

# เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการแพทย์พยาบาล

## Data Mining Techniques for Nursing Data Analysis

ศุภามณ จันทร์สกุล

Suphamon Chansakul

คณะพยาบาลศาสตร์ มหาวิทยาลัยอีสเทิร์นเอเซีย

Faculty of Nursing, Eastern Asia University

### บทคัดย่อ

ปัจจุบันมีข้อมูลต่าง ๆ และฐานข้อมูลที่เกี่ยวข้องในวิชาชีพการพยาบาลเป็นจำนวนมากแต่การนำข้อมูลต่าง ๆ และฐานข้อมูลเหล่านั้นมาใช้ประโยชน์โดยค้นหารูปแบบความสัมพันธ์ของข้อมูลซึ่งเรียกว่า “การวิเคราะห์เหมืองข้อมูล” ยังมีอยู่น้อย วัตถุประสงค์ของบทความวิชาการเพื่อนำเสนอความรู้เบื้องต้นในการทำเหมืองข้อมูลสำหรับพยาบาล การวิเคราะห์เหมืองข้อมูลแบ่งออกเป็นสองกลุ่มใหญ่คือ (1) การวิเคราะห์เหมืองข้อมูลเพื่อการทำนายมีการเรียนรู้แบบมีการสอน และ (2) การวิเคราะห์เหมืองข้อมูลเพื่อการอธิบายมีการเรียนรู้แบบไม่มีการสอน เทคนิคเหมืองข้อมูล 3 เทคนิคที่ได้รับความนิยมได้แก่ การจำแนกประเภท (ต้นไม้ตัดสินใจ และเครือข่ายประสาท) การจัดกลุ่มข้อมูล และการค้นหา กฎความสัมพันธ์ของข้อมูล ผลการวิเคราะห์ที่ได้จากเหมืองข้อมูลสร้างองค์ความรู้ และเป็นข้อค้นพบที่เป็นประโยชน์ในการนำไปใช้เพื่อการตัดสินใจ

**คำสำคัญ:** เหมืองข้อมูล, การจำแนกประเภท, เครือข่ายประสาท, ต้นไม้ตัดสินใจ, การวิเคราะห์ข้อมูลทางการแพทย์พยาบาล

### Abstract

In the nursing profession, there are various information and databases. But the identifying patterns of data relationships hidden in the data set are called “Data mining”, there is a few utilization. The purpose of this article is to provide an introduction to data mining for nursing. Data mining algorithms are divided into two major groups: (1) predictive (supervised learning) and (2) descriptive (unsupervised learning). Three of the most widely-used data mining algorithms such as classification (decision tree and neuron networks), clustering and association rule. Analysis results from data mining reveal, knowledge creation and useful discovery for decision making.

**Keywords:** data mining, classification, neuron networks, decision tree, nursing data analysis



## บทนำ

ปัจจุบันเป็นยุคแห่งข้อมูลข่าวสารมีสารสนเทศมากมายในทุกศาสตร์ ทั้งทางด้านการศึกษา ด้านธุรกิจ การตลาด ด้านสุขภาพ (healthcare) ซึ่งประกอบด้วย ข้อมูลทางการแพทย์และการพยาบาล ข้อมูลในวิชาชีพ การพยาบาลมีหลากหลาย ทั้งทางด้านการศึกษาทางการพยาบาล การบริหารทางการพยาบาล การปฏิบัติการพยาบาล และการวิจัยทางการพยาบาล โดยฐานข้อมูลทางการแพทย์ต่าง ๆ ได้แก่ Pubmed, CINALH, ProQuest nursing & Allied health source, Clinical key nursing, Ovid, DAO, ScienceDirect เป็นต้น (Mahidol University, 2015; Jirathummakoon, 2006) การเกิดขึ้นของข้อมูลต่าง ๆ และฐานข้อมูลจำนวนมากนำไปสู่การพัฒนาเทคโนโลยีสารสนเทศด้วยปัญญาประดิษฐ์ (Artificial Intelligence--AI) อันเกี่ยวข้องกับวิธีการที่ทำให้คอมพิวเตอร์มีความสามารถคล้ายมนุษย์หรือเลียนแบบพฤติกรรมมนุษย์ การทำเหมืองข้อมูล (data mining) เป็นหนึ่งในการประยุกต์ใช้ปัญญาประดิษฐ์ตั้งแต่ประมาณปี ค.ศ.1960 (Liao, Chu & Hsiao, 2012) เมื่อค้นหาคำสำคัญว่า “data mining technique” จากบทคัดย่อบนฐานข้อมูลออนไลน์ของ Elsevier SCOPUS, Springerlink, IEEE Xplore, EBSCO และ Wiley InterScience online database ตั้งแต่เดือนมกราคม ค.ศ.2000 ถึงเดือนสิงหาคม ค.ศ.2011 พบว่าเทคนิคเหมืองข้อมูลที่ได้รับความนิยม 4 อันดับแรกได้แก่ Decision tree, Artificial neural network, Clustering และ Association rule ตามลำดับ (Liao, Chu & Hsiao, 2012)

การทำเหมืองข้อมูลด้วยเทคนิควิธีการจำแนกกลุ่ม (classification) ได้แก่ การวิเคราะห์เครือข่ายประสาท (neuron networks) และการวิเคราะห์ต้นไม้ตัดสินใจ (decision tree) ใช้การเรียนรู้แบบมีการสอน (supervised learning) มีหลักการทำงานโดยแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนแรกเป็นข้อมูลสำหรับการเรียนรู้ (training data) และส่วนที่สองคือข้อมูลสำหรับใช้ทดสอบ (testing data) เพื่อประเมินความถูกต้องของตัวแบบ (model evaluation) โดยใช้ข้อมูลที่ไม่เคยเห็น

มาก่อน (unseen data) เพื่อกำหนดกลุ่มให้กับข้อมูลใหม่หรือทำนายค่าของการจัดหมวดหมู่ (Sripaoraya & Sinsomboonthong, 2017)

สำหรับบทความวิชาการนี้มีวัตถุประสงค์เพื่อนำเสนอสาระเกี่ยวกับเทคนิคเหมืองข้อมูล การนำไปใช้และประโยชน์ของเหมืองข้อมูลทางการแพทย์ ส่วนตัวอย่างของการวิเคราะห์ผู้เขียนขอแนะนำตัวอย่างเหมืองข้อมูลที่ใช้เทคนิควิธีการจำแนกกลุ่ม (classification) ได้แก่ การวิเคราะห์เครือข่ายประสาท (neuron networks) และการวิเคราะห์ต้นไม้ตัดสินใจ (decision tree) ด้วยโปรแกรมสำเร็จรูป SPSS โดยข้อมูลที่นำมาเป็นตัวอย่างในการวิเคราะห์เป็นข้อมูลทางการศึกษาทางการแพทย์

