ่านข้อมูลจากฐานข้อมูลครั้งที่สองเพื่อสร้างต้นไม้ FP-Tree โดยอ่านข้อมูลจากฐานข้อมูลที่ละรายการข้อมูล จากนั้นคัดชั้นข้อมูลในรายการข้อมูลนั้นที่ไม่ปรากฏอยู่ในตาราง Header ทิ้งไป แล้วเรียงชั้นข้อมูลที่เหลือตามลำดับในตาราง Header แล้วนำชั้นข้อมูลดังกล่าวไปสร้างโหนด (Node Tree) เพิ่มเข้าไปในต้นไม้ FP-Tree แล้วเชื่อมแต่ละโหนดที่เป็นชั้นข้อมูลเดียวกันเพิ่มเข้าไปกับตาราง Header

3) สร้าง Conditional pattern base และสร้าง Conditional FP-Tree ของแต่ละชั้นข้อมูล เพื่อใช้ในขั้นตอนการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย โดยการพิจารณาจะเริ่มจากชั้นข้อมูลต่ำสุดจนถึงชั้นข้อมูลที่อยู่บนสุดในตาราง Header ซึ่ง Conditional pattern base

หมายถึงเซตของชั้นข้อมูลที่เกิดขึ้นพร้อมกับชั้นข้อมูลที่กำลังพิจารณาในแต่ละเส้นทาง (Path Tree) และกำหนดให้ทุกชั้นข้อมูลมีค่าความถี่เท่ากับค่าความถี่ของชั้นข้อมูลที่กำลังพิจารณาจากต้นไม้ FP-Tree หลังจากนั้นสร้างต้นไม้ FP-Tree บน Conditional pattern base นี้ เรียกว่า Conditional FP-Tree ซึ่งเกิดจากการนำค่าความถี่ของแต่ละชั้นข้อมูลในทุกเส้นทางมารวมกัน และเลือกเฉพาะชั้นข้อมูลที่ผ่านค่าสนับสนุนขั้นต่ำจาก Conditional FP-Tree เพื่อนำไปสร้างกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยต่อไป

4) ค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยจากการสร้าง Conditional pattern base และสร้าง Conditional FP-Tree ของแต่ละชั้นข้อมูล โดยใช้หลักการงานแบบแบ่งแยกแล้วเอาชนะ (Divide and Conquer)

Input: FP-tree constructed based on Algorithm 1, using *DB* and a minimum support threshold ξ .

Output: The complete set of frequent patterns.

Method: Call FP-tree (FP-tree, null).

Procedure FPF-tree (FP-tree, α)

```
{
1  IF Tree contains a single path P
2  THEN for each combination (denoted as  $\beta$ ) of the nodes in the path P do
3     generate pattern  $\beta \cup \alpha$  with support = minimum support of node in  $\beta$ ;
4  ELSE for each  $a_i$  in the header of Tree do {
5     generate pattern  $\beta = a_i \cup \alpha$  with support =  $a_i$ .support;
6     construct  $\beta$ 's conditional pattern base and then  $\beta$ 's conditional FP-tree
       Tree $_{\beta}$ ;
7     If Tree $_{\beta} \neq \emptyset$ 
8     THEN call FP-Tree (Tree $_{\beta}$ ,  $\beta$ )    }
}
```

ภาพที่ 2.20 ขั้นตอนการทำงานของขั้นตอนวิธี FP-Growth

ที่มา: (Han *et al.*, 2006)

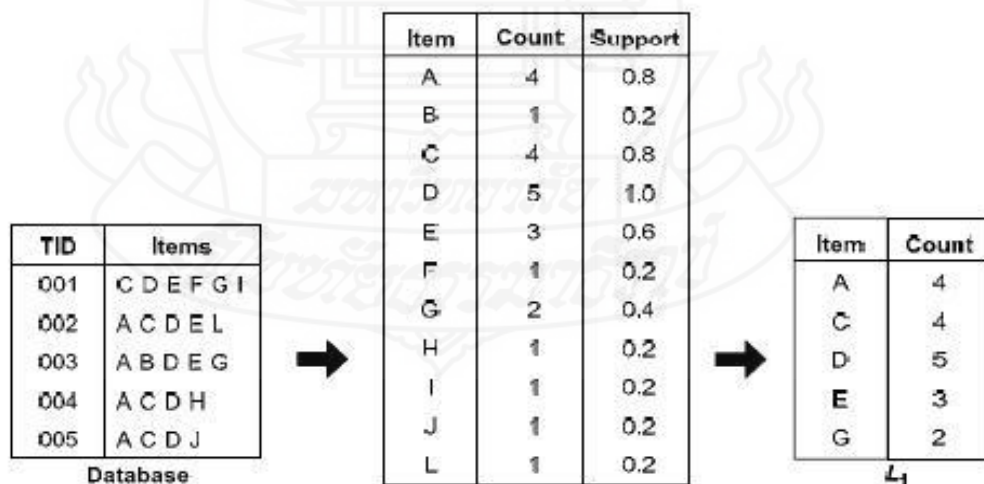
ตารางที่ 2.4 ฐานข้อมูลการซื้อสินค้า

รหัสรายการ (TID)	ชิ้นข้อมูล (Items)
001	C D E F G I
002	A C D E L
003	A B D E G
004	A C D H
005	A C D J

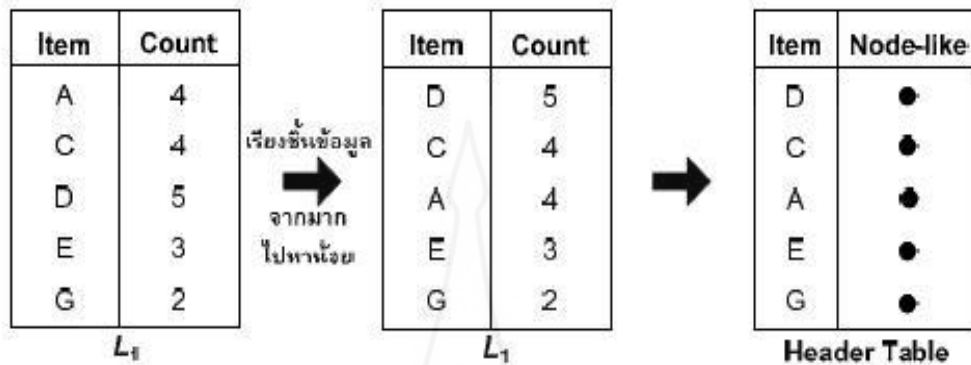
ที่มา (Han *et al.*, 2006)

5) ตัวอย่างการทำงานของขั้นตอนวิธี FP-Growth เพื่อให้เข้าใจถึงกระบวนการทำงานของขั้นตอนวิธี FP-Growth ได้อย่างชัดเจนขึ้น จึงขอยกตัวอย่างการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยของขั้นตอนวิธีนี้ โดยใช้ฐานข้อมูลการซื้อสินค้าในตารางที่ 2.4 ประกอบการอธิบาย พร้อมทั้งกำหนดค่าสนับสนุนขั้นต่ำเป็น 0.4 (มีค่าความถี่อย่างน้อยเท่ากับ 2 รายการข้อมูลปรากฏในฐานข้อมูล) ดังนี้

(1) อ่านข้อมูลจากฐานข้อมูลครั้งแรกเพื่อนับค่าความถี่ของแต่ละชิ้นข้อมูล และคำนวณค่าความถี่เพื่อกำจัดชิ้นข้อมูลที่ปรากฏไม่บ่อย ซึ่งผลลัพธ์ที่ได้คือ L_1 ดังภาพประกอบ 2.21

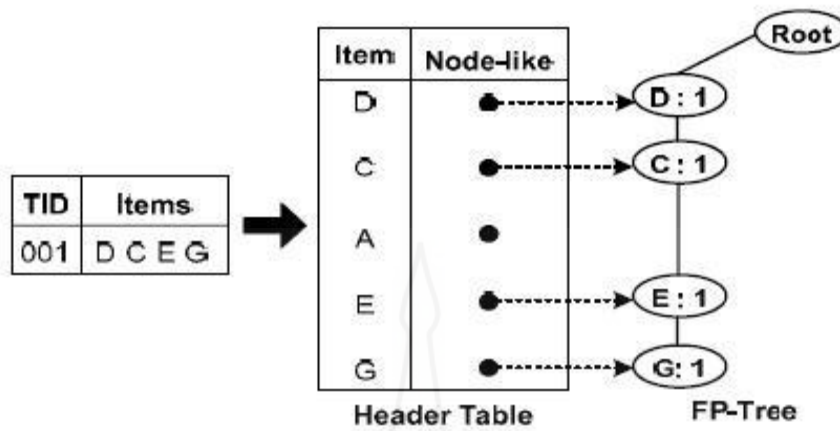
ภาพที่ 2.21 การอ่านข้อมูลจากฐานข้อมูลเพื่อค้นหา L_1 ของขั้นตอนวิธี FP-Growthที่มา: (Han *et al.*, 2006)

(2) จัดเรียง L_1 ใหม่ตามค่าความถี่ของแต่ละชั้นข้อมูลจากมากไปหาน้อยเพื่อสร้างตาราง Header ดังภาพที่ 2.22

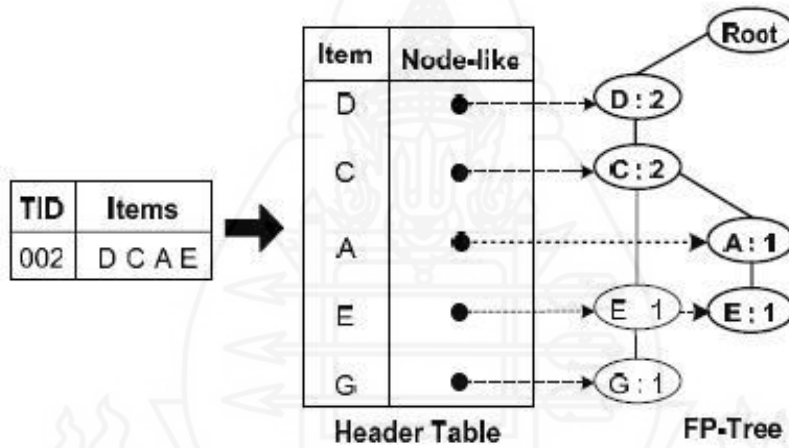


ภาพที่ 2.22 การสร้างตาราง Header ของขั้นตอนวิธี FP-Growth (Han *et al.*, 2006)

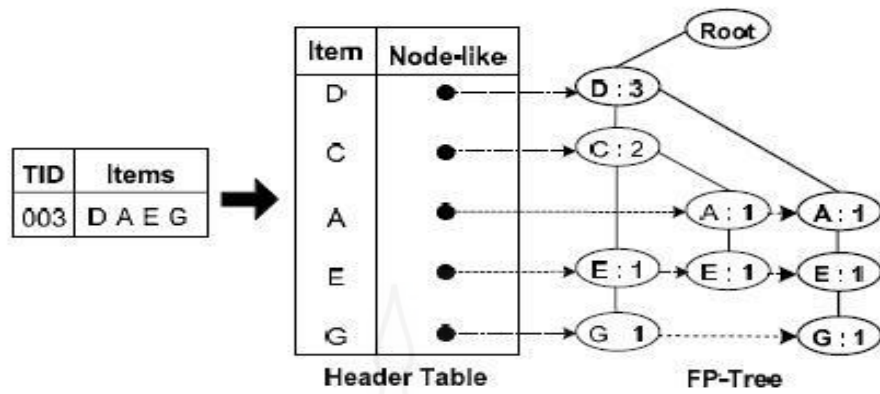
(3) อ่านข้อมูลจากฐานข้อมูลครั้งที่สอง โดยเริ่มอ่านชั้นข้อมูลทั้งหมดจากรายการข้อมูลแรกในฐานข้อมูลซึ่งประกอบด้วยชั้นข้อมูล C D E F G และ I จากนั้นตัดชั้นข้อมูลที่ไม่ปรากฏในตาราง Header ออกไป แล้วเรียงลำดับชั้นข้อมูลที่เหลือใหม่ตามลำดับในตาราง Header จะได้ลำดับของชั้นข้อมูลของรายการข้อมูลแรกคือ D C E และ G จากนั้นนำชั้นข้อมูลที่ได้ไปสร้างโหนดของชั้นข้อมูลเพิ่มเข้าไปในต้นไม้ FP-Tree แล้วเชื่อมแต่ละโหนดที่เพิ่มเข้าไปกับตาราง Header ผลลัพธ์แสดงดังภาพประกอบ 2.16 โดยตัวเลขที่อยู่หลังเครื่องหมาย “:” ในแต่ละโหนดหมายถึงค่าความถี่ของกลุ่มข้อมูล และทำตามขั้นตอนข้างต้นนี้กับทุกรายการข้อมูลในฐานข้อมูล จะสามารถแสดงผลลัพธ์ของการเพิ่มโหนดชั้นข้อมูลของแต่ละรายการข้อมูลเข้าไปในต้นไม้ FP-Tree ดังภาพประกอบที่ 2.23 ถึง 2.28



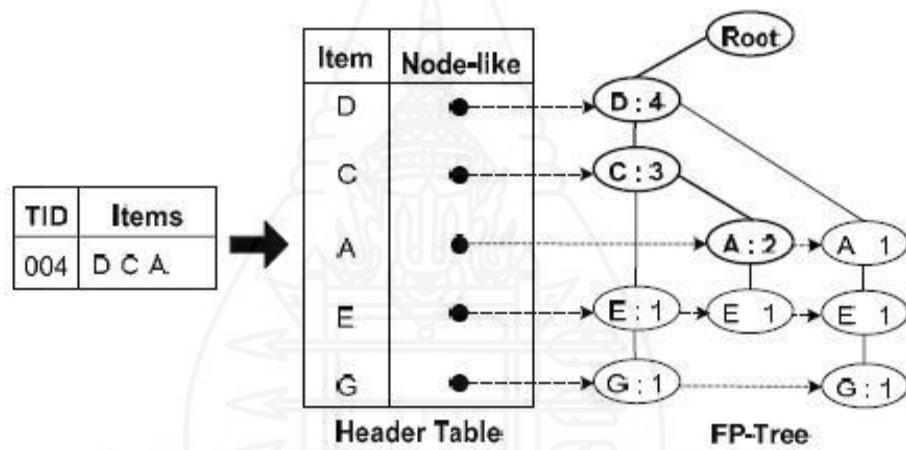
ภาพที่ 2.23 การอ่านรายการข้อมูลแรกจากฐานข้อมูลของขั้นตอนวิธี FP-Growth
ที่มา: (Han *et al.*, 2006)



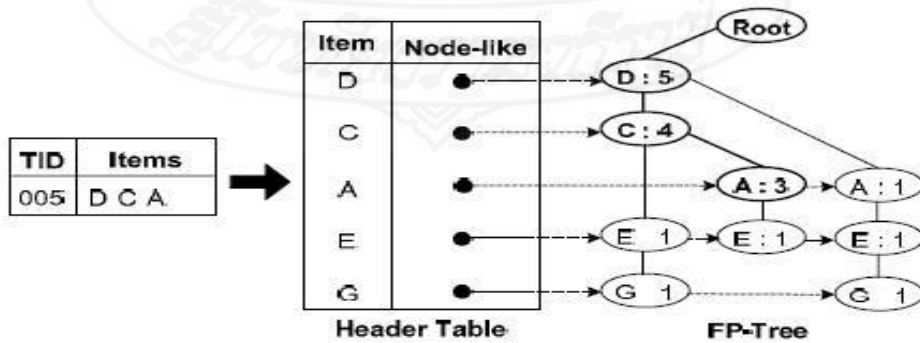
ภาพที่ 2.24 การอ่านรายการข้อมูลที่สองจากฐานข้อมูลของขั้นตอนวิธี FP-Growth
ที่มา: (Han *et al.*, 2006)



ภาพที่ 2.25 การอ่านรายการข้อมูลทีสามจากฐานข้อมูลของขั้นตอนวิธี FP-Growth ที่มา: (Han et al., 2006)

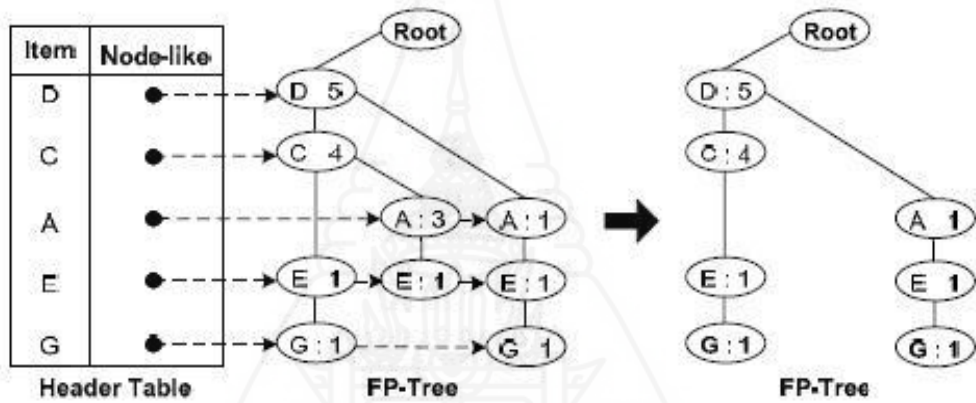


ภาพที่ 2.26 การอ่านรายการข้อมูลทีสี่จากฐานข้อมูลของขั้นตอนวิธี FP-Growth ที่มา: (Han et al., 2006)



ภาพที่ 2.27 การอ่านรายการข้อมูลทีห้าจากฐานข้อมูลของขั้นตอนวิธี FP-Growth ที่มา: (Han et al., 2006)

(4) สร้าง Conditional pattern base และสร้าง Conditional FP-Tree โดยเริ่มพิจารณาจากชั้นข้อมูลสุดท้ายในตาราง Header Table นั่นคือชั้นข้อมูล G ซึ่งจากภาพที่ 2.28 จะเห็นได้ว่าต้นไม้ FP-Tree มีโหนด G ปรากฏอยู่ 2 เส้นทางได้แก่เส้นทาง D C E มีค่าความถี่เป็น 1 (เกิดร่วมกับชั้นข้อมูล G จำนวน 1 ครั้ง) และเส้นทาง D A E มีค่าความถี่เป็น 1 ดังนั้นจะได้ Conditional pattern base ของชั้นข้อมูล G คือ {(DCE:1),(DAE:1)} ซึ่งจากทั้ง 2 เส้นทางจะเห็นว่าชั้นข้อมูล D และ E ปรากฏร่วมกันกับชั้นข้อมูล G เหมือนกันทั้ง 2 เส้นทาง ดังนั้นจะสร้าง Conditional FP-Tree ได้เป็น {(DE:2)} | G และทำเช่นนี้กับทุกชั้นข้อมูลในตาราง Header จะสามารถหา Conditional pattern base และสร้าง Conditional FP-Tree ของทุกชั้นข้อมูลได้ดังแสดงในตารางที่ 2.5



ภาพที่ 2.28 การสร้างตาราง Header ของขั้นตอนวิธี FP-Growth

ที่มา: (Han et al., 2006)

ตารางที่ 2.5 ค่า Conditional pattern base และค่า Conditional FP-Tree ของขั้นตอนวิธี FP-Growth

Items	Conditional pattern base	Conditional FP-Tree
G	{(DCE:1),(DAE:1)}	{(DE:2)} G
E	{(DC:1),(DCA:1),(DA:1)}	{(D:3),(DC:2),(DA:2)} E
A	{(DC:3),(D:1)}	{(D:4),(DC:3)} A
C	{(D:4)}	{(D:4)} C
D	∅	∅

ที่มา: (Han et al., 2006)

(5) ค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยจากการสร้าง Conditional pattern base และสร้าง Conditional FP-Tree โดยเริ่มพิจารณาจากชั้นข้อมูล G ซึ่งจากตารางที่ 2.5 จะเห็นว่าชั้นข้อมูล G มี Conditional pattern base เป็น $\{(DCE:1),(DAE:1)\}$ และ Conditional FP-Tree เป็น $\{(DE:2)\} | G$ ทำให้การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยของชั้นข้อมูล G แบ่งการทำงานของ $\{(DE:2)\} | G$ เป็น 2 ส่วนคือ EG:2 และ DG:2 โดยส่วนแรก EG:2 จะแยกได้อีกเป็น $\{(D:2)\} | EG$ ซึ่งสุดท้ายจะได้เป็น DEG:2 และส่วนที่สอง DG:2 ไม่สามารถแบ่งได้อีกดังนั้นจะได้กลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยของชั้นข้อมูล G เป็น G:2 EG:2 DG:2 และ DEG:2 ผลลัพธ์ของการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด แสดงดังตารางที่ 2.6

ตารางที่ 2.6 ผลลัพธ์ของการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยของขั้นตอนวิธี FP-Growth

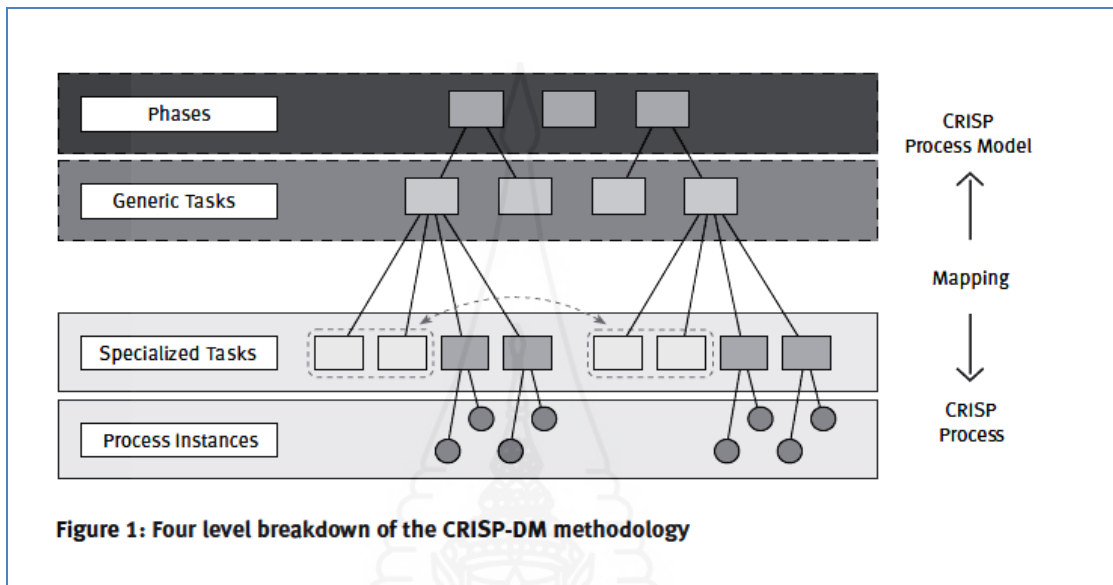
Items	Frequent Itemsets
G	G:2, EG:2, DG:2, DEG:2
E	E:3, DE:3, CE:2, AE:2, DCE:2, DAE:2
A	A:4, DA:4, CA:3, DCA:3
C	C:4, DC:4
D	D:5

ที่มา: (Han *et al.*, 2006)

4 หลักการของคริสป์-ดีเอ็ม

การจัดทำเหมืองข้อมูลมีขั้นตอนหรือกระบวนการทำเหมืองข้อมูล ซึ่งมีประกอบด้วยรายละเอียดต่างๆ ในการดำเนินการแต่ละขั้นตอนอาจจะแตกต่างกันไป ดังนั้นเพื่อให้มีแนวทางการทำงานร่วมกันอย่างชัดเจน จึงได้มีกำหนดกระบวนการที่เป็นมาตรฐานในการดำเนินการ ซึ่งมาตรฐานที่วันนี้ เรียกว่า คริสป์-ดีเอ็ม (CRISP-DM – Cross Industry Standard Process for Data Mining) ก่อเกิดขึ้นด้วยเหตุผลเพื่อให้มีขั้นตอนมาตรฐาน ที่มีกรอบงานแต่ละส่วนช่วยให้ผู้ทำเหมืองข้อมูลใช้เป็นแนวทาง เพื่อช่วยให้ทำงานมีประสิทธิภาพขึ้น

4.1 การแบ่งงานเป็นลำดับชั้น ในการกำหนดวิธีของการทำเหมืองข้อมูลจะใช้หลักการพื้นฐาน โดยการแบ่งงานเป็นลำดับชั้น (Hierarchical Breakdown) เป็นกลุ่มงานย่อย (Set of Tasks) 4 ระดับ ดังภาพที่ 2.29



ภาพที่ 2.29 การแบ่งระดับงานของคริสป์-ดีเอ็ม

ที่มา : <http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> June 30,2014

4.1.1 **ระดับที่ 1 งานหลัก (Phase)** เป็นส่วนบนสุด ประกอบด้วยงานย่อยต่างๆ ที่ต้องดำเนินการในการทำเหมืองข้อมูล ซึ่งเป็นงานทั่วไป (Generic Task) ทั้งหมด เพื่อให้เห็นภาพรวมของงานที่มีอยู่

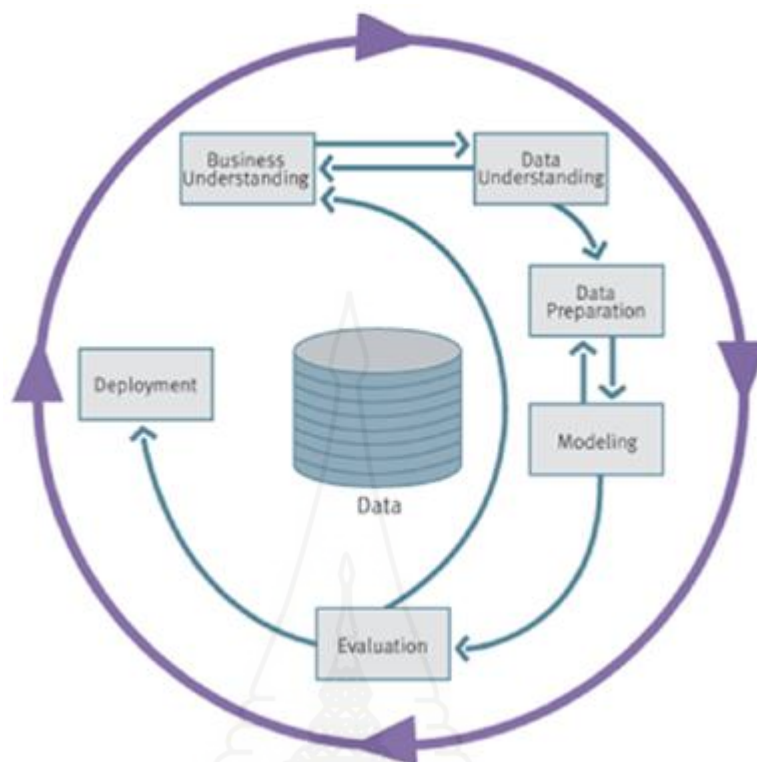
4.1.2 **ระดับที่ 2 งานทั่วไป (Generic Task)** ในส่วนนี้จะเป็นการกำหนดว่างานแต่ละงานจะทำเหมืองข้อมูลด้วยเทคนิคอะไรบ้าง ซึ่งบางครั้งอาจมีปัญหาคาดไม่ถึงเกิดขึ้นในระหว่างการทำงานก็ตาม แต่ก็ให้เห็นภาพของงานที่จะต้องทำได้ในระดับหนึ่ง และจะเชื่อมโยงกับระดับถัดไปได้ ตัวอย่างเช่น ระบุว่ามีการทำความสะอาดข้อมูล (Clean Data) จะต้องมีการกำหนดเทคนิคการทำเหมืองข้อมูลที่คาดว่าจะเหมาะสมสำหรับงานนั้น

4.1.3 **ระดับที่ 3 งานเฉพาะ (Specific Task)** ในส่วนนี้เป็นการขยายรายละเอียดของงานย่อยในระดับ 2 ว่าแต่ละงานย่อยมีวิธีการทำอะไรบ้างในแต่ละสถานการณ์ เช่น การทำความสะอาดข้อมูลที่เป็นตัวเลข ดังนั้นหากมีข้อมูลประเภทอื่นที่ไม่ใช่ตัวเลขแล้วจะมีวิธีการจัดการ

อย่างไร หรือตัวอย่างจากระดับที่สอง ระบบเทคนิคการทำเหมืองข้อมูลให้ชัดเจนว่าจะใช้เทคนิคใด เช่น การจัดกลุ่ม (Clustering) การจำแนกประเภท (Classification)

4.1.4 ระดับที่ 4 รายละเอียดวิธีการ (Process Instance) เป็นข้อมูลที่กำหนดรายละเอียดแต่ละงานย่อยในระดับที่เหนือกว่าให้มีรายละเอียดมากขึ้น ซึ่งจะเป็นรายละเอียดเกี่ยวกับการดำเนินการ (Action) ของแต่ละงานย่อย เงื่อนไข (Decision) ผลลัพธ์ที่เกิดขึ้น (Result of The Actions) ของงานย่อยแต่ละอัน

