

กระบวนการดัดไม้หนึ่งเพื่อตรวจสอบพฤติกรรมการไปใช้บริการไม่สมเหตุสมผล
ของผู้ป่วยนอก

ประไพพร อุตมา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมการด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2563

**The Data Mining Process to Investigate Non-Reasonable Behavior of Medical
Service Utilization in Hospitals**

Praprai Utama

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering
Department of Big Data Engineering,
College of Innovative Technology and Engineering,
Dhurakij Pundit University**

2020

หัวข้อวิทยานิพนธ์	กระบวนการค้นหาไมน์นิ่งเพื่อตรวจสอบพฤติกรรมการไปใช้ บริการไม่สมเหตุสมผลของผู้ป่วยนอก
ชื่อผู้เขียน	ประไพพร อุดมา
อาจารย์ที่ปรึกษา	ดร.เอกสิทธิ์ พัทธวงษ์ศักดิ์
สาขาวิชา	วิศวกรรมข้อมูลขนาดใหญ่
ปีการศึกษา	2562

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการใหม่ในการช่วยคัดกรองผู้ป่วยนอกที่มีพฤติกรรมผิดปกติของการไปใช้บริการสถานพยาบาลไม่สมเหตุสมผล โดยการประยุกต์ใช้เทคนิคการแบ่งกลุ่ม (RFMHPE Segmentation) และเทคนิคค้นหาไมน์นิ่ง (Data Mining) นำมาสร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) เพื่อจำแนกกลุ่มพฤติกรรมผิดปกติและพฤติกรรมปกติ ปัจจุบันการตรวจสอบข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลของผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ ทำได้โดยสถานพยาบาลแจ้งพฤติกรรมบุคคลที่ผิดปกติให้กรมบัญชีกลางตรวจสอบ จากการอ่านรายการเวชระเบียนโดยผู้เชี่ยวชาญ ซึ่งการตรวจสอบข้อมูลค่อนข้างยากและใช้เวลานาน

การศึกษานี้จึงเกิดขึ้นซึ่งมีวัตถุประสงค์เพื่อพัฒนารูปแบบการตรวจสอบโดยใช้เทคนิคค้นหาไมน์นิ่ง (Data Mining) ตรวจสอบวิเคราะห์ข้อมูลที่ไม่มีลาเบล (Unsupervised Learning) จากฐานข้อมูลผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ ระบบเบิกจ่ายตรงปีงบประมาณ 2556 จำนวน 24 ล้าน records ประกอบด้วย 1) ข้อมูลการเบิกจ่ายเงินรายบุคคล (OPDATA) 2) ข้อมูลการเบิกจ่ายเงินรายหมวด (OPBILL) 3) ข้อมูลใบสั่งยา (OPDISP) เลือกข้อมูลที่มีการเบิกจ่ายเงินค่ายานำกลับไปใช้ที่บ้าน (หมวด 4) มีค่ามากกว่า 30,000 บาท/ปี คัดผู้ป่วยนอกที่ใช้ยามะเร็งออกเหลือจำนวนทั้งหมด 31,001 คน ใช้ตัวแปร 6 ตัว R, F, M, H, P, E แบ่งกลุ่มข้อมูลด้วยเทคนิค RFMHPE และเทคนิคเคมีน (K-means)

ผลจากการแบ่งกลุ่มเทคนิค RFMHPE แบ่งได้ 413 กลุ่ม พบ 48 รูปแบบ ส่วนการแบ่งกลุ่มด้วยเทคนิค K-means แบ่งได้ 39 กลุ่ม พบ 18 รูปแบบ เลือกรูปแบบที่มีค่า F, H, E จากทั้ง 2

เทคนิค แล้วนำมาอินเตอร์เซ็กกันเพื่อกำหนดลาเบลประเภท (Type) A (กลุ่มที่น่าสนใจเป็นพิเศษ) ประเภท B (กลุ่มที่น่าสนใจรองลงมา) ประเภท C (กลุ่มพฤติกรรมปกติ) และทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ได้ค่าความแม่นยำ (accuracy) จากข้อมูลใช้ทดสอบ (test data) 24 คน เท่ากับ 91.67% และจากข้อมูลใช้ทดสอบ 74 คน ได้ความแม่นยำ เท่ากับ 97.30 % ทำการเปรียบเทียบความแตกต่างระหว่าง 2 กลุ่มด้วยค่าสถิติ t-test กลุ่มพฤติกรรมผิดปกติ 24 คน และกลุ่มพฤติกรรมปกติ 50 คน พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญในระดับสูงมาก (p-value= 0.000) และตัวแปรที่มีอำนาจในการทำนายแบบจำลองคือ F, M, H, P สามารถบอกลักษณะพฤติกรรมของของกลุ่มตัวอย่างได้ดี

อย่างไรก็ตามจากจำนวนผู้ป่วยนอกทั้งหมด 2.7 ล้านคน เบิกจ่ายเงินค่ายากลับไปใช้ต่อที่บ้าน จำนวน 36,652 ล้านบาท ผู้ใช้บริการที่มีการเบิกจ่ายเงินค่ายากลับไปใช้ต่อที่บ้านเกิน 30,000 บาท/ปี จำนวน 31,001 คน ร้อยละ 1.13 ของผู้ไปใช้บริการทั้งหมด เบิกจ่ายเงินค่ายากลับไปใช้ต่อที่บ้านร้อยละ 6.97 โดยแบ่งเป็นกลุ่มที่น่าสนใจเป็นพิเศษ ร้อยละ 0.8 กลุ่มที่น่าสนใจรองลงมา ร้อยละ 0.15 และกลุ่มพฤติกรรมปกติ ร้อยละ 0.18 พบว่าพฤติกรรมผิดปกติซึ่งเป็นเจ้าของสิทธิที่ไปใช้บริการมากในช่วงอายุ 50-59 ปี ดังนั้นการนำเทคนิคดาต้าไมน์นิ่งมาใช้สามารถช่วยค้นหากลุ่มคนที่ผิดปกติได้อย่างรวดเร็วและประหยัดเวลา ช่วยให้องค์กรเฝ้าระวังบุคคลที่มีพฤติกรรมผิดปกติในลักษณะเดียวกันได้

Thesis Title	The Data Mining Process to Investigate Non-Reasonable Behavior of Medical Service Utilization in Hospitals
Author	Praprai Utama
Thesis Advisor	Dr. Eakasit Pacharawongsakda
Department	Big Data Engineering
Academic Year	2019

ABSTRACT

The objective of this study is to propose the alternative aspects for screening of outpatients having non-reasonable medical service utilization behavior in hospitals by using RFMHPE Segmentation and Data Mining. Decision Tree is established to separate medical service utilization behavior between the normal groups and non-reasonable groups. At the present, the medical reimbursement in the Civil Servant Medical Benefit Scheme (CSMBS) system in Thailand is usually investigated from medical records by expert consideration. In other words, the hospitals have reported the Comptroller General's Department, Ministry of Finance the medical records with reporting of non-reasonable medical service utilization behavior. However, this method usually spend a lot of time and it is very complicated for investigations.

In order to development of this appraisal, data mining is manipulated for unsupervised learning investigations from the 24 million records of outpatient database in the CSMBS's Direct Billing System of fiscal year 2013. Including with 1) Outpatient medical reimbursement for individual data (OPDATA). 2) Public health service rate category. 3) Prescribing data. Selecting cases that having the medical expense reimbursement for discharge that is higher than 30,000 Thai Baht (THB) per year without outpatient taking cancer treatment drugs. This found 31,001 persons and were divided into different groups by six variables particularly R, F, M, H, P, and E with RFMHPE, and K-means.

The results of RFMHPE can be divided into 413 groups with 48 patterns. In case of K-means can be split into 39 groups with 18 patterns. In these results, selecting the patterns with having F, H, and E from both technics and intersect to set some labels including with type A (special interesting groups; outstandingly non-reasonable medical service utilization behavior

groups), type B (the second most interesting groups; moderately non-reasonable medical service utilization behavior groups), and type C (reasonable medical service utilization behavior groups)

In terms of decision trees, testing data from 24 persons can be presented the accuracy value as 91.67 percentages. Moreover, 74 persons of testing data can be illustrated the accuracy value as 97.30 percentages. The comparison between non-reasonable medical service utilization behavior groups with 24 persons and reasonable medical service utilization behavior groups with 50 persons by using statistics with independent sample T-test. The results of this study showed the significant differences ($p\text{-value} = 0.000$). The variables such as F, M, H, and P, all of them are the substantial variables that can predict model and evidently describe the behavioral characteristics of samples.

Additionally, the amount of 2.7 million patient's reimbursement as THB 36.652 million and there are 31,001 persons (1.13 %) who have amount of reimbursement is higher than THB 30,000 per year (6.97 %). Furthermore, 0.8 % of type A, 0.15 % of type B, and 0.18% of type C showed that the most of users is owner of medical welfare rights aged 50-59. The non-reasonable medical service utilization behavior groups will be expeditiously observed by implementing data mining. This technic can be an aspect to save time and helping organizations monitor group of persons with non-reasonable behavior of medical service utilization in hospitals.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยความกรุณาเป็นอย่างยิ่งจาก ดร.เอกสิทธิ์ พัทธวงค์ ศักดา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยให้คำแนะนำ ตลอดจนเปิดโลกทรรศน์ในการค้นคว้า ข้อมูลให้แก่ผู้วิจัย ขอขอบพระคุณ ศ.ดร. ธนารักษ์ ธีระมันคง ดร. สรรพฤทธิ์ มฤคทัต ดร. ธนภัทร นังคะจิตร และผศ.ดร. ดวงใจ จิตคงชื่น กรรมการสอบวิทยานิพนธ์ ซึ่งสละเวลามาเป็นกรรมการ สอบวิทยานิพนธ์และได้ให้ข้อคิดเห็นที่เป็นประโยชน์ต่อการศึกษาครั้งนี้ นอกจากนี้ผู้วิจัย ขอขอบพระคุณ ผศ.ดร. วรพล พงษ์เพ็ชร ผู้ก่อตั้งหลักสูตร Big data engineering และ ขอขอบพระคุณคณาจารย์ทุกท่านในคณะวิศวกรรมศาสตร์ ที่ได้ถ่ายทอดความรู้แก่ผู้วิจัยตลอด ระยะเวลาการศึกษา

ผู้วิจัยขอขอบพระคุณน้องแดง และเจ้าหน้าที่ที่เกี่ยวข้องทุกคน ในคณะวิศวกรรมศาสตร์ ที่คอยให้ความช่วยเหลือตลอดจนคำแนะนำต่างๆ ให้แก่ผู้วิจัยด้วยดีเสมอมา

ผู้วิจัยขอขอบพระคุณเพื่อนๆ ร่วมรุ่นพี่น้อง ที่ช่วยเหลือและให้กำลังใจกันเสมอมา

ผู้วิจัยขอขอบพระคุณ นพ.ถาวร สกุลพานิช หัวหน้างานที่เคาร์พ่นบถือ นายภาสกร สวน เรือง นางสาวอรารรณ ประสิทธิ์ผล เพื่อนร่วมงานที่น่ารัก ที่ช่วยเหลือและให้คำแนะนำที่ดีเสมอมา

สุดท้ายนี้ ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ครอบครัวและนายคณิน สาระสุข ที่คอย ให้กำลังใจและให้การสนับสนุนผู้วิจัยในทุกๆ ด้านเสมอมาจนสำเร็จการศึกษา

ประไพ อดุมมา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	๗
บทคัดย่อภาษาอังกฤษ.....	๗
กิตติกรรมประกาศ.....	๗
สารบัญตาราง.....	๗
สารบัญภาพ.....	๗
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	7
1.3 ประโยชน์ที่คาดว่าจะได้รับ.....	7
1.4 ขอบเขตการวิจัย.....	7
1.5 ระเบียบวิธีวิจัย.....	8
1.6 นิยามคำศัพท์.....	8
2. ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง.....	10
2.1 ขั้นตอนการสมัครเข้าร่วมโครงการเบิกจ่ายตรงผู้ป่วยนอก.....	10
2.2 ระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลาง.....	12
2.3 ทฤษฎีการทำเหมืองข้อมูล.....	13
2.4 วัตถุประสงค์ของการทำเหมืองข้อมูล.....	13
2.5 เป้าหมายหลักของการทำเหมืองข้อมูล.....	13
2.6 เทคนิคการวิเคราะห์เหมืองข้อมูล.....	14
2.7 การใช้อัลกอริทึมแมชชีน เลิร์นนิง (Machine Learning) กับข้อมูลสุขภาพ (Healthcare).....	14
2.8 รูปแบบการละเมิด (Abuse).....	17
2.9 รูปแบบการหลอกลวง (Fraud).....	17
2.10 การเก็บรวบรวมข้อมูล การสุ่มตัวอย่าง และการประมวลผล.....	18
2.11 การแบ่งกลุ่มด้วยวิธีเคมีน (K-means).....	23

สารบัญ (ต่อ)

บทที่	หน้า
2.12 การสร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจเพื่อทำนายการหลอกหลวง.....	25
2.13 บทความและงานวิจัยที่เกี่ยวข้อง.....	28
3. ระเบียบวิธี.....	31
3.1 ประชากรและตัวอย่าง.....	31
3.2 เครื่องมือที่ใช้ในการรวบรวมข้อมูล.....	32
3.3 ขั้นตอนดาต้าไมน์นิ่ง Framework.....	32
4. ผลการศึกษา.....	41
4.1 การจำแนกพฤติกรรมด้วยอัลกอริทึมการทำเหมืองข้อมูล.....	41
4.2 อัลกอริทึมการทำเหมืองข้อมูล.....	41
4.3 ข้อดีของการนำอัลกอริทึมการทำเหมืองข้อมูลมาใช้ในการตรวจสอบ พฤติกรรม.....	41
4.4 การหาความสัมพันธ์ระหว่างตัวแปร (Correlation).....	42
4.5 การแบ่งกลุ่มข้อมูล.....	44
4.6 การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการด้วยวิธีเซกเมนต์เทชั่น (Segmentation).....	45
4.7 ผลการวิเคราะห์การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการด้วยวิธี เซกเมนต์เทชั่น.....	45
4.8 รูปแบบที่พบวิธีเซกเมนต์เทชั่น.....	46
4.9 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน.....	48
4.10 การหาค่า Elbow point เพื่อหาค่า K ที่เหมาะสมที่สุด.....	48
4.11 รูปแบบที่พบจากการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน.....	50
4.12 การอินเตอร์เซกข้อมูล.....	52
4.13 การกำหนดลาเบล.....	53
4.14 การทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ.....	57
4.15 ผลการทำนายกลุ่มตัวอย่าง.....	64
4.16 การทดสอบค่าสถิติเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม.....	68

สารบัญ (ต่อ)

บทที่	หน้า
4.15 ผลการทำนายกลุ่มตัวอย่าง.....	64
4.16 การทดสอบค่าสถิติเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม.....	68
4.17 การทดสอบทางสถิติเพื่อเปรียบเทียบกลุ่มตัวอย่างด้วย T-Test.....	69
5. สรุปผลการศึกษา.....	72
5.1 สรุปผล.....	72
5.2 ข้อเสนอแนะ.....	75
บรรณานุกรม	77
ภาคผนวก.....	79
ก	80
ข	99
ประวัติผู้เขียน.....	115

สารบัญตาราง

ตารางที่	หน้า
2.1 อัลกอริทึมแมชชีน เลินนิ่ง กับข้อมูลสุขภาพ.....	15
3.1 ตัวแปรที่ใช้แบ่งกลุ่ม.....	37
4.1 ตัวแปรที่ใช้แบ่งกลุ่มพฤติกรรมผู้มารับบริการ.....	42
4.2 รูปแบบพฤติกรรมผู้มารับบริการวิธีเชกเมนต์เทชั่น.....	46
4.3 แสดงรูปแบบการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน.....	49
4.4 รูปแบบพฤติกรรมผู้ไปใช้บริการจากวิธีเคมีน.....	51
4.5 จำนวนคนหลังการอินเตอร์เซ็ก.....	55
4.6 รูปแบบที่พบแยกสายประเภท.....	55
4.7 สัดส่วนรูปแบบที่พบแต่ละเทคนิค.....	56
4.8 รูปแบบที่พบในแต่ละประเภท.....	56
4.9 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 24 คน.....	65
4.10 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 74 คน.....	66
4.11 แสดงค่าสถิติจากผลการทำนายกลุ่มตัวอย่าง.....	69
4.12 การเปรียบเทียบความแตกต่างระหว่างกลุ่มตัวอย่าง 2 กลุ่ม ด้วยสถิติ t-test.....	71
5.1 สรุปผลการศึกษา.....	73
5.2 แสดงการแบ่งกลุ่มรายช่วงอายุรายประเภท (คน).....	74
5.3 แสดงรายกลุ่มประเภทรายสิทธิการรักษา.....	75

สารบัญภาพ

ภาพที่	หน้า
1.1 ค่ารักษาพยาบาลผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ ปีงบประมาณ 2556-2560	1
1.2 จำนวนครั้งที่ไปใช้บริการผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ.....	2
1.3 จำนวนการเบิกจ่ายเงินค่ารักษาพยาบาลทั้ง 16 หมวด.....	3
1.4 จำนวนการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้านของผู้ป่วยนอกชายช่วงอายุ...	4
1.5 จำนวนการเบิกจ่ายเงินค่ายากลับบ้านและความถี่ที่ใช้บริการสถานพยาบาล....	5
1.6 จำนวนครั้งที่ไปใช้บริการหลายสถานพยาบาล (แคนตั้ง) จำนวนสถานพยาบาล ที่ใช้บริการ (แคนนอน).....	5
1.7 ร้อยละการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้าน ของผู้ป่วยนอกสิทธิ สวัสดิการข้าราชการจำแนกเป็น 10 กลุ่ม.....	6
1.8 การกระจายข้อมูลการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้าน.....	7
2.1 ขั้นตอนการสมัครเข้าร่วมโครงการเบิกจ่ายตรง ณ สถานพยาบาล.....	11
2.2 ระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลาง..	12
2.3 แสดงเครื่องมือที่ใช้วิเคราะห์แก้ปัญหาข้อมูลขนาดใหญ่.....	16
2.4 ขั้นตอนการวิเคราะห์แบบจำลองการหลอกลวง.....	18
2.5 แสดงการแบ่งกลุ่มพฤติกรรมลูกค้าวิธีเชกเมนต์เทชั่น.....	20
2.6 กราฟแสดงกลุ่มพฤติกรรมลูกค้าด้วยวิธีเชกเมนต์เทชั่น.....	21
2.7 แสดงการแบ่งกลุ่มพฤติกรรมด้วยแบบจำลอง LRFMP.....	22
2.8 การจัดกลุ่มด้วยวิธีเคมีน.....	24
2.9 การแบ่งข้อมูลใช้ทดลองและข้อมูลใช้ทดสอบ.....	25
2.10 แบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ.....	26
2.11 ผลการทำนายของแบบจำลอง.....	27
2.12 การวัดประสิทธิภาพ Cross-Validation.....	28
3.1 ขั้นตอนดาต้าไมนนิ่ง Framework.....	33
3.2 โครงสร้างข้อมูลผู้ป่วยนอก.....	34
3.3 การเชื่อมข้อมูลผู้ป่วยนอกระบบเบิกจ่ายตรง.....	35
3.4 ขั้นตอนการแบ่งฐานข้อมูลผู้ป่วยนอก.....	36

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
3.5 ขั้นตอนการจัดการฐานข้อมูล.....	36
3.6 การแบ่งข้อมูลออกเป็น 2 วิธี.....	38
3.7 แสดงค่าที่ดีที่สุดของการแบ่งกลุ่มเคมีนเท่ากับ 39 กลุ่ม.....	39
4.1 แสดงค่าความสัมพันธ์ระหว่างตัวแปร.....	43
4.2 ขั้นตอนการแบ่งกลุ่ม RFMHPE, K-means.....	44
4.3 การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการ.....	45
4.4 แสดงรูปแบบที่แบ่งด้วยวิธี RFMHPE.....	47
4.5 กราฟแสดงจุด K ที่เหมาะสมจาก Elbow point.....	48
4.6 การทำนอร์มอลไลซ์ (NORMALLIZATION) ข้อมูล.....	49
4.7 แสดงรูปแบบที่แบ่งด้วยวิธีเคมีน.....	52
4.8 ขั้นตอนการอินเตอร์เซ็ก.....	53
4.9 การกำหนดคลาเบล.....	54
4.10 การทดสอบแบบจำลอง.....	57
4.11 ขั้นตอนการทดสอบแบบจำลอง.....	58
4.12 การทดสอบแบบจำลองด้วยกลุ่มตัวอย่าง 24 คน.....	58
4.13 แสดงขั้นตอนการทำงานบนโปรแกรม Rapid Miner Studio.....	59
4.14 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดลอง 30,997 คน.....	59
4.15 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดสอบ 24 คน.....	60
4.16 แสดงการแบ่งลำดับชั้นของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ 24 คน.....	60
4.17 แสดงการแตกกิ่งของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจกลุ่มตัวอย่าง 24 คน.....	61
4.18 ขั้นตอนการทดสอบแบบจำลองกลุ่มตัวอย่าง 74 คน.....	62
4.19 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดลอง 30,947 คน.....	62
4.20 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดสอบ 74 คน.....	63
4.21 แสดงการแบ่งลำดับชั้นของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ 74 คน.....	63

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.22 แสดงการแตกกิ่งของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจกลุ่มตัวอย่าง 74 คน	64
4.23 ตัวอย่างการทดสอบค่า F ด้วยสถิติ t-test.....	70



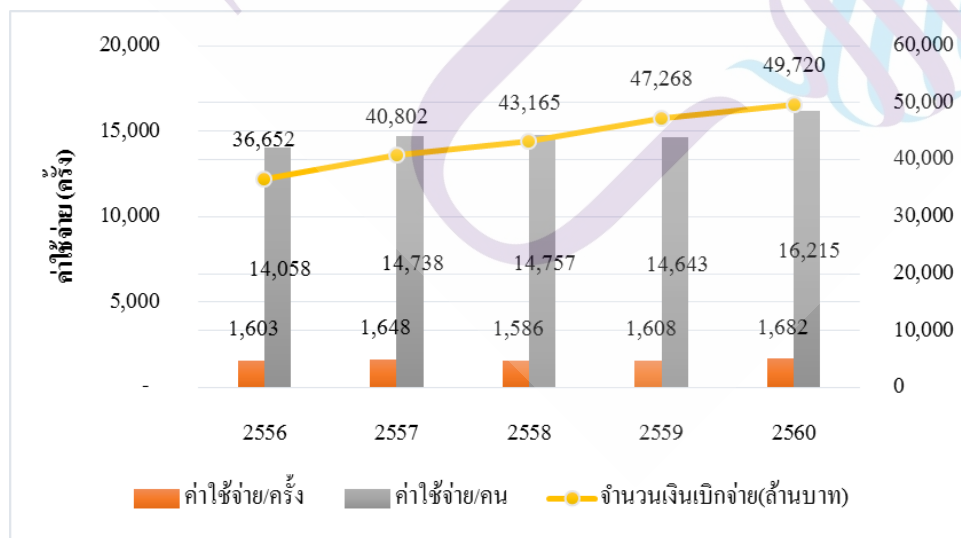
บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

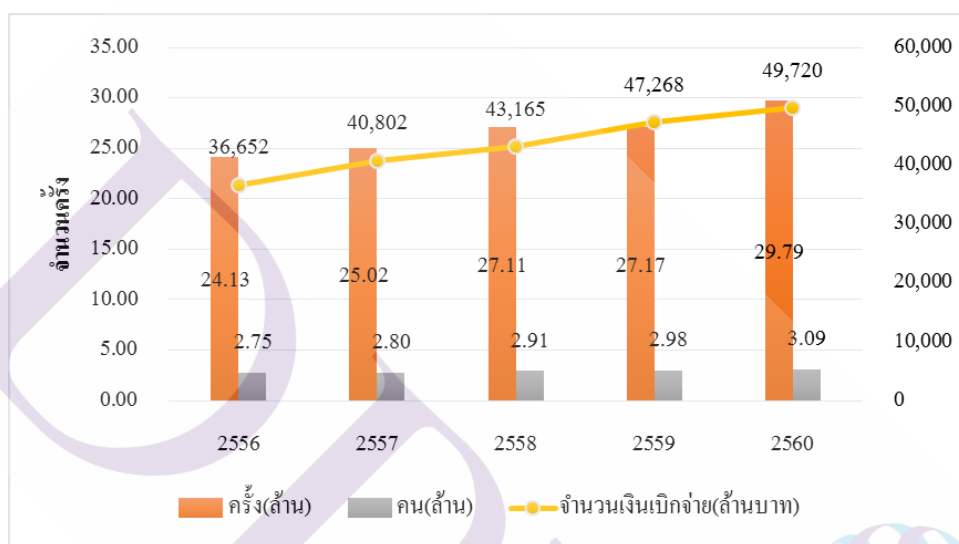
1.1.1 ภาพรวมของระบบการเบิกจ่ายเงินค่ารักษาพยาบาล

ปัจจุบันการเบิกจ่ายเงินค่ารักษาพยาบาลสิทธิสวัสดิการข้าราชการระบบเบิกจ่ายตรง จากข้อมูลของสำนักสารสนเทศบริการสุขภาพ (สทส) ซึ่งดูแลระบบการเบิกจ่ายค่ารักษาพยาบาลของข้าราชการ จำนวน 5 ล้านคน ครอบคลุม บิดา มารดา บุตรและข้าราชการบำนาญ จากข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ ปีงบประมาณ 2556-2560 พบว่าในปี 2560 ค่าใช้จ่ายต่อครั้งเท่ากับ 1,682 บาทต่อครั้ง และค่าใช้จ่ายต่อหัวเท่ากับ 16,215 บาทต่อคน และค่าใช้จ่ายรวม 49,720 ล้านบาท ดังภาพที่ 1.1



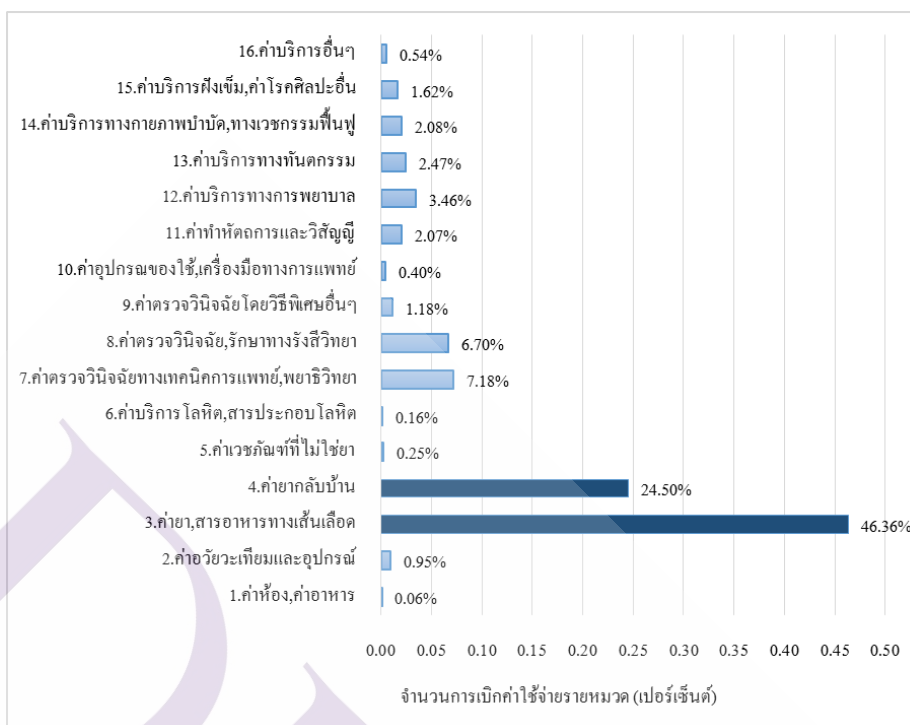
ภาพที่ 1.1 ค่ารักษาพยาบาลผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ ปีงบประมาณ 2556-2560

ภาพที่ 1.2 แสดงครั้งการไปใช้บริการของผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ พบว่าจำนวนการเบิกจ่ายเงินค่ารักษาพยาบาลมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ในปี 2560 ผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ ไปใช้บริการจำนวน 29.79 ล้านครั้ง/ปี จำนวนผู้ไปใช้บริการจำนวน 3.09 ล้านคน มีการเบิกจ่ายเงินค่ารักษาพยาบาล จำนวน 49,720 ล้านบาท



ภาพที่ 1.2 จำนวนครั้งที่ไปใช้บริการผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ

ภาพที่ 1.3 จากการเบิกจ่ายเงินค่ารักษาพยาบาลรายคนของผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการรายหมวดทั้ง 16 หมวด ในระบบเบิกจ่ายตรงปีงบประมาณ 2556 รวมเป็นเงินจำนวน 36,652 ล้านบาท หมวดที่มีการเบิกจ่ายสูงสุดคือ หมวดค่ายาที่ใช้ในสถานพยาบาล คิดเป็นร้อยละ 46 และยากลับไปใช้ที่บ้าน มีการเบิกจ่ายคิดเป็นร้อยละ 24 ของค่าใช้จ่ายทั้งหมด



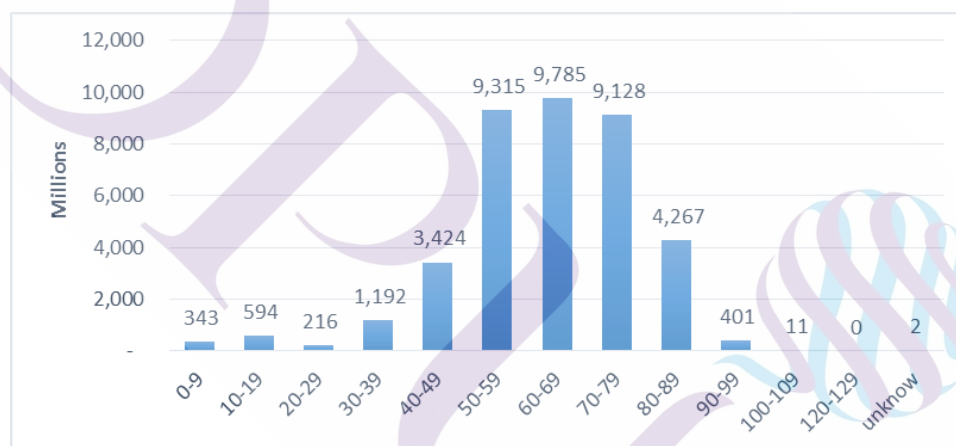
ภาพที่ 1.3 จำนวนการเบิกจ่ายเงินค่ารักษาพยาบาลทั้ง 16 หมวด

จากข้อมูลที่กล่าวมาข้างต้นเพื่อศึกษาหมวดค่ายานำกลับไปใช้ที่บ้าน (หมวด 4) พบความผิดปกติของการไปใช้บริการหลายสถานพยาบาล จำนวนสูงสุด 21 แห่ง และจำนวนครั้งที่ไปใช้บริการสูงสุด 597 ครั้ง เพื่อตรวจสอบความผิดปกติดังกล่าวต้องอาศัยผู้เชี่ยวชาญจากการตรวจสอบการเบิกค่าใช้จ่าย ซึ่งไม่สามารถเห็นพฤติกรรมภาพรวมของผู้เบิกค่าใช้จ่ายได้ (Rashidian et al., 2012) เดิมวิธีการตรวจสอบการเบิกค่าใช้จ่ายนี้ใช้เวลาในการตรวจสอบนานและไม่มีประสิทธิภาพ หลายประเทศใช้วิธีการนี้ (Copeland, Edberg, Panorska, & Wendel, 2013; Aral, Güvenir, Sabuncuoglu, & Akar, 2012; Ortega, Figueroa, & Ruz, 2006) อย่างไรก็ตาม การตรวจสอบการเบิกจ่ายเงินของประเทศไทยทำได้ยากและใช้เวลาในการตรวจสอบนานเช่นกัน ดังนั้นการนำอัลกอริทึมการทำเหมืองข้อมูล (Data mining) มาใช้ในการตรวจสอบวิเคราะห์ข้อมูลด้วยเทคนิคการเรียนรู้โดยไม่มีผู้สอน (Unsupervised learning) จะสามารถช่วยค้นหาความผิดปกติในข้อมูลได้ งานวิจัยนี้จึงเกิดขึ้นเพื่อช่วยให้สามารถตรวจสอบรูปแบบผิดปกติได้อย่างมีประสิทธิภาพและรวดเร็วมากขึ้น

1.1.2. ภาพรวมของการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้าน

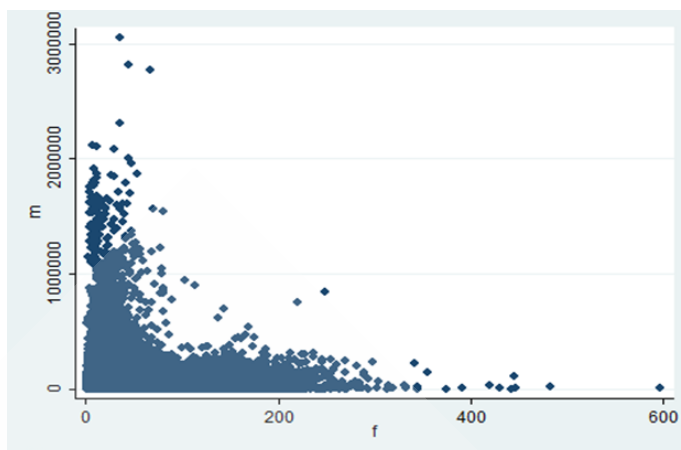
ภาพรวมการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้านของผู้ป่วยนอกสิทธิรักษาพยาบาล สวัสดิการข้าราชการระบบเบิกจ่ายตรง โดยชี้ให้เห็นถึงการไปใช้บริการที่สถานพยาบาล ซึ่งข้อมูลที่นำมาวิเคราะห์ภาพรวมเป็นข้อมูลการเบิกจ่ายค่ายานำกลับไปใช้ต่อที่บ้านระบบเบิกจ่ายตรงจาก กรมบัญชีกลาง ปีงบประมาณ 2556

จากภาพที่ 1.4 การเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้านของผู้ป่วยนอกสิทธิรักษาพยาบาล สวัสดิการข้าราชการผู้ป่วยนอก ปีงบประมาณ 2556 รายช่วงอายุช่วงอายุตั้งแต่ 50-79 ปี มีจำนวนการเบิกจ่ายค่ายานำกลับไปจำนวนมาก เบิกจ่ายสูงสุดในช่วงอายุ 60-69 ปี คิดเป็นร้อยละ 25.3 รองลงคือช่วงอายุ 50-59 ปี คิดเป็นร้อยละ 24 และช่วงอายุ 70-79 ปี คิดเป็นร้อยละ 23



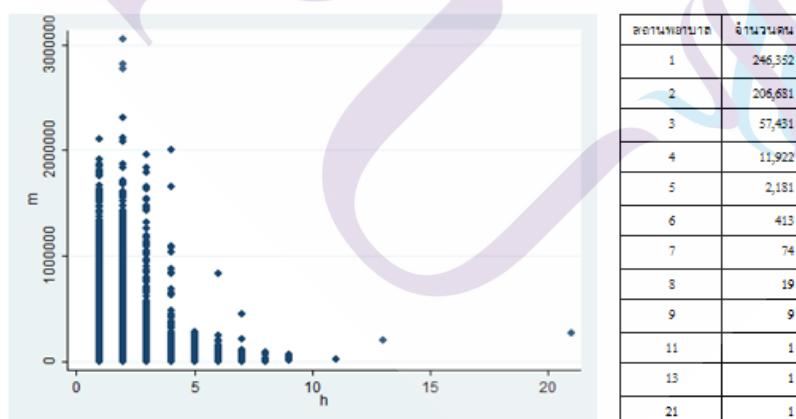
ภาพที่ 1.4 จำนวนการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้านของผู้ป่วยนอก รายช่วงอายุ

จากภาพที่ 1.5 แสดงการกระจายตัวของข้อมูลการเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้าน หมวด 4 และความถี่ของการไปใช้บริการที่สถานพยาบาล พบว่ามีจำนวนความถี่สูงสุดที่ไปใช้บริการในรอบ 1 ปี เท่ากับ 597 ครั้ง/ปี และจำนวนเงินที่เบิกจ่ายค่ายาสูงสุดจำนวน 3,055,881.8 บาท



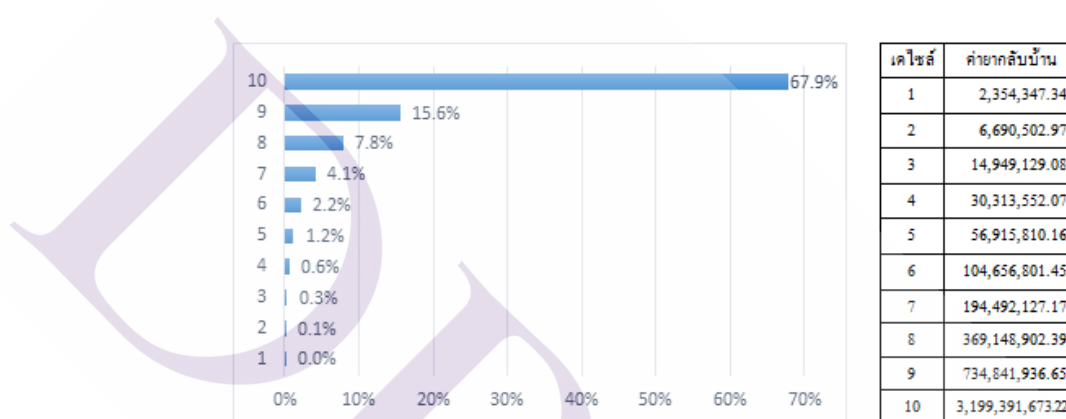
ภาพที่ 1.1 จำนวนการเบิกจ่ายเงินค่ายากลับบ้านและความถี่ที่ใช้บริการสถานพยาบาล

จากภาพที่ 1.6 พบว่าผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการที่ใช้บริการสถานพยาบาลแห่งเดียวมีจำนวน 246,352 คน คิดเป็นร้อยละ 47 ผู้มีสิทธิไปใช้บริการสถานพยาบาล 2-5 แห่ง มีจำนวน 278,215 คน และผู้มีสิทธิไปใช้บริการสถานพยาบาลมากกว่า 6 แห่ง มีจำนวนทั้งหมด 518 คน



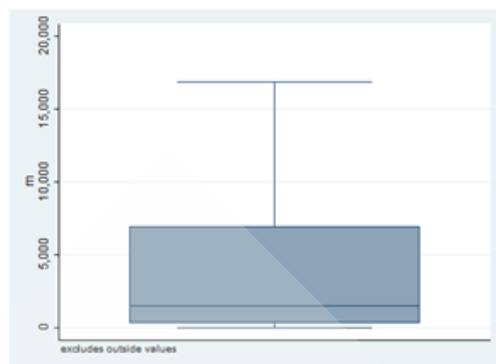
ภาพที่ 1.6 จำนวนครั้งที่ไปใช้บริการหลายสถานพยาบาล (แกนตั้ง) จำนวนสถานพยาบาลที่ใช้บริการ (แกนนอน)

ภาพที่ 1.7 แสดงกลุ่มการเบิกจ่ายเงินค่ายาคลับบ้านแบ่งออกเป็น 10 กลุ่มเดซิล์เท่ากัน (Decline) ผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการเบิกจ่ายเงินค่ายาคลับไปใช้ต่อที่บ้าน รวมทั้งสิ้น จำนวน 4,713 ล้านบาท จากจำนวนผู้ป่วยนอกทั้งหมด 525,085 คน พบว่าผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการกลุ่มที่ 10 คิดเป็นร้อยละ 10 ของผู้ป่วยนอกทั้งหมด มีการเบิกจ่ายเงิน ร้อยละ 67 ด้วยจำนวนเงิน 3,199 ล้านบาท ส่วนผู้มีสิทธิรักษาพยาบาลสวัสดิการข้าราชการอีก 9 กลุ่มที่เหลือร้อยละ 90 เบิกจ่ายค่ายาคลับบ้านเป็นเงินจำนวน 1,514 ล้านบาท



ภาพที่ 1.7 ร้อยละการเบิกจ่ายเงินค่ายาคลับไปใช้ต่อที่บ้าน ของผู้ป่วยนอกสิทธิสวัสดิการข้าราชการ จำแนกเป็น 10 กลุ่ม

จากภาพที่ 1.8 แสดงจำนวนผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการทั้งหมดที่มีการเบิกจ่ายเงินค่ายาคลับไปใช้ต่อที่บ้านจำนวน 525,085 คน ค่าเฉลี่ยในการเบิกจ่ายเงินค่ายาต่อคนเท่ากับ 8,977 บาท ค่ามัธยฐาน (Median) เท่ากับ 1,447 บาท และผู้มีสิทธิเบิกจ่ายเงินค่ายาคลับไปใช้ต่อที่บ้านสูงสุดอยู่ที่ 16,832 บาท



Median	1,447
P25	281
P75	6,901
IQR	6,620
Mean	8,977
Min	0
Max	16,832

ภาพที่ 1.8 การกระจายข้อมูลการเบิกจ่ายเงินค่ายานำกลับไปใช้ที่บ้าน

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อค้นหาพฤติกรรมที่ผิดปกติ ของการไปใช้บริการที่สถานพยาบาล โดยการประยุกต์ใช้เทคนิคการแบ่งกลุ่ม
2. สร้างโมเดลเพื่อจำแนกกลุ่มผู้ที่มีพฤติกรรมผิดปกติและกลุ่มปกติ

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบรูปแบบของกลุ่มที่มีพฤติกรรมต้องสงสัยที่ไปใช้บริการมากเกินไปจนเกิดความจำเป็น
2. สามารถลดระยะเวลาในการตรวจสอบข้อมูล
3. เพิ่มประสิทธิภาพในการตรวจสอบ เครื่องมือดาต้าไมน์นิ่งสามารถช่วยให้ค้นหากลุ่มคนต้องสงสัยได้ง่ายขึ้น

1.4 ขอบเขตของงานวิจัย

เพื่อค้นหารูปแบบที่ผิดปกติจากข้อมูลผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการระบบเบิกจ่ายตรง กรมบัญชีกลาง ปีงบประมาณ 2556 สถานพยาบาล จำนวน 168 แห่ง จำนวนผู้ให้บริการ 2,751,425 ล้านคน จำนวนครั้งที่ไปใช้บริการทั้งหมด 24,125,143 ล้านครั้ง

1.5 ระเบียบวิธีวิจัย

ศึกษาทฤษฎีและผลการวิจัยที่เกี่ยวข้อง โดยใช้อัลกอริทึมการทำเหมืองข้อมูล (Data mining) เพื่อค้นหารูปแบบ (pattern) การไปใช้บริการสถานพยาบาลและเบิกจ่ายเงินค่ายานำกลับไปใช้ที่บ้านของผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ กรณีไปสถานพยาบาลหลายแห่งในเวลาใกล้เคียงกันและมีความถี่ที่ไปใช้บริการจำนวนมาก โดยใช้เทคนิคการแบ่งกลุ่มลูกค้า (Segmentation) RFMHPE Model และการแบ่งกลุ่มข้อมูลเทคนิคเคมีน (K-means clustering) เพื่อเปรียบเทียบระหว่าง 2 เทคนิค ให้มั่นใจว่าผู้ไปใช้บริการมีพฤติกรรมผิดปกติจริงตามสมมุติฐาน และทำนายผลด้วยแบบจำลองการเรียนรู้แบบต้นไม้ตัดสินใจ (Decision tree)

1.6 นิยามคำศัพท์

ผู้ป่วยนอก หมายถึง ผู้เข้ารับการรักษาพยาบาลไม่ได้พักค้างในสถานพยาบาล (ตรวจและรับยาแล้วให้กลับบ้านได้)

การรักษาพยาบาล หมายถึง การรักษาคนที่ไม่สบาย เจ็บป่วย มีความบกพร่องหรือผิดปกติทางร่างกายหรือจิต

สถานพยาบาล หมายถึง สถานพยาบาลของทางราชการ เช่น โรงพยาบาลรัฐ ศูนย์อนามัย และสถานพยาบาลเอกชน

สวัสดิการ หมายถึง ผลประโยชน์ที่รัฐจัดให้ ซึ่งสามารถเปลี่ยนแปลง เพิ่มเติมได้ตามความเหมาะสม ขึ้นอยู่กับสถานะการเงินการคลังของประเทศไทยในขณะนั้น เช่น ค่ารักษาพยาบาลค่ารักษาพยาบาล เป็นต้น

ค่ารักษาพยาบาล หมายถึง ค่าใช้จ่ายที่เกิดขึ้นจากการรักษาพยาบาล ดังต่อไปนี้

1. ค่ายา ค่าเวชภัณฑ์ ค่าอุปกรณ์ทางการแพทย์ ค่าเลือดและส่วนประกอบของเลือดหรือสารทดแทน ค่าน้ำยาหรืออาหารทางเส้นเลือด ค่าออกซิเจน และอื่น ๆ ทำนองเดียวกันที่ใช้ในการบำบัดรักษาโรค

2. ค่าอวัยวะเทียมและอุปกรณ์ในการบำบัดรักษาโรค รวมทั้งค่าซ่อมแซมอวัยวะเทียมและอุปกรณ์