## เทคนิคเหมืองข้อมูล

### (Data mining techniques)

1. เหมืองข้อมูล (data mining) เป็นเทคนิคการวิเคราะห์ข้อมูลเพื่อค้นหารูปแบบ (pattern) ความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติและการเรียนรู้ของเครื่องคอมพิวเตอร์แล้วนำผลวิเคราะห์ที่ได้ไปใช้ประโยชน์เพื่อการตัดสินใจ (Wongwiwatanon, 2008; Na-Wichien, 2017b; Srisa-An, 2017) ในการเรียนรู้ของเครื่องคอมพิวเตอร์เพื่อให้ใกล้เคียงกับสมองมนุษย์ (machine learning) มีชนิดของการเรียนรู้ข้อมูล 2 แบบคือ (1) การเรียนรู้แบบมีการสอน (supervised learning) เป็นการเรียนรู้จากข้อมูลที่มีผู้สอน หรือกล่าวได้ว่าข้อมูลนั้นมีการกำหนดคลาสหรือค่าเฉลย และ (2) การเรียนรู้แบบไม่มีการสอน (unsupervised learning) เป็นการเรียนรู้แบบไม่มีผู้สอน หรือกล่าวได้ว่าข้อมูลนั้นไม่มีการกำหนดคลาสซึ่งส่วนใหญ่เป็นข้อมูลธรรมชาติที่พบได้โดยทั่วไป (Prachuabsupakij, 2018 ; Yoo et al., 2012) กระบวนการทำเหมืองข้อมูลใช้โมเดลในการเรียนรู้ จึงต้องนำอัลกอริทึมของการเรียนรู้ของเครื่อง (machine learning) มาใช้โดยชนิดของอัลกอริทึมในการทำเหมืองข้อมูล (data mining algorithms) แบ่งได้เป็นสองกลุ่มใหญ่ ดังนี้ (Na-Wichien, 2017b; Srisa-An, 2017; Yoo et al., 2012)

1.1 การวิเคราะห์เหมืองข้อมูลเพื่อการทำนาย (predictive) มีการเรียนรู้แบบมีการสอน (supervised learning) โดยเรียนรู้จากข้อมูลที่มีอยู่เดิมเพื่อทำนายข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต เช่น ถ้าข้อมูลมีการจัดลำดับของผู้เรียนไว้แล้วว่าใครเป็นผู้เรียนกลุ่มเก่ง ใครเป็นผู้เรียนกลุ่มปานกลาง และใครเป็นผู้เรียนกลุ่มอ่อน เหมืองข้อมูลสามารถเรียนรู้จากข้อมูลเหล่านี้และค้นหาโมเดลที่สามารถใช้อธิบายลักษณะของผู้เรียนกลุ่มต่าง ๆ จากโมเดลที่ได้นี้สามารถนำไปใช้ทำนายผู้เรียนใหม่ว่าน่าจะเป็นผู้เรียนกลุ่มประเภทใด เทคนิคการวิเคราะห์ได้แก่ การจำแนกประเภท (classification) หรือการสร้างตัวแบบทำนาย (prediction models)

1.2 การวิเคราะห์เหมืองข้อมูลเพื่อการอธิบาย (descriptive) มีการเรียนรู้แบบไม่มีการสอน (unsupervised learning) โดยค้นหารูปแบบความสัมพันธ์หรือลักษณะที่เชื่อมโยงกันของข้อมูล การวิเคราะห์ต่างจากแบบแรกตรงที่ไม่ได้กำหนดล่วงหน้าว่าจะให้โปรแกรมวิเคราะห์เหมืองข้อมูลค้นหาหรือโมเดลแต่ให้ค้นหาทุกรูปแบบที่น่าสนใจจากข้อมูล เทคนิคการวิเคราะห์ได้แก่ การจัดกลุ่ม (clustering) และการหากฎความสัมพันธ์ (association rule)

2. เทคนิคเหมืองข้อมูลที่ได้รับความนิยมในการนำไปใช้มี 3 เทคนิค ดังนี้ (Na-Wichien, 2017b; Srisa-An, 2017; Yoo et al., 2012)

2.1 การจำแนกประเภท (classification & prediction) มีการเรียนรู้แบบมีการสอนซึ่งกำหนดคลาส (class) ไว้แล้วในการจัดประเภท คลาสในการจำแนกประเภทเป็นคุณลักษณะ (attribute or feature) ในชุดข้อมูลนั้น ๆ ที่ผู้วิจัยสนใจ หรืออีกนัยหนึ่งคือตัวแปรตาม (dependent variable) ในการวิเคราะห์ทางสถิติ ตัวอย่าง เช่น การสูบบุหรี่ (smoking) กำหนดคลาสไว้ 2 อย่างคือ Yes หรือ No เป็นต้น กระบวนการในการจัดแบ่งประเภทของข้อมูลมี 2 ขั้นตอนคือ (1) การหาชุดต้นแบบซึ่งสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (training data) และ (2) ชุดข้อมูลทดสอบ (testing data) โดยตรวจสอบความตรงในการจำแนกเมื่อนำต้นแบบที่ได้มาใช้ทำนาย

ข้อมูลจำแนกกลุ่ม เช่น การหาความสัมพันธ์ระหว่างผลการตรวจร่างกาย ต่าง ๆ กับการเกิดโรคโดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย สำหรับเทคนิควิธีการจำแนกกลุ่ม (classification) ที่นิยมได้แก่ การวิเคราะห์เครือข่ายประสาท (neuron networks) และการวิเคราะห์ต้นไม้ตัดสินใจ (decision tree)

2.2 การจัดกลุ่มข้อมูล (clustering) มีการเรียนรู้แบบไม่มีการสอนซึ่งแตกต่างจากการจำแนกประเภทโดยการจัดกลุ่มจะไม่มีกำหนดคลาสไว้ก่อนจึงไม่มีต้นแบบที่สร้าง สำหรับขั้นตอนของการจัดกลุ่มมีการจัดข้อมูลที่มีคุณลักษณะคล้ายกันจัดเข้ากลุ่มเดียวกัน เช่น การจัดกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน เป็นต้น การจัดกลุ่มเหมาะสำหรับการศึกษาหรือสำรวจข้อมูลตามธรรมชาติที่มีจำนวนมากและไม่ทราบคุณลักษณะของข้อมูลเหล่านั้น สำหรับอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูลได้แก่ Hierarchical clustering algorithms (single-link, complete-link, average-link) และ Partitional clustering algorithms (K-means, K-medoids)

