

เปรียบเทียบประสิทธิภาพอัลกอริทึมดาต้าไมนนิ่งเพื่อทำนาย
โอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน

ปาณิสรา อ่อนละมุล

วิทยานิพนธ์เสนอต่อมหาวิทยาลัยรามคำแหง
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต (ระบบสนับสนุนการตัดสินใจ)
ปีการศึกษา 2561
ลิขสิทธิ์ของมหาวิทยาลัยรามคำแหง

A COMPARISON OF DATA MINING ALGORITHMS FOR PREDICTING
INPATIENT FALLS

PANISSARA ONLAMOON

A THESIS PRESENTED TO RAMKHAMHAENG UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
(DECISION SUPPORT SYSTEM)

2018

COPYRIGHTED BY RAMKHAMHAENG UNIVERSITY

ชื่อเรื่องวิทยานิพนธ์ เปรียบเทียบประสิทธิภาพอัลกอริทึมคาค่าไมนั้ิงเพื่อทำนาย
โอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน

ชื่อผู้เขียน นางปานิสรา อ่อนละมุล

สาขาวิชา ระบบสนับสนุนการตัดสินใจ

คณะกรรมการที่ปรึกษาวิทยานิพนธ์

รองศาสตราจารย์ ดร. ระพีพรรณ พิริยะกุล ประธานกรรมการ

รองศาสตราจารย์ ดร. มนตรี พิริยะกุล

มหาวิทยาลัยรามคำแหงอนุมัติให้วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญามหาบัณฑิต



.....คณะบดีบัณฑิตวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร. สืบพงษ์ ปราบใหญ่)

คณะกรรมการสอบวิทยานิพนธ์



.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. จวีวรรณ เพ็ชรศิริ)



.....กรรมการ

(รองศาสตราจารย์ ดร. ระพีพรรณ พิริยะกุล)



.....กรรมการ

(รองศาสตราจารย์ ดร. มนตรี พิริยะกุล)

บทคัดย่อ

ชื่อเรื่องวิทยานิพนธ์ เปรียบเทียบประสิทธิภาพอัลกอริทึมค้ำไมน์นิ่งเพื่อ
ทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน
ชื่อผู้เขียน นางปานิสรา อ่อนละมุล
ชื่อปริญญา วิทยาศาสตร์มหาบัณฑิต
สาขาวิชา ระบบสนับสนุนการตัดสินใจ
ปีการศึกษา 2561

คณะกรรมการที่ปรึกษาวิทยานิพนธ์

1. รองศาสตราจารย์ ดร. ระพีพรรณ พิริยะกุล ประธานกรรมการ
2. รองศาสตราจารย์ ดร. มนต์รี พิริยะกุล

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน และเพื่อเปรียบเทียบประสิทธิภาพระหว่างอัลกอริทึมเหมืองข้อมูล ได้แก่ วิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธี โครงข่ายประสาทเทียมเชิงลึก ข้อมูลที่ใช้ในการพัฒนาตัวแบบ เป็นข้อมูลของผู้ป่วยในที่ได้รับการรักษาในโรงพยาบาลเอกชนแห่งหนึ่ง ในช่วงเดือน มกราคม 2557 ถึง เดือนธันวาคม 2559 จำนวน 400 คน แบ่งข้อมูลด้วยวิธี Split Test ออกเป็น 2 ส่วน โดยข้อมูลส่วนที่หนึ่งร้อยละ 70 ใช้ในการสร้างตัวแบบคิดเป็นชุดข้อมูล จำนวน 280 คน ใช้วิธี 10-Fold Cross-validation ในการทดสอบประสิทธิภาพของตัวแบบ และข้อมูลส่วนที่สองร้อยละ 30 คิดเป็นชุดข้อมูลจำนวน 120 คน ใช้เป็นข้อมูลทำนาย โอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน ที่ได้รับการรักษาที่โรงพยาบาล ตัวแบบที่มี ประสิทธิภาพสูงสุดวัดจาก ค่าความถูกต้อง ค่าความผิดพลาดของการจำแนกประเภทข้อมูล ค่าพื้นที่ใต้กราฟ ROC ค่าความแม่นยำ ค่าความระลึกลับ ค่าวัดประสิทธิภาพรวมของค่าความ แม่นยำและค่าความระลึกลับ ค่าความไว และค่าความจำเพาะ โดยใช้โปรแกรม RapidMiner Studio เวอร์ชัน 8.1

ผลการศึกษา พบว่า ทั้ง 3 อัลกอริทึมมีประสิทธิภาพที่สูงใกล้เคียงกัน แต่วิธีถดถอย โลจิสติก ให้ค่าความถูกต้องสูงที่สุด คือ ร้อยละ 85.26 ค่าความผิดพลาดของการจำแนก

ประเภทข้อมูลน้อยที่สุด คือ ร้อยละ 14.74 ค่าพื้นที่ใต้กราฟ ROC สูงที่สุด คือ เท่ากับ 0.897 ค่าความแม่นยำสูงที่สุด คือ ร้อยละ 87.01 ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำ และค่าความระลึกสูงที่สุด คือ ร้อยละ 84.12 และมีค่าความจำเพาะสูงที่สุด คือ ร้อยละ 87.80 ในขณะที่วิธีต้นไม้ตัดสินใจนั้นมีค่าความระลึกสูงที่สุด คือ ร้อยละ 82.86 และค่าความไวสูงที่สุด คือ ร้อยละ 82.86 ดังนั้นอัลกอริทึมเหมือนข้อมูลวิธีถดถอยโลจิสติกจึงเป็นตัวแทนทำนายที่มีประสิทธิภาพสูงที่สุดในการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน

ABSTRACT

Thesis Title	A Comparison of Data Mining Algorithms for Predicting Inpatient Falls	
Student's Name	Mrs. Panissara Onlamoon	
Degree Sought	Master of Science	
Major	Decision Support System	
Academic Year	2018	
Advisory Committee		
	1. Asso. Prof. Dr. Rapeepan Piriyaikul	Chairperson
	2. Asso. Prof. Dr. Montree Piriyaikul	

In this thesis, the researcher predicts chances for inpatient falls and compares the efficiency of different data mining algorithms, *viz.*, logistic regression, decision tree, and deep artificial neural networks.

Data used for the development of a model were obtained from 400 inpatients treated at a private hospital from January 2014 to December 2016. Data were divided into two parts using the split test method.

The first part of the data (70 percent) was used in the construction of a model as a data set for 280 persons. The method of 10-Fold Cross-Validation was used to test the efficiency of the model. The second part of the data (30 percent) was the data set for 120 persons. Data were used to predict chances for falls of inpatients treated at a hospital.

The model with the highest efficiency was measured by accuracy, classification error, area falling under the receiver operating characteristic (ROC) curve, precision, recall, the f-measure, sensitivity, and specificity. Data were analyzed using the Rapid Miner Studio version 8.1 software computer program.

Findings are as follows:

All three algorithms exhibited more or less the same high efficiency. The method of logistic regression yielded accuracy at the highest level (85.26 percent). Classification error was at the lowest level (14.74 percent). The area under the ROC curve was at the highest level (0.897). Precision was at the highest level (87.01 percent). The f-measure was at the highest level (84.12 percent). Specificity was at the highest level (87.80 percent). Decision tree exhibited recall at the highest level (82.86 percent). Sensitivity was at the highest level (82.86 percent).

Therefore, the data mining algorithm of logistic regression was a predictive model with the highest efficiency for predicting inpatient falls.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาอย่างสูงและขอขอบพระคุณ
รองศาสตราจารย์ ดร. ระพีพรรณ พิริยะกุล ประธานที่ปรึกษาวิทยานิพนธ์ และ
รองศาสตราจารย์ ดร. มนตรี พิริยะกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่กรุณาให้คำปรึกษา
และแก้ไขข้อบกพร่องต่าง ๆ พร้อมทั้งให้ข้อเสนอแนะที่เป็นประโยชน์จนกระทั่ง
วิทยานิพนธ์สำเร็จลุล่วงไปด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร. นพ. กฤษณ์ พงศ์พิรุณห์ นายแพทย์ที่ปรึกษา
ด้านงานคุณภาพและวิจัยประจำโรงพยาบาลบำรุงราษฎร์ อินเตอร์เนชั่นแนล และ
ดร. เอกสิทธิ์ พัทธวงศ์ศักดิ์ ผู้อำนวยการหลักสูตรวิศวกรรมข้อมูลขนาดใหญ่ วิทยาลัย
นวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ ที่คอย
ช่วยเหลือและให้คำแนะนำพร้อมทั้งให้ข้อเสนอแนะที่เป็นประโยชน์ตั้งแต่เริ่มทำวิจัยจน
เสร็จสิ้น

ขอกราบขอบคุณมารดา และครอบครัวที่คอยสนับสนุนและให้กำลังใจตั้งแต่
เริ่มต้นศึกษาจนกระทั่งทำวิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี

ปาณิสรา อ่อนละมุล

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	(4)
บทคัดย่อภาษาอังกฤษ.....	(6)
กิตติกรรมประกาศ.....	(8)
สารบัญตาราง.....	(9)
สารบัญภาพประกอบ.....	(12)
บทที่	
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	4
ขอบเขตของการวิจัย.....	4
นิยามศัพท์เฉพาะ.....	5
เครื่องมือที่ใช้ในการวิจัย.....	6
ประโยชน์ที่คาดว่าจะได้รับ.....	6
2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	7
การพลัดตกหกล้ม.....	7
การทำเหมืองข้อมูล.....	14
เทคนิคต้นไม้ตัดสินใจ.....	19
เทคนิคการถดถอยโลจิสติก.....	25
เทคนิคโครงข่ายประสาทเทียม.....	28
ตัววัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล.....	37
งานวิจัยที่เกี่ยวข้อง.....	40
3 วิธีดำเนินการวิจัย.....	47
การรวบรวมข้อมูล.....	50
การเตรียมข้อมูล.....	51

บทที่	หน้า
การค้นหาตัวแบบ.....	53
การทดสอบประสิทธิภาพตัวแบบทำนาย.....	54
4 ผลการวิเคราะห์ข้อมูล.....	55
ผลการวิเคราะห์การคัดเลือกปัจจัย.....	55
ผลการวิเคราะห์และเปรียบเทียบประสิทธิภาพตัวแบบ.....	64
5 สรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะ.....	68
สรุปผลการวิจัย.....	68
การอภิปรายผล.....	69
ข้อเสนอแนะ.....	71
ภาคผนวก	
ก คู่มือการใช้งานโปรแกรม RapidMiner Studio 8.1.....	72
ข ผลการวิเคราะห์ข้อมูลจากโปรแกรม RapidMiner Studio 8.1.....	90
บรรณานุกรม.....	100
ประวัติผู้เขียน.....	108