3. ค่าบริการทางการแพทย์ ค่าบริการทางการแพทย์ ค่าตรวจวินิจฉัยโรค ค่าวิเคราะห์โรค แต่ไม่รวมถึงค่าธรรมเนียมแพทย์พิเศษ ค่าจ้างผู้พยาบาลพิเศษ ค่าธรรมเนียมพิเศษ และค่าบริการอื่นทำนองเดียวกันที่มีลักษณะเป็นเงินตอบแทนพิเศษ

4. ค่าตรวจครรภ์ ค่าคลอดบุตรและการดูแลหลังคลอดบุตร
5. ค่าห้องและค่าอาหาร ตลอดระยะเวลาที่เข้ารับการรักษาพยาบาล
6. ค่าใช้จ่ายเพื่อเป็นการเสริมสร้างสุขภาพและป้องกันโรค
7. ค่าฟื้นฟูสมรรถภาพร่างกายและจิตใจ
8. ค่าใช้จ่ายอื่นที่จำเป็นแก่การรักษาพยาบาลตามที่กระทรวงการคลังกำหนด

ผู้มีสิทธิ หมายถึง ข้าราชการ ลูกจ้างประจำ ผู้รับเบี้ยหวัดบำนาญ และลูกจ้างชาวต่างประเทศ ตามพระราชบัญญัติการเงินสวัสดิการเกี่ยวกับการรักษาพยาบาล พ.ศ. 2523

บุคคลในครอบครัว หมายถึง บิดา มารดา คู่สมรส และบุตร ที่ถูกต้องตามกฎหมาย (ของผู้มีสิทธิ)

หลอกลวง (Fraud) หมายถึง เจตนาทำการส่งรายการเรียกเก็บเงิน (Claim) หรือเป็นสาเหตุให้มีส่งรายการเรียกเก็บเงินที่ไม่ถูกต้อง หรือทำไม่ตรงความเป็นจริง เพื่อให้ได้รับการจ่ายชดเชยโดยรัฐสำหรับรายการที่สมควรต้องจ่าย

ละเมิด (Abuse) หมายถึง การดำเนินการโดยตรงหรือโดยอ้อมที่ส่งผลให้เกิดค่าใช้จ่ายที่ไม่จำเป็นต่อโปรแกรมการรักษา รวมถึงการให้บริการที่ไม่เป็นไปตามบริการทางสุขภาพที่จำเป็น ตามมาตรฐานที่วิชาชีพ และราคาที่เหมาะสม

บทที่ 2

แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิเคราะห์พฤติกรรมผู้ไปใช้บริการที่สถานพยาบาลแบบไม่สมเหตุสมผล จากข้อมูลการเบิกจ่ายเงินค่ายานำกลับไปใช้ที่บ้าน ของผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิการข้าราชการ จำนวน 168 แห่ง ผู้วิจัยได้ทำการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องเพื่อสามารถนำทฤษฎีและงานวิจัยที่เกี่ยวข้องเหล่านี้เป็นแม่แบบและประยุกต์ใช้ต่อไป เพื่อให้เห็นภาพรวมของระบบการเบิกจ่ายเงินค่ารักษาพยาบาลสวัสดิการข้าราชการ ระบบเบิกจ่ายตรง ในส่วนแรกจะชี้ให้เห็นขั้นตอนการเบิกจ่ายค่ารักษาพยาบาล ขั้นตอนการสมัครเข้าร่วมโครงการเบิกจ่ายตรง ณ สถานพยาบาล และระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลาง (คู่มือสวัสดิการรักษายาบาลข้าราชการ, 2551, 2553) ในส่วนที่สองคือการนำอัลกอริทึมการทำเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้กับข้อมูลการเบิกจ่ายเงินค่ายานำกลับไปใช้ที่บ้าน

2.1 ขั้นตอนการสมัครเข้าร่วมโครงการเบิกจ่ายตรงผู้ป่วยนอก

1. ผู้มีสิทธิ์หรือบุคคลในครอบครัวที่ประสงค์จะสมัครเข้าร่วมโครงการเบิกจ่ายตรงผู้ป่วยนอกนั้น จะต้องจัดทำฐานข้อมูลบุคลากรภาครัฐของตนให้สมบูรณ์ โดยผู้มีสิทธิ์สามารถปรับปรุง แก้ไข หรือเพิ่มเติมข้อมูลของตน หรือบุคคลในครอบครัวได้ โดยผ่าน “นายทะเบียนของส่วนราชการ” ที่ผู้มีสิทธิ์สังกัดอยู่

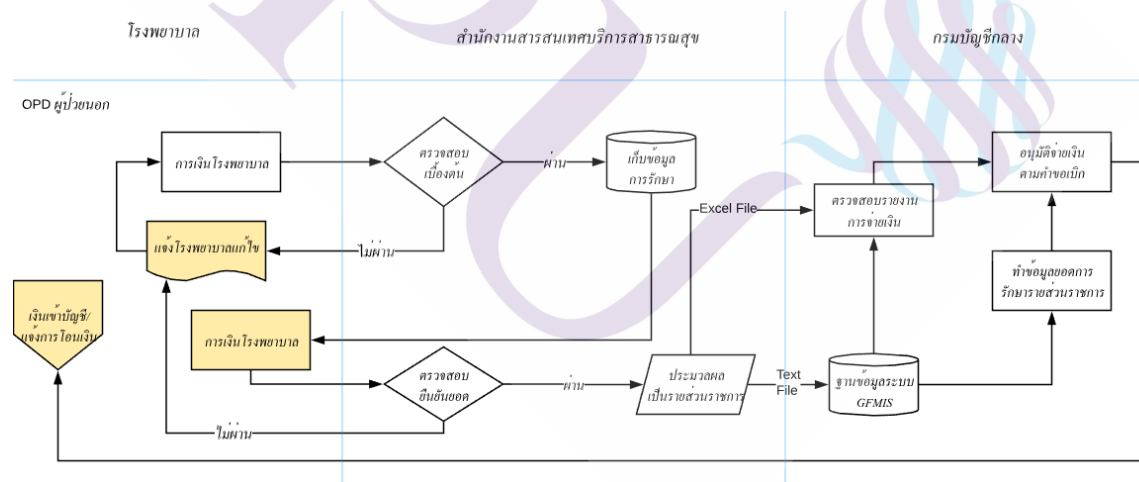
2. เมื่อนายทะเบียนปรับปรุง แก้ไข หรือเพิ่มเติมข้อมูลให้ผู้มีสิทธิ์และบุคคลในครอบครัวเรียบร้อยแล้ว ข้อมูลดังกล่าวจะถูกส่งมารวมใน “ฐานข้อมูลบุคลากรภาครัฐ” ของกรมบัญชีกลาง ในทันทีที่กรมบัญชีกลางจะนำฐานข้อมูลดังกล่าวส่งให้สถานพยาบาลของทางราชการทั้ง 991 แห่ง ทุก 15 วัน (ประมาณ วันที่ 4 และ 18 ของทุกเดือน) เพื่อใช้ในการตรวจสอบสิทธิผ่านเว็บไซต์ของกรมบัญชีกลาง (www.cgd.go.th) และเว็บไซต์ของ สกส. (www.chi.or.th)

3. ผู้มีสิทธิและบุคคลในครอบครัวสามารถตรวจสอบสิทธิว่าตนและบุคคลในครอบครัวมีสิทธิสมัครเข้าร่วมโครงการเบิกจ่ายตรงได้แล้วหรือไม่ หากพบว่ามีสิทธิ แสดงว่าสามารถไปสมัครเข้าร่วมโครงการเบิกจ่ายตรงฯ ณ สถานพยาบาลที่ผู้มีสิทธิ หรือบุคคลในครอบครัวประสงค์จะใช้บริการ สามารถสมัครได้มากกว่า 1 แห่ง

4. ภายหลังจากการสมัครเข้าร่วมโครงการเบิกจ่ายตรงฯ ณ สถานพยาบาลแล้วจะต้องการอนุมัติการเข้าร่วม คือได้ “เลขอนุมัติ” ซึ่งจะใช้เวลาประมาณ 15 วัน ผู้มีสิทธิหรือบุคคลในครอบครัวจึงจะใช้ระบบได้ (เลขอนุมัติที่ได้รับสามารถใช้ในการเข้ารักษาพยาบาลผู้ป่วยนอกสถานพยาบาลนั้นๆ ได้ทุกครั้ง)

5. กรณีผู้มีสิทธิ หรือบุคคลในครอบครัวไม่ได้เข้ารับการรักษาสถานพยาบาลที่สมัครเข้าร่วมโครงการเบิกจ่ายตรงฯ ใ้ให้ทดลองจ่ายเงินค่ารักษาพยาบาลไปก่อน และสามารถนำใบเสร็จค่ารักษาพยาบาลมาเบิกจากหน่วยงานต้นสังกัดได้ตามปกติ

6. ทั้งนี้ ผู้มีสิทธิ หรือบุคคลในครอบครัวที่มีสิทธิซ้ำซ้อน (เช่น มีสิทธิประกันสังคม สิทธิองค์กรปกครองส่วนท้องถิ่น สิทธิรัฐวิสาหกิจ สิทธิองค์กรอิสระ เป็นต้น) ไม่สามารถเข้าร่วมโครงการเบิกจ่ายตรงฯ ได้ กลุ่มมือสวัสดิการรักษายาบาลข้าราชการ (2551, เล่ม 1 น. 39)

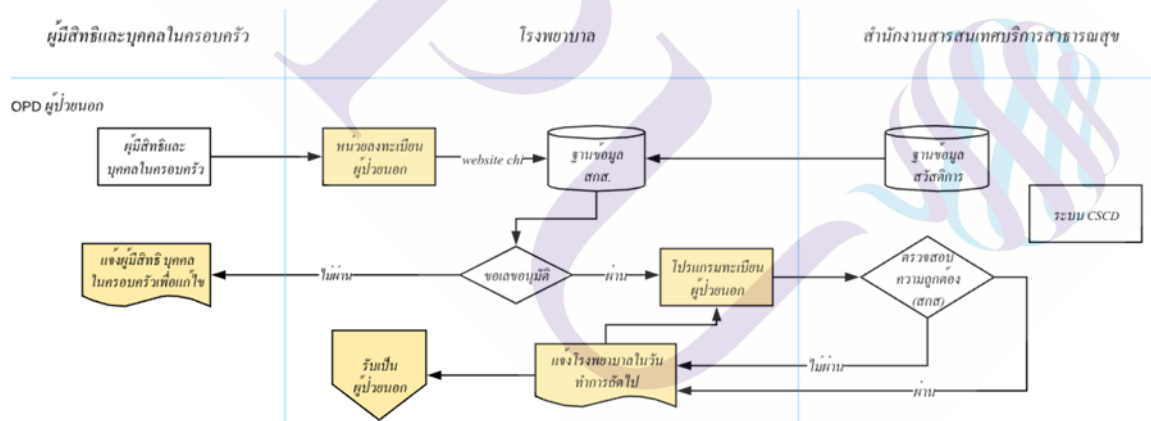


ภาพที่ 2.1 ขั้นตอนการสมัครเข้าร่วมโครงการเบิกจ่ายตรง ณ สถานพยาบาล

ที่มา: คู่มือสวัสดิการรักษายาบาลข้าราชการ (2551, เล่ม 1 น. 39)

2.2 ระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลาง

ระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลางนั้น ได้เริ่มดำเนินการเบิกจ่ายด้วยระบบอิเล็กทรอนิกส์ (ระบบเบิกจ่ายตรง) ตั้งแต่ปีงบประมาณ พ.ศ. 2545 ต่อมาได้มีการปรับเปลี่ยนเชื่อมโยงวิธีดังกล่าวเข้ากับระบบบริการงานการเงินการคลังภาครัฐ ซึ่งสำนักงานกลางสารสนเทศบริการสุขภาพ (สกส.) เป็นผู้วางระบบและบริการจัดการระบบดังกล่าว ได้พัฒนาระบบการเบิกจ่ายเงินกรณีผู้ป่วยนอก ทั้งนี้ เมื่อข้อมูลการรักษาพยาบาลได้ผ่านการตรวจสอบเบื้องต้นจาก สกส. แล้ว จะถูกส่งกลับไปยังสถานพยาบาลเพื่อยืนยัน และรับรองความถูกต้องของข้อมูลดังกล่าว พร้อมทั้งให้สถานพยาบาลจัดทำเอกสารคำขอเบิกเงินสวัสดิการรักษายาพยาบาลในระยะเวลาที่กำหนด เพื่อแสดงเจตนาในการขอเบิกเงินกับกรมบัญชีกลางผ่านระบบ GFMS การเบิกจ่ายค่ารักษาพยาบาลให้กับสถานพยาบาลนั้น กรมบัญชีกลางจะจ่ายให้ กรณีผู้ป่วยนอก : สัปดาห์ละ 1 ครั้ง (เดือนละ 4 ครั้ง) คู่มือสวัสดิการรักษายาพยาบาลข้าราชการ (2551, เล่ม 1 น. 53)



ภาพที่ 2.2 ระบบเบิกจ่ายเงินค่ารักษาพยาบาลระหว่างสถานพยาบาลกับกรมบัญชีกลาง

ที่มา: คู่มือสวัสดิการรักษายาพยาบาลข้าราชการ (2551, เล่ม 1 น. 53)

2.3 ทฤษฎีการทำเหมืองข้อมูล

อัลกอริทึมการทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น (Han,J., et al 2006) การทำเหมืองข้อมูล เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถ ค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำมาประยุกต์ใช้ในงานหลาย ประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์ การแพทย์ ด้าน เศรษฐกิจและสังคม

2.4 วัตถุประสงค์ของการทำเหมืองข้อมูล

1. เพื่อการค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge Discovery in Database)
2. เพื่อการสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge Extraction)
3. เพื่อจัดการกับข้อมูลในอดีต (Data Archeology)
4. เพื่อสำรวจข้อมูล (Data Exploration)
5. เพื่อค้นหาแพตเทิร์นของข้อมูลที่อยู่ (Data Pattern Processing)
6. เพื่อใช้ขุดเจาะข้อมูล (Data Dredging)
7. เพื่อให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

2.5 เป้าหมายหลักของการทำเหมืองข้อมูล

คุณลักษณะและเป้าหมายหลัก คือ ใช้ค้นหารูปแบบของข้อมูลที่ฝังลึกและซ่อนเร้นอยู่ภายใต้ ฐานข้อมูลขนาดใหญ่ โดยไม่ต้องอาศัยความชำนาญของผู้ใช้ ซึ่งสามารถค้นพบข้อมูลไม่คาดหวังมา ก่อน

2.6 เทคนิคการวิเคราะห์เหมืองข้อมูล

เทคนิคการวิเคราะห์เหมืองข้อมูลนั้นสามารถแบ่งได้เป็น 2 ประเภท

1. เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เทคนิคนี้จะเน้นที่การพิจารณาข้อมูลเป็นหลัก เช่น พิจารณาว่าข้อมูลมีความสัมพันธ์กันในลักษณะใดบ้าง เทคนิคในประเภทนี้จะแบ่งย่อยได้อีก คือ เทคนิคการค้นหากฎความสัมพันธ์ (association rule) และการแบ่งกลุ่มข้อมูล (clustering)

2. เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised learning) เทคนิคในประเภทนี้จะเน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตเพื่อนำมาสร้างแบบจำลองสำหรับทำนายหรือคาดการณ์สิ่งที่เกิดขึ้นในอนาคต แบบจำลองในที่นี้อาจจะเป็นสมการทางคณิตศาสตร์หรือ กฎต่างๆ ก็ได้ เทคนิคการเรียนรู้แบบมีผู้สอนนี้สามารถแบ่งย่อยได้ คือ การจำแนกประเภทข้อมูล (classification) และการประมาณค่าข้อมูล (regression) ซึ่งทั้งเทคนิคจะมีลักษณะที่คล้ายกันมากแต่แตกต่างกันที่คำตอบที่ต้องการทำนาย ซึ่งการจำแนกประเภทข้อมูลทำนายข้อมูลที่มีค่าเป็นนอมินอล (nominal) เช่น เพศชาย หญิง หรือค่าที่ไม่ใช่ตัวเลข ส่วนการประมาณค่าข้อมูลจะใช้กับข้อมูลคำตอบที่เป็นตัวเลขเท่านั้น (เอกสิทธิ์ พัชรวงศ์ศักดิ์, 2557, น. 15)

2.6.1 การวิเคราะห์ข้อมูลเพื่อป้องกันการหลอกลวง (Fraud Detection)

การนำเทคนิคการเรียนรู้แบบไม่มีผู้สอนมาใช้เพื่อค้นหารูปแบบผิดปกติในข้อมูล ซึ่งพฤติกรรมที่ค่าแตกต่างจากค่าเฉลี่ย สามารถกำหนดเป็นพฤติกรรมของลูกค้าโดยเฉลี่ยได้ หรือวิเคราะห์ทำนาย (predictive) กับข้อมูลที่มีลาเบล การสร้างแบบจำลองตรวจสอบรูปแบบข้อมูลที่ผิดปกติได้ การตรวจสอบการหลอกลวงกับข้อมูลที่ไม่รู้คำตอบ เพื่อค้นหาความผิดปกติที่น่าสงสัย มีประโยชน์สำหรับองค์กรที่เริ่มตรวจสอบ และไม่มีข้อมูลทดสอบ (test data) ซึ่งเป็นประวัติข้อมูลผิดปกติ สามารถใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอนมาใช้ในการตรวจสอบความผิดปกติที่ใช้กับข้อมูลประเภทนี้ได้ (Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke , 2015, p. 123) จะกล่าวถึงต่อไป

2.7 การใช้อัลกอริทึมแมชชีน เลิร์นนิง (Machine Learning) กับข้อมูลสุขภาพ (Healthcare)

การรวบรวมอัลกอริทึมแมชชีน เลิร์นนิง ที่สำคัญและใช้ในการจำแนกลักษณะข้อมูลสุขภาพ ซึ่งอัลกอริทึมเหล่านี้มีรูปแบบที่แตกต่างกัน ตามประเภทของข้อมูลที่ใส่เข้าไป เช่น อัลกอริทึมการเรียนรู้แบบมีผู้สอน คือ ข้อมูลที่มีการลาเบล (label) กำกับ และอัลกอริทึมการเรียนรู้แบบไม่มีผู้สอน คือ

ข้อมูลที่ไม่รู้คำตอบ โดยใส่เครื่องหมาย (*) ดังตารางข้างล่างนี้ (Prashant Natarajan, John C Frenzel, Detlev H Smaltz, 2017, p. 89-98)

ตารางที่ 2.1 อัลกอริทึมแมชชีน เลิร์นนิ่ง กับข้อมูลสุขภาพ

วิธี	อัลกอริทึม
1. Classifier	<ul style="list-style-type: none"> - Logistic regression - Support Vector Machine (SVM) - Decision tree, random forest, boosted trees - Deep Learning
2. Memory-Based Learning*	<ul style="list-style-type: none"> - Associative memory
3. Topic Modeling	<ul style="list-style-type: none"> - Latent Dirichlet Allocation (LDA). - Probabilistic Latent Semantic Analysis (pLSA), Latent Semantic Analysis (LSA).
4. Forecasting	<ul style="list-style-type: none"> - Linear regression* - Neural networks* - Exponential smoothing* - Auto-Regressive Integrated Moving Average (ARIMA) modeling*
5. Probability Estimation	<ul style="list-style-type: none"> - Probabilistic Graph Model (PGM) - Logistic regression.
6. Image and Video Understanding	<ul style="list-style-type: none"> - Deep learning
7. Speech to Text	<ul style="list-style-type: none"> - Hidden Markov Models (HMMs) - Deep learning, especially Long Short-Term Memory (LSTM)
8. Recommender Systems	<ul style="list-style-type: none"> - Collaborative filtering - Memory-based learning - Association rules.
9. Clustering	<ul style="list-style-type: none"> - Unsupervised Clustering* - Hierarchical clustering*
10. Text Understanding	<ul style="list-style-type: none"> - Natural Language Processing (NLP) - Text mining. - Deep learning.

จากภาพที่ 2.3 แสดงเครื่องมือที่ใช้วิเคราะห์แก้ปัญหาเกี่ยวกับข้อมูลขนาดใหญ่ (Prashant Natarajan, John C Frenzel, Detlev H Smaltz, 2017, p. 98) ในงานวิจัยชิ้นนี้ได้ใช้การแบ่งกลุ่มด้วยเทคนิคเคมีน



ภาพที่ 2.3 แสดงเครื่องมือที่ใช้วิเคราะห์แก้ปัญหาข้อมูลขนาดใหญ่

ที่มา: Demystifying Big Data and Machine Learning for Healthcare , (Prashant Natarajan, John C Frenzel, Detlev H Smaltz, 2017, p. 89-98)

2.8 รูปแบบการละเมิด (Abuse)

การละเมิด หมายถึง การดำเนินการโดยตรงหรือโดยอ้อมที่ส่งผลให้เกิดค่าใช้จ่ายที่ไม่จำเป็นต่อ โปรแกรมการรักษา (Medicare Program) รวมถึงการให้บริการที่ไม่เป็นไปตามบริการทางสุขภาพที่จำเป็น ตามมาตรฐานที่วิชาชีพ และราคาที่เหมาะสม

2.9 รูปแบบการหลอกลวง (Fraud)

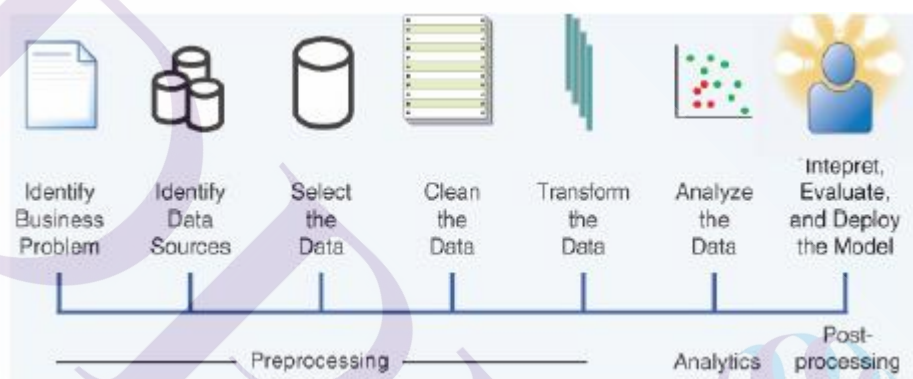
การหลอกลวง (fraud) หมายถึง การหลอกลวงในรูปแบบพฤติกรรมที่ไม่ตรงตามความเป็นจริง โดยมีเจตนาบิดเบือนไม่เหมาะสม สามารถจำแนกพฤติกรรมหลอกลวงได้ ดังเช่น การหลอกลวงทางด้านการดูแลสุขภาพ (healthcare fraud) เพื่อเบิกค่าใช้จ่ายชดเชยการรักษาที่ไม่ถูกต้อง โดยมีเจตนาหรือไม่เจตนา เพื่อหวังผลกำไร รวมไปถึงการเบิกยาที่ไม่จำเป็น แล้วนำไปขายเพื่อทำกำไร และนำไปเสริมมาเบิกในระบบ หรือการเบิกค่ารักษาพยาบาลซ้ำซ้อน และการลงบันทึกข้อมูลเท็จในระบบ หรือการปลอมการขายยาตามใบสั่งแพทย์ และการใช้บัตรประกันของผู้อื่น (Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke , 2015, p. 41-44)

2.9.1 การป้องกันและตรวจสอบการหลอกลวง (Fraud Detection)

การตรวจสอบและป้องกันการหลอกลวง มี 2 วิธีที่มีประสิทธิภาพในการตรวจสอบและป้องกันการ 1) การตรวจสอบจากผู้เชี่ยวชาญ โดยทำความเข้าใจและค้นพบรูปแบบการหลอกลวง ซึ่งเป็นการป้องกันการสูญเสียได้ในอนาคต เช่น การเบิกค่ารักษาพยาบาลจากอุบัติเหตุ กับบริษัทประกัน ถ้าไม่มีบันทึกรายงานจากตำรวจหรือการบาดเจ็บสาหัสแต่ไม่มีรายงานทางการแพทย์ รวมไปถึงการส่งใบเสร็จหลายรายการ ผู้เชี่ยวชาญตรวจสอบให้เป็นกรณีที่น่าสงสัย แต่วิธีการตรวจสอบของผู้เชี่ยวชาญพบข้อดีหลายประการ เช่น เครื่องมือที่ใช้ในการตรวจสอบมีราคาสูงและต้องใส่เงื่อนไขเพื่อตรวจสอบด้วยตัวเอง ซึ่งเป็นเรื่องยากที่จะจัดการและปรับปรุงให้ทันสมัยอยู่เสมอ 2) ระบบการตรวจจับและป้องกันอัตโนมัติ โดยอาศัยผู้เชี่ยวชาญเป็นผู้ประเมินและตรวจสอบ เพื่อป้องกันการหลอกลวง (Bart Baesens, Veronique Van Vlasselaer, Wouter Verbeke , 2015, p. 50)

2.9.2 ขั้นตอนการวิเคราะห์แบบจำลองการหลอกลวง

จากรูป 2.4 แสดงภาพรวมของขั้นตอนการวิเคราะห์แบบจำลองการหลอกลวง (Han and Kamber 2011; Hand, Mannila, and Smyth 2001; Tan, Steinbach, and Kumar 2005) มีขั้นตอนดังนี้ 1) ทำความเข้าใจปัญหาของธุรกิจ 2) ระบุแหล่งที่มาของข้อมูล 3) การเลือกข้อมูลมาวิเคราะห์และสร้างแบบจำลอง 4) ทำความสะอาดข้อมูลจากค่าข้อมูลที่สูญหาย (missing value) และข้อมูลที่ซ้ำซ้อน 5) ทำการแปลงข้อมูล (transformed data) ในขั้นตอนนี้มีการสร้างแบบจำลองตรวจจับการหลอกลวง 6) ประมวลผลโดยผู้เชี่ยวชาญ



ภาพที่ 2.4 ขั้นตอนการวิเคราะห์แบบจำลองการหลอกลวง

ที่มา: Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques, A Guide to Data Science for Fraud Detection, 2017, น. 67)

2.10 การเก็บรวบรวมข้อมูล การสุ่มตัวอย่าง และการประมวลผล

การแบ่งกลุ่มข้อมูลก่อนนำไปใช้สร้างแบบจำลองวิเคราะห์กลยุทธ เช่น ธนาคารต้องการใช้กลยุทธเฉพาะลูกค้าบางกลุ่ม การแบ่งกลุ่มสามารถใช้แบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ (Decision trees) ด้วยเทคนิคเคมีน (K-means clustering) หรือทฤษฎีแผนผังการจัดระเบียบตัวเอง (Self-Organizing Maps : SOM) (Bart Baesens, Veronique Van Vlasselaer, Wouter Verbeke , 2015, p. 118)

2.10.1 การแบ่งกลุ่ม (Clustering) การแบ่งกลุ่ม (Clustering) คือ การแบ่งข้อมูลออกเป็นกลุ่ม จากลักษณะทางสังคมพฤติกรรม วิถีชีวิต และลักษณะการทำธุรกรรม เพื่อเป็นเกณฑ์ในการตั้งค่าการตรวจจับความผิดปกติ เช่น 1) การแบ่งกลุ่มลูกค้าตามรายการจากข้อมูลการใช้บัตรเครดิต 2) การแบ่งกลุ่มการเบิกค่ารักษาพยาบาลเพื่อตรวจสอบการเบิกจ่าย 3) การแบ่งกลุ่มภาษีและการตรวจสอบภาษี 4) การแบ่งกลุ่มการโอนเงินเพื่อตรวจสอบการฟอกเงิน 5) การแบ่งกลุ่มการป้องกันการ ที่ได้รับความนิยมนคือ RFM (Recency, Frequency, Monetary)

2.10.2 การแบ่งกลุ่มพฤติกรรมที่คล้ายกันด้วยวิธีเซกเมนต์เทชั่น (Segmentation) การจัดกลุ่มลูกค้าที่มีพฤติกรรมคล้ายคลึงกันไว้ด้วยกัน ช่วยให้สามารถดูภาพรวมของลูกค้าได้ง่ายขึ้น เช่น กลุ่มลูกค้าที่มีการซื้อสินค้าบ่อยๆและมีค่าใช้จ่ายเยอะ หรือกลุ่มลูกค้าที่ไม่ค่อยมาซื้อสินค้า แต่เมื่อมาซื้อสินค้าแต่ละครั้งจะซื้อจำนวนมาก การจัดกลุ่มพฤติกรรมที่คล้ายกัน ผู้ค้นพบ (Hughes, 1996) เพื่อวิเคราะห์พฤติกรรมความต้องการของลูกค้า ซึ่งใช้ตัวแปร (R) วันล่าสุดที่มา (F) ความถี่ (M) จำนวนเงิน หากลูกค้า มีระยะเวลาสั้นๆ แสดงโอกาสกลับมาซื้อสินค้าสูงและความถี่ในการซื้อบ่งบอกได้ว่าเป็นลูกค้าประจำ และจำนวนเงินที่ใช้ไป เมื่อนำมาวิเคราะห์ ต้องทำการแบ่งข้อมูล R, F, M เป็น 5 กลุ่ม มีจำนวนเท่าๆกัน ควินไทล์ (quintile) ร้อยละ 20

2.10.3 แบบจำลอง RFM ต่อมาได้มีการนำวิธีมาใช้โดย (Fader, 2005) เรียกแบบย่อว่า RFM วิธีการจัดกลุ่มตามพฤติกรรมการซื้อสินค้า โดยดูจาก 1. ระยะเวลา (จำนวนวัน) จากการซื้อล่าสุดที่ผ่านมา (Recency) 2. ความถี่ของการซื้อสินค้า (Frequency) 3. การใช้จ่ายของลูกค้า (Monetary)

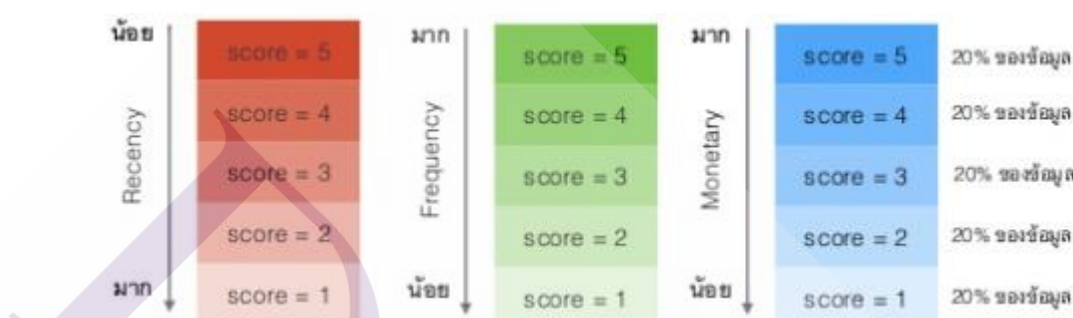
ซึ่งวิธีการนี้จะแบ่งค่า Recency (R) , Frequency (F) และ Monetary (M) ออกเป็น 5 ส่วนเท่าๆกัน โดยเลขที่มีค่ามากที่สุด (คือเลข 5) จะมีความสำคัญที่สุด และใช้ค่าตัวเลข 3 หลักเป็นตัวแทนของแต่ละกลุ่ม เช่น 555 คือ กลุ่มที่มีค่า R=5, F=5 และ M=5 หมายความว่า เป็นลูกค้าที่มักจะมาซื้อสินค้าบ่อยๆและมีค่าใช้จ่ายที่สูง ต่อมา มีการประยุกต์ใช้ LRFM มีตัวแปร (L) ระยะเวลาที่เป็นลูกค้า โดย (Reinartz, Kumar, 2000) ซึ่งแสดงถึงเป็นลูกค้าที่มีความภักดี

2.10.4 ขั้นตอนการแบ่งกลุ่มพฤติกรรมที่คล้ายกันด้วยวิธีเซกเมนต์เทชั่น

1. เริ่มจากเรียงลำดับการซื้อล่าสุดที่ผ่านมา จากน้อยไปมาก ค่าการซื้อล่าสุดที่ผ่านมา ที่มีค่าน้อย หมายความว่า ลูกค้ามาซื้อสินค้าเมื่อไม่นานมานี้ นั่นอาจจะบอกได้ว่าเป็นลูกค้าใหม่หรือลูกค้าที่มาซื้อสินค้าบ่อยๆ และถ้าค่าการซื้อล่าสุดที่ผ่านมา มีค่าน้อยย่อมแสดงว่าสำคัญ

2. เรียงลำดับค่าความถี่ของการซื้อสินค้าจากมากไปน้อย เนื่องจากลูกค้าที่ซื้อสินค้าบ่อยจะมีค่าความถี่มากและอาจหมายถึงว่าเป็นลูกค้าประจำ

3. เรียงลำดับค่าการใช้จ่ายของลูกค้าจากมากไปน้อย ยิ่งลูกค้าใช้จ่ายซื้อสินค้ามากค่าการใช้จ่ายของลูกค้าจะมากขึ้นตอนการแบ่งข้อมูลเป็น 5 กลุ่ม โดยแต่ละกลุ่มมีจำนวนเท่าๆกัน จากภาพที่ 2.5



ภาพที่ 2.1 แสดงการแบ่งกลุ่มพฤติกรรมลูกค้าวิธีเซกเมนต์เทชั่น

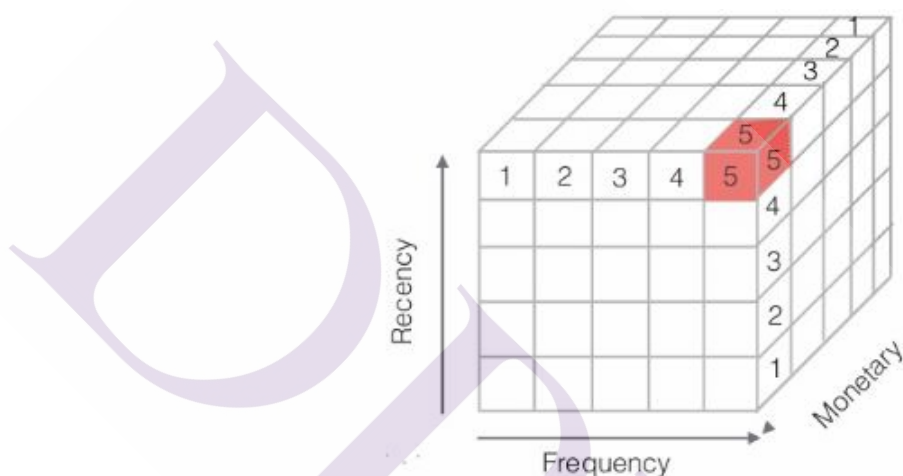
ที่มา: เอกสิทธิ์ พัทธวงศ์ศกดา. (2560). การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) และเทคนิคการค้าไมน์นิ่ง (Data Mining). (pp. 69). สืบค้นจาก https://www.slideshare.net/sitake/introduction-to-data-mining-and-big-data-analytics-70430643?next_slideshow=1

แบบจำลอง RFM ผู้ค้นพบ Hughes (1996) เพื่อวิเคราะห์พฤติกรรมความต้องการของลูกค้าจากการแบ่งกลุ่มพฤติกรรมลูกค้าด้วยวิธี RFM สามารถแสดงจากภาพที่ 2.6 ซึ่งกล่องสี่เหลี่ยมเล็กๆ แสดงว่า RFM ที่ได้ เช่น RFM = 555 (R = 5, F = 5, M = 5) ซึ่งเป็นกลุ่มที่น่าสนใจ ลูกค้าในแต่ละกลุ่มจะมีลักษณะแตกต่างกัน เช่น

1. ลูกค้ากลุ่ม RFM = 555 (R = 5, F = 5, M = 5) เป็นกลุ่มลูกค้าที่น่าสนใจที่สุดเนื่องจากการซื้อสินค้าเยอะและมาบ่อย
2. ลูกค้ากลุ่ม RFM = 551 (R = 5, F = 5, M = 1) เป็นกลุ่มลูกค้าที่มีการซื้อบ่อย แต่มีค่าใช้จ่ายที่น้อย ควรออกแคมเปญ (campaign) กระตุ้นให้ลูกค้าซื้อสินค้าที่มีราคาสูงขึ้น (up-selling)

3. ลูกค้ากลุ่ม RFM = 115 (R = 1, F = 1, M = 5) เป็นกลุ่มลูกค้าที่นานๆจะมาซื้อสินค้าสักครั้ง แต่ซื้อสินค้าที่มีราคาสูง ควรออกแคมเปญกระตุ้นให้ลูกค้าซื้อสินค้าบ่อยขึ้น

จากภาพที่ 2.6 แสดงการแบ่งกลุ่มลูกค้า โดยลูกค้ากลุ่ม RFM = 555 ลูกค้ามาซื้อสินค้าบ่อย และมีการใช้จ่ายที่สูง ลูกค้ากลุ่ม RFM = 551 เป็นกลุ่มลูกค้าที่มีการซื้อบ่อย แต่ซื้อจำนวนน้อย และลูกค้ากลุ่ม RFM = 115 เป็นกลุ่มลูกค้าที่นานๆ จะซื้อสักครั้ง แต่ซื้อสินค้าราคาสูง



ภาพที่ 2.2 กราฟแสดงกลุ่มพฤติกรรมลูกค้าด้วยวิธีเซกเมนต์แท่ง

ที่มา: เอกสิทธิ์ พันธ์วงศ์ศักดิ์. (2560). การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) และเทคนิคดาต้า ไมนิ่ง (Data Mining). (pp. 70). สืบค้นจาก https://www.slideshare.net/sitake/introduction-to-data-mining-and-big-data-analytics-70430643?next_slideshow=1

2.10.5 การแบ่งกลุ่มพฤติกรรมด้วยแบบจำลอง RFM เหมาะสมกับงานวิจัยครั้งนี้อย่างไร

1. เป็นการแบ่งกลุ่มพฤติกรรมผู้ไปใช้บริการที่เฉพาะเจาะจงออกเป็นกลุ่มพฤติกรรมที่ชัดเจนเพื่อให้เห็นรูปแบบของพฤติกรรมผู้ไปใช้บริการ
2. สามารถเลือกกลุ่มผู้ไปใช้บริการที่น่าสงสัยแล้วนำไปตรวจสอบเพิ่มเติม

2.10.6 การประยุกต์ใช้แบบจำลอง LRFMP

เนื่องจากการแบ่งกลุ่มพฤติกรรมด้วยแบบจำลอง RFM แบบเดิมไม่ได้สะท้อนถึงการซื้อตามความเป็นจริงของลูกค้า เป็นเพียงการทำธุรกรรมครั้งล่าสุดที่อาจจะทำให้เข้าใจผิดได้ ทำให้มีการประยุกต์แบบจำลอง LRFMP โดยการเพิ่มตัวแปร (L) ระยะเวลาที่เป็นลูกค้า แสดงความภักดีของลูกค้า ขึ้นมา การประยุกต์แบบจำลองนี้เกิดจาก Ha and Park (1998) ใช้สัญญาณลักษณะลูกค้าเพื่อดูสถานะผู้มาใช้บริการเป็นประจำ

1. ลูกค้าขึ้น \uparrow กับค่าเฉลี่ยของตัวแปร L, R, F, M, P ที่มีค่ามากกว่าค่าเฉลี่ยรวม
2. ลูกค้าลง \downarrow ถ้าค่าเฉลี่ยของตัวแปร L, R, F, M, P ที่มีค่าน้อยกว่าค่าเฉลี่ยรวม

Table V.
Clustering results by
K-means method

Cluster	Sample size	Average L	Average R	Average F	Average M	Average P	LRFMP Scores
1	538	633.29	39.67	175.24	24.32	4.99	L \uparrow R \downarrow F \uparrow M \downarrow P \downarrow
2	4,681	564.50	90.19	33.44	31.73	31.49	L \uparrow R \downarrow F \uparrow M \downarrow P \downarrow
3	1,091	482.17	301.18	5.33	34.21	159.32	L \uparrow R \uparrow F \downarrow M \downarrow P \uparrow
4	818	374.01	220.24	11.85	104.18	45.81	L \downarrow R \uparrow F \downarrow M \uparrow P \uparrow
5	3,343	173.70	399.79	10.34	30.14	27.85	L \downarrow R \uparrow F \downarrow M \downarrow P \downarrow
Average		419.81	218.59	28.74	36.76	43.41	

ภาพที่ 2.3 แสดงการแบ่งกลุ่มพฤติกรรมด้วยแบบจำลอง LRFMP

ที่มา: Serhat Peker, Altan Kocyigit and P. Erhan Eren, “LRFMP model for customer segmentation in the grocery retail industry a case”, February (2017)

2.10.7 วิธีการคำนวณค่าตัวแปร P ของแบบจำลอง LRFMP

แบบจำลอง LRFMP มีการนำค่า P มาใช้ในการแบ่งกลุ่มพฤติกรรมของแบบจำลอง LRFM เพื่ออธิบายลักษณะความสม่ำเสมอของลูกค้า การเพิ่มตัวแปร P (Periodicity) เข้าไปในแบบจำลองเพื่อแสดงลักษณะว่าเป็นลูกค้าประจำ โดยกำหนดระยะเวลาเป็นส่วนเบี่ยงเบนมาตรฐานของมาซื้อสินค้า

$$\text{Periodicity} = \text{stdev} (IVT_1, IVT_2, \dots, IVT_{n-1}, IVT_n) \quad (1)$$

IVT หมายถึง ระยะเวลาห่างระหว่างวันที่เข้ามาซื้อสินค้าล่าสุดลบด้วยวันถัดมา ถ้าค่า IVT ต่ำ แสดงว่าลูกค้าเข้ามาซื้อสินค้าเป็นประจำ

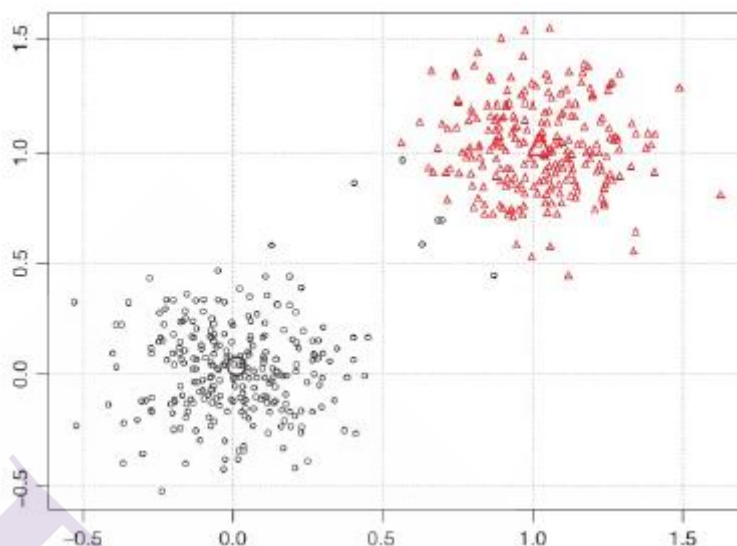
$$IVT_i = \text{date_diff}(t_{i+1}, t_i) \quad (2)$$

2.11 การแบ่งกลุ่มด้วยวิธีเคมีน (K-means)

การแบ่งกลุ่มด้วยวิธีเคมีน เป็นวิธีที่นิยมใช้ในการแบ่งกลุ่มข้อมูลและใช้เปรียบเทียบความคล้ายคลึงของข้อมูลต่างๆ กับจุดศูนย์กลางของแต่ละคลัสเตอร์ (centroid) หรือค่าเฉลี่ย (mean) ภายในคลัสเตอร์นั่นเอง การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีนนี้เป็นการแบ่งกลุ่มแบบ partitional clustering คือ แบ่งกลุ่มข้อมูลออกเป็นส่วนๆ ตามจำนวนกลุ่มที่ระบุ ซึ่งจะแบ่งกลุ่มออกเป็นจำนวน K คลัสเตอร์ตามที่กำหนดขึ้นมา โดยใช้การวัดระยะห่างระหว่างข้อมูลแต่ละตัวกับจุดศูนย์กลาง (centroid) ของแต่ละคลัสเตอร์ซึ่งจุดศูนย์กลางของแต่ละคลัสเตอร์คือค่าเฉลี่ยของแต่ละแอตทริบิวต์ของข้อมูลที่อยู่ในคลัสเตอร์นั้น (เอกสิทธิ์ พ็ชรวงศ์ศักดิ์, 2557, น. 31)

2.11.1 ขั้นตอนการแบ่งกลุ่มด้วยวิธีเคมีน

1. กำหนดจุดศูนย์กลางของแต่ละคลัสเตอร์โดยทำการสุ่ม (random cluster center) ตามจำนวนที่กำหนด K (cluster)
2. คำนวณระยะห่างระหว่างข้อมูลแต่ละตัวกับจุดศูนย์กลางของแต่ละคลัสเตอร์และกำหนดให้ข้อมูลอยู่ในคลัสเตอร์ที่ใกล้กับจุดศูนย์กลางของคลัสเตอร์นั้น
3. คำนวณหาจุดศูนย์กลางของแต่ละคลัสเตอร์ใหม่อีกครั้ง
4. ทำขั้นตอนที่ 2-3 จนกว่าข้อมูลทุกตัวอยู่ในคลัสเตอร์เดิม