2.3 การค้นหากฎความสัมพันธ์ของข้อมูล (association rule discovery) มีการเรียนรู้แบบไม่มีการสอนโดยค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลมากกว่าสองชุดขึ้นไป บางครั้งเรียกเทคนิคนี้ว่า “วิเคราะห์ตะกร้าตลาด (market basket analysis)” ซึ่งศึกษาพฤติกรรมผู้บริโภคโดยวิเคราะห์ความสัมพันธ์จากใบรายการซื้อสินค้า (transaction) เช่น เมื่อลูกค้าซื้อขนมปังจะซื้อเนยด้วย เป็นต้น สำหรับอัลกอริทึมที่ใช้ในการค้นหาความสัมพันธ์ของข้อมูลได้แก่ Apriori algorithm โดยหากความสัมพันธ์ที่มีค่าสนับสนุน (support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้มีความถูกต้อง และค่าความมั่นใจ (confidence) ซึ่งเป็นจำนวนของกรณีที่ถูกถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่ถูกสามารถนำไปใช้ได้ ซึ่งแตกต่างจากการจำแนกประเภทและการจัดกลุ่มข้อมูลที่ทำให้ความสำคัญใน

การตรวจสอบความตรง (accuracy) ของการจำแนกประเภท และการจัดกลุ่ม การหาความสัมพันธ์จากข้อมูลทางการแพทย์ ได้แก่ การหาความสัมพันธ์ระหว่างอาการ (symptoms) ภาวะสุขภาพ (health conditions) และโรค (diseases) เป็นต้น

## การนำไปใช้และประโยชน์ของเหมืองข้อมูลทางการแพทย์

จากการทบทวนวรรณกรรมจากบทความวิจัยทั้งในไทยและต่างประเทศพบมีการทำเหมืองข้อมูลทางด้าน สุขภาพ การแพทย์ และการพยาบาล โดยรายละเอียดดังตาราง 1 และตาราง 2

### ตาราง 1

เทคนิคเหมืองข้อมูลทางด้านสุขภาพ การแพทย์ และการพยาบาลในประเทศไทย

ผู้วิจัย	งานวิจัย	เทคนิคเหมืองข้อมูล	โปรแกรมวิเคราะห์
เรวดี ศักดิ์ดุยธรรม (Sakduyatam, 2010)	การใช้เทคนิคดาต้าไมน์นิ่งในการวิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษาโรคนิ้วล็อกในแบบต่าง ๆ ของคณะแพทยศาสตร์ วชิรพยาบาล มหาวิทยาลัยกรุงเทพมหานคร	Clustering, Association rule	WEKA
นิเวศ จิระวิชิตชัย (Chirawichitchai, 2010)	การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ	Classification : Naïve Bayes, MLP, Radial Basis, Function network, SVM, K-nearest neighbor, Dicision tree, Ripper	WEKA
กิตติศักดิ์ สุมามัลย์ (Sumamal, 2012)	การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล	Association rule ศึกษาความสัมพันธ์ของการเป็นโรค Classification ศึกษาการคัดกรองสุขภาพเบื้องต้น	SPSS, WEKA
ทัศนัท ฐานประเสริฐกุล และกมล เกียรติเรืองมลา (Thanprasertkul & Keatruangkamala, 2014)	การศึกษาศาเหตุการเกิดโรคความดันโลหิตสูงชนิดไม่ทราบสาเหตุ กรณีศึกษาโรงพยาบาลภาครัฐแห่งหนึ่งในประเทศไทย	Clustering	-
วิไลรัตน์ วิศวไพศาล, บุญช่วย ศรีธรรมศักดิ์ และสาธิษฐ์ นากกระแสน์ (Wisvapaisan, Srithammasak & Nakkrasae, 2016)	ปัจจัยที่มีผลต่อการทำนายการคลอดก่อนกำหนดในหญิงตั้งครรภ์ในโรงพยาบาลตำรวจ	Decission tree จำแนกกลุ่มที่คลอดครบกำหนดและคลอดก่อนกำหนด	WEKA

## ตาราง 1

เทคนิคเหมืองข้อมูลทางด้านสุขภาพ การแพทย์ และการพยาบาลในประเทศไทย (ต่อ)

ผู้วิจัย	งานวิจัย	เทคนิคเหมืองข้อมูล	โปรแกรมวิเคราะห์
สุรวัชร ศรีเปารยะ และ สายชล สิ้นสมบุญทอง (Sripaoraya & Sinsomboonthong, 2017)	การเปรียบเทียบประสิทธิภาพวิธีการ จำแนกกลุ่มการเป็นโรคไตเรื้อรัง: กรณีศึกษาโรงพยาบาลแห่งหนึ่งใน ประเทศอินเดีย	Classification: Decision tree, NN, SVM, rule-based, logistic regression, Naïve Bayes จากฐานข้อมูล UCI Machine Learning Repository	SPSS, WEKA

## ตาราง 2

เทคนิคเหมืองข้อมูลทางด้านสุขภาพ การแพทย์ และการพยาบาลในต่างประเทศ

ผู้วิจัย	งานวิจัย	เทคนิคเหมืองข้อมูล	โปรแกรมวิเคราะห์
Madigan and Curet (2006)	A data mining approach in home healthcare: outcomes and service use	CART (Classification and Regression Trees)	-
Yoon et al. (2010)	Prediction models for burden of caregivers applying data mining techniques	classification: algorithms (J48, RandomForest, MultilayerPerceptron, AdaboostM1)	SPSS, WEKA
Barker and Pasupathy (2011)	Impact of fatigue on performance in registered nurses: Data mining and implications for practice	CHAID	-
Lee et al. (2011)	Application of data mining to the identification of critical factors in patient falls using a web-based reporting system	artificial neural network (ANN)	SPSS, STATISTICA 8.0
Santos et al. (2013)	A data mining system for providing analytical information on brain tumors to public health decision makers	Clustering, Classification, Association	Microsoft Analysis Services 2005®

ตาราง 2

เทคนิคเหมืองข้อมูลทางด้านสุขภาพ การแพทย์ และการพยาบาลในต่างประเทศ (ต่อ)

ผู้วิจัย	งานวิจัย	เทคนิคเหมืองข้อมูล	โปรแกรมวิเคราะห์
Almasalha et al. (2013)	Data mining nursing care plans of end of life patients: A study to improve healthcare decision making	Clustering	Rapid Miner
Raju et al. (2015)	Exploring factors associated with pressure ulcers: A data mining approach	Decision Tree	Program R
Ghaibeh, Setoguchi and Moriguchi (2016)	Mining nurse care data: A study case on pressure ulcer prediction	Decision Tree	-
Alexander and Wang (2017)	Big data analytics in heart attack prediction	Classification: Naïve Bayes, Decision Tree and ANN	WEKA