4.2 ขั้นตอนของ คริสป์-ดีเอ็ม แบบจำลอง CRISP-DM (Cross- Industry Standard Process for Data Mining) เป็นแบบจำลองแสดงวงจรชีวิตของโครงการทำเหมืองข้อมูลหรือการค้าไมน์นิ่ง (วิชา เจริญภัณฑาทาร์กซ์, 2555) ดังแสดงในภาพที่ 2.30 โดยแบบจำลอง CRISP-DM จะครอบคลุมระยะเวลาหรือเฟส (Phase) การทำงานในโครงการตามลำดับ ของความสัมพันธ์ระหว่างงานทั้งหมดในโครงการ โดยวงจรชีวิตของโครงการทำเหมืองข้อมูล มีด้วยกัน 6 ระยะเวลาหรือเฟส โดยลำดับของงานในแต่ละเฟสไม่ตายตัว การเคลื่อนย้ายไปข้างหน้าหรือถอยหลังระหว่างเฟสสามารถกระทำได้ตลอดเวลา ผลที่ได้รับจากแต่ละเฟส ทำให้ทราบว่าต้องดำเนินการเฟสใดต่อไป ลูกศรที่แสดงในภาพชี้ให้เห็นความสำคัญและการขึ้นต่อกันระหว่างเฟสต่างๆ วงกลมที่อยู่รอบนอกเป็นสัญลักษณ์ของการทำเหมืองข้อมูลซึ่งต้องวนเป็นวงกลม โดยการทำเหมืองข้อมูลจะไม่สิ้นสุดเมื่อมีการนำผลไปใช้ แต่สิ่งที่ได้รับหรือเรียนรู้ในระหว่างกระบวนการและจากการนำผลไปใช้จะสามารถก่อให้เกิดสิ่งใหม่ๆ ซึ่งมักจะมุ่งเน้นเพื่อตอบคำถามให้แก่การดำเนินงานของหน่วยงานได้ โดยผลที่ได้รับจากกระบวนการทำเหมืองข้อมูลจะเป็นประโยชน์มากเมื่อนำประสบการณ์ที่ได้รับจากการวนซ้ำของกระบวนการนี้มาประกอบด้วยในกระบวนการทำเหมืองข้อมูลแบบคริสป์-ดีเอ็ม ได้กำหนดไว้ 6 ขั้นตอน



ภาพที่ 2.30 แบบจำลอง CRISP-DM

ที่มา : <http://www.crisp-dm.org/Process/index.htm> June 30, 2014

4.2.1 ความเข้าใจธุรกิจ เป็นการวิเคราะห์ความต้องการธุรกิจ จะช่วยให้เข้าใจสถานะของธุรกิจในปัจจุบัน และประเด็นความต้องการที่ผู้บริหารนำมาใช้ตัดสินใจเพื่อช่วยให้การดำเนินธุรกิจประสบความสำเร็จ และตัดสินใจได้ดียิ่งขึ้น เช่น ความต้องการเกี่ยวกับพฤติกรรมลูกค้า เพื่อพยากรณ์ว่าลูกค้ารายใดน่าจะซื้อสินค้าในรายการสินค้าใหม่ ทำให้สามารถวางแผนธุรกิจได้ตรงกับความต้องการ

4.2.2 ความเข้าใจข้อมูล เป็นการวิเคราะห์ความต้องการข้อมูลที่เกี่ยวข้อง เนื่องจากแหล่งข้อมูล ที่นำมาจัดทำเหมืองข้อมูลนั้นมีมากมาย และจัดเก็บในคลังข้อมูลหรือฐานข้อมูลขนาดใหญ่ มีข้อมูลที่หลากหลายซึ่งมีทั้งข้อมูลที่จำเป็นต้องใช้ในการทำเหมืองข้อมูล ปะปนกับข้อมูลอื่นซึ่งไม่ต้องการในการทำเหมืองข้อมูลในบางครั้ง จึงต้องมีขั้นตอนการกำหนดรายการและประเภทของข้อมูลที่จะนำมาใช้ โดยมีการตรวจสอบในด้านของคุณภาพของข้อมูล จำนวน ปริมาณเนื้อหา และการเข้าถึงข้อมูล เพื่อกำหนดเป็นข้อมูลที่ตรงกับความต้องการทำเหมืองข้อมูล ซึ่งอาจจะมีการคัดเลือก โดยการในคำสั่งสอบถามข้อมูลจากคลังข้อมูลที่มีข้อมูลจำนวนมาก โดยเลือกตัวอย่างข้อมูลมาทำเหมืองก่อน ซึ่งช่วยลดค่าใช้จ่าย และประหยัดเวลาในการดำเนินการได้

4.2.3 การเตรียมข้อมูล เป็นขั้นตอนเพื่อทำการแปลงหรือปรับเปลี่ยนข้อมูล ให้อยู่ในรูปแบบที่เหมาะสมกับวิธีการที่จะใช้ในการทำเหมืองข้อมูล เช่น ข้อมูล วัน/เดือน/ปี ที่มาจากแต่ละฐานข้อมูลมีรูปแบบแตกต่างกัน ได้แก่ 9 สิงหาคม 2556 9/08/56 08/9/2013 ในการจัดเตรียมข้อมูลจึงต้องแปลงให้มีรูปแบบเดียวกันเพื่อให้ใช้งานร่วมกันได้ โดยปกติแล้วการแปลงข้อมูลจะถูกกำหนดโดยเงื่อนไขของการปฏิบัติงาน และวิธีการทำเหมืองข้อมูลที่ต้องจัดเตรียมข้อมูลไว้ล่วงหน้าให้มีรูปแบบที่พร้อมใช้งาน

4.2.4 การจัดทำแบบจำลอง ในขั้นนี้เป็นการใช้เทคนิคการทำเหมืองข้อมูล แต่ละเทคนิคที่มีรูปแบบและวิธีการในการทำแตกต่างกัน เพื่อสร้างแบบจำลองขึ้นมา ซึ่งแบบจำลองที่ได้จะนำช่วยสกัดสาระสำคัญหรือความรู้ที่ซ่อนเร้นอยู่อกจากฐานข้อมูล เช่น การค้นหาคำตอบว่าลูกค้าจะซื้อสินค้าหรือไม่ อาจต้องทำการวิเคราะห์โดยใช้เทคนิคการจัดกลุ่ม (Clustering) ของลูกค้า หรือเทคนิคการจำแนกประเภทลูกค้าแต่ละคน (Classification) ว่าจะซื้อหรือไม่ซื้อสินค้าในการสร้างแบบจำลองด้วยเทคนิคต่างๆ ในการทำเหมืองข้อมูล จะแบ่งกลุ่มข้อมูลที่เกี่ยวข้องคือ ข้อมูลที่ใช้ในการสร้างแบบจำลอง (Training Data Set) และกลุ่มข้อมูลที่ใช้ในการทดสอบแบบจำลอง (Testing Data Set) เพื่อตรวจสอบความเที่ยงตรงหรือความน่าเชื่อถือของตัวแบบจำลองที่สร้างขึ้น การใช้เทคนิคต่างๆ ทำให้ได้แบบจำลองที่ช่วยตอบคำถามที่เกิดขึ้น ทั้งนี้ในขณะที่ทำเหมืองข้อมูล อาจมีความจำเป็นต้องเข้าถึงข้อมูลอื่นในคลังข้อมูล รวมทั้งการแปล หรือทดลองนำผลไปประยุกต์ใช้กับธุรกิจเพื่อทดสอบความน่าเชื่อถือ มีการนำเสนอสารสนเทศที่ทำเหมืองข้อมูลที่ได้มาวิเคราะห์เพื่อตอบคำถามที่ผู้ตัดสินใจกำลังค้นหาคำตอบ ซึ่งการวิเคราะห์ในส่วนนี้จะครอบคลุมการกรองสารสนเทศที่เหมาะสมกับผู้ใช้งานและการแปลผล

4.2.5 การประเมินผล การประเมินผลลัพธ์จากการทำเหมืองข้อมูลเกี่ยวข้องกับ 2 ประเด็น คือ

1) การประเมินความรู้ที่ค้นพบจากการทำเหมืองข้อมูล ว่าตรงกับที่ต้องการหรือไม่ หรือครอบคลุมประเด็นปัญหาที่มีหรือไม่

2) การประเมินวิธีการนำเสนอผลลัพธ์ในรูปแบบที่เหมาะสมและง่ายต่อการใช้งาน เช่น แผนภูมิต่างๆ ในเชิงรูปภาพ ถ้าผลลัพธ์ที่ได้ไม่เป็นที่พอใจ ก็จะต้องทำขั้นตอนนี้ซ้ำอีก รวมทั้งขั้นตอนก่อนหน้าด้วย หากเป็นคำตอบที่พอใจหรือน่าเชื่อถือ ก็จะสามารถดำเนินการขั้นต่อไป

4.2.6 การนำแบบจำลองนำไปใช้งาน เป็นขั้นตอนสุดท้ายในการทำเหมืองข้อมูล หลังจากการตรวจสอบจนแน่ใจว่าแบบจำลองที่สร้างขึ้นมาน่าเชื่อถือ ผู้ใช้งานซึ่งเป็นผู้ทำงานเกี่ยวข้องกับการวิเคราะห์และตัดสินใจ สามารถนำไปใช้งานได้

5. โปรแกรมเวก้า (WEKA)

โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) : เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จ ภาพประกอบประเภทฟรีแวร์ อยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม WEKA ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และ ใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน (Run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำดาต้าไมน์นิ่งที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกได้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization

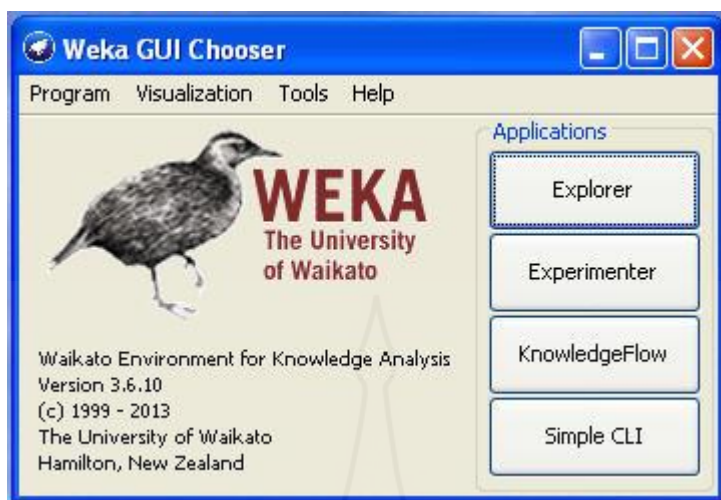
ที่มา : <http://cheekswellilive.wordpress.com> July 26, 2014



ภาพที่ 2.31 โลโก้ของซอฟต์แวร์ WEKA เป็นรูปนกเฉพาะถิ่นของประเทศนิวซีแลนด์

ที่มา : [http://en.wikipedia.org/wiki/Weka_\(machine_learning\)#mediaviewer/File:Weka_\(software\)_logo.png](http://en.wikipedia.org/wiki/Weka_(machine_learning)#mediaviewer/File:Weka_(software)_logo.png)

July 26, 2014



ภาพที่ 2.32 หน้าการหลักของซอฟต์แวร์ WEKA

ที่มา : <http://nuiza.files.wordpress.com/2012/03/d2.jpg> July 26, 2014

5.1 รายละเอียดเมนูหลักในโปรแกรม WEKA หน้าต่าง WEKA GUI Chooser มีเมนูในการทำงานประกอบไปด้วย

5.1.1 Explorer เป็นส่วนที่ผู้ใช้สามารถใช้ฟังก์ชันการทำงานต่างๆ ของ WEKA ผ่านทางหน้าจอ GUI ซึ่งเป็นส่วนที่เหมาะสมสำหรับผู้เริ่มต้นใช้งาน WEKA เพราะผู้ใช้จะสามารถเรียกฟังก์ชันการทำงานต่างๆ ได้เพียงแค่คลิกและเปลี่ยนแปลงพารามิเตอร์ในหน้าฟอร์มเท่านั้น ต้องยอมรับว่าส่วนนี้เป็นจุดขายอีกอย่างของ WEKA ที่ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลได้อย่างสะดวกมากยิ่งขึ้น

5.1.2 Experimenter เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถทดลองเปลี่ยนแปลงเทคนิคในการวิเคราะห์ข้อมูลและค่าพารามิเตอร์ต่างๆ ที่เกี่ยวข้องได้หลากหลายรูปแบบ จนได้ผลลัพธ์ที่น่าพอใจ ซึ่งการใช้ Explorer นั้นไม่สะดวกเนื่องจากต้องมาทำการเปลี่ยนแปลงค่าต่างๆ เอง

5.1.3 Knowledge Flow เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถนำเทคนิคต่างๆ ของ WEKA มาเรียงต่อกันเพื่อช่วยให้การวิเคราะห์ข้อมูลทำงานได้ตามที่ต้องการ

5.1.4 Simple CLI เป็นส่วนที่ผู้ใช้สามารถเรียกฟังก์ชันการทำงานของ WEKA มาใช้ผ่านทาง command line ได้ซึ่งการเรียกใช้ฟังก์ชันผ่านทาง command line นี้จะช่วยให้ผู้ใช้เข้าใจการเรียกฟังก์ชันต่างๆ เมื่อหลังหน้าจอ GUI ของ WEKA ได้และสามารถนำไปประยุกต์ใช้ในการเขียนโปรแกรมเพื่อเรียก WEKA ในการใช้งานได้อีกด้วย

5.2 ข้อดีของโปรแกรม WEKA คือ มีอัลกอริทึมที่รู้จักกันดีของการทำค้ำไม้หนึ่งให้เลือกใช้อย่างครบถ้วน และสามารถเขียนฟังก์ชันเพิ่มเข้าไปในโปรแกรมเองได้ โดยสรุปดังนี้

- 5.2.1 เป็นซอฟต์แวร์เสรีที่สามารถดาวน์โหลดได้ฟรี โดยสามารถใช้ได้ทุกระบบปฏิบัติการ
- 5.2.2 สามารถเชื่อมต่อข้อมูล SQL Database โดยใช้ Java Database Connectivity
- 5.2.3 การเตรียมข้อมูลและเทคนิคในการสร้างแบบจำลองที่ครอบคลุม มีลักษณะที่ง่ายต่อการใช้งาน
- 5.2.4 สนับสนุนเกี่ยวกับการทำเหมืองข้อมูล (Data Mining)
- 5.2.5 การเตรียมข้อมูล (Data Preprocessing)
- 5.2.6 การทำเหมืองข้อมูลด้วยเทคนิคการจำแนกข้อมูล (Classification)
- 5.2.7 การทำเหมืองข้อมูลด้วยเทคนิคการจัดกลุ่ม (Clustering)
- 5.2.8 การทำเหมืองข้อมูลด้วยเทคนิคการวิเคราะห์ความสัมพันธ์ (Associating)
- 5.2.9 เทคนิคการคัดเลือกข้อมูล (Selecting Attributes)
- 5.2.10 เทคนิคการนำเสนอข้อมูลด้วยรูปภาพ (Visualization)

6. งานวิจัยที่เกี่ยวข้อง

จรรยาศักดิ์ ศรีอ่อน (2552) ได้ศึกษาเกี่ยวกับเรื่อง การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้องโดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง ระบบจำหน่ายส่วนใหญ่ของการไฟฟ้าส่วนภูมิภาค เป็นระบบเหนือนดิน พาดไปตามแนวถนนทั่วประเทศ จึงมีปัจจัยเสี่ยงที่ส่งผลให้เกิดปัญหากระแสไฟฟ้าขัดข้องได้บ่อยครั้งกว่าระบบใต้ดิน โดยเหตุการณ์กระแสไฟฟ้าขัดข้องมักมีสาเหตุและระยะเวลาแตกต่างกันไปในแต่ละพื้นที่ การไฟฟ้าส่วนภูมิภาคได้ดำเนินการจัดเก็บข้อมูลกระแสไฟฟ้าขัดข้อง ด้วยโปรแกรม จฟ.3 ตั้งแต่ปีพ.ศ. 2537 เพื่อใช้ในการคำนวณดัชนีความเชื่อถือได้ คือค่าดัชนี SAIFI SAIDI และ MAIFI เพื่อบอกถึงคุณภาพของการจ่ายไฟฟ้าให้กับผู้ใช้ไฟฟ้าและใช้ในการวางแผนระบบไฟฟ้า นอกจากนี้ยังมีการนำข้อมูลดังกล่าวไปใช้ในโครงการวิจัยเพื่อปรับปรุงด้านต่างๆของการไฟฟ้าส่วนภูมิภาค ข้อมูลเหตุการณ์กระแสไฟฟ้าขัดข้องของการไฟฟ้าส่วนภูมิภาค ปีละประมาณ 30% ไม่สามารถระบุสาเหตุที่ชัดเจนของการเกิดเหตุการณ์กระแสไฟฟ้าขัดข้องได้ จึงใช้การสันนิษฐานสาเหตุจากประสบการณ์ของผู้บันทึกในการเกิดเหตุการณ์แต่ละครั้ง ซึ่งอาจทำให้ข้อมูลมีความคลาดเคลื่อนจากความเป็นจริงและเกิดผลเสียเมื่อนำข้อมูลดังกล่าวไปใช้ในการวางแผนปรับปรุงความเชื่อถือได้

ในภายหลัง การศึกษาและใช้การทำเหมืองข้อมูล (Data Mining) ในการวิเคราะห์ข้อมูลเหตุการณ์ กระแสไฟฟ้าขัดข้อง เพื่อระบุสาเหตุที่ไม่ทราบ ทำให้ได้ข้อมูลกระแสไฟฟ้าขัดข้องมีความน่าเชื่อถือมากขึ้นและสามารถนำไปใช้ได้ดังต่อไปนี้ในการวิเคราะห์และวางแผนระบบไฟฟ้าของการไฟฟ้า ส่วนภูมิภาคต่อไปในอนาคต มีวัตถุประสงค์ 1) ศึกษาฐานข้อมูลเหตุการณ์ไฟฟ้าขัดข้องของ กฟภ. และสาเหตุต่างๆ ที่ทำให้เกิดเหตุการณ์กระแสไฟฟ้าขัดข้อง 2) ศึกษาวิธีการทำเหมืองข้อมูล ในส่วน ของเทคนิคการจำแนกประเภท 3) หาแบบจำลองที่มีประสิทธิภาพในการทำนายสาเหตุจาก 3 เทคนิคที่แตกต่างกัน 4) หาสาเหตุที่ทำให้เกิดกระแสไฟฟ้าขัดข้องในกรณีที่ไม่ทราบสาเหตุ โดยใช้ แบบจำลองที่เหมาะสมที่สุด ผลสรุปการวิจัย ผลจากการทดสอบแบบจำลองโดยการทดสอบกับชุด ข้อมูลทดสอบที่ทราบสาเหตุและ ไม่ทราบสาเหตุพบว่าความคลาดเคลื่อนส่วนใหญ่เกิดจาก ผลการ ทำนายของสาเหตุอื่นๆ กับสาเหตุ จากภัยธรรมชาติ การทดสอบแบบจำลองโดยการทดสอบกับชุด ข้อมูลที่ไม่ทราบสาเหตุ โดยตัดสาเหตุที่ สันนิษฐานออก ผลที่ได้สามารถแสดงค่าความน่าจะเป็น ของแต่ละสาเหตุ ทำให้สะดวกในการนำไปใช้ประกอบการพิจารณาเพื่อป้องกันเหตุการณ์ที่มีความ เป็นไปได้อย่างรวดเร็วในขณะที่เกิดเหตุการณ์ โดยการนำข้อมูลแวดล้อมที่เกี่ยวข้องมาใช้กับแบบจำลอง ที่ได้ ส่งผลให้แก้ไขได้ตรงสาเหตุและสามารถนำระบบคืนสู่สภาพเดิมในระยะเวลาอันรวดเร็ว จากผลการวิจัยจะเห็นได้ว่าการใช้ต้นไม้ตัดสินใจเหมาะสมกับการนำมาใช้ทำแบบจำลองเนื่องจาก ให้ค่าความถูกต้องสูงและมีค่าความคลาดเคลื่อนน้อย ดังนั้นการนำหลักการการทำเหมืองข้อมูลและ ผลที่ได้จากงานวิจัยมาประยุกต์ใช้กับการทำงานจะมีส่วนช่วย กฟภ. ในการบันทึกข้อมูล กระแสไฟฟ้าขัดข้องมีความแม่นยำถูกต้อง และมีหลักเกณฑ์การระบุสาเหตุมากขึ้น

อังคณา พิจาร โชติ (2552) ได้ทำการศึกษาเกี่ยวกับเรื่อง ระบบสนับสนุนการตัดสินใจ สำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น โรคเบาหวาน โดยใช้เทคนิคดาต้าไมนนิ่ง เพื่อช่วยในการ วิเคราะห์หา ความสัมพันธ์ของปัจจัยเสี่ยงต่างๆ ที่ก่อให้เกิดโรคเบาหวาน ข้อมูลความสัมพันธ์ของ ปัจจัยเสี่ยงที่ได้มาจากการวิเคราะห์จะเป็นสารสนเทศที่มีความสำคัญที่จะเป็นตัวช่วยให้หน่วยงาน ทางด้านการแพทย์นำไปใช้สำหรับการวางแผนป้องกันโรค โดยระบบที่พัฒนานี้จะประกอบด้วย 3 ส่วน คือ ส่วนของคลังข้อมูลที่ใช้สำหรับเก็บข้อมูลปัจจัยเสี่ยงที่ได้จากการคัดกรองกลุ่มเสี่ยง ส่วน ของเหมืองข้อมูลที่เป็นตัวค้นหาความสัมพันธ์ของปัจจัยเสี่ยงต่างๆ และส่วนการแสดงผล รายงานซึ่งจะจัดทำในรูปแบบของเว็บแอปพลิเคชันเพื่อสะดวกต่อการเรียกใช้งาน โดยใช้ข้อมูล การคัดกรองหลอดเลือดดี ที่คัดกรองผู้ป่วยในแต่ละ โรงพยาบาลส่งเสริมสุขภาพตำบล ในเขตพื้นที่ อำเภอนองบัวระเหว จังหวัดชัยภูมิ ผลการศึกษาโดยเริ่มจากการหาความสัมพันธ์ด้วยอัลกอริทึม และนำไปสร้างโปรแกรม ระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น โรคเบาหวาน ผลการค้นหากหาความสัมพันธ์ปัจจัยเสี่ยงที่ก่อให้เกิดโรคเบาหวานจะพบว่าประวัติ

ของครอบครัวเป็นปัจจัยเสี่ยงที่มีอัตราการเสี่ยงที่มากที่สุดซึ่งพบคู่กับค่าดัชนีมวลกาย นอกจากนี้ในการค้นหาความสัมพันธ์นั้น ผู้พัฒนาได้คัดเลือกกฎหรือความสัมพันธ์ที่ดีที่สุดเพียง 10 กฎ แต่ละกฎนั้นจะมีค่าความถูกต้องหรือค่าความเชื่อมั่นในระดับที่ดี ได้กำหนดค่าของความถูกต้องสูงสุดของข้อมูลต่ำสุดที่ 50% และ 100% ตามลำดับ ข้อเสนอแนะแนวทางในการศึกษาต่อ ควรพัฒนาระบบที่มีความสามารถในการสนับสนุนการตัดสินใจได้หลายโรค ควรมีแพทย์ผู้เชี่ยวชาญร่วมในการพัฒนาระบบ ควรมีการนำเทคนิคการสร้างเหมือนข้อมูลแบบอื่นเข้ามาใช้ในการค้นหาความสัมพันธ์ของปัจจัยเสี่ยง

เกวลี ผาใต้ (2553) ได้ทำการศึกษาเกี่ยวกับเรื่อง การพยากรณ์โอกาสเกิดภาวะเสี่ยงของผู้ป่วยเบาหวาน โดยใช้โครงข่ายประสาทเทียม ปัจจุบันผู้ป่วยเบาหวานมีแนวโน้มเพิ่มจำนวนขึ้นเรื่อยๆ และผู้ป่วยที่ป่วยเป็นโรคเบาหวานแล้วมักจะมีความเสี่ยงเกิดโรคแทรกซ้อนตามมา ซึ่งบางรายกว่าจะรู้ว่าตนเองนั้น มีอาการภาวะแทรกซ้อน ก็มักจะเกิดโรคแทรกซ้อนขึ้นแล้วและทำให้การรักษาทำได้ยาก การศึกษาอิสระฉบับนี้จึงนำเสนอวิธีการจำแนกข้อมูลผู้ป่วยเบาหวานซึ่งได้แก่เพศ อายุ ข้อมูลปัจจัยที่มีผลที่ทำให้มีความเสี่ยงต่อการเกิดภาวะแทรกซ้อน โดยประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) เพื่อวิเคราะห์ความเสี่ยงของการเกิดอาการของโรคแทรกซ้อน 2 โรค คือ โรคไตวายเรื้อรัง และโรคไขมันในเลือดสูง จากผลการทดลองพบว่าวิธีการที่นำเสนอมีความแม่นยำและสามารถนำไปใช้ในการประกอบการตัดสินใจของสาขาชีพในการตรวจรักษาอาการในระยะแรกเริ่มของโรคได้ ข้อมูลเป้าหมายที่ต้องการ กล่าว คือ ข้อมูลผู้ป่วยเบาหวานย้อนหลัง ดังนั้นในการศึกษาอิสระนี้จึงได้เตรียมข้อมูลผู้ป่วยเบาหวานย้อนหลังที่เก็บรวบรวมโดยโรงพยาบาลวชิรพยาบาล โดยกลุ่มข้อมูลตัวอย่างเป็นข้อมูลของผู้ป่วยจำนวน 3 เดือนระหว่างเดือนเมษายน 2553 ถึงเดือน มิถุนายน 2553 นอกจากนี้ในส่วนของปัจจัยนำเข้าที่คาดว่าจะมีความสัมพันธ์ ซึ่งส่งผลกระทบต่อพยากรณ์โอกาสเกิดภาวะเสี่ยงของผู้ป่วยเบาหวานซึ่งประกอบไปด้วย เพศ อายุ ดัชนีมวลกาย ความดันโลหิต ระดับน้ำตาลสะสม ระดับน้ำตาลในเลือด ค่าครีตินิน ค่าไตรกลีเซอไรด์ ค่าคลอเลสเทอรอล ค่าไขมันที่มีความหนาแน่นสูง ค่าไขมันที่มีความหนาแน่นต่ำ วัตถุประสงค์ของการวิจัย เพื่อศึกษาความสัมพันธ์ของปัจจัยเสี่ยงต่างๆ กับการเกิดภาวะเสี่ยงโรคแทรกซ้อนในผู้ป่วยเบาหวาน เพื่อสร้างแบบจำลองการคาดการณ์ภาวะเสี่ยงที่จะเกิดโรคแทรกซ้อนขึ้นในผู้ป่วยเบาหวาน โดยใช้แบบจำลองโครงข่ายประสาทเทียม โดยผลการศึกษาจะใช้ข้อมูลนำเข้าปัจจัยทั้งหมดของแต่ละปัจจัย ในการเกิดโรคแทรกซ้อน โดยปัจจัยที่เกี่ยวข้องกับโรคไตวายเรื้อรัง ได้แก่ เพศ อายุ ดัชนีมวลกาย ความดันโลหิต ระดับน้ำตาลสะสม ระดับน้ำตาลในเลือด ค่าครีตินิน และปัจจัยที่เกี่ยวข้องกับโรคไขมันในเลือดสูง ได้แก่ เพศ อายุ ดัชนีมวลกาย ค่า

ไตรกลีเซอไรด์ ค่าคลอเลสเตอรอล ค่าไขมันที่มีความหนาแน่นสูง ค่าไขมันที่มีความหนาแน่นต่ำ ข้อเสนอแนะ ในการศึกษาครั้งต่อไปอาจพิจารณาเพิ่มปัจจัยที่คาดว่าจะส่งผลต่อการพยากรณ์การเกิดภาวะแทรกซ้อนในผู้ป่วยเบาหวาน ซึ่งคาดว่าจะช่วยเพิ่มประสิทธิภาพการพยากรณ์ของตัวแบบโครงข่ายประสาทเทียมได้เนื่องจากการวัดประสิทธิภาพการพยากรณ์สามารถทำได้หลากหลายวิธีการ ดังนั้นในการศึกษาครั้งต่อไป ผู้ศึกษาจะมุ่งศึกษาเพื่อหาวิธีการที่ดีที่สุดเพื่อเพิ่มประสิทธิภาพในการวัดความถูกต้องของการพยากรณ์ต่อไป