ภาพที่ 2.8 การจัดกลุ่มด้วยวิธีเคมีน

ที่มา: Fraud Analytics Using Descriptive Predictive and Social Network Techniques A Guide to Data Science for Fraud Detection, (Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke, 2015, p. 133-149)

2.11.2 การหาระยะห่างระหว่างข้อมูล (Distance Function)

ทางคณิตศาสตร์เรียกความแตกต่างระหว่างข้อมูลนี้ว่า “ระยะห่างระหว่างข้อมูล (distance)” โดยมีฟังก์ชันในการคำนวณหาระยะห่างระหว่างข้อมูลที่ใช้กันอย่างแพร่หลาย ในการศึกษาครั้งนี้ใช้วิธีการหาระยะห่างด้วยวิธี Euclidean Distance

2.11.3 การหาระยะห่างด้วยวิธี Euclidean (Euclidean distance)

การหาระยะห่างด้วยวิธี Euclidean เกิดจากรากที่สองของผลต่างระหว่างแอตทริบิวต์ต่างๆ ยกกำลังสองตั้งสมการ

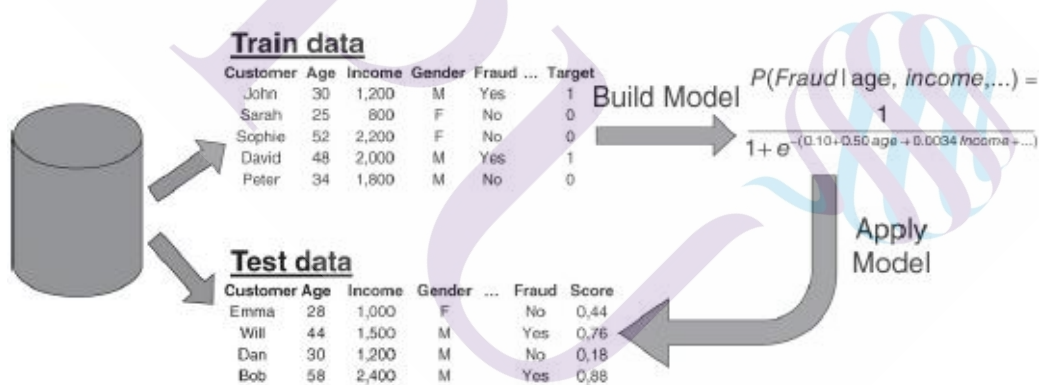
$$D_{Euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_L - y_L)^2}$$

โดยที่ x_1 คือ แอตทริบิวต์ที่ 1 ของข้อมูลจุดที่ 1 และ y_1 คือ แอตทริบิวต์ 1 ของข้อมูลจุดที่ 2 โดยข้อมูลทั้งสองตัว (x และ y) มีจำนวนแอตทริบิวต์เท่ากับ L (เอกสิทธิ์ พัชรวงศ์ศักดิ์, 2557, น. 30)

2.12 การสร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจเพื่อทำนายการหลอกลวง

เทคนิคการจำแนกประเภทข้อมูลเป็นการนำข้อมูลที่มีอยู่ในอดีตมาสอนระบบเพื่อให้เรียนรู้รูปแบบที่เกิดขึ้นในข้อมูลแล้วจึงสร้างเป็นสมการแบบจำลองขึ้นมาเพื่อหาคำตอบให้สำหรับข้อมูลใหม่ คำตอบที่ได้เรียกว่าคลาส (class) หรือลาเบล (label) เรียกว่าเทคนิคการจำแนกประเภทข้อมูล (classification) ด้วยวิธีการเรียนรู้แบบต้นไม้ตัดสินใจ ชั้นแรกจะทำการคัดเลือกแอตทริบิวต์ ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็น โหนด (node) บนสุดของต้นไม้ (root node) หลังจากนั้นทำการหาแอตทริบิวต์ถัดไปเรื่อยๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) การแตกกิ่งของต้นไม้ เกิดจากการเลือกแอตทริบิวต์ ที่มีความสัมพันธ์กับ คลาสมากที่สุด ซึ่งค่า IG ที่มีความสัมพันธ์กับคลาสมากที่สุดจะให้ เป็น โหนดบนสุด หลังจากนั้นแอตทริบิวต์ ที่มีความสัมพันธ์น้อยกว่าจะแตกกิ่งออกไปเรื่อยๆ จนทุกข้อมูลใน โหนด นั้นเป็นคลาสดียวกัน

การแบ่งข้อมูลเพื่อทำการทดสอบใช้วิธี Cross-validation Test ชั้นแรกต้องทำการแบ่งข้อมูล ออกเป็น 2 ชุด คือ ข้อมูลที่ใช้ทดลอง (training data) ร้อยละ 70 และข้อมูลที่ใช้ทดสอบ (testing data) ร้อยละ 30



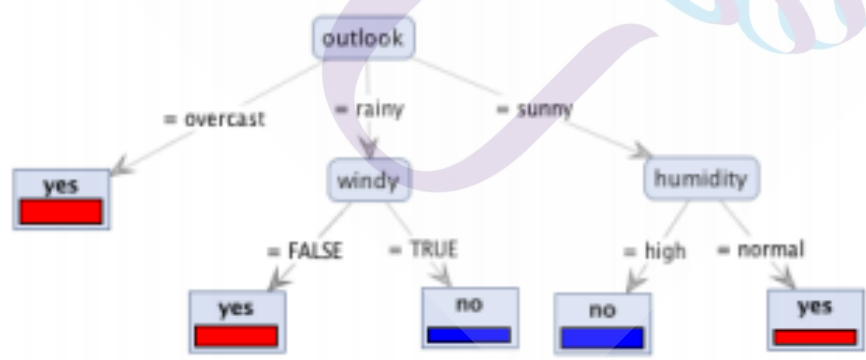
ภาพที่ 2.4 การแบ่งข้อมูลใช้ทดลองและข้อมูลใช้ทดสอบ

ที่มา: Fraud Analytics Using Descriptive Predictive and Social Network Techniques A Guide to Data Science for Fraud Detection, (Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke, 2015, p. 212)

2.12.1 ขั้นตอนการแบ่งกลุ่มด้วยแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจการแบ่งกลุ่มข้อมูลด้วยแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจแบ่งได้ 3 ขั้นตอนดังนี้

1. ขั้นตอนการสร้างแบบจำลอง ขั้นตอนนี้เป็นขั้นตอนแรกในการจำแนกประเภทข้อมูล โดยการนำข้อมูลเทรนนิ่งค่าค่า หรือข้อมูลที่ใช้ในการเรียนรู้ มาสร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ เพื่อให้คอมพิวเตอร์ได้เรียนรู้จากตัวอย่างข้อมูล โดยข้อมูลเทรนนิ่งค่าค่านี้คือ ข้อมูลที่มีในอดีตโดยจะประกอบด้วย 2 ส่วนคือ แอตทริบิวต์ทั่วไปและแอตทริบิวต์ที่เป็นคลาสคำตอบที่สนใจ ตัวอย่างเช่น ข้อมูลที่เก็บสภาพอากาศย้อนหลัง 14 วันเพื่อดูว่าในแต่ละวันจะมีการจัดการแข่งขันเบสบอลขึ้นหรือไม่ ซึ่งแอตทริบิวต์ Outlook, Temperature, Humidity และ Windy จะเป็นแอตทริบิวต์ประเภททั่วไป ที่ใช้ในการพิจารณาว่าถ้าค่าในแอตทริบิวต์เหล่านี้เป็นลักษณะไหนแล้วจึงมีการจัดแข่งเบสบอล ส่วนแอตทริบิวต์ที่เป็นคลาสคำตอบที่สนใจคือ แอตทริบิวต์ Play หลังจากนั้นเทคนิคจะเรียนรู้จากข้อมูลเทรนนิ่งค่าค่าและสร้างเป็นแบบจำลองในรูปแบบต่างๆ ออกมา ดังภาพที่ 2.10 จากแบบจำลองการเรียนรู้แบบต้นไม้ตัดสินใจที่สร้าง สามารถสรุปเป็นกฎได้ เช่น (1) IF Outlook = overcast THEN play = yes (2) IF Outlook = sunny AND Humidity = high THEN play = no

จากกฎข้อที่ 1 แปลความหมายได้ว่าถ้าสภาพอากาศเป็นแบบ overcast จะมีการจัดแข่งขันเบสบอล แต่จากกฎข้อที่ 2 แปลความหมายได้ว่าถ้าสภาพอากาศเป็นแบบ sunny และความชื้นมากแล้ว จะไม่สามารถจัดแข่งเบสบอลได้ (เอกสิทธิ์ พัทธรงค์ศักดิ์ดา, 2557, น. 57)



ภาพที่ 2.5 แบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ

ที่มา: An Introduction to Data Mining Techniques, (เอกสิทธิ์ พัทธรงค์ศักดิ์ดา, 2557, น. 27)

2. ขั้นตอนการทดสอบประสิทธิภาพของแบบจำลอง หลังจากที่เราสร้างแบบจำลองขึ้นมาได้แล้ว ขั้นตอนถัดมาจะต้องทำการวัดประสิทธิภาพของแบบจำลองที่เราสร้างได้ โดยมีตัววัดประสิทธิภาพของแบบจำลองจากตาราง confusion matrix เช่น ค่า precision, ค่า recall, ค่า f-measure และ ค่า accuracy เป็นต้น โดยค่า accuracy เป็นการวัดความถูกต้องของแบบจำลอง พิจารณารวมทุกคลาส คือ จำนวน True Positive ของทุกคลาสรวมกัน

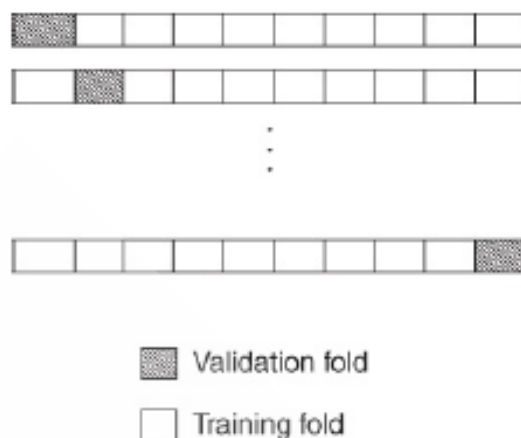
3. ขั้นตอนนำแบบจำลองไปใช้งานเพื่อทำนายข้อมูลใหม่ หลังจากที่เราวัดประสิทธิภาพของแบบจำลองได้ผลเป็นที่น่าเชื่อถือหรือพอใจแล้ว เราสามารถนำแบบจำลองที่เราสร้างได้ นี้ไปใช้ในการทำนายข้อมูลที่เราเข้ามาใหม่ เช่น ถ้าพบว่าสภาพอากาศในวันปัจจุบัน ทำนายได้ว่าจะมีการจัดแข่งเบสบอลขึ้น เนื่องจากแอตทริบิวต์ Outlook มีค่าเป็น sunny และ Humidity เป็น normal แล้ว Play จะเป็น yes

No.	Outlook	Temperature	Humidity	Windy	Play
15	sunny	hot	normal	false	?

ภาพที่ 2.6 ผลการทำนายของแบบจำลอง

ที่มา: An Introduction to Data Mining Techniques, (เอกลีทรี พัทธวงศ์ศักดิ์, 2557, น. 28)

2.12.2 การวัดประสิทธิภาพของแบบจำลองการจำแนกประเภทข้อมูล การทดสอบด้วย Cross-validation วิธีนี้เป็นวิธีที่นิยมใช้ในการทดสอบประสิทธิภาพของแบบจำลอง เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธีการตรวจสอบไขว้กัน (Cross-validation) เป็นการสุ่มตัวอย่างโดยเริ่มจากแบ่งชุดข้อมูลออกเป็น ส่วนๆ และนำบางส่วนจากชุดข้อมูลนั้นมาตรวจสอบผลลัพธ์จากการทำ ในกรณีการทำ K-fold cross-validation จะแบ่งข้อมูลออกเป็น K ส่วนเท่าๆ กัน เช่น 10 ส่วน (10-fold cross-validation) จะทำการคำนวณค่าความผิดพลาด 10 รอบ โดยแต่ละรอบการคำนวณข้อมูลชุดหนึ่งจากข้อมูล 10 ส่วน จะถูกเลือกออกมาเพื่อเป็นข้อมูลทดสอบ และข้อมูลอีก 9 ส่วนจะถูกใช้เป็นข้อมูลสำหรับการเรียนรู้ ทำวนไปจนครบจำนวนที่แบ่งไว้ ดังภาพที่ 2.12



ภาพที่ 2.7 การวัดประสิทธิภาพ Cross-Validation

ที่มา: Fraud Analytics Using Descriptive Predictive and Social Network Techniques A Guide to Data Science for Fraud Detection, (Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke, 2015, p. 213)

2.13 บทความและงานวิจัยที่เกี่ยวข้อง

Serhat Peker, Altan Kocyigit and P. Erhan Eren (2017) ได้ศึกษาเรื่อง LRFMP model for customer segmentation in the grocery retail industry: a case study โดยพัฒนาแบบจำลอง RFM ในรูปแบบใหม่เรียกว่า LRFMP ประกอบด้วย Length , Recency, Frequency, Monetary, Periodicity มาประยุกต์ใช้ในการจัดกลุ่มลูกค้าวงการค้าปลีกร้านขายของชำในประเทศตุรกี ใช้เทคนิคดาต้าไมนึ่ง การแบ่งกลุ่มลูกค้าด้วยเทคนิคเคมีน ผลการวิจัยสามารถแบ่งกลุ่มลูกค้าได้ 5 กลุ่ม คือ ลูกค้าประจำแฟนพันธุ์แท้ ลูกค้าประจำ ลูกค้าที่ไม่แน่นอน ลูกค้ารายได้สูงและลูกค้ารายได้น้อย จากความแตกต่างในพฤติกรรมการซื้อสินค้าตามแบบจำลอง LRFMP ทำให้กำหนดกลยุทธ์ทางการตลาดได้

Bijan Geraili, Mahdi Nasiri, Mohammad Arab (2015) ได้ศึกษาเรื่อง Improving Fraud and Abuse Detection in General Physician Claim A Data Mining Study โดยศึกษาตัวชี้วัด 13 ตัว ของสิทธิการรักษาประกันสังคม มีสถานพยาบาลในเครือ 70 แห่ง คลินิก 280 แห่ง เพื่อวัดประสิทธิภาพในการตรวจสอบแพทย์ที่ต้องสงสัยว่าทุจริตร้อยละ 98 และแพทย์จ่ายยาไม่สมเหตุผล (abuse) ร้อยละ 85 วิธีการวิจัยใช้การทำเหมืองข้อมูล โดยใช้ข้อมูลที่มีการลาเบลและข้อมูลที่ไม่มีลาเบล แล้วนำไปสร้าง

แบบจำลอง แบ่งเป็น 2 กลุ่ม 1) พฤติกรรมการจ่ายยาของแพทย์ที่ไม่เหมาะสมหรือเกินความจำเป็น วิธีนี้อาจจะเป็นการสร้างรายได้ทางอ้อม หรืออาจจะส่งจ่ายที่แพทย์ ได้รับผลประโยชน์จากหน่วยงานภายนอก 2) พฤติกรรมที่แพทย์ไม่เขียนรายการยาในใบสั่งยา แต่ให้คนไข้ไปซื้อจากร้านขายยาข้างนอก แล้วนำใบเสร็จไปเบิกกับหน่วยงานที่ทำงาน วิธีการนี้อาจจะมีการเขียนใบสั่งยาปลอม แล้วนำใบสั่งยาไปซื้อจากร้านขายยาที่สมรู้ร่วมคิด และนำใบไปเบิกจากหน่วยงานหรือบริษัทประกัน สุดท้ายคนไข้และแพทย์นำเงินที่ได้มาแบ่งกัน ผลการวิจัยพบว่า มีตัวชี้วัดที่มีความสัมพันธ์กับเรื่องค่าใช้จ่าย คือ ความถี่มารับบริการ ซึ่งผู้ป่วยไปรับบริการกับแพทย์มากกว่า 1 ครั้งในช่วงเวลาสั้น และแพทย์ได้เขียนใบสั่งยาให้ไปซื้อจากร้านขายยาข้างนอก จากนั้นก็ทำการรักษารายการยาในใบสั่งยาร้อยละ 30 ให้ไปซื้อจากร้านขายยาข้างนอกสถานพยาบาล ทั้งนี้แพทย์ ร้อยละ 54 มีพฤติกรรมไม่เหมาะสม และร้อยละ 2 สงสัยว่าทุจริต

Dallas Thornton, Guido van Capelleveen, Mannes Poel, Jos van Hillegersberg and Roland M. Mueller (2014) ได้ศึกษาเรื่อง Outlier-based Health Insurance Fraud Detection for U.S. Medicaid โดยใช้อัลกอริทึมการทำเหมืองข้อมูล ในการตรวจสอบการทุจริตด้วยเทคนิคการเรียนรู้แบบไม่มีผู้สอน เพื่อพยากรณ์การทุจริต ข้อมูลการเบิกค่ารักษาพยาบาลด้านทันตกรรม โดยใช้ข้อมูล 2012 -2013 จำนวน 11 เดือน พบว่ามีคนไข้ส่วนน้อยที่ทำการทุจริต แต่ที่ต้องสงสัยคือ สถานพยาบาล ที่มีผู้ป่วยมารับบริการมากกว่า 300 คนต่อสัปดาห์ อาจต้องสงสัยว่าทำทุจริต ดังเช่น มีการใส่ข้อมูลปลอม หรืออาจจะเป็นการให้บริการทันตกรรมเคลื่อนที่ วิธีการตรวจสอบการทุจริต จะทำนายผู้ให้บริการที่มีค่าผิดปกติจากหลายตัวชี้วัด 14 ตัว วิเคราะห์ข้อมูลด้วยโปรแกรม R กรณีที่มีการเบิกเงินค่ารักษาพยาบาล 10,000 ดอลลาร์ต่อเดือน และการวิเคราะห์ด้วยเทคนิคการแบ่งกลุ่มเคมีน เพื่อดูค่าที่ผิดปกติจากข้อมูลผู้ให้บริการจำนวน 500 แห่ง ผลการวิจัยพบว่าผู้ให้บริการร้อยละ 71 มีพฤติกรรมที่น่าสงสัยและผิดปกติ จากหลักฐานที่มีทำให้สามารถเข้าตรวจสอบได้ตามกฎหมาย

Hsin-Hung Wu, Shih-Yen Lin, and Chih-Wei Liu (2014) ได้ศึกษาเรื่อง Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan โดยศึกษาแบบจำลอง LRFM Length, Recency, Frequency, Monetary ในคลินิกทันตกรรมเด็กของประเทศไต้หวัน ใช้เทคนิค SOM และเทคนิคเคมีนในการแบ่งกลุ่ม 12 กลุ่ม ด้วยค่าเฉลี่ยของ L, R, F, M และใช้เมทริกซ์พิจารณาความสัมพันธ์จากระยะเวลาและผู้ป่วยใหม่ ผลที่ได้คือสามารถแบ่งกลุ่มผู้ป่วยได้ 3 กลุ่ม คือกลุ่มผู้ป่วยที่เป็นคนไข้ประจำแฟนพันธ์แท้ ได้จากค่า L, R, F มีค่ามากกว่าค่าเฉลี่ยรวม กลุ่มที่เป็นผู้ป่วยเก่า จากค่า L มีค่าต่ำมากและอาจจะกลายเป็นผู้ป่วยประจำ แฟนพันธ์แท้ได้ใน

อนาคต เนื่องจากค่า L, R มีค่าต่ำ จากผลการวิเคราะห์ที่ได้ สามารถนำมาวางแผนกลยุทธ์การตลาดเพื่อ ออกแบบความต้องการของผู้ป่วยที่แตกต่างกันไปและพัฒนาผู้ป่วยให้กลายเป็นลูกค้าประจำในอนาคต

Qi Liu, Miklos Vasarhelyi (2013) ได้ศึกษาเรื่อง Healthcare fraud detection A survey and a clustering model incorporating Geo-location information ของประเทศอเมริกาในปีงบประมาณ 2010 วัตถุประสงค์เพื่อตรวจสอบการทุจริตด้านการเบิกค่ารักษาพยาบาล จากการศึกษาสามารถช่วยลด ค่าใช้จ่ายด้านสุขภาพได้มากขึ้น โดยให้ความสนใจในการป้อนข้อมูลที่ผิดพลาดหรือไม่สมบูรณ์ในการ เบิกเงินค่ารักษาพยาบาลและการให้บริการทางการแพทย์ที่ไม่ซ้ำซ้อน วิธีการวิจัยใช้ข้อมูลประกันสังคม โดยแบ่งกลุ่มผู้มารับบริการเป็น 3 ประเภท 1) อายุมากกว่า 65 ปี 2) อายุน้อยกว่า 65 ปีและคนพิการ 3) ผู้ป่วยโรคไตวายเรื้อรังระยะสุดท้าย และแบ่งกลุ่มผู้ให้บริการสถานพยาบาลเป็น 3 ประเภท 1) Hospital insurance 2) Medical insurance 3) Prescription drug coverage (เฉพาะผู้มีรายได้น้อย) ได้แบ่งการ ทุจริตเป็น 4 ประเภท 1) การทุจริตของผู้มารับบริการ 2) การทุจริตสมาชิกประกันภัย 3) การทุจริตของผู้ ให้บริการประกันภัย 4) การสมรู้ร่วมคิดทุจริตที่เกี่ยวข้องกันมากกว่าหนึ่งฝ่าย วิธีศึกษาคือการนำ เทคนิคเหมืองข้อมูลมาใช้ในการตรวจสอบทุจริตกับข้อมูลที่มีเวลาเบล ซึ่งใช้เทคนิคโครงข่ายประสาท เทียมทำการตรวจสอบ ผลที่ได้พบว่ามีการทุจริตเงินจำนวน 47.9 พันล้าน ดอลลาร์ของค่าใช้จ่าย ประกันสุขภาพภาครัฐ อย่างไรก็ตามงานวิจัยนี้สามารถพัฒนาวิธีการตรวจสอบการทุจริตเพื่อป้องกัน ไม่ได้เกิดขึ้นได้ในอนาคต

Karca Duru Aral a, Halil Altay Güvenir b, I hsan Sabuncuog luc, Ahmet Ruchan Akard,e (2012) ได้ศึกษาเรื่อง A prescription fraud detection model วัตถุประสงค์ของงานวิจัยเพื่อพัฒนา แบบจำลองสำหรับตรวจสอบการทุจริตใบสั่งยา ที่เป็นปัญหาใหญ่ทำให้สูญเสียเงินจำนวนมาก วิธีการ วิจัยโดยการสุ่มตัวอย่างใบสั่งยาแพทย์ และใช้เทคนิคเหมืองข้อมูลกับข้อมูลที่มีการเวลาเบลและเทคนิค โครงข่ายประสาทเทียม (neural network) เพื่อวิเคราะห์การทุจริตใบสั่งยา ซึ่งสามารถช่วยลดค่าใช้จ่าย ในการจ้างผู้เชี่ยวชาญมาตรวจสอบ โดยทดสอบกับฐานข้อมูลการผ่าตัดหัวใจในผู้ใหญ่ พบว่ามีการ ทุจริตร่วมกันระหว่างบริษัทยา แพทย์ ร้านขายยาและผู้ป่วย การนำเทคนิคเหมืองข้อมูลมาใช้วิเคราะห์ ผล ทำให้สามารถตรวจสอบข้อมูลได้อย่างมีประสิทธิภาพมากขึ้นและลดค่าใช้จ่ายจากการจ้าง ผู้เชี่ยวชาญมาตรวจสอบเอกสาร

บทที่ 3

ระเบียบวิธีวิจัย

3.1 ประชากร และตัวอย่าง

การศึกษาระบบการทำเหมืองข้อมูล (Data Mining) เพื่อตรวจสอบพฤติกรรมการใช้บริการไม่สมเหตุสมผล จากฐานข้อมูลการเบิกจ่ายเงินค่ายาน่ากลับไปใช้ที่บ้าน ของผู้ป่วยนอกสิทธิรักษาพยาบาลสวัสดิข้าราชการในระบบเบิกจ่ายตรง ปีงบประมาณ 2556 โดยใช้ข้อมูลจากสถานพยาบาล จำนวน 168 แห่ง จำนวนผู้ใช้บริการ 2,751,425 คน มีความถี่ในการไปใช้บริการ จำนวน 24,125,143 ครั้ง จำนวนคนที่มีการเบิกจ่ายเงินค่ายากลับไปใช้ที่บ้าน 525,085 คน การศึกษาใช้ข้อมูลตาราง Master (รายละเอียดการเบิกจ่ายเงินค่ารักษาพยาบาลรายคน) เชื่อมกับกับตาราง Bill (รายการเบิกจ่ายเงินค่ารักษาพยาบาลรายหมวดรายคน) แล้วเลือกเฉพาะหมวด 4 (ค่ายาน่ากลับไปใช้ที่บ้าน) ที่มีค่ายากลับบ้านมากกว่า 30,000 บาทต่อปี และตัดคนที่ใช้ยามะเร็งออก ทำให้เหลือข้อมูลที่ใช้ในการวิเคราะห์จำนวน 31,001 คน แล้วนำมาแบ่งกลุ่ม ตัวแปรที่ใช้แบ่งลักษณะรูปแบบพฤติกรรมมีทั้งหมด 6 ตัว ได้แก่ R, F, M, H, P, E คือ (1) จำนวนวันล่าสุดที่ใช้บริการ (2) ความถี่ที่ใช้บริการ (3) จำนวนเงินที่เบิกจ่ายเงินค่ายากลับบ้าน (4) จำนวนสถานพยาบาลที่ใช้บริการ (5) ช่วงห่างระหว่างวันที่ไปใช้บริการ (6) ค่ายาเฉลี่ยต่อครั้ง ซึ่งวิธีการตรวจสอบแบบเดิมคือสุ่มตรวจเลือกคนที่มีความถี่ไปใช้บริการสูง ไปหลายสถานพยาบาลและมีค่ายาเฉลี่ยต่อครั้งสูง ซึ่งวิธีการตรวจสอบแบบเดิมใช้เวลานานและมีความยากลำบากในการตรวจสอบฐานข้อมูลขนาดใหญ่ เพื่อค้นหาบุคคลส่วนน้อยที่ซ่อนอยู่ในฐานข้อมูล

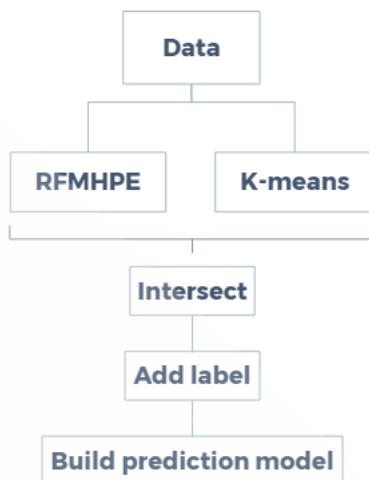
3.2 เครื่องมือที่ใช้ในการรวบรวมข้อมูล

เครื่องมือที่ใช้ในประมวลผลและการจัดการฐานข้อมูล คือ คอมพิวเตอร์ระบบปฏิบัติการ วินโดว์ 10, CPU Core i7, หน่วยความจำ (RAM) 40 กิกะไบต์ ฮาร์ดดิสก์ (SSD hard disk) 256 กิกะไบต์ และฮาร์ดดิสก์ (HHD hard disk) 2 เทระไบต์ ใช้โปรแกรม STATA เวอร์ชัน 11.2 ทำความสะอาดข้อมูล (Cleaning data) และวิเคราะห์รายการที่ผิดปกติ ใช้โปรแกรม R เวอร์ชัน 3.4.3 การจัดกลุ่มวิธีเชกเมนต์เทชั่น (Cluster Segmentation) การแบ่งกลุ่มวิธีเคมีน (Partitional Clustering) และโปรแกรม Rapid Miner studio เวอร์ชัน 7.6.003 สร้างโมเดลการเรียนรู้โดยการจำแนกประเภท (Classification) เพื่อทำนายหมวดหมู่ของข้อมูลเป็นกลุ่ม (Class)

3.3 ขั้นตอนดาต้าไมนนิ่ง Framework

ขั้นตอนการแบ่งกลุ่มพฤติกรรมความคล้ายคลึงของผู้ไปใช้บริการด้วยเทคนิคดาต้าไมนนิ่ง มีขั้นตอนดังนี้ ดังภาพที่ 3.1

1. ขั้นตอนการทำความสะอาดข้อมูล นำฐานข้อมูลผู้ป่วยนอกปีงบประมาณ 2556 ที่ผ่านการทำความสะอาดแล้ว เชื่อมตาราง Mater กับตาราง Bill เพื่อให้ได้ข้อมูลรายบุคคลหมวดค่าใช้ยานำกลับไปใช้ที่บ้าน (หมวด 4)
2. ขั้นตอนการแบ่งกลุ่มข้อมูลด้วยวิธีเชกเมนต์เทชั่น แยกข้อมูลเป็น 3 กลุ่ม และแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน โดยการหาค่าจุดแบ่งกลุ่มที่เหมาะสม (Best K) ที่ได้จากจุด Elbow point
3. ขั้นตอนการทำการอินเตอร์เซกชัน (intersection data) กลุ่มที่เลือกไว้จากการแบ่งกลุ่มด้วยวิธีเชกเมนต์เทชั่น และแบ่งกลุ่มด้วยวิธีเคมีน
4. ขั้นตอนการสร้างลาเบลเพื่อระบุประเภท (Type)
5. ขั้นตอนการทดสอบแบบจำลองการเรียนรู้แบบต้นไม้ตัดสินใจ (Decision Tree) เพื่อทำนายหมวดหมู่ของกลุ่มข้อมูล (Class) และดูความแม่นยำในการทำนาย



ภาพที่ 3.1 ขั้นตอนดาต้าไมนนิ่ง Framework

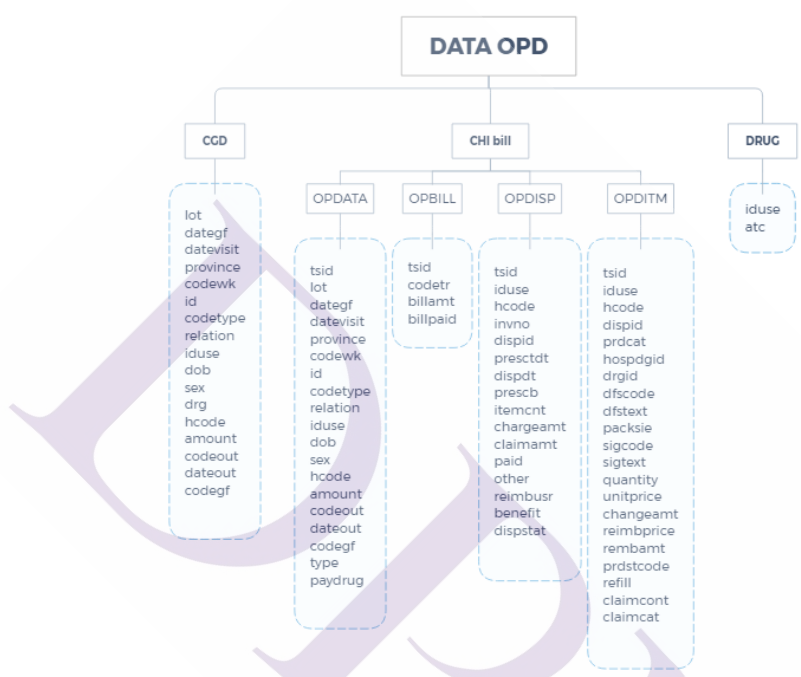
3.3.1 ขั้นตอนการทำความสะอาดข้อมูล

ก่อนนำข้อมูลมาใช้ในการศึกษา ต้องทำความเข้าใจฐานข้อมูลและตรวจสอบความสมบูรณ์ของข้อมูล รวมถึงทราบข้อจำกัดของข้อมูล ซึ่งฐานข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลผู้ป่วยนอก สวัสดิการข้าราชการ ระบบเบิกจ่ายตรง ปีงบประมาณ 2556 ข้อมูลมาจากกรมบัญชีกลาง CGD (The Comptroller General's Department) ทั้งหมด 2 ฐาน ได้แก่ 1) ข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลรายบุคคล 2) ข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลรายหมวดแสดงรายการยา (CHI Bill) ประกอบด้วยสถานพยาบาลที่ส่งรายการเรียกเก็บค่าใช้จ่ายจำนวน 168 แห่ง ฐานข้อมูล CGD เป็นข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลรายบุคคล ซึ่งไม่มีรายการหมวดค่ารักษาและข้อมูลยา ประกอบด้วยฐานข้อมูลผู้ป่วยนอก OPD (Out Patient Department) ซึ่งทุกรายการใน ได้ผ่านการอนุมัติจ่ายเงินผ่านระบบ GFMS (Government Fiscal Management Information System) จำนวนสถานพยาบาลที่ส่งรายการเรียกเก็บค่าใช้จ่ายจำนวน 1,380 แห่ง

3.3.1.1 โครงสร้างฐานข้อมูล

จากภาพที่ 3.2 แสดงโครงสร้างข้อมูล ประกอบด้วยฐานข้อมูล CGD, CHI Bill และฐานข้อมูลยา DRUG ซึ่งฐานข้อมูล CGD เป็นข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลรายบุคคล ที่มีการ

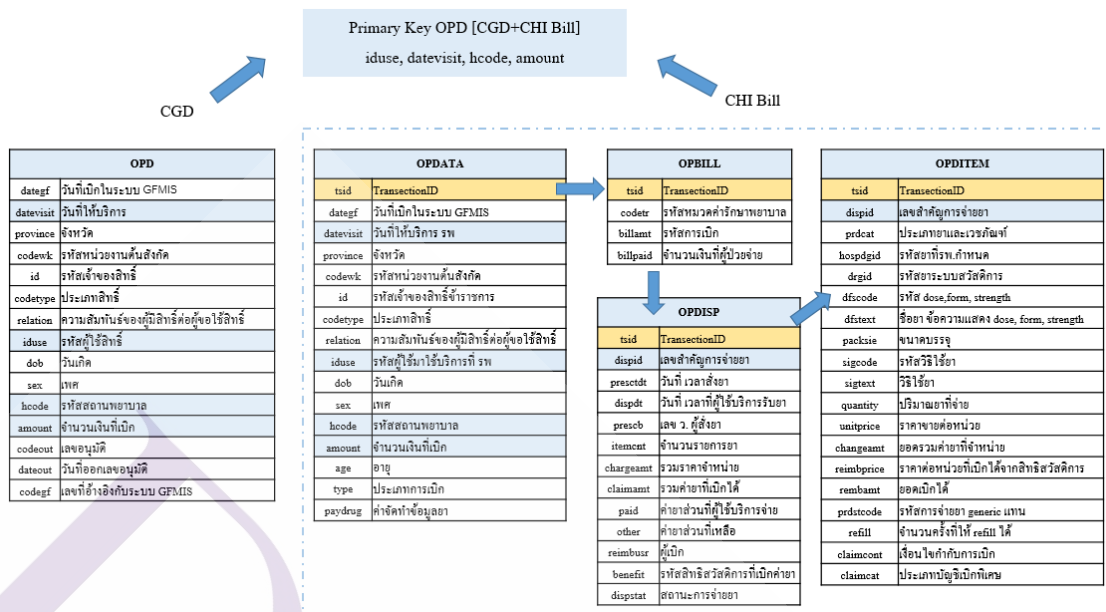
เบิกจ่ายเงินเรียบร้อยแล้ว จำนวน 1,380 แห่ง ส่วนฐานข้อมูล CHI Bill เป็นข้อมูลเบิกจ่ายค่ารักษาพยาบาลรายบุคคลมีรายละเอียดการเบิกค่ารักษาพยาบาลรายหมวดการรักษาที่รอการอนุมัติ จำนวน 168 แห่ง และข้อมูลยา DRUG แบ่งตามหมวดยา



ภาพที่ 3.2 โครงสร้างข้อมูลผู้ป่วยนอก

3.3.1.2 การเชื่อมฐานข้อมูลผู้ป่วยนอกระบบเบิกจ่ายตรง

การเชื่อมฐานข้อมูลผู้ป่วยนอกระบบเบิกจ่ายตรง มีวัตถุประสงค์เพื่อเชื่อมข้อมูลค่าใช้จ่ายรายหมวดและหมวดยาที่ใช้ในการรักษาจากฐานข้อมูล CGD, CHI Bill เข้าด้วยกัน โดยโครงสร้างข้อมูล CHI Bill ประกอบด้วยรายละเอียด 1) ข้อมูลการเบิกจ่ายเงินรายบุคคล (OPDDATA) 2) ข้อมูลการเบิกจ่ายเงินรายหมวด (OPBILL) 3) ข้อมูลใบสั่งยา (OPDISP) 4) ข้อมูลรายการยา (OPDITEM) ตัวแปรที่ใช้ในการเชื่อมระหว่างตารางได้แก่ iduse, datevisit, hcode, amount ดังภาพที่ 3.3

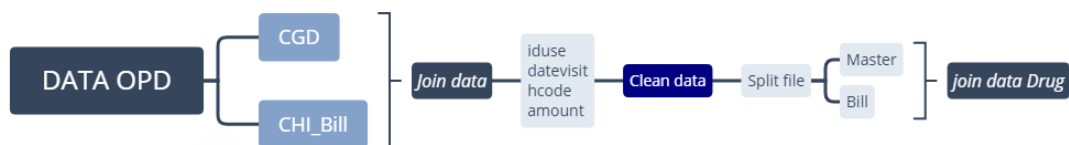


ภาพที่ 3.3 การเชื่อมข้อมูลผู้ป่วยนอกระบบเบิกจ่ายตรง

3.3.1.3 การแบ่งข้อมูลเพื่อใช้ในการศึกษา

การจัดการฐานข้อมูลที่ผ่านกระบวนการแปลงไฟล์และเชื่อม (Join) ข้อมูล 2 ฐานเข้าด้วยกัน ขึ้นต่อไปคือการทำความสะอาดข้อมูล (clean data) ตรวจสอบความถูกต้องของข้อมูล ดังต่อไปนี้ (1) ตรวจสอบความซ้ำซ้อนของข้อมูล (2) ตรวจสอบจำนวนการเบิกจ่ายเงินคำรักษาพยาบาลรายบุคคล (3) ตรวจสอบวันที่ไปใช้บริการ (datevisit) (4) ตรวจสอบความถี่การไปใช้บริการ (5) การเบิกจ่ายเงินของสถานพยาบาล (6) ตรวจสอบความสมบูรณ์ของแต่ละตัวแปร โดยใช้ลาเบล -8 กรณีที่ข้อมูลสูญหาย (missing value) แล้วแตกไฟล์ (split) ออกเป็น 2 ไฟล์ ดังภาพที่ 3.4

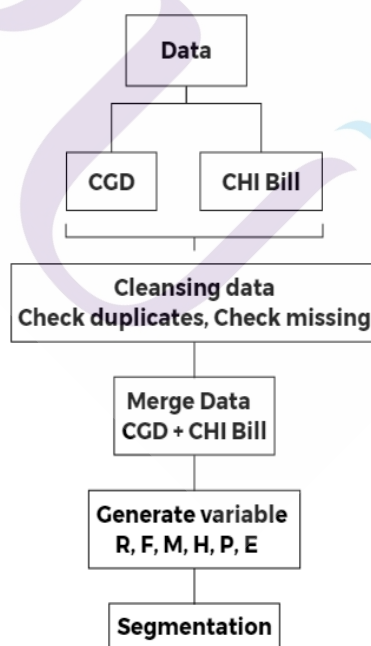
1. Master รายละเอียดการเบิกจ่ายเงินคำรักษาพยาบาลรายคน
2. Bill รายการเบิกจ่ายเงินคำรักษาพยาบาลรายหมวดรายคน



ภาพที่ 3.4 ขั้นตอนการแบ่งฐานข้อมูลผู้ป่วยนอก

3.3.1.4 ขั้นตอนการจัดการฐานข้อมูล

1. ฐานข้อมูล CGD, CHI Bill การเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้านของผู้ป่วยนอก ปีงบประมาณ 2556
2. ทำความสะอาดข้อมูล ตรวจสอบการเบิกจ่ายเงินซ้ำซ้อนและตรวจสอบข้อมูลที่ผิดพลาด
3. เชื่อมฐานข้อมูล CGD และฐาน CHI_Bill ด้วย Primary key เลือกเฉพาะหมวด 4 (จำนวนเงินเบิกจ่ายเงินค่ายานำกลับไปใช้ต่อที่บ้าน)
4. สร้างตัวแปร R, F, M, H, P, E เพื่อนำไปแบ่งกลุ่มต่อไป



ภาพที่ 3.5 ขั้นตอนการจัดการฐานข้อมูล

3.3.2 ขั้นตอนการแบ่งกลุ่มข้อมูล

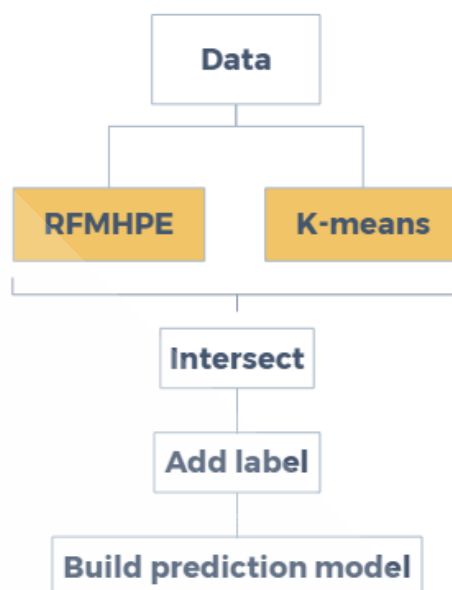
จากตารางที่ 3.1 ก่อนทำการแบ่งกลุ่มข้อมูล ตัวแปรที่นำมาใช้ในการแบ่งกลุ่มคือ R, F, M, H, P, E โดยตัดข้อมูลที่มีการเบิกจ่ายเงินค่ายากลับบ้านมากกว่า 30,000 บาท/ปี และไม่ใช่ผู้ป่วยที่ใช้รหัสมาตรฐานยา The Anatomical Therapeutic Chemical classification system (ATC) กลุ่มยามะเร็งรหัส L

ตารางที่ 3.1 ตัวแปรที่ใช้แบ่งกลุ่ม

ตัวแปร	ความหมาย
R	จำนวนวันล่าสุดที่ไปใช้บริการ/ปี
F	ความถี่ที่ไปใช้บริการ/ปี
M	ค่ายากลับบ้าน/ปี
H	จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี
P	ช่วงห่างระหว่างวันที่ไปใช้บริการ/ปี
E	ค่ายาเฉลี่ยต่อครั้ง

3.3.2.1 การแบ่งข้อมูล

1. แบ่งกลุ่มข้อมูลด้วยวิธีเซกเมนต์เทชั่น แยกข้อมูลเป็น 3 กลุ่ม กลุ่มละ 33.5% แล้วเลือกกลุ่มที่น่าสนใจที่มีค่ารูปแบบ F, H, E
2. แบ่งกลุ่มข้อมูลด้วยวิธีเคมีน โดยการหาค่าจุดแบ่งกลุ่มที่เหมาะสม (Best K) ที่ได้จากจุด Elbow point แล้วเลือกกลุ่มที่น่าสนใจที่มีค่ารูปแบบ F, H, E
3. นำข้อมูลที่ได้แต่ละเทคนิค ไปอินเตอร์เซก ดังภาพที่ 3.6



ภาพที่ 3.6 การแบ่งข้อมูลออกเป็น 2 วิธี

3.3.2.2 การแบ่งกลุ่มด้วยวิธีเซกเมนต์เทชั่น

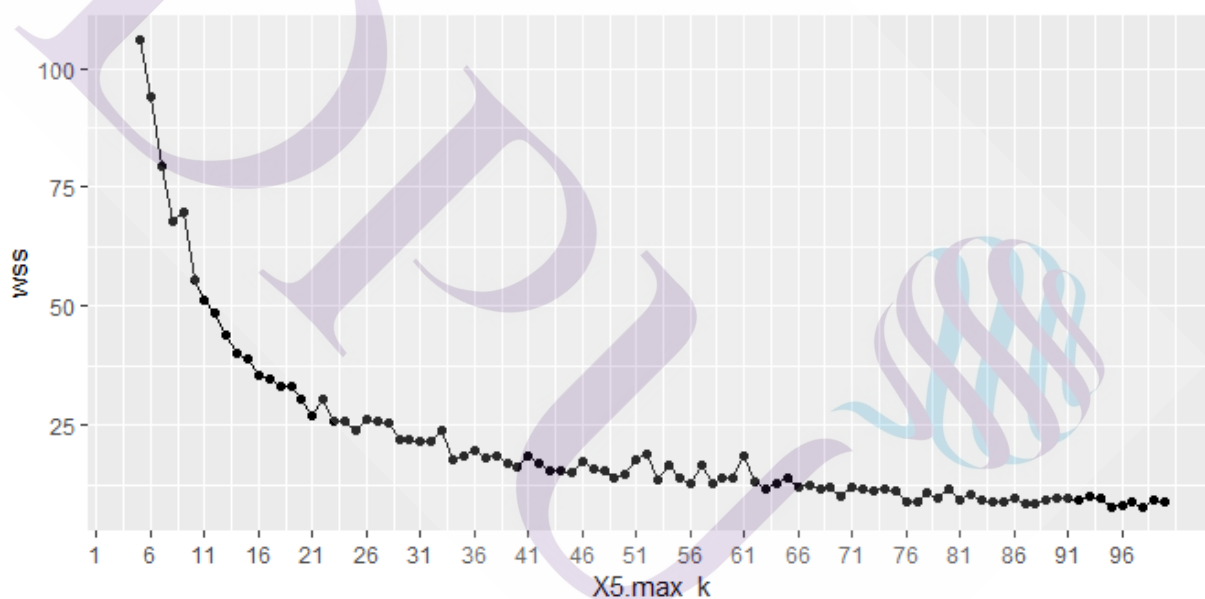
วิธีการแบ่งกลุ่มวิธีเซกเมนต์เทชั่น เป็นการจัดกลุ่มบุคคลที่มีความคล้ายคลึงกัน ตามพฤติกรรมการไปใช้บริการที่สถานพยาบาล ทำการเรียงข้อมูลจากน้อยไปมากและแบ่งเป็นกลุ่มละ 33.5% เท่าๆกัน ทั้ง 6 ตัวแปร ได้ค่า 1-3 โดยค่าคะแนน 3 เป็นค่าที่มากที่สุด ใช้ค่าตัวเลข 3 หลักเป็นตัวแทนของแต่ละกลุ่ม เช่น คะแนน 333 คือ กลุ่มที่มีค่า R=3, F=3 และ M=3