จากบทความวิจัยในตาราง 1 และ 2 แสดงการทำเหมืองข้อมูลทางด้านสุขภาพ โดยเทคนิคที่นิยม ได้แก่ การจัดประเภท และการจัดกลุ่ม จากการศึกษาดังกล่าว ทำให้เห็นประโยชน์ของเหมืองข้อมูลทางการพยาบาลดังนี้

1. ประโยชน์ทางคลินิก ได้แก่ การคัดกรองสุขภาพเบื้องต้น การวิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จในการรักษาพยาบาล การสร้างโมเดลการวิเคราะห์โรค การศึกษาสาเหตุของการเกิดโรค การศึกษาปัจจัยที่ทำให้ผู้ป่วยพลัดตกหกล้ม ปัจจัยที่ทำให้เกิดแผลกดทับ การพยากรณ์ภาวะหัวใจขาดเลือดเฉียบพลัน เป็นต้น ผลการวิเคราะห์ที่ได้เป็นประโยชน์ในการวางแผนการพยาบาลผู้ป่วยและการตัดสินใจให้การพยาบาลทางคลินิก

2. ประโยชน์ในด้านการบริหารการพยาบาล ได้แก่ ผลลัพธ์ของบริการเมื่อวิเคราะห์ต้นทุนและความคุ้มค่าในการรักษาพยาบาล การสร้างตัวแบบทำนายภาระของผู้ดูแลผู้ป่วย การพยากรณ์ความเหนื่อยล้าในการปฏิบัติงานของพยาบาลวิชาชีพ เป็นต้น ซึ่งผลวิเคราะห์ที่ได้จะเป็นประโยชน์ต่อผู้บริหารพยาบาลในการตัดสินใจและวางแผนงานทางการพยาบาลต่อไป

**ตัวอย่างการวิเคราะห์เหมืองข้อมูลโดยใช้เทคนิควิธีการจำแนกกลุ่ม (Classification)**

สำหรับตัวอย่างเทคนิควิธีการจำแนกประเภท (classification) ในครั้งนี้ใช้เทคนิคเครือข่ายประสาทและต้นไม้ตัดสินใจ วิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS เนื่องจากเป็นโปรแกรมที่นิยมใช้โดยทั่วไปในการวิจัยและผู้ใช้ซึ่งเป็นพยาบาลส่วนใหญ่มีความคุ้นเคยมากกว่าโปรแกรมอื่น ส่วนข้อมูลที่น่านำมาเป็นตัวอย่างวิเคราะห์เป็นข้อมูลทางการศึกษาพยาบาลประกอบด้วยข้อมูลระดับชั้นปี สไตส์การเรียนรู้แบบ VARK ความชอบต่อรูปแบบการจัดการเรียนการสอน แรงจูงใจใฝ่สัมฤทธิ์ในการเรียน การกำกับตนเองในการเรียน กำหนดให้เป็นตัวแปรต้นในการวิเคราะห์เพื่อทำนายตัวแปรตามคือ ผลสัมฤทธิ์ทางการเรียนซึ่งวัดจากเกรดเฉลี่ยสะสม (GPA) 1/2559 ของนักศึกษาพยาบาล มหาวิทยาลัยอีสเทิร์นเอเซีย รายละเอียดของข้อมูลดังนี้

ระดับชั้นปี (Level) มี 3 ชั้นปี คือ ปี 1, ปี 2 และ ปี 3



สไตล์การเรียนรู้แบบ VARK มี 4 รูปแบบ คือ (1) unimodal (V, A, R, K) (2) bimodal (VA, VR, VK, AR, AK, RK) (3) trimodal (VAR, VAK, VRK, ARK) และ (4) multimodal/VARK

ความชอบต่อรูปแบบการจัดการเรียนการสอน (LS\_PFL) มีความชอบ 3 รูปแบบ คือ (1) ผู้สอนบรรยายเนื้อหาเป็นหลัก (2) ผู้เรียนค้นคว้าด้วยตนเองเป็นหลัก และ (3) ผู้เรียนค้นคว้าด้วยตนเองร่วมกับผู้สอนบรรยายเนื้อหาในสัดส่วนพอ ๆ กัน

แรงจูงใจใฝ่สัมฤทธิ์ในการเรียน (MA\_gr) มีคะแนนเฉลี่ย 5 ระดับคือ มากที่สุด, มาก, ปานกลาง, น้อย และน้อยที่สุด นำข้อมูลมาจัดเป็น 3 กลุ่มโดย กลุ่มที่มีคะแนนเฉลี่ย 1-2.5 เป็นกลุ่มแรงจูงใจใฝ่สัมฤทธิ์ในการเรียนน้อย กลุ่มที่มีคะแนนเฉลี่ย 2.51-3.50 เป็นกลุ่มแรงจูงใจใฝ่สัมฤทธิ์ในการเรียนปานกลาง และกลุ่มที่มีคะแนนเฉลี่ย 3.51-5.00 เป็นกลุ่มแรงจูงใจใฝ่สัมฤทธิ์ในการเรียนมาก

การกำกับตนเองในการเรียน (SR\_gr) มีคะแนนเฉลี่ย 5 ระดับคือ มากที่สุด, มาก, ปานกลาง, น้อย และน้อยที่สุด นำข้อมูลมาจัดเป็น 3 กลุ่มโดย กลุ่มที่มีคะแนนเฉลี่ย 1-2.5 เป็นกลุ่มการกำกับตนเองในการเรียนน้อย กลุ่มที่มีคะแนนเฉลี่ย 2.51-3.50 เป็นการกำกับตนเองในการเรียนปานกลาง และกลุ่มที่มีคะแนนเฉลี่ย 3.51-5.00 เป็นกลุ่มการกำกับตนเองในการเรียนมาก

ผลสัมฤทธิ์ทางการเรียน (GPA\_gr) วัดจากเกรดเฉลี่ยสะสม (GPA) 1/2559 ของนักศึกษาพยาบาล นำข้อมูลมาจัดเป็น 2 กลุ่มโดยกำหนดเกรดเฉลี่ยสะสม > 2.86 เป็นกลุ่มผลสัมฤทธิ์ทางการเรียนดี และเกรดเฉลี่ยสะสม < 2.86 เป็นกลุ่มผลสัมฤทธิ์ทางการเรียนไม่ดี