สารบัญตาราง

ตาราง	หน้า
1 ประเภทของโครงข่ายประสาทเทียม.....	30
2 กลุ่มตัวอย่างจำแนกตามเพศ.....	56
3 กลุ่มตัวอย่างจำแนกตามประวัติการพลัดตกหกล้ม.....	57
4 กลุ่มตัวอย่างจำแนกตามการได้รับการคาสายเพื่อให้สารละลาย ทางหลอดเลือดดำ.....	57
5 กลุ่มตัวอย่างจำแนกตามการใช้อุปกรณ์ในการช่วยเดิน.....	58
6 กลุ่มตัวอย่างจำแนกตามภาวะความอ่อนแรง.....	58
7 กลุ่มตัวอย่างจำแนกตามความสามารถในการเคลื่อนไหวของระยางค์ ส่วนล่างลดลง.....	59
8 กลุ่มตัวอย่างจำแนกตามภาวะความผิดปกติทางด้านอารมณ์และจิตใจ.....	59
9 กลุ่มตัวอย่างจำแนกตามอาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะ ในระยะเวลา 24 ชม.....	60
10 กลุ่มตัวอย่างจำแนกตามอาการท้องเสียหรือได้รับยาระบาย ในระยะเวลา 24 ชม.....	60
11 กลุ่มตัวอย่างจำแนกตามอาการเวียนศีรษะบ้านหมุน.....	61
12 กลุ่มตัวอย่างจำแนกตามการดูแลของญาติหรือผู้ดูแลในโรงพยาบาล.....	61
13 กลุ่มตัวอย่างจำแนกตามการได้รับยาบรรเทาอาการปวด.....	62
14 กลุ่มตัวอย่างจำแนกตามการได้รับยานอนหลับ.....	62
15 กลุ่มตัวอย่างจำแนกตามการได้รับยากล่อมประสาท.....	63
16 กลุ่มตัวอย่างจำแนกตามการได้รับยากลายเครียด.....	63
17 กลุ่มตัวอย่างจำแนกตามสถานะการพลัดตกหกล้ม.....	63

สารบัญภาพประกอบ

ภาพ	หน้า
1 ขั้นตอนการทำเหมืองข้อมูล.....	17
2 ตัวอย่างโครงข่ายประสาทเทียม.....	32
3 การเชื่อมต่อระหว่างชั้นนำเข้า ชั้นซ่อนและชั้นผลลัพธ์.....	33
4 แผนภูมิลำดับของการทำงาน.....	50
5 ข้อมูลในรูปแบบไฟล์เอ็กเซล.....	51
6 ข้อมูลแอทริบิวต์ของชุดข้อมูลเรียนรู้ของผู้ป่วยในของโรงพยาบาลที่คัดเลือก	52
7 กลุ่มตัวอย่างจำแนกตามอายุ.....	55
8 กลุ่มตัวอย่างจำแนกตามระยะเวลาการรักษาตัวในโรงพยาบาล.....	56
9 ค่าเปรียบเทียบประสิทธิภาพตัวทำนายวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก.....	65
10 หน้าต่างผล Example set ของชุดข้อมูลทดสอบกับการทดสอบ ตัวแบบวิธีถดถอยโลจิสติก.....	66
11 หน้าต่างผลสถิติการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้รับบริการ รายใหม่ที่น่าเข้าไปทดสอบตัวแบบวิธีถดถอยโลจิสติก.....	66
12 หน้าต่างผลรูปแบบแผนภูมิของการทำนายโอกาสเกิดการพลัดตกหกล้ม ของผู้รับบริการรายใหม่ที่น่าเข้าไปทดสอบตัวแบบวิธีถดถอยโลจิสติก.....	66
13 หน้าต่างตัวแปรทำนายด้วยวิธีถดถอยโลจิสติกที่มีความสัมพันธ์กับ การพลัดตกหกล้ม.....	67
14 หน้าต่าง Welcome ของ RapidMiner Studio 8.1.....	73
15 หน้าต่างเริ่มต้นการทำงานของโปรแกรม.....	73
16 หน้าต่างหน้าจอการทำงานในโปรแกรม.....	74
17 หน้าต่างการสร้าง Repository ใหม่ในโปรแกรม.....	74
18 หน้าต่างการตั้งชื่อใหม่ใน Repository.....	75
19 หน้าต่างการนำข้อมูลเข้าในโปรแกรม.....	75

ภาพ	หน้า
20 หน้าต่างการเลือกไฟล์ที่จะนำเข้าไปในโปรแกรม.....	76
21 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรม.....	76
22 หน้าต่างการเปลี่ยนหน้าที่เป็น คลาสคำตอบ	77
23 หน้าต่างการเปลี่ยนหน้าที่เป็น ID ของ Primary Key.....	77
24 หน้าต่างการจัดเก็บข้อมูลใน Repository.....	78
25 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรมในรูปแบบตารางแสดงผลข้อมูล ที่ไหลคเข้าไปแสดงในรูปแบบตาราง.....	78
26 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรมในรูปแบบกราฟ.....	79
27 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรมในรูปแบบค่าสถิติ.....	79
28 หน้าต่างการเลือกไฟล์ที่ต้องการนำมาสร้างตัวแบบ.....	80
29 หน้าต่างการเลือกวิธีการทำนายและเลือกคลาส Fall Case.....	81
30 หน้าต่างสัดส่วนของข้อมูล Fall และ No Fall.....	82
31 หน้าต่างการคำนวณความสำคัญของแอทริบิวต์ที่จะนำไปสร้างตัวแบบ.....	83
32 หน้าต่างขั้นตอนการเลือกตัวแบบที่จะทำการสร้าง.....	84
33 หน้าต่างรายงานในส่วนการข้อมูลในรูปแบบตาราง.....	84
34 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในการหาตัวแบบวิธี ถดถอยโลจิสติกหน้าจอหลัก.....	85
35 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธีถดถอยโลจิสติก.....	85
36 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในการหาตัวแบบ วิธีต้นไม้ตัดสินใจหน้าจอหลัก.....	86
37 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในส่วน Find Optimal Depth ในการทดลองหาตัวแบบวิธีต้นไม้ตัดสินใจ.....	86
38 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธีต้นไม้ตัดสินใจ.....	87

ภาพ	หน้า
39 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในการหาตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึกหน้าจอหลัก.....	87
40 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึก.....	88
41 หน้าต่าง Example Set ของชุดข้อมูลทดสอบที่ยังไม่ทราบคลาสที่นำเข้ามา..	88
42 หน้าต่างการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบ.....	89
43 หน้าต่างการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบ.....	89
44 หน้าต่างรายงานผลชุดข้อมูลเรียนรู้ที่นำเข้าไปวิเคราะห์ในส่วน ของ Data ที่ได้รับการเปลี่ยนสำหรับเพื่อสร้างตัวแบบ.....	91
45 หน้าต่างรายงานผลค่าน้ำหนักของตัวแปรนำเข้า.....	92
46 หน้าต่างตัวแบบของวิธีถดถอยโลจิสติก.....	92
47 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบ วิธีถดถอยโลจิสติก.....	93
48 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบวิธีถดถอยโลจิสติก.....	93
49 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบวิธีถดถอยโลจิสติก.....	93
50 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบ วิธีถดถอยโลจิสติก.....	93
51 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบ วิธีถดถอยโลจิสติก.....	94
52 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบ วิธีถดถอยโลจิสติก.....	94
53 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบ วิธีถดถอยโลจิสติก.....	94
54 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบ วิธีถดถอยโลจิสติก.....	94

ภาพ	หน้า
55 หน้าต่างตัวแบบของวิธีต้นไม้ตัดสินใจในรูปแบบต้นไม้.....	95
56 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	95
57 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	95
58 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	96
59 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	96
60 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	96
61 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	96
62 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	97
63 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบ วิธีต้นไม้ตัดสินใจ.....	97
64 หน้าต่างรายงานรายละเอียดของตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึก....	97
65 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	97
66 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	98
67 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	98
68 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	98

ภาพ	หน้า
69 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	99
70 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	99
71 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	99
72 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบ วิธีโครงข่ายประสาทเทียมเชิงลึก.....	99

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การพลัดตกหกล้มเป็นปัญหาทางสาธารณสุขที่สำคัญ ซึ่งเป็นสาเหตุการเสียชีวิตจากทั่วโลกเป็นอันดับสองในกลุ่มของการบาดเจ็บโดยไม่ตั้งใจ รองจากการบาดเจ็บจากอุบัติเหตุบนท้องถนน ซึ่งมีแนวโน้มเพิ่มสูงขึ้นอย่างต่อเนื่อง องค์การอนามัยโลกรายงานในแต่ละปีจากทั่วโลกว่า มีประมาณ 646,000 คน เสียชีวิตจากการพลัดตกหกล้ม ซึ่งมากกว่าร้อยละ 80 อยู่ในประเทศที่มีรายได้ต่ำและปานกลาง ผู้ใหญ่ที่มีอายุมากกว่า 65 ปี ต้องทนทุกข์ทรมานภายหลังจากเกิดการพลัดตกหกล้มเป็นจำนวนมาก ในแต่ละปีมีเหตุการณ์พลัดตกหกล้มที่ได้รับบาดเจ็บรุนแรงในระดับที่จะต้องเข้าพบแพทย์ จำนวน 37.3 ล้านครั้ง กลยุทธ์ในการป้องกันการพลัดตกหกล้มควรเน้นที่การศึกษา การฝึกอบรม การสร้างสภาพแวดล้อมที่ปลอดภัย และการจัดลำดับความสำคัญของงานวิจัยที่เกี่ยวข้องกับการกำหนดนโยบายที่มีประสิทธิภาพ ในการลดความเสี่ยงของการพลัดตกหกล้ม (World Health Organization--WHO, 2018) โรงพยาบาล องค์การวิจัยสุขภาพ และคุณภาพแห่งประเทศสหรัฐอเมริกา (Agency for Healthcare Research and Quality--AHRQ, 2013) รายงานว่าแต่ละปีมีชาวอเมริกันเกิดการพลัดตกหกล้มในโรงพยาบาลประมาณ 700,000-1,000,000 คน และส่งผลให้ผู้ป่วยเกิดการบาดเจ็บตั้งแต่ระดับเล็กน้อยจนถึงระดับรุนแรง ทำให้ผู้ป่วยต้องนอนโรงพยาบาลนานขึ้นเฉลี่ย 6.3 วัน ส่งผลกระทบต่อค่าใช้จ่ายในการดูแลรักษาพยาบาล ผู้ป่วยที่มีการบาดเจ็บจากการพลัดตกหกล้มต้องเสียค่าใช้จ่ายในการรักษาพยาบาลต่อรายเฉลี่ย 14,000 เหรียญสหรัฐฯ (The Joint Commission International--JCI, 2015)

สำหรับประเทศไทยมีผู้สูงอายุเสียชีวิตมากกว่า 2,000 คน หรือเฉลี่ยวันละ 5 คน ความเสี่ยงต่อการพลัดตกหกล้มเพิ่มสูงขึ้นตามอายุและมีแนวโน้มเพิ่มสูงขึ้นอย่างต่อเนื่อง เพศชายมีอัตราการเสียชีวิตจากการพลัดตกหกล้มสูงกว่าเพศหญิง 3 เท่า เพศหญิงมีความชุก