3.3.2.3 การแบ่งกลุ่มด้วยวิธีเคมีน (K-means)

การแบ่งกลุ่มด้วยวิธีเคมีน สามารถแบ่งกลุ่มข้อมูลที่มีความคล้ายกันให้อยู่ในกลุ่มเดียวกัน โดยไม่ทราบจำนวนกลุ่มและไม่ได้กำหนดคลาสเบลของข้อมูลไว้ก่อน ซึ่งเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised classification) คือไม่มีการระบุคลาสเบลคำตอบไว้ และสามารถจัดกลุ่มกันตามระยะความห่างระหว่างจุด วิธีที่ใช้กันคือระยะทางแบบยูคลิด (Euclidean distance) การทำงานของวิธีเคมีนจะตัดแบ่งพาร์ติชัน (Partition) ออกเป็น K กลุ่ม และแทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ย (Mean) ของกลุ่ม ใช้เป็นจุดศูนย์กลาง (Centroid) ของกลุ่มเป็นตัววัดระยะห่างของข้อมูลในกลุ่ม

3.3.2.4 การค้นหาค่า K ที่ดีที่สุดด้วยเทคนิค Elbow method

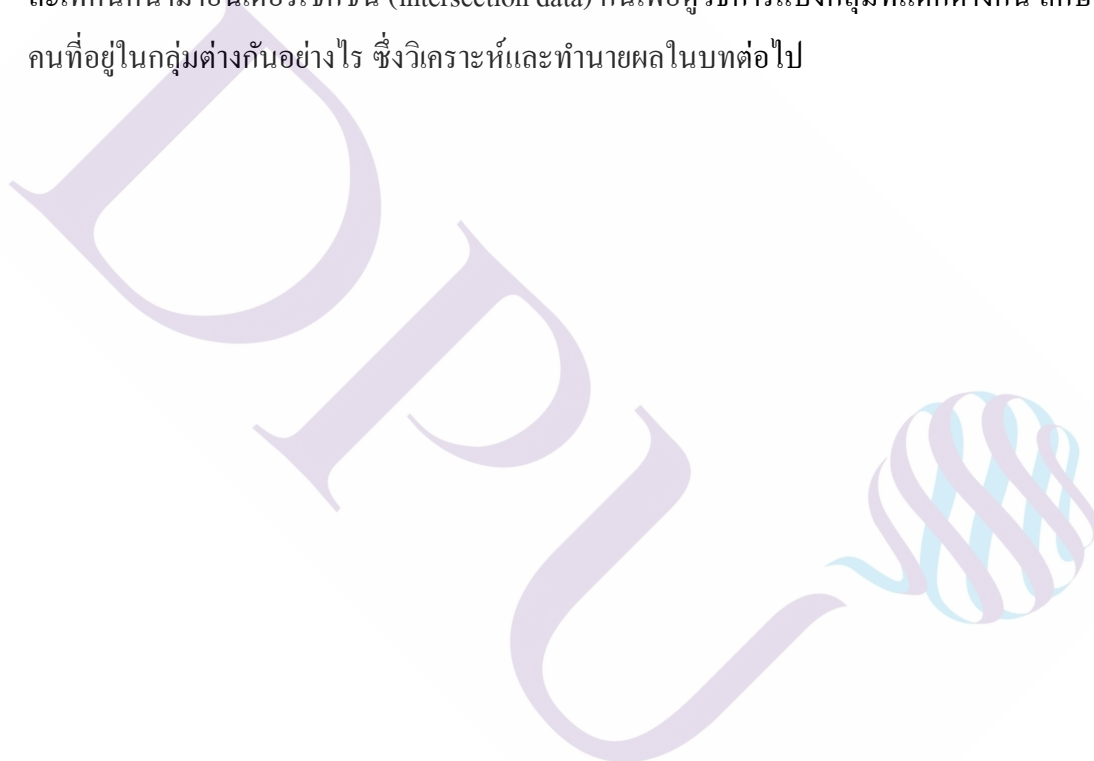
การค้นหาค่า K ที่เหมาะสมที่สุดในการแบ่งกลุ่มเคมีน ทำการเขียนสคริปคำสั่งโปรแกรม R เพื่อค้นหาค่า K ที่ดีที่สุดในการแบ่งกลุ่ม โดยใช้เทคนิค Elbow method เพื่อกำหนดค่า K ที่ดีที่สุดในการแบ่งกลุ่ม ทำซ้ำวน 200 รอบ กับกลุ่มจำนวนที่แตกต่างกันไปในแต่ละรอบ ได้ค่าเฉลี่ยของระยะห่างของ Centroids ของแต่ละ K จากนั้นเลือก K ที่มีลักษณะคล้ายคลึงกันและมีระยะห่างระหว่างจุดใกล้เคียงกัน เทคนิค Elbow method สามารถคำนวณค่า K ที่เหมาะสมที่สุด ได้ค่า Elbow point พบว่าค่า K ที่ดีที่สุดของการแบ่งกลุ่มเคมีน เท่ากับ 39 กลุ่ม ถึงแม้ว่าจะเพิ่มจำนวน K ให้มากกว่า 39 กลุ่ม ก็ไม่มีผลต่อการแบ่งกลุ่ม สังเกตจากกราฟลาดลง



ภาพที่ 3.7 แสดงค่าที่ดีที่สุดของการแบ่งกลุ่มเคมีนเท่ากับ 39 กลุ่ม

3.3.3 ความแตกต่างระหว่างการแบ่งกลุ่มวิธีเซกเมนต์เทียบกับวิธีเคมีน

จากเทคนิคการแบ่งกลุ่มที่แตกต่างกัน การแบ่งกลุ่มวิธีเซกเมนต์เทียบกับวิธีเคมีน เป็นการแบ่งกลุ่มผู้ไปใช้บริการเป็น 3 ช่วงเท่าๆกัน คิดเป็นร้อยละ 33.5 โดยเรียงข้อมูลก่อนแล้วแบ่งเป็นกลุ่ม โดยกลุ่มที่มีค่าคะแนน Score สูงจะเป็นกลุ่มที่มีความสำคัญ ส่วนการแบ่งด้วยวิธีเคมีน จะทำการแบ่งกลุ่มด้วยระยะห่างระหว่างจุด จุดไหนที่อยู่ใกล้กันจะถูกนำมาอยู่ในกลุ่มเดียวกัน จากวิธีการแบ่งกลุ่มที่แตกต่างกันนี้ เพื่อดูความแตกต่างในการแบ่งกลุ่มว่าเป็นไปทางเดียวกันหรือไม่ โดยเลือกกลุ่มที่น่าสนใจจากแต่ละเทคนิคนำมาอินเตอร์เซกชัน (intersection data) กันเพื่อดูวิธีการแบ่งกลุ่มที่แตกต่างกัน ลักษณะของคนที่อยู่ในกลุ่มต่างกันอย่างไร ซึ่งวิเคราะห์และทำนายผลในบทต่อไป



บทที่ 4

ผลการศึกษา

4.1 การจำแนกพฤติกรรมด้วยอัลกอริทึมการทำเหมืองข้อมูล

วัตถุประสงค์ของการจำแนกพฤติกรรมผู้ไปใช้บริการแบบไม่สมเหตุสมผล เพื่อค้นหาพฤติกรรมที่ผิดปกติของการรับยากลับบ้านแบบไม่สมเหตุสมผล โดยการประยุกต์ใช้อัลกอริทึมการทำเหมืองข้อมูลจำแนกกลุ่มผู้ที่มีพฤติกรรมผิดปกติและกลุ่มพฤติกรรมปกติ

4.2 อัลกอริทึมการทำเหมืองข้อมูล

การทำเหมืองข้อมูล หมายถึง กระบวนการวิเคราะห์ข้อมูล เพื่อค้นหารูปแบบ (patterns) หรือการหาความสัมพันธ์ (relation) ระหว่างข้อมูลที่มีประโยชน์ในฐานข้อมูลขนาดใหญ่ โดยมีขั้นตอนวิเคราะห์ข้อมูล ได้แก่ 1) การหาความสัมพันธ์ของข้อมูลที่เกิดขึ้นร่วมกัน (association analysis) 2) การจัดกลุ่มข้อมูล (clustering) เป็นการจัดกลุ่มข้อมูลตามความคล้ายคลึง เช่น การจัดกลุ่มวิธีเชกเมนต์เทชั่น (Segmentation) โดยแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการ 3) การจำแนกประเภทข้อมูล (classification) เป็นการคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต

4.3 ข้อดีของการนำอัลกอริทึมการทำเหมืองข้อมูลมาใช้ในการตรวจสอบพฤติกรรม

การทำเหมืองข้อมูลช่วยให้ประหยัดเวลา สามารถค้นหาบุคคลที่มีพฤติกรรมผิดปกติได้รวดเร็ว และใช้งานกับฐานข้อมูลที่มีขนาดใหญ่ ซึ่งช่วยลดภาระของเจ้าหน้าที่ตรวจสอบ จากผลการวิเคราะห์สามารถแบ่งกลุ่มและเห็นรูปแบบพฤติกรรมของบุคคลได้ชัดเจน ทำให้องค์กรสามารถเฝ้าระวังบุคคลที่มีพฤติกรรมผิดปกติในลักษณะเดียวกัน ช่วยลดความเสียหายที่อาจจะเกิดขึ้นให้น้อยลงได้

จากข้อมูลการเบิกจ่ายเงินค่ารักษาพยาบาลปีงบประมาณ 2556 พบว่ามีจำนวนผู้ป่วยที่ไปใช้บริการทั้งสิ้น 24,125,143 ครั้ง จำนวนคน 2,751,425 คน ในการตรวจสอบข้อมูลดังกล่าวทำได้โดยการเขียนสคริปคำสั่งด้วยภาษาโปรแกรม STATA เพื่อตรวจสอบรายการเรียกเก็บค่าใช้จ่าย (transection) ทั้งหมด 24 ล้านครั้ง ซึ่งการวิเคราะห์สามารถเห็นรายการผิดปกติค่าโดดเด่น แต่ไม่มีการแบ่งกลุ่มตามลักษณะพฤติกรรมของผู้ไปใช้บริการและรายการที่ผิดปกติซึ่งซ่อนอยู่ในข้อมูล

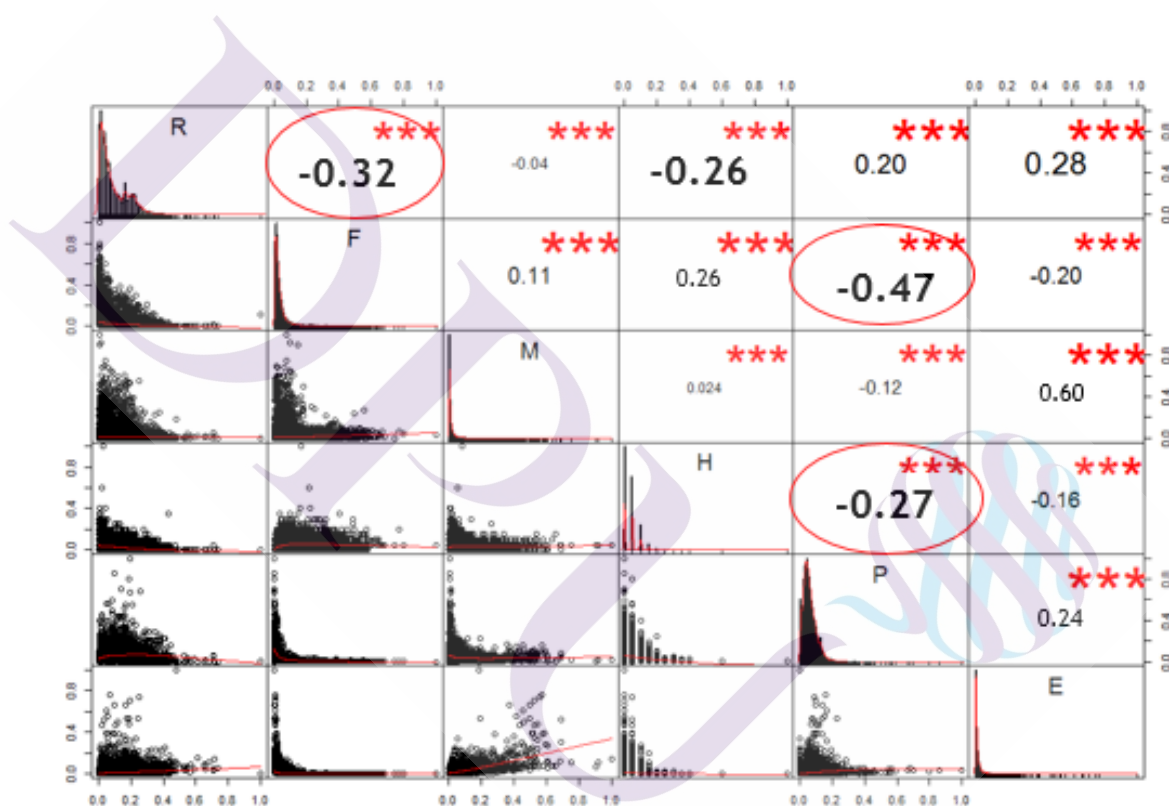
4.4 การหาความสัมพันธ์ระหว่างตัวแปร (Correlation)

การพิจารณาความสัมพันธ์ระหว่างตัวแปรที่นำมาใช้ในการศึกษาค้นคว้าครั้งนี้ เพื่อดูทิศทางความสัมพันธ์ โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient (r)) เป็นตัววัดความสัมพันธ์ระหว่างตัวแปร X และ Y มีความสัมพันธ์กันหรือไม่ โดยมีช่วงอยู่ที่ -1 ถึง 1 ตัวแปรที่นำมาใช้ทดสอบมีทั้งหมด 6 ตัว ได้แก่ R, F, M, H, P, E จากตาราง Master และตาราง Bill มาเชื่อมต่อกัน จำนวนผู้ป่วยบริการทั้งหมดที่มีการเบิกจ่ายค่ายานากลับไปใช้ที่บ้านจำนวน 525,085 คน เป็นเงินเบิกค่ายากลับบ้านทั้งสิ้น 4,713 ล้านบาท เลือกเฉพาะคนที่มีค่ายามากกว่า 30,000 บาทต่อปี ตัดคนที่ใช้ยามะเร็งออกจำนวน 437 คน เหลือจำนวน 31,001 คน เพื่อใช้ในทดสอบการแบ่งกลุ่ม ดังตารางที่ 4.1

ตารางที่ 4.1 ตัวแปรที่ใช้แบ่งกลุ่มพฤติกรรมผู้มารับบริการ

ตัวแปร	ความหมาย
R	จำนวนวันล่าสุดที่ไปใช้บริการ/ปี
F	ความถี่ที่ไปใช้บริการ/ปี
M	ค่ายากลับบ้าน/ปี
H	จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี
P	ช่วงห่างระหว่างวันที่ไปใช้บริการ/ปี
E	ค่ายาเฉลี่ยต่อครั้ง

การดูทิศทางความสัมพันธ์ระหว่างตัวแปรโดยมีค่าสัมประสิทธิ์สหสัมพันธ์เป็นตัวบ่งชี้ถึงระดับความสัมพันธ์ หากค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ -1 หรือ 1 แสดงถึงตัวแปรทั้งสองมีความสัมพันธ์กันอย่างมาก และหากมีค่าเข้าใกล้ 0 หมายความว่ามีความสัมพันธ์กันน้อยหรือไม่มีเลย โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ ใช้เกณฑ์ดังนี้ (Hinkle D. E, 1998, p. 118) จากภาพที่ 4.1 แสดงความสัมพันธ์ระหว่างตัวแปร F, P มีความสัมพันธ์กันน้อยร้อยละ -47 ส่วนตัวแปร R, F มีความสัมพันธ์น้อยร้อยละ -32 ตัวแปรที่นำมาใช้มีความสัมพันธ์ในทิศทางตรงกันข้าม



ภาพที่ 4.1 แสดงค่าความสัมพันธ์ระหว่างตัวแปร

4.5 การแบ่งกลุ่มข้อมูล

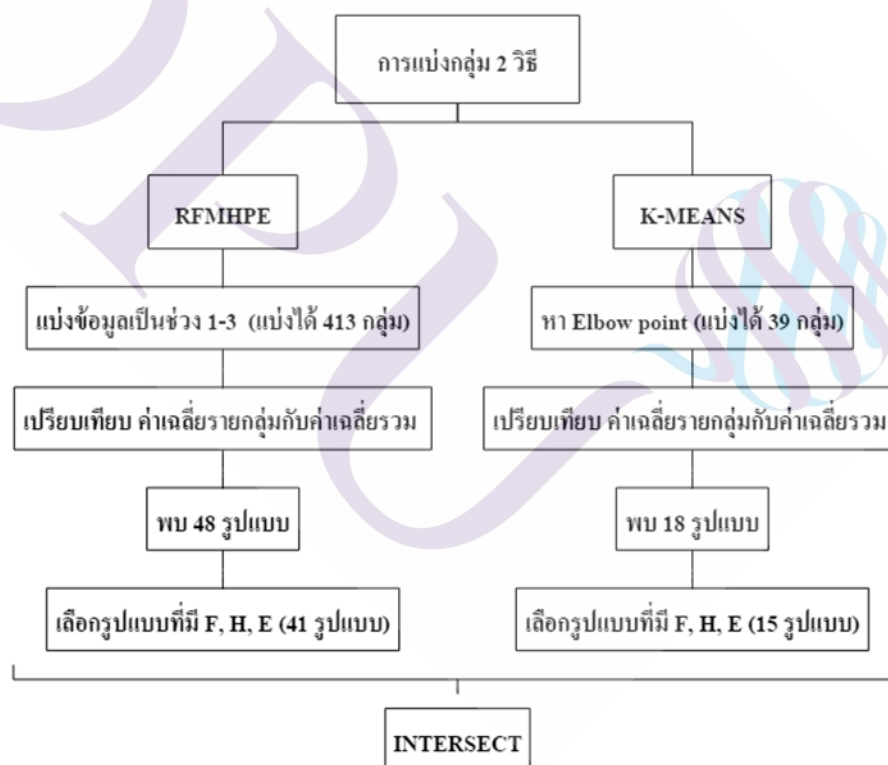
จากภาพที่ 4.2 ในการศึกษาครั้งนี้ ได้ทำการแบ่งกลุ่มข้อมูลเป็น 2 วิธี คือ RFMHPE และวิธีเคมีน K-means เพื่อจัดกลุ่มข้อมูลตามความคล้ายคลึง โดยอธิบายเป็นลำดับขั้นตอนดังต่อไปนี้

1. ขั้นตอนการแบ่งกลุ่ม

1.1 การแบ่งกลุ่มด้วยวิธี RFMHPE โดยแบ่งข้อมูลเป็น 3 ช่วง เท่าๆกัน แบ่งได้ 413 กลุ่ม หลังจากผ่านขั้นตอนเปรียบเทียบค่าเฉลี่ยรายกลุ่มกับค่าเฉลี่ยรวม พบรูปแบบ 48 รูปแบบ

1.2 การแบ่งกลุ่มด้วยวิธีเคมีน แบ่งข้อมูลเป็น 39 กลุ่ม หลังจากผ่านขั้นตอนเปรียบเทียบค่าเฉลี่ยรายกลุ่มกับค่าเฉลี่ยรวม พบรูปแบบ 18 รูปแบบ

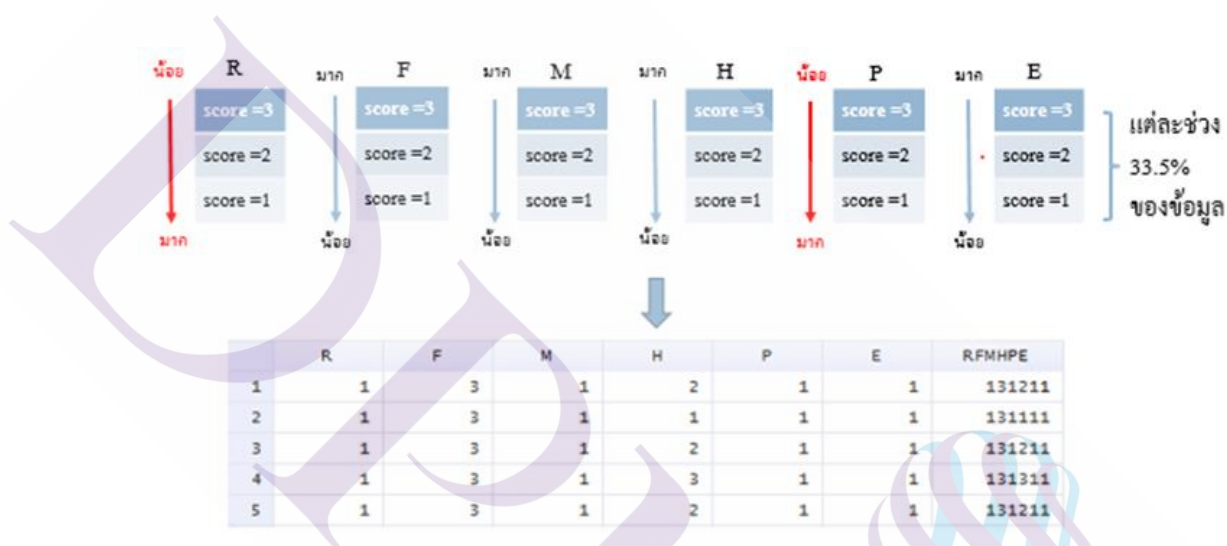
2. ขั้นตอนการเลือกรูปแบบ ได้ทำการเลือกกลุ่มที่มีรูปแบบ F, H, E ไปใช้ในการวิเคราะห์แบบจำลองต่อไป



ภาพที่ 4.2 ขั้นตอนการแบ่งกลุ่ม RFMHPE, K-means

4.6 การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการด้วยวิธีเซกเมนต์เทชั่น (Segmentation)

การแบ่งกลุ่มข้อมูลตามพฤติกรรมผู้ไปใช้บริการด้วยวิธีเซกเมนต์เทชั่น เป็นการจัดกลุ่มข้อมูลตามพฤติกรรมการไปใช้บริการที่สถานพยาบาล ผู้ที่มีพฤติกรรมคล้ายคลึงกันอยู่ในกลุ่มเดียวกัน ทำให้เห็นความเด่นชัดที่แตกต่างกันตามพฤติกรรมของผู้ไปใช้บริการ จากภาพที่ 4.3 ทำการเรียงข้อมูลจากน้อยไปมาก แล้วแบ่งเป็น 3 ช่วง ช่วงละ 33.5 เท่าๆกัน ทั้ง 6 ตัวแปร ได้ค่า 1-3 โดยค่าคะแนน 3 เป็นค่าที่มากที่สุด จากภาพช่อง RFMHPE คือ ค่าคะแนนที่ได้ของแต่ละเรคคอร์ด



ภาพที่ 4.3 การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการ

4.7 ผลการวิเคราะห์การแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการด้วยวิธีเซกเมนต์เทชั่น

จากการแบ่งกลุ่มตามพฤติกรรมผู้ไปใช้บริการที่สถานพยาบาล ได้ทั้งหมด 413 กลุ่ม วิธีการหารูปแบบได้มาจากการนำค่าเฉลี่ย (Mean) รายกลุ่ม เปรียบเทียบกับค่าเฉลี่ยรวม (แถวสุดท้าย) ถ้าค่าเฉลี่ยรายกลุ่มตัวแปรใดมีค่ามากกว่าค่าเฉลี่ยรวม (แถวสุดท้าย) ตัวแปรนั้นจะปรากฏค่าในช่องรูปแบบ เช่น ลำดับที่ 1 ค่าการจัดกลุ่ม 111112 ได้รูปแบบ RP มาจาก ค่า R เท่ากับ 20 วัน ซึ่งมีค่าน้อยกว่าค่าเฉลี่ยรวม 69 วัน ทำให้ค่า R ปรากฏในช่องรูปแบบ และค่า P เท่ากับ 16 วัน มีค่าน้อยกว่า 27 วัน ทำให้ค่า P (ซึ่งตัวแปร R, P ยึดค่าน้อย ส่วนตัวแปร F, M, H, E ยึดค่ามาก ดังภาพที่ 4.3)

4.8 รูปแบบที่พบวิธีเชกเมนต์แท่ง

จากตารางที่ 4.3 แสดงรูปแบบการแบ่งกลุ่มข้อมูลตามพฤติกรรมผู้ไปใช้บริการด้วยวิธีเชกเมนต์แท่ง พบรูปแบบทั้งหมด 48 รูปแบบ รูปแบบที่พบมากที่สุดคือ RHP 3,133 คน รองลงมาคือรูปแบบ RFHP จำนวน 3,118 คน และรูปแบบ RP จำนวน 2,373 คน ตามลำดับ จากรูปแบบที่ได้นี้สามารถบอกลักษณะพฤติกรรมได้ชัดเจน ว่าบุคคลไหนมีรูปแบบพฤติกรรมอย่างไร เช่น คนที่มีรูปแบบเท่ากับ H คือ คนกลุ่มนี้มีพฤติกรรมไปสถานพยาบาลหลายแห่ง ส่วนคนที่อยู่ในกลุ่มรูปแบบ MHE คือ คนกลุ่มนี้มีพฤติกรรมการเบิกจ่ายเงินค่ายากลับบ้านจำนวนมาก ไปใช้บริการหลายสถานพยาบาล และมีค่ายาเฉลี่ยต่อครั้งสูง เป็นต้น

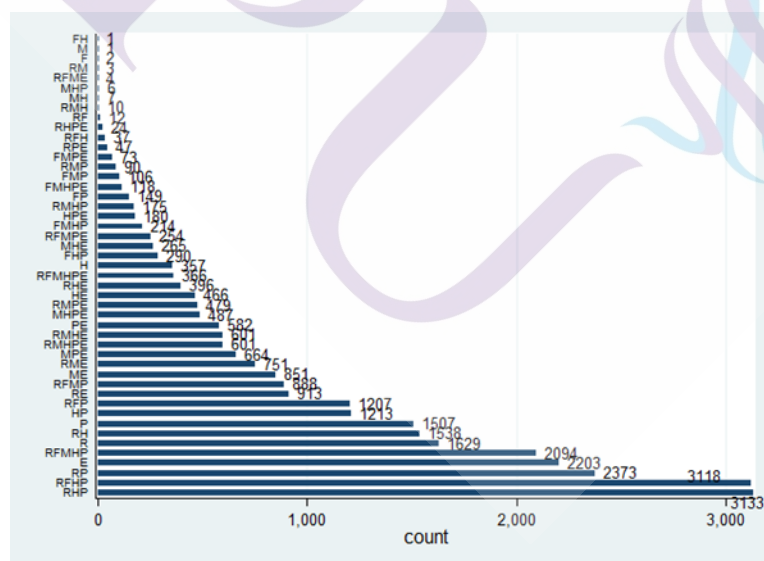
ตารางที่ 4.2 รูปแบบพฤติกรรมผู้มารับบริการวิธีเชกเมนต์แท่ง

ลำดับ	รูปแบบ	จำนวน	ร้อยละ	ลำดับ	รูปแบบ	จำนวน	ร้อยละ
1	E	2,203	7.11	21	P	1,507	4.86
2	F	2	0.01	22	PE	582	1.88
3	FH	1	0	23	R	1,629	5.25
4	FHP	290	0.94	24	RE	913	2.95
5	FMHP	214	0.69	25	RF	12	0.04
6	FMHPE	118	0.38	26	RFH	37	0.12
7	FMP	106	0.34	27	RFHP	3,118	10.06
8	FMPE	73	0.24	28	RFME	4	0.01
9	FP	149	0.48	29	RFMHP	2,094	6.75
10	H	357	1.15	30	RFMHPE	366	1.18
11	HE	466	1.5	31	RFMP	888	2.86
12	HP	1,213	3.91	32	RFMPE	254	0.82
13	HPE	180	0.58	33	RFP	1,207	3.89
14	M	1	0.000	34	RH	1,538	4.96
15	ME	851	2.75	35	RHE	396	1.28
16	MH	7	0.02	36	RHP	3,133	10.11
17	MHE	265	0.85	37	RHPE	24	0.08
18	MHP	6	0.02	38	RM	3	0.01
19	MHPE	487	1.57	39	RME	751	2.42
20	MPE	664	2.14	40	RMH	10	0.03

ตารางที่ 4.2 รูปแบบพฤติกรรมผู้มารับบริการวิธีเชกเมนต์เทชั่น (ต่อ)

ลำดับ	รูปแบบ	จำนวน
41	RMHE	601
42	RMHP	175
43	RMHPE	601
44	RMP	90
45	RMPE	479
46	RP	2,373
47	RPE	47
48	ไม่พบรูปแบบ	516
41	RMHE	601
42	RMHP	175

จากภาพที่ 4.4 แสดงภาพรูปแบบที่พบวิธี RFMHPE จำนวน 48 รูปแบบ รูปแบบที่พบมากที่สุด คือ RHP, RFHP, RP



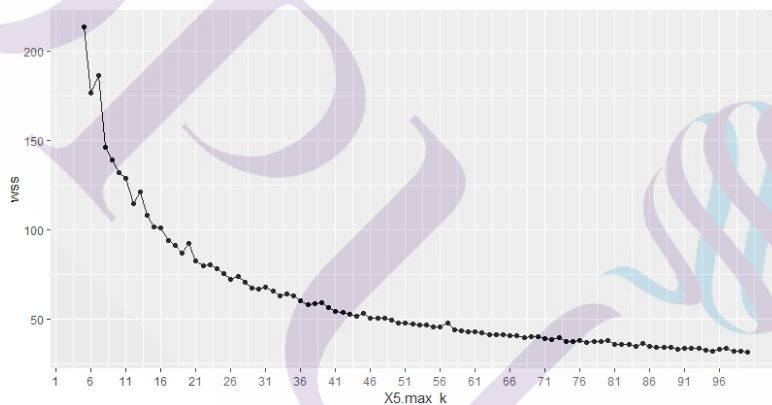
ภาพที่ 4.4 แสดงรูปแบบที่แบ่งด้วยวิธี RFMHPE

4.9 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน

การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน เป็นการแบ่งกลุ่มข้อมูลออกเป็น K กลุ่ม แล้วแทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ย โดยใช้จุดศูนย์กลาง (centroid) ของกลุ่มเป็นตัววัดระยะห่างของข้อมูล ถ้าระยะห่างของข้อมูลอยู่ใกล้กัน ข้อมูลบุคคลนั้นก็อยู่กลุ่มเดียวกัน ดังนั้นก่อนที่จะมีการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน ต้องหาจุด Elbow point ก่อน เพื่อให้ทราบว่าควรแบ่งกลุ่ม K จำนวนเท่าไรจึงเหมาะสมที่สุด

4.10 การหาค่า Elbow point เพื่อหาค่า K ที่เหมาะสมที่สุด

การค้นหาค่า K ที่เหมาะสมที่สุดในการแบ่งกลุ่มข้อมูล โดยเขียนสคริปคำสั่ง โปรแกรม R ทำการค้นหาค่า K จากตัวแปรทั้ง 6 ตัว คือ R, F, M, H, P, E พบจุดที่เหมาะสมคือ K=39 โดยแบ่งกลุ่มจากข้อมูลที่มีลักษณะคล้ายคลึงกันและมีระยะห่างระหว่างกลุ่มใกล้เคียงกัน ลักษณะของกราฟที่ลาดลง เกิดจากการเพิ่มจำนวน K ให้มากกว่า 39 ก็ไม่ส่งผลต่อการแบ่งกลุ่ม ดังภาพที่ 4.5



ภาพที่ 4.5 กราฟแสดงจุด K ที่เหมาะสมจาก Elbow point

จากภาพที่ 4.6 ก่อนทำการแบ่งกลุ่มข้อมูลผู้ป่วยนอกด้วยวิธีเคมีน ต้องทำนอร์มอลไลซ์ (NORMALIZATION) ข้อมูลให้อยู่ในช่วง 0-1 ทุกตัวแปร เพื่อให้อัลกอริทึมการทำเหมืองข้อมูลสามารถนำไปใช้ประมวลผลได้ และกำหนดค่า K เท่ากับ 39 กลุ่ม ตามพฤติกรรมที่คล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน

R	F	M	H	P	E	cluster
.0038511	.4527027	.000971	0	.0109492	.0001059	4
.0077022	.5022523	.0022803	.1	.0035242	.0001088	4
.0025674	.4076577	.0000321	.1	.004388	.0001099	29
.0038511	.6396396	.0059516	.05	.0028345	.0001152	4
.0025674	.4009009	.0001814	0	.0050659	.0001191	29
.0077022	.490991	.0029087	0	.0035521	.0001304	4
.0038511	.4009009	.0006821	0	.0042944	.0001338	29
.0038511	.4189189	.0011878	.15	.0040984	.0001347	22

ภาพที่ 4.6 การทำนอร์มอลไลซ์ (NORMALIZATION) ข้อมูล

4.11 รูปแบบที่พบจากการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน

จากตารางที่ 4.4 จากการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน นำค่าเฉลี่ยรายกลุ่มของแต่ละตัวแปร เปรียบเทียบกับค่าเฉลี่ยรวมแถวสุดท้าย ถ้าค่าเฉลี่ยของกลุ่มใด มีค่ามากกว่าค่าเฉลี่ยรวม ค่าตัวแปรนั้น จะปรากฏในช่องรูปแบบ เช่น ค่าเฉลี่ย E ของ กลุ่มที่ 2 เท่ากับ 9,843.54 บาท/ครั้ง มีค่ามากกว่าค่าเฉลี่ยรวม 8,221.87 บาท/ครั้ง ค่า E จึงปรากฏในช่องรูปแบบ เป็นต้น

ตารางที่ 4.3 แสดงรูปแบบการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน

ลำดับกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
1	878	91.24	12.03	60,814.05	1	22.73	5,824.68	P
2	1388	174.3	5.81	50,613.80	1	29.14	9,843.54	E
3	23	108.17	4.87	1,386,625.00	1.26	49.13	309,131.10	ME
4	159	6.18	230.25	135,317.50	2.06	1.77	595.89	RFMHP
5	454	244.39	9.37	77,936.64	1.38	13.16	16,005.94	PE
6	1033	112.71	15.28	71,372.43	2.3	17.65	5,518.02	HP
7	171	128.24	4.12	54,157.57	1.22	105.68	16,709.78	E
8	739	21.03	34.11	77,404.71	1	10	2,399.22	RFP
9	728	29.39	23.55	60,876.27	4.31	17.36	3,089.51	RFHP
10	2442	21.43	13.78	56,654.64	2	23.48	4,353.97	RHP
11	1829	15.65	17.15	60,684.16	1	18.34	3,745.90	RP
12	522	17.61	109.24	99,033.84	2.07	3.35	916.75	RFMHP

ตารางที่ 4.3 แสดงรูปแบบการแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน (ต่อ)

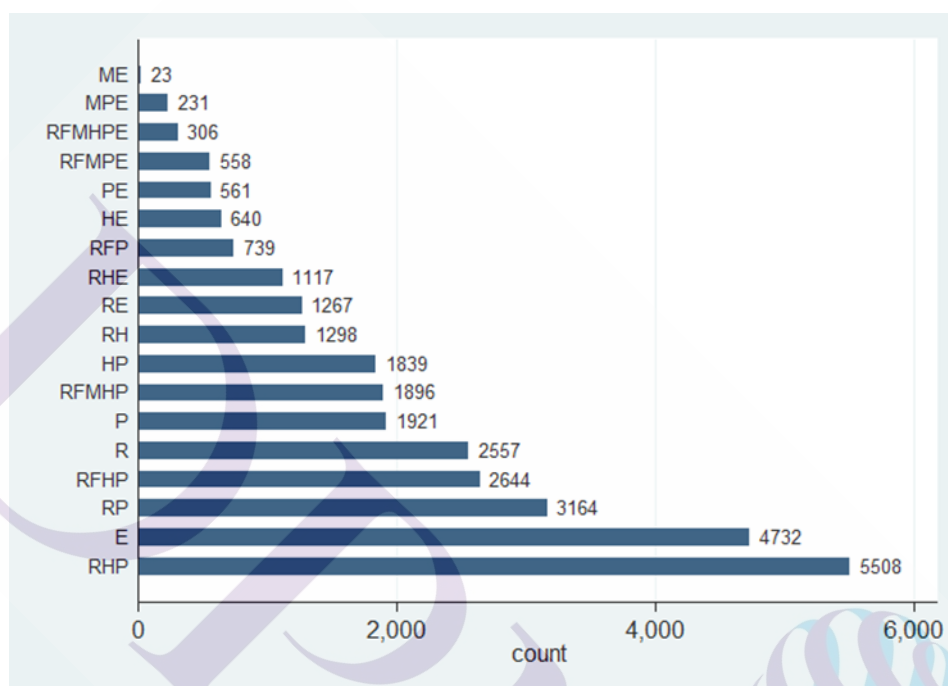
ลำดับกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
13	231	140.91	10.85	421,323.70	1.49	24.65	53,021.23	MPE
14	941	157.3	3.26	57,631.01	1.03	56.09	19,219.57	E
15	852	124.11	5.35	58,291.27	1	40.13	11,538.06	E
16	1043	139.18	11.68	66,674.92	1	18.19	8,142.19	P
17	806	170.27	11.71	65,622.47	2.17	19.92	7,501.34	HP
18	1298	68.26	8.42	58,234.48	2.17	33.56	7,326.31	RH
19	1798	24.37	19.3	59,862.97	3	18.94	3,576.78	RHP
20	32	93.03	4.34	46,680.31	1.41	225.39	13,444.79	E
21	1523	21.67	9.33	58,466.62	1	32.39	6,576.59	R
22	243	10.94	137.79	105,796.40	3.94	2.94	785.75	RFMHP
23	107	352.16	3.55	68,497.72	1.2	10.66	34,578.25	PE
24	585	21.15	58	81,152.00	3.42	6.36	1,469.78	RFHP
25	71	61.61	23.89	1,564,860.00	1.92	19.24	90,172.07	RFMHPE
26	517	22.99	65.23	91,683.52	1.74	5.46	1,454.41	RFMHP
27	640	130.87	5.92	55,017.73	2.07	39.16	10,152.29	HE
28	235	47.33	29.59	861,954.90	1.75	13.94	40,360.65	RFMHPE
29	455	8.91	163.05	118,540.40	1.77	2.33	735.96	RFMHP
30	558	32.27	23.07	373,873.30	1.68	17.44	20,658.73	RFMPE
31	1331	17.94	29.78	68,336.00	2	11.6	2,416.32	RFHP
32	1335	49.67	14.99	60,977.86	1	20.04	4,562.53	RP
33	770	211.72	3	51,919.90	1.07	44.42	18,886.29	E
34	1268	61.18	18.02	60,645.55	2.17	17.08	3,701.84	RHP
35	378	33.26	4.36	59,294.51	1.34	85.47	16,144.39	RE
36	1117	26.47	7.32	59,023.08	2.14	45.66	8,656.23	RHE
37	1034	61.14	7.39	53,626.39	1	37.44	7,811.06	R
38	578	79.31	4.93	57,153.13	1.3	61.72	13,378.17	E
39	889	27.04	5.6	60,043.53	1	53.31	11,398.09	RE
ค่าเฉลี่ยรวม	31,001	68.74	20.75	82,390.45	1.71	26.8	8,221.87	

จากตารางที่ 4.5 แสดงรูปแบบทั้งหมด 18 รูปแบบ ค่าแต่ละตัวแปร เป็นค่าเฉลี่ยของกลุ่มของแต่ละรูปแบบ รูปแบบที่พบมากที่สุด RHP จำนวน 5,508 คน รูปแบบ E จำนวน 4,732 คน และรูปแบบ RP จำนวน 3,164 คน

ตารางที่ 4.4 รูปแบบพฤติกรรมผู้ไปใช้บริการจากวิธีเคมีน

ลำดับ	รูปแบบ	จำนวน	ร้อยละ
1	E	4,732	15.26
2	HE	640	2.06
3	HP	1,839	5.93
4	ME	23	0.07
5	MPE	231	0.75
6	P	1,921	6.2
7	PE	561	1.81
8	R	2,557	8.25
9	RE	1,267	4.09
10	RFHP	2,644	8.53
11	RFMHP	1,896	6.12
12	RFMHPE	306	0.99
13	RFMPE	558	1.8
14	RFP	739	2.38
15	RH	1,298	4.19
16	RHE	1,117	3.6
17	RHP	5,508	17.77
18	RP	3,164	10.21

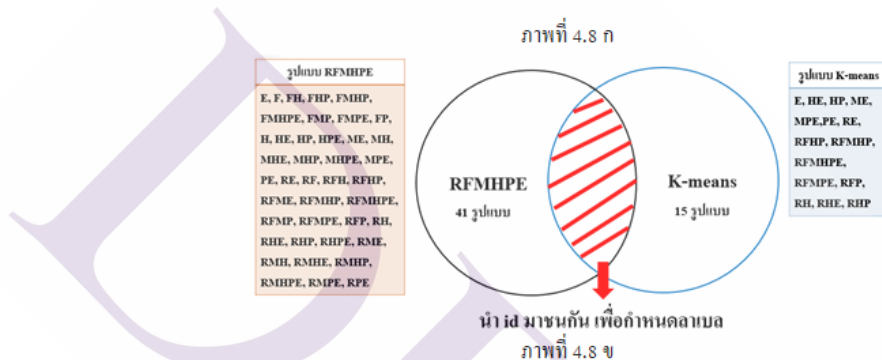
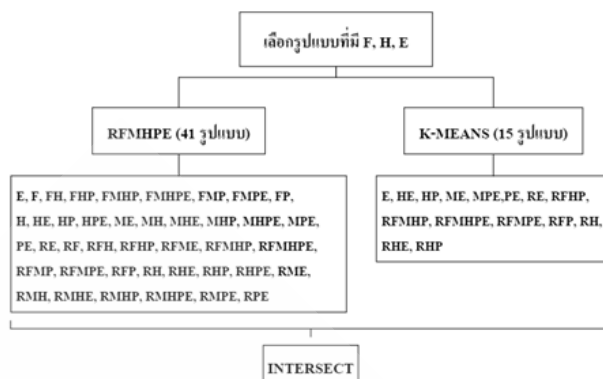
จากภาพที่ 4.7 แสดงภาพรูปแบบที่พบวิธีเคมีน จำนวน 18 รูปแบบ รูปแบบที่พบมากที่สุด คือ RHP, E, RP



ภาพที่ 4.7 แสดงรูปแบบที่แบ่งด้วยวิธีเคมีน

4.12 การอินเตอร์เซกข้อมูล (Intersect Data)

ขั้นตอนก่อนการอินเตอร์เซกข้อมูลผู้ไปใช้บริการจำนวน 31,001 คน แบ่งกลุ่มด้วยวิธี RFMHPE เลือกกลุ่มที่มีรูปแบบ F, H, E ได้ 41 รูปแบบ จำนวน 24,882 คน และการแบ่งกลุ่มด้วยวิธี เคมีน เลือกรูปแบบ F, H, E ได้ 15 รูปแบบ จำนวน 23,359 คน ไม่พบรูปแบบ F, H, E จำนวน 4,858 คน แล้วย้นำข้อมูลรายบุคคลจาก 2 วิธีมาอินเตอร์เซก ดังภาพที่ 4.8