จากตัวอย่างข้อมูลด้านการศึกษาทางการแพทย์ดังกล่าวที่นำมาวิเคราะห์ด้วยวิธีเครือข่ายประสาทเทียมและต้นไม้ตัดสินใจ ซึ่งใช้หลักการเรียนรู้แบบมีผู้สอนสามารถเรียนรู้จากข้อมูลต่าง ๆ ที่มีอยู่ แล้วค้นหาโมเดลที่สามารถใช้อธิบายลักษณะของนักศึกษาที่มีผลสัมฤทธิ์ทางการเรียนดีหรือไม่ดี โมเดลที่ได้จากการวิเคราะห์เหมือนข้อมูลสามารถนำไปใช้ทำนายนักศึกษาใหม่ได้ว่าน่าจะเป็นนักศึกษาประเภทใดระหว่างนักศึกษาที่มีผลสัมฤทธิ์ทางการเรียนดีหรือนักศึกษาที่มีผลสัมฤทธิ์

ทางการเรียนไม่ดี ผู้เขียนขอแนะนำหลักการวิเคราะห์เครือข่ายประสาทเทียมและต้นไม้ตัดสินใจและตัวอย่างในการวิเคราะห์ด้วยทั้ง 2 วิธี ในลำดับต่อไป

### หลักการวิเคราะห์เครือข่ายประสาท ตัวอย่างการวิเคราะห์และการแปลผล

การวิเคราะห์เครือข่ายประสาท (neuron network) มีโครงสร้างพื้นฐานและใช้หลักการคำนวณเลียนแบบการทำงานของระบบสมองและใยประสาทในสมองของมนุษย์ มีการส่งสัญญาณผ่านเซลล์ประสาทเรียกว่าโหนด (node) ที่เชื่อมโยงหลายชั้น การทำงานของเครือข่ายประสาทใช้โครงสร้างในการวิเคราะห์แบบ Multilayer Perception (MLP) ซึ่งเป็นโครงข่ายประสาทเทียมแบบหลายชั้นโหนด (node) ในแต่ละชั้นมีค่าน้ำหนัก (weight) ที่แตกต่างกัน การวิเคราะห์ที่ใช้โปรแกรมสำเร็จรูป SPSS แบ่งข้อมูล (partition) เป็นข้อมูลฝึก (training data) ข้อมูลทดสอบ (test data) และข้อมูลตรวจสอบ (holdout) ในสัดส่วน 2:2:1 กำหนดชนิดของข้อมูลฝึกแบบ Mini-batch ซึ่งเป็นวิธีผสมผสานแบบ Batch กับแบบ Online วิธีนี้เหมาะสมกับข้อมูลขนาดปานกลางโดยจะแบ่งข้อมูลเป็นกลุ่มเท่า ๆ กันและมีการปรับน้ำหนักการเชื่อมต่อ (update synaptic weight) เมื่อข้อมูลกลุ่มใดกลุ่มหนึ่งผ่านการฝึก (Na-Wichien, 2017a) จากข้อมูลทางการศึกษาพยาบาลที่นำมาเป็นตัวอย่างในการวิเคราะห์แสดงผลการวิเคราะห์และตัวอย่างการแปลผลการวิเคราะห์เครือข่ายประสาท MLP ดังนี้

#### Case Processing Summary

		N	Percent
Sample	Training	77	37.7%
	Testing	87	42.6%
	Holdout	40	19.6%
Valid		204	100.0%
Excluded		3	
Total		207	

ภาพ 1 สรุปจำนวนและชนิดของข้อมูลวิเคราะห์

**Network Information**

Input Layer	Factors	1	Level	
		2	VARK	
		3	LS_PFL	
		4	SR_gr	
		5	MA_gr	
Hidden Layer(s)	Number of Units <sup>a</sup>			14
	Number of Hidden Layers			1
	Number of Units in Hidden Layer 1 <sup>a</sup>			2
Output Layer	Activation Function		Hyperbolic tangent	
	Dependent Variables	1	GPA_gr	
	Number of Units			2
	Activation Function		Softmax	
	Error Function		Cross-entropy	

a. Excluding the bias unit

ภาพ 2 ข้อมูลตัวแปร

**Model Summary**

Training	Cross Entropy Error	49.245
	Percent Incorrect Predictions	36.4%
	Stopping Rule Used	1 consecutive step (s) with no decrease in error <sup>a</sup>
	Training Time	0:00:00.16
Testing	Cross Entropy Error	60.458
	Percent Incorrect Predictions	46.0%
Holdout	Percent Incorrect Predictions	50.0%

Dependent Variable: GPA\_gr

a. Error computations are based on the testing sample.

ภาพ 3 สรุปผลการวิเคราะห์โมเดล

**Classification**

Sample	Observed	Predicted		
		not good	good	Percent Correct
Training	not good	33	13	71.7%
	good	15	16	51.6%
	Overall Percent	62.3%	37.7%	63.6%
Testing	not good	25	9	73.5%
	good	31	22	41.5%
	Overall Percent	64.4%	35.6%	54.0%
Holdout	not good	18	7	72.0%
	good	13	2	13.3%
	Overall Percent	77.5%	22.5%	50.0%