ของการปลัดตกหกล้มสูงกว่าเพศชาย 1.5 เท่า ส่วนใหญ่หกล้มนอกบ้าน การบาดเจ็บหลัง การปลัดตกหกล้มมีตั้งแต่อาการฟกช้ำ แผลถลอก กระดูกหัก และรุนแรงถึงเสียชีวิต ผลการ คาดการณ์ปี พ.ศ. 2560 ถึงปีพ.ศ. 2564 จะมีประชากรผู้สูงอายุไทย อายุ 60 ปี ขึ้นไปเกิด ปลัดตกหกล้มประมาณ 3,030,900 ถึง 4,714,800 คน และจะเสียชีวิตประมาณ 5,700 ถึง 8,900 คน (นิพา ศรีช้าง และ ลวิตรา กำวิ, 2559) จากรายงานอุบัติการณ์การปลัดตกหกล้ม ศูนย์การแพทย์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี มีอัตราการเกิดอุบัติการณ์ ปลัดตกหกล้มในปี พ.ศ. 2551 เท่ากับ 1.05 ครั้ง ต่อ 1,000 วันนอน และปี พ.ศ. 2552 เท่ากับ 1.26 ครั้ง ต่อ 1,000 วันนอน (ชารทิพย์ คนตรง, นงนุช แยมวงษ์, นฤมล คงสาคร และ สายรุ้ง ดีนก, 2553) จะเห็นได้ว่า อัตราเกิดปลัดตกหกล้มของผู้ป่วยในมีแนวโน้มเพิ่มสูงขึ้น แต่ทั้งนี้ขึ้นอยู่กับสภาพแวดล้อมด้วย คือ ข้อมูลในแต่ละแห่งมีความแตกต่างกัน ขึ้นอยู่กับ สถานที่ ลักษณะของผู้ป่วย ภาวะการเจ็บป่วย การรักษาพยาบาล และวัฒนธรรมการรายงาน อุบัติการณ์ของแต่ละองค์กร

การปลัดตกหกล้มใน โรงพยาบาล เป็นตัวชี้วัดคุณภาพเชิงผลลัพธ์ที่สำคัญ ในระบบ บริการสุขภาพ ปัจจัยเสี่ยงที่มีความสัมพันธ์กับการปลัดตกหกล้ม จากการศึกษาของ รัตนพร ทามิ พวงทิพย์ ชัยพิบาลสฤณี และปานตา อภิรักษ์ภานนท์ (2561) พบว่าปัจจัย ส่วนบุคคล (อายุ เพศ ประวัติปลัดตกหกล้ม การใช้ถ้ำมในการสื่อสาร และระยะห่างของ ประเทศที่อาศัยตามเส้นแบ่งเวลา) ปัจจัยภายใน (ภาวะจิตใจ/อารมณ์ อาการสับสน/มึนงง อาการวิงเวียนบ้านหมุน ความถี่ในการขยับถ่ายอุจจาระ/ปัสสาวะ ได้รับยา กลุ่ม Hypnotic /Benzodiazepine Anticonvulsants และความบกพร่องในการเคลื่อนไหวทรงตัว) และ ปัจจัยภายนอก (การมีญาติหรือผู้ดูแล ระยะวันนอน โรงพยาบาล และระดับความเสี่ยงต่อ การปลัดตกหกล้ม) มีความสัมพันธ์กับการเกิดปลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ ระดับ .05

การปลัดตกหกล้มมีแนวโน้มจะเกิดขึ้นได้ภายหลังจากที่เคยหกล้มมาแล้ว ซึ่ง พยาบาลสามารถป้องกันการเกิดครั้งต่อไปได้ โดยให้การพยาบาลเพื่อลดปัจจัยเสี่ยง รวมทั้งได้มีการพัฒนาแนวทางการป้องกันการปลัดตกหกล้มตามปัจจัยเสี่ยงหลายด้าน ที่ เฉพาะเจาะจงในผู้ป่วยแต่ละราย รวมทั้งมีการจัดการด้านสิ่งแวดล้อมให้มีความปลอดภัย แต่ยังคงพบว่า มีอุบัติการณ์ของการปลัดตกหกล้มในโรงพยาบาลอยู่อย่างต่อเนื่องและ

บางรายเกิดการบาดเจ็บรุนแรง ส่งผลให้ต้องนอนโรงพยาบาลนานขึ้น และกระทบต่อค่าใช้จ่ายในการรักษาพยาบาลที่มากขึ้น ซึ่งเป็นความเสี่ยงต่อการเกิดการฟ้องร้องและเสียภาพลักษณ์ขององค์กร เนื่องด้วยผู้รับบริการมีความหลากหลายของปัญหาความเจ็บป่วย รูปแบบการใช้ชีวิตประจำวัน และความไม่คุ้นเคยต่อสถานที่ ทำให้ความเสี่ยงต่อการเกิดการพลัดตกหกล้มในโรงพยาบาลแตกต่างกันไป การเพิ่มปริมาณของจำนวนผู้ป่วยที่มากขึ้นทุก ๆ ปี ทำให้มีนักวิจัยให้ความสำคัญกับการศึกษาวิจัยทาง ด้านการเรียนรู้ของเครื่องจักร (machine learning) เพิ่มมากขึ้น ซึ่งเป็นสิ่งที่จำเป็นในการบริหารจัดการดูแลผู้ป่วยเพื่อเป็นการเอื้อประโยชน์กับบุคลากรทางการแพทย์ ปัจจุบันเครื่องมือดังกล่าวที่เข้ามามีบทบาทเป็นอย่างมากในการวิเคราะห์ข้อมูลของผู้ป่วย คือ เทคนิคการเรียนรู้ของเครื่องที่มีหลากหลายวิธี การวิเคราะห์โรคหรือจำแนกข้อมูลทางการแพทย์ส่วนใหญ่ใช้การเรียนรู้ด้วยเครื่องจักรที่ได้รับความนิยม คือ วิธีโครงข่ายประสาทเทียม (Artificial Neural Network--ANN) วิธีต้นไม้ตัดสินใจ (decision tree) และวิธีถดถอยโลจิสติก (logistic regression) เนื่องจากมีความสามารถในการจำแนกข้อมูลอยู่ในระดับที่ดีและประหยัดทรัพยากร มีหลายงานวิจัยที่ศึกษาและได้ผลการทำนายของเทคนิคจำแนกประเภทข้อมูล (classification) ที่อยู่ในระดับสูง จากงานวิจัยของ Kangwanariyakul, Nantasenamat, Tantimongcolwat, and Naenna (2010) ใช้เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network--ANN) และ Support Vector Machine (SVM) โปรแกรม Matlab และ WEKA ในการพยากรณ์โรคกล้ามเนื้อหัวใจขาดเลือด จากงานวิจัยของ สุรวัชร ศรีเปารยะ และสายชล สิ้นสมบุญทอง (2560) เรื่องการเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง: กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย ใช้เทคนิค Decision Tree, NN, SVM, Rule-based, Logistic Regression, Naïve Bayes จากฐานข้อมูล UCI Machine Learning Repository ใช้โปรแกรม SPSS, WEKA วิเคราะห์

ดังนั้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาเทคนิคเหมืองข้อมูล วิธีถดถอยโลจิสติก (logistic regression) วิธีต้นไม้ตัดสินใจ (decision tree) และวิธีโครงข่ายประสาทเทียมเชิงลึก (deep learning) นำมาเปรียบเทียบประสิทธิภาพตัวแบบเพื่อทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน (inpatient) เพื่อหาตัวแบบทำนายที่มีประสิทธิภาพสูงที่สุด โดยใช้โปรแกรม RapidMiner Studio เวอร์ชัน 8.1 ในการวิเคราะห์โดยศึกษาข้อมูล

ย้อนหลังจากแฟ้มประวัติของผู้ป่วยและรายงานอุบัติการณ์การพลัดตกหกล้ม เนื่องจากโรงพยาบาลมีระบบการบันทึกข้อมูลเกี่ยวกับการประเมินความเสี่ยงและรายงานทางการแพทย์เพื่อป้องกันการพลัดตกหกล้มในแฟ้มประวัติของผู้ป่วย ที่เข้ารับการรักษาพยาบาลทุกรายและมีระบบรายงานอุบัติการณ์ที่สามารถสืบค้นได้ เพื่อนำข้อมูลที่ได้มาเป็นหลักฐานเชิงประจักษ์ในการพัฒนาแนวทางการป้องกันการพลัดตกหกล้มให้มีประสิทธิภาพต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน (inpatient) ของโรงพยาบาล
2. เพื่อเปรียบเทียบประสิทธิภาพอัลกอริทึมข้อมูลระหว่างวิธีถดถอยโลจิสติก (logistic regression) วิธีต้นไม้ตัดสินใจ (decision tree) และวิธีโครงข่ายประสาทเทียมเชิงลึก (deep learning)

ขอบเขตของการวิจัย

การวิจัยครั้งนี้ เป็นการศึกษาข้อมูลย้อนหลัง (retrospective) โดยวิเคราะห์ข้อมูลจากแฟ้มประวัติผู้ป่วย รายงานอุบัติการณ์ และข้อมูลการทบทวนเหตุการณ์การพลัดตกหกล้มของผู้ป่วยใน (inpatient) ระหว่างที่เข้ารับการรักษาในโรงพยาบาลเอกชน ในช่วงเดือนมกราคม 2557 ถึงเดือนธันวาคม 2559 มาทำการวิเคราะห์จำนวน 400 ราย

นิยามศัพท์เฉพาะ

ดาต้าไมนิง (data mining): คือ เหมือนข้อมูล ซึ่งเป็นการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases--KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อมูลจำนวนมากมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีทางวิชา สถิติ (statistics) การเรียนรู้ของเครื่อง (machine learning) และการรู้จำแบบ (pattern recognition)

โครงข่ายประสาทเทียมเชิงลึก (deep learning): คือ ตัวแบบหนึ่งของโครงข่ายประสาทเทียมที่ใช้วิธีการเรียนรู้แบบหลายชั้น (multilayer perceptron) ได้รับการเรียนรู้โดยการไล่ระดับการสุ่มโดยใช้อัลกอริทึมกระบวนการย้อนกลับของการรู้จำ (back-propagation) เครือข่ายประกอบด้วยชั้นซ่อน (hidden layer) จำนวนมากซึ่งประกอบด้วยเซลล์ประสาท (neuron) ที่ใช้ฟังก์ชันกระตุ้น (activation function) การทำงานแบบ Hyperbolic Tangent, ตัวเรียงกระแสไฟฟ้า (rectifier) และ Maxout

ต้นไม้ตัดสินใจ (decision tree) คือ กระบวนการจำแนกข้อมูลด้วยการสร้างต้นไม้ขึ้นโดยใช้กระบวนการเวียนเกิด (recursive) ในการแตกกิ่งก้านของต้นไม้ไปเรื่อย ๆ เพื่อใช้ในการตัดสินใจ ที่ประกอบไปด้วยโหนดต่าง ๆ ที่ไม่ใช่โหนดใบ ถูกใช้ในการแสดงแอทริบิวต์หนึ่ง ๆ ของข้อมูล แต่ละกิ่งก้านของโหนดหนึ่ง ๆ จะหมายถึงค่าที่เป็นไปได้จากการทดสอบกับแอทริบิวต์นั้น ๆ และจะประกอบไปด้วยโหนดใบที่จะมีหมวดหมู่ข้อมูลจัดเก็บอยู่ มีตัวชี้วัดการแบ่งข้อมูล ได้แก่ ค่าเกนความรู้ (information gain) ค่าอัตราส่วนเกน (gain ratio) และดัชนีจีนิ (gini index)