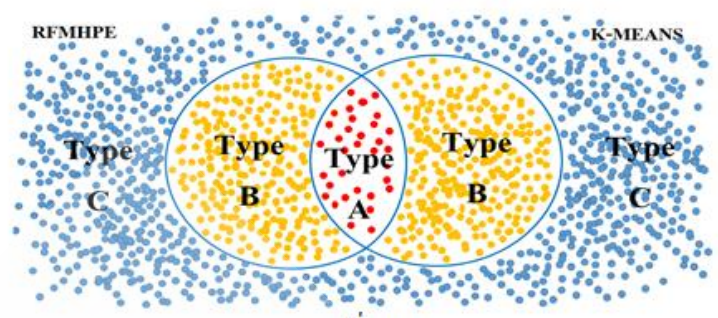


ภาพที่ 4.8 ขั้นตอนการอินเตอร์เซก

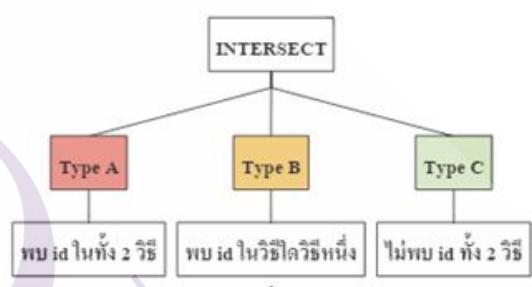
4.13 การกำหนดคลาส

จากภาพที่ 4.9 ก แสดงขั้นตอนการอินเตอร์เซกแล้วกำหนดคลาส โดยกำหนดให้ประเภท A คือ id บุคคลนั้นพบทั้ง 2 วิธี ส่วน id บุคคลที่พบในวิธีใดวิธีหนึ่ง กำหนดให้เป็นประเภท B และ id บุคคลที่ไม่พบในวิธีไหนเลยกำหนดให้เป็นประเภท C

จากภาพที่ 4.9 ค คือวิธีการกำหนดประเภทจากวิธีการแบ่งกลุ่ม RFMHPE รูปแบบที่มีค่า F, H, E ไล่ 1 ในช่อง T1 ส่วนการแบ่งกลุ่มด้วยวิธีเคมีน เลือกรูปแบบที่มีค่า F, H, E ไล่ 1 ในช่อง T2 และไม่พบรูปแบบค่า F, H, E ทั้ง 2 วิธีไล่ 1 ในช่อง T3



ภาพที่ 4.9 ก



ภาพที่ 4.9 ข

id	scoreRFMPHE	scoreKmeans	T1	T2	T3	type
4164	RME	R	.	.	1	C
4165	RME	R	.	.	1	C
4166	RME	RFMPE	.	1	.	B
4167	RME	R	.	.	1	C
4168	RMHP	RHP	1	1	.	A
4169	RMHP	RHP	1	1	.	A

ภาพที่ 4.9 ค

ภาพที่ 4.9 การกำหนดคลาส

จากตารางที่ 4.5 แสดงจำนวนคนที่ได้หลังจากการอินเตอร์เซก โดยประเภท A คือกลุ่มที่พบ id ทั้ง 2 วิธี เป็นกลุ่มที่น่าสนใจเป็นพิเศษ มีจำนวน 22,098 คน คิดเป็นร้อยละ 71.28 ประเภท B คือกลุ่มที่พบ id ในวิธีใดวิธีหนึ่ง เป็นกลุ่มที่น่าสนใจรองลงมา มีจำนวน 4,045 คน คิดเป็นร้อยละ 13.05 และประเภท C คือกลุ่มที่ไม่พบ id ในทั้ง 2 วิธี เป็นกลุ่มพฤติกรรมปกติ มีจำนวน 4,858 คน คิดเป็นร้อยละ 15.67 ตามลำดับ

ตารางที่ 4.5 จำนวนคนหลังการอินเตอร์เซ็ก

ประเภท	จำนวน (คน)	ร้อยละ		ความหมาย
A	22,098	71.28	พบ id ทั้ง 2 วิธี	กลุ่มที่น่าสนใจเป็นพิเศษ
B	4,045	13.05	พบ id ในวิธีใดวิธีหนึ่ง	กลุ่มที่น่าสนใจรองลงมา
C	4,858	15.67	ไม่พบ id ทั้ง 2 วิธี	กลุ่มพฤติกรรมปกติ
รวม	31,001	100		

จากตารางที่ 4.6 แสดงกลุ่มประเภทข้อมูลหลังจากทำการอินเตอร์เซ็ก ประเภท A พบรูปแบบทั้ง 40 รูปแบบ ประเภท B มีทั้งหมด 22 รูปแบบ และประเภท C มีทั้งหมด 5 รูปแบบ

ตารางที่ 4.6 รูปแบบที่พบแยกรายประเภท

ประเภท A	ประเภท B	ประเภท C
E, F, FH, FHP, FMHP, FMHPE, FMP, FMPE, FP, H, HE, HP, HPE, ME, MH, MHE, MHP, MHPE, MPE, PE, RE, RF, RFH, RFHP, RFME, RFMHP, RFMHPE, RFMP, RFMPE, RFP, RH, RHE, RHP, RHPE, RME, RMH, RMHE, RMHP, RMHPE, RMPE	E, F, FMP, FMPE, FP, M, ME, MPE, P, PE, R, RE, RF, RFME, RFMP, RFMPE, RFP, RME, RMP, RMPE, RP, RPE	P, R, RM, RMP, RP

จากตารางที่ 4.7 สัดส่วนร้อยละของรูปแบบประเภท A (กลุ่มที่น่าสนใจเป็นพิเศษ) ได้มาจากวิธีการแบ่งกลุ่ม RFMHPE และวิธีการแบ่งกลุ่มเคมिन คิดเป็นร้อยละ 71.28 รูปแบบ B (กลุ่มที่น่าสนใจรองลงมา) ได้มาจากวิธี RFMHPE คิดเป็นร้อยละ 8.98 ส่วนรูปแบบที่ได้จากวิธีการแบ่งกลุ่มเคมिन คิดเป็นร้อยละ 4.07 และรูปแบบ C (กลุ่มพฤติกรรมปกติ) ได้มาจากวิธี RFMHPE คิดเป็นร้อยละ 15.25 สัดส่วนของรูปแบบที่ได้จากวิธีเคมिन คิดเป็นร้อยละ 15.67

ตารางที่ 4.7 สัดส่วนรูปแบบที่พบแต่ละเทคนิค

เทคนิค	รูปแบบ A	รูปแบบ B	รูปแบบ C
RFMHPE	71.28%	8.98%	15.25%
K-means	71.28%	4.07%	15.67%

จากตารางที่ 4.8 หลังจากการอินเตอร์เซ็กพบว่ประเภท A มาจากการแบ่งกลุ่มวิธี RFMHPE จำนวน 40 รูปแบบ รูปแบบที่พบมากที่สุด คือ (RHP, RFHP, E) มาจากการแบ่งกลุ่มวิธีเคมีน จำนวน 15 รูปแบบ รูปแบบที่พบมากที่สุด (RHP, E, RFHP)

ประเภท B มาจากการแบ่งกลุ่มวิธี RFMHPE จำนวน 17 รูปแบบ รูปแบบที่พบมากที่สุด (RFP, RMPE, RME) มาจากการแบ่งกลุ่มวิธีเคมีน จำนวน 4 รูปแบบ รูปแบบที่พบมากที่สุด (E, RE, PE)

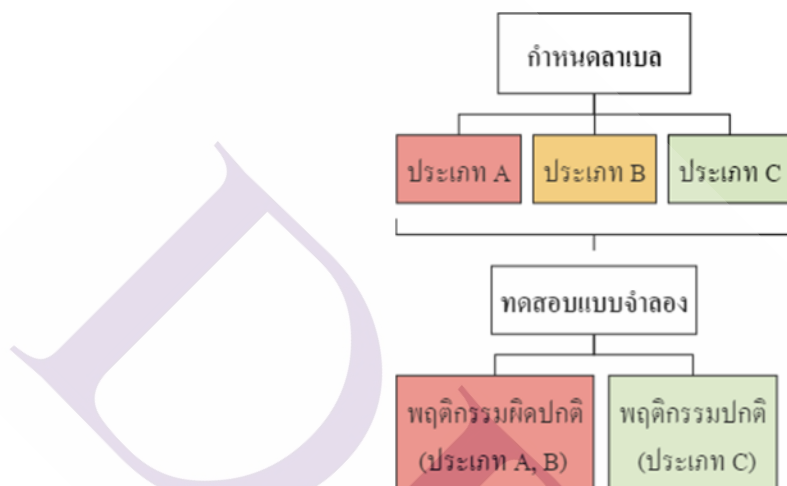
ประเภท C มาจากการแบ่งกลุ่มวิธี RFMHPE จำนวน 5 รูปแบบ รูปแบบที่พบมากที่สุด (RP, R, P) มาจากการแบ่งกลุ่มวิธีเคมีน จำนวน 3 รูปแบบ รูปแบบที่พบมากที่สุด (RP, R, P)

ตารางที่ 4.8 รูปแบบที่พบในแต่ละประเภท

ประเภท	เทคนิค RFMHPE	เทคนิคเคมีน
A	40 รูปแบบ (RHP, RFHP, E)	15 รูปแบบ (RHP, E, RFHP)
B	17 รูปแบบ (RFP, RMPE, RME)	4 รูปแบบ (E, RE, PE)
C	5 รูปแบบ (RP, R, P)	3 รูปแบบ (RP, R, P)

4.14 การทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ

หลังจากกำหนดคลาสเบลเรียบร้อยแล้ว ขั้นตอนต่อไปเป็นการทดสอบแบบจำลองเพื่อทำนายพฤติกรรมของผู้ป่วยกลุ่มพฤติกรรมผิดปกติและพฤติกรรมปกติ ดังภาพที่ 4.10

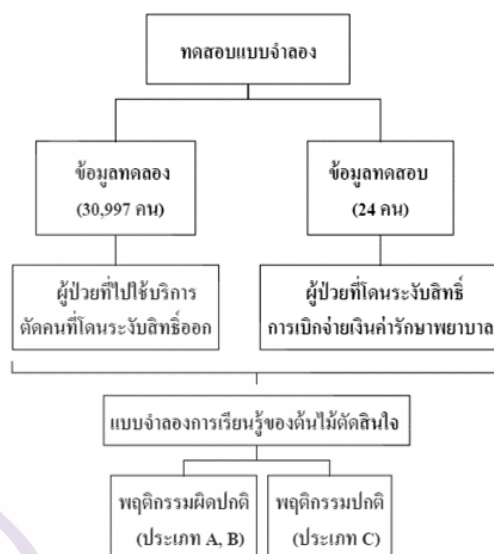


ภาพที่ 4.9 การทดสอบแบบจำลอง

จากภาพที่ 4.11 แสดงขั้นตอนการทดสอบแบบจำลอง โดยนำข้อมูลที่ผ่านการลาเบลแล้ว แบ่งข้อมูลเป็น 2 ชุด คือข้อมูลที่ใช้ทดลอง (train data) และข้อมูลที่ใช้ทดสอบ (test data) โดยนำข้อมูลที่ใช้ทดลองมาสอนระบบให้รู้ว่ามีข้อมูลใดอยู่ในกลุ่มเดียวกันบ้าง ผลลัพธ์ที่ได้จากการเรียนรู้คือแบบจำลองจัดประเภทข้อมูล (Classifier model) ส่วนข้อมูลที่ใช้ทดสอบจะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากแบบจำลอง เพื่อทดสอบความถูกต้อง หลังจากนั้นเมื่อมีข้อมูลใหม่เข้ามาก็นำข้อมูลมาทดสอบผ่านแบบจำลองนี้ โดยแบบจำลองจะสามารถทำนายกลุ่มของข้อมูลนี้ได้

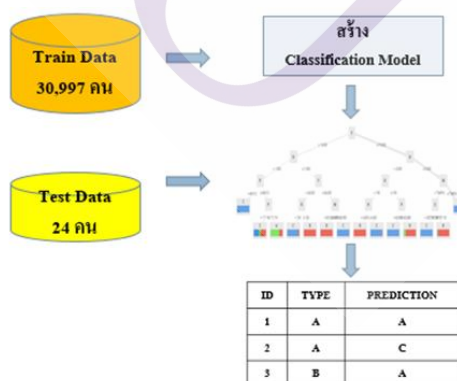
4.14.1 การทดสอบแบบจำลองด้วยกลุ่มตัวอย่าง 24 คน

ข้อมูลที่ใช้ทดลอง ตัดคนที่โดนระงับสิทธิ์ออกเหลือจำนวน 30,997 คน และข้อมูลที่ใช้ทดสอบ จำนวน 24 คน คือคนที่โดนระงับสิทธิ์การเบิกจ่ายเงินค่ารักษาพยาบาล หลังจากนั้นนำไปทดสอบแบบจำลอง



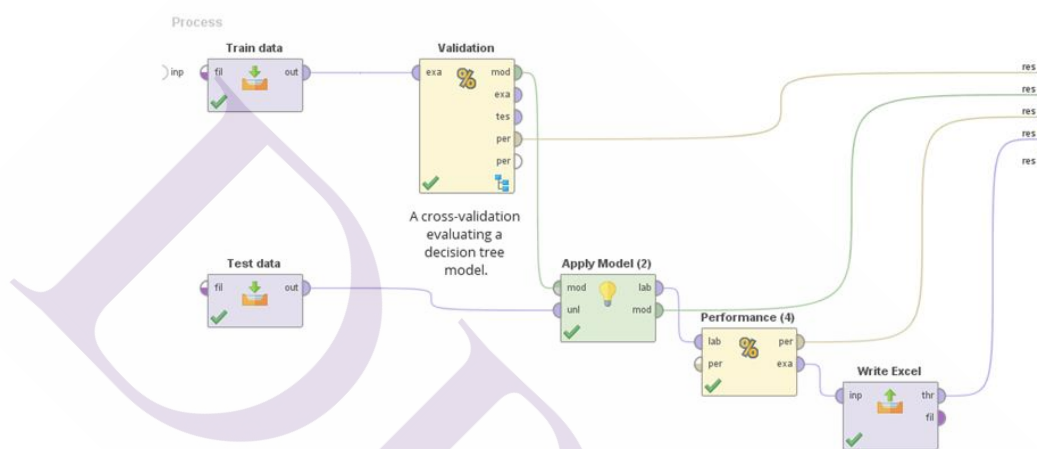
ภาพที่ 4. 11 ขั้นตอนการทดสอบแบบจำลอง

ในการทดสอบประสิทธิภาพของแบบจำลอง ได้แบ่งข้อมูลเรียนรู้ออกเป็น k ส่วนเท่า ๆ กัน (folds) เพื่อใช้ในการทดสอบ (Cross-validation) 10 ส่วน โดยข้อมูลที่ใช้ทดลอง และข้อมูลที่ใช้ทดสอบ สำหรับสร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ ด้วยการจำแนกประเภท (Classification) ข้อมูลออกเป็นกลุ่ม (class) จากลักษณะของตัวแปรที่ใส่เข้าไป ทำให้สามารถวิเคราะห์ข้อมูลและตัดสินใจได้ถูกต้องยิ่งขึ้น



ภาพที่ 4. 12 การทดสอบแบบจำลองด้วยกลุ่มตัวอย่าง 24 คน

จากภาพที่ 4.13 แสดงขั้นตอนการทำงานของโปรแกรม Rapid Miner Studio เพื่อทำนายความแม่นยำของแบบจำลอง โดยข้อมูลที่ใช้ทดลอง คือ ข้อมูลการเบิกจ่ายเงินค่ายากลับบ้านรายบุคคล ปีงบประมาณ 2556 ที่ผ่านการแบ่งกลุ่มและกำหนดคลาสประเภทไว้แล้ว ตัดคนที่โดนระงับสิทธิการเบิกจ่ายเงินค่ารักษาพยาบาลออกเหลือจำนวน 30,997 คน ส่วนข้อมูลที่ใช้ทดสอบ คือ ข้อมูลรายบุคคลที่ถูกระงับสิทธิการเบิกค่ารักษาพยาบาลจำนวน 24 คน



ภาพที่ 4.10 แสดงขั้นตอนการทำงานบนโปรแกรม Rapid Miner Studio

จากภาพที่ 4.14 ผลการทดสอบโดยแบ่งข้อมูลเรียนรู้ออกเป็น 10 ส่วนเท่าๆ กัน จำนวน 30,997 คน ใช้ในการทดสอบ พบว่าทำนายประเภท A เป็น A ถูก จำนวน 16,271 คน ทำนาย A เป็น B จำนวน 106 คน และทำนาย A เป็น C จำนวน 14 คน เป็นต้น

accuracy: 96.62% +/- 0.18% (mikro: 96.62%)

	true C	true B	true A	class precision
pred. C	12424	424	14	96.59%
pred. B	33	1253	106	90.01%
pred. A	2	470	16271	97.18%
class recall	99.72%	58.36%	99.27%	

ภาพที่ 4.11 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดลอง 30,997 คน

ผลการวัดประสิทธิภาพของแบบจำลองจากข้อมูลที่ใช้ทดสอบ จำนวน 24 คน ได้ความแม่นยำ ร้อยละ 91.67 สามารถทำนายประเภท A ถูก จำนวน 22 คน และทำนายประเภท B เป็นประเภท A ผิด 2 คน ดังภาพที่ 4.15

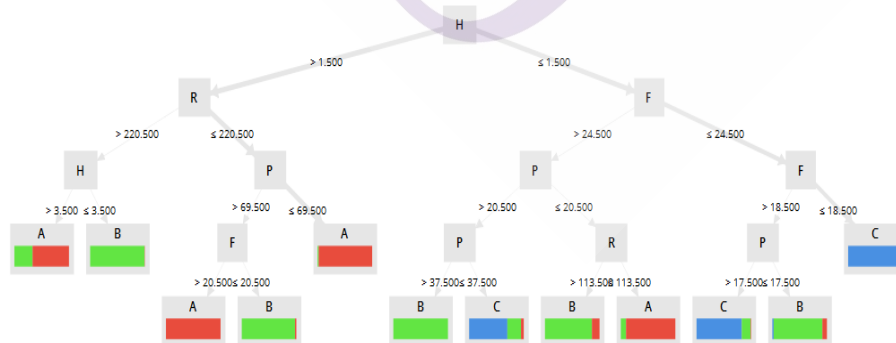
accuracy: 91.67%

	true A	true B	true C	class precision
pred. A	22	2	0	91.67%
pred. B	0	0	0	0.00%
pred. C	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

ภาพที่ 4.12 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดสอบ 24 คน

หลังจากการวัดประสิทธิภาพของแบบจำลอง โปรแกรมได้สร้างแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ และทำการคัดเลือกตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ (root node) หลังจากนั้นก็จะหาตัวแปรถัดไปเรื่อยๆ ในการหาความสัมพันธ์ของตัวแปรนี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG)

จากภาพที่ 4.16 แสดงการแตกกิ่งของต้นไม้ เกิดจากการเลือกตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุด คือตัวแปร H (จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี) มีค่าน้อยกว่าเท่ากับ 1.5 แห่ง/ปี จะแตกกิ่งไปทางขวา ตัวแปร F (ความถี่ที่ไปใช้บริการ/ปี) มีความสัมพันธ์กับกลุ่มลำดับถัดมา แตกกิ่งไปทางซ้ายซึ่งมีตัวแปร R (จำนวนวันล่าสุดที่ไปใช้บริการ/ปี) มีความสัมพันธ์ลำดับถัดมา และแตกกิ่งออกไปเรื่อยๆ จนทุกข้อมูลในโหนดนั้นเป็นกลุ่มเดียวกัน



ภาพที่ 4.13 แสดงการแบ่งลำดับชั้นของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ 24 คน

จากภาพที่ 4.17 ตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุดของกลุ่มตัวอย่าง 24 คน คือ H (จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี) ลำดับถัดมาคือ R (จำนวนวันล่าสุดที่ไปใช้บริการ/ปี) และ F (ความถี่ที่ไปใช้บริการ/ปี) สามารถอธิบายถึงทางซ้ายได้ว่า ถ้าตัวแปร H มีค่ามากกว่า 1.5 แห่ง/ปี ตัวแปร R มีค่ามากกว่า 220 วัน และตัวแปร H มีค่ามากกว่า 3.5 แห่ง/ปี ทำนายว่าบุคคลนั้นเป็นประเภท A จำนวน 2 คน เป็นประเภท B จำนวน 1 คน แต่ถ้าตัวแปร H มีค่าน้อยกว่าเท่ากับ 3.5 แห่ง/ปี ทำนายว่าบุคคลนั้นเป็นประเภท A จำนวน 1 คน และเป็นประเภท B จำนวน 166 คน เป็นต้น

Tree

```

H > 1.500
| R > 220.500
| | H > 3.500: A {C=0, B=1, A=2}
| | H ≤ 3.500: B {C=0, B=166, A=1}
| R ≤ 220.500
| | P > 69.500
| | | F > 20.500: A {C=0, B=0, A=2}
| | | F ≤ 20.500: B {C=0, B=192, A=5}
| | P ≤ 69.500: A {C=0, B=328, A=14996}
H ≤ 1.500
| F > 24.500
| | P > 20.500
| | | P > 37.500: B {C=0, B=2, A=0}
| | | P ≤ 37.500: C {C=14, B=5, A=1}
| | P ≤ 20.500
| | | R > 113.500: B {C=0, B=68, A=11}
| | | R ≤ 113.500: A {C=2, B=136, A=1282}
| F ≤ 24.500
| | F > 18.500
| | | P > 17.500: C {C=82, B=16, A=1}
| | | P ≤ 17.500: B {C=23, B=831, A=78}
| | F ≤ 18.500: C {C=12338, B=402, A=12}

```

ภาพที่ 4.14 แสดงการแตกกิ่งของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจกลุ่มตัวอย่าง 24 คน

4.14.2 การทดสอบแบบจำลองด้วยกลุ่มตัวอย่าง 74 คน

ข้อมูลที่ใช้ทดลอง ตัดคนที่โดนระงับสิทธิออกจำนวน 74 คน ซึ่งเดิมผู้ป่วยที่โดนระงับสิทธิจำนวน 24 คน ทำการสุ่มตัวอย่างจากข้อมูลชุดทดลองที่เป็นประเภท C จำนวน 50 คน ดังนั้นข้อมูลที่ใช้ทดลองเหลือจำนวน 30,947 คน และข้อมูลที่ใช้ทดสอบ จำนวน 74 คน หลังจากนั้นนำไปทดสอบแบบจำลอง ดังภาพที่ 4.18



ภาพที่ 4. 15 ขั้นตอนการทดสอบแบบจำลองกลุ่มตัวอย่าง 74 คน

จากภาพที่ 4.19 ผลการทดสอบโดยแบ่งข้อมูลเรียนรู้ออกเป็น 10 ส่วนเท่าๆ กัน จำนวน 30,947 คน ใช้ในการทดสอบ พบว่าทำนายประเภท A เป็น A ถูก จำนวน 21,592 คน ทำนาย A เป็น B จำนวน 424 คน และทำนาย A เป็น C จำนวน 41 คน เป็นต้น

accuracy: 89.63% +/- 0.51% (mikro: 89.63%)

	true C	true A	true B	class precision
pred. C	4835	41	2178	68.54%
pred. A	3	21592	551	97.50%
pred. B	13	424	1310	74.99%
class recall	99.67%	97.89%	32.43%	

ภาพที่ 4. 16 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดลอง 30,947 คน

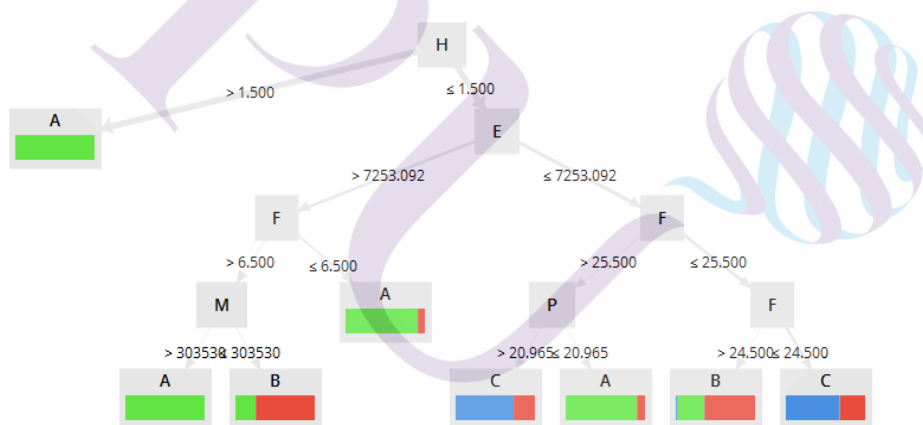
ผลการวัดประสิทธิภาพของแบบจำลองจากข้อมูลที่ใช้ทดสอบ จำนวน 74 คน ได้ความแม่นยำ ร้อยละ 97.30 สามารถทำนายประเภท A ถูก จำนวน 22 คน ทำนายผิดเป็นประเภท B จำนวน 2 คน และกลุ่มคนประเภท C ทำนายถูก จำนวน 50 คน ดังภาพที่ 4.20

accuracy: 97.30%

	true A	true B	true C	class precision
pred. A	22	2	0	91.67%
pred. B	0	0	0	0.00%
pred. C	0	0	50	100.00%
class recall	100.00%	0.00%	100.00%	

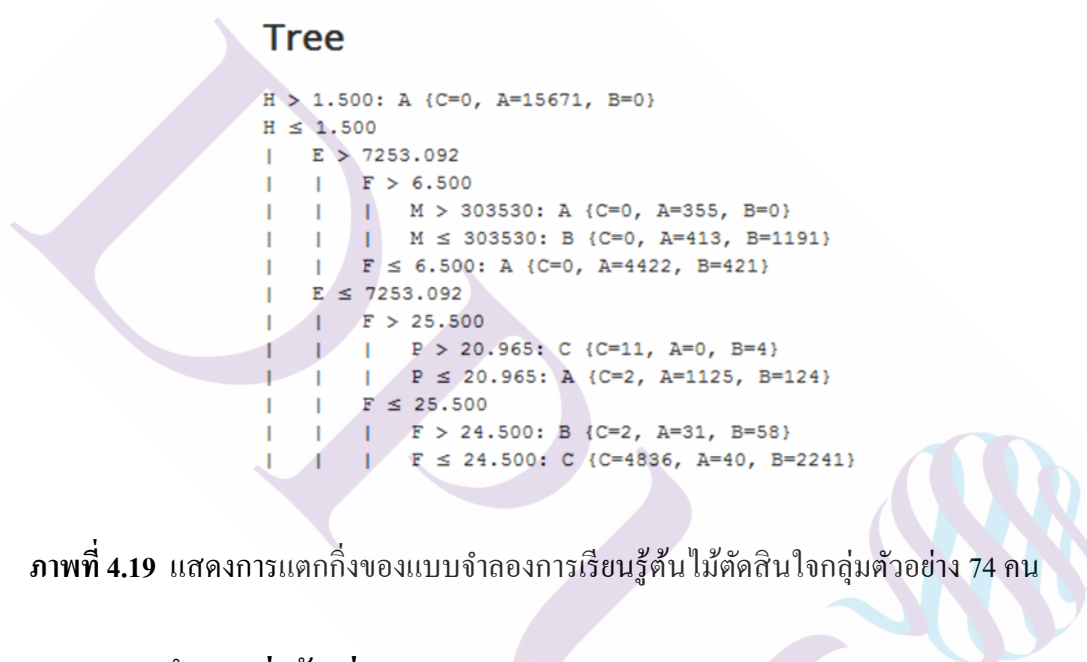
ภาพที่ 4.17 ผลการวัดประสิทธิภาพของแบบจำลองข้อมูลที่ใช้ทดสอบ 74 คน

จากภาพที่ 4.21 ตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุด คือตัวแปร H (จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี) ถัดลงมาคือตัวแปร E (ค่ายาเฉลี่ยต่อครั้ง/ปี) และตัวแปร F (ความถี่ที่ไปใช้บริการ/ปี) มีความสัมพันธ์กับกลุ่มลำดับถัดมา



ภาพที่ 4.18 แสดงการแบ่งลำดับชั้นของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ 74 คน

จากภาพที่ 4.22 ตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุดของกลุ่มตัวอย่าง 74 คน คือ H (จำนวนสถานพยาบาลที่ไปใช้บริการ/ปี) มีค่ามากกว่า 1.5 แห่ง/ปี ทำนายเป็นประเภท A จำนวน 15,671 คน ลำดับถัดมาคือ E (ค่ายาเฉลี่ยต่อครั้ง/ปี) มีค่ามากกว่า 7,253 บาท และค่า F (ความถี่ที่ไปใช้บริการ/ปี) มีค่าน้อยกว่าเท่ากับ 6.5 ครั้ง/ปี ทำนายเป็นประเภท A จำนวน 4,422 คน ทำนายผลเป็นประเภท B จำนวน 421 คน และค่า F (ความถี่ที่ไปใช้บริการ/ปี) มีค่ามากกว่า 6.5 ครั้ง/ปี แยกกิ่งต่อมาที่ค่า M (ค่ายากลับบ้าน/ปี) มีค่ามากกว่า 303,530 บาท ทำนายผลเป็นประเภท A จำนวน 355 คน เป็นต้น



ภาพที่ 4.19 แสดงการแตกกิ่งของแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจกลุ่มตัวอย่าง 74 คน

4.15 ผลการทำนายกลุ่มตัวอย่าง

จากตารางที่ 4.9 แสดงผลการทำนายรายบุคคลจากข้อมูลชุดทดสอบ (ผู้ที่ถูกระงับสิทธิ์) จำนวน 24 คน ที่ได้จากโปรแกรม RapidMiner Studio พบว่า มี 22 คนที่ทำนายถูก และทำนายผิด จำนวน 2 คน

ตารางที่ 4.9 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 24 คน

ID	R	F	M	H	P	E	Confidence (A)	Confidence (B)	Confidence (C)	ประเภท	ผลการ ทำนาย
id1	21	75	41,269	15	4	550	1	0	0	A	A
id2	179	20	140,877	5	9	7,044	1	0	0	A	A
id3	165	46	342,098	5	11	7,437	1	0	0	A	A
id4	178	30	99,549	5	6	3,318	1	0	0	A	A
id5	158	82	79,519	7	7	970	1	0	0	A	A
id6	172	18	183,865	4	11	10,215	1	0	0	A	A
id7	196	21	88,262	3	8	4,203	1	0	0	B	A
id8	161	28	77,510	5	61	2,768	1	0	0	B	A
id9	43	41	237,054	5	9	5,782	1	0	0	A	A
id10	8	52	512,012	5	9	9,846	1	0	0	A	A
id11	7	94	49,817	6	7	530	1	0	0	A	A
id12	30	36	325,467	4	9	9,041	1	0	0	A	A
id13	288	30	125,999	6	10	4,200	0.7	0.3	0	A	A
id14	37	50	131,617	11	8	2,632	1	0	0	A	A
id15	15	53	121,531	5	7	2,293	1	0	0	A	A
id16	7	60	134,988	6	6	2,250	1	0	0	A	A
id17	4	100	46,437	6	4	464	1	0	0	A	A
id18	8	47	405,994	5	7	8,638	1	0	0	A	A
id19	21	117	408,491	7	3	3,491	1	0	0	A	A
id20	89	109	346,727	41	4	3,181	1	0	0	A	A
id21	5	68	41,468	4	5	610	1	0	0	A	A
id22	130	36	131,320	5	15	3,648	1	0	0	A	A
id23	23	77	267,474	21	5	3,474	1	0	0	A	A
id24	4	99	63,346	9	4	640	1	0	0	A	A

จากตารางที่ 4.10 แสดงผลการทำนายรายบุคคลจากข้อมูลชุดทดสอบ (ผู้ที่ถูกระงับสิทธิ์) จำนวน 74 คน ที่ได้จากโปรแกรม RapidMiner Studio พบว่าทำนายถูก 22 คน เป็นประเภท A และทำนายประเภท B เป็น A ผิดจำนวน 2 คน และทำนายประเภท C ถูกทั้งหมด จำนวน 50 คน

ตารางที่ 4.10 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 74 คน

ID	R	F	M	H	P	E	Confidence (A)	Confidence (B)	Confidence (C)	ประเภท	ผลการ ทำนาย
id1	179	20	140,877	5	9	7,044	1	0	0	A	A
id2	43	41	237,054	5	9	5,782	1	0	0	A	A
id3	15	53	121,531	5	7	2,293	1	0	0	A	A
id4	165	46	342,098	5	11	7,437	1	0	0	A	A
id5	8	52	512,012	5	9	9,846	1	0	0	A	A
id6	7	60	134,988	6	6	2,250	1	0	0	A	A
id7	178	30	99,549	5	6	3,318	1	0	0	A	A
id8	158	82	79,519	7	7	970	1	0	0	A	A
id9	7	94	49,817	6	7	530	1	0	0	A	A
id10	4	100	46,437	6	4	464	1	0	0	A	A
id11	5	68	41,468	4	5	610	1	0	0	A	A
id12	172	18	183,865	4	11	10,215	1	0	0	A	A
id13	30	36	325,467	4	9	9,041	1	0	0	A	A
id14	8	47	405,994	5	7	8,638	1	0	0	A	A
id15	130	36	131,320	5	15	3,648	1	0	0	A	A
id16	288	30	125,999	6	10	4,200	1	0	0	A	A
id17	21	117	408,491	7	3	3,491	1	0	0	A	A
id18	89	109	346,727	41	4	3,181	1	0	0	A	A
id19	23	77	267,474	21	5	3,474	1	0	0	A	A
id20	37	50	131,617	11	8	2,632	1	0	0	A	A
id21	21	75	41,269	15	4	550	1	0	0	A	A
id22	4	99	63,346	9	4	640	1	0	0	A	A
id23	196	21	88,262	3	8	4,203	1	0	0	B	A
id24	161	28	77,510	5	61	2,768	1	0	0	B	A
id25	108	8	35,331	1	21	4,416	0	0.3	0.7	C	C
id26	16	15	33,533	1	18	2,236	0	0.3	0.7	C	C
id27	46	9	57,919	1	30	6,435	0	0.3	0.7	C	C
id28	38	12	59,880	1	23	4,990	0	0.3	0.7	C	C

ตารางที่ 4.10 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 74 คน (ต่อ)

ID	R	F	M	H	P	E	Confidence (A)	Confidence (B)	Confidence (C)	ประเภท	ผลการ ทำนาย
id29	6	12	40,762	1	26	3,397	0	0.3	0.7	C	C
id30	27	8	36,038	1	37	4,505	0	0.3	0.7	C	C
id31	9	9	49,955	1	37	5,551	0	0.3	0.7	C	C
id32	11	17	39,920	1	18	2,348	0	0.3	0.7	C	C
id33	84	17	41,323	1	19	2,431	0	0.3	0.7	C	C
id34	159	7	41,100	1	17	5,871	0	0.3	0.7	C	C
id35	81	9	45,412	1	33	5,046	0	0.3	0.7	C	C
id36	5	13	85,501	1	24	6,577	0	0.3	0.7	C	C
id37	108	10	60,169	1	23	6,017	0	0.3	0.7	C	C
id38	23	15	30,266	1	21	2,018	0	0.3	0.7	C	C
id39	76	9	30,661	1	29	3,407	0	0.3	0.7	C	C
id40	59	6	34,200	1	37	5,700	0	0.3	0.7	C	C
id41	57	10	51,605	1	24	5,161	0	0.3	0.7	C	C
id42	4	17	35,655	1	19	2,097	0	0.3	0.7	C	C
id43	62	10	49,083	1	25	4,908	0	0.3	0.7	C	C
id44	47	15	32,579	1	17	2,172	0	0.3	0.7	C	C
id45	61	8	32,725	1	52	4,091	0	0.3	0.7	C	C
id46	90	6	36,324	1	40	6,054	0	0.3	0.7	C	C
id47	24	17	38,568	1	19	2,269	0	0.3	0.7	C	C
id48	12	15	52,865	1	21	3,524	0	0.3	0.7	C	C
id49	4	10	39,175	1	25	3,918	0	0.3	0.7	C	C
id50	61	5	34,605	1	49	6,921	0	0.3	0.7	C	C
id51	72	12	42,686	1	29	3,557	0	0.3	0.7	C	C
id52	82	10	37,465	1	26	3,747	0	0.3	0.7	C	C
id53	93	7	33,630	1	26	4,804	0	0.3	0.7	C	C
id54	45	12	34,550	1	23	2,879	0	0.3	0.7	C	C
id55	155	8	44,898	1	24	5,612	0	0.3	0.7	C	C
id56	130	8	32,752	1	22	4,094	0	0.3	0.7	C	C

ตารางที่ 4.10 ผลการทำนายข้อมูลชุดทดสอบกลุ่มตัวอย่าง 74 คน (ต่อ)

ID	R	F	M	H	P	E	Confidence (A)	Confidence (B)	Confidence (C)	ประเภท	ผลการ ทำนาย
id57	47	9	37,300	1	29	4,144	0	0.3	0.7	C	C
id58	34	16	47,755	1	23	2,985	0	0.3	0.7	C	C
id59	8	18	36,261	1	17	2,015	0	0.3	0.7	C	C
id60	34	16	32,689	1	19	2,043	0	0.3	0.7	C	C
id61	144	18	65,290	1	24	3,627	0	0.3	0.7	C	C
id62	47	12	36,105	1	23	3,009	0	0.3	0.7	C	C
id63	8	7	40,871	1	42	5,839	0	0.3	0.7	C	C
id64	23	18	52,766	1	32	2,931	0	0.3	0.7	C	C
id65	48	18	84,232	1	19	4,680	0	0.3	0.7	C	C
id66	16	11	34,627	1	23	3,148	0	0.3	0.7	C	C
id67	26	9	65,007	1	32	7,223	0	0.3	0.7	C	C
id68	58	5	31,972	1	48	6,394	0	0.3	0.7	C	C
id69	44	10	53,939	1	29	5,394	0	0.3	0.7	C	C
id70	17	18	62,403	1	16	3,467	0	0.3	0.7	C	C
id71	125	6	41,334	1	19	6,889	0	0.3	0.7	C	C
id72	33	12	30,773	1	22	2,564	0	0.3	0.7	C	C
id73	68	5	30,065	1	42	6,013	0	0.3	0.7	C	C
id74	5	10	39,545	1	33	3,955	0	0.3	0.7	C	C

4.16 การทดสอบค่าสถิติเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม

การทดสอบสมมติฐานด้วยวิธี t-test นั้นเป็นการทดสอบสมมติฐานที่ใช้บ่อยที่สุดในงานวิจัย โดยเปรียบเทียบความแตกต่างของตัวแปรผลลัพธ์ระหว่าง 2 กลุ่ม การทดสอบสมมติฐานด้วยวิธี t-test ควรมีการกระจายตัวแปรแบบปกติ (normal distribution) แต่เนื่องด้วยข้อมูลประชากรมีขนาดใหญ่พอ จึงใช้ทฤษฎีแนวโน้มนำเข้าสู่ศูนย์กลาง (Central limit theorem) มาอธิบายได้ว่าเมื่อขนาดของกลุ่มตัวอย่างที่ถูกสุ่มเลือกมาจากประชากรมีขนาดใหญ่พอ ค่าประมาณค่าเฉลี่ยของประชากรจากกลุ่มตัวอย่างนั้น (mean estimator) จะมีการกระจายตัวแบบปกติและใกล้เคียงกับค่าเฉลี่ยที่แท้จริงในประชากร (ดิเรก ถิรมมธุรสกุล, 2554 น. 512)

จากตารางที่ 4.11 แสดงค่าสถิติระหว่าง 2 กลุ่ม คือกลุ่มพฤติกรรมผิดปกติและกลุ่มพฤติกรรมปกติ ความแตกต่างระหว่าง 2 กลุ่มเปรียบเทียบได้จากค่าเฉลี่ย (Mean) ระหว่างตัวแปรและค่าเบี่ยงเบนมาตรฐาน (SD)

กลุ่มตัวอย่างที่นำมาเปรียบเทียบมีทั้งหมด 74 คน แบ่งเป็นกลุ่มตัวอย่างพฤติกรรมผิดปกติ 24 คน และกลุ่มพฤติกรรมปกติ 50 คน โดยค่าเฉลี่ยของกลุ่มพฤติกรรมผิดปกติตัวแปร R, F, M, H มีค่าเฉลี่ยสูงกว่ากลุ่มพฤติกรรมปกติ เช่น ค่าเฉลี่ยความถี่ที่ใช้บริการของกลุ่มพฤติกรรมผิดปกติ (F) 57 ครั้ง/ปี สูงกว่ากลุ่มพฤติกรรมปกติมีค่าเฉลี่ย 11 ครั้ง/ปี และค่าเฉลี่ยของจำนวนสถานพยาบาลที่ใช้บริการ/ปี (H) ของกลุ่มพฤติกรรมผิดปกติ จำนวน 8 แห่ง/ปี สูงกว่ากลุ่มพฤติกรรมปกติ 1 แห่ง/ปี เป็นต้น

ตารางที่ 4.11 แสดงค่าสถิติจากผลการทำนายนกลุ่มตัวอย่าง

ตัวแปร	พฤติกรรมผิดปกติ n= 24		พฤติกรรมปกติ n= 50		รวมทั้งสิ้น n= 74	
	ค่าเฉลี่ย	SD	ค่าเฉลี่ย	SD	ค่าเฉลี่ย	SD
R = จำนวนวันล่าสุดที่ใช้บริการ/ปี	81.20	84.79	52.8	41.61	62.01	60.05
F = ความถี่ที่ใช้บริการ/ปี	57.87	29.98	11.28	4.08	26.39	27.87
M = จำนวนเงินที่เบิกจ่าย/ปี	183,445.5	137,731.3	43,481.38	12,967.9	88,875.14	102,183.2
H = จำนวนสถานพยาบาลที่ใช้บริการ/ปี	8.12	8.04	1	0	3.31	5.62
P = ช่วงห่างระหว่างวันที่ใช้บริการ/ปี	9.54	11.31	26.92	8.90	21.28	12.67
E = ค่ายากลับบ้าน/ครั้ง	4,051.04	3117.74	4,261.46	1,537.45	4,193.21	2,158.48

4.17 การทดสอบทางสถิติเพื่อเปรียบเทียบกลุ่มตัวอย่างด้วย T-Test

การทดสอบทางสถิติเพื่อเปรียบเทียบกลุ่มตัวอย่างทั้ง 2 กลุ่มจากการคำนวณทางสถิติ t-test เพื่อหาค่าสถิติ t (t-value หรือ t statistics) ซึ่งแสดงถึงผลลัพธ์ของกลุ่มตัวอย่าง จากภาพที่ 4.23 แสดงการเปรียบเทียบค่าเฉลี่ยระหว่าง 2 กลุ่มพฤติกรรมผิดปกติและพฤติกรรมปกติจากตัวแปร F เปรียบเทียบ

ระหว่าง 2 กลุ่ม กำหนดความเชื่อมั่น 95% ของตัวแปรที่มีค่า p-value ≤ 0.05 แตกต่างกันอย่างมีนัยสำคัญทางสถิติ sig

```

- ttest f ,by(grp)
Two-sample t test with equal variances

```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	24	57.875	6.128858	29.98596	45.21304	70.53696
2	50	11.28	.5771429	4.881816	10.12819	12.43981
combined	74	26.39189	3.239874	27.87045	19.93483	32.84895
diff		46.595	4.298866		38.04132	55.14868

```

diff = mean(1) - mean(2)
Ho: diff = 0
Ha: diff < 0
Pr(T < t) = 1.0000

```

P-value ≤ 0.05

```

t = 10.8591
degrees of freedom = 72
Ha: diff != 0
Pr(|T| > |t|) = 0.0000
Ha: diff > 0
Pr(T > t) = 0.0000

```

ภาพที่ 4. 20 ตัวอย่างการทดสอบค่า F ด้วยสถิติ t-test

จากตารางที่ 4.12 เป็นการเปรียบเทียบความแตกต่างระหว่าง 2 กลุ่ม คือกลุ่มพฤติกรรมผิดปกติจำนวน 24 คน และกลุ่มพฤติกรรมปกติ จำนวน 50 คน เช่น ค่าเฉลี่ยความถี่ที่ไปใช้บริการ/ปี (F) เท่ากับ 57.87 ครั้ง/ปี และ 11.28 ครั้ง/ปี ความถี่เฉลี่ยที่ไปใช้บริการของกลุ่มพฤติกรรมผิดปกติ มีค่ามากกว่ากลุ่มพฤติกรรมปกติอยู่ 46.59 ครั้ง/ปี ช่วงความเชื่อมั่นของค่าแตกต่างของค่าเฉลี่ย 95% Conf. Interval คือช่วงระหว่าง 38.04 ถึง 55.14 ซึ่งความถี่เฉลี่ยที่ไปใช้บริการของกลุ่มพฤติกรรมผิดปกติและกลุ่มพฤติกรรมปกติ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ (p-value=0.000)

การเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มพฤติกรรมผิดปกติกับกลุ่มพฤติกรรมปกติแตกต่างกันอย่างมีนัยสำคัญระดับสูงมาก (p-value=0.0000) ได้แก่ ตัวแปร F, M, H, P ส่วนค่า p-value ≥ 0.05 แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติไม่ sig คือ กลุ่มพฤติกรรมผิดปกติกับกลุ่มพฤติกรรมปกติไม่ต่างกัน ได้แก่ ตัวแปร (R) จำนวนวันล่าสุดที่ไปใช้บริการ/ปี มีค่าเฉลี่ยเท่ากับ 81.20 วัน และ 52.8 วัน จำนวนวันล่าสุดที่ไปใช้บริการเฉลี่ยของกลุ่มพฤติกรรมผิดปกติมีค่ามากกว่ากลุ่มพฤติกรรมปกติ 28.4 วัน ช่วงความเชื่อมั่นของค่าแตกต่างของค่าเฉลี่ย 95% Conf. Interval คือช่วงระหว่าง 0.77 ถึง 57.59 ซึ่งแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (p-value = 0.0562) และตัวแปร (E) ค่ายากกลับบ้าน/ครั้ง ที่มีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (p-value = 0.6976)