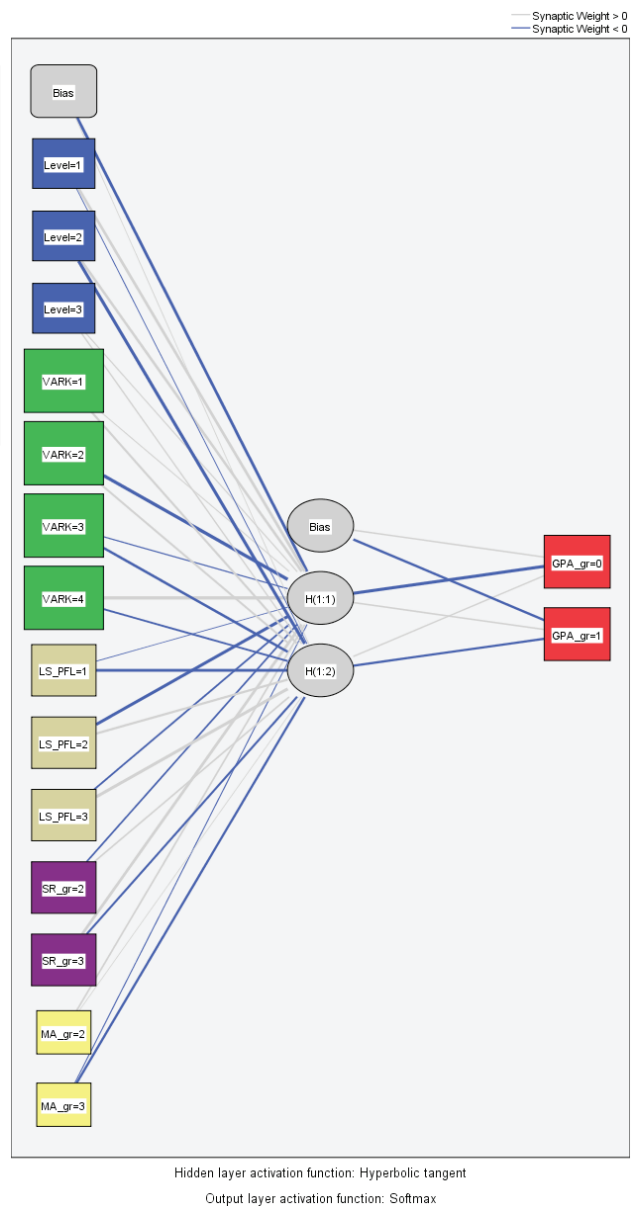
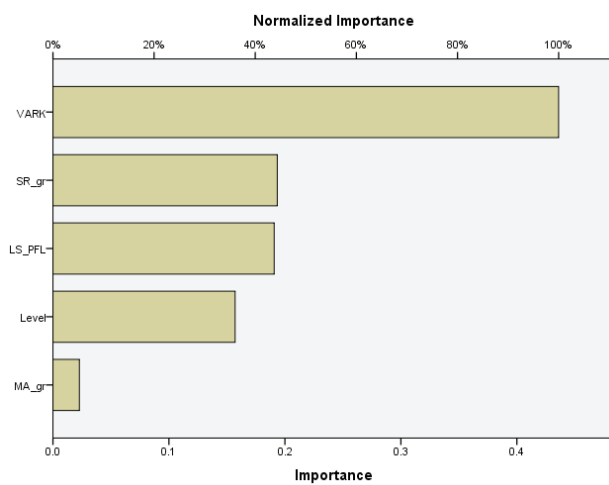
Dependent Variable: GPA\_gr

ภาพ 4 การพยากรณ์ผลสัมฤทธิ์ทางการเรียน



การแปลความหมายของผลการวิเคราะห์ข้อมูล ดังภาพ 2 ดังนี้ ระดับชั้นปี (Level) สไตล์การเรียนรู้แบบ VARK (VARK) ความชอบต่อรูปแบบการจัดการเรียน การสอน (LS\_PFL) แรงจูงใจใฝ่สัมฤทธิ์ในการเรียน (MA\_gr) และการกำกับตนเองในการเรียน (SR\_gr) ใช้การวิเคราะห์ เครือข่ายใยประสาทแบบ MLP แสดงดังภาพ 4 พบว่า มีจำนวนข้อมูลวิเคราะห์ทั้งหมด 207 ข้อมูล แบ่งเป็น ข้อมูลฝึก 77 ข้อมูล ข้อมูลทดสอบ 87 ข้อมูล และข้อมูล ตรวจสอบ 40 ข้อมูล ข้อมูลสามารถระบุตัวแปรสำคัญที่ ใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียน (GPA\_gr) ได้

คิดเป็นร้อยละ 50 พิจารณา จาก Holdout ในภาพ 3 และ 4 โดยตัวแปรที่มีความสำคัญลำดับมากพิจารณาจาก ค่าความสำคัญที่ใช้ในการพยากรณ์ ได้แก่ สไตล์การเรียนรู้ แบบ VARK (VARK) การกำกับตนเองในการเรียน (SR\_gr) ความชอบต่อรูปแบบการจัดการเรียนการสอน (LS\_PFL) ระดับชั้นปี (Level) และแรงจูงใจใฝ่สัมฤทธิ์ในการเรียน (MA\_gr) ตามลำดับ และแสดงโครงข่ายประสาทเทียม แบบหลายชั้นโหนด (node) โดยในแต่ละชั้นมีค่าน้ำหนัก (weight) ที่แตกต่างกันดังภาพ 5



ภาพ 5 แสดงค่าความสำคัญในการพยากรณ์และโครงข่ายประสาทเทียมแบบ MLP

## หลักการวิเคราะห์ต้นไม้ตัดสินใจ

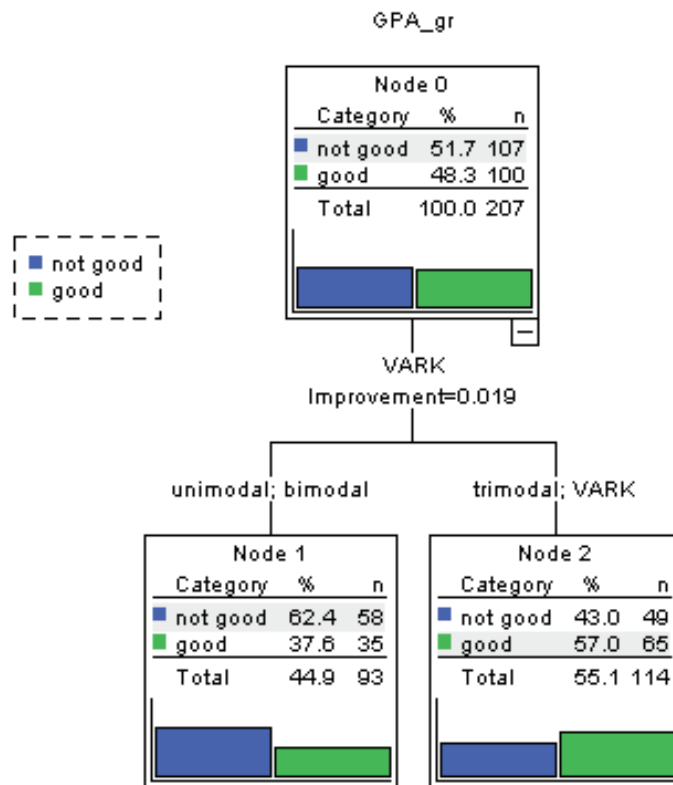
### ตัวอย่างการวิเคราะห์และการแปลผล

การวิเคราะห์ต้นไม้ตัดสินใจเป็นการนำข้อมูลมาสร้างแบบจำลองพยากรณ์ที่มีลักษณะคล้ายโครงสร้างต้นไม้ โดยใช้ข้อมูลตัวแปรต้นทำนายผลของตัวแปรตามแล้วมีการสร้างกฎต่าง ๆ ขึ้นเพื่อใช้ในการตัดสินใจ การวิเคราะห์ต้นไม้ตัดสินใจใช้อัลกอริทึมที่ไม่ซับซ้อนในการวิเคราะห์ทำให้สามารถตีความและเข้าใจลักษณะของรูปแบบข้อมูล (pattern) ได้ง่ายเพราะมีการแยกออกเป็นกฎต่าง ๆ (Na-Wichien, 2017b ; Srisa-An, 2017) ในการวิเคราะห์ต้นไม้ตัดสินใจใช้โปรแกรมสำเร็จรูป SPSS เลือก growing method โดยใช้วิธี CRT (Classification and Regression Trees) ในการจำแนกและทำนายซึ่งการคำนวณวิธีนี้จะคำนวณโดยแบ่งข้อมูลเป็นส่วน ๆ แล้ว