การพลัดตกหกล้ม (falls) หมายถึง การเปลี่ยนตำแหน่งของร่างกายจากการล้ม ถลา หรือตกลงไปสู่พื้น หรือพื้นผิวอื่นที่ต่ำกว่าโดยไม่ได้ตั้งใจ และไม่สามารถควบคุมได้ อาจส่งผลให้ร่างกายได้รับบาดเจ็บ หรือไม่ได้รับบาดเจ็บก็ได้

ผู้ป่วยใน (inpatient) หมายถึง ผู้ป่วยที่มารับการตรวจรักษาและต้องนอนพักรักษาตัวในโรงพยาบาลตั้งแต่ 6 ชั่วโมงขึ้นไป ซึ่งจะพักอยู่ตามห้องผู้ป่วย (วอร์ด) ต่าง ๆ

เครื่องมือที่ใช้ในการวิจัย

1. ด้านฮาร์ดแวร์ (hardware)
 - 1.1 เครื่องคอมพิวเตอร์ CPU Intel Core i3
 - 1.2 หน่วยความจำหลัก Random Access Memory: RAM 4 GB
 - 1.3 หน่วยความจำสำรอง (hard disk) 500 GB
2. ด้านซอฟต์แวร์ (software)
 - 2.1 ระบบปฏิบัติการ Windows 8
 - 2.2 โปรแกรม Microsoft Excel 2010
 - 2.3 โปรแกรมสำเร็จรูป RapidMiner Studio 8.1
 - 2.4 โปรแกรม Stata 13.1

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้รูปแบบตัวแบบทำนายที่เหมาะสมและมีประสิทธิภาพสูงสุดในการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยในที่เข้ารับการรักษาในโรงพยาบาล
2. สามารถประยุกต์ใช้ตัวแบบในการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยในที่เข้ารับการรักษาในโรงพยาบาลได้

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องมาใช้ประกอบการศึกษาเปรียบเทียบตัวแบบทำนายนการพลัดตกหกล้มของผู้ป่วยที่เข้ารับบริการรักษาเป็นผู้ป่วยในของโรงพยาบาล โดยประยุกต์ใช้เทคนิคดาต้าไมน์นิง ตามหัวข้อดังนี้

1. การพลัดตกหกล้ม (falls)
2. การทำเหมืองข้อมูล
3. เทคนิคต้นไม้ตัดสินใจ
4. เทคนิคการถดถอยโลจิสติก
5. เทคนิคโครงข่ายประสาทเทียม
6. ตัววัดประสิทธิภาพของตัวแบบการจำแนกประเภทข้อมูล
7. งานวิจัยที่เกี่ยวข้อง

การพลัดตกหกล้ม (falls)

ความหมายของการพลัดตกหกล้ม

World Health Organization (2018) ให้ความหมายของการพลัดตกหกล้มว่า เป็นเหตุการณ์ที่บุคคลหลุดตัวลงไปที่พื้นหรือในระดับที่ต่ำกว่าโดยไม่ได้ตั้งใจ อาจส่งผลให้ได้รับบาดเจ็บรุนแรงถึงเสียชีวิตได้

Canadian Institute for Health Information (2014) ให้ความหมาย การพลัดตกหกล้มว่าเป็นการเปลี่ยนท่าของร่างกายลงไปที่พื้นหรือระดับที่ต่ำกว่าโดยไม่ได้ตั้งใจ รวมทั้งที่มีผู้พบเห็นหรือ ไม่มีก็ตาม แต่พบว่าผู้ป่วยนอนอยู่บนพื้น ส่งผลให้ร่างกายผู้ป่วยได้รับบาดเจ็บหรือไม่ได้รับบาดเจ็บก็ได้

โดยสรุปในการวิจัยครั้งนี้ ผู้วิจัยให้ความหมายเชิงปฏิบัติการของการพลัดตกหกล้ม

หมายถึง การเปลี่ยนแปลงตำแหน่งของร่างกายจากการลื่น ถลาหรือตกลงไปสู่พื้น หรือพื้นผิวอื่นๆ ที่ต่ำกว่า โดยไม่ได้ตั้งใจและไม่สามารถควบคุมได้ อาจส่งผลให้ร่างกายได้รับบาดเจ็บหรือไม่ได้รับบาดเจ็บก็ได้

ประเภทของการลัดตกหกล้ม

จากการศึกษาย้อนหลังของ Morse (2009) ที่ศึกษาเปรียบเทียบลักษณะของผู้ป่วยที่มีการลัดตกหกล้มและไม่มีการลัดตกหกล้มในช่วงเวลาเดียวกัน โดยวิธีการสุ่มตัวอย่างจำนวน 100 คนพบว่า การลัดตกหกล้มของผู้ป่วยในโรงพยาบาลเกิดขึ้นใน 3 ลักษณะ คือ

1. การลัดตกหกล้มที่เกิดขึ้นจากอุบัติเหตุ (accidental falls) การลัดตกหกล้มในลักษณะนี้คิดเป็นร้อยละ 14 เป็นการลัดตกหกล้มจากการลื่นไถล สะดุด หรือเหตุอื่นที่ทำให้เกิดอุบัติเหตุ ซึ่งสาเหตุส่วนใหญ่ที่พบเกิดจากปัจจัยด้านสิ่งแวดล้อม เช่น หยดน้ำของเหลวหรือปัสสาวะที่หกอยู่บนพื้น และการเคลื่อนย้ายผู้ป่วยที่ไม่ถูกต้อง การลัดตกหกล้มที่เกิดจากอุบัติเหตุนี้ไม่สามารถใช้การทำนายจากแบบประเมินได้ เนื่องจากได้เกิดจากปัจจัยทางด้านร่างกาย ดังนั้น กลยุทธ์ในการป้องกันการลัดตกหกล้มในลักษณะนี้สามารถทำได้โดยการออกแบบให้สิ่งแวดล้อมมีความปลอดภัยหรือแจ้งให้ผู้ป่วยทราบเกี่ยวกับสภาพแวดล้อม การใช้อุปกรณ์ต่าง ๆ รวมทั้งอุปกรณ์ช่วยในการเดินและการแนะนำการเคลื่อนย้ายผู้ป่วยอย่างถูกต้อง

2. การลัดตกหกล้มจากสภาพร่างกายที่คาดการณ์ได้ (anticipated physiological falls) เป็นการลัดตกหกล้มที่เกิดขึ้นเมื่อผู้ป่วยถูกระบุว่า มีแนวโน้มจากการให้คะแนนว่า “มีความเสี่ยงสูงต่อการลัดตกหกล้ม” จากการใช้แบบประเมินการลัดตกหกล้มจากสภาพร่างกายที่คาดการณ์ได้ เกิดขึ้นร้อยละ 78 ของการลัดตกหกล้มทั้งหมดซึ่งสามารถทำนายและวางแผนดำเนินการป้องกันได้

3. การลัดตกหกล้มจากสภาพร่างกายที่คาดการณ์ไม่ได้ (unanticipated physiological falls) เป็นการลัดตกหกล้มที่เกิดจากสาเหตุทางสรีรวิทยา แต่เป็นสถานการณ์ที่ไม่สามารถทำนายได้ก่อนเกิดเหตุการณ์ครั้งแรก ซึ่งเกิดขึ้นประมาณร้อยละ 8 ของการลัดตกหกล้มทั้งหมด ยกตัวอย่างเช่น จากโรคที่เกิดขึ้นกระทันหัน หน้ามืด หรือเกิดการเกิดกระดูกสะโพกหักจากพยาธิสภาพ (pathological fracture of the hip) เมื่อเกิด

พลัดตกหกล้มในลักษณะนี้ผู้ป่วยมีแนวโน้มจะเกิดขึ้นอีก ซึ่งพยาบาลสามารถป้องกันการเกิดครั้งต่อไปได้ เช่น พยาบาลอาจจะแนะนำผู้ป่วยให้ป้องกันอาการหน้ามืดที่เกิดจากการลุกขึ้นอย่างรวดเร็ว โดยเตือนให้ผู้ป่วยลุกขึ้นช้า ๆ ก็จะสามารลดความเสี่ยงต่อการพลัดตกหกล้มได้

จากการศึกษาของ Quigley and Goff (อ้างอิงใน Quigley & Miceli, 2011) พบว่าการพลัดตกหกล้มในโรงพยาบาล แบ่งเป็น 4 ประเภท คือ

1. การพลัดตกหกล้มที่เกิดจากอุบัติเหตุ (accidental falls) เกิดจากปัจจัยจากสิ่งแวดล้อม เช่น การรั่วไหลของน้ำในท่อน้ำ ทำให้เป็นสาเหตุให้ผู้สูงอายุลื่นหรือไถลได้
2. การพลัดตกหกล้มที่เกิดจากปัจจัยภายใน (anticipated physiological falls) เช่น ภาวะความดันโลหิตต่ำจากการเปลี่ยนท่า ภาวะสมองเสื่อม และความบกพร่องของการทรงตัวหรือปัจจัยภายนอก เช่น จากการใช้ยาบางชนิดหรือการช่วยเหลือผู้สูงอายุที่ไม่เหมาะสม
3. การพลัดตกหกล้มสาเหตุเกิดจากเหตุการณ์ที่ไม่คาดคิดหรือไม่ทราบว่าอาการจะเกิดขึ้น (unanticipated physiological falls) เช่น อาการชัก หน้ามืด หัวใจขาดเลือด เป็นต้น ซึ่งการพลัดตกหกล้มประเภทนี้ไม่สามารถที่จะป้องกันได้
4. การพลัดตกหกล้มโดยเจตนา (intentional falls) เกิดขึ้นเมื่อผู้ป่วยมีเจตนาที่จะล้มลงไปที่พื้น

ปัจจัยที่เกี่ยวข้องกับการพลัดตกหกล้ม

จากการทบทวนวรรณกรรม มีงานวิจัยจำนวนมากที่ได้ศึกษาเกี่ยวกับปัจจัยที่เกี่ยวข้องกับการพลัดตกหกล้มของผู้ป่วยขณะอยู่ในโรงพยาบาลพบว่า ปัจจัยที่มีอิทธิพลต่อการพลัดตกหกล้มส่วนมากมาจากปัจจัยภายในมากกว่าปัจจัยภายนอก Perell et al. (2001) ศึกษารายละเอียดปัจจัยภายในและภายนอกมีดังนี้

1. ปัจจัยภายใน (intrinsic factors) เป็นลักษณะหรือเป็นปัจจัยที่เกิดจากสภาพร่างกายหรือความเจ็บป่วยของผู้ป่วย ได้แก่

1.1 อายุ (age) มีรายงานการวิจัยที่ผ่านมาหลายรายงานยืนยันว่า อายุเป็นปัจจัยที่มีผลต่อการพลัดตกหกล้มของผู้ป่วยอย่างมีนัยสำคัญ ผู้ป่วยที่มีอายุมากกว่าจะมีภาวะ