Cunningham & Frank (1999) ได้ทำการศึกษาเกี่ยวกับเรื่อง Market basket analysis of library circulation data ซึ่งเป็นการวิเคราะห์ความสัมพันธ์ โดยใช้หลักการ Market basket analysis ที่ใช้ในการหาความสัมพันธ์ของสินค้าที่ถูกค้ำมักจะซื้อพร้อมกัน ซึ่งใช้ข้อมูลการยืม – คืนหนังสือช่วงปี 1998 จำนวน 52,518 เรคคอร์ด ของห้องสมุดมหาวิทยาลัยไวคาโต (University of Waikato) ประเทศนิวซีแลนด์มาทำการประมวลผล และใช้โปรแกรม WEKA เป็นเครื่องมือในการประมวลผลข้อมูล เพื่อค้นหาความสัมพันธ์ของหนังสือที่มักถูกยืมไปด้วยกัน ด้วยขั้นตอนวิธี Apriori โดยการหาความสัมพันธ์ของหมวดหมู่หนังสือ ซึ่งมีการจัดหมวดหมู่หนังสือแบบระบบการจัดหมวดหมู่หนังสือของหอสมุดรัฐสภาอเมริกัน(L.C.) ใช้ค่า Minimum support = 0.01 และค่า Minimum confidence = 0.01 การวิจัยเป็นการหาค่าความสัมพันธ์ของหมวดหมู่หนังสือ หัวเรื่องหนังสือ ที่มีข้อมูลเกิดขึ้นร่วมกันหรือมีความสัมพันธ์กันของหนังสือที่มีการยืมคืนจากห้องสมุด ผลของการวิจัย ได้นำไปใช้ประโยชน์การจัดวางตำแหน่งของหนังสือบนชั้นหนังสือให้เหมาะสมและตรงตามความต้องการของสารสนเทศของผู้ใช้บริการ เพื่อให้ผู้ใช้บริการเห็นหนังสือที่ต้องการโดยง่าย

Chien-Lung Chan (2008) ได้ทำการศึกษาเกี่ยวกับเรื่องการสอบสวนของภาวะแทรกซ้อนของเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล ใช้ทฤษฎีและการวิเคราะห์เชิงตัวเลขสำรวจความสัมพันธ์ระหว่างข้อมูลทางสรีรวิทยาและ ภาวะแทรกซ้อนของโรคหลอดเลือด ภาวะแทรกซ้อนของโรคไต ภาวะแทรกซ้อนของโรกระบบประสาท ภาวะแทรกซ้อนจอประสาทตา และปัญหาเท้าพิการ ผลการวิเคราะห์ของผู้ป่วยโรคเบาหวาน 8,736 ในภาคเหนือของไต้หวัน โดยใช้การวิเคราะห์ของการทำเหมืองข้อมูลโดยใช้ C5.0 และโครงข่ายประสาทเทียม โดยนำผลมาเปรียบเทียบกับผลพบว่า Creatinine เป็นปัจจัยบ่งชี้ที่สำคัญที่สุดสำหรับผู้ป่วยโรคเบาหวานจอประสาทตา ถ้า Creatinine ผลการทำนาย ภาวะแทรกซ้อนจอประสาทตาโดยใช้ C5.0 เป็น 58.62 และ 74.73 และ โครงข่ายประสาทเทียม เป็น 59.48 และ 99.86 นอกจากนี้ ภาวะแทรกซ้อนของโรคไต ผู้ป่วยโรคเบาหวานเพศหญิง ที่มีประวัติครอบครัวเป็นโรคเบาหวานจะง่ายต่อการได้รับการ ภาวะแทรกซ้อน ผลการทำนาย ภาวะแทรกซ้อนของโรคไต โดยใช้ C5.0 เป็น 69.44 และ 81.36 และโครงข่ายประสาทเทียม เป็น 74.44

และ 98.55 ผลการทำนายภาวะแทรกซ้อนเกี่ยวกับเท้า โดยใช้ C5.0 เป็น 64.71 และ 83.48 และโครงข่ายประสาทเทียม เป็น 67.63 และ 99.70 สำหรับผู้ป่วยโรคเบาหวานเพศหญิงที่มีค่าดัชนีมวลกายไม่ปกติ น้ำตาลในเลือดสูง และผู้ป่วยโรคเบาหวานเพศชายส่วนใหญ่ จะเป็นโรคความดันโลหิตสูง นอกจากนี้ผู้ป่วยโรคเบาหวานที่สูบบุหรี่เป็นปัจจัยเสี่ยงในการเกิดภาวะแทรกซ้อน

จริยาพร จำปา (2553) ศึกษาหาทฤษฎีความสัมพันธ์ของการยืมทรัพยากร ห้องสมุดที่มีผลกับหมวดหมู่ของทรัพยากรห้องสมุด กฟผ. และพยากรณ์ความต้องการทรัพยากร ห้องสมุด โดยผลที่ได้จะมีประโยชน์ต่อการบริหารจัดการห้องสมุด โดยเฉพาะการจัดสรรงบประมาณที่มีอยู่จำกัด เพื่อการจัดหาทรัพยากรที่เหมาะสมและตรงต่อความต้องการของผู้ใช้บริการการวิจัยนี้ใช้ข้อมูลการยืมทรัพยากรห้องสมุดตั้งแต่ ปี พ.ศ. 2548-2553 สำหรับศึกษาพฤติกรรมการใช้ทรัพยากรห้องสมุด และพยากรณ์ความต้องการ จากการหาทฤษฎีความสัมพันธ์ด้วยอัลกอริทึม Apriori พบกฎทั้งสิ้น 4 กฎ ดังนี้ กฎที่ 1 ผู้ใช้บริการเมื่อยืมหมวดหมู่ 300 (สังคมศาสตร์) ก็จะยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือเทคโนโลยี) กฎข้อที่ 2 เมื่อยืมหมวดหมู่ 000 (เบ็ดเตล็ดหรือความรู้ทั่วไป) ก็จะยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือเทคโนโลยี) กฎข้อที่ 3 เมื่อยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์ หรือเทคโนโลยี) ก็จะยืมหมวดหมู่ 000 (เบ็ดเตล็ดหรือความรู้ทั่วไป) กฎข้อที่ 4 เมื่อยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือเทคโนโลยี) ก็จะยืมหมวดหมู่ 300 (สังคมศาสตร์) ด้วยค่าความเชื่อมั่น 18, 12, 7 และ 7 เปอร์เซ็นต์ตามลำดับ หลังจากนั้นนำกฎที่ได้ไปวิเคราะห์การถดถอย และวิเคราะห์อนุกรมเวลาเพื่อหาสมการพยากรณ์ความต้องการ โดยแต่ละสมการมีการประเมินความแม่นยำโดยหาค่าความคลาดเคลื่อนสัมพัทธ์ (MRE) ได้เท่ากับ 36.36 16.47 16.46 และ 10.67 เปอร์เซ็นต์ตามลำดับ และนำไปสร้างโปรแกรมการพยากรณ์ความต้องการทรัพยากรห้องสมุด โดยวิธีทฤษฎีความสัมพันธ์ของห้องสมุด กฟผ. และให้เจ้าหน้าที่ห้องสมุดประเมินผลการใช้โปรแกรมผลการประเมินความพึงพอใจมีค่าเฉลี่ยเท่ากับ 4.52 จากคะแนนเต็ม 5 แสดงว่าเจ้าหน้าที่ห้องสมุดมีความพึงพอใจระบบอยู่ในระดับดีมาก

วิชัย เอกพลากร การศึกษาพัฒนาดัชนีความเสี่ยงต่อเบาหวาน การศึกษานี้ได้ใช้ข้อมูลการศึกษาทางระบาดวิทยาในกลุ่มพนักงานการไฟฟ้าแห่งประเทศไทย (EGAT study) ในการวิเคราะห์ปัจจัยเสี่ยงของการเกิดเบาหวานและการให้คะแนนตามปัจจัยเสี่ยงที่มีความสำคัญต่อการเกิดโรค การวิเคราะห์ปัจจัยเสี่ยงที่มีนัยสำคัญมาพัฒนาคะแนนความเสี่ยง โดยใช้เทคนิคทางสถิติ Logistic Regression Analysis

บทที่ 3

วิธีการดำเนินการวิจัย

งานวิจัยนี้เป็นการสร้างแบบจำลองเพื่อพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยการใช้เทคนิคดาต้าไมน์นิ่งการพัฒนาแบบจำลองครั้งนี้ต้องการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นเบาหวาน ใช้ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน จึงกำหนดขั้นตอนการดำเนินการดังต่อไปนี้

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการดำเนินงานวิจัย
3. ขั้นตอนการดำเนินงาน

1. ประชากรและกลุ่มตัวอย่าง

ข้อมูลที่น่ามาสร้างแบบจำลองเพื่อพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ใช้ข้อมูลผลการตรวจวินิจฉัยโรค ที่มาตรวจรักษาที่ โรงพยาบาลปทุมธานี ตั้งแต่วันที่ 1 มกราคม 2553 ถึง วันที่ 31 ธันวาคม 2556 โดยกรองข้อมูลเริ่มตั้งแต่ผู้ป่วยเป็นโรคเบาหวาน ตามรหัสโรค E10-E14 จำนวนผู้ป่วยเบาหวานที่กรองข้อมูลแล้วทั้งหมด 17,043 คน มาทำการรักษาทั้งหมด 106,317 ครั้ง มีการลงรหัสโรคที่ไม่ซ้ำกัน 4,179 รหัสโรค ตัดข้อมูลโรคที่ทำการลงผลการตรวจวินิจฉัยให้กับผู้ป่วยเบาหวานที่น้อยกว่า 50 ครั้ง ตัดข้อมูลผู้ป่วยที่มาตรวจรักษาไม่ถึง 10 ครั้ง หลังจากตัดข้อมูลออกแล้ว เหลือจำนวนผู้ป่วยเบาหวาน 6,484 คน ผลการตรวจวินิจฉัยตามรหัสโรค 467 รหัสโรค

2. เครื่องมือที่ใช้ในการดำเนินงานวิจัย

การใช้เครื่องมือทำเหมืองข้อมูล สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อน โรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน เครื่องมือที่ใช้ในการดำเนินงานวิจัยประกอบด้วย

2.1 ฮาร์ดแวร์ คือ เครื่องคอมพิวเตอร์ Notebook จำนวน 1 เครื่องเป็นเครื่องส่วนตัวของผู้วิจัย

2.2 ซอฟต์แวร์ ได้แก่ MicroSoft Visual Foxpro จัดเก็บข้อมูลผลตรวจการวินิจฉัยโรค Microsoft office 2010 สร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อน โรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โปรแกรม WEKA ใช้ในการหาความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm

2.3 ระบบสารสนเทศ พัฒนาระบบสารสนเทศเพื่อทดลองแบบพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี ข้อมูลที่นำเสนอมีประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์ ระบบสารสนเทศเพื่อทดสอบแบบพยากรณ์ ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

3. ขั้นตอนการดำเนินงาน

การสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในโรงพยาบาลปทุมธานี มาวิเคราะห์ตามขั้นตอนการทำเหมืองข้อมูล เพื่อค้นหารูปแบบ แนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดของข้อมูลนั้น โดยอาศัยหลักสถิติและหลักคณิตศาสตร์ โดยใช้ขั้นตอนการประมวลผลข้อมูลตามขั้นตอน CRISP-DM การวิเคราะห์ข้อมูลแบ่งออกเป็น 6 ขั้นตอนดังนี้

3.1 ขั้นตอนการทำความเข้าใจปัญหา (Research Understanding Phase)

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องและศึกษาข้อมูลเกี่ยวกับการรักษาของผู้ป่วยหวาน และการลงข้อมูลผลตรวจการรักษา ในระบบสารสนเทศของโรงพยาบาลปทุมธานี ปัญหาที่เกิดขึ้นจริงกับการปฏิบัติงานของการลงผลตรวจรักษาของผู้ป่วยเบาหวาน ดังนี้

3.1.1 แพทย์ไม่สามารถลงผลการตรวจวินิจฉัยโรค ในระบบคอมพิวเตอร์ได้ทันที ซึ่งแพทย์จะเขียนผลการตรวจวินิจฉัยโรคของผู้ป่วยในประวัติการตรวจรักษา และเจ้าหน้าที่บันทึกข้อมูล นำไปบันทึกข้อมูลในระบบคอมพิวเตอร์ อาจจะไม่ถูกต้องของข้อมูลได้ จากลายมือแพทย์ที่อ่านยาก และความเชี่ยวชาญด้านความรู้เรื่องรหัสโรคของเจ้าหน้าที่บันทึกข้อมูล

3.1.2 จำนวนผู้ป่วยเบาหวานที่มาทำการรักษาในแต่ละวันมีจำนวนมาก ทำให้แพทย์ผู้ทำการตรวจรักษา อาจลืมเขียนผลการตรวจวินิจฉัยโรคหรือเจ้าหน้าที่บันทึกข้อมูลลืมลงผลการตรวจวินิจฉัยโรคในระบบคอมพิวเตอร์ ทำให้ข้อมูลไม่ครบถ้วน

3.2 การทำความเข้าใจข้อมูล (Data Understanding Phase)

ผู้วิจัยดำเนินการรวบรวมข้อมูลที่เกี่ยวข้อง เพื่อใช้ในการวิเคราะห์ด้วยเทคนิคดาต้าไมนิ่ง (Data Mining) โดยเริ่มจากการเก็บรวบรวมข้อมูล ทะเบียนประวัติผู้ป่วยผลการตรวจรักษาของผู้ป่วยเบาหวานทั้งหมด ทั้งสถานะผู้ป่วยนอก (OPD: Outpatient Department) ผู้ป่วยใน (IPD: Inpatient Department) และคัดกรองเลือกข้อมูลเฉพาะผู้ป่วยเบาหวานเท่านั้น โดยเจ้าหน้าที่ที่ดูแลระบบสารสนเทศของโรงพยาบาลปทุมธานี ทำการส่งข้อมูลมาให้ ซึ่งอยู่ในรูปแบบไฟล์ ดาต้าเบส (DBF: Database File) เพื่อนำข้อมูลที่ได้ ไปใช้งานต่อไป ซึ่งมีรายละเอียดข้อมูลดังต่อไปนี้

3.2.1 โครงสร้างข้อมูลทะเบียนประวัติผู้ป่วย (Patients) ที่มารับบริการในโรงพยาบาลปทุมธานี ประกอบไปด้วย ดังนี้

ตารางที่ 3.1 โครงสร้างข้อมูลแฟ้มประวัติผู้ป่วย

No.	Name	Data Type	Value	Description
1	Hn	Character		หมายเลขทะเบียนจำตัวผู้ป่วย ไม่ซ้ำกัน
2	Title	Character		คำนำหน้า
3	Name	Character		ชื่อผู้ป่วย
4	Surname	Character		นามสกุลผู้ป่วย
5	Sex	Int	1=ชาย 2=หญิง	เพศ
7	Age	Int		อายุ

3.2.2 โครงสร้างข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก
(OPD: Outpatient Department) ที่มารับบริการในโรงพยาบาลปทุมธานี ประกอบไปด้วย ดังนี้

ตารางที่ 3.2 โครงสร้างข้อมูลแฟ้มผลการตรวจรักษาวินิจฉัยโรคของผู้ป่วยเบาหวานผู้ป่วยนอก

No.	Name	Data Type	Value	Description
1	Hn	Character		หมายเลขทะเบียนจำตัวผู้ป่วย
2	Vn	Character		หมายเลขทะเบียนรับบริการในแต่ละครั้งไม่ซ้ำกัน
3	Date	Date		วันที่มาทำการรักษา
4	Diag	Character		ผลการตรวจวินิจฉัยโรค

3.2.3 โครงสร้างข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน (IPD:
Inpatient Department) ที่มารับบริการในโรงพยาบาลปทุมธานี ประกอบไปด้วย ดังนี้

ตารางที่ 3.3 โครงสร้างข้อมูลแฟ้มผลการตรวจรักษาวินิจฉัยโรคของผู้ป่วยเบาหวานผู้ป่วยใน

No.	Name	Data Type	Value	Description
1	Hn	Character		หมายเลขทะเบียนจำตัวผู้ป่วย
2	An	Character		หมายเลขทะเบียนรับเป็นผู้ป่วยใน แต่ละครั้งไม่ซ้ำกัน
3	Date	Date		วันที่ลงผลการรักษา
4	Dx	Character		ผลการตรวจวินิจฉัยโรค

3.2.4 โครงสร้างข้อมูลรหัสโรค (ICD101: ICD-10-TM International Statistical Classification
of Diseases and Related Health Problems 10th Revision, Thai Modification,2012) ประกอบไป
ด้วย ดังนี้

ตารางที่ 3.4 โครงสร้างข้อมูล รหัสโรค

No.	Name	Data Type	Value	Description
1	CODE	Character		รหัสโรค
2	DESCT	Character		ชื่อรหัสโรค ภาษาไทย
3	DESC	Character		ชื่อรหัสโรค ภาษาอังกฤษ

3.2.5 โครงสร้างข้อมูลกลุ่มโรค (Groupdx : Disease Group) ประกอบไปด้วย ดังนี้

ตารางที่ 3.5 โครงสร้างข้อมูล รหัสกลุ่มโรค

No.	Name	Data Type	Value	Description
1	CODE	Character		รหัสกลุ่มโรค
2	NAME	Character		ชื่อกลุ่มโรค

3.3 การเตรียมข้อมูล (Data Preparation Phase)

การจัดเตรียมข้อมูลให้มีความถูกต้อง เพื่อให้แบบจำลองมีความแม่นยำ โดยใช้ข้อมูลผลการตรวจวินิจฉัยโรค ตั้งแต่ที่ผู้ป่วยเป็นโรคเบาหวาน นำไปสร้างแบบจำลอง ตัวแปรที่สำคัญได้แก่ จำนวนครั้งที่มาตรวจรักษากับแพทย์ จำนวนโรคที่ทำการลงผลการตรวจวินิจฉัยโรคให้กับผู้ป่วยเบาหวาน และเป็นผู้ป่วยเบาหวานเท่านั้น ทำการเลือกข้อมูลโรคผู้ป่วยเบาหวาน ตามภาพที่ 3.1 กรองข้อมูลเริ่มตั้งแต่ผู้ป่วยเป็นโรคเบาหวาน ตามรหัสโรค E10-E14 จำนวนผู้ป่วยเบาหวาน ตัดข้อมูลรายการโรคที่ทำการลงผลการตรวจวินิจฉัยโรคให้กับผู้ป่วยเบาหวานที่น้อยกว่า 50 ครั้ง ตัดข้อมูลผู้ป่วยที่มาตรวจรักษาไม่ถึง 10 ครั้ง และทำการส่งข้อมูลออกมาเป็นรูปแบบ File Microsoft Excel 2010

3.3.1 ตารางข้อมูลทะเบียนประวัติผู้ป่วย (Patients) ที่มารับบริการในโรงพยาบาลปทุมธานี ดังนี้

Hn	Title	Name	Surname	Sex	Age
00613806	นาย			1	75
00613807	นาย			1	61
00613808	นาย			1	39
00613831	นาง			2	75
00613832	นาย			1	43
00613833	นาย			1	54
00613838	นาย			1	35
00613840	นาง			2	51
00613841	นาง			2	56
00613844	นาง			2	71
00613845	นาง			2	79
00613848	นาย			1	51
00613850	นาง			2	51
00613851	นาย			1	36
00613854	นาง			2	57
00613855	นาง			2	37
00613857	นาง			2	53
00613858	นาง			2	36
00613859	นาง			2	54
00613862	นาย			1	62
00613864	นาง			2	78
00613866	นาง			2	50
00613868	นาย			1	64

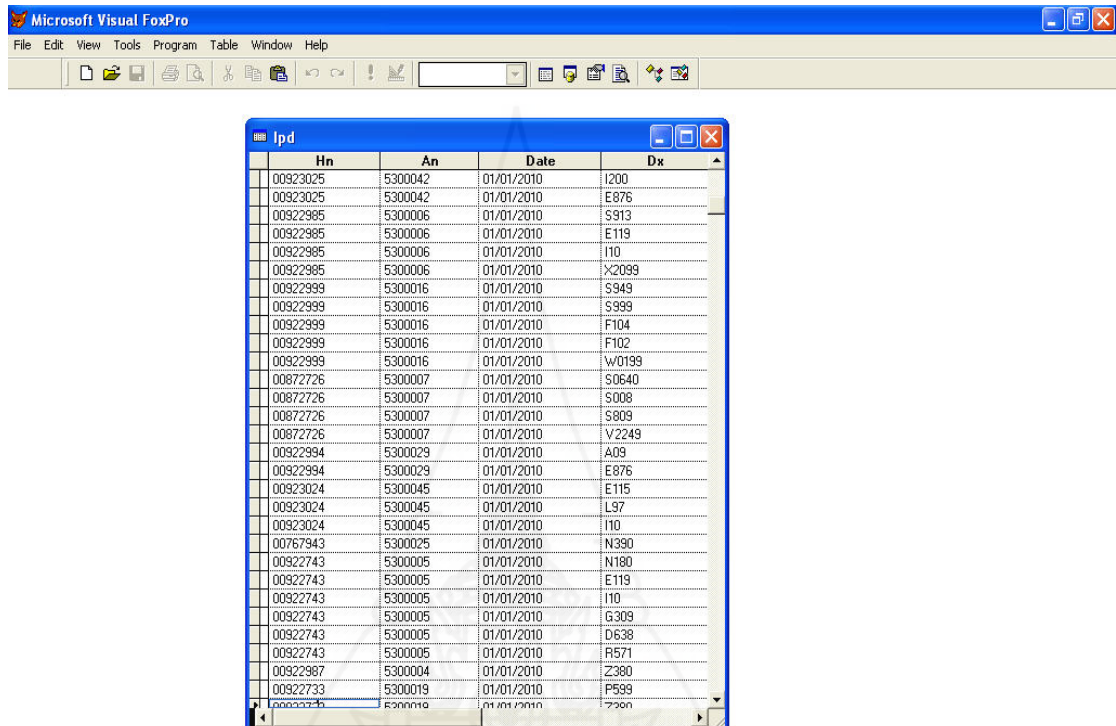
ภาพที่ 3.1 ข้อมูลทะเบียนประวัติผู้ป่วย

3.3.2 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก (OPD: Outpatient Department) ที่มารับบริการในโรงพยาบาลปทุมธานี ดังนี้

Hn	Vn	Date	Diag
00614035	03191394	01/10/2009	O800
00914097	03191455	01/10/2009	O800
00165925	03191592	01/10/2009	R33
00914086	03191426	01/10/2009	A09
00866938	03191595	01/10/2009	K30
00267161	03191594	01/10/2009	H814
00914098	03191596	01/10/2009	Z380
00597347	03191593	01/10/2009	R104
00898023	03191598	01/10/2009	J120
00283855	03191597	01/10/2009	A09
00278687	03191599	01/10/2009	S099
00278687	03191599	01/10/2009	V229
00914089	03191655	01/10/2009	O800
00782596	03191661	01/10/2009	A09
00912332	03191706	01/10/2009	N10
00599403	03191734	01/10/2009	G409
00637568	03191774	01/10/2009	O800
00380490	03191812	01/10/2009	C322
00911743	03191666	01/10/2009	J189
00238869	03191677	01/10/2009	J069
00401063	03191739	01/10/2009	J449
00190984	03191647	01/10/2009	J069
00510755	03191626	01/10/2009	I10
00510755	03191626	01/10/2009	E785
00298335	03191610	01/10/2009	I10
00298335	03191610	01/10/2009	F785

ภาพที่ 3.2 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก

3.3.3 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน (IPD: Inpatient Department) ที่มารับบริการในโรงพยาบาลปทุมธานี ดังนี้



The screenshot shows a Microsoft Visual FoxPro window with a table named 'ipd'. The table contains the following data:

Hn	An	Date	Dx
00923025	5300042	01/01/2010	I200
00923025	5300042	01/01/2010	E876
00922985	5300006	01/01/2010	S913
00922985	5300006	01/01/2010	E119
00922985	5300006	01/01/2010	I10
00922985	5300006	01/01/2010	X2099
00922999	5300016	01/01/2010	S949
00922999	5300016	01/01/2010	S999
00922999	5300016	01/01/2010	F104
00922999	5300016	01/01/2010	F102
00922999	5300016	01/01/2010	w0199
00872726	5300007	01/01/2010	S0640
00872726	5300007	01/01/2010	S008
00872726	5300007	01/01/2010	S809
00872726	5300007	01/01/2010	V 2249
00922994	5300029	01/01/2010	A09
00922994	5300029	01/01/2010	E876
00923024	5300045	01/01/2010	E115
00923024	5300045	01/01/2010	L97
00923024	5300045	01/01/2010	I10
00767943	5300025	01/01/2010	N390
00922743	5300005	01/01/2010	N180
00922743	5300005	01/01/2010	E119
00922743	5300005	01/01/2010	I10
00922743	5300005	01/01/2010	G309
00922743	5300005	01/01/2010	D638
00922743	5300005	01/01/2010	R571
00922987	5300004	01/01/2010	Z390
00922733	5300019	01/01/2010	P599
00922733	5300019	01/01/2010	Z390

ภาพที่ 3.3 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน

3.3.4 ข้อมูลรหัสโรค (ICD101: ICD-10-TM) International Statistical Classification of Diseases and Related Health Problems 10th Revision, Thai Modification, 2012 ดังนี้

Code	Desc	Desc
E10	เบาหวานชนิดที่ต้องพึ่งอินซูลิน	Insulin-dependent diabetes mellitus
E100	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับโคม่า	Insulin-dependent diabetes mellitus type 1 at with coma
E101	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะเลือดเป็นกรดจากคีโตน	Insulin-dependent diabetes mellitus type 1 at with ketoacidosis
E102	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางไต	Insulin-dependent diabetes mellitus type 1 at with renal complications
E103	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา	Insulin-dependent diabetes mellitus type 1 at with ophthalmic complication
E104	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท	Insulin-dependent diabetes mellitus type 1 at with neurological complications
E105	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย	Insulin-dependent diabetes mellitus type 1 at with peripheral circulatory complicati
E106	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด	Insulin-dependent diabetes mellitus type 1 at with other specified complications
E107	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนหลายอย่าง	Insulin-dependent diabetes mellitus type 1 at with multipla complications
E108	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด	Insulin-dependent diabetes mellitus type 1 at with unspecified complications
E109	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ไม่มีภาวะแทรกซ้อน	Insulin-dependent diabetes mellitus type 1 at without complications
E11	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน	Non-insulin-dependent diabetes mellitus
E112	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางไต	Non-insulin-dependent diabetes mellitus type 2 at with renal complications
E113	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา	Non-insulin-dependent diabetes mellitus type 2 at with ophthalmic complication
E119	เบาหวานแบบที่ 2 ชนิดไม่พึ่งอินซูลินในคนอายุน้อย	Non-insulin-dependent diabetes mellitus type 2 at without complications
E12	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ	Malnutrition-related diabetes mellitus
E129	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ไม่มีภาวะแทรกซ้อน	Malnutrition-related diabetes mellitus at without complications
E13	เบาหวานชนิดอื่นที่ระบุรายละเอียด	Other specified diabetes mellitus
E131	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะเลือดเป็นกรดจากคีโตน	Other specified diabetes mellitus at with ketoacidosis
E139	เบาหวานชนิดอื่นที่ระบุรายละเอียด ไม่มีภาวะแทรกซ้อน	Other specified diabetes mellitus at without complications
E14	เบาหวานที่ไม่ระบุรายละเอียด	Unspecified diabetes mellitus
E140	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับโคม่า	Unspecified diabetes mellitus at with coma
E141	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะเลือดเป็นกรดจากคีโตน	Unspecified diabetes mellitus at with ketoacidosis
E142	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางไต	Unspecified diabetes mellitus at with renal complications

ภาพที่ 3.4 ข้อมูลรหัสโรค ICD101

3.3.5 ข้อมูลกลุ่มโรค (Groupdx : Disease Group)

Code	Name
A80-A89	การติดเชื้อไวรัสที่ระบบประสาทส่วนกลาง
B00-B09	การติดเชื้อไวรัสที่มีอาการที่ผิวหนังและเยื่อเมือก
B20-B24	โรคภูมิคุ้มกันบกพร่องจากไวรัส [เอชไอวี]
B25-B34	โรคจากไวรัสอื่น
B35-B49	โรคติดเชื้อรา
B90-B94	ผลที่ตามมาของโรคติดเชื้อและปรสิต
B95-B97	แบคทีเรีย ไวรัส และสิ่งก่อโรคติดเชื้ออื่น
C15-C26	เนื้องอกร้ายของอวัยวะย่อยอาหาร
C30-C39	เนื้องอกร้ายของอวัยวะหายใจและอวัยวะในช่องอก
C50-C58	เนื้องอกร้ายของเต้านมและอวัยวะสืบพันธุ์สตรี
D10-D36	เนื้องอกไม่ร้าย
D50-D53	โลหิตจางจากการขาดสารอาหาร
D55-D59	โลหิตจางจากเม็ดเลือดแดงแตก
D60-D64	โลหิตจางจากไขกระดูกฝ่อและโลหิตจางชนิดอื่น
D65-D69	ภาวะแข็งตัวของเลือดผิดปกติ จำเลือด และภาวะเลือดออกอื่น
E10-E14	เบาหวาน
E15-E16	ความผิดปกติอื่นของการควบคุมกลูโคสและการสังเคราะห์อินซูลินในตับอ่อน
E70-E79	ความผิดปกติทางเมตาบอลิกของโปรตีน ไขมัน และคาร์โบไฮเดรต
E79-E90	ความผิดปกติทางเมตาบอลิกอื่น