พิจารณาข้อมูลที่มีลักษณะคล้ายคลึงกับตัวแปรตามมากที่สุด เนื่องจากตัวแปรตามมีมาตรวัดเป็นข้อมูลไม่ต่อเนื่องซึ่งสามารถ Impurity Measure ได้ 2 แบบ คือ GINI Twoing และ Ordered Twoing โดย GINI เป็นค่าที่ระบุความแตกต่างระหว่างข้อมูลที่สัมพันธ์กับตัวแปรตาม Twoing เป็นการคำนวณที่แบ่งข้อมูลเป็น 2 กลุ่มแล้วเลือกตัวแปรต้นที่สามารถแบ่งกลุ่มได้สอดคล้องกับตัวแปรตาม ส่วน Ordered Twoing แบ่งข้อมูลเป็น 2 กลุ่มคล้าย Twoing แต่ตัวแปรตามเป็นมาตรวัดแบบเรียงอันดับ (Na-Wichien, 2017b) การประเมินค่าความแม่นยำ (cross validation) โดยแบ่งข้อมูลออกเป็น ส่วน ๆ จำนวนรอบของการตรวจสอบ (number of sample folds) ความถูกต้องในการจำแนกประเภทด้วยชุดข้อมูลใช้ 10 รอบ ตัวอย่างการวิเคราะห์ต้นไม้ตัดสินใจด้วยวิธี CRT และการแปลผลต้นไม้ตัดสินใจ ดังนี้

Model Summary			Risk		
Specifications	Growing Method	CRT	Method	Estimate	Std. Error
	Dependent Variable	GPA_gr	Resubstitution	.406	.034
	Independent Variables	Level, VARK, LS_PFL, MA_gr, SR_gr	Cross-Validation	.444	.035
	Validation	Cross Validation	Growing Method: CRT Dependent Variable: GPA_gr		
	Maximum Tree Depth	5			
	Minimum Cases in Parent Node	100			
	Minimum Cases in Child Node	50			
Results	Independent Variables Included	VARK			
	Number of Nodes	3			
	Number of Terminal Nodes	2			
	Depth	1			

ภาพ 6 สรุปผลการวิเคราะห์โมเดล



ภาพ 7 แผนภูมิการวิเคราะห์ต้นไม้ตัดสินใจ

**Classification**

Observed	Predicted		
	not good	good	Percent Correct
not good	58	49	54.2%
good	35	65	65.0%
Overall Percentage	44.9%	55.1%	59.4%

Growing Method: CRT  
Dependent Variable: GPA\_gr

ภาพ 8 ความถูกต้องในการจัดประเภท

**Tree Table**

Node	not good		good		Total		Predicted Category	Parent Node	Primary Independent Variable		
	N	Percent	N	Percent	N	Percent			Variable	Improvement	Split Values
0	107	51.7%	100	48.3%	207	100.0%	not good				
1	58	62.4%	35	37.6%	93	44.9%	not good	0	VAR K	.019	unimodal; bimodal
2	49	43.0%	65	57.0%	114	55.1%	good	0	VAR K	.019	trimodal; VAR K

Growing Method: CRT  
Dependent Variable: GPA\_gr

ภาพ 9 ข้อมูลตารางต้นไม้ตัดสินใจ

การแปลความหมายของผลการวิเคราะห์ต้นไม้ตัดสินใจด้วยวิธี CRT ดังภาพ 6 และ 7 พบว่ามีจำนวนข้อมูลวิเคราะห์ทั้งหมด 207 ข้อมูล มีตัวแปรต้นคือ ระดับชั้นปี (level) สไตล์การเรียนรู้แบบ VARK ความชอบต่อรูปแบบการจัดการเรียนการสอน (LS\_PFL) แรงจูงใจใฝ่สัมฤทธิ์ในการเรียน (MA\_gr) และการกำกับตนเองในการเรียน (SR\_gr) ผลการวิเคราะห์พบตัวแปรสามารถจัดประเภทบุคคลด้วยการทำนายผลสัมฤทธิ์ทางการเรียน ได้ถูกต้องคิดเป็นร้อยละ 59.4 ดังภาพ 8 และมีค่าความคลาดเคลื่อนจากการทำนายร้อยละ 40.6 ดังภาพ 6 ได้กฎในการจัดประเภทจำนวน 2 กฎดังภาพ 7 และ 9 ดังนี้

1. ถ้าผู้เรียนมีสไตล์การเรียนรู้แบบ VARK ในรูปแบบ trimodal หรือ multimodal/VARK จะมีผลสัมฤทธิ์ทางการเรียนดีคิดเป็นร้อยละ 57
2. ถ้าผู้เรียนมีสไตล์การเรียนรู้แบบ VARK ในรูปแบบ unimodal หรือ bimodal จะมีผลสัมฤทธิ์ทางการเรียนไม่ดีคิดเป็นร้อยละ 62.4

สรุปผลการวิเคราะห์ทั้ง 2 วิธีคือ การวิเคราะห์เครือข่ายประสาทแบบ MLP และการวิเคราะห์ต้นไม้ตัดสินใจด้วยวิธี CRT ให้ผลวิเคราะห์ที่สอดคล้องกัน โดยการวิเคราะห์เครือข่ายประสาทแบบ MLP สามารถระบุตัวแปรสำคัญที่ใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้คิดเป็นร้อยละ 50 และตัวแปรที่มีความสำคัญลำดับมากพิจารณาจากค่าความสำคัญที่ใช้ในการพยากรณ์พบว่าตัวแปรสไตล์การเรียนรู้มีค่าความสำคัญ (normalized important) สูงสุดคือร้อยละ 100 แสดงดังภาพ 8 ผลการวิเคราะห์สอดคล้องกับการวิเคราะห์ต้นไม้ตัดสินใจด้วยวิธี