ดังนั้นตัวแปรที่มีอำนาจในการทำนายแบบจำลอง คือตัวแปร F, M, H, P สามารถบอกลักษณะพฤติกรรมของกลุ่มตัวอย่างได้และยังสะท้อนให้เห็นกลุ่มพฤติกรรมผิดปกติและกลุ่มพฤติกรรมปกติได้อย่างชัดเจน

ตารางที่ 4.12 การเปรียบเทียบความแตกต่างระหว่างกลุ่มตัวอย่าง 2 กลุ่ม ด้วยสถิติ t-test

ตัวแปร	ค่าเฉลี่ย	Mean difference	95% Conf.	P-value
R = จำนวนวันล่าสุดที่ใช้บริการ/ปี	1) 81.20 2) 52.8	28.40	1) 0.77 2) 57.59	0.0562
F = ความถี่ที่ใช้บริการ/ปี	1) 57.87 2) 11.28	46.59	1) 38.04 2) 55.14	0.0000
M = จำนวนเงินที่เบิกจ่าย/ปี	1) 183,445.5 2) 43,481.24	139,964.2	1) 101,066.3 2) 178,862.2	0.0000
H = จำนวนสถานพยาบาลที่ใช้บริการ/ปี	1) 8.12 2) 1	7.12	1) 4.87 2) 9.37	0.0000
P = ช่วงห่างระหว่างวันที่ไปใช้บริการ/ปี	1) 9.54 2) 26.92	-17.37	1) -22.19 2) -12.55	0.0004
E = ค่าชากลับบ้าน/ครั้ง	1) 4,051.04 2) 4,261.40	-210.36	1) -1,285.15 2) 8,64.42	0.6976

บทที่ 5

สรุปผลการศึกษา

5.1 สรุปผล

การศึกษานี้มีวัตถุประสงค์ เพื่อนำเสนอวิธีการใหม่ ในการช่วยคัดกรองผู้ป่วยที่มีพฤติกรรมผิดปกติ โดยการใช้เทคนิคการแบ่งกลุ่มและเทคนิคดาต้าไมน์นิ่งเพื่อหาคนที่มีความผิดปกติและพฤติกรรมปกติ แล้วสร้างเป็นแบบจำลองขึ้นมาใช้ตรวจสอบได้ง่ายขึ้น

ผลการศึกษากระบวนการดาต้าไมน์นิ่งการเบิกจ่ายเงินสวัสดิการรักษายาบาลข้าราชการผู้ป่วยนอก ในปีงบประมาณ 2556 จากข้อมูลการเบิกจ่ายเงินค่ารักษายาบาลระบบจ่ายตรงกรมบัญชีกลาง โดยมีสถานพยาบาลจำนวน 168 แห่ง เรียกเก็บค่ารักษายาบาลผ่านระบบเบิกจ่ายตรงตัวแปรที่ใช้หารูปแบบพฤติกรรม 6 ตัว คือ R, F, M, H, P, E โดยเลือกผู้ป่วยนอกที่มีการเบิกจ่ายเงินค่ายา นำกลับไปใช้ที่บ้าน (หมวด 4) มากกว่า 30,000 บาท/ปี และไม่ใช้ผู้ป่วยที่ใช้ยามะเร็ง ทำการแบ่งกลุ่มด้วยวิธี RFMHPE และการแบ่งกลุ่มด้วยเคมีน แล้วนำมาอินเตอร์เซกกันเพื่อกำหนดคลาสแยกประเภทกลุ่มพฤติกรรมผู้ไปใช้บริการ สามารถแบ่งกลุ่มได้ 3 กลุ่ม คือประเภท A กลุ่มที่น่าสนใจเป็นพิเศษ ประเภท B คือ กลุ่มที่น่าสนใจรองลงมา และประเภท C คือ กลุ่มพฤติกรรมปกติ

หลังจากกำหนดคลาสแล้วนำมาทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจกับกลุ่มตัวอย่างที่ใช้ทดสอบจำนวน 24 คน ตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุด คือตัวแปร H, F, R ต่อมาได้เพิ่มกลุ่มตัวอย่างที่ใช้ทดสอบเป็น 74 คน โดยสุ่มผู้ป่วยพฤติกรรมปกติประเภท C จำนวน 50 คน นำมารวมกับ 24 คน แล้วทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ พบว่าตัวแปรที่มีความสัมพันธ์กับกลุ่มมากที่สุด คือตัวแปร H, E, F

ทำการเปรียบเทียบความแตกต่างระหว่าง 2 กลุ่ม คือกลุ่มพฤติกรรมผิดปกติและกลุ่มพฤติกรรมปกติ มีแตกต่างกันอย่างมีนัยสำคัญระดับสูงมาก ($p\text{-value}=0.0000$) ดังนั้นตัวแปรที่มีอำนาจในการทำนายแบบจำลอง คือตัวแปร F, M, H, P สามารถบอกลักษณะพฤติกรรมของกลุ่มตัวอย่างได้ดี

จากตารางที่ 5.1 แสดงผู้ใช้บริการผู้ป่วยนอกทั้งหมด 2.7 ล้านคน คิดเป็นร้อยละ 100 มีการเบิกจ่ายเงินค่ายากลับบ้านจำนวน 36,652 ล้านบาท ผู้ใช้บริการที่มีการเบิกจ่ายเงินค่ายากลับบ้านเกิน 30,000 บาท/ปี จำนวน 31,001 คน คิดเป็นร้อยละ 1.13 ของผู้มารับบริการทั้งหมด เบิกค่ายากลับบ้าน ร้อยละ 6.97 โดยแบ่งเป็นกลุ่มที่น่าสนใจเป็นพิเศษ ร้อยละ 0.8 กลุ่มที่น่าสนใจรองลงมา ร้อยละ 0.15 และกลุ่มพฤติกรรมปกติ ร้อยละ 0.18

ตารางที่ 5.1 สรุปผลการศึกษา

รายการ	คน	ร้อยละ	เบิกค่า รักษาพยาบาล	ร้อยละ
1. ผู้ใช้บริการผู้ป่วยนอกทั้งหมด	2,751,425	100%	36,652,250,202.92	100%
2. ผู้ใช้บริการผู้ป่วยนอกที่มีการเบิกค่ายาเกิน 30,000 บาท/ปี	31,001	1.13%	2,554,186,482.50	6.97%
2.1 ประเภท A กลุ่มที่น่าสนใจเป็นพิเศษ	22,098	0.80%	2,046,182,829.6	5.58%
2.2 ประเภท B กลุ่มที่น่าสนใจรองลงมา	4,045	0.15%	285,587,171.5	0.78%
2.3 ประเภท C กลุ่มพฤติกรรมปกติ	4,858	0.18%	222,416,481.3	0.61%

จากตารางที่ 5.2 แสดงการแบ่งกลุ่มพฤติกรรมรายช่วงอายุ จำนวนคนสูงสุดในช่วงอายุ 70-79 ปี พบพฤติกรรมผิดปกติจำนวน 6,898 คน และพฤติกรรมปกติจำนวน 1,413 คน อย่างไรก็ตามปัจจัยที่เป็นตัวกำหนดรูปแบบพฤติกรรมมาจากตัวแปร R, F, H, E ที่ได้จากทดสอบแบบจำลองการเรียนรู้ต้นไม้ตัดสินใจ ทั้งนี้ควรทำการตรวจสอบในรายละเอียดเชิงลึกต่อไป

ตารางที่ 5.2 แสดงการแบ่งกลุ่มรายช่วงอายุรายประเภท (คน)

ช่วงอายุ	พฤติกรรมผิดปกติ	พฤติกรรมปกติ	รวม
0-9	185	20	205
10-19	384	46	430
20-29	79	11	90
30-39	613	86	699
40-49	1,850	290	2,140
50-59	5,484	949	6,433
60-69	6,603	1,217	7,820
70-79	6,898	1,413	8,311
80-89	3,617	745	4,362
90-99	410	76	486
100-109	16	5	21
รวม	26,139	4,858	30,997

จากตารางที่ 5.3 แสดงพฤติกรรมรายสิทธิการเบิกจ่ายเงินค่ารักษาพยาบาล พบว่าเจ้าของสิทธิมาใช้บริการมากที่สุดและพบความผิดปกติสูงสุดจำนวน 4,023 คน ในช่วงอายุ 50-59 ปี รองลงมาคือกลุ่มบิดา/มารดาและกลุ่มคู่สมรส ตามลำดับ ทั้งนี้ควรทำการตรวจสอบในรายละเอียดเชิงลึกต่อไป

ตารางที่ 5.3 แสดงรายกลุ่มประเภทรายสิทธิการรักษา

ช่วงอายุ	เจ้าของสิทธิ		บิดา/มารดา		บุตร		คู่สมรส		รวม
	ผิดปกติ	ปกติ	ผิดปกติ	ปกติ	ผิดปกติ	ปกติ	ผิดปกติ	ปกติ	
0-9	21	5	7	2	154	13	3	.	205
10-19	384	46	.	.	430
20-29	55	6	.	.	16	2	8	3	90
30-39	489	67	.	.	9	1	115	18	699
40-49	1,405	220	43	7	.	.	402	63	2,140
50-59	4,023	744	607	61	1	.	853	144	6,433
60-69	3,641	677	2,146	342	.	.	816	198	7,820
70-79	2,612	502	3,816	790	.	.	470	121	8,311
80-89	1,070	251	2,456	475	.	.	91	19	4,362
90-99	99	13	309	63	.	.	2	.	486
100-109	4	1	11	4	.	.	1	.	21
รวม	13,419	2,486	9,395	1,744	564	62	2,761	566	30,997

5.2 ข้อเสนอแนะ

การเลือกรูปแบบเพื่อทำการวิเคราะห์เป็นสิ่งสำคัญ และขึ้นอยู่กับวัตถุประสงค์ของแต่ละงาน อาทิเช่น ผลวิเคราะห์ครั้งนี้มุ่งหาบุคคลที่มีพฤติกรรมการไปใช้บริการไม่สมเหตุสมผล ดังนั้นรูปแบบที่สนใจ คือ เป็นผู้ที่ไปใช้บริการบ่อย ไปสถานพยาบาลหลายแห่งในระยะเวลาที่ไปไม่ห่างกันมาก และมีค่าเฉลี่ยขาดต่อครั้งสูง การศึกษานี้จึงมุ่งหาคนที่มีพฤติกรรมผิดปกติจากกลุ่ม แต่ก็ไม่ได้หมายความว่าทุกคนที่อยู่ในกลุ่มจะมีโอกาสไปใช้บริการไม่สมเหตุสมผล ซึ่งต้องทำการตรวจสอบเชิงลึกในรายบุคคลต่อไป ดังนั้นการนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้สามารถช่วยค้นหากลุ่มคนที่ผิดปกติได้อย่างรวดเร็วและประหยัดเวลา ช่วยลดภาระการทำงานของเจ้าหน้าที่ และองค์กรสามารถเฝ้าระวังบุคคลที่มีพฤติกรรมผิดปกติในลักษณะเดียวกันและช่วยลดความเสียหายให้น้อยลงได้

เนื่องด้วยฐานข้อมูลได้มาจากหลายแหล่ง การนำฐานข้อมูลขนาดใหญ่มาเชื่อมโยงกัน ต้องใช้เครื่องคอมพิวเตอร์ที่มีประสิทธิภาพสูง และด้วยลักษณะข้อมูลที่ต้องเชื่อมโยงกันทุกฐาน ถึงนำข้อมูลมาใช้วิเคราะห์ได้ครบถ้วน เช่น ฐานข้อมูล CGD, CHI Bill การเชื่อมกันต้องหาตัวแปรเชื่อมในแต่ละตาราง ซึ่งต้องใช้ความเข้าใจในข้อมูลและความชำนาญพิเศษในการจัดการฐานข้อมูลขนาดใหญ่ ในอนาคตข้อมูลเหล่านี้หากมีตัวแปรเชื่อมเข้าหากันได้ จะทำให้การจัดการฐานข้อมูลง่ายขึ้นและช่วยลดความผิดพลาดในการเชื่อมต่อฐานข้อมูลได้



DPU

ภาคผนวก





ภาคผนวก ก

6.1.1 ผู้ป่วยนอก CGD

ลำดับ	ข้อมูล	ตัวแปร	คำอธิบาย
1	งวดเบิก	lot	5 หลัก
2	วันที่เบิกในระบบ GFMS	dategf	YYYYMMDD (คศ.)
3	วันที่ให้บริการ	datevisit	YYYYMMDD (คศ.)
4	จังหวัด	province	
5	รหัสหน่วยงานต้นสังกัด	codewk	
6	รหัสเจ้าของสิทธิ์	id	
7	ประเภทสิทธิ์	codetype	1.ข้าราชการ 2.บำนาญ 3.ลูกจ้างประจำ
8	ความสัมพันธ์ของผู้มีสิทธิ์ต่อผู้ใช้สิทธิ์	relation	1. ผู้มีสิทธิ์ 2. บิดา/มารดา 3. บุตร 4. คู่สมรส 5. บุตรของคู่สมรสเดิม
9	รหัสผู้ใช้สิทธิ์	iduse	
10	วันเกิด	dob	YYYYMMDD (คศ.)
11	เพศ	sex	1. ชาย 2. หญิง
12	รหัสสถานพยาบาล	hcode	
13	จำนวนเงินที่เบิก	amount	
14	เลขอนุวัติ	codeout	
15	วันที่ออกเลขอนุวัติ	dateout	YYYYMMDD (คศ.)
16	เลขที่อ้างอิงกับระบบ GFMS	codegf	

6.1.2 ผู้ป่วยนอก OPDATA

ลำดับ	ข้อมูล	ตัวแปร	คำอธิบาย
1	TransactionID	tsid	
2	งวดเบิก	lot	
3	วันที่เบิกในระบบ GFMIS	dategf	YYYYMMDD (คศ.)
4	วันที่ให้บริการ	datevisit	YYYYMMDD (คศ.)
5	จังหวัด	province	
6	รหัสหน่วยงานต้นสังกัด	codewk	
7	รหัสเจ้าของสิทธิ์	id	
8	ประเภทสิทธิ์	codetype	1.ข้าราชการ/ลูกจ้างประจำ 2. บำนาญ 3. พนักงานของรัฐ 4. ผู้อาศัยสิทธิ์
9	ความสัมพันธ์ของผู้มีสิทธิ์ต่อผู้ใช้สิทธิ์	relation	1. เจ้าตัว 2. บุตร 3. บิดา/มารดา 4. คู่สมรส
10	รหัสผู้ใช้สิทธิ์	iduse	
11	วันเกิด	dob	YYYYMMDD (คศ.)
12	เพศ	sex	1. ชาย 2. หญิง
13	รหัสสถานพยาบาล	hcode	
14	จำนวนเงินที่เบิก	amount	
15	เลขอนุมัติ	codeout	
16	วันที่ออกเลขอนุมัติ	dateout	YYYYMMDD (คศ.)
17	เลขที่อ้างอิงกับระบบ GFMIS	codegf	
18	ประเภทการเบิก	type	B: เบิกผู้ป่วยนอกทั่วไป H: เบิกค่าฟอกเลือดล้างไต

			V: เบิกค่ารถส่งต่อ
			S: เบิกค่ารังสีรักษาผู้ป่วยมะเร็ง
ลำดับ	ข้อมูล	ตัวแปร	คำอธิบาย
19	ค่าจัดทำข้อมูลยา	paydrug	

6.1.3 ผู้ป่วยนอก OPBILL

ลำดับ	ข้อมูล	ตัวแปร	คำอธิบาย
1	TransctionID	tsid	
2	billamt		จำนวนเงินรวม
3	billamt1	room	1. ค่าห้อง
4	billamt2	prosthesis	2. อวัยวะเทียม/อุปกรณ์ในการบำบัดรักษา
5	billamt3	drug_hospital	3. ยาและสารอาหารทางเส้นเลือดที่ใช้ใน รพ
6	billamt4	drug_home	4. ยาที่นำไปใช้ที่บ้าน
7	billamt5	non_pharmaceutical	5. เวชภัณฑ์ที่ไม่ใช่ยา
8	billamt6	blood	6. บริการโลหิตและส่วนประกอบของโลหิต
9	billamt7	pathology	7. ตรวจวินิจฉัยทางเทคนิคและพยาธิวิทยา
10	billamt8	xray	8. ตรวจวินิจฉัยทางรังสี
11	billamt9	other_diag	9. ตรวจวินิจฉัยอื่นๆ
12	billamtA	equipment	10. อุปกรณ์ของใช้และเครื่องมือทางการแพทย์
13	billamtB	anesthesia	11. ทำหัตถการ และบริการวิสัญญี
14	billamtC	nurse	12. ค่าบริการทางการพยาบาล
15	billamtD	dental	13. บริการทางทันตกรรม
16	billamtE	physical_medicine	14. บริการทางกายภาพบำบัดและเวชกรรมฟื้นฟู
17	billamtF	acupuncture	15. บริการฝังเข็ม/ผู้ประกอบโรคศิลปะ
18	billamtG	other1	16. บริการอื่นๆ
19	billpaid	จำนวนเงินที่ผู้ป่วยจ่าย	
20	billpaid1	p_room	1. ค่าห้อง

ลำดับ	ข้อมูล	ตัวแปร	คำอธิบาย
21	billpaid2	p_prosthesis	2. อวัยวะเทียม/อุปกรณ์ในการบำบัดรักษา
22	billpaid3	p_drug_hospital	3. ยาและสารอาหารทางเส้นเลือดที่ใช้ใน รพ
23	billpaid4	p_drug_homeP	4. ยาที่นำไปใช้ที่บ้าน
24	billpaid5	p_non_pharmaceutical	5. เวชภัณฑ์ที่ไม่ใช่ยา
25	billpaid6	p_blood	6. บริการโลหิตและส่วนประกอบของโลหิต
26	billpaid7	p_pathology	7. ตรวจวินิจฉัยทางเทคนิคและพยาธิวิทยา
27	billpaid8	p_xray	8. ตรวจวินิจฉัยทางรังสี
28	billpaid9	p_other_diag	9. ตรวจวินิจฉัยอื่นๆ
29	billpaidA	p_equipment	10. อุปกรณ์ของใช้และเครื่องมือทางการแพทย์
30	billpaidB	p_anesthesia	11. ทำหัตถการ และบริการวิสัญญี
31	billpaidC	p_nurse	12. ค่าบริการทางการพยาบาล
32	billpaidD	p_dental	13. บริการทางทันตกรรม
33	billpaidE	p_physical_medicine	14. บริการทางกายภาพบำบัดและเวชกรรมฟื้นฟู
34	billpaidF	p_acupuncture	15. บริการฝังเข็ม/ผู้ประกอบโรคศิลปะ
35	billpaidG	p_other1	16. บริการอื่นๆ

ตารางที่ 6.1 รูปแบบที่พบจากการแบ่งกลุ่มข้อมูลด้วยวิธีเซกเมนต์เทชั่น

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
1	111112	6	20	8	34,322	1	16	4,485	RP
2	111113	2	12	1	34,288	1	-	34,288	RPE
3	111122	9	17	7	34,286	1	26	5,069	RP
4	111123	2	18	4	34,350	1	27	8,872	RE
5	111132	250	15	7	34,298	1	45	5,316	R
6	111133	90	16	4	34,866	1	70	9,828	RE
7	111222	4	9	7	34,006	2	25	4,822	RHP
8	111232	175	15	7	34,026	2	45	5,021	RH

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
9	111233	27	14	4	35,066	2	74	9,102	RHE
10	111322	1	15	8	37,216	3	27	4,652	RHP
11	111332	34	14	7	33,991	3	45	4,933	RH
12	112112	2	18	8	46,492	1	16	6,227	RP
13	112113	2	13	6	49,734	1	15	9,187	RPE
14	112122	5	11	8	42,315	1	27	5,597	R
15	112123	5	17	6	51,868	1	22	10,117	RPE
16	112132	108	15	7	45,828	1	41	6,221	R
17	112133	178	15	5	51,308	1	58	11,186	RE
18	112222	5	19	8	47,553	2	26	6,250	RHP
19	112223	4	17	6	50,228	2	23	8,607	RHPE
20	112232	60	17	8	45,679	2	41	6,036	RH
21	112233	86	16	6	52,215	2	50	9,457	RHE
22	112322	1	23	8	48,365	3	28	6,046	RH
23	112323	1	26	8	58,950	3	29	7,369	RH
24	112332	17	11	8	45,537	3	41	6,082	RH
25	112333	8	15	6	52,987	3	47	9,202	RHE
26	113113	3	18	6	117,081	1	16	21,184	RMPE
27	113123	9	18	6	111,974	1	25	21,683	RMPE
28	113133	197	14	6	133,556	1	51	23,545	RME
29	113213	2	21	7	132,370	2	11	18,910	RMHPE
30	113223	2	18	6	81,870	2	19	14,549	RHPE
31	113233	137	16	7	119,400	2	49	18,556	RMHE
32	113333	22	12	7	105,015	3	42	14,373	RMHE
33	121111	12	15	15	35,219	1	16	2,368	RP
34	121121	303	14	13	34,209	1	23	2,580	RP
35	121122	75	15	10	36,426	1	27	3,607	R
36	121131	18	16	11	32,775	1	34	2,977	R
37	121132	87	16	9	35,058	1	34	3,740	R
38	121211	8	18	14	34,101	2	14	2,558	RHP
39	121221	293	15	13	33,936	2	24	2,583	RHP

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
40	121222	42	17	10	36,228	2	27	3,616	RH
41	121231	29	15	11	32,576	2	37	2,927	RH
42	121232	64	15	9	34,797	2	39	3,726	RH
43	121311	9	13	15	33,540	3	15	2,200	RHP
44	121312	1	4	9	37,773	3	16	4,197	RHP
45	121321	127	14	14	34,248	3	23	2,505	RHP
46	121322	11	13	10	36,732	3	28	3,610	RH
47	121331	23	12	12	33,026	3	39	2,793	RH
48	121332	20	14	9	35,419	3	43	3,794	RH
49	122111	3	15	16	45,378	1	17	2,836	RP
50	122112	11	17	12	53,931	1	14	4,557	RP
51	122121	77	13	15	44,232	1	21	2,967	RP
52	122122	309	14	12	50,705	1	25	4,181	RP
53	122132	90	15	10	49,620	1	34	5,217	R
54	122211	3	17	16	43,707	2	15	2,787	RHP
55	122212	11	14	14	55,406	2	14	4,131	RHP
56	122221	94	14	15	43,764	2	21	2,941	RHP
57	122222	250	15	12	50,937	2	25	4,131	RHP
58	122231	3	10	15	43,706	2	48	2,971	RH
59	122232	108	15	10	48,441	2	36	4,996	RH
60	122312	3	15	12	53,487	3	16	4,783	RHP
61	122321	37	14	15	42,756	3	22	2,878	RHP
62	122322	105	15	13	50,239	3	24	3,995	RHP
63	122331	2	17	14	44,222	3	46	3,165	RH
64	122332	29	13	10	49,549	3	36	5,052	RH
65	123112	7	18	14	80,385	1	14	5,702	RP
66	123113	4	15	12	194,219	1	11	18,868	RMPE
67	123122	181	15	13	76,911	1	23	5,802	RP
68	123123	177	15	12	228,413	1	25	18,436	RMPE
69	123132	13	16	10	66,633	1	35	6,735	R
70	123133	74	14	10	145,124	1	35	14,961	RME

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
71	123212	4	6	16	85,413	2	16	5,508	RMHP
72	123213	8	18	14	215,628	2	13	16,354	RMHPE
73	123222	142	14	13	73,758	2	24	5,586	RHP
74	123223	132	15	13	238,076	2	24	18,341	RMHPE
75	123232	16	13	11	71,386	2	35	6,673	RH
76	123233	65	15	10	136,527	2	37	13,574	RMHE
77	123312	3	22	15	80,896	3	14	5,407	RHP
78	123322	57	16	14	76,639	3	22	5,549	RHP
79	123323	46	15	14	180,533	3	23	13,634	RMHPE
80	123332	17	16	12	71,868	3	40	6,005	RH
81	123333	14	12	11	127,559	3	34	12,214	RMHE
82	131111	332	13	36	34,630	1	12	1,281	RFP
83	131121	71	14	18	34,698	1	19	1,934	RP
84	131211	479	13	44	34,509	2	10	1,122	RFHP
85	131221	84	14	18	35,052	2	19	1,934	RHP
86	131231	5	15	20	35,631	2	43	1,768	RH
87	131311	401	12	49	34,335	3	10	1,007	RFHP
88	131321	69	12	20	33,991	3	19	1,780	RHP
89	131331	6	9	32	36,735	4	34	1,313	RFH
90	132111	421	13	42	49,896	1	11	1,737	RFP
91	132112	5	16	18	59,710	1	16	3,353	RP
92	132121	70	14	19	48,258	1	19	2,624	RP
93	132122	20	13	17	59,367	1	18	3,424	RP
94	132131	2	18	18	51,446	1	32	2,940	R
95	132211	575	12	49	49,904	2	10	1,562	RFHP
96	132212	5	19	17	57,289	2	16	3,370	RHP
97	132221	74	13	20	48,181	2	19	2,458	RHP
98	132222	8	18	18	59,397	2	20	3,371	RHP
99	132231	8	12	30	53,149	2	36	1,900	RFH
100	132232	1	19	17	58,593	2	38	3,447	RH
101	132311	510	11	56	49,177	3	9	1,335	RFHP

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
102	132312	4	18	18	60,525	3	16	3,317	RHP
103	132321	51	14	22	47,987	3	20	2,345	RFHP
104	132322	8	19	17	58,581	3	19	3,397	RHP
105	132331	6	20	20	45,415	4	49	2,289	RH
106	132332	1	25	18	62,640	3	31	3,480	RH
107	133111	424	9	106	104,136	1	6	1,531	RFMP
108	133112	237	13	26	125,546	1	13	4,708	RFMP
109	133113	163	13	30	463,657	1	12	16,057	RFMPPE
110	133121	3	24	28	71,610	1	18	2,625	RFP
111	133122	57	14	18	91,009	1	19	5,035	RMP
112	133123	29	13	18	385,887	1	19	21,631	RMPE
113	133132	3	23	20	118,946	1	34	5,765	RM
114	133133	4	17	21	221,659	1	32	9,852	RFME
115	133211	662	9	104	121,950	2	5	1,533	RFMHP
116	133212	228	13	33	149,093	2	12	4,609	RFMHP
117	133213	133	14	33	564,685	2	11	17,226	RFMHPE
118	133221	8	13	38	69,561	2	21	2,291	RFHP
119	133222	68	14	20	94,536	2	20	4,810	RMHP
120	133223	39	14	20	329,094	2	19	16,707	RMHPE
121	133231	2	23	27	68,979	2	48	2,617	RFH
122	133232	4	16	19	72,504	2	33	3,868	RH
123	133233	2	14	18	454,108	2	32	25,571	RMHE
124	133311	527	9	102	113,609	4	5	1,446	RFMHP
125	133312	139	13	37	163,665	3	12	4,612	RFMHP
126	133313	53	14	30	454,458	3	12	15,116	RFMHPE
127	133321	4	11	27	76,157	4	24	2,874	RFHP
128	133322	35	12	20	86,353	3	19	4,370	RMHP
129	133323	16	19	19	228,768	3	20	12,202	RMHPE
130	133331	2	14	27	80,394	4	32	3,024	RFH
131	133332	6	15	19	77,431	3	35	4,041	RH
132	133333	3	17	20	255,172	3	48	12,850	RMHE

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
133	211112	3	53	6	34,563	1	12	5,480	RP
134	211113	3	59	2	36,760	1	5	27,931	RPE
135	211122	30	55	7	35,339	1	28	4,820	R
136	211123	9	51	3	34,685	1	26	11,444	RPE
137	211132	464	50	7	34,097	1	44	5,300	R
138	211133	246	52	4	34,689	1	68	9,912	RE
139	211213	1	47	2	33,975	2	13	16,988	RHPE
140	211222	21	49	7	34,431	2	27	4,700	RHP
141	211223	2	60	3	38,555	2	24	16,015	RHPE
142	211232	295	50	7	33,909	2	42	5,143	RH
143	211233	67	51	4	35,317	2	68	9,625	RHE
144	211322	4	39	8	32,524	3	29	4,066	RH
145	211323	1	51	3	31,790	3	24	10,597	RHPE
146	211332	72	49	7	33,959	3	40	4,968	RH
147	211333	2	61	4	35,436	3	53	8,859	RHE
148	212112	2	48	8	46,616	1	14	6,211	RP
149	212113	5	47	4	50,193	1	4	27,772	RPE
150	212122	25	60	8	45,850	1	28	5,820	R
151	212123	21	50	4	52,214	1	24	14,369	RPE
152	212132	206	48	7	46,056	1	39	6,335	R
153	212133	397	50	5	50,554	1	55	10,931	RE
154	212213	2	51	7	53,540	2	9	8,496	RHPE
155	212222	14	56	8	47,324	2	29	5,972	RH
156	212223	11	53	5	50,050	2	25	11,279	RHPE
157	212232	140	50	7	45,043	2	39	6,220	RH
158	212233	178	50	5	50,412	2	52	9,917	RHE
159	212313	1	46	5	46,740	3	17	9,348	RHPE
160	212322	5	52	8	47,845	3	29	6,150	RH
161	212323	2	48	8	60,711	3	26	8,120	RHP
162	212332	34	48	7	45,312	3	38	6,148	RH
163	212333	28	49	6	51,486	3	48	8,723	RHE

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
164	213113	6	45	4	83,581	1	8	38,972	RMPE
165	213123	22	55	7	115,700	1	27	16,920	RME
166	213133	353	50	6	155,784	1	52	29,893	RME
167	213213	1	69	8	222,489	2	15	27,811	MHPE
168	213223	20	56	7	117,166	2	26	16,940	RMHPE
169	213233	229	50	6	140,030	2	46	21,862	RMHE
170	213323	5	58	8	298,948	3	27	40,750	RMHE
171	213333	36	50	7	132,020	3	41	18,514	RMHE
172	221111	29	53	15	33,641	1	15	2,322	RP
173	221112	1	38	11	35,972	1	10	3,270	RP
174	221121	311	48	13	33,746	1	22	2,642	RP
175	221122	145	53	10	35,675	1	26	3,650	RP
176	221131	26	50	12	32,641	1	39	2,867	R
177	221132	82	46	9	34,799	1	37	3,724	R
178	221211	35	55	15	33,703	2	15	2,307	RHP
179	221221	335	46	13	34,039	2	22	2,637	RHP
180	221222	133	53	10	35,665	2	26	3,642	RHP
181	221231	35	49	13	33,327	2	50	2,719	RH
182	221232	76	44	9	34,855	2	37	3,758	RH
183	221311	14	50	15	36,313	3	15	2,500	RHP
184	221312	1	30	10	33,518	3	17	3,352	RHP
185	221321	127	47	13	33,006	3	22	2,557	RHP
186	221322	29	50	10	35,690	3	26	3,565	RHP
187	221331	12	48	12	32,864	3	39	2,851	RH
188	221332	28	47	10	35,473	3	35	3,750	RH
189	222111	11	43	16	44,650	1	16	2,824	RP
190	222112	13	52	15	53,517	1	16	3,626	RP
191	222121	67	47	15	43,855	1	21	2,995	RP
192	222122	417	48	12	50,267	1	24	4,381	RP
193	222131	11	50	15	44,550	1	37	3,032	R
194	222132	115	44	10	49,172	1	36	5,079	R

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
195	222211	13	53	15	42,968	2	16	2,780	RHP
196	222212	10	50	14	51,256	2	15	3,813	RHP
197	222221	74	45	15	43,238	2	21	2,954	RHP
198	222222	343	48	12	49,687	2	24	4,369	RHP
199	222231	4	47	15	44,014	2	35	2,889	RH
200	222232	87	45	10	49,632	2	39	4,962	RH
201	222311	5	52	16	45,237	3	17	2,898	RHP
202	222312	2	41	16	53,665	3	16	3,462	RHP
203	222321	38	46	15	44,415	3	20	2,954	RHP
204	222322	135	49	12	50,510	3	24	4,262	RHP
205	222331	4	69	16	42,730	4	51	2,764	H
206	222332	31	45	10	48,148	3	39	4,719	RH
207	223112	12	55	15	75,545	1	15	5,073	RP
208	223113	12	47	15	225,877	1	15	15,099	RMPE
209	223122	156	45	13	74,887	1	23	5,780	RP
210	223123	228	49	12	191,108	1	24	16,279	RMPE
211	223132	22	48	12	72,426	1	43	6,226	R
212	223133	104	45	10	131,698	1	39	13,461	RME
213	223212	8	60	16	90,258	2	16	5,773	RMHP
214	223213	15	53	14	277,978	2	15	19,654	RMHPE
215	223222	173	48	13	74,519	2	23	5,767	RHP
216	223223	217	50	12	205,142	2	24	16,820	RMHPE
217	223232	18	44	13	76,268	2	50	6,122	RH
218	223233	58	46	10	161,730	2	37	16,023	RMHE
219	223312	6	55	16	69,577	4	16	4,495	RHP
220	223313	7	56	15	526,809	3	15	36,395	RMHPE
221	223322	76	45	13	76,401	3	22	5,771	RHP
222	223323	75	47	13	178,826	3	23	13,835	RMHPE
223	223332	5	36	13	72,176	3	41	5,965	RH
224	223333	25	42	10	184,769	3	39	19,407	RMHE
225	231111	218	46	28	34,583	1	12	1,422	RFP

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจับกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
226	231121	44	43	21	34,952	1	20	1,764	RP
227	231131	4	56	21	32,714	1	33	1,579	RF
228	231211	283	45	31	34,705	2	12	1,406	RFHP
229	231221	47	42	20	34,858	2	20	1,781	RHP
230	231231	1	74	28	34,972	2	31	1,249	FH
231	231311	254	44	30	34,607	3	12	1,366	RFHP
232	231321	35	41	19	34,714	3	20	1,847	RHP
233	231331	3	55	19	38,788	3	56	2,095	RH
234	232111	227	45	29	49,884	1	12	2,048	RFP
235	232112	6	44	17	59,650	1	16	3,443	RP
236	232121	65	44	19	48,404	1	20	2,576	RP
237	232122	5	37	17	58,847	1	21	3,462	RP
238	232131	7	50	21	48,291	1	38	2,307	RF
239	232211	279	46	31	49,116	2	12	1,960	RFHP
240	232212	11	51	17	59,359	2	15	3,437	RHP
241	232221	52	42	20	48,377	2	21	2,510	RHP
242	232222	6	48	17	58,737	2	21	3,455	RHP
243	232231	8	49	22	49,476	2	43	2,334	RFH
244	232311	221	45	34	49,268	3	11	1,761	RFHP
245	232312	3	52	17	58,483	3	16	3,373	RHP
246	232321	41	40	22	49,059	3	20	2,344	RFHP
247	232322	2	50	17	57,893	3	18	3,405	RHP
248	232331	8	45	24	50,985	3	37	2,281	RFH
249	233111	104	44	53	92,515	1	9	2,226	RFMP
250	233112	123	46	25	115,253	1	13	4,669	RFMP
251	233113	91	47	27	410,407	1	13	15,743	RFMPPE
252	233121	6	36	25	75,195	1	22	3,060	RFP
253	233122	33	46	19	86,359	1	20	4,515	RMP
254	233123	11	41	19	307,922	1	21	15,568	RMPE
255	233131	1	32	21	63,653	1	47	3,031	RF
256	233133	1	52	17	804,947	1	31	47,350	RME

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจับกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
257	233211	143	43	57	92,626	2	8	2,067	RFMHP
258	233212	160	46	28	129,454	2	12	4,676	RFMHP
259	233213	126	47	27	482,612	2	12	18,090	RFMHPE
260	233221	7	48	31	79,166	2	21	2,604	RFHP
261	233222	38	46	19	91,394	2	21	4,860	RMHP
262	233223	18	40	21	306,655	2	20	14,544	RMHPE
263	233231	3	58	24	70,136	2	35	2,957	RFH
264	233232	4	44	19	87,407	2	42	4,623	RMH
265	233233	3	39	18	235,870	2	33	13,121	RMHE
266	233311	132	44	61	98,688	3	7	1,959	RFMHP
267	233312	103	42	27	118,084	3	12	4,409	RFMHP
268	233313	54	48	29	526,000	3	12	17,633	RFMHPE
269	233321	5	43	31	72,447	3	21	2,378	RFHP
270	233322	22	40	19	85,040	3	22	4,508	RMHP
271	233323	6	41	19	360,953	3	21	19,359	RMHPE
272	233332	6	44	19	93,367	3	43	4,932	RMH
273	233333	2	49	19	164,761	4	63	8,826	RMHE
274	311112	19	222	7	33,739	1	14	5,163	P
275	311113	91	228	2	34,323	1	2	28,490	PE
276	311122	383	155	7	34,338	1	25	5,236	P
277	311123	126	206	4	34,894	1	26	10,312	PE
278	311132	343	126	6	33,984	1	41	5,867	
279	311133	1,059	167	3	34,523	1	51	11,856	E
280	311212	23	204	7	33,831	2	13	5,026	HP
281	311213	10	247	3	34,432	2	10	12,391	HPE
282	311222	141	147	7	34,430	2	25	5,050	HP
283	311223	38	187	4	35,689	2	26	9,572	HPE
284	311232	189	117	6	34,075	2	41	5,706	H
285	311233	169	145	4	34,612	2	50	9,695	HE
286	311312	2	222	7	33,308	3	13	4,894	HP
287	311322	30	122	7	34,127	3	25	4,635	HP

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจັกดุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
288	311323	4	186	5	34,740	3	26	7,739	HP
289	311332	19	116	6	33,252	3	41	5,667	H
290	311333	8	126	4	37,640	3	51	9,432	HE
291	312112	21	204	8	45,045	1	13	5,934	P
292	312113	89	234	3	48,583	1	6	30,998	PE
293	312122	215	144	7	46,151	1	24	6,211	P
294	312123	276	175	5	50,878	1	26	10,324	PE
295	312132	83	117	7	43,851	1	42	6,446	
296	312133	1,144	158	4	48,766	1	50	14,953	E
297	312212	9	193	8	44,413	2	15	5,812	HP
298	312213	25	226	4	49,757	2	12	12,964	HPE
299	312222	97	137	7	45,924	2	25	6,161	HP
300	312223	92	172	6	52,057	2	26	9,477	HPE
301	312232	54	109	7	45,282	2	41	6,421	H
302	312233	271	134	5	50,448	2	46	11,371	HE
303	312312	1	136	8	51,198	3	17	6,400	HP
304	312313	2	152	3	47,572	3	14	15,857	HPE
305	312322	13	142	7	47,154	3	24	6,402	HP
306	312323	13	159	6	53,843	3	27	9,534	HPE
307	312332	15	119	7	44,224	3	41	6,356	H
308	312333	18	119	6	49,139	3	45	8,883	HE
309	313113	86	260	4	129,010	1	8	56,235	MPE
310	313123	309	166	6	134,281	1	25	22,631	MPE
311	313133	821	153	4	120,664	1	49	32,423	ME
312	313213	20	229	6	143,791	2	11	24,101	MHPE
313	313223	139	157	7	128,345	2	25	19,300	MHPE
314	313233	215	131	5	137,234	2	45	27,708	MHE
315	313313	2	187	8	153,712	3	13	20,919	MHPE
316	313323	22	131	7	109,914	3	25	15,075	MHPE
317	313333	17	115	6	130,247	3	44	22,598	MHE
318	321111	86	148	14	34,800	1	14	2,565	P

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
319	321112	18	198	10	34,622	1	13	3,605	P
320	321121	96	111	12	33,631	1	20	2,861	P
321	321122	145	123	10	35,347	1	23	3,729	P
322	321131	25	116	12	34,223	1	44	2,783	
323	321132	18	114	9	34,829	1	57	3,764	
324	321211	81	142	14	34,357	2	14	2,515	HP
325	321212	14	185	10	36,601	2	13	3,602	HP
326	321221	68	108	12	33,133	2	21	2,849	HP
327	321222	96	119	10	35,063	2	23	3,659	HP
328	321231	5	123	12	33,868	2	49	2,843	H
329	321232	10	114	10	37,745	2	46	3,950	H
330	321311	36	126	14	34,943	3	14	2,460	HP
331	321312	5	131	10	36,766	3	16	3,703	HP
332	321321	31	114	12	33,395	3	21	2,786	HP
333	321322	15	109	10	36,137	3	23	3,801	HP
334	321331	7	109	12	33,906	3	46	2,944	H
335	321332	4	109	10	34,851	3	52	3,574	H
336	322111	31	130	15	43,970	1	14	2,937	P
337	322112	102	151	12	50,535	1	15	4,256	P
338	322121	7	112	15	43,685	1	21	2,949	P
339	322122	238	121	10	49,574	1	22	4,854	P
340	322131	2	91	14	44,233	1	49	3,159	
341	322132	36	124	10	47,834	1	50	4,785	
342	322211	25	135	15	43,274	2	13	2,917	HP
343	322212	85	147	13	50,654	2	15	4,103	HP
344	322221	16	99	14	41,961	2	20	2,986	HP
345	322222	182	115	11	49,538	2	22	4,777	HP
346	322231	2	102	15	43,915	2	44	2,928	H
347	322232	31	119	11	49,850	2	44	4,725	H
348	322311	13	126	15	43,871	3	15	2,942	HP
349	322312	23	138	13	53,107	3	15	4,227	HP

ตารางที่ 6.1 (ต่อ)

ลำดับ	ค่าการจัดกลุ่ม	จำนวนคน	R	F	M	H	P	E	รูปแบบ
350	322321	3	86	15	42,556	3	18	2,918	HP
351	322322	62	108	11	49,441	3	22	4,682	HP
352	322331	1	125	14	41,579	3	32	2,970	H
353	322332	3	124	10	52,090	3	42	5,155	H
354	323112	58	145	14	76,639	1	14	5,737	P
355	323113	88	162	13	231,257	1	14	18,299	MPE
356	323122	62	117	12	72,043	1	21	6,220	P
357	323123	181	126	11	210,058	1	21	20,271	MPE
358	323132	9	140	12	71,011	1	40	6,157	
359	323133	29	130	11	292,165	1	38	25,529	ME
360	323212	51	155	14	77,057	2	13	5,636	HP
361	323213	99	160	13	249,999	2	14	19,918	MHPE
362	323222	38	113	12	72,303	2	21	6,090	HP
363	323223	157	117	11	230,187	2	22	20,817	MHPE
364	323232	6	100	13	75,289	2	64	5,981	H
365	323233	28	119	10	270,321	2	47	25,880	MHE
366	323312	17	148	14	76,318	3	14	5,458	HP
367	323313	18	161	13	208,778	3	14	16,277	MHPE
368	323322	14	116	12	72,017	3	21	6,195	HP
369	323323	28	113	11	223,442	3	21	20,319	MHPE
370	323332	2	104	13	80,721	3	46	6,293	H
371	323333	5	90	12	712,925	3	37	54,548	MHE
372	331111	60	147	29	34,188	1	9	1,481	FP
373	331121	12	113	21	35,396	1	22	1,774	FP
374	331211	88	131	30	34,557	2	9	1,466	FHP
375	331221	8	110	22	34,493	2	22	1,676	FHP
376	331231	4	140	18	34,317	2	35	1,893	H
377	331311	47	121	37	34,343	3	9	1,261	FHP
378	331321	4	147	22	33,507	3	19	1,579	FHP
379	332111	87	136	26	47,858	1	10	2,111	FP
380	332112	7	115	18	59,417	1	13	3,382	P

ตารางที่ 6.1 (ต่อ)

381	332121	2	97	20	41,753	1	20	2,132	P
382	332131	2	88	22	48,230	1	41	2,346	F
383	332211	92	124	32	48,175	2	9	1,903	FHP
384	332212	4	111	18	60,020	2	13	3,385	HP
385	332221	7	124	21	45,832	2	22	2,249	HP
386	332222	1	117	19	62,210	2	23	3,274	HP
387	332231	1	80	20	46,734	2	43	2,337	H
388	332311	47	115	31	49,608	3	9	1,937	FHP
389	332312	1	99	18	59,652	3	13	3,314	HP
390	332321	5	111	20	48,656	3	21	2,453	HP
391	333111	41	142	61	90,744	1	6	2,007	FMP
392	333112	65	137	23	106,112	1	10	4,597	FMP
393	333113	71	131	26	487,836	1	9	18,737	FMPE
394	333121	2	108	23	69,107	1	19	3,006	FP
395	333122	5	104	18	77,374	1	23	4,349	P
396	333123	2	106	23	1,215,616	1	21	56,418	FMPE
397	333132	1	110	17	99,263	1	51	5,839	M
398	333133	1	104	17	207,653	1	51	12,215	ME
399	333211	71	133	65	108,718	2	5	1,930	FMHP
400	333212	82	129	25	118,933	2	10	4,814	FMHP
401	333213	78	123	25	519,752	2	10	20,487	FMHPE
402	333221	2	115	24	71,619	2	22	2,984	FHP
403	333222	5	97	18	95,048	2	23	5,316	MHP
404	333223	5	118	22	255,753	2	20	11,002	FMHPE
405	333232	5	139	19	97,870	2	34	5,240	MH
406	333233	1	111	17	135,417	2	34	7,966	MH
407	333311	29	127	57	95,033	3	5	1,897	FMHP
408	333312	32	153	25	113,464	3	10	4,523	FMHP
409	333313	35	127	25	432,970	3	10	17,621	FMHPE
410	333321	2	96	30	77,480	3	18	2,718	FHP
411	333322	1	109	20	96,050	3	21	4,802	MHP