ภาพที่ 3.5 ข้อมูลกลุ่มโรค

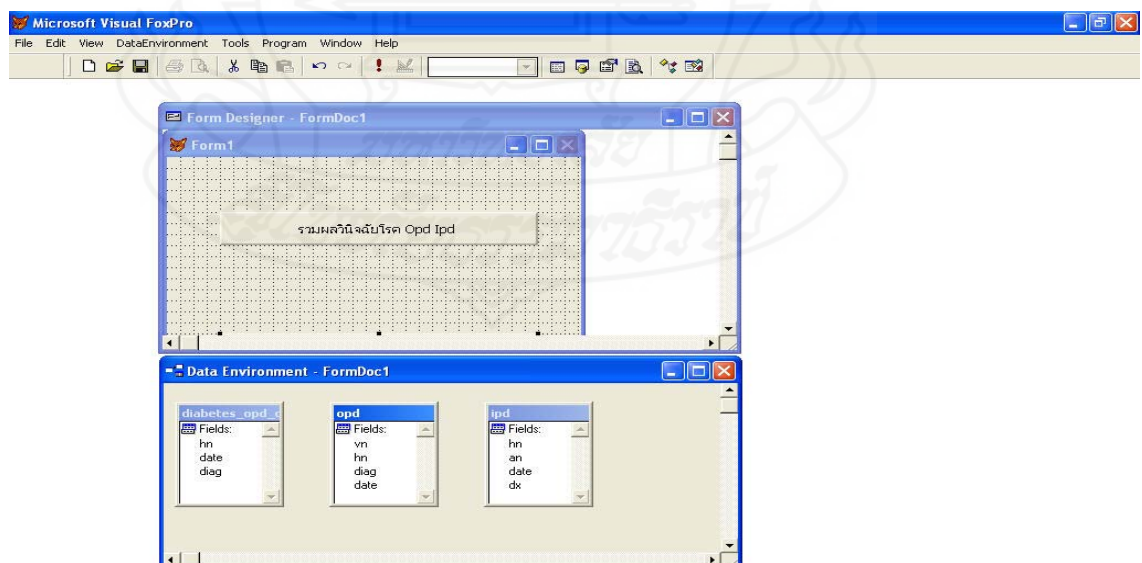
3.3.6 ขั้นตอนการนำข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานจากตารางข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก (OPD: Outpatient Department) และ ตารางข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน (IPD: Inpatient Department) ทั้ง 2 ตารางให้อยู่ตารางเดียว โดยการสร้างโครงสร้างฐานข้อมูลใหม่

1) โครงสร้างฐานเก็บผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานทั้งหมด (Diabetes_Opd_DX) ประกอบไปด้วย

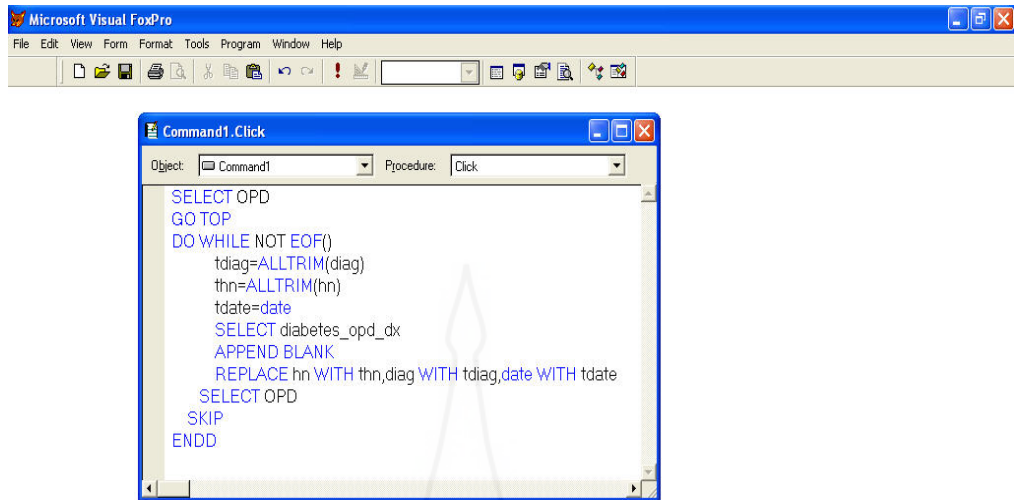
ตารางที่ 3.6 โครงสร้างฐานเก็บผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน

No.	Name	Data Type	Value	Description
1	HN	Character		หมายเลขทะเบียนจำตัวผู้ป่วย
2	Diag	Character		ผลการตรวจวินิจฉัยโรค
3	Date	Date		วันที่ลงผลการตรวจวินิจฉัยโรค

2) การนำข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานจากตารางข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยนอก (OPD: Outpatient Department) และ ตารางข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน ผู้ป่วยใน (IPD: Inpatient Department) ไปยัง ตาราง ผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานทั้งหมด (Diabetes_Opd_DX)



ภาพที่ 3.6 การนำข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน

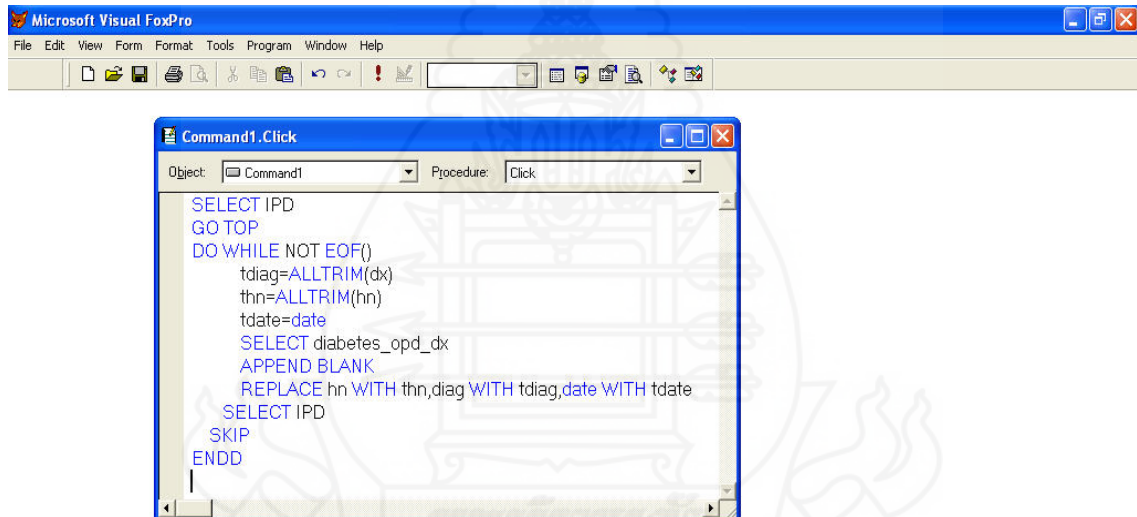


```

SELECT OPD
GO TOP
DO WHILE NOT EOF()
    tdiag=ALLTRIM(diag)
    thn=ALLTRIM(hn)
    tdate=date
    SELECT diabetes_opd_dx
    APPEND BLANK
    REPLACE hn WITH thn,diag WITH tdiag,date WITH tdate
    SELECT OPD
    SKIP
ENDD

```

ภาพที่ 3.7 ซอร์ซโค้ด (Source Code) การนำข้อมูลผลตรวจผู้ป่วยเบาหวาน ผู้ป่วยนอก



```

SELECT IPD
GO TOP
DO WHILE NOT EOF()
    tdiag=ALLTRIM(dx)
    thn=ALLTRIM(hn)
    tdate=date
    SELECT diabetes_opd_dx
    APPEND BLANK
    REPLACE hn WITH thn,diag WITH tdiag,date WITH tdate
    SELECT IPD
    SKIP
ENDD

```

ภาพที่ 3.8 ซอร์ซโค้ด (Source Code) การนำข้อมูลผลตรวจผู้ป่วยเบาหวาน ผู้ป่วยใน

Hn	Date	Diag
00094388	01/01/2010	E119
00922991	01/01/2010	Z532
00922992	01/01/2010	A09
00910690	01/01/2010	S014
00910690	01/01/2010	E119
00299458	01/01/2010	O800
00913789	01/01/2010	A09
00922993	01/01/2010	S80
00922993	01/01/2010	V2849
00709938	01/01/2010	A09
00388448	01/01/2010	S400
00388448	01/01/2010	E119
00409061	01/01/2010	E146
00836199	01/01/2010	J069
00790646	01/01/2010	C539
00319077	01/01/2010	F20
00367807	01/01/2010	Z532
00521111	01/01/2010	Z532
00922449	01/01/2010	Z480
00379065	01/01/2010	Z480
00594926	01/01/2010	R42
00757247	01/01/2010	Z235
00827517	01/01/2010	R509
00379364	01/01/2010	Z235

ภาพที่ 3.9 ผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวานทั้งหมด (Diabetes_Opd_DX)

3) ทำการเลือกข้อมูลผู้ป่วยเบาหวาน ตาม รูปที่ 3.8 ตามขั้นตอนดังนี้

(1) เลือกข้อมูลผลการตรวจวินิจฉัยโรค E10-E14 ที่แพทย์ลงผลเป็นครั้งแรก มาเป็นตัวแปรต้น และเลือกข้อมูลผลการตรวจวินิจฉัยโรค หลังจากที่เป็นโรคเบาหวาน

(2) ทำการตัดรายการรหัสโรคที่ลงทะเบียนน้อยกว่า 50 รายการและตัดข้อมูลผู้ป่วยที่มาทำการรักษาไม่ถึง 10 ครั้งออก

(3) ส่งข้อมูลมาเป็นรูปแบบ File Microsoft Excel 2010 เพื่อเป็นแบบจำลองพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวาน

(4) การแปลงรูปแบบข้อมูล ปรับโครงสร้างข้อมูลให้เหมาะสมนำไปใช้จัดทำ Data Mining ได้หลังจากคัดเลือกข้อมูลแล้ว ใส่มูลค่าในแต่ละคอลัมน์ ดังนี้

ค่า t แสดงว่า ผู้ป่วย มีการลงผลวินิจฉัยโรค

ค่าว่าง แสดงว่า ผู้ป่วย ไม่มีการลงผลวินิจฉัยโรค

จากนั้นดำเนินการแปลงไฟล์ข้อมูลที่มีการปรับแก้ไข ให้เป็นไฟล์ที่เหมาะสมกับโปรแกรมที่จะใช้ในการวิเคราะห์ข้อมูล ซึ่งผู้วิจัยเลือกใช้โปรแกรม WEKA เป็นโปรแกรมสำหรับวิเคราะห์ข้อมูล ดังนั้นจึงได้ดำเนินการเปลี่ยนไฟล์ข้อมูลเดิมที่อยู่ในรูปแบบ File Microsoft Excel 2010 ให้เป็นไฟล์ที่สามารถเปิดใช้งานได้ในโปรแกรม WEKA คือ บันทึกเป็นไฟล์ .arff

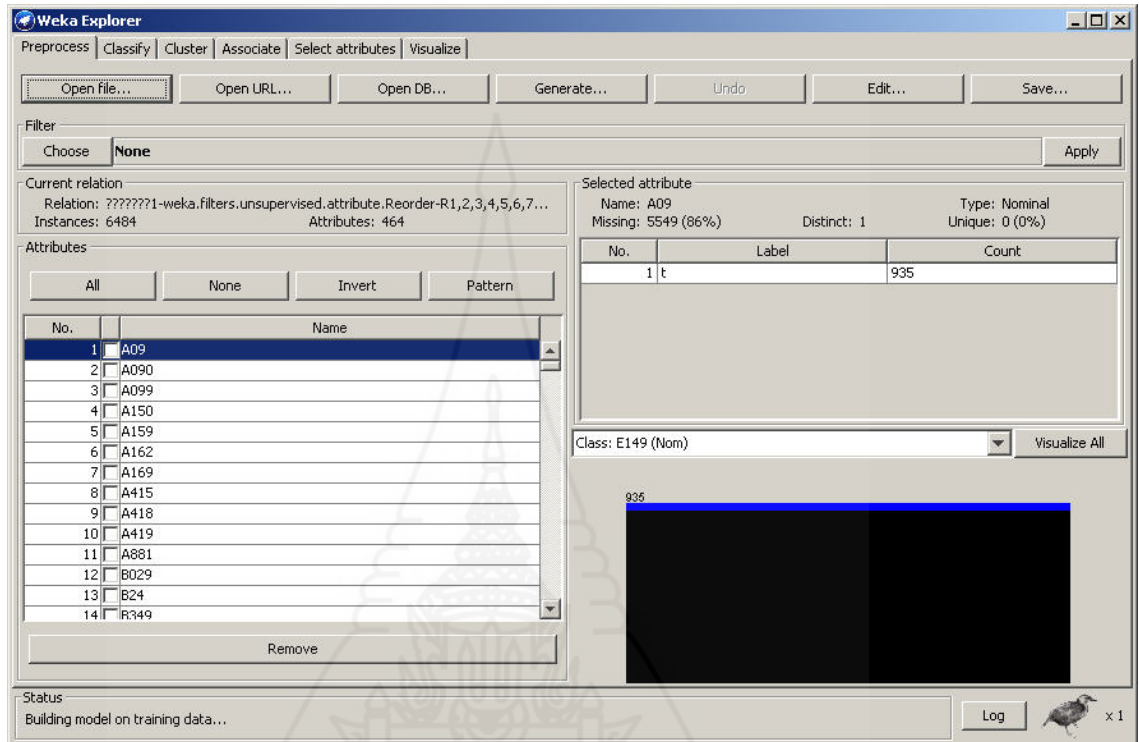
	A	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1	HN	E041	E059	E10	E100	E103	E109	E113	E114	E115	E119	E144	E145	E149	E162	E780	E785	E871	E876	F009	F200
2		1043				t	t		t		t						t				
3		1399				t	t		t						t		t				
4		1441					t														
5		1733			t						t										
6		1890					t	t	t		t								t		
7		1915									t		t						t		
8		1991		t							t						t		t		
9		2140					t				t								t		
10		2341					t				t				t				t		
11		2540					t				t				t				t		
12		2765					t				t								t		
13		3093				t	t				t								t		
14		3139					t				t								t		
15		3281													t		t		t		
16		3522					t				t								t		
17		3536					t	t			t								t		
18		3724			t						t								t		
19		3930					t				t				t				t		
20		4195		t							t								t		
21		4219					t				t								t	t	t
22		4246				t	t	t	t		t								t		
23		4257					t				t								t		t
24		4259				t	t		t	t	t								t		
25		4576		t			t		t	t	t								t		t
26		5234					t		t		t								t		
27		5511	t				t														

ภาพที่ 3.10 ข้อมูลออกมาเป็นรูปแบบ File Microsoft Excel 2010

3.4 การสร้างแบบจำลอง (Modeling Phase)

การจัดทำแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวานใช้โปรแกรม WEKA ในการวิเคราะห์ข้อมูลเพื่อหาภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยการหาความสัมพันธ์ของข้อมูล (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm แล้วนำผลพยากรณ์มาเปรียบเทียบกับค่าความถูกต้องแม่นยำ โดยใช้ผลที่มีความเชื่อถือมากที่สุด จึงได้ทำการกำหนดค่าสนับสนุนหรือค่าความถูกต้อง ขั้นต่ำที่ 0.1 (10%) สูงสุดที่ 1.0 (100%) ให้แสดงค่า กฎ (Rules) ที่ค้นพบให้แสดงรายการ 20 กฎ (Rules) ค่าความเชื่อถือที่เท่ากับหรือมากกว่า 0.5 (50 % ถึง 100 %) ในการค้นหาความสัมพันธ์ของข้อมูลเพื่อหาความสัมพันธ์การเกิดภาวะแทรกซ้อนของโรคผู้ป่วยเบาหวาน

3.4.1 ขั้นตอนการนำข้อมูลเข้าโปรแกรม WEKA เพื่อทำการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์



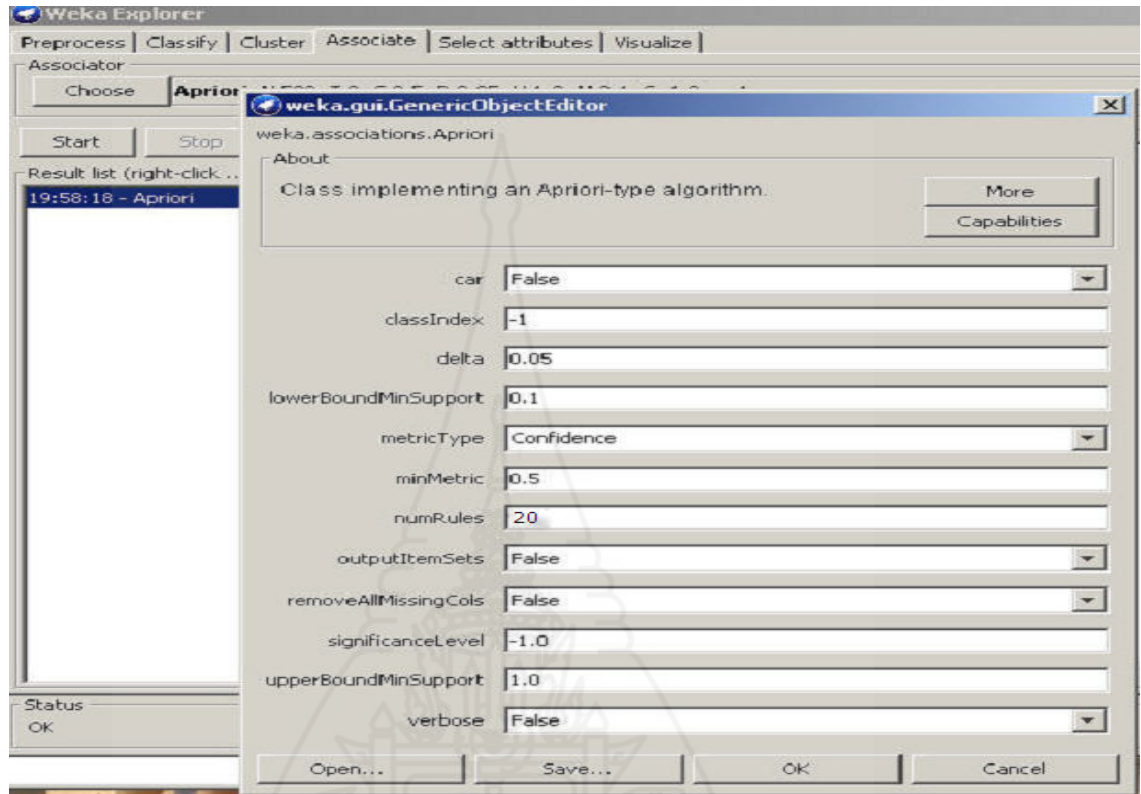
ภาพที่ 3.11 การนำข้อมูลเข้าโปรแกรม WEKA เพื่อทำการวิเคราะห์ข้อมูล

3.4.2 ขั้นตอนการกำหนดค่า *MinSupport*, *Confidence*, *numRules* โปรแกรม WEKA

ในการหาความสัมพันธ์ของข้อมูล (Association) ระหว่าง Apriori Algorithm และ FP Growth Algorithm การกำหนดค่าพารามิเตอร์ปรับค่า Min Support หมายถึง ค่าสนับสนุนต่ำสุดหรือค่าความถูกต้องในกล่อง lower Bound Min Support เช่น 0.1 (ค่าสนับสนุนต่ำสุด 10%) ปรับค่า Min Confidence หมายถึง ค่าความเชื่อมั่นต่ำสุด ในกล่อง min Metric โดย metric Type เป็น Confidence เช่น 0.5 (ค่าความเชื่อมั่นต่ำสุด 50%) ปรับค่า num Rules หมายถึง จำนวนกฎที่ต้องการแสดง ในกล่อง num Rules เช่น 20 กฎ

เพื่อเป็นการค้นหาความสัมพันธ์ที่มีความเชื่อถือมากที่สุด จึงได้ทำการกำหนดค่าสนับสนุนหรือค่าความถูกต้อง ขั้นต่ำที่ 0.1 (10%) สูงสุดที่ 1.0 (100%) ให้แสดงค่ากฎ (Rules) ที่ค้นพบให้แสดงรายการ 20 กฎ (Rules) ค่าความเชื่อถือที่เท่ากับหรือมากกว่า 0.5 (50% ถึง 100%) เพื่อค้นหาความสัมพันธ์ของข้อมูลเพื่อหาความสัมพันธ์การเกิดภาวะแทรกซ้อนของโรคผู้ป่วย

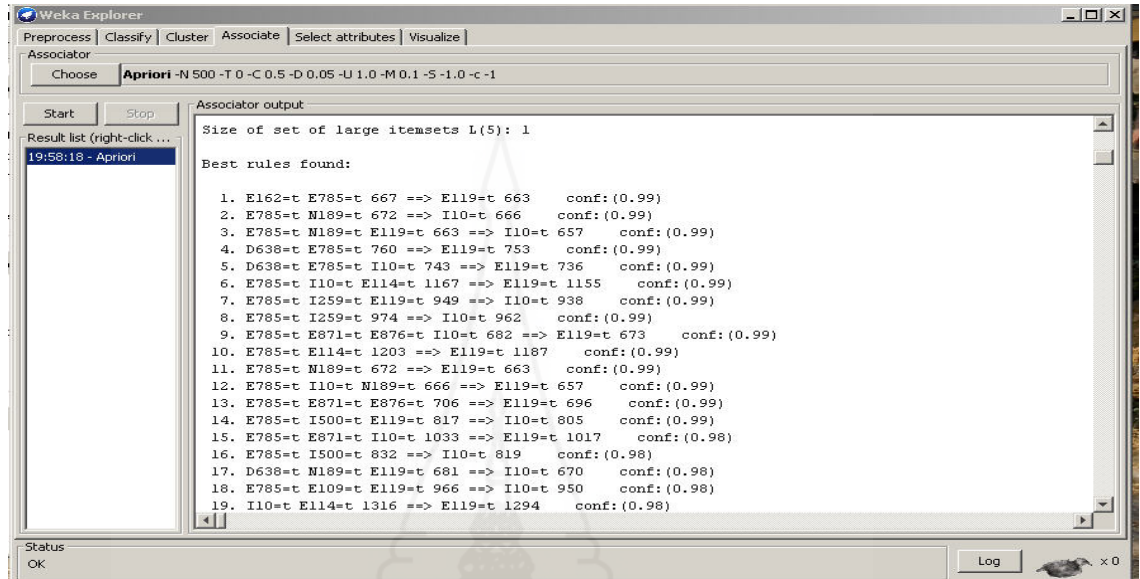
เบหวน โดยใช้โปรแกรม WEKA นั้นสามารถกำหนดจำนวนความสัมพันธ์ที่ต้องการได้ดังนี้
 ภาพที่ 3.12



ภาพที่ 3.12 การกำหนดค่า MinSupport , Confidence , numRules

3.4.3 ผลการพยากรณ์ โดยวิธีการหาความสัมพันธ์ของข้อมูล (Association) Apriori

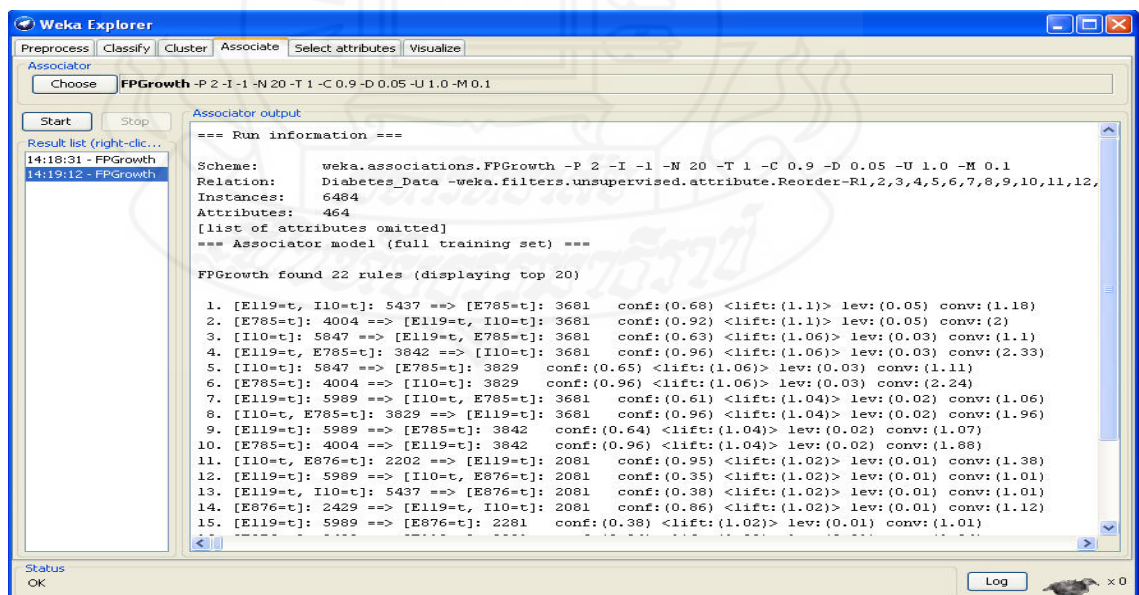
Algorithm



ภาพที่ 3.13 ผลการพยากรณ์ด้วย Apriori Algorithm

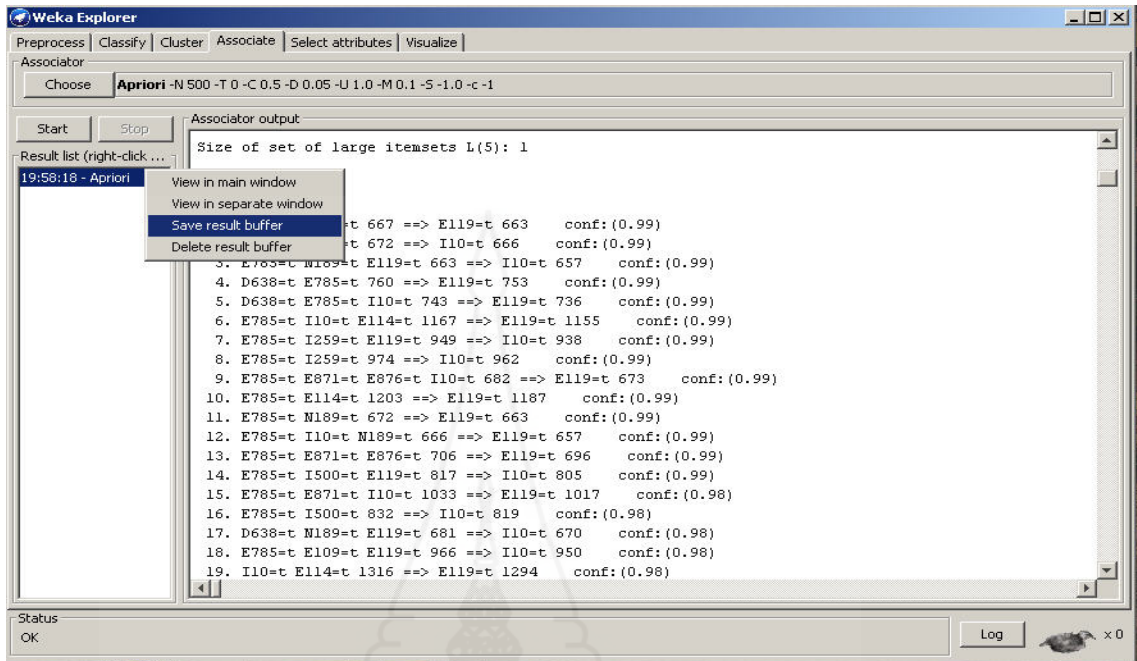
3.4.4 ผลการพยากรณ์ โดยวิธีการหาความสัมพันธ์ของข้อมูล (Association)

FPGrowth Algorithm

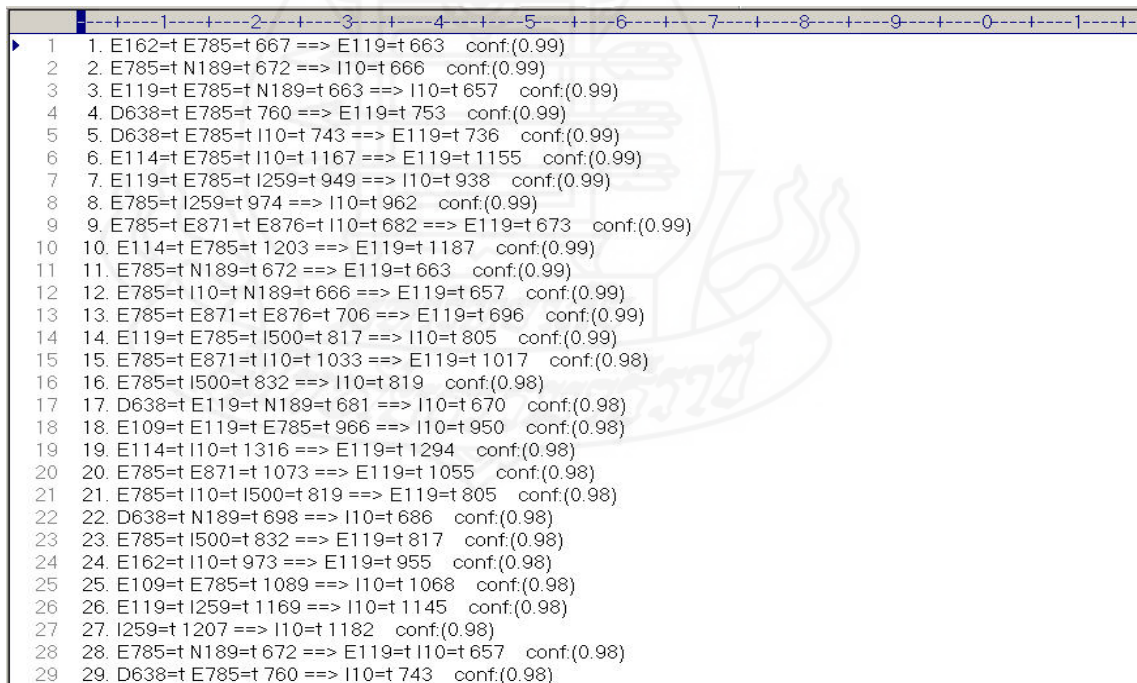


ภาพที่ 3.14 ผลการพยากรณ์ด้วย FPGrowth Algorithm

3.4.5 การบันทึกผลการวิเคราะห์ข้อมูล ส่งข้อมูลออกในรูปแบบ ไฟล์ TEXT FILE



ภาพที่ 3.15 การบันทึก ส่งข้อมูลออกในรูปแบบ ไฟล์ TEXT FILE



ภาพที่ 3.16 ข้อมูลในรูปแบบ ไฟล์ TEXT FILE

3.5 การประเมินผล (Evaluation Phase)