เสี่ยงต่อการพลัดตกหกล้มสูงกว่าผู้ป่วยที่มีอายุน้อย เนื่องจากการเปลี่ยนแปลงสรีระโดยผู้ป่วยที่มีอายุ 60-65 ปี หรือสูงกว่าจะมีภาวะเสี่ยงต่อการพลัดตกหกล้มสูง และผู้ป่วยที่มีอายุ 80 ปี หรือสูงกว่าจะมีภาวะเสี่ยงต่อการพลัดตกหกล้มสูงที่สุด (Evan, Hodgkinson, Lambert, Wood, & Kowanko, 1998; WHO, 2018)

1.2 เพศ (sex) จากการศึกษาพบว่า เพศเป็นปัจจัยที่มีความสัมพันธ์ต่อการพลัดตกหกล้ม โดยเฉพาะในผู้สูงอายุ เนื่องจากความจำกัดทางสรีระของเพศหญิง เช่น จากความแข็งแรงของกล้ามเนื้อลดลง ความหนาแน่นของมวลกระดูกต่ำ การมีน้ำหนักตัวมาก และการเปลี่ยนแปลงในการเคลื่อนไหวของสะโพกในการรับน้ำหนักของขาขณะเดินในเพศหญิงและเพศชายแตกต่างกัน โดยผู้สูงอายุหญิงจะมีท่าเดินที่มีลักษณะคล้ายเป็ด (waddling gait) กล่าวคือ ฐานเดินแคบ ส่วนผู้สูงอายุชายจะมีท่าเดินที่มีฐานกว้างกว่าและระยะก้าวสั้น จึงเป็นอีกเหตุผลหนึ่งที่ทำให้ผู้สูงอายุหญิงหกล้มได้มากกว่าผู้สูงอายุชาย (ลัดดา เทียมวงศ์ และเรวดี เพชรศิริสัมพันธ์, 2552)

1.3 สภาวะทางจิตใจ (mental status) มีรายงานการวิจัยหลายรายงานยืนยันว่าการเปลี่ยนแปลงของสภาพจิตใจของผู้ป่วยมักมีความสัมพันธ์ต่อการพลัดตกหกล้ม โดยสภาวะทางจิตใจที่เกี่ยวข้องกับการพลัดตกหกล้ม ได้แก่ อาการสับสน การเปลี่ยนแปลงการรับรู้ หรือมีอาการเห็นภาพหลอน (Conley, Schultz, & Selvin, 1999; Evan et al., 1998; Hendrich, 2007; Oliver, Daly, Martin, & McMurdo, 2004)

1.4 ประวัติพลัดตกหกล้ม (history of falls) ผู้ป่วยที่เคยมีประวัติพลัดตกหกล้มมากกว่า 1 ครั้งจะเพิ่มความเสี่ยงในการพลัดตกหกล้มมากยิ่งขึ้น เนื่องจากผู้ป่วยจะกลัวการหกล้มซ้ำ ขาดความมั่นใจในการเคลื่อนไหว ทำให้เสี่ยงต่อการพลัดตกหกล้มเป็นอย่างมาก มีรายงานการวิจัยหลายรายงานยืนยันว่า ผู้ป่วยที่เคยมีประวัติพลัดตกหกล้มมีความสัมพันธ์กับการพลัดตกหกล้มอย่างมีนัยสำคัญ (Conley et al., 1999; Oliver et al., 2004)

1.5 การได้รับยา (medication) เป็นปัจจัยหนึ่งที่มีผลเกี่ยวข้องกับการพลัดตกหกล้ม ซึ่งมีสาเหตุมาจากฤทธิ์และอาการข้างเคียงของยาที่ผู้ป่วยได้รับ เช่น ทำให้เกิดอาการมึนงง สับสน ควบคุมการขับถ่ายปัสสาวะไม่ได้ และความดันโลหิตต่ำลง ส่งผลให้เกิดอาการหน้ามืดเป็นลมหรือล้มลงได้ (Miceli, 2008) มีหลายรายงานที่พบว่า การได้รับยาบางชนิด

หรือได้รับยาหลายชนิดมีผลทำให้เกิดภาวะเสี่ยงสูงต่อการพลัดตกหกล้ม ประเภทของยา มีความสัมพันธ์กับการพลัดตกหกล้มในหอผู้ป่วย มีงานวิจัยหลายเรื่อง que ศึกษาพบว่า ยาที่มีความสัมพันธ์กับการพลัดตกหกล้ม ได้แก่ ยาที่ออกฤทธิ์ต่อจิตประสาท (antidepressant) เช่น ยานอนหลับ (hypnotics) ยาแก้ปวด (PCA/opiates) ยาระงับประสาท ยากล่อมประสาท โดยเฉพาะอย่างยิ่ง ยาในกลุ่มของ Benzodiazepine และ Sedative (Hendrich et al., 2003; Hitcho et al., 2004; Krauss et al., 2005; Oliver et al., 1997; Passaro et al., 2000) ยาควบคุมการเต้นของหัวใจ (antiarrhythmic) เช่น Digoxin, Beta-blocker เนื่องจากยาในกลุ่มนี้ จะทำให้หัวใจเต้นช้า เกิดความดันโลหิตต่ำขณะเปลี่ยนท่า ทำให้เวียนศีรษะ เป็นลมได้ง่าย (Gales & Menard, 1995; Hitcho et al., 2004) ส่วนยากลุ่ม Anti-hypertensive Diuretics และ Antipsychotics ไม่มีความสัมพันธ์ต่อการพลัดตกหกล้มในผู้ป่วยผู้ใหญ่ (Richardson, 2002)

1.6 ลักษณะการเคลื่อนไหว (mobility) เป็นปัจจัยที่มีความสัมพันธ์ทั้งทางตรง และทางอ้อมที่เกี่ยวข้องกับการพลัดตกหกล้ม โดยประเมินจากความอ่อนแรงของร่างกาย (weak) หรือความสามารถในการเคลื่อนไหวของร่างกายส่วนล่างที่ลดลง (impaired gait) ซึ่งอาจจะเกิดจากโรค หรืออายุที่มากขึ้น หรือประเมินจากการเดินว่าสามารถเดินได้ปกติ โดยไม่ต้องใช้อุปกรณ์ช่วย หรือต้องใช้อุปกรณ์ในการช่วยเดิน เช่น ไม้เท้า (cane) ไม้พยุง 4 ขา (walker) หรือใช้ไม้ค้ำยัน (crutches) เป็นต้น ซึ่งผู้ป่วยที่ต้องใช้อุปกรณ์ช่วยเดินเหล่านี้ จะทำให้เพิ่มความเสี่ยงต่อการพลัดตกหกล้มสูงขึ้น (Morse, 2009) นอกจากนี้ ยังมีงานวิจัย ที่ศึกษาเกี่ยวกับผู้ป่วยที่มีปัญหาด้านการเคลื่อนไหวพบว่า ผู้ป่วยที่มีปัญหาการเคลื่อนไหว มีความสัมพันธ์เชิงบวกกับการเกิดการพลัดตกหกล้ม (Lahmann et al., 2015; Meijers, Halfens, Neyens, Luiking, & Verlaan, 2012; Hendrich et al., 2007) และจากการทบทวนวรรณกรรมอย่างเป็นระบบของ Shah, Williams, Joshi, and Aziz (2012) พบว่า ผู้ป่วยที่เกิดการพลัดตกหกล้มร้อยละ 70 มีการใช้อุปกรณ์ช่วยเดินขณะอยู่ที่บ้าน

1.7 ความต้องการเกี่ยวกับการขับถ่าย (toileting need) Hendrich et al. (2007) ศึกษาการพัฒนาเครื่องมือเพื่อใช้ทำนายการพลัดตกหกล้มที่ใช้ปฏิบัติการทางคลินิกในโรงพยาบาลพบว่า การเปลี่ยนแปลงของการขับถ่ายเป็นปัจจัยที่มีนัยสำคัญต่อการพลัดตกหกล้ม ได้แก่ ขับถ่ายปัสสาวะบ่อยในตอนกลางคืน (nocturia) และกลั้นปัสสาวะไม่ได้ (urinary

incontinence) ซึ่งสอดคล้องกับผลการศึกษาของ Lahmann et al. (2015) ที่พบทวนวรรณกรรมอย่างเป็นระบบพบว่า ผู้ป่วยที่มีปัญหาเรื่องการขับถ่าย ท้องเสีย และการขับถ่าย ปัสสาวะบ่อยร่วมกับมีปัญหาด้านการเคลื่อนไหวมีความสัมพันธ์กับการพลัดตกหกล้ม

1.8 ภาวะความเจ็บป่วย มีการศึกษาที่สนับสนุนปัจจัยดังกล่าว ได้แก่ การศึกษาของ Mion et al. (1989) ที่ศึกษาอุบัติการณ์ของการพลัดตกหกล้มและลักษณะของผู้ป่วยของหน่วยฟื้นฟูสภาพพบว่า มีโรคที่เจ็บป่วยทางจิตเพียงโรคเดียวเท่านั้นที่มีความสัมพันธ์ต่อการพลัดตกหกล้มอย่างมีนัยสำคัญ นอกจากนี้ ยังมีการศึกษาการพลัดตกหกล้มในผู้ป่วยเฉพาะโรคโดย Nytrom and Hellstrom (2013) ได้ศึกษาอุบัติการณ์การพลัดตกหกล้มในผู้ป่วยโรคเส้นเลือดในสมองตีบ (stroke) โดยให้เหตุผลว่า ผู้ป่วยเหล่านี้มีปัญหาบกพร่องด้านโครงสร้าง ความจำ การพูด การสื่อสาร และการเคลื่อนไหว ซึ่งไม่สามารถที่จะควบคุมตัวเองได้เนื่องจากมีปัญหาของการสั่งงานจากสมอง ซึ่งขึ้นอยู่กับบริเวณของสมองที่ได้รับอันตรายพบว่า อุบัติการณ์การพลัดตกหกล้มในผู้ป่วยโรคเส้นเลือดในสมองตีบมีการพลัดตกหกล้มอย่างน้อย 1 ครั้ง ในระยะเวลา 6 สัปดาห์ หลังจากมีอาการ ซึ่งสอดคล้องกับการศึกษาของ พัชรินทร์ คณะพล (2557) ที่ศึกษาความชุกของปัจจัยทำนายการเกิดการพลัดตกหกล้มของผู้ป่วยโรคหลอดเลือดสมองหลังการจำหน่ายออกจากโรงพยาบาลพบว่า มีความชุกของการเกิดการพลัดตกหกล้มร้อยละ 34.15 โดยมีปัจจัยภายในบุคคลที่เกิดจากพยาธิสภาพของโรค ได้แก่ การทรงตัว กล้ามเนื้ออ่อนแรง ความบกพร่องในการดูแลตนเอง และความรู้กิดบกพร่อง มีความสัมพันธ์ต่อการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.9 ปัจจัยภายในอื่น ๆ นอกจากปัจจัยภายในที่มีผลต่อการพลัดตกหกล้มดังกล่าวแล้ว ยังมีงานวิจัยที่แสดงให้เห็นว่า ยังมีปัจจัยภายในอื่น ๆ ที่มีผลต่อการพลัดตกหกล้ม เช่น การศึกษาของ Mion et al. (1989) ในหน่วยฟื้นฟูสภาพพบว่า เพศชายมีจำนวนการพลัดตกหกล้มมากกว่าเพศหญิงอย่างมีนัยสำคัญ พบว่า ผู้ป่วยที่นอนโรงพยาบาลนานกว่าจะมีจำนวนการพลัดตกหกล้มมากกว่าผู้ป่วยที่นอนในโรงพยาบาลสั้นกว่าอย่างมีนัยสำคัญ และยังมีงานวิจัยที่ศึกษาพบว่า อาการง่วงซึม รูปแบบที่ทำงานของพยาบาล การได้รับสารที่เป็นอันตราย การพักผ่อนไม่ได้ของผู้ป่วย และผู้ป่วยที่เคยได้รับการรักษาในหอผู้ป่วยหนักมีผลต่อการพลัดตกหกล้มอย่างมีนัยสำคัญ (Aston et al. อ้างถึงใน Evans et al., 1998)