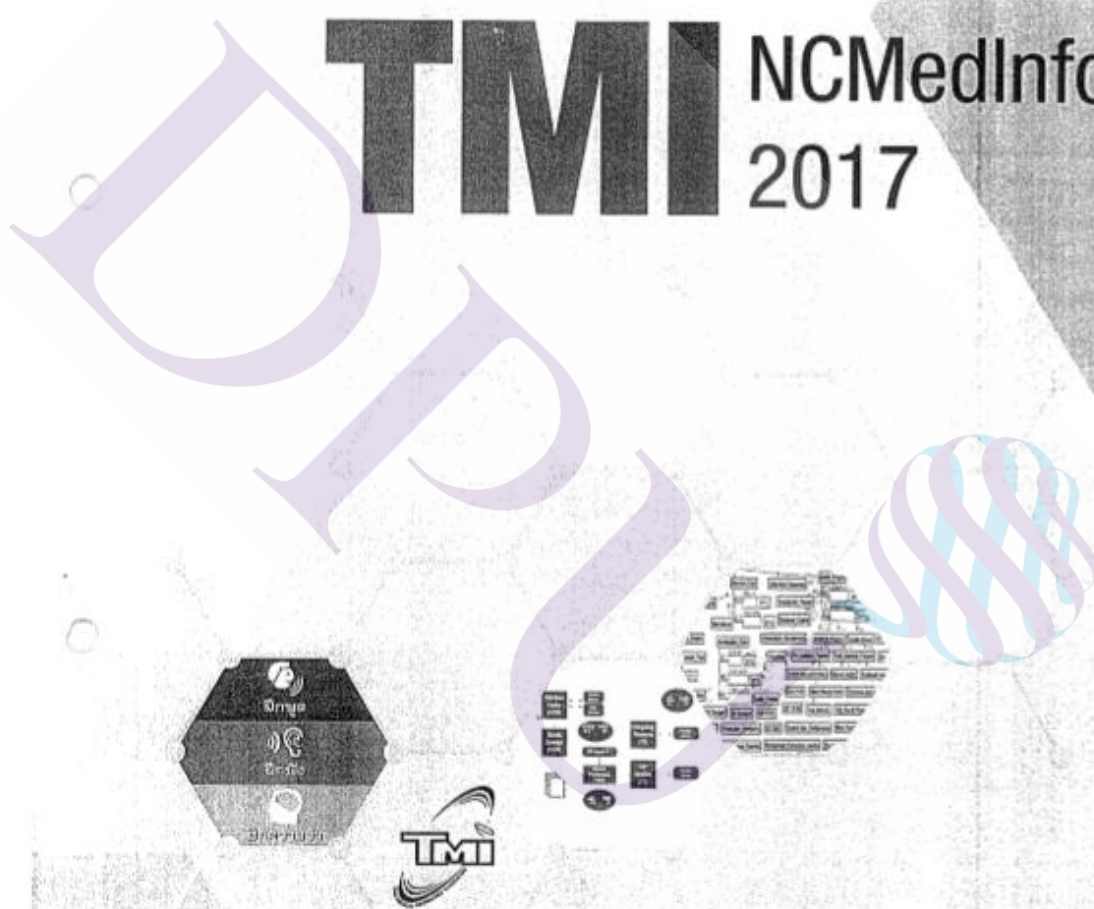
DRU

ภาคผนวก ข



Proceedings The Sixth National Conference on Medical Informatics

TMI NCMedInfo 2017



กระบวนการค้นหาไม่จำเป็นเพื่อศึกษาพฤติกรรมผิดปกติ การเบิกจ่ายค่ายานากลับไปใช้ที่บ้าน กรณีผู้ป่วยนอกสิทธิสวัสดิการข้าราชการระบบเบิกจ่ายตรง

ประไพระ อุดม

สาขาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาเขตนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยอัสสัมชัญ
อีเมล: 585162020019@ajtu.ac.th

เอกสิทธิ์ พัทธวงศ์ศักดิ์

สาขาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาเขตนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยอัสสัมชัญ
อีเมล: eakasit.pac@ajtu.ac.th

บทคัดย่อ -- การเบิกจ่ายเงินประกันที่ไม่เหมาะสมให้กับผู้มีสิทธิเบิกจ่ายคืน
เนื่องจากการตรวจผิดพลาดหรือการใช้ไปเป็นทางที่ผิด ปัญหาที่เกิดขึ้นเป็น
ปัญหาใหญ่ที่สำคัญของระบบสุขภาพที่ให้บริการเชิงปริมาณและคุณภาพ
แก่คนไข้จำนวนมาก ปัจจุบันการตรวจสอบการเคลมสามารถทำได้จากแร่แจ้ง
พฤติกรรมผิดปกติของคนที่ส่งสำเนาจากโรงพยาบาล แล้วตรวจสอบบัญชีเงินได้
ตรวจสอบจากกรรมการประเมินเงินได้ของผู้ป่วย ซึ่งมีความยากลำบาก
และใช้เวลานานในการตรวจสอบก่อนจ่ายเงิน งานวิจัยนี้มีวัตถุประสงค์เพื่อ
พัฒนารูปแบบการตรวจสอบความผิดปกติของการเบิกจ่ายค่ายานากลับไป
ใช้ที่บ้าน โดยการใช้ Data Mining ตรวจสอบวิเคราะห์ข้อมูลด้วย
เทคนิค Unsupervised Learning จากฐานข้อมูลผู้ป่วยนอกสิทธิข้าราชการ
ระบบเบิกจ่ายตรงปีงบประมาณ 2556 จำนวน 21 ล้าน records ประกอบด้วย 1)
แฟ้มข้อมูลวินิจฉัยของผู้ป่วยนอก (OPDATA) 2) แฟ้มจ่ายค่ายานากลับ
(OPBILL) 3) แฟ้มการรับส่งค่ายา (OPDISP) 4) แฟ้มราคาจ่าย
(OPDTIEM) รวบรวมจากการเบิกจ่ายกรมบัญชีกลาง การทดสอบประกอบด้วย
เทคนิค 2 เทคนิค เพื่อทำการเปรียบเทียบ คือ RFM Segmentation และ
เทคนิค Partitional Clustering ด้วย K-means, K-medians ผลที่ได้คือพบว่า
ผู้ป่วยที่มารับบริการจำนวน 2.6 ล้านคน ความถี่ในการมารับบริการจำนวน 21
ล้านครั้ง นำใช้บริการสถานพยาบาลสูงสุด 32 แห่ง จำนวน 710 ครั้ง และ
จำนวนที่มีการเบิกจ่ายค่ายานากลับบ้านจำนวน 5.3 แสนคน ผลจาก RFM score
(555555) มีจำนวน 112 คน 0.02% ผลจากเทคนิค Partitional Cluster ค่าส่วน
ใหญ่กลุ่มหนึ่งกลุ่ม เพื่อให้การตรวจสอบพฤติกรรมที่ผิดปกติงานวิจัยนี้
สามารถช่วยแก้ปัญหาแบบความผิดปกติที่เกิดขึ้นได้อย่างมีประสิทธิภาพมากขึ้น
คำสำคัญ : RFM, K-means, K-medians, Unsupervised, Euclidean
distance

Abstract— The improper insurance fund reimbursement occurring from uncommon unappropriated aspects for person entitled to reimbursement is a significant problem of health system which loses a huge amount of money to provide care for this group. At the presents, although claim system can be verified by The Comptroller general's Department (CGD), patient's medical records especially abnormal behavior or reports from hospitals will be reviewed by CGD's experts with considerably difficulty and taking a long time to check.

The objective of this study to develop patterns of abnormal behavior of outpatient drug reimbursement investigations which use Data Mining 21 million records of outpatient database under Civil Servant Medical Benefit Scheme (CSMBS)'s Direct Billing System in fiscal year 2013 are analyzed by Unsupervised Learning technique. In this reimbursement database, consisting of 4 sections such as outpatient diagnostic (OPDATA), expenses category (OPBILL), dispensing (OPDISP) and medical record (OPDTIEM) which are collected and examined by 2 techniques to compare such as RFM Segmentation Model and Partitional Clustering with K-means, K-medians.

The result illustrated that there were 2.6 million patients attended the service with 21 million times of services frequency, and 32 hospitals in total with 710 times are accessed the health services. Additionally, approximately 5.3 hundred thousand people of outpatient reimbursement analyzed by RFM score (555555) showed that there were 112 patients, amounting to 0.02%. With regard to the result of Partitional Cluster, the vast majority of reimbursement fall into one group. To investigate uncommon behavior of reimbursement, this study will be explore the unusual patterns more efficiently.

Keywords- RFM, K-means, K-medians, Unsupervised, Euclidean distance

1. บทนำ

ในปัจจุบันประเทศไทยได้มีการเปิดช่องทางจำหน่ายอาหารสุขภาพเป็นจำนวนมากเพิ่มขึ้นอย่างต่อเนื่อง จากข้อมูลจากรายงานของทุกฝ่ายที่...

ภาพที่ 1.1 ค่าใช้จ่ายในการรักษาพยาบาล



รูปที่ 1.1 ค่าใช้จ่ายในการรักษาพยาบาลในประเทศไทย

จากการเปิดจำหน่ายของผลิตภัณฑ์สุขภาพ ซึ่งส่วนใหญ่ในการจำหน่ายเรารวมไปถึงการเปิดช่องทางจำหน่ายที่ 16 หมวด รวมเป็นเงินจำนวน 34 พันล้านบาท...

2. วัตถุประสงค์ของการวิจัย

ศึกษาพฤติกรรมการมารับบริการที่สถานพยาบาล... เพื่อวิเคราะห์แนวโน้มการเข้ารับบริการ...

3. ขอบเขตการวิจัย

ศึกษาพฤติกรรมและผลการวิจัยที่เกี่ยวข้อง [1-3] ใช้เทคนิคการเรียนรู้ของเครื่อง (Data mining) เพื่อค้นหา รูปแบบ (pattern) ของผู้ป่วย...

4. แนวคิดที่เกี่ยวข้อง

Seluc Palun, Altan Kooyigit and P. Erhan Eras (2017) ได้ศึกษาเรื่อง LRFMP model for customer segmentation in the grocery retail industry...

Hue-Hung Wu, Shih-Yen Lin, and Chia-Py Liu (2014) ได้ศึกษาเรื่อง Analyzing Patients' Values by Applying CLUSTER Analysis and LRFMP Model in a Pediatric Dental Clinic in Taiwan...

Biyan Geraili, Mahdi Nasiri, Mohammad Aash (2015) ได้ศึกษาเรื่อง Improving Fraud and Abuse Detection in General Physician Claims A Data Mining Study...

แพทย์จ่ายยาไม่เหมาะสม สมควรคิดเงินค่ายามาขึ้น วิธีใหม่แพทย์จะสร้างรายชื่อให้
 มาเชื่อม ผลกระทบถ้าจะทำได้ไม่พอนี้ ได้วิธีคนละประเภท โฆษณาราคาหน่วยจ่าย
 ภายนอก 2) ทศุติกรรมที่แพทย์ไม่เขียนรายการยาในใบสั่งยา แต่ให้คนไข้ไป
 ซื้อจากร้านขายยาข้างนอก แล้วไปใบเสร็จไปเบิกกับหน่วยงานที่ต้นสังกัด
 วิธีการนี้จะจะมีวิธีการอื่น ในสิ่งของนอก และนำไปซื้อร้านขายยาที่ถูกรวมคิด
 และนำเงินไปเบิกจากหน่วยงานหรือบริษัทประกัน แล้วนำเงินที่ได้มานำกับ
 ผลการวิจัยพบว่า มีตัวชี้วัดที่มีความสัมพันธ์กับเรื่องค่าใช้จ่าย คือ ความถี่มารับ
 บริการ ซึ่งผู้ป่วยไปรับบริการกับแพทย์มากกว่า 1 ครั้งในช่วงเวลาสั้น และแพทย์
 ใด้เขียนใบสั่งยาไปซื้อที่ร้านขายยาข้างนอก จากบันทึกการรักษารายการยา
 ในใบสั่งยา 30% ให้ไปซื้อจากร้านขายยาข้างนอกสถานพยาบาล ทั้งนี้ แพทย์ 54%
 มีพฤติกรรมการไม่เหมาะสม และ 2% คงยั้วว่าถูกต้อง

Karea Dahi Anal a, Halli Alay Givner b, Ihsan Subenceng Inc, Ahmet Ruchan Akande (2012) ได้ศึกษาเรื่อง 4 A prescription fraud detection
 model วัตถุประสงค์ของงานวิจัยเพื่อพัฒนาโมเดลสำหรับการตรวจสอบการทุจริตใน
 ซึ่งอาจเป็นปัญหาใหญ่ทำให้สูญเสียเงินจำนวนมาก วิธีการวิจัยโดยการค้นคว้าด้วย
 ไปดัดแปลง และใช้ data mining เทคนิค supervised และ neural networks
 เพื่อประเมินการทุจริตในสิ่งยา ซึ่งสามารถช่วยค้นหาข้อผิดพลาดในการสั่งจ่ายยา
 ตรวจตรวจสอบ โดยทดสอบกับฐานข้อมูลขนาดใหญ่ในเชิงรู้ หน่วยงานการทุจริต
 ร่วมกับระหว่างบริษัทยา, แพทย์, ร้านขายยาและผู้ป่วย การนำ data mining มา
 ใช้ ทำให้สามารถตรวจสอบข้อมูลได้โดยไม่มีประสิทธิภาพมากจนเกินไปและลด
 ค่าใช้จ่ายของการจ้างผู้เชี่ยวชาญตรวจสอบเอกสาร

Dallas Thornton, Guido van Capelleven, Marnen Poel, Joe van
 Hillegestberg and Roland M. Mueller (2014) ได้ศึกษาเรื่อง 4 Outlier-based
 Health Insurance Fraud Detection for U.S. Medicaid โดยศึกษาเรื่องข้อมูลใน
 การตรวจสอบการทุจริตด้วยเทคนิค Unsupervised เพื่อค้นหาการทุจริต
 ข้อมูลการเคลมประกันกรรม โดยใช้ข้อมูล 2012-2013 จำนวน 11 เดือน พบว่า
 คนไข้ไม่ใช่ว่าทั้งหมดที่ทุจริต แต่ที่สงสัยสงสัยคือ ๕๐ คนพยาบาลที่มีผู้ป่วยมารับ
 บริการมากกว่า 300 คน ต่อปีค่าที่ อาจต้องสงสัยว่าทุจริต เช่น มีการใส่ข้อมูล
 ปลอม หรืออาจจะเป็นการให้บริการที่เกินควรกรณีอื่น ๆ วิธีการตรวจสอบการ
 ทุจริต จะหาผู้ให้บริการที่มีค่าผิดปกติจาก multiple predictors โดยการวิเคราะห์
 ตัวชี้วัด 14 ตัว วิเคราะห์ข้อมูลด้วยโปรแกรม R กรณีที่มีการเบี่ยงเบนของ
 10,000 ผลลัพธ์ที่ผิดปกติ และการวิเคราะห์ด้วยเทคนิค K-means clustering เพื่อ
 ค้นหาที่ผิดปกติจากข้อมูลผู้ป่วยให้บริการจำนวน 500 แห่ง ผลการวิจัยพบว่าผู้
 ให้บริการ 71% มีพฤติกรรมการไม่เหมาะสมและผิดปกติ จากหลักฐานทำให้สามารถ
 ตรวจสอบได้ตามกฎหมาย

Qi Liu, Miklos Vasarhelyi (2013) ได้ศึกษาเรื่อง 4 Healthcare fraud
 detection: A survey and a clustering model incorporating Geo-location
 information ของประเทศอเมริกาในปีงบประมาณ 2010 วัตถุประสงค์เพื่อกรว
 ลกษณการทุจริตมีประสิทธิผลมากขึ้นและช่วยลดค่าใช้จ่ายด้านสุขภาพ โดยให้
 ความสนใจในการป้อนข้อมูลที่ผิดพลาดวิธีไม่สมบูรณ์ในบทความและผลการ
 ให้บริการทางการแพทย์ที่นำเข้าสู่ วิธีการวิจัยใช้ข้อมูล SSO ให้แบ่งกลุ่มผู้มา

รับบริการ เป็น 3 ประเภท 1) อายุมากกว่า 65 ปี 2) อายุตั้งแต่ 65 ปีและชน
 พิศาว 3) ผู้ป่วยโรคไตวายเรื้อรังระยะสุดท้าย และแบ่งเข้าสู่ให้บริการเป็น 3
 ประเภท 1) Hospital insurance 2) Medical insurance 3) prescription drug
 coverage (เฉพาะผู้มีรายได้ปีละ) ได้แบ่งการทุจริตเป็น 4 ประเภท 1) การทุจริต
 ของผู้ให้บริการ 2) การทุจริตสมาชิกประกันภัย 3) การโกงค่าใช้จ่ายการประกันภัย
 4) สมาชิกสมาชิกทุจริต การทุจริตที่เกี่ยวข้องกันมากกว่าหนึ่ง การนำแบบ
 พลาได้นำมาใช้ในการตรวจสอบการทุจริตด้วยเทคนิค Score card ซึ่งใช้ social
 network ในการตรวจสอบ ผลที่ได้พบว่ามีการทุจริตจำนวน 47.9 พันล้าน
 ดอลลาร์ต่อปีซึ่งค่าใช้จ่ายประกันสุขภาพเหล่านี้จะรวมไว้ที่ค่าสมาชิกซึ่งสามารถ
 พัฒนารูปแบบการตรวจสอบการทุจริตที่ป้องกันไม่ได้คือเงิน ๖ ใน ๑๐๐

4.1 RFM Model

แบบจำลอง RFM ผู้เขียน Hughes(1996) เพื่อวิเคราะห์พฤติกรรมการ
 ซื้อของของลูกค้า ซึ่งได้ตัวแปร (R) วันล่าสุดที่มา (F) ความถี่ (M) จำนวนเงิน
 ทางการค้า มีระยะเวลาเข้ามาแสดงโอกาสสัมพันธ์กับลูกค้า และส่วนนี้ในการ
 จัดแบ่งคนได้ว่าเป็นลูกค้าประจำ และจำนวนวันที่ใช้ไปเรียกว่าวิธีการ
 ซื้อของ แบ่งร้อยละ R, F, M เป็น 5 กลุ่ม มีจำนวนเท่าๆกัน (quintile) 20% ซึ่ง
 การจัดกลุ่มลูกค้าที่มีพฤติกรรมการซื้อที่ไว้ใจกันจะช่วยให้สามารถคาดการณ์
 ของลูกค้าได้เพิ่มขึ้น ส่วน กลุ่มลูกค้าที่มีการซื้อสินค้าบ่อย แต่มีค่าใช้จ่าย
 อยู่น้อย และมีการใช้จ่ายเฉพาะหรือกลุ่มลูกค้าที่ไม่ค่อยซื้อสินค้าบ่อย แต่มี
 มาซื้อบ่อยครั้งจะอยู่ในจำนวนที่มาก วิธีการจัดกลุ่มตามพฤติกรรมการซื้อสินค้า
 โดยดูจาก 1) ระยะเวลา (จำนวนวัน) จากการซื้อครั้งล่าสุดมา 2) ความถี่ของ
 การซื้อสินค้า 3) การใช้จ่ายของลูกค้า วิธีการนี้ใช้โดย John Feder ในปี 2005
 มีอีกแบบชื่อว่า RFM ซึ่งวิธีการนี้จะแบ่งค่า Recency (R) , Frequency (F) และ
 Monetary (M) ออกเป็น 5 ส่วนเท่าๆกัน โดยดูจากที่มากที่สุด (คือค่า 5) จะมี
 ความสำคัญที่สุดและใช้ค่าตัวเลข 3 หลักเป็นตัวแทน จะแบ่งกลุ่มเป็น 555 คือ
 กลุ่มที่มีค่า R=5, F=5 และ M=5 พบว่ามีความสัมพันธ์กับลูกค้าที่มีมูลค่า
 บ่อยๆ และมีการใช้จ่ายที่สูงสุดตามการประยุกต์ใช้แบบอื่นๆ LRFM มีค่าแบ่ง
 (L) ระยะเวลาที่เป็นลูกค้า John Reinartz and Kumar (2000) ซึ่งเสนอวิธีจัดลูกค้า
 ที่มีมูลค่าสูง

4.2 LRFPM Model

โดยมีการนำ (P) มาใช้ร่วมกับ LRFM เพื่ออธิบายลักษณะการ
 รับประทานของลูกค้าเนื่องจากแบบจำลอง RFM ดั้งเดิมไม่ได้ใช้วิธีดูถึงลักษณะการ
 ความเป็นจริงของลูกค้า เป็นที่สังเกตว่าการรวมครั้งนี้ดูว่าจะจะทำให้เข้าใจ
 คิดได้ การเพิ่มตัวแปร (P) เข้าไปในโมเดลเพื่อแก้ปัญหาที่ควรคำนวณค่า
 Periodicity (P) แสดงลักษณะว่าเป็นลูกค้าประจำ โดยลักษณะระยะเวลาเป็น
 ส่วนที่อธิบายมาตรฐานของการมาซื้อสินค้า

$$Periodicity = \text{stdev}(IVT_1, IVT_2, \dots, IVT_{t-1}, IVT_t) \quad (1)$$

IVT หมายถึง ระยะเวลาที่ระหว่างวันที่เข้ามาซื้อสินค้า ฐานค่าตัวแปร
 ตัวแปร IVT คำนวณว่าลูกค้าจะเข้าซื้อสินค้าเป็นประจำ

$$IVT_t = \text{date_diff}(t, t+1, d) \quad (2)$$

5. การเตรียมข้อมูล

ข้อมูลที่มีนำมาใช้เป็นการเลือกค่ากลับป้อนให้ผู้ใช้บริการของผู้ป่วยออก รายนอกลด ประกอบด้วยตัวแปรที่มีข้อมูล 4 เ็นคือ OPDATA, OPBILL, OPDISP และ OPTEMM ซึ่งเป็น Normalized form โดยมี Primary key field เป็น InteractionID ดัง ภาพที่ 3.1 เมื่อดึงข้อมูลที่มีทั้งหมด 2 ฐาน จากทราเวลงานกรมบัญชีกลาง ทั้ง รวมข้อมูลเข้าด้วยกันเพื่อให้ได้ตัวแปรครบพร้อมทั้งการวิจัยได้ โดยนำโค ดัรตามทฤษฎีการสุภาพ เป็นหน่วยงานตรวจสอบและมีข้อมูลจากระบบ GPMS นำมาใช้ได้ช่วยซึ่ง ส่วนการนำข้อมูลมาใช้โดยกรมบัญชีกลางจะนำข้อมูล นี้มาใช้เพื่อกระบวนการตรวจสอบความถูกต้องของข้อมูลโดยตรวจสอบความ จีรังซ้อนของข้อมูล, ตรวจสอบจำนวนเงินเบิกจ่ายให้ตรงกับยอดเบิก, วันที่มารับ บริการ (datevisit) และความสัมพันธ์ของมารับบริการ

ภาพที่ 3.1 ฐานข้อมูลผู้ป่วยนอกลักษณะ Normalized form



6. วิธีดำเนินการวิจัย

การคำนวณตัวแปร R (Recency) จากวันสุดท้ายที่มารับบริการสถาน พยาบาล โดยกำหนดวันสุดท้าย (30 กันยายน 2556) ลงวันมารับบริการครั้ง สุดท้าย (datevisit) จำนวนครั้งในกรมมารับบริการที่สถานพยาบาล (F), จำนวน สถานพยาบาลที่ไปรับบริการรายบุคคล (H) และค่าเฉลี่ยค่าเฉลี่ยค่ารายคนต่อ ครั้ง (M) จากตัวแปร (billamt) และทำการแบ่งกลุ่ม Cluster RFM, K-means และ K-medians เพื่อแบ่งกลุ่มข้อมูลที่มีลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน มีความคล้ายคลึงกันมากที่สุดและข้อมูลที่อยู่ต่างกลุ่มกันมีความแตกต่างกัน มากที่สุด โดยแบ่งเป็น 5 กลุ่ม หัวข้อโปรแกรม STATA

6.1 Partitional clustering

ก่อนทำการแบ่งกลุ่มต้อง Normalize ข้อมูล มิฉะนั้นจะเกิดความก่อน ทรนจากนั้นทำการแบ่ง Cluster ด้วย K-means, K-medians ใช้วิธีการวัดระยะห่าง Euclidean distance ในการทำ RFM, RFMPLHQ โมเดลตัวแปรที่นำมาแบ่ง Cluster ประกอบด้วยตัวแปรดังต่อไปนี้

- M (Monetary) ค่าเฉลี่ยการเบิกจ่ายค่ารายคนครั้ง
- R (Recency) จำนวนวันจากกรมมารับบริการครั้งสุดท้าย
- F (Frequency) ความถี่ของการมารับบริการ

- P (Periodicity) ส่วนเบี่ยงเบนมาตรฐานของช่วงเวลาที่มารับบริการ
- H (HoopVisit) จำนวนสถานพยาบาลที่ไปรับบริการ
- Q (Quantity) ปริมาณยาที่เบิกจ่าย
- Age 0%*

7. ผลการทดลอง

7.1 ค่าสถิติและความสัมพันธ์

จากตารางที่ 7.1 พบว่า (Q) มีความสัมพันธ์อย่างมีนัยสำคัญกับค่าราย รายครั้ง (M) เท่ากับ 9% โคตที่ (H) มีความสัมพันธ์กับจำนวนครั้ง (F) เท่ากับ 32% ซึ่งมีความสัมพันธ์ที่มีความสัมพันธ์กันค่อนข้างน้อย

ตารางที่ 7.1 แสดงความสัมพันธ์ค่าสถิติ Cramer's on ระหว่างตัวแปร

	Q	M	R	F	P	H	Age
Q	1						
M	0.090	1					
R	-0.032	0.011	1				
F	0.005	-0.017	-0.142	1			
P	0.049	0.025	-0.074	0.301	1		
H	-0.028	-0.090	-0.140	0.323	0.156	1	
Age	0.125	0.072	0.108	0.147	0.089	0.125	1

ที่มา: สำนักรวบรวมข้อมูลการวิจัยกรมบัญชีกลาง (สวสช)

จากตารางที่ 7.2 ข้อมูลการมารับบริการที่สถานพยาบาล จำนวน 21 ล้านครั้ง จำนวนผู้มารับบริการ 2.6 ล้านคน ค่าเฉลี่ยผู้มารับบริการที่มีค่า รายจ่ายค่ารายคนมากที่สุด 5.3 แสนคน จำนวนเฉลี่ยที่มีการเบิกจ่ายค่ารายคน (M) เฉลี่ยต่อคนมากที่สุด 340,907 บาทต่อครั้ง (R) ครั้ง (F) ในการมามากที่สุดที่ 710 ครั้ง ช่วงระยะเวลาที่มารับบริการ (P) สูงกว่า 255 วัน มีค่าเฉลี่ยของ ข้อมูลผู้ป่วยมารับบริการเป็นประจำ จำนวนสถานพยาบาล ที่ไปใช้บริการมากที่สุด 32 แห่ง มีการเบิกจ่ายค่ารายคนสูงสุด (Max) 2.5 ล้านบาท ปริมาณยาที่ได้รับจำนวนมากที่สุด 5,700 หน่วย

ตารางที่ 7.2 แสดงค่าเฉลี่ยและตัวแปร

Stats	Mean	Median	Max	Min	SD
M	300	184	340,907	0	3,269
MMax	8,580	1,410	2,943,471	1	27,133
R	98	62	345	1	89
F	12	8	710	1	17
P	34	29	255	0	28
H	3	2	33	1	1
Q	67	30	5,700	1	99
Age	57	50	100	0	19

ที่มา: สำนักรวบรวมข้อมูลการวิจัยกรมบัญชีกลาง (สวสช)

7.2 RFMPLHQ score

การจัดอันดับผู้มารับบริการด้วยคะแนนของกลุ่มตามโมเดล RFM Segmentation ค่า score ซึ่งแบ่งได้ 8,693 กลุ่ม ที่มีผู้มารับบริการสูงสุด (RFMPLHQ = 1125 (2) จำนวน 3,145 คน คิดเป็น 0.58% ความหมายคือ

ผู้รับบริการมาไม่ก่อนนัดมีการเบิกจ่ายค่ายาตามรับจำนวนมาและไปไม่ส่งมอบยาตามนัด) ค่า score (RFMPHQ = 555555) มีจำนวน 112 คน คิดเป็น 0.02% ความหมายคือ ผู้มารับบริการมาเป็นประจำ, มีการเบิกจ่ายค่ายาจำนวนมาก, ไปหลายสถานพยาบาล และปริมาณการเบิกจ่าย

การประยุกต์ใช้โมเดล LRFMP เป็นเทคนิคที่เกิดจาก Ho and Park (1998) ซึ่งได้ศึกษาเกณฑ์ (T) กับค่าเฉลี่ยของตัวแปร L, R, P, M, IVT หรือ P ที่มีค่ามากกว่าค่าเฉลี่ยรวม และจะใช้ (L) ถ้าค่าที่ได้น้อยกว่าค่าเฉลี่ยรวม จากตารางที่ 7.3, 7.4 เป็นการประยุกต์ใช้ค่า (H) แทนค่า (L), (P) เพื่อดูสถานะผู้มารับบริการเป็นประจำ ค่าที่ได้จะค่าเฉลี่ย (H) จำนวนสถานพยาบาลที่ไปรับบริการ เข้าใน RFMPHQ

7.3 RFMPHQ Score แบ่งกลุ่มด้วย K-means, K-medians

จากตารางที่ 7.3. การคำนวณค่า RFMPHQ score ค่าที่ได้ถูกหั่น (T) จะปรากฏตัวแปรในค่า score ตัวแปรที่มีถูกหั่น (L) จะไม่ปรากฏตัวแปรใน Cluster จากการแบ่งกลุ่มด้วยเทคนิค K-means สามารถจำแนกได้ 4 ประเภท ประเภทที่ 1 (FHQ) คิดเป็น 50% ใน Cluster 1 มีค่าเฉลี่ยความถี่สถานพยาบาล (P) 15 ครั้ง ค่าเฉลี่ยจำนวนสถานพยาบาลที่ไป (H) 1.95 โรงพยาบาล ต่อปริมาณยาที่ได้รับกลับบ้านเฉลี่ย (Q) จำนวนเฉลี่ย 69.7 หน่วย ประเภทที่ 2 (R) ใน Cluster 4.5 จำนวนวันเฉลี่ยที่ผู้รับบริการรับบริการ ค่าสุด 299, 107 วัน คิดเป็น 9% และ 19% ประเภทที่ 3 (RMP) ใน Cluster 3 คิดเป็น 16% มีการเบิกจ่ายค่ามากกว่าค่าเฉลี่ยเท่ากับ 1,082 บาท ประเภทที่ 4 (P) ใน Cluster 4 คิดเป็น 71% จะลดเวลาเฉลี่ยในการรับบริการประจำ 106 วัน

ตารางที่ 7.3. ค่าเฉลี่ยโมเดล RFMPHQ score จำนวน K-means

Cluster	N	H	P	M	F	R	Q	RFMPHQ	%
C1	267268	11.29	38.06	801.16	19.81	1.83	69.78	1102	50%
C2	31162	69.97	5.07	820.31	106.88	1.97	61.37	8	3%
C3	91581	103.89	0.75	1001.22	12.81	1.01	68.82	334	16%
C4	46187	103.81	0.97	919.29	37.62	1.00	78.32	8	4%
C5	108281	107.07	10.91	918.32	39.81	1.70	68.52	8	10%
Total	538810	87.88	13.01	808.81	34.06	1.68	68.82		100%

ที่มา: ส่วนวิเคราะห์เพื่อการประเมินผลเชิงประจักษ์จากข้อมูล (SPSS)

จากตารางที่ 7.4 การคำนวณค่า RFMPHQ score แบ่งด้วย K-medians สามารถจำแนกได้ 5 ประเภท ประเภทแรก (FHQ) ใน Cluster 1 คิดเป็น 37% โดยมีค่า F สูงสุด 18 ครั้ง ค่าเฉลี่ยจำนวนสถานพยาบาลที่ไป (H) 2.05 แห่ง ประเภทที่ 2 (MPQ) ใน Cluster 2 คิดเป็น 18% โดยมี (M) เท่ากับ 1,010 บาท และ (Q) ปริมาณยาเฉลี่ยที่ได้รับ 75 หน่วย ประเภทที่ 3 (RMP) ใน Cluster 3 คิดเป็น 17% โดยมี (M) เฉลี่ยสูงสุด 1,064 บาท ประเภทที่ 4 (R) ใน Cluster 4 คิดเป็น 19% ประเภทที่ (RH) ใน Cluster 5 คิดเป็น 18% โดยมีค่า (H) เฉลี่ยเท่ากับ 1.83 แห่ง

ตารางที่ 7.4. ค่าเฉลี่ยโมเดล RFMPHQ score จำนวน K-medians

Cluster	N	H	P	M	F	R	Q	RFMPHQ	%
C1	197783	21.11	18.18	787.47	12.81	3.82	61.57	118	37%
C2	95410	62.91	7.77	1008.36	67.71	1.37	72.71	362	18%
C3	91317	170.68	0.41	1094.27	27.22	1.01	68.28	334	17%
C4	30075	201.38	0.90	837.06	31.47	1.00	61.14	8	10%
C5	36126	103.89	11.11	871.28	20.10	1.82	61.26	311	18%
Total	538810	93.60	10.01	808.81	34.06	1.68	68.82		100%

ที่มา: ส่วนวิเคราะห์เพื่อการประเมินผลเชิงประจักษ์จากข้อมูล (SPSS)

จากค่าที่ได้ตารางที่ 7.3, 7.4 เมื่อเลือกผู้คนที่มีความเสี่ยงสูงใน กลุ่มที่สนใจคือ (FHQ) และ (RH) ทำการ intersect กันพบว่า K-means = C1, K-medians = C1 พบกลุ่มคนที่มีลักษณะเดียวกัน 197,783 คน ในตารางที่ 7.5 และเมื่อตรวจสอบว่ากลุ่มคนดังกล่าวอยู่ในกลุ่มเดียวกับโมเดล RFMPHQ จึงทำการ intersect อีกครั้ง

ตารางที่ 7.5. Intersect K-means, K-medians

K-means	K-medians					Total
	C1	C2	C3	C4	C5	
C1	197,783	64,262	0	0	1,102	263,147
C2	0	18,107	3,388	0	1,237	22,732
C3	0	0	95,257	7,848	3	103,108
C4	0	0	0	10,387	3	10,390
C5	0	3,444	1,267	0	88,170	92,881
Total	197,783	67,612	98,667	10,391	90,412	465,865

ที่มา: ส่วนวิเคราะห์เพื่อการประเมินผลเชิงประจักษ์จากข้อมูล (SPSS)

ทำการ intersect กลุ่ม FHQ กับโมเดล RFMPHQ พบกลุ่ม (Q)-5 ค่า FHQ=555 มีจำนวน 4,394 คน (P=5) คือ ความถี่ของการมารับบริการ อยู่ระหว่าง 17-710 ครั้ง (P=5) จำนวนสถานพยาบาลที่ไปรับบริการ 3-32 แห่ง และ (Q=5) ปริมาณยาที่ได้รับจ่าย 101-5700 หน่วย

ตารางที่ 7.6 รูปแบบพฤติกรรมกรรมการรับบริการ

กลุ่ม	ลักษณะ
FHQ	มาบ่อย, ไปหลายแห่ง, ปริมาณยาจ่ายไม่พอ
PM	มาบ่อย, เข้มงวดจ่ายยา
MPQ	มักกลับมาจำนวนมาก, มาประจำ, ปริมาณยาจำนวนมาก
RMP	มาครั้งเดียว, เข้มงวดจ่ายยา

จากตารางที่ 7.6 แสดงรูปแบบพฤติกรรมกรรมการรับบริการ ที่ได้จากการวิเคราะห์โมเดล RFMPHQ สามารถเห็นความสัมพันธ์ระหว่างจำนวนสถานพยาบาลที่ได้รับจ่ายได้เป็นอย่างดี (P) มีความสัมพันธ์กับจำนวนวันเฉลี่ย (M) กลุ่มที่มารับและเบิกจ่ายจำนวนมาก (P=1M=1) กลุ่มที่มารับและเบิกจ่ายมาก (P=1M=1) กลุ่มที่มารับและเบิกจ่ายน้อย (P=1M=1) และกลุ่มที่มาครั้งเดียวและเบิกจ่ายน้อย (P=1M=1) และกลุ่ม

(RFM) มาแบ่งระดับค่าตามนี้เอง อีกในกรณีที่ผู้มารับบริการมาแล้วสุด (R) และ ความถี่ (F) กลุ่ม (RF²L) มาจากมาครั้งความถี่ต่ำ และ (R²F²L) มาใช้วิธีการ เดียวกันแต่มีความถี่ต่ำ กลุ่มที่ (R²L²F) มาใช้วิธีการเหมือนเดียวกันแต่มีความถี่สูง ความสัมพันธ์อีกในกรณีที่ระหว่างความถี่ (F) และจำนวนสถานพยาบาล ที่ไปรับบริการ (R) กลุ่ม (F²R²L) มาใช้วิธีการแบ่งตัวไปหลายโรงพยาบาล กลุ่มที่ไม่ต้องมาโรงพยาบาล (R²L²F²) และกลุ่มที่มีค่าใช้จ่ายมากไปรักษา พยาบาล (R²L²F) กลุ่ม (M²R²L) ผู้ป่วยมีค่าตามนี้เองและไปรักษานอกหน่วยเดียว และกลุ่มผู้ป่วยที่เบิกค่ายาและไปแพทย์ต่าง (M²R²L)

8. บทสรุป

งานวิจัยชิ้นนี้ นำเสนอผลการศึกษาพฤติกรรมการมารับบริการที่สถาน พยาบาล โดยทำการเก็บข้อมูลจำนวนมากขึ้นไปจัดอันดับก่อนเริ่มเก็บความจำเป็น โดยการใช้โปรแกรมที่ใช้โมเดล RFM กับการเคลือบสีฟันสุขภาพ ซึ่งค้นโมเดล RFM ใช้จัดกลุ่มลูกค้าที่มีมาหรือสินค้า เพื่อนำไปกำหนดกลยุทธ์ทางการตลาด การนำโมเดล RFM มาประยุกต์ใช้ครั้งนี้ใช้ตัวแปร วันล่าสุดที่มารับบริการ, ความถี่, และการเบิกจ่ายเมื่อครั้งก่อน อาจพบรูปแบบการมารับบริการจากค่า RFM score และเพื่อให้ชัดเจนมากขึ้นได้แก่เทคนิค K-means , K-medians แบ่งกลุ่ม 5 กลุ่ม (k=5) เพื่อให้สามารถเปรียบเทียบกับโมเดล RFM 5 กลุ่มได้ พบกลุ่มที่แปลกใจคือ RFM คือ กลุ่มที่ไปรับบริการบ่อย ไปหลายสถานพยาบาล และปริมาณยาที่ได้รับกลับบ้านจำนวนมาก เพื่อให้ทราบว่าเป็นกลุ่มคนเดียวกัน ทำการ intersect กลุ่ม K-means, K-medians ได้กลุ่มคนที่มีลักษณะเดียวกัน 197,783 คน อย่างไรก็ตามกลุ่มนี้กล่าวถึงกว้าง เหมือนเป็นการทำตัวโมเดล RFM ที่สามารถจัดกลุ่มได้ชัดเจนกว่า ดังนั้นผู้วิจัยจึงทำการ intersect อีกครั้งกับ โมเดล RFM ทำให้ได้ค่า RFM score =555 จำนวน 4,594 คน สามารถแยกกลุ่ม ที่มีลักษณะเดียวกันได้ชัดเจนมากขึ้น และเพื่อเพิ่มประสิทธิภาพในการศึกษา รูปแบบที่ชัดเจนขึ้น การประยุกต์ใช้โมเดล RFM RFM ทำการวิเคราะห์และอ่าน มีความสัมพันธ์ สามารถทำนายรูปแบบลักษณะของผู้มารับบริการได้ชัดเจนมากขึ้น ที่กล่าวมาทั้งหมดนี้การใช้เทคนิคดังกล่าวนี้ เป็นการแบ่งกลุ่มผู้เข้ารับ บริการที่มีปริมาณข้อมูลจำนวนมาก สามารถทำได้ที่ในรูปแบบที่ซ่อนอยู่ใน

ข้อมูลได้ และตรวจสอบข้อมูลได้มากขึ้น ใช้เวลาไม่นาน และลดค่าใช้จ่ายในการ จัดผู้รับบริการด้านการตรวจสอบ

ข้อจำกัดในการวิจัยคือ การที่ได้มีข้อมูลปริมาณมากที่มีอีก หน่วยของเอา ที่แตกต่างออกไป ทำให้ได้มีความในเชิงลึกต่อไป และการวิเคราะห์ที่มี ประสิทธิภาพสูงอาจมีหลายเทคนิคเพื่อทำการเปรียบเทียบกัน จึงจะได้ผล วิเคราะห์ที่แม่นยำมากที่สุด

REFERENCES

- [1] กรมบัญชีกลาง, คู่มือสวัสดิการรักษายาเสพติด ฉบับ 1, 2551, หน้า 6-38
- [2] เอกสิทธิ์ พัชรวงศ์ศักดิ์, "Introduction to Business Analytics with RapidMiner Studio 6", ปีพ.ศ. 2558, หน้า 57-69
- [3] เอกสิทธิ์ พัชรวงศ์ศักดิ์, "An Introduction to Data Mining Techniques", ปีพ.ศ. 2557, หน้า 27-49
- [4] Bijan Geraili, Mahdi Nasiri, Mohammad Arab "Improving Fraud and Abuse Detection in General Practice Claims: A Data Mining Study", (2015)
- [5] Dallas Thornton, Guido van Capelleveen, Manjiv Poel, Jos van Hillegersberg and Roland M. Mueller, "Outlier-based Health Insurance Fraud Detection for U.S. Medicaid", (2014)
- [6] Hsin-Hung Wu, Shih-Yen Lin, and Chih-Wei Lin, "Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan", (2014)
- [7] Hughes AM, "Boosting reposit with RFM", (1596)
- [8] Karca Duru Aral a, Halil Altay Güvenir b, Ebu az Sabuncuoğlu, Ahmet Ruchan Akard, "A prescription fraud detection model", (2012)
- [9] Qi Liu, Miklos Vasarhelyi, "Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information", (2013)
- [10] Richard C, Alessandro Re, "Customer segmentation and value proposition", (2015)
- [11] Sung Ho Ha, and Sang Chan Park, "Application of data mining tools to hotel data mart on the intranet for database marketing", (1998)
- [12] Serbat Peker, Altan Kocuyigit and P. Erhan Ezer, "LRFMP model for customer segmentation in the grocery retail industry a case", February (2017)

The use of data mining to investigate abnormal behavior of outpatient drug reimbursement under civil servant medical benefit scheme

Prapai Utama, Eakasit Pacharawongseakda

Big Data Engineering, College of Innovative Technology and Engineering, Dhurakij Pundit University, Bangkok, Thailand

Abstract

The improper insurance fund reimbursement occurring from uncommon unappropriated aspects for person entitled to reimbursement is a significant problem of health system which loses a huge amount of money to provide care for this group. At the presents, although claim system can be verified by The Comptroller general's Department (CGD), patient's medical records especially abnormal behavior reports from hospitals are reviewed by CGD's experts with considerably difficulties and taking a long time to check. The objective of this study is to develop patterns of abnormal behavior of outpatient drug reimbursement investigations using Data Mining. 21 million records of outpatient database under Civil Servant Medical Benefit Scheme (CSMBS)'s Direct Billing System in fiscal year 2013 were analyzed by Unsupervised Learning technique. In this reimbursement database, consisting of 4 sections such as outpatient diagnostic (OPDATA), expenses category (OPBILL), dispensing (OPDISP) and medical record (OPDIEM)

which are collected and examined by 2 techniques to compare such as RFM Segmentation Model and Partitional Clustering with K-means, K-medians. The result illustrated that there were 2.6 million patients attended the service with 21 million times of services frequency, and 32 hospitals in total with 710 times are accessed the health services. Additionally, approximately 6.3 hundred thousand people of outpatient reimbursement analyzed by RFM score (656656) showed that there were 112 patients, amounting to 0.02%. As a result of Partitional Cluster, the vast majority of reimbursement fall into one group. To investigate uncommon behavior of reimbursement, this study explored the unusual patterns more efficiently.

Keywords: RFM, K-means, K-medians, unsupervised, euclidean distance.