เป็นการประเมินผลผลลัพธ์จากแบบจำลองและอัลกอริทึมที่ใช้วิเคราะห์ข้อมูลว่าผลลัพธ์ที่ได้นั้นครอบคลุมและสามารถตอบวัตถุประสงค์ที่กำหนดไว้หรือไม่ จากนั้นจึงนำแบบจำลองที่ได้จากการทำเหมืองข้อมูล มาประเมินพิจารณาถึงความเหมาะสม ในการนำแบบจำลองไปประยุกต์ใช้ ว่าแบบจำลองที่ได้นั้นมีความแม่นยำในการทำนายผลพยากรณ์มากน้อยเพียงใด โดยในการประเมินแบบจำลองนั้น จะแบ่งตามลักษณะของการทำเหมืองข้อมูล กฎความสัมพันธ์สามารถประเมินผลที่ได้จากการพิจารณาค่าความเชื่อมั่น และค่าสนับสนุน ประเมินผลโดยการเปรียบเทียบผลพยากรณ์ที่ได้กับข้อมูลจริงที่เกิดขึ้น

3.6 การนำแบบจำลองไปใช้งาน (Deployment Phase)

3.6.1 นำเสนอผลลัพธ์ในรูปแบบที่เหมาะสม นำแบบจำลองที่ได้ไปทดลองใช้ โดยพัฒนาระบบสารสนเทศ ระบบสารสนเทศนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี ข้อมูลที่นำเสนอประกอบไปด้วยประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์

3.6.2 ระบบสารสนเทศนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ประกอบไปด้วย ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ประกอบด้วย

1) ระบบชื่อ “ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน”

2) หน้าจอหลักระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ประกอบไปด้วยรายละเอียดของการวิจัย และที่มาของโครงการ วัตถุประสงค์ของการวิจัยครั้งนี้ กลุ่มประชากรตัวอย่างที่ใช้ในการวิจัย ทฤษฎีที่เกี่ยวข้องเครื่องมือที่ใช้ในการดำเนินการวิจัย ผู้วิจัย และ อาจารย์ที่ปรึกษาของวิทยานิพนธ์นี้ หน้าจอหลัก ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน แยกออกเป็น 3 ส่วน คือ

- (1) ชื่อโรงพยาบาล โลโก้ ที่อยู่ ที่ทำการวิจัย
- (2) รายละเอียดของการวิจัย
 - ชื่อระบบโปรแกรมที่พัฒนา
 - ที่มาของโครงการ

- วัตถุประสงค์ของการวิจัย
- ทฤษฎีเกี่ยวข้อง
- เครื่องมือที่ใช้ในการดำเนินงาน
- ชื่อผู้วิจัย ระดับการศึกษา สาขาวิชา ชื่อมหาวิทยาลัย
- ชื่ออาจารย์ที่ปรึกษาหลัก และอาจารย์ที่ปรึกษาร่วม

(3) ส่วนการทำงานของระบบโปรแกรม

- การเข้าใช้โปรแกรม
- ปิดโปรแกรม

หน้าเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน			
โรงพยาบาลปทุมธานี เลขที่ 7 ถนนปทุมธานี - ลาดหลุมแก้ว ด.บางปรอก อ.เมือง จ.ปทุมธานี 12000			
ระบบ นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน			
ที่มาของโครงการ วิทยานิพนธ์ เรื่องการพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล Predicting complications. Of diabetic patients by using data mining techniques			
วัตถุประสงค์ของการวิจัย การสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรค ของกลุ่มผู้ป่วยโรคเบาหวาน ด้วยใช้ข้อมูลประวัติการรักษาในโรงพยาบาล ด้วยเทคนิคการทำเหมืองข้อมูล			
ประชากรและกลุ่มตัวอย่าง ประชากรกลุ่มตัวอย่าง ข้อมูลสารสนเทศด้านการให้บริการและประวัติการรักษาพยาบาลของผู้ป่วยเบาหวาน ที่ขึ้นทะเบียนกับโรงพยาบาลปทุมธานี จำนวนผู้ป่วย 17,043 คน ตั้งแต่วันที่ 1 มกราคม 2553 ถึง วันที่ 31 ธันวาคม 2556			
ทฤษฎีที่เกี่ยวข้อง การใช้เครื่องมือทำเหมืองข้อมูล (Data Mining Tool) เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล โดยการใช้ออฟต์แวร์ทำเหมืองข้อมูล โดยใช้ กฎกฎความสัมพันธ์ (Association Rule) และ กฎการแบ่งกลุ่ม (Clustering) โดยการคำนวณทางสถิติ (Descriptive Statistics)			
เครื่องมือที่ใช้ในการดำเนินงานวิจัย การใช้เครื่องมือทำเหมืองข้อมูล (Data mining Tool) เพื่อวิเคราะห์รูปแบบของข้อมูล และความสัมพันธ์ของข้อมูล โดยนำข้อมูลสารสนเทศด้านการให้บริการ และประวัติการรักษาพยาบาลของผู้ป่วยเบาหวาน เครื่องมือที่ใช้ในการดำเนินงานวิจัยประกอบด้วย 2.1 ฮาร์ดแวร์ คือ เครื่องคอมพิวเตอร์ Notebook จำนวน 1 เครื่อง เป็น เครื่องส่วนตัวของผู้วิจัย 2.2 ซอฟต์แวร์ ซอฟต์แวร์ที่ใช้ ได้แก่ ซอฟต์แวร์ระบบจัดการฐานข้อมูล MicroSoft Visual Foxpro และซอฟต์แวร์ Microsoft office 2010 สำหรับการพัฒนาคอลเลกชันข้อมูล (data warehouse) และใช้เทคนิคการทำเหมืองข้อมูล (data mining technique) โดยใช้โปรแกรม WEKA เป็นโปรแกรมแถมฟรีวิเคราะห์ข้อมูล หารูปแบบของคลังข้อมูล และหาความสัมพันธ์ของข้อมูล			
ผู้วิจัย นายปฏิกรณ์ ทรายหมอ	รหัสประจำตัว	2549600142	นักศึกษาปริญญาโท คณะวิทยาศาสตร์และเทคโนโลยี แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.วิภา เจริญกิตติหารักษ์ อาจารย์วิทยา พรพัชรพงศ์	อาจารย์ที่ปรึกษาหลัก อาจารย์ที่ปรึกษาร่วม	

ภาพที่ 3.17 หน้าจอหลักระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

3) หน้าจอทำงานระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน เป็นระบบสืบค้นข้อมูลประวัติการรักษาของผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี โดยทำการค้นหาสถิติข้อมูลการลงผลการตรวจวินิจฉัยโรคเบาหวาน จากทะเบียนประวัติการรักษา เฉพาะรหัสกลุ่มโรคเบาหวาน และนำผลการตรวจวินิจฉัยโรคที่มีความถี่มากที่สุด ไปเป็นตัวแปรในการสืบค้นหาภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็น

โรคเบาหวาน หน้าจอทำงานระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน แบ่งออกเป็น 5 ส่วน ดังนี้

- (1) ชื่อโรงพยาบาล โลโก้ ที่อยู่ ที่ทำการวิจัย
- (2) ส่วนของการค้นหา
 - ค้นหาตาม หมายเลขประจำตัวผู้ป่วย H.N. (Hospital Number)
 - ค้นหาตาม รหัสโรค
- (3) ส่วนของผลการพยากรณ์
 - ผลการพยากรณ์ภาวะแทรกซ้อนของโรค
 - ข้อมูลตามกลุ่มโรค
- (4) ข้อควรปฏิบัติ/คำแนะนำของแพทย์
 - อาการของโรค
 - ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต และระดับเสี่ยงต่อการเสียชีวิต
- (5) ส่วนการทำงานของระบบโปรแกรม
 - กลุ่มงานห้องตรวจผู้ป่วยเบาหวาน
 - ผลการพยากรณ์โอกาสการเกิด โรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน
 - ปิดโปรแกรม



หน้าเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน : 19:13:52

โรงพยาบาลปทุมธานี
เลขที่ 7 ถนนปทุมธานี-ลาดหลุมแก้ว ต.บางปรอก อ.เมือง จ.ปทุมธานี 12000

นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

ค้นหาข้อมูลผู้ป่วย
หมายเลข H.N.
รหัสโรค E119 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ที่ไม่มีภาวะแทรกซ้อน

รายละเอียดชื่อโรค NON-INSULIN-DEPENDENT DIABETES MELLITUS / NON-INSULIN-DEPENDENT DIABETES MELLITUS
เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ที่ไม่มีภาวะแทรกซ้อน

ผลการพยากรณ์ภาวะแทรกซ้อนของโรค

ผลการทำนาย	รหัสโรค	รายการ
7.E119=t1.726 ==> I10-I16.17 conf.(0.94)	E10-E14	เบาหวาน
	I10-I15	โรคความดันโลหิตสูง

ถ้าผู้ป่วยเป็นโรค เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ที่ไม่มีภาวะแทรกซ้อน นั้นแล้ว ผู้ป่วยจะมีโอกาสที่จะเป็นโรค ความดันโลหิตสูง มีความเป็นไปได้ 94%

อาการของโรค

อาการแทรกซ้อนต่างๆ มักจะเกิดขึ้นเมื่อเป็นเบาหวานมานาน โดยไม่ได้รับการรักษาอย่างจริงจัง หรือปล่อยปละละเลย ทั้งนี้ โรคแทรกซ้อนที่อาจพบได้ เช่น

- ตา อาจเป็นต้อกระจกก่อนวัย ประสาทตาหรือจอตา (retina) เสื่อม หรือเลือดออกในน้ำวุ้นลูกตา (vitreous hemorrhage) ทำให้มีอาการตามัวลงเรื่อยๆ หรือมองเห็นจุดดำลอยไปลอยมา และอาจทำให้ตามอดในที่สุด
- ระบบประสาท ผู้ป่วยอาจเป็นปลายประสาทอักเสบ

ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์

ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต สูง
ระดับเสี่ยงต่อการเสียชีวิต ปานกลาง

- เบาหวานเป็นโรคเรื้อรังที่ต้องรักษาติดต่อกันเป็นเวลานาน หรือตลอดชีวิต ซึ่งหากได้รับการรักษาอย่างจริงจัง อาจมีชีวิตเหมือนคนปกติได้ แต่การรักษาไม่จริงจังก็อาจมีอันตรายจากโรคแทรกซ้อนได้มาก
- ควบคุมอาหาร การลดน้ำหนัก (ถ้าอ้วน) และการออกกำลังกายอย่างสม่ำเสมอ มีความสำคัญมาก ในรายที่เป็นไม่มาก

กลุ่มงานห้องตรวจอายุรกรรม โรงพยาบาลปทุมธานี

พิมพ์รายการ ESC. ยกเลิก

ภาพที่ 3.18 หน้าจอระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

4) การพิมพ์ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวานระบบพิมพ์ ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็น เป็นโรคเบาหวานแยกออกเป็น 5 ส่วน ดังนี้

- (1) ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน
- (2) ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต และ ระดับเสี่ยงต่อการเสียชีวิต
- (3) อาการตามกลุ่มโรค
- (4) ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์
- (5) กลุ่มงานห้องตรวจผู้ป่วยเบาหวาน

ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

รายงานวันที่ : 07 สิงหาคม 2557 19:14:30

ผลการพยากรณ์ภาวะแทรกซ้อน

7. E119=t 1726 ==> I10=t 1617 conf:(0.94)

ถ้าผู้ป่วยเป็นโรค เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ที่ไม่มีภาวะแทรกซ้อน นี้แล้ว ผู้ป่วยจะมีโอกาสที่จะเป็นโรค ความดันโลหิตสูง มีความเป็นไปได้ 94%

ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต สูง

ระดับเสี่ยงต่อการเสียชีวิต ปานกลาง

อาการของโรค เบาหวาน

อาการแทรกซ้อนต่างๆ มักจะเกิดเมื่อเป็นเบาหวานมานาน โดยไม่ได้รับการรักษาอย่างจริงจัง หรือปล่อยปละละเลย ทั้งนี้ โรคแทรกซ้อนที่อาจพบได้ เช่น

1. ตา อาจเป็นต้อกระจกก่อนวัย ประสาทตาหรือจอตา (retina) เลื่อม หรือเลือดออกในน้ำวุ้นลูกตา (vitreous hemorrhage)

ทำให้มีอาการตามัวลงเรื่อยๆ หรือมองเห็นจุดดำลอยไปลอยมา และอาจทำให้ตาบอดในที่สุด

2. ระบบประสาท ผู้ป่วยอาจเป็นปลายประสาทอักเสบ มีอาการชาหรือปวดร้อนตามปลายมือปลายเท้า

ซึ่งอาจทำให้มีแผลเกิดขึ้นที่เท้าได้ง่าย (อาจลุกลามจนเท้าเน่า) บางคนอาจมีอาการวิงเวียนเนื่องจากมีภาวะความดันตกในท้ายัน

บางคนอาจไม่มีความรู้สึกทางเพศ ท้องเดินตอนกลางคืนบ่อย หรือกระเพาะปัสสาวะไม่ทำงาน (กลั้นปัสสาวะไม่อยู่ หรือไม่มีแรงเบ่งปัสสาวะ)

3. ไต มักจะเสื่อม จนเกิดภาวะไตวาย มีอาการ บวม ชีด ความดันโลหิตสูง

ซึ่งเป็นสาเหตุการตายของผู้ป่วยเบาหวานที่พบได้ค่อนข้างบ่อย

4. ผนังหลอดเลือดแดงแข็ง (atherosclerosis) ทำให้เป็นโรคความดันโลหิตสูง, อัมพาต, โรคหัวใจขาดเลือด ถ้าหลอดเลือดที่เท้าตีบแข็ง

เลือดไปเลี้ยงเท้าไม่พอ อาจทำให้เท้าเย็นเป็นตะคริว หรือปวดขณะเดินมากๆ หรืออาจทำให้เป็นแผลหายยาก หรือเท้าเน่า

(ซึ่งอาจเกิดร่วมกับการติดเชื้อ)

ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์

1. เบาหวานเป็นโรคเรื้อรังที่ต้องรักษาติดต่อกันเป็นเวลานาน หรือตลอดชีวิต ซึ่งหากได้รับการรักษาอย่างจริงจัง อาจมีชีวิตเหมือนคนปกติได้ แต่ถ้ารักษาไม่จริงจังก็อาจมีอันตรายจากโรคแทรกซ้อนได้มาก

2. ควบคุมอาหาร การลดน้ำหนัก (ถ้าอ้วน) และการออกกำลังกาย มีความสำคัญมาก ในรายที่เป็นไม่มาก ถ้าปฏิบัติในเรื่องเหล่านี้ได้ดี

อาจหายจากเบาหวานได้โดยไม่ต้องพึ่งยา ทั้งนี้ ผู้ป่วยควรอาหารที่มีผลต่อโรค ดังต่อไปนี้ ลดการกินน้ำตาล และของหวานทุกชนิด

รวมทั้งผลไม้หวานและน้ำผึ้ง และควรเลิกกินน้ำหวาน น้ำอัดลม ขนมหวาน เหล้าเบียร์ ลดการกินอาหารพวกแป้ง เช่น ข้าว ข้าวเหนียว ขนมปัง

กล้วยเดี่ยว กล้วยน้ำว้า กล้วยหอม กล้วยตาก กล้วยตากแห้ง กล้วยตากสุก กล้วยตากแห้ง กล้วยตากสุก กล้วยตากแห้ง กล้วยตากสุก กล้วยตากแห้ง กล้วยตากสุก

หันไปกินอาหารพวกโปรตีน เนื้อแดง ไข่ นม ถั่วต่างๆ รวมทั้งเพิ่มผักและผลไม้ที่ไม่หวานจัดให้มากขึ้น ออกกำลังกายอย่างสม่ำเสมอ

แต่ไม่ควรหักโหม เช่น เดินเร็ว วิ่งเหยาะ ซี่งกีรยาน ว่ายน้ำ รำมวยจีน เล่นโยคะ ภายบริหาร เป็นต้น

3. เลิกสูบบุหรี่โดยเด็ดขาด มิเช่นนั้น อาจทำให้ผนังหลอดเลือดแดงแข็งเร็วขึ้น ซึ่งเป็นต้นเหตุของโรคแทรกซ้อนต่างๆ

4. หมั่นดูแลรักษาเท้าเป็นพิเศษ ระวังอย่าให้เกิดบาดแผลหรือการอักเสบ เพราะอาจลุกลามจนกลายเป็นแผลเน่าจนต้องตัดนิ้วหรือขาทั้ง

ควรล้างเท้าให้สะอาดด้วยสบู่ เช็ดให้แห้ง โดยเฉพาะตรงซอกเท้า อย่าถูแรงๆ เวลาตัดเล็บเท้า ควรตัดออกตรงๆ อย่าตัดโค้งหรือตัดถูกเนื้อ

อย่าเดินเท้าเปล่า ระวังเหยียบลูกของมีคม นวม หรือของร้อน อย่าสวมรองเท้าคับไป หรือใส่ถุงเท้ารัดแน่นเกินไป ถ้าเป็นหูดหรือตาปลาที่เท้า

กลุ่มงานห้องตรวจอายุรกรรม โรงพยาบาลปทุมธานี

ภาพที่ 3.19 พิมพ์ผลการพยากรณ์โอกาสการเกิด โรคอื่นที่เกี่ยวข้องจากการเป็น โรคเบาหวาน

ผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน

รายงานวันที่ : 07 สิงหาคม 2557 19:14:30

ผลการพยากรณ์ภาวะแทรกซ้อน

7. E119=t 1726 ==> I10=t 1617 conf:(0.94)

ถ้าผู้ป่วยเป็นโรค เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ที่ไม่มีภาวะแทรกซ้อน นี้แล้ว ผู้ป่วยจะมีโอกาสที่จะเป็นโรค ความดันโลหิตสูง มีความเป็นไปได้ 94%

ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต สูง

ระดับเสี่ยงต่อการเสียชีวิต สูง

อาการของโรค โรคความดันโลหิตสูง

หัวใจทำงานหนักขึ้น ทำให้ผนังหัวใจหนาตัวและถ้าไม่ได้รับการรักษาอย่างถูกต้อง ผนังหัวใจจะยืดออกและเสียหายที่ ทำให้เกิดหัวใจโต และหัวใจวายได้ในที่สุด อาจเกิดภาวะหลอดเลือดในสมองตีบตันหรือแตก ทำให้เป็นอัมพาตหรือเสียชีวิตได้ ถ้าเป็นเรื้อรัง อาจกลายเป็นโรคความจำเสื่อม สมาธิลดลง เลือดอาจไปเลี้ยงไตไม่พอ เนื่องจากหลอดเลือดเสื่อม ทำให้ไตวายเรื้อรังและภาวะไตวายจะยิ่งทำให้ความดันโลหิตสูงขึ้นอีก หลอดเลือดแดงในตาจะเสื่อมลงอย่างช้าๆ อาจมีเลือดที่จอตา ทำให้ประสาทตาเสื่อม ตามัวลงเรื่อยๆ จนตาบอดได้ บางที่อาจปวดต้อกัน

โรคความดันโลหิตสูงเป็นปัจจัยเสี่ยงทำให้เกิดโรคหัวใจ โรคไต โรคหลอดเลือดหัวใจตีบ โรคอัมพาต โรคหัวใจเป็นโรคที่มีอัตราตายสูง ดังนั้นการป้องกันความดันโลหิตสูงสามารถป้องกันอัตราการตายจากโรคหัวใจ และโรคอัมพาต โรคความดันโลหิตสูงเป็นภัยเงียบที่คุกคามชีวิตของทุกท่านเนื่องจากไม่มีอาการเตือนดังนั้น การจะทราบว่าเป็นความดันโลหิตสูงจำเป็นต้องวัดความดันโลหิต

ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์

ออกกำลังกายสม่ำเสมอ ตัวอย่างเช่น การเดินเร็วๆ วิ่งเหยาะๆ หรือว่ายน้ำ ปั่นจักรยาน ควรออกกำลังกายประมาณ 15-20 นาที อย่างน้อย 3-4 ครั้ง/สัปดาห์ ควรหลีกเลี่ยงอาหารที่มีรสเค็มจัด เพื่อลดปริมาณเกลือซึ่งจะทำให้ความดันโลหิตสูงได้ ลดเครื่องดื่มแอลกอฮอล์ และงดสูบบุหรี่ ลดความเครียดของงานและภาวะแวดล้อม ลดน้ำหนักตัว โดยเฉพาะในรายที่น้ำหนักเกินมาตรฐาน ความอ้วนถือเป็นปัจจัยเสี่ยงที่สำคัญของโรคความดันโลหิตสูง รับประทานยาและพบแพทย์สม่ำเสมอ เพื่อตรวจวัดความดันโลหิตและปรับยาให้เหมาะสม

กลุ่มงานห้องตรวจอายุรกรรม โรงพยาบาลปทุมธานี

ภาพที่ 3.20 พิมพ์ผลการพยากรณ์โอกาสการเกิด โรคอื่นที่เกี่ยวข้องจากการเป็น โรคเบาหวาน

หน้า 2

บทที่ 4

ผลการวิจัย

การสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยโรคเบาหวานโรงพยาบาลปทุมธานี ด้วยนำข้อมูลผลการตรวจวินิจฉัยโรค หลังจากผู้ป่วยเป็นโรคเบาหวานแล้ว ด้วยเทคนิคการทำเหมืองข้อมูล แบบการหาความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm เป็นการหาความสัมพันธ์ของภาวะแทรกซ้อนของโรคอื่น สามารถสรุปผลการดำเนินงานประกอบด้วย 2 ส่วน ดังต่อไปนี้

1. ผลการวิเคราะห์ข้อมูล
2. พัฒนาระบบสารสนเทศ นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน เพื่อนำแบบจำลองที่ได้ไปทดลองใช้

1. ผลการวิเคราะห์ข้อมูล

เมื่อนำข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน เข้าสู่กระบวนการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยใช้กระบวนการการทำเหมืองข้อมูลในการค้นหาความสัมพันธ์ดังกล่าว ข้อมูลจะถูกนำมาเปลี่ยน Format ให้อยู่ในรูปแบบของไฟล์ WEKA Attribute-Relation File Format (.arff) เพื่อให้สามารถนำไปใช้งานกับโปรแกรม WEKA ในการค้นหาความสัมพันธ์นั้นได้กำหนดค่าความน่าเชื่อถือของความสัมพันธ์ที่ได้ดังต่อไปนี้

ค่า Lower Bound Minimum support= 0.1

ค่า Upper Bound Minimum support= 1.0

ค่า Confidence = 0.2

ค่า Rules =20

เพื่อเป็นการค้นหาความสัมพันธ์ที่มีความเชื่อถือ จึงได้ทำการกำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.1 (10%) และค่าสนับสนุนสูงสุดหรือค่าความถูกต้องสูงสุดไว้ที่ 1.0 (100%) ให้แสดงค่ากฎ (Rules) ที่ค้นพบให้แสดงรายการ 20 กฎ (Rules) ความสัมพันธ์ที่ได้จะมีความเชื่อถือตั้งแต่ 20% - 100% การค้นหาความสัมพันธ์ของข้อมูลเพื่อหาความสัมพันธ์การเกิดภาวะแทรกซ้อนของผู้ป่วยโรคเบาหวาน โดยใช้โปรแกรม WEKA

Relation: ??????1-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,4...

D696	E039	E041	E059	E160	E162	E169	E274	E780	E785	E789	E834	E835	E86	E870	E871	E872	E875	E876	E877	E88
					t				t									t		
t									t											
								t	t									t		
									t	t								t		
									t	t							t			
					t		t	t	t					t			t		t	
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					
									t						t					

ภาพที่ 4.1 ข้อมูลผลการตรวจวินิจฉัยโรคของผู้ป่วยเบาหวาน

จากข้อมูลผู้ป่วยเบาหวาน เมื่อนำมาค้นหาความสัมพันธ์ การสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน โดยใช้รหัสโรคตั้งแต่ รหัส E10- E14 มาทำการวิเคราะห์ข้อมูลที่ละรหัสเพื่อหาภาวะแทรกซ้อน ของโรคเบาหวานแต่ละโรค ต้องการหาว่าเมื่อผู้ป่วยเป็นโรคนี้อแล้วจะเป็นโรคแทรกซ้อนอะไรบ้าง ด้วยโปรแกรม WEKA ในการวิเคราะห์ข้อมูลหาความสัมพันธ์ภาวะแทรกซ้อน ของโรคเบาหวาน

1.1 รหัสโรคเบาหวาน ตั้งแต่ รหัส E10 - E14 มีทั้งหมด ตามข้อมูลรหัสโรค (ICD10: ICD-10-TM International Statistical Classification of Diseases and Related Health Problems 10th Revision, Thai Modification, 2012) 56 รหัสโรคตามตารางที่ 4.1 ผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี พบว่ามีการลงผลตรวจวินิจฉัยโรคเบาหวานทั้งหมด 11 รหัสโรคจากข้อมูลผลการตรวจวินิจฉัยของแพทย์ตามตารางที่ 4.2 ดังนี้

ตารางที่ 4.1 รหัสโรคเบาหวานทั้งหมด

No	รหัส	ชื่อโรค
1	E10	เบาหวานชนิดที่ต้องพึ่งอินซูลิน
2	E100	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับ โคมา
3	E101	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะเลือดเป็นกรดจากคีโตน
4	E102	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางไต
5	E103	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา
6	E104	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท
7	E105	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย
8	E106	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด
9	E107	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนหลายอย่าง
10	E108	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด
11	E109	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ไม่มีภาวะแทรกซ้อน
12	E11	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน
13	E110	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับ โคมา
14	E111	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะเลือดเป็นกรดจากคีโตน
15	E112	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางไต
16	E113	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา
17	E114	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท
18	E115	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย
19	E116	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด
20	E117	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนหลายอย่าง
21	E118	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด
22	E119	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ไม่มีภาวะแทรกซ้อน
23	E12	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ
24	E120	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับ โคมา
25	E121	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะเลือดเป็นกรดจากคีโตน
26	E122	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนทางไต

ตารางที่ 4.1 (ต่อ)

No	รหัส	ชื่อโรค
27	E123	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนทางตา
28	E124	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนทางประสาท
29	E125	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย
31	E126	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด
32	E127	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนหลายอย่าง
33	E128	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด
34	E129	เบาหวานเกี่ยวกับภาวะทุพโภชนาการ ไม่มีภาวะแทรกซ้อน
35	E13	เบาหวานชนิดอื่นที่ระบุรายละเอียด
36	E130	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับโคมา
37	E131	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะเลือดเป็นกรดจากคีโตน
38	E132	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางไต
39	E133	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางตา
40	E134	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางประสาท
41	E135	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย
42	E136	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด
43	E137	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนหลายอย่าง
44	E138	เบาหวานชนิดอื่นที่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด
45	E139	เบาหวานชนิดอื่นที่ระบุรายละเอียด ไม่มีภาวะแทรกซ้อน
46	E14	เบาหวานที่ไม่ระบุรายละเอียด
47	E140	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับโคมา
48	E141	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะเลือดเป็นกรดจากคีโตน
49	E142	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางไต
50	E143	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางตา
51	E144	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางประสาท

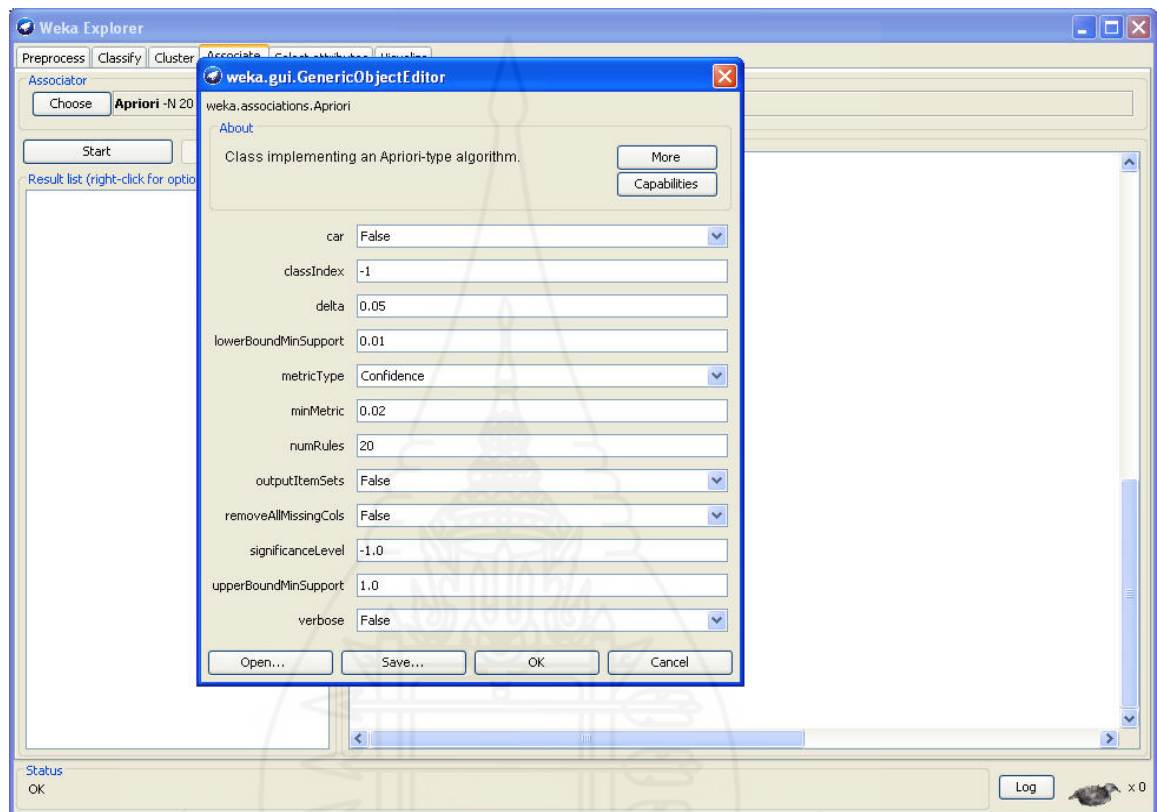
ตารางที่ 4.1 (ต่อ)

No	รหัส	ชื่อโรค
52	E145	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย
53	E146	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนอื่นที่ระบุรายละเอียด
54	E147	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนหลายอย่าง
55	E148	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนที่ไม่ระบุรายละเอียด
56	E149	เบาหวานที่ไม่ระบุรายละเอียด ไม่มีภาวะแทรกซ้อน

ตารางที่ 4.2 ผลตรวจวินิจฉัยโรคเบาหวาน ในผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

No	รหัส	ชื่อโรค	ครั้ง
1	E10	เบาหวานชนิดที่ต้องพึ่งอินซูลิน	146
2	E100	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับโคม่า	28
3	E103	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา	54
4	E109	เบาหวานชนิดที่ต้องพึ่งอินซูลิน ไม่มีภาวะแทรกซ้อน	1,621
5	E113	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา	146
6	E114	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท	379
7	E115	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย	75
8	E119	เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ไม่มีภาวะแทรกซ้อน	1,729
9	E144	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางประสาท	152
10	E145	เบาหวานที่ไม่ระบุรายละเอียด ร่วมกับภาวะแทรกซ้อนทางระบบไหลเวียนส่วนปลาย	251
11	E149	เบาหวานที่ไม่ระบุรายละเอียด ไม่มีภาวะแทรกซ้อน	146

1.2 ทำการวิเคราะห์ข้อมูลที่ तरहที่สเพื่อหาภาวะแทรกซ้อน ของโรคเบาหวานแต่ละโรค โดยต้องการหาว่าเมื่อผู้ป่วยเป็นโรคนี้แล้ว จะเป็นโรคแทรกซ้อนอะไรบ้าง การนำกฎความสัมพันธ์ที่ได้จากการวิเคราะห์ ผู้วิจัยจะเลือกเฉพาะกฎความสัมพันธ์ที่มีค่าสนับสนุนหรือค่าความถูกต้องขั้นต่ำที่ 10 % สูงสุดที่ 100% ค่าความเชื่อถือที่เท่ากับหรือมากกว่า 20 % ถึง 100 % เท่านั้น



ภาพที่ 4.2 กำหนดค่าสนับสนุน กำหนดค่าความเชื่อถือ

1.3 ผลการวิเคราะห์ข้อมูลหาความสัมพันธ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

1.3.1 $E103 \Rightarrow H3600$ $\text{conf:}(0.65)$ แสดงว่า ถ้าผู้ป่วยเป็น E103 เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็น โรคแทรกซ้อน H3600 จอตาจากโรคเบาหวานชนิดไม่ออกขยาย (เอ็นพีดีอาร์) คิดเป็น 65 %

1.3.2 $E103 \Rightarrow H251$ $\text{conf:}(0.46)$ แสดงว่า ถ้าผู้ป่วยเป็นโรค E103 เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรกซ้อน H251 ต้อกระจกในผู้สูงอายุ คิดเป็น 46 %

1.3.3 $E103=t 54 \implies H3602=t 24$ conf:(0.44) แสดงว่า ถ้าผู้ป่วยเป็นโรค E103 เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรกซ้อน H3602 จอตาจากโรคเบาหวานชนิดงอกขยาย คิดเป็น 44 %

1.3.4 $E113=t 162 \implies H3600=t 124$ conf:(0.77) แสดงว่า ถ้าผู้ป่วยเป็นโรค E113 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรก H3600 จอตาจากโรคเบาหวานชนิดไม่งอกขยาย (เอ็นพีดีอาร์) คิดเป็น 77 %

1.3.5 $E113=t 162 \implies H251=t 79$ conf:(0.49) แสดงว่า ถ้าผู้ป่วยเป็นโรค E113 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรก H251 ต □ อกระจกในผู้สูงอายุ คิดเป็น 49 %

1.3.6 $E113=t 162 \implies H3602=t 58$ conf:(0.36) แสดงว่า ถ้าผู้ป่วยเป็นโรค E113 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรก H3602 จอตาจากโรคเบาหวานชนิดงอกขยาย คิดเป็น 36 %

1.3.7 $E113=t 162 \implies N189=t 33$ conf:(0.2) แสดงว่า ถ้าผู้ป่วยเป็นโรค E113 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา แล้วจะเป็นโรคแทรก N189 ไตวายเรื้อรังที่มีไต □ ระบายละเอียด คิดเป็น 20 %

1.3.8 $E114=t 379 \implies H251=t 92$ conf:(0.24) แสดงว่า ถ้าผู้ป่วยเป็น E114 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท แล้วจะเป็นโรคแทรก H251 ต □ อกระจกในผู้สูงอายุ คิดเป็น 24 %

1.3.9 $E114=t 379 \implies I259=t 77$ conf:(0.2) แสดงว่า ถ้าผู้ป่วยเป็น E114 เบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท แล้วจะเป็นโรคแทรก I259 โรคหัวใจขาดเลือดเรื้อรังที่มีไต □ ระบายละเอียด คิดเป็น 20 %

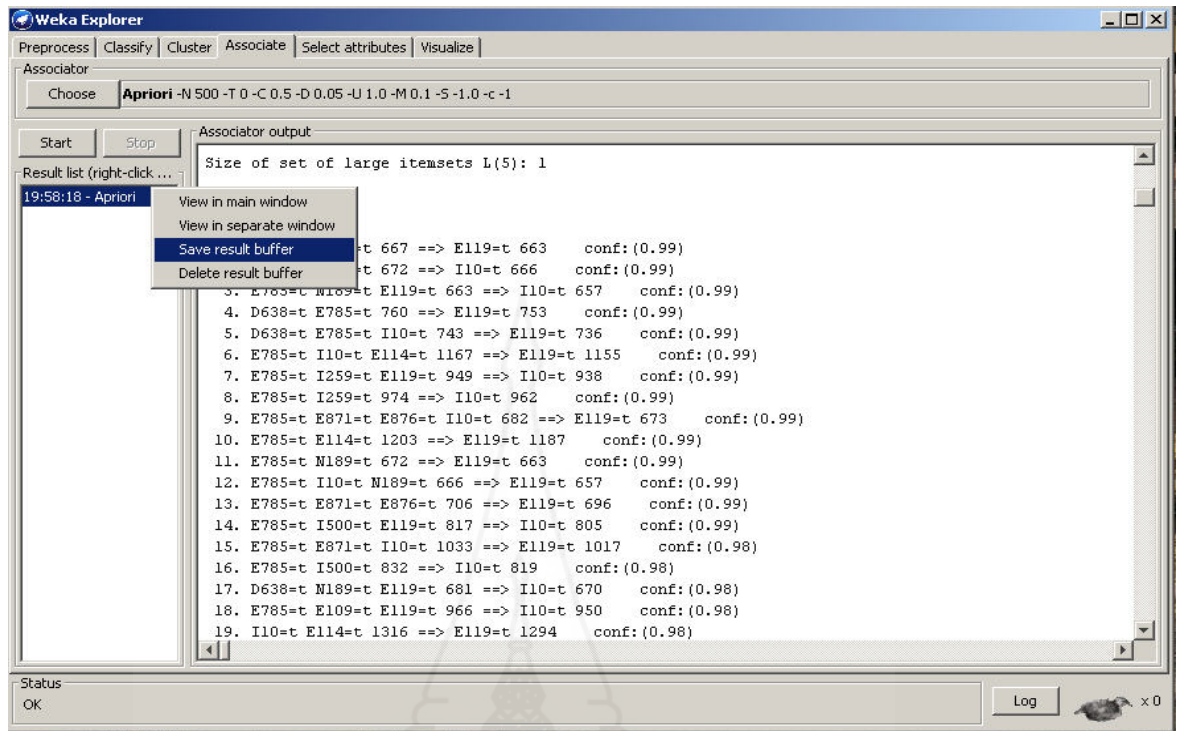
1.4 สรุปผลการวิเคราะห์ข้อมูลภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ผู้ป่วยจะมีโอกาสที่จะเป็น โรคดังนี้

ตารางที่ 4.3 โรคที่ผู้ป่วยมีโอกาสที่จะเป็นหลังจากเป็นโรคเบาหวาน

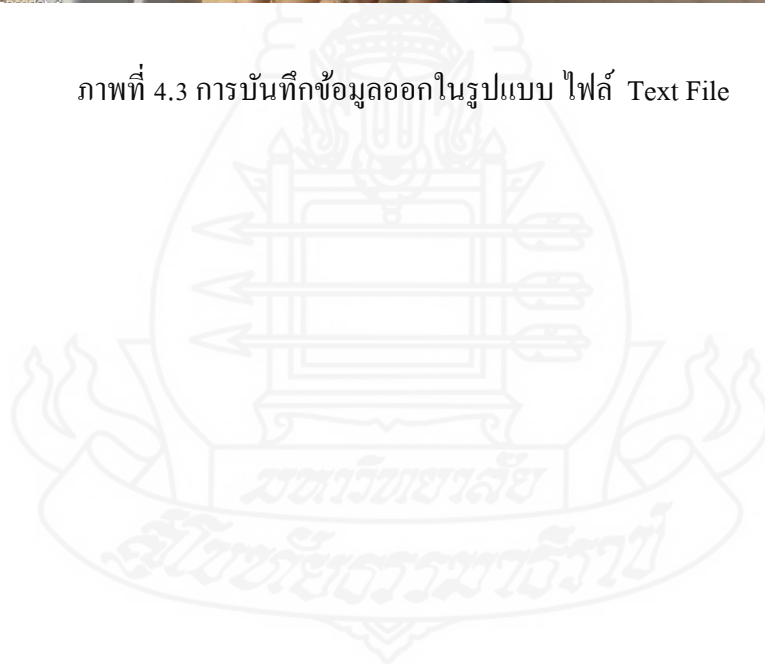
No	รหัส	ชื่อโรค	ครั้ง
1	H251	ต ^๓ อกระจกในผู้สูงอายุ	393
2	H3600	จอตตาจากโรคเบาหวานชนิดไม่ดออกซาย (เอ็นพีดีอาร์)	180
3	H3602	จอตตาจากโรคเบาหวานชนิดดออกซาย	99
4	I259	โรคหัวใจขาดเลือดเรื้อรังที่มีไต ^๓ ระบุนรายละเอียด	268
5	N189	ไตวายเรื้อรังที่มีไต ^๓ ระบุนรายละเอียด	245

2. การพัฒนาระบบโปรแกรม นำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน เพื่อนำแบบจำลองที่ได้ไปทดลองใช้

นำแบบจำลองที่ได้ไปทดลองใช้ โดยพัฒนาระบบสารสนเทศ ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 ให้กับเจ้าหน้าที่ทดลองใช้ระบบสารสนเทศ นำเสนอผลการพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ข้อมูลที่นำเสนอมีประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็น อัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์ ได้พัฒนาระบบโปรแกรมชื่อ ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน



ภาพที่ 4.3 การบันทึกข้อมูลออกในรูปแบบ ไฟล์ Text File



	1	2	3	4	5	6	7	8	9	0	1
1	1.	E162=t	E785=t	667 ==>	E119=t	663	conf.(0.99)				
2	2.	E785=t	N189=t	672 ==>	I10=t	666	conf.(0.99)				
3	3.	E119=t	E785=t	N189=t	663 ==>	I10=t	657	conf.(0.99)			
4	4.	D638=t	E785=t	760 ==>	E119=t	753	conf.(0.99)				
5	5.	D638=t	E785=t	I10=t	743 ==>	E119=t	736	conf.(0.99)			
6	6.	E114=t	E785=t	I10=t	1167 ==>	E119=t	1155	conf.(0.99)			
7	7.	E119=t	E785=t	I259=t	949 ==>	I10=t	938	conf.(0.99)			
8	8.	E785=t	I259=t	974 ==>	I10=t	962	conf.(0.99)				
9	9.	E785=t	E871=t	E876=t	I10=t	682 ==>	E119=t	673	conf.(0.99)		
10	10.	E114=t	E785=t	1203 ==>	E119=t	1187	conf.(0.99)				
11	11.	E785=t	N189=t	672 ==>	E119=t	663	conf.(0.99)				
12	12.	E785=t	I10=t	N189=t	666 ==>	E119=t	657	conf.(0.99)			
13	13.	E785=t	E871=t	E876=t	706 ==>	E119=t	696	conf.(0.99)			
14	14.	E119=t	E785=t	I500=t	817 ==>	I10=t	805	conf.(0.99)			
15	15.	E785=t	E871=t	I10=t	1033 ==>	E119=t	1017	conf.(0.98)			
16	16.	E785=t	I500=t	832 ==>	I10=t	819	conf.(0.98)				
17	17.	D638=t	E119=t	N189=t	681 ==>	I10=t	670	conf.(0.98)			
18	18.	E109=t	E119=t	E785=t	966 ==>	I10=t	950	conf.(0.98)			
19	19.	E114=t	I10=t	1316 ==>	E119=t	1294	conf.(0.98)				
20	20.	E785=t	E871=t	1073 ==>	E119=t	1055	conf.(0.98)				
21	21.	E785=t	I10=t	I500=t	819 ==>	E119=t	805	conf.(0.98)			
22	22.	D638=t	N189=t	698 ==>	I10=t	686	conf.(0.98)				
23	23.	E785=t	I500=t	832 ==>	E119=t	817	conf.(0.98)				
24	24.	E162=t	I10=t	973 ==>	E119=t	955	conf.(0.98)				
25	25.	E109=t	E785=t	1089 ==>	I10=t	1068	conf.(0.98)				
26	26.	E119=t	I259=t	1169 ==>	I10=t	1145	conf.(0.98)				
27	27.	I259=t	1207 ==>	I10=t	1182	conf.(0.98)					
28	28.	E785=t	N189=t	672 ==>	E119=t	I10=t	657	conf.(0.98)			
29	29.	D638=t	E785=t	760 ==>	I10=t	743	conf.(0.98)				

ภาพที่ 4.4 ข้อมูลในรูปแบบ ไฟล์ Text File

2.1 โครงสร้างฐานข้อมูล ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นเบาหวาน

2.1.1 โครงสร้างข้อมูลรหัสโรค (ICD101: ICD-10-TM International Statistical Classification of Diseases and Related Health Problems 10th Revision, Thai Modification, 2012) ประกอบไปด้วย ดังนี้

ตารางที่ 4.4 โครงสร้างข้อมูลรหัสโรค

No.	Name	Data Type	Value	Description
1	CODE	Character		รหัสโรค
2	DESCT	Character		ชื่อรหัสโรค ภาษาไทย
3	DESC	Character		ชื่อรหัสโรค ภาษาอังกฤษ

2.1.2 โครงสร้างข้อมูลทะเบียนประวัติผู้ป่วย (Patients) ที่มารับบริการในโรงพยาบาล ปทุมธานี ประกอบไปด้วย ดังนี้

ตารางที่ 4.5 โครงสร้างข้อมูลเพิ่มประวัติผู้ป่วย

No.	Name	Data Type	Value	Description
1	Hn	Character		หมายเลขทะเบียนจำตัวผู้ป่วย ไม่ซ้ำกัน
2	Title	Character		คำนำหน้า
3	Name	Character		ชื่อผู้ป่วย
4	Surname	Character		นามสกุลผู้ป่วย

2.1.3 โครงสร้างข้อมูลกลุ่มโรค (Groupdx : Disease Group) ประกอบไปด้วย ดังนี้

ตารางที่ 4.6 โครงสร้างข้อมูลรหัสกลุ่มโรค

No.	Name	Data Type	Value	Description
1	CODE	Character		รหัสกลุ่มโรค
2	NAME	Character		ชื่อกลุ่มโรค
3	COMMENT	Character		อาการของโรค
4	COMMENT1	Character		ข้อควรปฏิบัติ/คำแนะนำ จากแพทย์
5	TYPE1	Numeric	1=ต่ำ 2=ปานกลาง 3=สูง	ระดับเสี่ยงต่อการเป็น อัมพฤกษ์-อัมพาต
6	TYPE2	Numeric	1=ต่ำ 2=ปานกลาง 3=สูง	ระดับเสี่ยงต่อการเสียชีวิต

2.1.3 โครงสร้างข้อมูลข้อมูลผลการวิเคราะห์ข้อมูลจากโปรแกรม WEKA รูปแบบไฟล์ประกอบไปด้วย ดังนี้

ตารางที่ 4.7 โครงสร้างข้อมูลข้อมูลผลการวิเคราะห์ข้อมูลจากโปรแกรม WEKA

No	Name	Data Type	Value	Description
1	NAME	Character		ผลการวิเคราะห์ข้อมูล
2	NUM	Numeric		จำนวนผลการวิเคราะห์ข้อมูล
3	BDX1	Character		ลำดับ
4	BDX2	Character		ผลการวิเคราะห์ข้อมูลที่ 1
5	BDX3	Character		ผลการวิเคราะห์ข้อมูลที่ 2
6	BDX4	Character		ผลการวิเคราะห์ข้อมูลที่ 3
7	BDX5	Character		ผลการวิเคราะห์ข้อมูลที่ 4
8	BDX6	Character		ผลการวิเคราะห์ข้อมูลที่ 5
9	BDX7	Character		ผลการวิเคราะห์ข้อมูลที่ 6
10	BDX8	Character		ผลการวิเคราะห์ข้อมูลที่ 7
11	BDX9	Character		ผลการวิเคราะห์ข้อมูลที่ 8
12	BDX10	Character		ผลการวิเคราะห์ข้อมูลที่ 9
13	Conf	Character		ค่าความเชื่อมั่น

2.2 การนำเข้าข้อมูล รูปแบบ Text File ให้เป็นรูปแบบฐานข้อมูล

ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน โดยทำการอ่านข้อมูลจาก ผลการวิเคราะห์ข้อมูล ที่ทำส่งออกเป็นรูปแบบ Text File ให้เป็นรูปแบบฐานข้อมูล ดาต้าเบส (DBF: Database File) พัฒนาโดย Microsoft Visual Foxpro ข้อมูลตามภาพที่ 4.4 ระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน จะทำการอ่านข้อมูล และจัดแยกข้อมูลผลการวิเคราะห์ให้เป็นรูปแบบฐานข้อมูลตามภาพที่ 4.5

2.2.1 ผลการวิเคราะห์ข้อมูลจากโปรแกรม WEKA ในรูปแบบฐานข้อมูล

Name	Num	Bdx1	Bdx2	Bdx3	Bdx4	Bdx5	Bdx6	Bdx7	Bdx8	Bdx9	Bdx10	Conf
11. E103=t 54 ==> H251=t 25 conf.(0.46)	7	11.	E103	54	==>	H251	25					conf.(0.46)
6. E103=t 54 ==> H3600=t 35 conf.(0.65)	7	6.	E103	54	==>	H3600	35					conf.(0.65)
12. E103=t 54 ==> H3602=t 24 conf.(0.44)	7	12.	E103	54	==>	H3602	24					conf.(0.44)
170. E109=t 1621 ==> H103=t 39 conf.(0.02)	7	170.	E109	1621	==>	H103	39					conf.(0.02)
164. E109=t 1621 ==> H110=t 44 conf.(0.03)	7	164.	E109	1621	==>	H110	44					conf.(0.03)
166. E109=t 1621 ==> H111=t 41 conf.(0.03)	7	166.	E109	1621	==>	H111	41					conf.(0.03)
89. E109=t 1621 ==> H251=t 316 conf.(0.19)	7	89.	E109	1621	==>	H251	316					conf.(0.19)
161. E109=t 1621 ==> H251=t H3600=t 78 conf.(0.05)	8	161.	E109	1621	==>	H251	H3600	78				conf.(0.05)
163. E109=t 1621 ==> H251=t I259=t 45 conf.(0.03)	8	163.	E109	1621	==>	H251	I259	45				conf.(0.03)
165. E109=t 1621 ==> H251=t N189=t 42 conf.(0.03)	8	165.	E109	1621	==>	H251	N189	42				conf.(0.03)
173. E109=t 1621 ==> H264=t 35 conf.(0.02)	7	173.	E109	1621	==>	H264	35					conf.(0.02)
140. E109=t 1621 ==> H3600=t 148 conf.(0.09)	7	140.	E109	1621	==>	H3600	148					conf.(0.09)
160. E109=t 1621 ==> H3602=t 78 conf.(0.05)	7	160.	E109	1621	==>	H3602	78					conf.(0.05)
167. E109=t 1621 ==> H401=t 41 conf.(0.03)	7	167.	E109	1621	==>	H401	41					conf.(0.03)
162. E109=t 1621 ==> I219=t 48 conf.(0.03)	7	162.	E109	1621	==>	I219	48					conf.(0.03)
151. E109=t 1621 ==> I249=t 118 conf.(0.07)	7	151.	E109	1621	==>	I249	118					conf.(0.07)
159. E109=t 1621 ==> I249=t I259=t 81 conf.(0.05)	8	159.	E109	1621	==>	I249	I259	81				conf.(0.05)
168. E109=t 1621 ==> I251=t 40 conf.(0.02)	7	168.	E109	1621	==>	I251	40					conf.(0.02)
114. E109=t 1621 ==> I259=t 223 conf.(0.14)	7	114.	E109	1621	==>	I259	223					conf.(0.14)
172. E109=t 1621 ==> I259=t N189=t 39 conf.(0.02)	8	172.	E109	1621	==>	I259	N189	39				conf.(0.02)
157. F109=t 1621 ==> I64=t 97 conf.(0.06)	7	157.	F109	1621	==>	I64	97					conf.(0.06)

ภาพที่ 4.5 ตารางเก็บผลการวิเคราะห์ข้อมูลรูปแบบฐานข้อมูล



บทที่ 5

สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาวิจัยเรื่อง การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล ผู้วิจัยทำการทดสอบ และวิเคราะห์หาความรู้ใหม่โดยการทำเหมืองข้อมูลได้ข้อสรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะดังนี้

1. สรุปการวิจัย

ผลการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อน โรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน และนำแบบจำลองที่ได้ไปทดลองใช้ โดยพัฒนาระบบสารสนเทศ ทดลองใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 ให้กับเจ้าหน้าที่ทดลองใช้ระบบสารสนเทศ นำเสนอผลการพยากรณ์ โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ข้อมูลที่นำเสนอมีประวัติผู้ป่วย ผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์

1.1 จัดเตรียมข้อมูลเพื่อใช้ในการพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน การจัดเตรียมข้อมูลให้มีความถูกต้อง เพื่อให้แบบจำลองมีความแม่นยำมากที่สุด ทั้งนี้ข้อมูลที่ใช้ในการสร้าง ข้อมูลผลตรวจการวินิจฉัยโรค เริ่มตั้งแต่ที่ผู้ป่วยเป็นโรคเบาหวานเพื่อนำไปใช้ในการสร้างแบบจำลอง ตัวแปรที่สำคัญได้แก่ จำนวนครั้งการมาตรวจรักษากับแพทย์ จำนวนโรคที่ทำการลงทะเบียน เฉพาะผู้ป่วยเบาหวานเท่านั้น ทำการเลือกข้อมูลผลการตรวจวินิจฉัยโรคผู้ป่วยเบาหวาน ตามรหัสโรค E10-E14 ตัดข้อมูลโรคที่ทำการลงผลการตรวจวินิจฉัยให้กับผู้ป่วยเบาหวานที่น้อยกว่า 50 ครั้ง ตัดข้อมูลผู้ป่วยที่มาตรวจรักษาไม่ถึง 10 ครั้ง และจัดเก็บข้อมูลผลการตรวจวินิจฉัยโรคเบาหวานตามเงื่อนไขดังกล่าว โดยโปรแกรม MicroSoft Visual Foxpro และส่งข้อมูลออกมาเป็นรูปแบบ File Microsoft Excel 2010 เพื่อเป็นข้อมูลแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน แปลงรูปแบบข้อมูลปรับโครงสร้างข้อมูลให้เหมาะสม

1.2 สร้างแบบจำลองการพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากเป็นโรคเบาหวาน การจัดทำแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ใช้โปรแกรม WEKA ในการวิเคราะห์ข้อมูลเพื่อหาภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็น

โรคเบาหวาน โดยวิธีการหาความสัมพันธ์ของข้อมูล (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm แล้วนำผลพยากรณ์มาเปรียบเทียบหาค่าความถูกต้องแม่นยำ โดยใช้ผลที่มีความเชื่อถือมากที่สุด จึงได้ทำการกำหนดค่าสนับสนุนหรือค่าความถูกต้อง ขั้นต่ำที่ 0.1 (10%) สูงสุดที่ 1.0 (100%) ให้แสดงค่ากฎ (Rules) ค่าความเชื่อถือที่เท่ากับหรือมากกว่า 0.2 (20 % ถึง 100 %) การค้นหาความสัมพันธ์ของข้อมูล

1.3 วิเคราะห์ข้อมูลหาความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โดยวิธีการวิเคราะห์หาความสัมพันธ์ของข้อมูล (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm จากผลที่วิเคราะห์ข้อมูลทั้ง 2 อัลกอริทึมนั้นได้ผลมีความค่าความเชื่อถือใกล้เคียงกัน เลือกผลที่มีความค่าสนับสนุนตั้งแต่ 20 % ขึ้นไป จากผลการวิเคราะห์ข้อมูล ดังนี้

1.3.1 วิเคราะห์ข้อมูลด้วยขั้นตอน Apriori Algorithm เป็นขั้นตอนวิธีพื้นฐานในการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยและสร้างกฎความสัมพันธ์ได้ผลดังนี้

- (1) $E103=t\ 54 \implies H3600=t\ 35$ conf:(0.65)
- (2) $E103=t\ 54 \implies H251=t\ 25$ conf:(0.46)
- (3) $E103=t\ 54 \implies H3602=t\ 24$ conf:(0.44)
- (4) $E113=t\ 162 \implies H3600=t\ 124$ conf:(0.77)
- (5) $E113=t\ 162 \implies H251=t\ 79$ conf:(0.49)
- (6) $E113=t\ 162 \implies H3602=t\ 58$ conf:(0.36)
- (7) $E113=t\ 162 \implies N189=t\ 33$ conf:(0.2)
- (8) $E114=t\ 379 \implies H251=t\ 92$ conf:(0.24)
- (9) $E114=t\ 379 \implies I259=t\ 77$ conf:(0.2)

1.3.2 วิเคราะห์ข้อมูลด้วยขั้นตอน FPGrowth Algorithm ขั้นตอนเพิ่มเติมเพื่อลดจำนวนของการอ่านข้อมูลจากฐานข้อมูล พร้อมทั้งนำเสนอโครงสร้างข้อมูลแบบใหม่ขึ้นมาที่มีชื่อว่า FP-Tree โดยใช้ชื่อว่าขั้นตอนวิธีแบบ FP-Growth เป็นขั้นตอนวิธีที่ช่วยให้มีการอ่านข้อมูลข้อมูลเพียง 2 ครั้งและไม่มีการสร้างกลุ่มข้อมูลที่ซิงได้ผลดังนี้