CRT ซึ่งพบว่าตัวแปรสามารถจัดประเภทบุคคลด้วยการทำนายผลสัมฤทธิ์ทางการเรียนได้ถูกต้องคิดเป็นร้อยละ 59.4 ดังภาพ 8 โดยข้อมูลตารางต้นไม้ดังภาพ 9 แสดงให้เห็นว่าตัวแปรสไตล์การเรียนรู้แบบ VARK เป็นตัวแปรต้นหลักในการทำนายผลสัมฤทธิ์ทางการเรียน และพบกฎในการจัดประเภทมีจำนวน 2 กฎคือ (1) ถ้าผู้เรียนมีสไตล์การเรียนรู้แบบ VAR ในรูปแบบ trimodal หรือ VARK/multimodal จะมีผลสัมฤทธิ์ทางการเรียนดีคิดเป็นร้อยละ 57 และ (2) ถ้าผู้เรียนมีสไตล์การเรียนรู้แบบ VARK ในรูปแบบ unimodal หรือ bimodal จะมีผลสัมฤทธิ์ทางการเรียนไม่ดีคิดเป็นร้อยละ 62.4

## บทสรุป

การทำเหมืองข้อมูลเพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูลนั้น การวิเคราะห์แบ่งเป็นการวิเคราะห์เหมืองข้อมูลเพื่อการทำนายและการวิเคราะห์เหมืองข้อมูลเพื่อการอธิบาย โดยเทคนิคเหมืองข้อมูลที่ได้รับความนิยมได้แก่ การจำแนกประเภท (ต้นไม้ตัดสินใจ และเครือข่ายประสาท) การจัดกลุ่มข้อมูล และการค้นหาความสัมพันธ์ของข้อมูล ในปัจจุบันมีข้อมูลต่าง ๆ และฐานข้อมูลที่เกี่ยวข้องในวิชาชีพการพยาบาลจำนวนมาก พยาบาลสามารถนำข้อมูลต่าง ๆ จากฐานข้อมูลนำไปใช้ในการวิเคราะห์เหมืองข้อมูล ผลจากการวิเคราะห์เหมืองข้อมูลทำให้เกิดองค์ความรู้ สามารถนำไปใช้เพื่อการตัดสินใจอันเป็นประโยชน์ทางคลินิกและประโยชน์ในด้านการบริหารการพยาบาลต่อไป



## References

- Alexander, C. A., & Wang, L. (2017). Big data analytics in heart attack prediction. *Journal of Nursing and Care, 6*(2), 1-9.
- Almasalha, F., Xu, D., Keenan, G. M., Khokhar, A., Yao, Y., Chen, Y. C., . . . Wilkie, D. J. (2013). Data mining nursing care plans of end of life patients: A study to improve healthcare decision making. *International Journal of Nursing Knowledge, 24*(1), 15-24.
- Chirawichitchai, N. (2010). *Data mining for automatical disease analysis*. Bangkok: Suan Sunandha Rajabhat University. (in Thai)
- Ghaibeh, A., Setoguchi, Y., & Moriguchi, H. (2016). Mining nurse care data: A study case on pressure ulcer prediction. *2016 IEEE International Conference on Healthcare Informatics*, Illinois.
- Jirathummakoon, S. (2006). Information search for evidence based in nursing. *Thai Journal of Cardio-Thoracic Nursing, 17*(2), 37-51. (in Thai).
- Lee, T. T., Liu, C. Y., Kuo, Y. H., Mills, M. E., Fong, J. G., & Hung, C. (2011). Application of data mining to the identification of critical factors in patient falls using a web-based reporting system. *International journal of medical informatics, 80*(2), 141-150.
- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications, 39*(12), 11303-11311.
- Madigan, E. A., & Curet, O. L. (2006). *A data mining approach in home healthcare: outcomes and service use*. Retrieved from <https://bmchealthservres.biomedcentral.com/track/pdf/10.1186/1472-6963-6-18>
- Mahidol University. (2015). *Nursing databases*. Retrieved from <https://med.mahidol.ac.th/learningresources/th/content/e-databases/nursing> (in Thai)
- Na-Wichien, S. (2017a). *Neuron networks and decision tree analysis*. Bangkok: Chulalongkorn University. (in Thai)
- Na-Wichien, S. (2017b). *Data mining*. Bangkok: Chulalongkorn University. (in Thai)
- Pasupathy, K. S., & Barker, L. M. (2011). Impact of fatigue on performance in registered nurses: Data mining and implications for practice. *Journal for Healthcare Quality, 34*(5), 22-30.
- Prachuabsupakij, W. (2018). *Data mining with WEKA*. Prachinburi: King Mongkut's University of Technology North Bangkok. (in Thai)
- Raju, D., Su, X., Patrician, P. A., Loan, L. A., & McCarthy, M. S. (2015). Exploring factors associated with pressure ulcers: A data mining approach. *International Journal of Nursing Studies, 52*(1), 102-111.

- Santos, R. S., Malheiros, S. M., Cavalheiro, S., & De Oliveira, J. M. (2013). A data mining system for providing analytical information on brain tumors to public health decision makers. *Computer Methods and Programs in Biomedicine*, 109(3), 269-282.
- Sakduyatham, R. (2010). *Factors for successful analysis of trigger finger treatment of faculty of medicine, Vajira hospital, university of Bangkok using data mining techniques*. Nonthaburi: Rajapruek College. (in Thai)
- Sripaoraya, S., & Sinsomboonthong, S. (2017). Efficiency comparison of data mining, classification methods for chronic kidney disease: A case study of a hospital in India. *Journal of Science and Technology*, 25(5). 839-853. (in Thai)
- Srisa-An, C. (2017). *Intelligent data sources...everyone can do (data warehouse and data mining)*. Retrieved from [https://www.thaimooc.org/courses/course-v1:RSU-MOOC+rsu001+2017\\_T1/info](https://www.thaimooc.org/courses/course-v1:RSU-MOOC+rsu001+2017_T1/info) (in Thai)
- Sumamal, K. (2012). *Basic health screening by using data mining techniques*. Master of Science (Web Engineering) Thesis, Dhurakij Pundit University. (in Thai)
- Thanprasertkul, T., & Keatruangkamala, K. (2014). The cause predicting to blood pressure essential hypertension: A case study of public hospital in Thailand. *34<sup>th</sup> the National Graduate Research Conference*, Khon Kaen. (in Thai)
- Wisvapaisan, W., Srithammasak, B., & Nakkrasae, S. (2016). Factors affecting the prediction of gravid women giving preterm births at police general hospital. *Journal of the Police Nurse*, 8(2), 83-89. (in Thai).
- Wongwiwatanon, P. (2008). Data mining in medical informatics. *Data Management & Biostatistics Journal*, 4(2), 12-21. (in Thai)
- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data Mining in healthcare and biomedicine: A survey of the literature. *Journal of Medical Systems*, 36(4), 2431-2448.
- Yoon, S., Patrao, M., Schauer, D., & Gutierrez, J. (2010). *Prediction models for burden of caregivers applying data mining techniques*. Retrieved from <http://aimsclences.org/article/doi/10.3934/bdia.2017014?viewType=html>