Received 31 August 2017; Accepted 30 November 2017

Correspondence: Prapai Utama, College of Innovative Technology and Engineering, Dhurakij Pundit University, Bangkok, Thailand (Tel.: +66-2-954-7300; E-mail address: prapai@hisro.or.th)

กระบวนการดำเนินงานเพื่อศึกษาพฤติกรรมผิดปกติ การเบิกจ่ายค่ายาน่ากลับ ไปใช้ที่บ้านกรณีผู้ป่วยนอกสิทธิสวัสดิการข้าราชการระบบเบิกจ่ายตรง

ประไพ ฤตมา, เอกอภีร์ พงษ์พงศ์พิภพ

สาขาวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยวิศวกรรมศาสตร์เทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต กรุงเทพมหานคร

บทคัดย่อ

การเบิกเคลมเงินประกันที่ไม่เหมาะสมให้กับผู้มีสิทธิเบิกค่ารักษา เนื่องจากการรวมผิดปกติหรือการใช้ใบในทางที่ผิด ปัญหานี้ได้กลายเป็นปัญหาใหญ่ที่สำคัญของระบบสุขภาพทำให้สูญเสียเงินจำนวนมากในการดูแลกลุ่มคนดังกล่าว ปัจจุบันการตรวจสอบการเคลมสามารถทำได้จากการแจ้งพฤติกรรมผิดปกติของคนดังกล่าวจากโรงพยาบาลแล้วทางกรมบัญชีกลางได้ตรวจสอบจากการอ่านเวชระเบียนคนไข้โดยผู้เชี่ยวชาญ ซึ่งมีความยากลำบาก และใช้เวลาในการตรวจสอบค่อนข้างนาน งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อพัฒนารูปแบบการตรวจสอบความผิดปกติของการเบิกจ่ายค่ายาน่ากลับออกที่เบิกมาใช้ที่บ้าน โดยการใช้ Data Mining ตรวจสอบวิเคราะห์ข้อมูลด้วยเทคนิค Unsupervised Learning จากฐานข้อมูลผู้ป่วยนอกสิทธิข้าราชการระบบเบิกจ่ายตรงปีงบประมาณ 2556 จำนวน 21 ล้าน records ประกอบด้วย 1) เพิ่มข้อมูลวินิจฉัยโรครของผู้ป่วยนอก (OPDATA) 2) เพิ่มค่าใช้จ่ายรายหมวด (OPBILL) 3) เพิ่มการเบิกจ่ายค่ายา (OPDISP) 4) เพิ่มรายการยา (OPDIEM) รวบรวมจากการเบิกจ่ายกรมบัญชีกลาง การทดสอบประกอบด้วยเทคนิค 2 เทคนิค

บทนำ

ในปัจจุบันประเทศไทยได้มีการเบิกจ่ายค่ารักษาพยาบาลในการดูแลสุขภาพเป็นจำนวนเพิ่มขึ้นอย่างต่อเนื่อง จากข้อมูลค่ารักษาพยาบาลของทุกสิทธิการรักษา จากภาพที่ 1 ระบบหลักประกันสุขภาพถ้วนหน้า (UC) มีค่าใช้จ่าย 82,907 ล้านบาท ซึ่งดูแลผู้มีสิทธิจำนวน 49 ล้านคน กองทุนสวัสดิการข้าราชการบำนาญข้าราชการ (CSMBS) มีค่าใช้จ่าย 58,832 ล้านบาท ครอบคลุมบิดา, มารดา, บุตรและข้าราชการบำนาญดูแลผู้มีสิทธิ จำนวน 8 ล้านคน และกองทุนหลักประกันสุขภาพแห่งชาติ (SSO) ดูแลผู้มีสิทธิจำนวน 11 ล้านคน มีค่าใช้จ่ายในการรักษาพยาบาลผู้ประกันตนทั้งสิ้น 41,288 ล้านบาท

ผู้นิพนธ์ประสานงาน: ประไพ ฤตมา, สาขาวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัยวิศวกรรมศาสตร์เทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต กรุงเทพมหานคร (โทร. +66-2-954-7300; E-mail address: prapai@hiso.or.th)

เพื่อทำการเปรียบเทียบ คือโมเดล RFM Segmentation และเทคนิค Partitional Clustering ด้วย K-means, K-medians ผลลัพธ์ที่ได้พบว่าผู้ป่วยที่มารับบริการจำนวน 2.6 ล้านคน ความถี่ในการมารับบริการจำนวน 21 ล้านครั้ง มาใช้บริการสถานพยาบาลสูงสุด 32 แห่ง จำนวน 710 ครั้ง และจำนวนที่มีการเบิกเคลมค่ายากลับบ้านจำนวน 5.3 แสนคน ผลจาก RFM score (555555) มีจำนวน 112 คน 0.02% ผลจากเทคนิค Partitional Cluster ค่าส่วนใหญ่ตกอยู่ในหนึ่งกลุ่ม เพื่อให้การตรวจสอบพฤติกรรมที่ผิดปกติงานวิจัยชิ้นนี้สามารถช่วยค้นหารูปแบบความผิดปกติที่เกิดขึ้นได้อย่างมีประสิทธิภาพมากขึ้น

คำสำคัญ: RFM, K-means, K-medians, Unsupervised, Euclidean distance.

วันที่รับต้นฉบับ 31 สิงหาคม 2560; วันที่ตอบรับ 30 พฤศจิกายน 2560



ภาพที่ 1 ค่าใช้จ่ายในการรักษาพยาบาล
ที่มา: สำนักงานวิจัยเพื่อการพัฒนาลูกประกันสุขภาพไทย (สวภ)

จากการเบิกจ่ายรายคนของสิทธิข้าราชการ ซึ่งนำเสนอในการนำมาวิเคราะห์ ลงไปถึงการเบิกจ่ายค่าพยาบาลทั้ง 16 หมวด รวมเป็นเงินจำนวน 34 หมื่นล้านบาท ในระบบเบิกจ่ายตรงปีงบประมาณ 2558 ค่าใช้จ่ายสูงสุดอยู่ที่หมวดค่ายาที่ใช้ในสถานพยาบาล คิดเป็นร้อยละ 48 และยากลับไปใช้ที่บ้าน มีการเบิกจ่ายคิดเป็นร้อยละ 24 ของค่าใช้จ่ายทั้งหมด ซึ่งสาเหตุอาจจะมีการเบิกจ่ายยาในปริมาณเกินความจำเป็นหรือไม่เหมาะสม จากข้อมูลทีกล่าวมาข้างต้น เพื่อศึกษาหมวดยาที่นำกลับไม่ใช้ที่บ้าน พบความผิดปกติของกำรมารับบริการหลายสถานพยาบาล จำนวนสูงสุด 32 แห่ง และจำนวนครั้งที่ไปรับบริการสูงสุด 710 ครั้ง เพื่อตรวจสอบความผิดปกติดังกล่าวด้วยตัวชี้วัดเชิงวิชาการการเบิกเคลม และไม่สามารถเห็นพฤติกรรมภาพรวมของผู้เบิกเคลมเพิ่มเติมวิธีการตรวจสอบนี้ใช้เวลาในการตรวจสอบนานและไม่มีประสิทธิภาพ หลายประเทศใช้วิธีการนี้ อย่างไรก็ตาม การตรวจสอบการเคลมของประเทศไทย ทำได้ยากและใช้เวลานานในการตรวจสอบนานเช่นกัน ดังนั้นการนำเหมืองข้อมูล Data mining มาช่วยในการตรวจสอบวิเคราะห์ด้วยเทคนิค (unsupervised learning) จะสามารถช่วยค้นหาความผิดปกติในข้อมูลได้ งานวิจัยนี้จึงเกิดขึ้น เพื่อช่วยให้สามารถตรวจสอบรูปแบบผิดปกตินี้ให้มีประสิทธิภาพและรวดเร็วมากขึ้น การศึกษานี้มีวัตถุประสงค์เพื่อค้นหาพฤติกรรมที่ผิดปกติของการไปรับบริการที่สถานพยาบาล โดยการประยุกต์ใช้เทคนิคการแบ่งกลุ่ม

ระเบียบวิธีวิจัย

ศึกษาพฤติกรรมและผลการวิจัยที่เกี่ยวข้อง โดยใช้เทคนิค Data Mining (Data mining) เพื่อค้นหารูปแบบ (pattern) ของผู้ป่วยนอกที่มาใช้บริการที่สถานพยาบาลและรับยาที่บ้าน ที่อาจจะมีความผิดปกติที่ซ่อนอยู่ กรณีไม่หลายสถานพยาบาลในเวลาใกล้เคียงกัน ด้วยเทคนิค RFM Analysis, K-means, K-medians ในการตรวจสอบเปรียบเทียบกัน 3 เทคนิค เพื่อให้มั่นใจว่าผู้ป่วยที่มีพฤติกรรมผิดปกติจริงตามสมมติฐาน

Serhat Peker, Altan Kocyigit และ Erhan Eren³ ได้ศึกษาเรื่อง LRFMP model for customer segmentation in the grocery retail industry: a case study โดยพัฒนาโมเดล RFM ในรูปแบบใหม่เรียกว่า LRFMP ประกอบด้วย length, recency, frequency, monetary, periodicity มาประยุกต์ใช้ในการจัดกลุ่ม ถูกวางการค้าปลีกร้านขายของชำในประเทศตุรกี การแบ่งกลุ่มลูกค้าใช้เทคนิค data mining โดยใช้ K-means ผลการวิจัยสามารถแบ่งกลุ่มลูกค้าได้ 5 กลุ่ม คือ ลูกค้าประจำแฟนพันธ์แท้, ลูกค้าประจำ, ลูกค้าที่ไม่แน่นอน, ลูกค้ารายได้สูง, ลูกค้ารายได้น้อย

จากความแตกต่างในพฤติกรรมการใช้สินค้าตามโมเดล LRFMP ทำให้กำหนดกลยุทธ์ทางการตลาดได้

Hsin-Hung Wu, Shih-Yen Lin และ Chih-Wei Liu⁴ ได้ศึกษาเรื่อง Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan โดยศึกษาโมเดล LRFM (length, recency, frequency, monetary) ในคลินิกทันตกรรมเด็กของประเทศไต้หวัน ใช้เทคนิค SOM และใช้ K-means จัดกลุ่มใหม่ในการแบ่งกลุ่ม 12 กลุ่ม ด้วยค่าเฉลี่ยของ L, R, F, M และใช้เมทริกซ์ที่จรรยาความสัมพันธ์จากระยะเวลา และผู้ป่วยใหม่ ผลที่ได้คือสามารถแบ่งกลุ่มผู้ป่วยได้ 3 กลุ่ม คือ กลุ่มผู้ป่วยคนไข้ประจำแฟนพันธ์แท้ ได้จากค่า L,R,F มีค่ามากกว่าค่าเฉลี่ยรวมกลุ่มที่เป็นผู้ป่วยเก่า จากค่า L มีค่ามาก และอาจจะกลายเป็นผู้ป่วยประจำแฟนพันธ์แท้ได้ในอนาคต และกลุ่มผู้ป่วยที่หายไปเนื่องจากค่า L, R มีค่าต่ำ จากผลการวิเคราะห์ที่ได้สามารถนำมาวางแผนกลยุทธ์การตลาดเพื่อออกแบบความต้องการของผู้ป่วยที่แตกต่างกันไปและพัฒนาผู้ป่วยให้กลายเป็นลูกค้าประจำในอนาคต

Bijan Geralli, Mahdi Nasiri, Mohammad Arab⁵ ได้ศึกษาเรื่อง Improving Fraud and Abuse Detection in General Physician Claim A Data Mining Study โดยศึกษาตัวชี้วัด 13 ตัว ของสิทธิการรักษาประกันสังคมมี สถานพยาบาลในเครือ 70 แห่ง คลินิก 260 แห่ง เพื่อวัดประสิทธิภาพในการตรวจสอบแพทย์ ที่ต้องสงสัยว่าทุจริต 98% และแพทย์จ่ายยาไม่เหมาะสม (abuse) 85% วิธีการวิจัยใช้การทำเหมืองข้อมูล เทคนิค supervised และ unsupervised แล้วนำไปสร้างแบบจำลอง ได้แบ่งเป็น 2 กลุ่ม 1) พฤติกรรมแพทย์จ่ายยาไม่เหมาะสมหรือเกินความจำเป็น วิธีนี้แพทย์อาจจะสร้างรายได้ทางซอม และแพทย์อาจจะให้ยาที่แพทย์ ได้รับผลประโยชน์จากหน่วยงาน ภายนอก 2) พฤติกรรมที่แพทย์ไม่เขียนรายการยาในใบสั่งยา แต่ให้คนไข้ไปซื้อยาจากร้านขายยาข้างนอก แล้วนำไปเคาะไปเบิกกับหน่วยงานที่ต้นสังกัด วิธีการนี้อาจจะมีการเขียน ใบสั่งยาปลอม และนำไปซื้อจากร้านขายยาที่สมัครร่วมคิด และนำเงินไปเบิกจากหน่วยงานหรือบริษัท ประกัน แล้วนำเงินที่ได้มาแบ่งกัน ผลการวิจัยพบว่า มีตัวชี้วัดที่มีความสัมพันธ์กับเรื่องค่าใช้จ่าย คือ ความถี่มารับบริการ ซึ่งผู้ป่วยไปรับบริการกับแพทย์มากกว่า 1 ครั้งในช่วงเวลาสั้น และแพทย์ ได้เขียนใบสั่งยาให้ไปซื้อที่ร้านขายยาข้างนอก จากบันทึกการรักษารายการยา ในใบสั่งยา 30% ให้ไปซื้อจากร้านขายยาข้างนอกสถานพยาบาล ทั้งนี้ แพทย์ 84% มีพฤติกรรมไม่เหมาะสม และ 2% สงสัยว่าทุจริต

Dallas Thomson และคณะ⁶ ได้ศึกษาเรื่อง Outlier - based Health Insurance Fraud Detection for U.S. Medicaid

โดยใช้เหมืองข้อมูลในการตรวจสอบการทุจริตด้วยเทคนิค Unsupervised เพื่อพยากรณ์การทุจริตข้อมูลการเคลมทันตกรรม โดยใช้ข้อมูล 2012 - 2013 จำนวน 11 เดือนพบว่า คนไข้ไม่ใช่ทั้งหมดที่ทำการทุจริต แต่ที่ต้องสงสัยคือสถานพยาบาลที่มีผู้ป่วยมารับบริการมากกว่า 300 คนต่อสัปดาห์ อาจต้องสงสัยว่าทุจริต เช่น มีการใส่ข้อมูลปลอมหรืออาจจะเป็นการให้บริการทันตกรรมเคลื่อนที่ วิธีการตรวจสอบการทุจริต จะหาผู้ให้บริการที่มีค่าผิดปกติจาก multiple predictors โดยการวิเคราะห์ตัวชี้วัด 14 ตัว วิเคราะห์ข้อมูลด้วยโปรแกรม R กรณีที่มีการเบิกเงินเคลม 10,000 ดอลลาร์ต่อเดือน และการวิเคราะห์ด้วยเทคนิค K - means clustering เพื่อดูค่าที่ผิดปกติจากข้อมูลผู้ให้บริการจำนวน 600 แห่ง ผลการวิจัยพบว่าผู้ให้บริการ 71% มีพฤติกรรมที่น่าสงสัยและมีผิดปกติ จากหลักฐานทำให้สามารถเข้าตรวจสอบได้ตามกฎหมาย Qi Liu และ Miklos Vasarhelyi⁷ ได้ศึกษาเรื่อง Healthcare fraud detection A survey and a clustering model incorporating Geo-location information ของประเทศอเมริกาในปีงบประมาณ 2010 วัตถุประสงค์เพื่อตรวจสอบการทุจริตให้ประสิทธิภาพมากขึ้นและช่วยลดค่าใช้จ่ายด้านสุขภาพ โดยให้ความสนใจในการป้องกันข้อมูลที่ผิดพลาดหรือไม่สมบูรณ์ในการเคลมเงินและการให้บริการทางการแพทย์ที่ไม่เข้าเงื่อนไข วิธีการวิจัยใช้ข้อมูล SSO ได้แบ่งกลุ่มผู้มารับบริการเป็น 3 ประเภท 1) อายุมากกว่า 66 ปี 2) อายุน้อยกว่า 66 ปีและคนพิการ 3) ผู้ป่วยโรคไตวายเรื้อรังระยะสุดท้าย และแบ่งกลุ่มผู้ให้บริการเป็น 3 ประเภท 1) Hospital insurance 2) Medical insurance 3) prescription drug coverage (เฉพาะผู้มีรายได้น้อย) ได้แบ่งการทุจริตเป็น 4 ประเภท 1) การทุจริตของผู้ให้บริการ 2) การทุจริตสมาชิกประกันภัย 3) การโกงผู้ให้บริการประกันภัย 4) สมรู้ร่วมคิดทุจริต การทุจริตที่เกี่ยวข้องกันมากกว่าหนึ่งฝ่าย การนำเทคนิคดาต้าไมนิ่งมาใช้ในการตรวจสอบทุจริตด้วยเทคนิค Supervised ซึ่งใช้ neural network ในการตรวจสอบผลที่ได้พบว่ามีการทุจริตเงินจำนวน 47.9 พันล้านดอลลาร์ของค่าใช้จ่ายประกันสุขภาพภาครัฐอย่างไรก็ตามงานวิจัยนี้สามารถพัฒนาวิธีการตรวจสอบการทุจริตเพื่อป้องกันไม่ให้เกิดขึ้นได้ในอนาคต

RFM Model

แบบจำลอง RFM ผู้ค้นพบ Hughes (1998) เพื่อวิเคราะห์พฤติกรรมความต้องการของลูกค้า ซึ่งใช้ตัวแปร (R) วันล่าสุดที่มา (F) ความถี่ (M) จำนวนเงิน หากลูกค้ามีระยะเวลาสั้นๆ แสดงโอกาสกลับมาซื้อสินค้าสูง และความถี่ในการซื้อบ่งบอกได้ว่าเป็นลูกค้าประจำ และจำนวนเงินที่ใช้ไปเมื่อคำนวณวิเคราะห์

ต้องทำการ แบ่งข้อมูล R, F, M เป็น 5 กลุ่ม มีจำนวนเท่าๆ กัน (quintile) 20% ซึ่งการจัดกลุ่มลูกค้าที่มีพฤติกรรมคล้ายๆ กันไว้ด้วยกันจะช่วยให้สามารถดูภาพรวมของลูกค้าได้ง่ายขึ้น เช่น กลุ่มลูกค้าที่มีการซื้อสินค้าของเรา เมื่อมาซื้อสินค้าของเราบ่อยๆ และมีการใช้จ่ายเยอะหรือกลุ่มลูกค้าที่ไม่ค่อยมาซื้อสินค้าของเราแต่เมื่อมาซื้อแต่ละครั้งจะซื้อในจำนวนที่มาก วิธีการจัดกลุ่มตามพฤติกรรมในการซื้อสินค้าโดยดูจาก 1) ระยะเวลา (จำนวนวัน) จากการซื้อล่าสุดที่ผ่านไปมา 2) ความถี่ของการซื้อสินค้า 3) การใช้จ่ายของลูกค้า วิธีการนี้นำเสนอ โดย Feder ในปี 2006 เรียกแบบย่อว่า RFM ซึ่งวิธีการนี้จะแบ่งค่า Recency (R), Frequency (F) และ Monetary (M) ออกเป็น 5 ส่วนเท่าๆ กัน โดยเลขที่มีค่ามากที่สุด (คือเลข 5) จะมีความสำคัญที่สุด และใช้ค่าตัวเลข 3 หลักเป็นตัวแทน ของแต่ละกลุ่ม เช่น 555 คือ กลุ่มที่มีค่า R=5, F=5 และ M=5 หมายความว่า เป็นลูกค้าที่มักจะมาซื้อสินค้าบ่อยๆ และมีการใช้จ่ายที่สูง ต่อมามีการประยุกต์ใช้แบบจำลอง LRFM มีตัวแปร (L) ระยะเวลาที่เป็นลูกค้า โดย Reinartz and Kumar (2000) ซึ่งแสดงถึงเป็นลูกค้าที่มีความภักดี

LRFMP Model

โดยมีการนำ (P) มาใช้ร่วมกับ LRFM เพื่ออธิบายลักษณะความสม่ำเสมอของลูกค้าเนื่องจากแบบจำลอง RFM เดิมไม่ได้สะท้อนถึงการซื้อตามความเป็นจริงของลูกค้า เป็นเพียงการทำธุรกรรมครั้งล่าสุดที่อาจจะทำให้เข้าใจผิดได้ การเพิ่มตัวแปร (P) เข้าไปในโมเดลเพื่อแก้ปัญหาที่ การคำนวณค่า Periodicity (P) แสดงลักษณะว่าเป็นลูกค้าประจำ โดยกำหนดระยะเวลาเป็นส่วนเบี่ยงเบนมาตรฐานของการมาซื้อสินค้า

$$\text{Periodicity} = \text{stdev}(IVT_1, IVT_2, \dots, IVT_n, IVT_n) \quad (1)$$

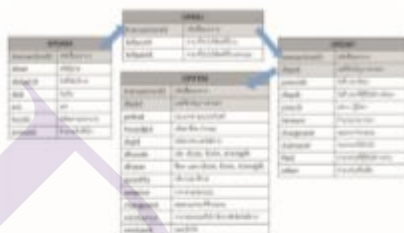
IVT หมายถึง ระยะเวลาห่างระหว่างวันที่เข้ามาซื้อสินค้าล่าสุดลบด้วยวันถัดมา ถ้าค่า IVT ต่ำแสดงว่าลูกค้าเข้ามาซื้อสินค้าเป็นประจำ

$$IVT_i = \text{date_diff}(t_{i+1}, t_i) \quad (2)$$

การเตรียมข้อมูล

ข้อมูลที่น่าสนใจเป็นการเบิกจ่ายกลับไปที่บ้านของผู้ป่วยนอกรายบุคคล ประกอบด้วยแฟ้มข้อมูล 4 แฟ้ม คือ OPDATA, OPBILL, OPDISP และ OPITEM ซึ่งเป็น Normalized form โดยมี Primary key field เป็น transactionID ดังภาพที่ 3 เนื่องจากข้อมูลดิบมีทั้งหมด 2 ล้าน จากหน่วยงานกรมบัญชีกลาง ต้องรวมข้อมูลเข้าด้วยกันเพื่อให้ได้ตัวแปรครบพร้อม

ทำการวิจัยได้โดยสำนักสารสนเทศบริการสุขภาพเป็นหน่วยงานตรวจสอบและนำข้อมูลเข้าระบบ GFMS แต่ไม่ได้จ่ายจริง ส่วนการจ่ายเคลมอนุมัติโดยกรมบัญชีกลางหลังจากนั้นก็เข้าสู่กระบวนการตรวจสอบความถูกต้องของข้อมูลโดยตรวจสอบความซ้ำซ้อนของข้อมูล, ตรวจสอบจำนวนเงินเบิกจ่ายให้ตรงกับขอเบิก, วันที่มารับบริการ (datevisit) และความถี่ของการมารับบริการ



ภาพที่ 3 ฐานข้อมูลผู้ป่วยวนชกลักษณะ Normalized form

วิธีดำเนินการวิจัย

การคำนวณตัวแปร R (Recency) จากวันสุดท้ายที่มารับบริการสถานพยาบาลโดยกำหนดวันสุดท้าย (30 กันยายน 2568) ลบวันที่มารับบริการครั้ง สุดท้าย (datevisit) จำนวนครั้งในการมารับบริการที่สถานพยาบาล (F), จำนวนสถานพยาบาลที่ไม่มารับบริการรายบุคคล (H) และค่าเฉลี่ยค่ายาที่เบิกรายคนต่อครั้ง (M) จากตัวแปร (billamt4) และทำการแบ่งกลุ่ม Cluster RFM, K-means และ K-medians เพื่อแบ่งกลุ่มข้อมูลที่มีลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน มีความคล้ายคลึงกันมากที่สุด และข้อมูลที่อยู่ต่างกลุ่มกันมีความแตกต่างกันมากที่สุดโดยแบ่งเป็น 6 กลุ่ม ด้วยโปรแกรม STATA

Partitional clustering

ก่อนทำการแบ่งกลุ่มต้อง Normalize ข้อมูลให้มีขนาดเท่ากันก่อน หลังจากนั้นทำการแบ่ง Cluster ด้วย K-means, K-medians ใช้วิธีการวัดระยะห่าง Euclidean distance ในการทำ RFM, RFMPHQ โมเดลตัวแปรที่นำมาแบ่ง Cluster ประกอบด้วยตัวแปรดังต่อไปนี้

- M (Monetary) ค่าเฉลี่ยการเบิกจ่ายค่ายารายครั้ง
- R (Recency) จำนวนวันจากการมารับบริการล่าสุด
- F (Frequency) ความถี่ของการมารับบริการ
- P (Periodicity) ส่วนเบี่ยงเบนมาตรฐานของระยะเวลาที่มาใช้บริการ
- H (HospVisit) จำนวนสถานพยาบาลที่ไม่มารับบริการ
- Q (Quantity) ปริมาณยาที่เบิกจ่าย
- Age อายุ

ผลการวิจัย

ค่าสถิติและความสัมพันธ์

จากตารางที่ 1 พบว่า (Q) มีความสัมพันธ์กับค่าเฉลี่ยการเบิกจ่ายค่ายารายครั้ง (M) เท่ากับ 9% โดยที่ (H) มีความสัมพันธ์กับความถี่ (F) เท่ากับ 32% ซึ่งมีตัวแปรทั้งหมดมีความสัมพันธ์กันค่อนข้างน้อย

ตารางที่ 1 แสดงค่าสัมพัทธ์ค่าสถิติ Correlation ระหว่างตัวแปร

	Q	M	R	F	P	H	Age
Q	1						
M	0.090	1					
R	-0.032	0.011	1				
F	-0.008	-0.037	-0.142	1			
P	0.049	0.025	-0.074	-0.301	1		
H	-0.028	-0.090	-0.140	0.322	-0.186	1	
Age	0.125	0.072	-0.108	0.143	-0.089	0.125	1

ที่มา: สำนักงานวิจัยเพื่อการพัฒนาหลักประกันสุขภาพไทย (สวปก)

จากตารางที่ 2 ข้อมูลการมารับบริการที่สถานพยาบาลจำนวน 21 ล้านครั้ง จำนวนผู้มารับบริการ 2.6 ล้านคน คิดข้อมูลผู้มารับบริการที่มีการเบิกจ่ายค่ายาฉบับบ้านเฉลี่ย 6.3 แสนคน จำนวนเงินที่มีการเบิกจ่ายค่ายารายครั้ง (M) เฉลี่ยต่อคนมากที่สุด 240,906 บาทต่อครั้ง มีความถี่ (F) ในการมามากสุดที่ 710 ครั้ง ช่วงระยะเวลาที่มารับบริการ (P) สูงสุด 266 วัน ถ้าคำนวณแสดงถึงผู้มารับบริการเป็นประจำ จำนวนสถานพยาบาล ที่ไปใช้บริการมากที่สุด 32 แห่ง มีการเบิกจ่ายฉบับบ้านสูงสุด (Mdrug) 2.9 ล้านบาท ปริมาณยาที่ได้รับจำนวนมากสุด 6,700 หน่วย

ตารางที่ 2 แสดงหาค่าสถิติของตัวแปร

Stats	Mean	Median	Max	Min	SD
M	890	184	340,907	1	3,269
Mdrug	8,580	1,410	2,943,456	1	37,133
R	98	62	366	1	89
F	12	8	710	1	17
P	34	29	266	0	28
H	2	2	32	1	1
Q	67	30	6,700	1	99
Age	57	59	100	0	19

ที่มา: สำนักงานวิจัยเพื่อการพัฒนาหลักประกันสุขภาพไทย (สวปก)

RFMPHQ score

การจัดอันดับผู้มารับบริการด้วยการแบ่งกลุ่มตามโมเดล RFM Segmentation ค่า score ซึ่งแบ่งได้ 8,893 กลุ่ม ที่มีผู้มารับบริการสูงสุด (RFMPHQ = 112612) จำนวน 3,145 คน คิดเป็น 0.68% ความหมายคือผู้รับบริการมาไม่บ่อยแต่มีการเบิกจ่ายค่ายากลับบ้านจำนวนมากและไปแค่สถานพยาบาลเดียว) ค่า score (RFMPHQ = 566666) มีจำนวน 112 คน คิดเป็น 0.02% ความหมายคือ ผู้มารับบริการมาเป็นประจำ ,มีการเบิกจ่ายค่ายาจำนวนมาก ไปหลายสถานพยาบาล และปริมาณการเบิกจ่ายสูง การประยุกต์ใช้โมเดล LRFMP เป็นเทคนิคที่เกิดจาก Ha and Park (1998) ซึ่งใช้สถิติคุณลักษณะ (x) กับค่าเฉลี่ยของตัวแปร L, R, F, M, IVT หรือ P ที่มีค่ามากกว่าค่าเฉลี่ยรวม และจะใช้ (j) ถ้าค่าที่ได้มีน้อยกว่าค่าเฉลี่ยรวมจากตารางที่ 3, 4 เป็นการประยุกต์ใช้ค่า (H) แทนค่า (L), (P) เพื่อสถานะผู้มารับบริการเป็นประจำ ค่าที่ได้จะค่าและ (H) จำนวนสถานพยาบาลที่ไม่รับบริการเข้าใน RFMPH

ตารางที่ 3 ค่าเฉลี่ยโมเดล RFMPHQ score ด้วย K-means

Cluster	N	R	F	M	P	H	Q	RFMPHQ	Size
C1	267,968	31.69	15.16	861.36	26.81	1.93	69.76	FHQ	50%
C2	36,102	66.97	3.47	820.61	106.98	1.47	63.67	P	7%
C3	87,961	189.04	8.79	1082.72	33.81	1.63	66.03	RM	16%
C4	48,387	299.61	8.94	856.30	22.66	1.60	66.26	R	9%
C5	100,601	107.07	10.91	836.72	30.01	1.79	66.52	R	19%
Total	639,919	97.66	12.01	890.01	34.38	1.80	66.93		100%

ที่มา: สำนักงานวิจัยเพื่อการพัฒนาหลักประกันสุขภาพไทย (สวปก)

จากตารางที่ 4 การคำนวณค่า RFMPHQ score แบ่งด้วย K-medians สามารถจำแนกได้ 6 ประเภท ประเภทแรก (FH) ใน Cluster 1 คิดเป็น 37% โดยมีค่า F สูงสุด 18 ครั้ง ค่าเฉลี่ยจำนวนสถานพยาบาลที่ไป (H) 2.06 แห่ง ประเภทที่ 2 (MPQ) ใน Cluster 2 คิดเป็น 18% โดยมี (M) เท่ากับ 1,010 บาท

ตารางที่ 4 ค่าเฉลี่ยโมเดล RFMPHQ score ด้วย K-medians

Cluster	N	R	F	M	P	H	Q	RFMPHQ	Size
C1	197,783	29.17	18.16	787.63	22.01	2.06	66.97	FH	37%
C2	96,618	42.81	6.37	1010.36	67.71	1.63	76.71	MPQ	16%
C3	92,327	179.48	8.61	1064.37	37.22	1.61	66.68	RMP	17%
C4	66,066	291.20	8.98	866.08	23.67	1.60	66.74	R	10%
C5	96,136	102.86	11.32	823.66	29.16	1.63	66.20	RH	16%
Total	639,919	97.66	12.01	890.01	34.38	1.80	66.93		100%

ที่มา: สำนักงานวิจัยเพื่อการพัฒนาหลักประกันสุขภาพไทย (สวปก)

RFMPHQ Score ด้วยการแบ่งกลุ่ม K-means, K-medians

จากตารางที่ 3 การคำนวณค่า RFMPHQ score ค่าที่มีลูกศรขึ้น (↑) จะปรากฏตัวแปรในค่า score ส่วนค่าที่มีลูกศรลง (↓) จะไม่ปรากฏตัวแปรใน Cluster จากการแบ่งกลุ่มด้วยเทคนิค K-means สามารถจำแนกได้ 4 ประเภท ประเภทแรก (FHQ) คิดเป็น 50% ใน Cluster 1 มีค่าเฉลี่ยความถี่มาสถานพยาบาล (F) 15 ครั้ง ค่าเฉลี่ยจำนวนสถานพยาบาลที่ไป (H) 1.93 โรงพยาบาล และปริมาณยาที่เบิกกลับบ้านเฉลี่ย (Q) จำนวนเฉลี่ย 69.7 หน่วย ประเภทที่ 2 (R) ใน Cluster 4,5 จำนวนวันเฉลี่ยที่ผู้ป่วยมารับบริการ ล่าสุด 299, 107 วัน คิดเป็น 9% และ 19% ประเภทที่ 3 (RM) ใน Cluster 3 คิดเป็น 16% มีการเบิกจ่ายค่ายามากกว่าค่าเฉลี่ย เท่ากับ 1,082 บาท ประเภทที่ 4 (P) ใน Cluster 2 คิดเป็น 7% ระยะเวลาเฉลี่ยที่มาใช้บริการประจำ 106 วัน

และ (Q) ปริมาณยาเฉลี่ยที่ได้รับ 76 หน่วย ประเภทที่ 3 (RMP) ใน Cluster 3 คิดเป็น 17% โดยมี (M) เฉลี่ยสูงสุด 1,064 บาท ประเภทที่ 4 (R) ใน Cluster 4 คิดเป็น 10% ประเภทที่ (RH) ใน Cluster 5 คิดเป็น 16% โดยมีค่า (H) เฉลี่ยเท่ากับ 1.63 แห่ง

จากค่าที่ได้ตารางที่ 3, 4 เพื่อเลือกกลุ่มคนที่ลักษณะคล้ายกัน กลุ่มที่สนใจคือ (FHO) และ (FH) ทำการ intersect ระหว่าง K-means = C1, K-medians= C1 พบกลุ่มคนที่มีลักษณะเดียวกัน 197,783 คน ในตารางที่ 5 และเพื่อตรวจสอบว่ากลุ่มคนดังกล่าวอยู่ในกลุ่มเดียวกับโมเดล RFMPHQ จึงทำการ intersect อีกครั้ง

ตารางที่ 5 Intersect K-means, K-medians

K-medians	K-means					Total
	C1	C2	C3	C4	C5	
C1	197,783	64,865	0	0	5,320	267,968
C2	0	30,309	3,568	0	1,225	35,102
C3	0	0	80,292	7,668	1	87,961
C4	0	0	0	48,387	0	48,387
C5	0	3,444	8,487	0	88,590	100,501
Total	197,783	98,618	92,327	56,055	95,136	539,919

ที่มา: สำนักงานวิจัยเพื่อการพัฒนาหลักประกันสุขภาพไทย (สวปก)

ทำการ intersect กลุ่ม FHO กับโมเดล RFMPHQ จากกลุ่ม FHO=665 มีจำนวน 4,694 คน (F=6) คือ ความถี่ของการมารับบริการ อยู่ระหว่าง 17-710 ครั้ง (H=6) จำนวนสถานพยาบาลที่ไม่มารับบริการ 3-32 แห่ง และ (Q=6) ปริมาณยาที่เบิกจ่าย 101-5700 หน่วย

ตารางที่ 6 รูปแบบพฤติกรรมกรรมการมารับบริการ

กลุ่ม	คำอธิบาย
FHO	มาบ่อย, ไปหลายพ. ปริมาณยาจำนวนมาก
FM	มาบ่อย, เบิกเคลมจำนวนมาก
MPQ	เบิกเคลมจำนวนมาก, มาประจำ, ปริมาณยาจำนวนมาก
RMP	นานครั้งมา, เบิกเคลมจำนวนมาก

จากตารางที่ 6 แสดงรูปแบบพฤติกรรมกรรมการมารับบริการ ที่ได้จากการวิเคราะห์โมเดล RFMPOH สามารถเห็นความสัมพันธ์ระหว่างตัวแปรและอ่านตีความหมายได้เป็นคู่ พบว่า ความถี่ (F) มีความสัมพันธ์กับจำนวนเงินที่เบิก (M) กลุ่มที่มาบ่อยและเบิกจ่ายจำนวนมาก (F↑ M↑) กลุ่มที่นานครั้งมาแต่เบิกจ่ายมาก (F↓ M↑) กลุ่มที่นานครั้งมาและเบิกจ่ายน้อย (F↓ M↓) และกลุ่ม (F↑ M↓) มาบ่อยแต่เบิกจ่ายน้อย อีกในหนึ่งผู้มารับบริการมากที่สุด (R) และความถี่ (F) กลุ่ม (R↑ F↓) นานๆ มาครั้งความถี่ต่ำ และ (R↓ F↓) มาใช้บริการเร็วๆ นี้ แต่มีความถี่ต่ำ กลุ่มที่ (R↓ F↑) มาใช้บริการเมื่อเร็วๆ นี้และมีความถี่สูง

ความสัมพันธ์ซิกในหนึ่งระหว่างความถี่ (F) และจำนวนสถานพยาบาล ที่ไปรับบริการ (H) กลุ่ม (F↑ H↑) มาใช้บริการบ่อยแล้วไปหลายโรงพยาบาล กลุ่มที่ไม่ค่อยมาโรงพยาบาล (F↓ H↓) และกลุ่มที่เบิกจ่ายเงินมากไปรักษาหลายแห่ง (M↑ H↑) กลุ่ม (M↓ H↓) ผู้ป่วยเบิกจ่ายน้อยและไม่รักษาแต่แห่งเดียว และกลุ่มผู้ป่วยที่เบิกจ่ายมากและไปแค่แห่งเดียว (M↑ H↓)

วิจารณ์ผลและสรุป

งานวิจัยชิ้นนี้นำเสนอผลการศึกษาพฤติกรรมกรรมการมารับบริการที่สถานพยาบาลโดยทำการเบิกจ่ายค่ายานากลับไปใช้ที่บ้านบ่อยครั้งเกินความจำเป็น โดยการวิจัยได้ประยุกต์ใช้โมเดล RFM กับการเคลมด้านสุขภาพ ซึ่งเดิมโมเดล RFM ใช้จัดกลุ่มลูกค้าที่มาซื้อสินค้า เพื่อนำไปกำหนดกลยุทธ์ทางการตลาด การนำโมเดล RFM มาประยุกต์ใช้ครั้งนี้ใช้ตัวแปร วันล่าสุดที่มารับบริการ, ความถี่, และการเบิกจ่ายเฉลี่ยต่อครั้ง อาจพบรูปแบบการมารับบริการจากค่า RFM score แต่เพื่อให้ชัดเจนมากขึ้น ได้นำเทคนิค K-means, K-medians แบ่งกลุ่ม 6 กลุ่ม (k=6) เพื่อให้สามารถเปรียบเทียบกับโมเดล RFM 5 กลุ่มได้ พบกลุ่มที่น่าสนใจคือ FHO คือ กลุ่มที่ไม่มารับบริการบ่อย, ไปหลายสถานพยาบาล และปริมาณยาที่ได้รับกลับมามีจำนวนมาก เพื่อให้ทราบว่าเป็นกลุ่มคนเดียวกัน ทำการ intersect กลุ่ม K-means, K-medians ได้กลุ่มคนที่มีลักษณะเดียวกัน 197,783 คน อย่างไรก็ตามกลุ่มดังกล่าวยังกว้าง เมื่อเทียบกับการทำด้วยโมเดล RFM ที่สามารถเรียกกลุ่มได้ชัดเจนกว่า ดังนั้นผู้วิจัยจึงทำการ intersect อีกครั้งกับโมเดล RFM ทำให้ได้ค่า FHO score = 665 จำนวน 4,694 คน สามารถแยกกลุ่มที่มีลักษณะเดียวกันได้ชัดเจนมากขึ้น และเพื่อเพิ่มประสิทธิภาพในการศึกษารูปแบบที่ชัดเจนขึ้น การประยุกต์ใช้โมเดล RFMPOH ทำการวิเคราะห์ในแต่ละด้านตีความค่าเป็นคู่ สามารถเห็นรูปแบบลักษณะของผู้มารับบริการได้ชัดเจนมากยิ่งขึ้น ทักล่าวมาทั้งหมดนี้การใช้เทคนิคค่าตัวโมเดลในการแบ่งกลุ่มผู้มารับบริการที่มีปริมาณข้อมูลจำนวนมาก สามารถทำให้เห็นรูปแบบที่ชัดเจนอยู่ในข้อมูลได้ และตรวจสอบข้อมูลได้ง่ายขึ้น ใช้เวลาไม่มาก และลดค่าใช้จ่ายในการจ้างผู้เชี่ยวชาญมาทำการตรวจสอบ

ข้อจำกัดในการวิจัยครั้งนี้คือ หน่วยของยาที่แตกต่างกันไป เช่น ยาเม็ด ยาขวด ยาผง และอื่นๆ ซึ่งการนำมาใช้ในงานวิจัยครั้งนี้เป็นการรวมยอดจำนวนหน่วยของยา ถ้ามีการแบ่งหน่วยยาให้ชัดเจนกว่านี้ อาจจะทำให้การวิเคราะห์แม่นยำมากขึ้น แนวทางการพัฒนาในอนาคตได้มีการเพิ่มเทคนิคค่าตัวโมเดล เช่น เทคนิค DBSCAN Clustering เพื่อแบ่งกลุ่มข้อมูลที่แตกต่างกันอยู่ในกลุ่มเดียวกัน และสร้างโมเดล Decision tree ซึ่งต้องหาตัวอย่างกลุ่มคนที่มีพฤติกรรมผิดปกติเพื่อใช้ในการ

รันโมเดล เดิมทีโมเดล RFMPQH ก็ให้ผลที่ดีและแม่นยำแล้ว แต่ผู้วิจัยต้องการเปรียบเทียบหลายเทคนิค เพื่อค้นหากลุ่ม ผู้มีพฤติกรรมผิดปกติที่ซ่อนอยู่ในข้อมูลโดยใช้เวลาไม่มากและมีประสิทธิภาพในการตรวจสอบมากที่สุด

เอกสารอ้างอิง

1. กรมบัญชีกลาง, คู่มือสวัสดิการรักษายาพยาบาลข้าราชการ เล่ม 1, 2551, หน้า 8-38
2. เอกสิทธิ์ พิธรวงศ์ศึกษา, Introduction to Business Analytics with RapidMiner Studio 6, 2558, หน้า 67-90
3. Peker S, Kocyigit A, Eren PE. LRFMP model for customer segmentation in the grocery retail industry a case. *Market Intel Plan.* 2017;36(4) :644-69.
4. Wu HH, Lin SY, Liu CW. Analyzing Patients' Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan. *Sci World J.* 2014; DOI: 10.1155/2014/886496.
5. Geraili B, Nasiri M, Arab M. Improving Fraud and Abuse Detection in General Physician Claim A Data Mining Study. *Int J Health Politic Manag.* 2016;6(3):166-72.
6. Thornton D, van Capelleveen G, Poel M, van Hillegersberg J, Mueller RM. Outlier-based Health Insurance Fraud Detection for U.S. Medicaid. In *Proceedings of the 18th International Conference on Enterprise Information Systems.* 2014:684-94.
7. Liu Q, Vasarhelyi M. Healthcare fraud detection A survey and a clustering model incorporating Geo-location information. in *29th World Continuous Auditing and Reporting Symposium (29WCARS), Brisbane, Australia.* 2013:1-10.

DDP

บรรณานุกรม



บรรณานุกรม

- ดิเรก ถิรมจรุสกุล, 2554 น. 512
- เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2557). An Introduction to Data Mining Techniques.
- เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2560). การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data) และเทคนิคการขุด
 ไม่นิ่ง (Data Mining). (pp. 69).
- Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke. (2015). Fraud Analytics Using
 Descriptive Predictive and Social Network Techniques A Guide to Data Science for Fraud
 Detection.
- Bijan Geraili, Mahdi Nasiri, Mohammad Arab. (2015). Improving Fraud and Abuse Detection in
 General Physician Claim A Data Mining Study.
- Dallas Thornton, Guido van Capelleveen, Mannes Poel, Jos van Hillegersberg and Roland M. Mueller.
 (2014). Outlier-based Health Insurance Fraud Detection for U.S. Medicaid.
- Hsin-Hung Wu, Shih-Yen Lin, and Chih-Wei Liu. (2014). Analyzing Patients' Values by
 Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan.
- Hughes (1996). Boosting response with RFM.
- Karca Duru Aral a, Halil Altay Güvenir b, İhsan Sabuncuoğlu c, Ahmet Ruchan Akar d, e.
 (2012). A prescription fraud detection model.
- Prashant Natarajan, John C Frenzel, Detlev H Smaltz. (2017). (p. 89-98). Demystifying Big Data
 and Machine Learning for Healthcare.
- Qi Liu, Miklos Vasarhelyi. (2013). Healthcare fraud detection A survey and a clustering model
 incorporating Geo-location information.
- Reinartz and Kumar. (2000). On The Profitability of Long Lifetime Customer: An Empirical
 Investigation and Implications for Marketing.
- Serhat Peker, Altan Kocyigit and P. Erhan Eren. (2017). LRFMP model for customer
 segmentation in the grocery retail industry: a case study.

ประวัติผู้เขียน

ชื่อ-นามสกุล

ประวัติการศึกษา

ประไพพร อุดมา

เศรษฐศาสตรบัณฑิต สาขาการเงินและธนาคาร

มหาวิทยาลัยรามคำแหง

เศรษฐศาสตรมหาบัณฑิต สาขาการเงินและธนาคาร

มหาวิทยาลัยรามคำแหง

ตำแหน่งและสถานที่ทำงานปัจจุบัน

ประสบการณ์ ผลงานทางวิชาการ

นักวิจัย สถาบันวิจัยระบบสาธารณสุข

- งานตีพิมพ์ The use of data mining to investigate abnormal behavior of outpatient drug reimbursement under civil servant medical benefit scheme
- ปัจจัยที่มีผลต่อต้นทุนค่ารักษาพยาบาลของผู้ป่วยสิทธิสวัสดิการรักษายาบาลข้าราชการ และพนักงานส่วนท้องถิ่น
- Effect of Cost Containment Strategies to Patterns of Patient Services in the Civil Servant Medical Benefit Scheme
- Performance of health care services under Universal health care coverage scheme in Bangkok, Thailand
- Tax Efficiency in the Provinces of Thailand