- (1) $[E103=t]:\ 54 \implies [H3600=t]:\ 35$ \langle conf:(0.65) \rangle lift:(7.44) lev:(0.01) conv:(2.46)
- (2) $[E103=t]:\ 54 \implies [H251=t]:\ 25$ \langle conf:(0.46) \rangle lift:(2.43) lev:(0.01) conv:(1.46)

- (3) [E103=t]: 54 \implies [H3602=t]: 24 <conf:(0.44)> lift:(9.28) lev:(0.01) conv:(1.66)
 (4) [E113=t]: 162 \implies [H3600=t]: 124 <conf:(0.77)> lift:(8.79) lev:(0.05) conv:(3.79)
 (5) [E113=t]: 162 \implies [H251=t]: 79 <conf:(0.49)> lift:(2.56) lev:(0.02) conv:(1.56)
 (6) [E113=t]: 162 \implies [H3602=t]: 58 <conf:(0.36)> lift:(7.48) lev:(0.02) conv:(1.47)
 (7) [E113=t]: 162 \implies [N189=t]: 33 <conf:(0.2)> lift:(1.72) lev:(0.01) conv:(1.1)
 (8) [E114=t]: 379 \implies [H251=t]: 92 <conf:(0.24)> lift:(1.28) lev:(0.01) conv:(1.07)
 (9) [E114=t]: 379 \implies [I259=t]: 77 <conf:(0.2)> lift:(1.57) lev:(0.01) conv:(1.09)

2. อภิปรายผล

ผลการพยากรณ์ความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี พบว่าผู้ป่วยที่เป็นโรค 1) เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (E103 Insulin-dependent diabetes mellitus type 1 at with ophthalmic complication) แล้วจะเป็นโรคแทรกซ้อนดังนี้ 1.1) โรคจอตาจากโรคเบาหวานชนิดไม่องกขยาย (เอ็นพีดีอาร์) (H3600 Nonproliferative diabetic retinopathy (NPDR)) คิดเป็น 65 % 1.2) โรคต้อกระจกในผู้สูงอายุ (H251 Senile nuclear cataract) คิดเป็น 46 % 1.3) โรคจอตาจากโรคเบาหวานชนิดองกขยาย (พีดีอาร์) (H3602 Proliferative diabetic retinopathy (PDR)) คิดเป็น 44 % 2) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (E113 Non-insulin-dependent diabetes mellitus type 2 at with ophthalmic complication) แล้วจะเป็นโรคแทรกซ้อน 2.1) โรคจอตาจากโรคเบาหวานชนิดไม่องกขยาย (เอ็นพีดีอาร์) (H3600 Nonproliferative diabetic retinopathy (NPDR)) คิดเป็น 77 % 2.2) โรคต้อกระจกในผู้สูงอายุ (H251 Senile nuclear cataract) คิดเป็น 49 % 2.3) โรคจอตาจากโรคเบาหวานชนิดองกขยาย (พีดีอาร์) (H3602 Proliferative diabetic retinopathy (PDR)) คิดเป็น 36 % 2.4) โรคไตวายเรื้อรังที่มีไต □ ระบุรายละเอียด (N189 Chronic renal failure, unspecified) คิดเป็น 20 % 3) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท (E114 Chronic renal failure, unspecified) แล้วจะเป็นโรคแทรกซ้อน 3.1) โรคต้อกระจกในผู้สูงอายุ (H251 Senile nuclear cataract) คิดเป็น 24 % 3.2) โรคหัวใจขาดเลือดเรื้อรังที่มีได้ ระบุรายละเอียด (I251 Atherosclerotic heart disease) คิดเป็น 20 %

จากที่ได้กล่าวข้างต้นเป็นการสรุปผลลัพธ์ที่ได้จากการศึกษาวิจัย ซึ่งผลสรุปดังกล่าวสามารถนำมาอภิปรายผลได้ดังนี้ การเปรียบเทียบผลของการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โรงพยาบาลปทุมธานี พบว่า การวิเคราะห์หา

ความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm มีผลการพยากรณ์ที่มีค่าสนับสนุนตั้งแต่ 20 % ขึ้นไปใน 3 อันดับแรก ผลพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวานเท่านั้น แสดงว่าการหาความสัมพันธ์ (Association) ระหว่าง Apriori Algorithm และ FPGrowth Algorithm มีผลพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน โรงพยาบาลปทุมธานี มีค่าสนับสนุนและค่าความเชื่อถือ โดดเดี่ยวกัน จากผลการทดลองที่ได้ จะเห็นว่าผลการวิเคราะห์ข้อมูล ใช้ข้อมูลชุดเดียว ในการทดสอบจะพบว่าผลการวิเคราะห์ข้อมูลด้วย Apriori Algorithm ใช้เวลามากกว่า FPGrowth Algorithm ในการหาผลพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน จากการกำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อถือขั้นต่ำที่เท่ากัน จากผลพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน สอดคล้องสนับสนุนกับ เกวลี ผาใต้ (2553) ได้ทำการศึกษาเกี่ยวกับเรื่อง การพยากรณ์โอกาสเกิดภาวะเสี่ยงของผู้ป่วยเบาหวานโดยใช้โครงข่ายประสาทเทียม ด้วยวิธีการจำแนกข้อมูลผู้ป่วยเบาหวานซึ่งได้แก่ เพศ อายุ ข้อมูลปัจจัยที่มีผลที่ทำให้มีความเสี่ยงต่อการเกิดภาวะแทรกซ้อน โดยประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) เพื่อวิเคราะห์หาความเสี่ยงของการเกิดโรคแทรกซ้อน 2 โรค คือ โรคไตวายเรื้อรัง และโรคไขมันในเลือดสูง และสอดคล้องสนับสนุนกับ Chien-Lung Chan (2008) ได้ทำการศึกษาเกี่ยวกับเรื่อง การสอบสวนของภาวะแทรกซ้อนของโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล ใช้ทฤษฎีและการวิเคราะห์เชิงตัวเลขสำรวจความสัมพันธ์ระหว่างข้อมูลทางสรีรวิทยา โดยใช้การวิเคราะห์ของ การทำเหมืองข้อมูลโดยใช้ C5.0 และโครงข่ายประสาทเทียม พบว่า ผลการทำนาย ภาวะแทรกซ้อน จอประสาทตาโดยใช้ C5.0 เป็น 58.62 และ 74.73 และโครงข่ายประสาทเทียม เป็น 59.48 และ 99.86 และภาวะแทรกซ้อนของโรคไต โดยใช้ C5.0 เป็น 69.44 และ 81.36 และโครงข่ายประสาทเทียม เป็น 74.44 และ 98.55

3. ข้อเสนอแนะ

3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

3.1.1 การศึกษาครั้งนี้มีการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ด้วยใช้ข้อมูลประวัติผลการวินิจฉัยโรคโรงพยาบาลปทุมธานี โดยใช้เทคนิคการทำเหมืองข้อมูลหากนำไปใช้จริง ควรพัฒนาโปรแกรมให้สามารถดึงค่าปัจจัยต่างๆ มาอัตโนมัติเพื่อให้สะดวกกับผู้ใช้งาน

3.1.2 หากนำต้นแบบระบบแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวาน ควรประเมินต้นแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานอีกครั้ง เพื่อนำผลการผลการวิเคราะห์มาพัฒนาแนวทางการทำเหมืองข้อมูล และการประยุกต์ใช้ เพื่อให้ต้นแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่น หลังจากผู้ป่วยเป็นโรคเบาหวานมีความแม่นยำยิ่งขึ้น

3.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

3.2.1 ควรนำข้อมูลประวัติการรักษาด้านอื่น มาวิเคราะห์ข้อมูลในการสนับสนุนการตัดสินใจ เช่น ข้อมูลนัดหมายผู้ป่วย การได้รับยา/เวชภัณฑ์ ข้อมูลการทำหัตถการ เพศ สมรส การสูบบุหรี่ เป็นต้น

3.2.2 ควรมีการนำเทคนิคการสร้างเหมืองข้อมูลแบบอื่น เข้ามาใช้ในการค้นหาความสัมพันธ์ การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วยโรคเบาหวาน เพื่อที่จะได้นำผลของความสัมพันธ์ในแต่ละเทคนิคมาทำการเปรียบเทียบ เพื่อเลือกเทคนิคที่ให้ค่าความถูกต้องสูงสุด

3.2.3 ควรมีการนำข้อมูลปัจจัยเสี่ยงจากประชากรในหลายพื้นที่มาใช้ในการสร้างแบบ Model เพื่อที่จะทำให้ได้ข้อมูลที่มีความครอบคลุมมากขึ้น

3.2.4 เนื่องจากการวัดประสิทธิภาพการพยากรณ์สามารถทำได้หลากหลายวิธีการ ดังนั้นในการศึกษาครั้งต่อไป ผู้ศึกษาจะมุ่งศึกษาเพื่อหาวิธีการที่ดีที่สุดเพื่อเพิ่มประสิทธิภาพในการวัดความถูกต้องของการพยากรณ์ต่อไป

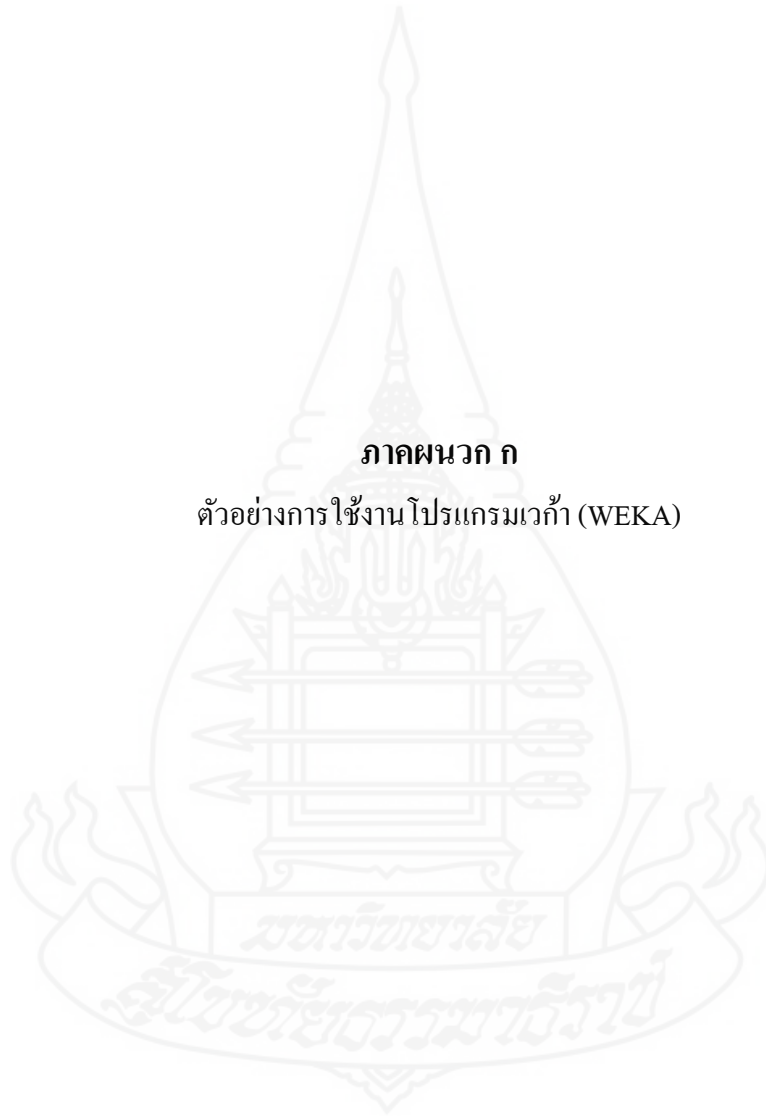


ภาคผนวก



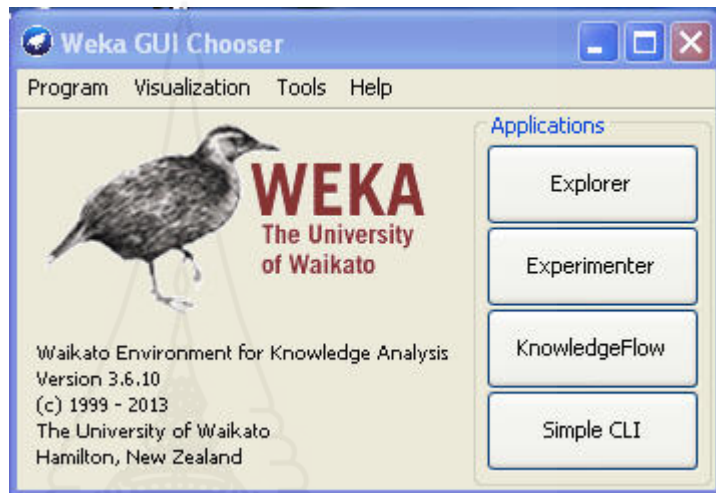
ภาคผนวก ก

ตัวอย่างการใช้งานโปรแกรมเวก้า (WEKA)



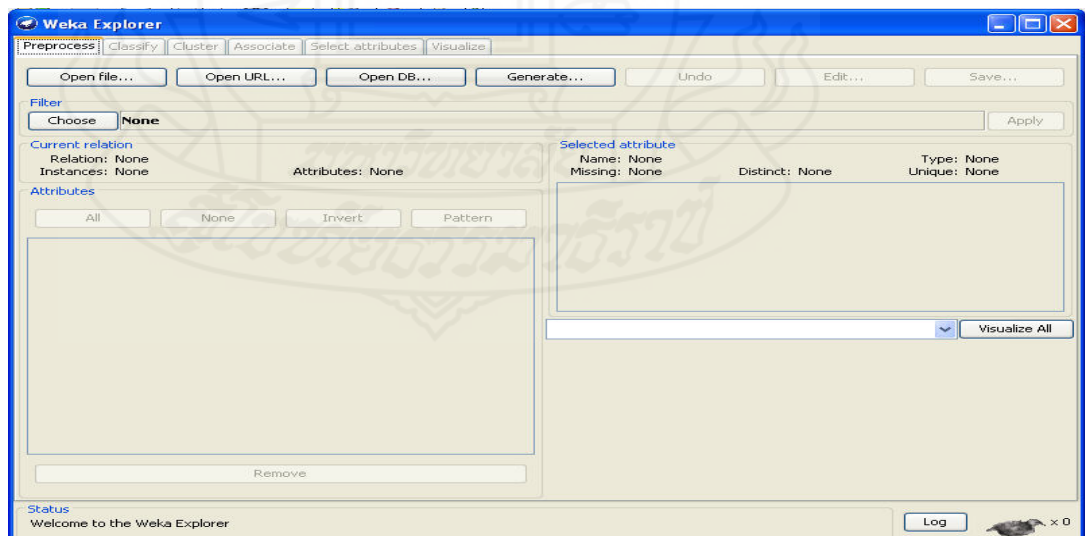
โปรแกรมเวก้า (WEKA)

WEKA ย่อมาจาก Waikato Environment for Knowledge Analysis พัฒนาด้วยภาษาจาวาทั้งหมดพัฒนาโดยเน้นทางด้าน Machine learning และ Data mining มีโมดูลย่อยสำหรับจัดการข้อมูล ใช้ GUI และคำสั่งในการสั่งให้ซอฟต์แวร์ประมวลผล ประกอบไปด้วย 4 ส่วน ดังนี้



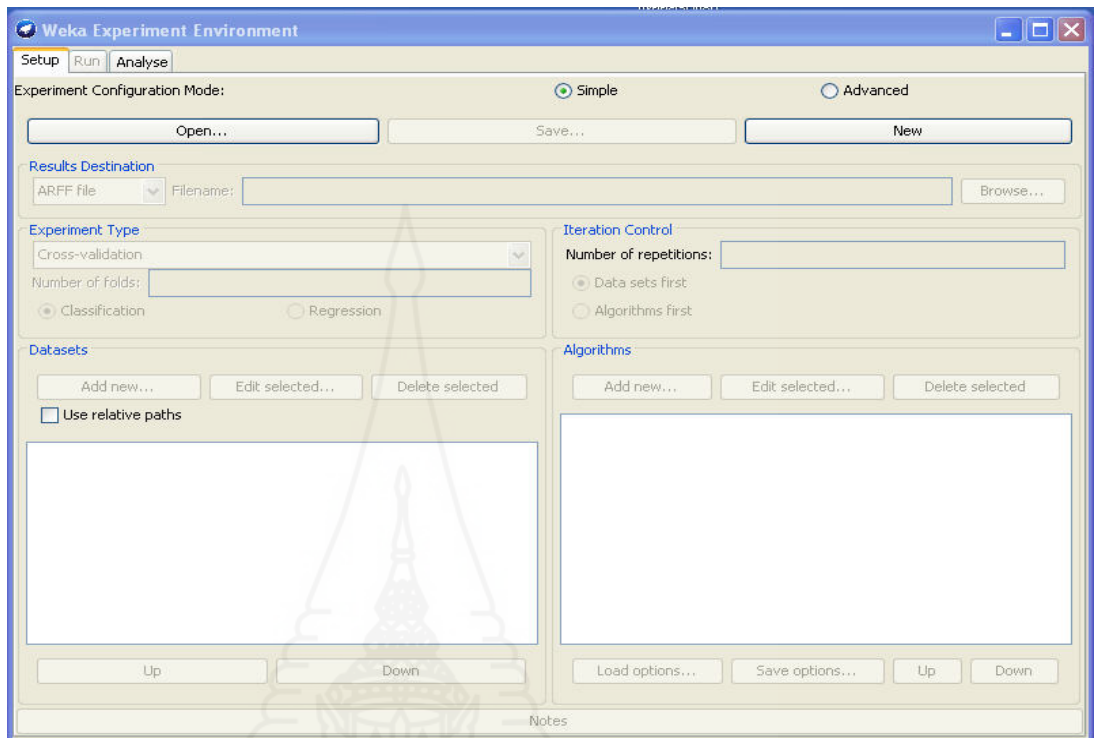
ภาพที่ 10 หน้าจอหลักของโปรแกรมเวก้า

Explorer ในส่วนนี้จะออกแบบให้มีลักษณะเป็น GUI เพื่อให้ง่ายต่อการใช้งาน เหมาะสมสำหรับผู้ไม่ค่อยมีพื้นฐานในการเขียนโปรแกรม



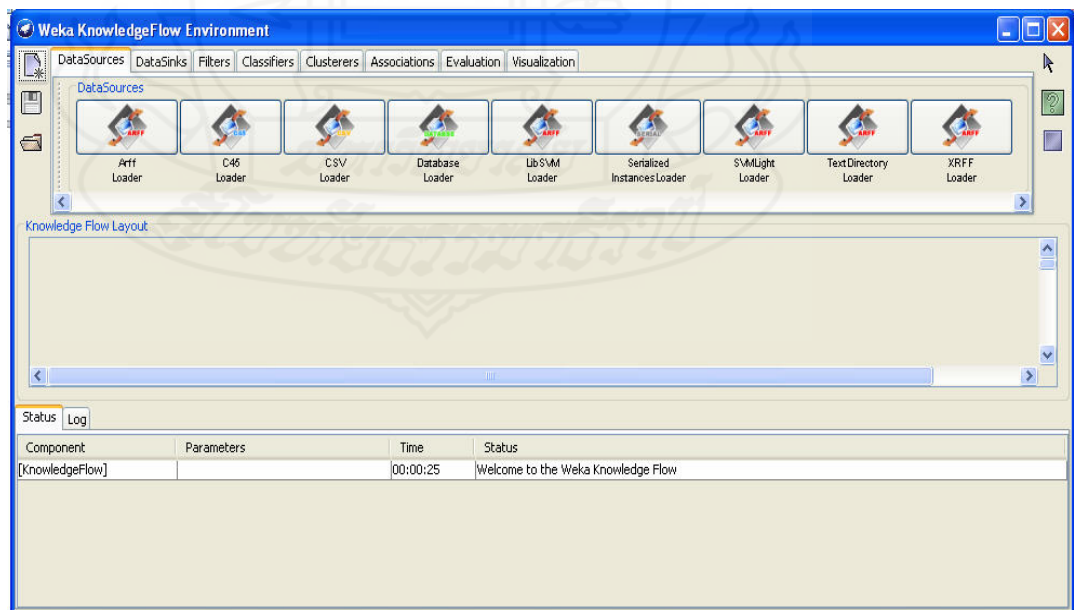
ภาพที่ 11 หน้าจอ Explorer

Experimenter ใช้เป็นส่วนในการออกแบบการทดลอง และทดสอบผลโดยใช้สถิติ



ภาพที่ 12 หน้าจอ Experimenter

KnowledgeFlow เป็นส่วนที่ใช้ในการออกแบบผังในการไหลของความรู้



ภาพที่ 13 หน้าจอ KnowledgeFlow

Simple CLI (Command Line Interface) เป็นส่วนที่ใช้รับคำสั่งผ่านการพิมพ์ที่เหมาะสมสำหรับผู้ที่มีพื้นฐานทางด้านการเขียน โปรแกรม

```

SimpleCLI
Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with './' or '~/ '
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

> help

Command must be one of:
  java <classname> <args> [ > file]
  break
  kill
  cls
  history
  exit
  help <command>

```

ภาพที่ 14 หน้าจอ Simple CLI

1. การใช้โปรแกรม WEKA

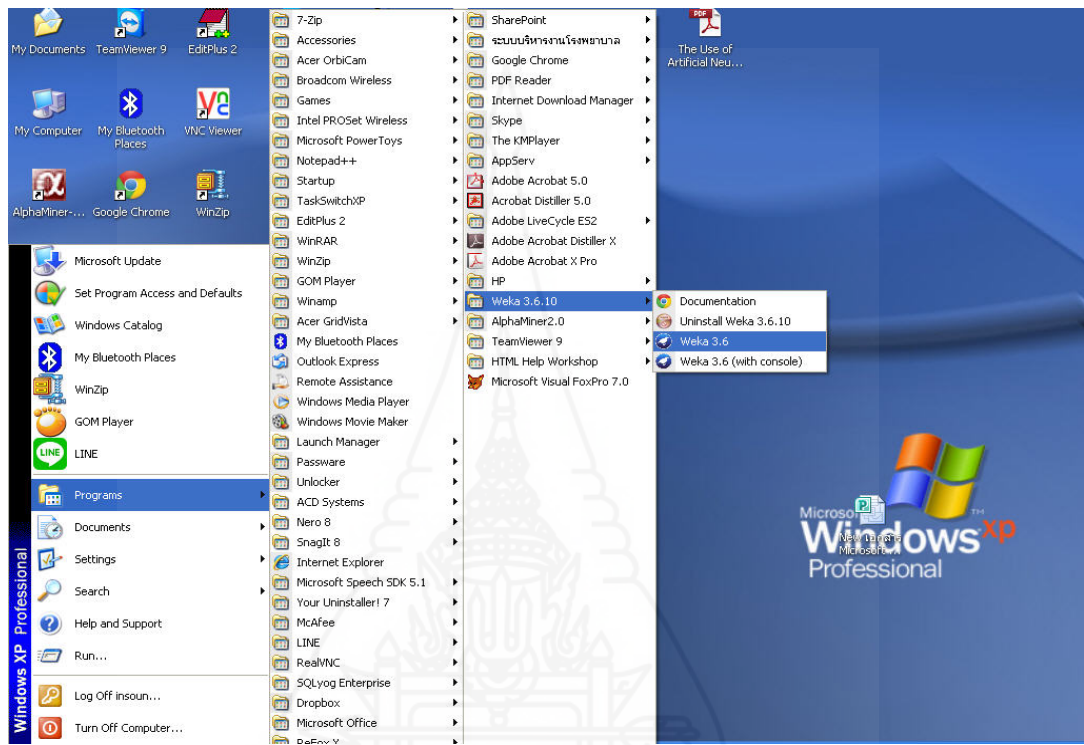
วิเคราะห์ข้อมูลการใช้งานโปรแกรมเวก้านั้นวิธีการจะขึ้นอยู่กับรูปแบบของการสร้างโมเดล แต่โดยหลักๆ แล้วจะคล้ายๆ กัน จะแตกต่างกันเฉพาะในส่วนของการเลือกปพลิเคชันเท่านั้น ในที่นี้ขอยกตัวอย่างการใช้เทคนิค Associate Rule Discovery เพื่อสนับสนุนการตัดสินใจในการจัดการกับธุรกิจ E-Commerce เพื่อใช้ในการจัดโปรโมชั่นการขายสินค้าโดยนำสินค้าที่มีการขายคู่กันสูงมาจับคู่กับสินค้าที่ขายได้ยากแล้วทำการลดราคา อีกทั้งยังเป็นประโยชน์ในการจัดวางสินค้าต่างๆ ที่หน้าเว็บไซต์ ได้นำโปรแกรม WEKA ซึ่งเป็นโปรแกรมที่สำคัญในการหาความสัมพันธ์ของสิ่งต่างๆ ที่ซ่อนอยู่ อีกทั้งยังเป็น โปรแกรมที่ไม่ซับซ้อนและสามารถทำความเข้าใจได้ง่าย ดังนี้

ก่อนที่จะทำการติดตั้งโปรแกรม WEKA ในเครื่องคอมพิวเตอร์จะเป็นต้องมี Java Runtime ก่อน เนื่องจากโปรแกรม WEKA เป็น Software ที่ถูกพัฒนาโดยภาษา Java

1.1 ขั้นตอนการติดตั้งโปรแกรมเวก้า

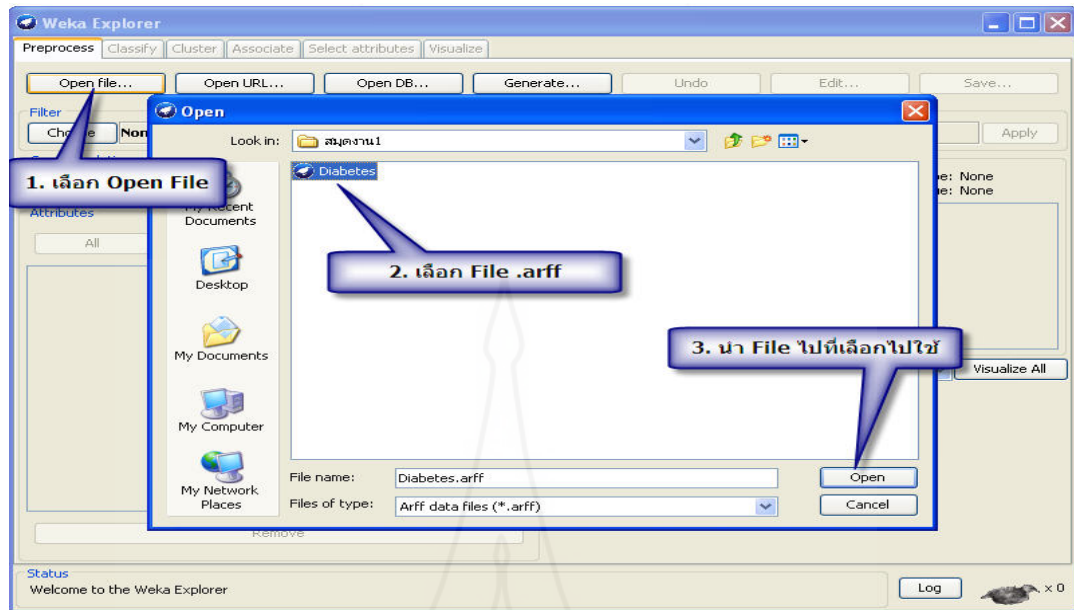
(1) ติดตั้งโปรแกรม WEKA Version 3.5.10

(2) เมื่อติดตั้งโปรแกรมเสร็จเรียบร้อยแล้ว ทำการเปิดโปรแกรกดังภาพ



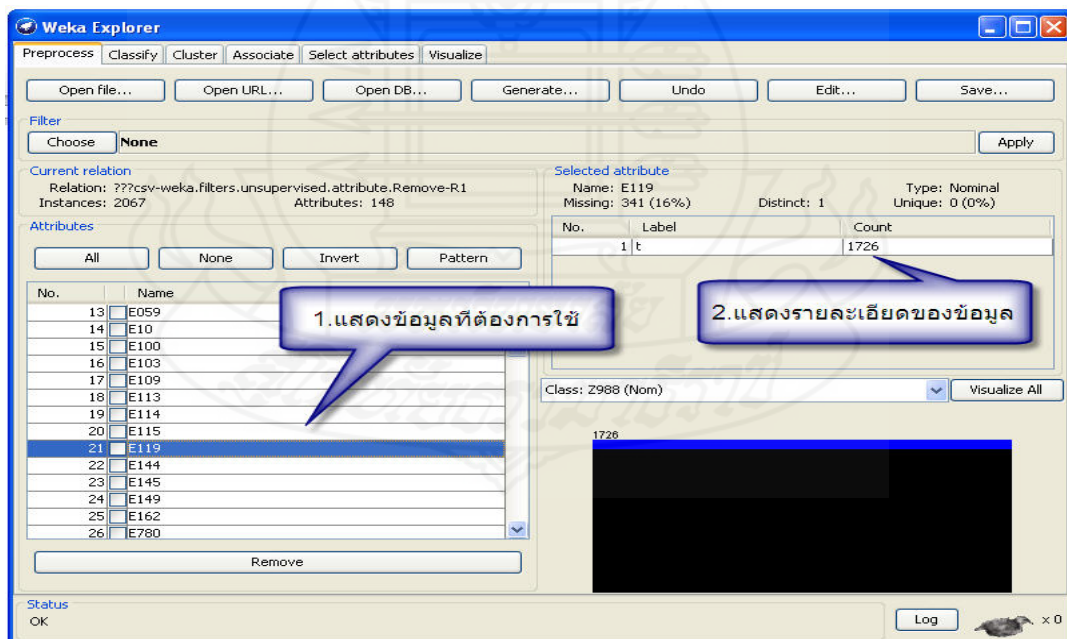
ภาพที่ 15 หน้าจอเรียกใช้โปรแกรม WEKA

(3) เมื่อครันโปรแกรมก็จะได้น้ำจอดังภาพ หลังจากนั้นเลือก Application แล้วเลือก ปุ่ม Explorer เพื่อเปิดหน้าจอใช้งานโปรแกรม เมื่อต้องการเริ่มวิเคราะห์ข้อมูลให้คลิกปุ่ม Open file เพื่อทำการเปิดไฟล์ข้อมูลที่ได้จัดเตรียมไว้ นั้นขึ้นมาใช้งานซึ่งโครงการนี้จะไฟล์ที่มีนามสกุล “arff” ซึ่งต้องเลือกรูปแบบของไฟล์ก่อน เมื่อปรากฏชื่อไฟล์ที่ต้องการเปิดชื่อ Diabetes.arff