นอกจากนี้ยังมีผลการศึกษาของ Morse (2009) พบว่า การได้รับการคาสายเพื่อให้สารละลายทางหลอดเลือดดำเป็นปัจจัยที่มีนัยสำคัญต่อการพลัดตกหกล้ม

1.10 ปัจจัยภายนอกอื่น ๆ ที่เกิดจากสิ่งแวดล้อมที่เป็นอันตรายหรือกิจกรรมที่ก่อให้เกิดอันตราย ได้แก่

1.10.1 ระยะเวลาอนรรักษาตัวในโรงพยาบาล เป็นปัจจัยหนึ่งที่มีงานวิจัยรายงานว่า มีผลต่อการพลัดตกหกล้มของผู้ป่วยในโรงพยาบาล Mion et al. (1989) แต่จากการศึกษาพบว่า ผลการวิจัยยังมีความไม่สอดคล้องกัน โดย Rapp et al. (2016) ได้ศึกษาระยะเวลาที่เกิดการพลัดตกหกล้มของผู้สูงอายุหลังจากรับไว้ในโรงพยาบาลพบว่า ในช่วง 2-4 วันแรก มีอัตราการพลัดตกหกล้มเท่ากับ 24.6 ต่อ 1,000 วันนอน และมีอัตราการลดลงเป็น 13.00 ต่อ 1,000 วันนอนหลังจากสัปดาห์ที่ 2 ส่วนงานวิจัยอื่นพบว่า เกิดในช่วงอื่น หรือส่วนใหญ่จะอยู่ในช่วงแรก หรือช่วงปลายของการอยู่ในโรงพยาบาล แต่ไม่ได้ขยายผลว่าเป็นช่วงใด แต่เป็นประเด็นในสิ่งแวดล้อมที่แตกต่างกัน อาจเป็นช่วงที่ต้องการความช่วยเหลือหรืออ่อนเพลียจากการพักผ่อนจากความเจ็บป่วย และการรักษาของโรงพยาบาล จึงนำไปสู่ปัจจัยที่เกี่ยวกับระยะเวลาขณะเกิดพลัดตกหกล้ม (Evans et al., 1998)

1.10.2 การดูแลของญาติหรือผู้ดูแลในโรงพยาบาล เป็นปัจจัยหนึ่งที่มีส่วนทำให้ผู้ป่วยเกิดการพลัดตกหกล้ม การมีส่วนร่วมในการเฝ้าระวังและการดูแลผู้ป่วย รวมทั้งการแจ้งเตือนหรือขอความช่วยเหลือจากเจ้าหน้าที่ในการทำกิจกรรมต่าง ๆ ไม่ละทิ้งผู้ป่วยให้อยู่โดยลำพังเพื่อป้องกันการพลัดตกหกล้ม มีรายงานวิจัยที่ศึกษาพบว่าส่วนมากการพลัดตกหกล้มของผู้ป่วยในโรงพยาบาล มักจะเป็นการพลัดตกหกล้มแบบไม่มีผู้พบเห็น (un witness falls) เกิดขึ้นขณะที่ผู้ป่วยอยู่เพียงลำพังหรือผู้ดูแลไม่อยู่ด้วย เช่น หกล้มในห้องน้ำขณะฉวยาหรืออยู่หน้าห้องน้ำ (Hitcho et al., 2004) และการศึกษาเกี่ยวกับการมีผู้ดูแล และการป้องกันการพลัดตกหกล้มพบว่า การมีส่วนร่วมของญาติหรือการมอบหมายให้มีผู้ดูแลผู้ป่วยจะช่วยลดอุบัติการณ์การพลัดตกหกล้มของผู้ป่วยได้ (Krauss et al., 2007)

การทำเหมืองข้อมูล (data mining)

การทำเหมืองข้อมูล คือ กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์ หรือสร้างตัวแบบสำหรับการจำแนกหน่วย หรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์ เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์ และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูล มีขั้นตอนการดำเนินงานหลายขั้นตอน ซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์ เป็นต้น การดำเนินงานมักอยู่ในลักษณะของการสร้างตัวแบบ (modeling) ที่อธิบายความเป็นไป หรือสภาพการณ์หนึ่งที่เกิดขึ้นแล้ว หรือที่เราทราบคำตอบ แล้วนำตัวแบบนี้มาใช้อธิบายสถานการณ์ที่ยังไม่เกิดขึ้น หรือที่ไม่ทราบคำตอบ ตัวแบบเหล่านี้ อาจเป็นตัวแบบที่เรียบง่ายไปจนถึงตัวแบบที่ยุ่ยากซับซ้อน และอาจใช้การผสมผสานแนวคิด หรือเครื่องมือต่าง ๆ เข้าด้วยกัน เพื่อที่จะสามารถสกัดข้อความรู้ที่อยู่ในข้อมูลขนาดใหญ่ได้ โดยใช้เทคโนโลยีคลังข้อมูล (data warehouse) เข้ามาช่วยในการจัดการข้อมูลเพื่อเพิ่มประสิทธิภาพของการทำเหมืองข้อมูล ดังนั้น ถ้ามีฐานข้อมูลขนาดใหญ่ที่มีข้อมูลคุณภาพดี เทคโนโลยีการทำเหมืองข้อมูลจะช่วยค้น หรือแสวงหาโอกาสทางธุรกิจใหม่ โดยการทำเหมืองข้อมูลจะก่อให้เกิดกระบวนการอัตโนมัติในการค้นพบสารสนเทศหรือข้อความรู้ในฐานข้อมูลขนาดใหญ่ ด้วยการใช้วิธีการ เช่น การพยากรณ์แนวโน้ม และพฤติกรรมผู้บริโภคแบบอัตโนมัติ หรือเกิดกระบวนการอัตโนมัติในการค้นพบรูปแบบที่ไม่เคยรู้จักมาก่อน ด้วยการใช้วิธีการค้นหาเข้าไปในรายละเอียดของฐานข้อมูลเพื่อหารูปแบบที่ซ่อนอยู่ในฐานข้อมูลนั้น การทำเหมืองข้อมูลมีประโยชน์ในการวิเคราะห์ฐานข้อมูลขนาดใหญ่เพื่อแก้ปัญหาการตัดสินใจทางธุรกิจ ตัวอย่างการประยุกต์การทำเหมืองข้อมูลในธุรกิจด้านต่าง ๆ ได้แก่ การใช้เหมืองข้อมูลเพื่อตอบคำถามว่า ลูกค้ารายใดมีโอกาสจะยุติการใช้บริการ หรือหันไปใช้บริการของธุรกิจคู่แข่ง ลักษณะของลูกค้าที่สร้างรายได้ให้กิจการมากที่สุด หรือน้อยที่สุดเป็นอย่างไร ควรนำเสนอสินค้า หรือบริการใดให้ลูกค้า

รายได้ ลูกค้ายรายใดมีโอกาสยุติการใช้บริการมากที่สุด ลูกค้ายสินเชื่อยรายใดมีโอกาสกลายเป็น NPL มาก ลักษณะของลูกค้ายที่มีความเสี่ยงสูงหรือต่ำเป็นอย่างไร รูปแบบการใช้จ่าย ลักษณะใดแสดงถึงการทุจริตการใช้บัตรเครดิต รูปแบบของโครงสร้างหนี้แบบใดที่แสดงแนวโน้มการไม่สามารถชำระหนี้ของลูกค้าย การพยากรณ์การใช้จ่ายผ่านบัตรเครดิตของลูกค้ายแต่ละกลุ่ม ลักษณะหรือรูปแบบการเบิกจ่ายที่แสดงถึงความเป็นไปได้ในการทุจริตเป็นอย่างไร การพยากรณ์ว่าลูกค้ายรายใดจะซื้อกรรมธรรม์เพิ่มเติม การค้นหาพฤติกรรมที่ต่อทุจริตในการขอสินไหมทดแทน การหารูปแบบของพฤติกรรมของลูกค้ายกลุ่มเสี่ยง ควรติดต่อลูกค้ายบ่อยเพียงใด ในโอกาสใด ผู้ใดมีโอกาสสูงในการตอบรับการเสนอขาย ครั้งนี้ การเสนอขายแบบใดจะได้รับการตอบสนองที่ดีจากลูกค้าย จะจัดกลุ่มลูกค้ายอย่างไร จึงจะได้ประโยชน์สูงสุด การหาความสัมพันธ์ระหว่างลักษณะของประชากรของลูกค้ายและการประยุกต์การทำเหมืองข้อมูลในการดำเนินการด้านต่าง ๆ ได้แก่ การพยากรณ์ยอดขาย การวิเคราะห์การส่งเสริมการขาย การวิเคราะห์ช่องทางจำหน่าย การรักษาลูกค้าย การจัดกลุ่มลูกค้าย การวิเคราะห์การสร้างกำไร การวิเคราะห์ความพึงพอใจของลูกค้าย ระบบอัตโนมัติ การบริการลูกค้าย ระบบช่วยแก้ปัญหา การตรวจสอบการทุจริต การวิเคราะห์การขอสินไหมทดแทน การวางแผนการผลิตสินค้า การจัดการทรัพยากร การพยากรณ์เครือข่าย การเลือกสถานที่ตั้งร้าน หรือสาขาของกิจการ และการวิเคราะห์ผลการดำเนินงานของร้านหรือสาขา การทำเหมืองข้อมูลจะช่วยฝ่ายการตลาด ในการกำหนดรูปแบบการส่งเสริมการขายได้อย่างแม่นยำ และยังจัดรูปแบบของการส่งเสริมการขายให้ตรงกับความต้องการ ความอยากได้ และความคาดหวังของลูกค้าย ถ้ามีข้อมูลที่จำเป็นอยู่ในฐานข้อมูล การทำเหมืองข้อมูลก็จะสามารถพยากรณ์พฤติกรรมของลูกค้ายได้ ฤกษ์แจสำคัญอยู่ที่จะต้องหารูปแบบที่เหมาะสมกับปัญหาธุรกิจปัจจุบัน ตัวอย่างเช่น การรวมการทำเหมืองข้อมูลเข้ากับการบริหารส่งเสริมการขายจะมีผลดีช่วยให้ผลของการส่งเสริมการขายมีประสิทธิภาพดีขึ้น โดยการใช้ระบบการให้คะแนน ซึ่งจะจำแนกลูกค้ายเป็นส่วนที่คาดหวังได้ วงจรการค้าที่เร็วขึ้นจะช่วยลดต้นทุน เพิ่มการเข้าถึงลูกค้ายส่วนที่คาดหวังได้ดีกว่าคู่แข่ง และเพิ่มความแม่นยำในการจำกัดความผิดพลาดที่เกิดจากการทำงานด้วยมือ โดยการใช้โปรแกรมที่ทำงานอัตโนมัติ ในขั้นตอนต่าง ๆ (สุชาดา กิระนันท์, 2545)