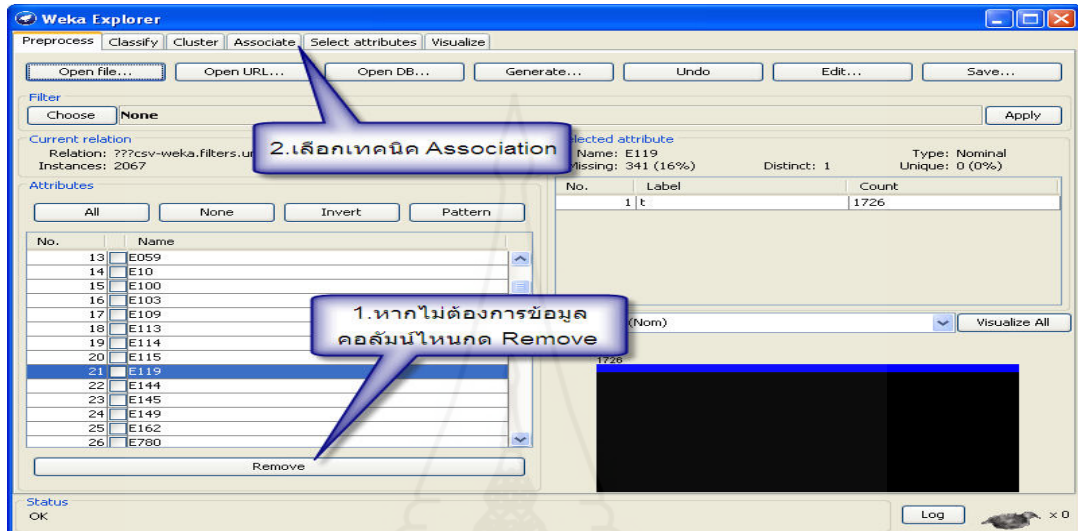
ภาพที่ 16 หน้าจอส่วนประกอบหลักของ Explorer

(4) เมื่อเปิดไฟล์แล้วจะได้หน้าจอที่แสดงคอลัมน์ของไฟล์ปรากฏให้เลือกใช้งาน



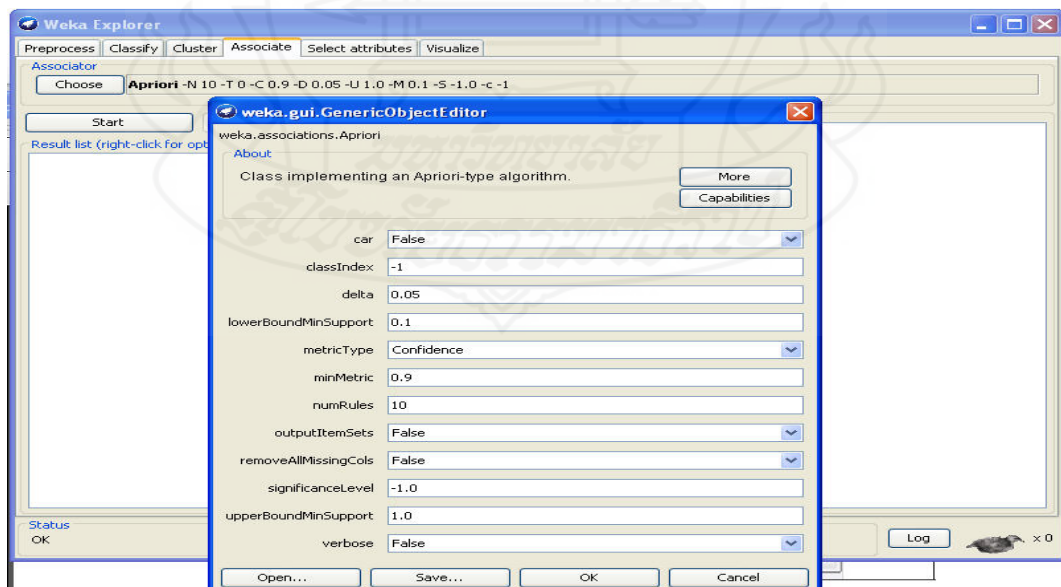
ภาพที่ 17 หน้าต่าง Exploere ที่แสดงรายละเอียดของข้อมูล

(5) นำข้อมูลเข้าไปทำการวิเคราะห์เฉพาะ คอลัมน์ที่สนใจหากไม่ต้องการข้อมูลทำการกดเลือกเพื่อเลือกคอลัมน์ที่ไม่ต้องการวิเคราะห์ แล้วคลิกที่ปุ่ม Remove หลังจากกด Remove คอลัมน์ที่ไม่ต้องการใช้วิเคราะห์แล้วก็ กดเลือก Associate ดังภาพ



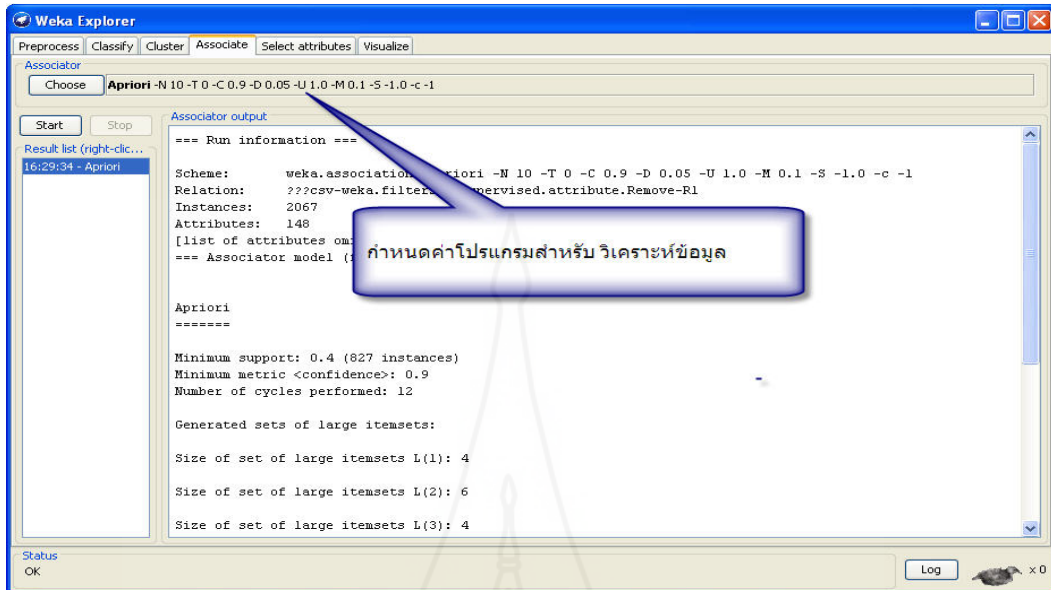
ภาพที่ 18 หน้าต่างแสดงการเลือกรูปแบบการสร้างโมเดล

(6) หลังจากเลือก Associate กดตรง Choose เลือก Apriori แล้วทำการกำหนดค่าโปรแกรมเพื่อวิเคราะห์ข้อมูลหากดูความสัมพันธ์ตามที่ต้องการจากนั้นกดเลือก Start เพื่อเริ่มทำการวิเคราะห์ ดังปรากฏภาพ โดยให้การหาแบบ Confident ปรับค่า LowerBoundMinSupport เป็น 0.1 minMetric เป็น 0.9 NumRules เป็น 10



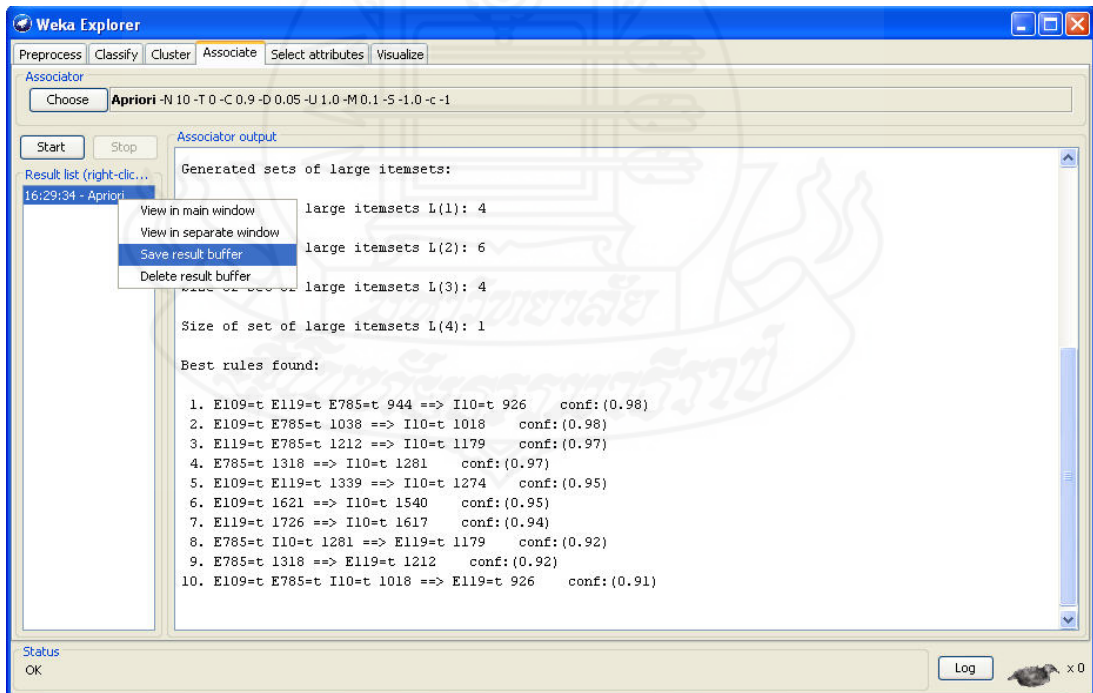
ภาพที่ 19 หน้าต่างแสดงการกำหนดค่าโปรแกรมสำหรับวิเคราะห์ข้อมูล

(7) โปรแกรมก็จะแสดงผลกฎความสัมพันธ์ของข้อมูลตามหน้าจอด้านล่าง



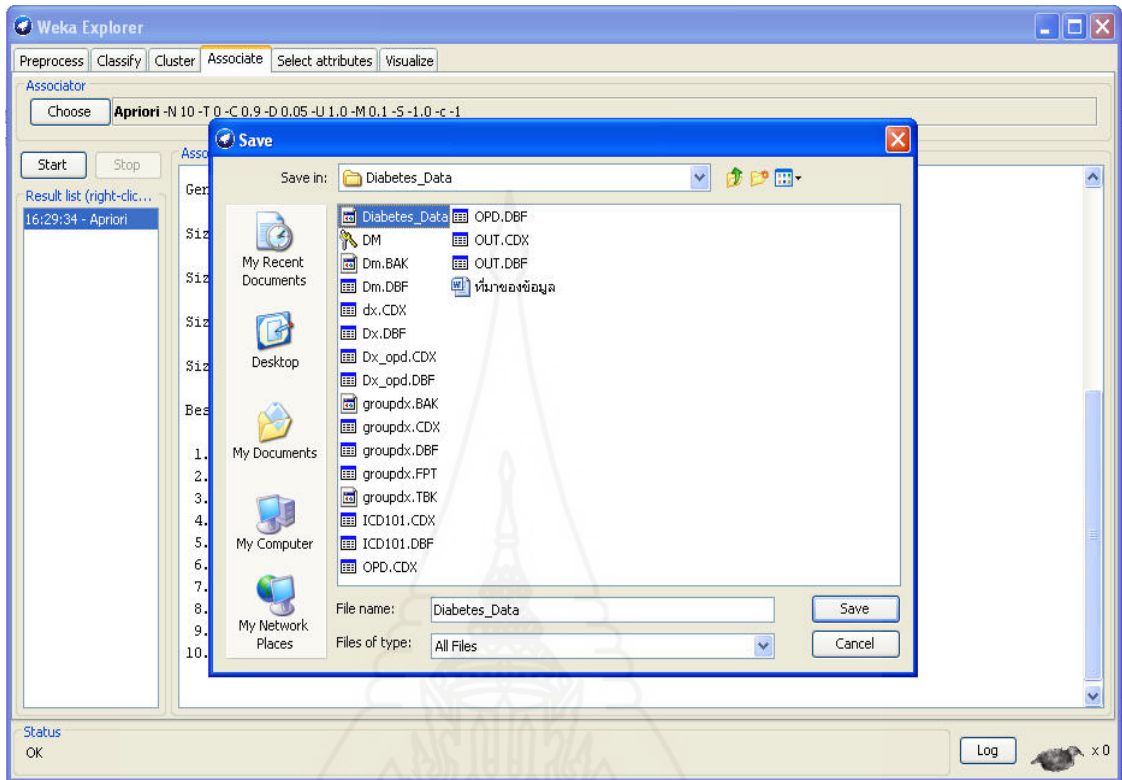
ภาพที่ 20 หน้าต่างแสดงการกำหนดรูปแบบอัลกอริทึมและกำหนดค่า

(9) การบันทึกเป็นรูปแบบ Text File



ภาพที่ 20 หน้าต่างบันทึกเป็นรูปแบบ Text File

(10) การ Save ชื่อ รูปแบบ Text File



ภาพที่ 21 หน้าต่างการ Save ชื่อ รูปแบบ Text File



ภาคผนวก ข

เอกสาร อนุญาตให้เก็บข้อมูลเพื่อการศึกษาวิจัย





บันทึกข้อความ

ส่วนราชการ กลุ่มงานการพยาบาลผู้ป่วยนอก/ศูนย์แปล โทร. ๐-๒๕๕๘-๘๘๐๔
 ที่ ปท ๐๐๓๒.๒๑๐.๒/๒๓๙ วันที่ ๒๐ กรกฎาคม ๒๕๕๗
 เรื่อง เอกสาร รับรอง ทดลองใช้ระบบนำเสนอมผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้อง
 จากการเป็นโรคเบาหวาน

อ้างตามเอกสารเลขที่ ปท ๐๐๓๒.๒๑๓.๓/๙ ลงวันที่ ๕ มีนาคม ๒๕๕๖ การขออนุญาตให้เก็บ
 ข้อมูลเพื่อการศึกษาวิจัย ของ นายชฎิภรณ์ ทราয়หมอ เรื่อง "การพยากรณ์ภาวะแทรกซ้อน ของผู้ป่วย
 โรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล" นั้น ขอรับรองว่าได้ทำการทดลองใช้ ระบบนำเสนอมผลการ
 พยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ตั้งแต่ ๑ เมษายน ๒๕๕๗ ถึง ๑๕
 กรกฎาคม ๒๕๕๗ ที่กลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี

(นางปัทมา วงศ์นคร)

พยาบาลวิชาชีพชำนาญการ

หัวหน้ากลุ่มงานการพยาบาลผู้ป่วยนอก/ศูนย์แปล





บันทึกข้อความ

ส่วนราชการ กลุ่มงานพัฒนาทรัพยากรบุคคล โรงพยาบาลปทุมธานี โทร.๘๗๐๒

ที่ ปท ๐๐๓๒.๒๐๓.๓/๙

วันที่ ๕ มีนาคม ๒๕๕๖

เรื่อง อนุญาตให้เก็บข้อมูลเพื่อการศึกษาวิจัย

เรียน นายชฎิภรณ์ ทราชมอ

ตามที่ ท่าน ได้ขออนุมัติเข้าเก็บข้อมูลเพื่อศึกษาวิจัย เรื่อง “การพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล” นั้น

ในการนี้ คณะกรรมการวิจัย โรงพยาบาลปทุมธานี ได้ประชุมพิจารณาแล้ว และมีมติอนุมัติ และรับรองด้านจริยธรรมการวิจัยในมนุษย์ในโครงการศึกษาดังกล่าว

จึงเรียนมาเพื่อทราบและดำเนินการ ต่อไป

(นายสุรัตน์ สุขประเสริฐ)

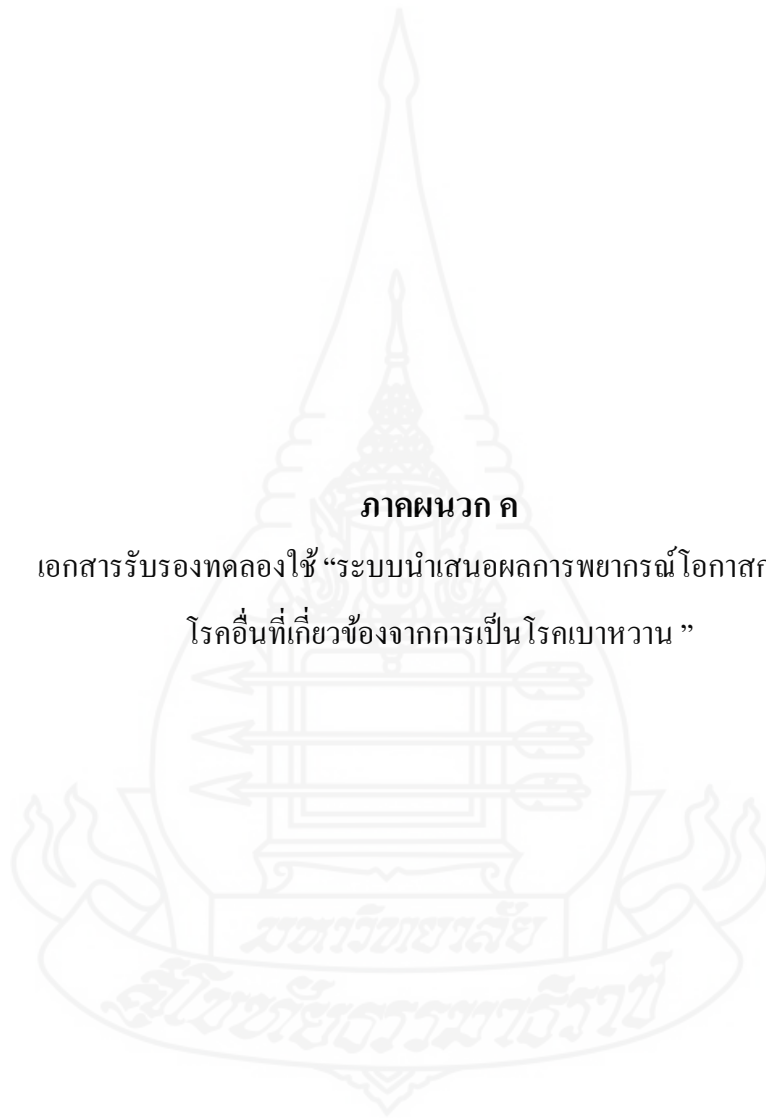
รองผู้อำนวยการฝ่ายการแพทย์

ประธานคณะกรรมการวิจัยโรงพยาบาลปทุมธานี



ภาคผนวก ค

เอกสารรับรองทดลองใช้ “ระบบนำเสนอผลการพยากรณ์โอกาสการเกิด
โรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน ”





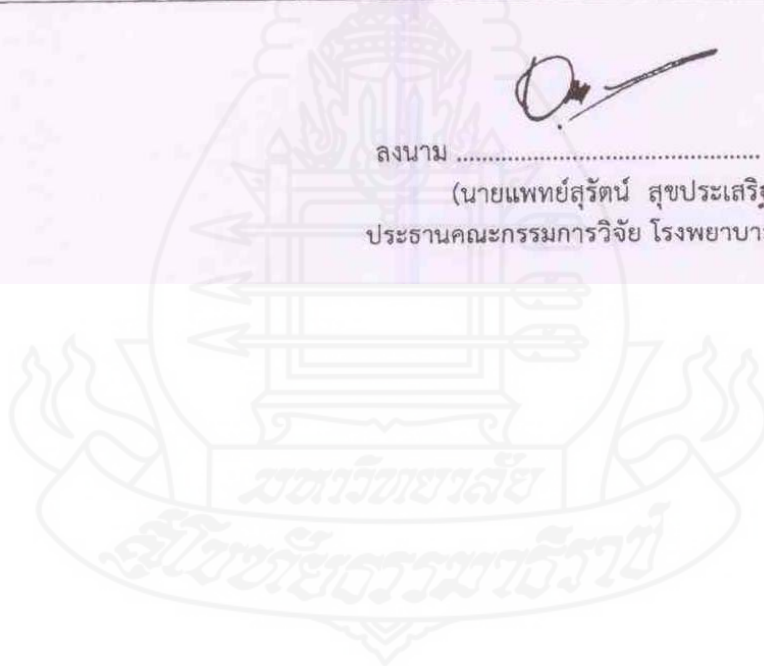
เอกสารรับรองโครงการวิจัย โดย โรงพยาบาลปทุมธานี

ชื่อโครงการวิจัย	“การพยากรณ์ภาวะแทรกซ้อนของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล”
ชื่อผู้วิจัย/หัวหน้าโครงการ	นายชฎิกกรณ์ ทรายหมอ
สถานที่ทำวิจัย	โรงพยาบาลปทุมธานี
การรับรอง	รับรองด้านจริยธรรมการวิจัยในมนุษย์และอนุมัติให้ดำเนินการวิจัยได้
วัน/เดือน/ปี ที่รับรอง	๕ มีนาคม ๒๕๕๖

ลงนาม

(นายแพทย์สุรัตน์ สุขประเสริฐ)

ประธานคณะกรรมการวิจัย โรงพยาบาลปทุมธานี



บรรณานุกรม



บรรณานุกรม

- เกวลี ผาใต้ (2553) “การพยากรณ์โอกาสเกิดภาวะเสี่ยงของผู้ป่วยเบาหวานโดยใช้โครงข่ายประสาทเทียม” การศึกษาอิสระปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
- ขจรศักดิ์ ศรีอ่อน (2552) “การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้องโดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง วิทยานิพนธ์ ปริญญาวิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมไฟฟ้า) สาขาวิศวกรรมไฟฟ้า ภาควิศวกรรมไฟฟ้า มหาวิทยาลัยเกษตรศาสตร์
- ข้อมูลรหัสโรค ICD101 (ICD101: ICD-10-TM International Statistical Classification of Diseases and Related Health Problems 10th Revision, Thai Modification, 2012)
- จริยาพร จำปา (2553) “การพยากรณ์ความต้องการทรัพยากรห้องสมุดโดยวิธีกฎความสัมพันธ์ของห้องสมุดการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิตภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- นงลักษณ์ พลอยปลื้ม (2540) “การแปลงข้อมูลเข้าสู่ Data Warehouse” *BCM Magazine* 8, 103 (กันยายน): 126
- ฝ่ายสารสนเทศ โรงพยาบาลปทุมธานี (2556) รายชื่อผู้ป่วยเบาหวาน ตั้งแต่วันที่ 1 มกราคม 2553 ถึงวันที่ 31 ธันวาคม 2556 จำนวน 17,043 คน
- ฟูไวดะห์ คีอมอง (2552) “AMFIST: ขั้นตอนวิธีสำหรับการค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย โดยรองรับรายการข้อมูลที่คล้ายคลึงกัน” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์
- วิชัย เอกพลากร. การศึกษาพัฒนาดัชนีความเสี่ยงต่อเบาหวาน กรุงเทพมหานคร ศูนย์เวชศาสตร์ชุมชน คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี
- วิภา เจริญกัณฑ์ธารักษ์ (2555) “หลักการพื้นฐานของการทำเหมืองข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ = Data warehouse data mining and business intelligence* 8, 1-33 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมมาธิราช สาขาวิชาวิทยาศาสตร์และเทคโนโลยี

สมพงษ์ สุวรรณวัลย์กร ความรู้พื้นฐานเกี่ยวกับโรคเบาหวานและโรคแทรกซ้อน กรุงเทพมหานคร:
สาขาวิชาต่อมไร้ท่อ และเมตะบอลิซึม ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

สุวรรณณี อัสวกุลชัย (2555) “หลักการพื้นฐานของคลังข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล
เหมืองข้อมูล และธุรกิจอัจฉริยะ = Data warehouse data mining and business
intelligence* หน่วยที่ 1 หน้า1-30 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมมาธิราช
บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี

อังคณา พิจารโชติ (2552) “ระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น
โรคเบาหวาน โดยใช้เทคนิคคาน่าไมน์นึ่ง” การศึกษาอิสระปริญญาวิทยาศาสตร
มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

Berberidis, G. Tzanis, and L. Vlahavas. “Mining for Contiguous Frequent Itemsets in
Transaction databases”. In *Proceedings of IEEE Workshop on Intelligent Data
Acquisition and Advanced Computing systems: Technology and Applications*,
Sofia, Bulgaria, pp. 679-685, 2005.

Charu C. Aggarwal, Jiawei Han. “Frequent Pattern Mining Pattern-Growth Methods”. In
Proceedings of Amazon Frequent Pattern Mining, pp. 65-79, 2014.

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.

Connolly, Thomas M. and Begg, Carolyn E. (2002). *Database Systems: A Practical Approach to
Design, Implementation, and Management 3rd.* MA : Addison-Wesley.

Cunningham, S.J. & Frank, E. (1999). Market basket analysis of library circulation data.
Proceedings of the sixth International Conference on Neural Information Processing
1999, Volume II, Perth, Western Australia.

<http://cheekswellilive.wordpress.com> July 26, 2014.

<http://dssresources.com/papers/features/brobst&rarey01062003.html> Retrieved July 25, 2014

[http://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_](http://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_Related_Health_Problems)
[Related_Health_Problems](http://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_Related_Health_Problems) Retrieved July 24, 2014.

[http://en.wikipedia.org/wiki/Weka_\(machine_learning\)#mediaviewer/File:Weka_\(software\)_logo](http://en.wikipedia.org/wiki/Weka_(machine_learning)#mediaviewer/File:Weka_(software)_logo.png)
[png](http://en.wikipedia.org/wiki/Weka_(machine_learning)#mediaviewer/File:Weka_(software)_logo.png) July 26, 2014.

- <http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> Retrieved June 30 , 2014.
- <http://gotoknow.org/posts/452189> Retrieved July 20, 2014.
- <http://siripornk.blogspot.com/2010/08/data-mining.html> Retrieved July 24, 2014.
- <http://www.crisp-dm.org/Process/index.htm> Retrieved June 30 , 2014.
- <http://www.vcharkarn.com/varticle/36790#chapter1> July 25, 2014.
- <http://nuiza.files.wordpress.com/2012/03/d2.jpg> July 26, 2014¹
- Jiawei, Han and Micheline, Kamber. (2006). *Data Mining: Concepts and Techniques*. Second Edition. San Francisco, CA: Morgan Kaufmann Publishers is an imprint of Elsevier.
- Larose, D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Indianapolis: Wiley.
- Nicholson, S. and Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. Hershey PA: Idea Group Publishing.
- R. Agrawal, T. Imielinski, and A. Sawami. "Mining Association Rules between Sets of Items in Large Databases". In Proceeding of ACM SIGMOD, pp. 207-216, 1993.
- R. Agrawal, R. Srikant. Fast Algorithms for Mining Association Rules (1994) Proc. 20th Int. Conf. Very Large Data Bases, VLDB.
- Roiger and Geatz. (2003). *Data mining – A tutorial-based primer*. Pearson International Edition, USA: Addison-Wesley.
- Rygielski, J. Wang, and D. Yen. "Data mining techniques for customer relationship management". *Published in Technology in Society* 24, pp. 483-502, 2002.

ประวัติผู้วิจัย

ชื่อ	นายชฎิภรณ์ ทรายหมอ
วัน เดือน ปี เกิด	16 ธันวาคม 2522
สถานที่เกิด	อำเภอแม่จัน จังหวัดเชียงราย
ประวัติการศึกษา	ร.บ. มหาวิทยาลัยรามคำแหง เมื่อปีการศึกษา 2548
สถานที่ทำงาน	ศูนย์สารสนเทศ โรงพยาบาลปทุมธานี จังหวัดปทุมธานี อำเภอเมืองปทุมธานี จังหวัดปทุมธานี
ตำแหน่ง	ที่ปรึกษาระบบโปรแกรม