การทำเหมืองข้อมูลในสารสนเทศทางการแพทย์

สารสนเทศทางชีวการแพทย์ (biomedical informatics) นับวันเป็นที่รู้จักและมีผู้ให้ความสนใจกันมากขึ้น (ปัทมวิษณุ วงศ์วิวัฒนานนท์, 2008) ข้อมูลสารสนเทศการวิจัยทางการแพทย์เติบโตขึ้นอย่างรวดเร็ว สืบเนื่องจากความก้าวหน้าทางเทคโนโลยีการแพทย์สมัยใหม่ เช่น เทคโนโลยีโมเลกุล พันธุกรรมศาสตร์ เทคนิคการจัดเรียงของยีน การจำแนกประเภทโปรตีน การถ่ายภาพทางการแพทย์ การบันทึกเวชระเบียนผู้ป่วย และข้อมูลงานวิจัยทางการแพทย์อื่น ๆ อีกจำนวนมาก การทำเหมืองข้อมูลกับสารสนเทศทางการแพทย์ (health/medical informatics) ยังคงเป็นเรื่องใหม่ ๆ นักวิจัยในการทำเหมืองข้อมูลเริ่มให้ความสนใจมากขึ้น โดยเหตุผลสำคัญ คือ แนวโน้มการแพทย์ที่เป็นส่วนบุคคลมากขึ้น (personalized medicine) ที่มุ่งเน้นไปที่การค้นหาคำรักษาและมาตรการป้องกันส่วนบุคคล แหล่งข้อมูลที่สำคัญในการทำเหมืองข้อมูลยังคงมาจากความก้าวหน้าทางเทคโนโลยีชีวภาพ เราอาจแบ่งงานทำเหมืองข้อมูลในสารสนเทศทางการแพทย์ออกได้เป็น 4 ด้าน ดังนี้

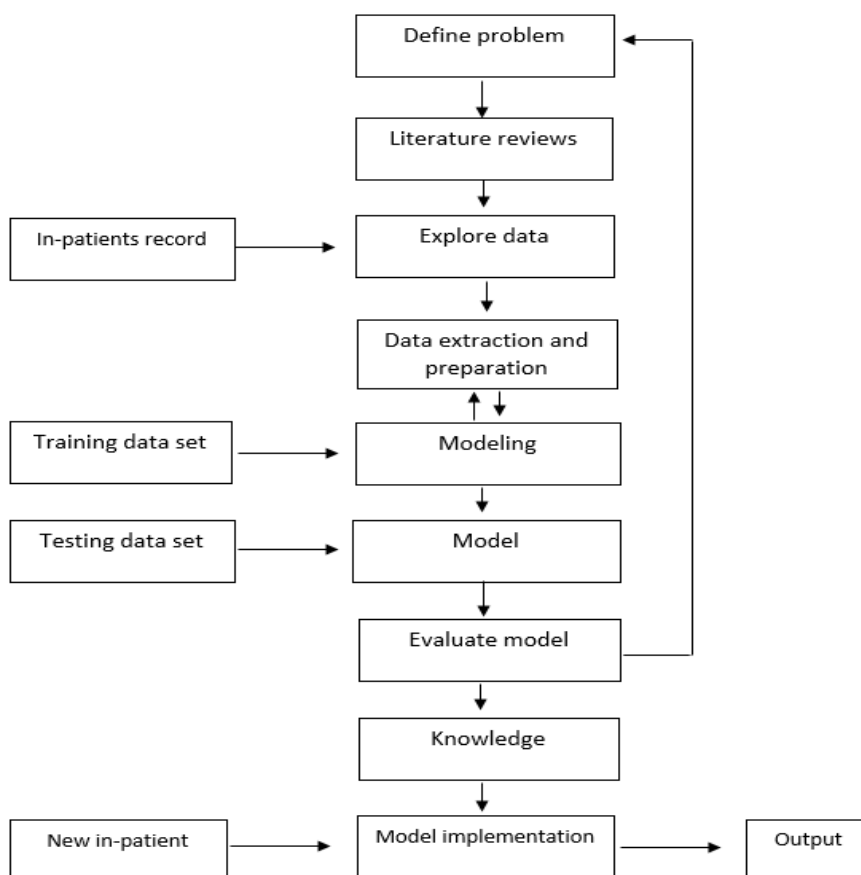
1. งานวินิจฉัยโรค เพื่อวิเคราะห์ว่าผู้ป่วยเจ็บป่วยจากภาวะทางการแพทย์อะไรบ้าง เช่น การวินิจฉัยผู้ป่วยมะเร็งเต้านมและลำไส้ระยะแรกเป็นเรื่องยาก หากกระทำโดยวิธีการแพทย์ปกติ การใช้ข้อมูลทางพันธุศาสตร์ช่วยได้มากในการวินิจฉัยที่รวดเร็วขึ้น รวมทั้งความแม่นยำในการวินิจฉัยด้วย

2. การพยากรณ์โรค เพื่อคาดคะเนว่าผู้ป่วยจะหายเจ็บป่วยได้กี่ขนาดไหนและโรคจะดำเนินไปอย่างไรตามระยะเวลา เช่น การใช้ Biomarker ในการคาดคะเนว่าอวัยวะที่ปลูกถ่ายสามารถทนอยู่ในร่างกายของผู้รับได้นานเท่าใด

3. การรักษาที่เหมาะสม เพื่อคาดคะเนผลการรักษาโรค เช่น การใช้ Biomarker ในการคาดคะเนว่าการรักษาทางเคมีบำบัดได้ผลอย่างไร

4. การทำความเข้าใจเกี่ยวกับกลไกของโรค เพื่อให้เกิดแนวคิดหรือความรู้ใหม่เกี่ยวกับสาเหตุการเกิดโรค เช่น การวิจัยเกี่ยวกับ Signaling Pathway ในระหว่างการติดเชื้อไวรัส เป็นต้น

ขั้นตอนการทำเหมืองข้อมูล



ภาพ 1 ขั้นตอนการทำเหมืองข้อมูล

1. การทำความเข้าใจปัญหา โดยตั้งเป้าหมายว่า การทำเหมืองข้อมูลครั้งนี้ต้องการที่จะแก้ปัญหาใด โดยงานวิจัยนี้มีเป้าหมายจะทำนายการการพลัดตกหกล้มของผู้ป่วยใหม่ที่เข้ามาใช้บริการเป็นผู้ป่วยในของโรงพยาบาล ตั้งเกณฑ์วัดความสำเร็จในการทำเหมืองข้อมูล ประเมินสถานการณ์ในด้านต่าง ๆ เช่น ความรู้พื้นฐานในเรื่องที่จะทำเหมืองข้อมูล มีเพียงพอหรือไม่ และประโยชน์จากการทำเหมืองข้อมูลจะคุ้มค่ากับต้นทุนที่เสียไปหรือไม่ วางแผนการทำเหมืองข้อมูลว่าจะเก็บข้อมูลอย่างไร และใช้อัลกอริทึมใดในการทำเหมืองข้อมูล

2. การศึกษาข้อมูลและงานวิจัย โดยศึกษาทฤษฎีและงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับการใช้เทคนิคโครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึก วิธีต้นไม้ตัดสินใจ และวิธีถดถอยโลจิสติก ทำนายการพลัดตกหกล้มของผู้ป่วยที่เข้ามาใช้บริการรักษาเป็นผู้ป่วยใน

ของโรงพยาบาล

3. การทำความเข้าใจข้อมูล ได้แก่ เก็บรวบรวมข้อมูลที่เกี่ยวข้อง ข้อมูลมีความถูกต้องน่าเชื่อถือ และมีปริมาณมากพอหรือยัง ข้อมูลมีความเหมาะสม มีรายละเอียดเพียงพอต่อการนำไปใช้วิเคราะห์หรือไม่ กำหนดคุณลักษณะของข้อมูลที่ได้เก็บมาได้ และตรวจสอบข้อมูลขั้นต้น โดยตรวจสอบทั้งความสมบูรณ์และความถูกต้องของข้อมูล

4. การเตรียมข้อมูล โดยการคัดเลือกข้อมูลที่จะนำมาใช้ กำหนดเป้าหมายว่าจะวิเคราะห์อะไร เลือกใช้เฉพาะข้อมูลที่เกี่ยวข้องกับสิ่งที่จะวิเคราะห์ ทำความสะอาดข้อมูล เป็นกระบวนการเตรียมข้อมูลให้เหมาะสมที่สุด เพื่อนำไปใช้ในขั้นตอนต่อไป ซึ่งมีวิธีการ ได้แก่ ลบข้อมูลที่ซ้ำซ้อน แก้ไขข้อมูลที่ผิดพลาด ได้แก่ ผิดรูปแบบ ขาดหาย และแปลกแยกจากข้อมูลอื่น การแก้ไขข้อมูลให้ถูกต้องสมบูรณ์ เช่น การแก้ไขค่าว่างของข้อมูล ปรับเปลี่ยนข้อมูลให้มีรูปแบบที่เหมาะสมพร้อมที่จะนำไปใช้ในการวิเคราะห์หาตัวแบบ ทำนายตามอัลกอริทึมที่เลือกใช้

5. การสร้างตัวแบบทำนาย โดยการเลือกอัลกอริทึมที่เหมาะสมในการทำเหมืองข้อมูล กำหนดรูปแบบการทดสอบผลลัพธ์ และสร้างแบบจำลองตามอัลกอริทึมที่เลือก

ในขั้นตอนนี้ถ้าพบว่า รูปแบบของข้อมูลไม่เหมาะสมในการสร้างตัวแบบทำนาย จะกลับไปยังขั้นตอนก่อนหน้า เพื่อปรับเปลี่ยนรูปแบบของข้อมูลใหม่ให้เหมาะสมกับอัลกอริทึมที่เลือกใช้ โดยการสร้างตัวแบบทำนายนั้นจะใช้ชุดข้อมูลเรียนรู้ในการสร้าง และเมื่อได้ผลลัพธ์ตัวแบบจำลองแล้วจะทดสอบตัวแบบด้วยชุดข้อมูลทดสอบ

6. การวัดประสิทธิภาพของตัวแบบ โดยเป็นการแบ่งข้อมูลเพื่อใช้ในการทดสอบ เพื่อดูว่าแบบจำลองนี้ได้ผลหรือไม่เพียงใด และมีความผิดพลาดตรงไหน ตอบโจทย์ที่ตั้งไว้หรือไม่ ถ้าไม่ตอบโจทย์ จะต้องดำเนินการแก้ไขในกระบวนการตั้งแต่การทำความเข้าใจปัญหาใหม่เป็นต้นมา

7. การนำไปใช้ เมื่อได้ตัวแบบทำนายที่เหมาะสมแล้ว จะนำตัวแบบที่ได้ไปใช้ทำนายการพลัดตกหกล้มของผู้ป่วยที่เข้ามาใช้บริการเป็นผู้ป่วยในของโรงพยาบาลว่ามีผลลัพธ์การทำนายเป็นอย่างไร โดยกำหนดผลลัพธ์ (output) เป็นแบบสองทางเลือก (binary class) มีคำตอบคือ “หกล้ม” และ “ไม่หกล้ม” เพื่อนำไปปรับปรุงรูปแบบการประเมินการพลัดตกหกล้มของผู้ป่วยและวางแผนการทำงานในการให้บริการผู้ป่วยต่อไป

เทคนิคการเรียนรู้แบบมีผู้สอน (supervised learning)

หรือเรียกว่า การสร้างตัวแบบในการทำนาย (predictive modeling) เทคนิคประเภทนี้จะเน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีต เพื่อนำมาสร้างตัวแบบสำหรับทำนายหรือคาดการณ์สิ่งที่เกิดขึ้นในอนาคต ตัวแบบในที่นี้อาจจะเป็นสมการทางคณิตศาสตร์หรือกฎต่าง ๆ ก็เป็นได้ เทคนิคนี้สามารถแบ่งย่อยได้อีก คือ การจำแนกประเภทข้อมูล (classification) และการประมาณค่าข้อมูลการถดถอย (regression) ซึ่งทั้งสองเทคนิคจะมีลักษณะที่คล้ายกันมาก แต่แตกต่างกันที่คำตอบที่ต้องการทำนาย ซึ่งการจำแนกประเภทข้อมูลจะทำนายข้อมูลที่มีค่าเป็นนามินอล (nominal) เช่น เพศชาย เพศหญิง หรือค่าที่ไม่ใช่ตัวเลขนั่นเอง ส่วนการประมาณค่าข้อมูลจะใช้กับข้อมูลคำตอบที่เป็นตัวเลขเท่านั้น

การจำแนกประเภทข้อมูล (classification)

การจำแนกประเภทข้อมูล คือ กระบวนการสร้างตัวแบบจำแนกประเภทข้อมูล (data classification model) เพื่อทำนายกลุ่มของข้อมูลใหม่ (unseen data) ตัวอย่างของกลุ่มเช่น กลุ่มของลูกค้าที่ซื้อคอมพิวเตอร์และไม่ซื้อคอมพิวเตอร์ กลุ่มของลูกค้าที่มีฐานะดี ปานกลาง และไม่ดี กลุ่มของการผลิตสินค้าที่ผ่านเกณฑ์และไม่ผ่านเกณฑ์ ซึ่งจะเรียกว่าคลาส (class) ของข้อมูล ซึ่งในคลาสเดียวกันนั้นจะต้องมีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ในคลาสที่แตกต่างกัน การสร้างตัวแบบจำแนกข้อมูลจะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลเรียนรู้ เป็นชุดข้อมูลที่มีบทบาทในการสร้างตัวแบบจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบเป็นชุดข้อมูลประเมินความถูกต้องของตัวแบบจำแนกประเภทข้อมูล

เทคนิคต้นไม้ตัดสินใจ (decision tree)

ในช่วงปลายยุค ค.ศ. 1970 และต้นปีค.ศ. 1980 Quinlan (1986) นักวิจัยทางด้านการเรียนรู้ของเครื่อง (machine learning) ได้พัฒนาอัลกอริทึมสำหรับสร้างต้นไม้ตัดสินใจ ที่มีชื่อว่า ID3 (iterative dichotomiser) ขึ้นมา ซึ่งต่อมาได้พัฒนาต่อยอด ID3 ไปเป็น C4.5 ที่ได้กลายมาเป็นอัลกอริทึมพื้นฐานที่ใช้สำหรับเปรียบเทียบประสิทธิภาพการทำงานของ

อัลกอริทึมต่าง ๆ ทางด้านการเรียนรู้แบบมีผู้สอน (supervised learning) ต่อมาได้มีกลุ่มของนักสถิติ Breiman, Friedman, Olshen, and Stone (1984) ได้ตีพิมพ์หนังสือ Classification and Regression Trees (CART) ขึ้น เพื่ออธิบายในรุ่นของต้นไม้ตัดสินใจที่มี 2 ทางเลือก (binary decision trees) ID3 C4.5 และ CART ได้ประยุกต์ใช้วิธีการเชิงละโมภ (greedy approach) สร้างต้นไม้ส่วนใหญ่ใช้วิธีการแบบ “Top-down Recursive Divide-and-conquer” โดยพิจารณาชุดข้อมูลสำหรับเรียนรู้ด้วยการแบ่งข้อมูลออกเป็นส่วนย่อย ๆ ในระหว่างกระบวนการสร้างต้นไม้ การเรียนรู้ของต้นไม้ตัดสินใจ เป็นการเรียนรู้โดยการจำแนกประเภทข้อมูล โดยแบ่งข้อมูลออกเป็นกลุ่ม (class) ต่าง ๆ โดยใช้คุณลักษณะแอทริบิวต์ข้อมูลในการจำแนกประเภท ต้นไม้ตัดสินใจที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภท และคุณลักษณะแต่ละตัวมีความสำคัญมากน้อยต่างกันอย่างไร เพราะฉะนั้น การจำแนกประเภทมีประโยชน์ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลและตัดสินใจได้อย่างถูกต้อง

ค่าเกนความรู้ (information gain)

สูตรที่ใช้การคำนวณค่าเกนความรู้ (information gain) ซึ่งจะเป็นตัวชี้วัดการแบ่งข้อมูลออกเป็นชุดข้อมูลย่อยที่ได้รับความนิยมอย่างแพร่หลาย โดยค่านี้จะถูกประยุกต์ใช้ในอัลกอริทึม ID3 ซึ่งจะเลือกแอทริบิวต์สำหรับแบ่งข้อมูลจากแอทริบิวต์ ที่มีค่าเกนความรู้สูงที่สุด จะเป็นการเลือกแอทริบิวต์ที่ต้องการข้อมูลที่น้อยที่สุด ระบุหรือแบ่งข้อมูลออกเป็นชุดข้อมูลย่อย การดำเนินงานเริ่มต้นจะต้องเริ่มจากการคำนวณ ค่า $\text{Info}(D)$ หรือเรียกอีกอย่างหนึ่งว่า ค่าเอนโทรปี (entropy of D) ที่หมายถึง ค่าเฉลี่ยของปริมาณข้อมูลที่ต้องการ ระบุถึงหมวดหมู่ของข้อมูลเรคคอร์ดหนึ่ง ๆ ในชุดข้อมูล ที่ จะขึ้นกับอัตราส่วนของจำนวนเรคคอร์ด ที่สอดคล้องกับแต่ละหมวดหมู่ สามารถคำนวณได้ดังนี้

$$\text{Info}(D) = - \sum_{i=1}^m p_i \log_2(p_i)$$

เมื่อ p_i คือ ค่าความน่าจะเป็นที่เรคคอร์ดหนึ่ง ๆ จะมีหมวดหมู่ของข้อมูลเป็นหมวดหมู่ C_i ที่ซึ่งสามารถคำนวณได้จาก $\frac{|C_i, D|}{|D|}$ สมมติว่า เราต้องการที่จะแบ่งข้อมูล D ออกเป็นชุดข้อมูลย่อย โดยใช้แอทริบิวต์ A ที่มีค่าที่เกิดขึ้นในชุดข้อมูล v ค่าที่แตกต่างกัน $\{a_1, a_2, \dots, a_v\}$ ถ้าแอทริบิวต์ A มีค่าที่เกิดขึ้นเป็นแบบไม่ต่อเนื่อง จะพิจารณาการแบ่ง

ชุดข้อมูลออกเป็น v ชุดข้อมูลย่อย โดยสามารถแบ่งได้เป็น $\{D_1, D_2, \dots, D_v\}$ โดยที่ D_j ใด ๆ จะบรรจุไปด้วยเซตของเรคคอร์ดที่แอทริบิวต์ A มีค่าเป็น a_j เป็นต้น โดยแต่ละชุดข้อมูลย่อยจะสอดคล้องกับกิ่งย่อยที่แตกออกจากโหนด N ที่กำลังพิจารณาอยู่ ดังนั้น ในการแบ่งชุดข้อมูลย่อยจะต้องพยายามทำให้แต่ละชุดข้อมูลย่อยประกอบไปด้วยเรคคอร์ดที่มีหมวดหมู่เหมือนกันทั้งหมด แต่อย่างไรก็ตาม มักเป็นการยากที่จะมีเรคคอร์ดที่มีหมวดหมู่เหมือนกันมีค่าในแอทริบิวต์เหมือนกัน เนื่องจากในชุดข้อมูลย่อยที่ถูกแบ่งมักจะมีเรคคอร์ดที่มีหมวดหมู่ที่แตกต่างกัน ดังนั้น จะพิจารณาว่าเรคคอร์ดที่มีค่าในแอทริบิวต์เหมือนกันนั้นมีความเหมือนกันของหมวดหมู่เป็นเช่นไร โดยจะต้องคำนวณค่า $\text{Info}_A(D)$ ที่หมายถึง จำนวนข้อมูลที่คาดว่าจะใช้สำหรับแบ่งข้อมูล D ออกเป็นชุดข้อมูลย่อย โดยพิจารณาแอทริบิวต์ A ซึ่งสามารถคำนวณได้ดังนี้

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{D} \times \text{Info}(D_j)$$

เมื่อ $\frac{|D_j|}{|D|}$ หมายถึง จำนวนเรคคอร์ดในชุดข้อมูล D ที่มีค่าในแอทริบิวต์ A เป็น a_j หารด้วยจำนวนเรคคอร์ดทั้งหมดในชุดข้อมูล D หลังจากคำนวณค่า $\text{Info}(D)$ และค่า $\text{Info}_A(D)$ ค่าเกณฑ์ความรู้สำหรับการพิจารณาแอทริบิวต์ A จะสามารถคำนวณได้จากค่าความแตกต่างระหว่างปริมาณข้อมูลที่ต้องการในการระบุถึงหมวดหมู่ของข้อมูลสำหรับเรคคอร์ดหนึ่ง ๆ กับจำนวนที่คาดว่าจะใช้สำหรับแบ่งข้อมูล D ออกเป็นชุดข้อมูลย่อย โดยพิจารณาแอทริบิวต์ A ที่สามารถแสดงการคำนวณได้ดังนี้

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

ค่า $\text{Gain}(A)$ จะบ่งบอกปริมาณเกณฑ์ความรู้ที่เราจะรับด้วยการพิจารณาแอทริบิวต์ A และการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อย ตามการพิจารณาแอทริบิวต์ A ถ้าค่า $\text{Gain}(A)$ เป็นค่าเกณฑ์ความรู้ที่มีค่ามากที่สุดในบรรดาค่าเกณฑ์ความรู้ของแอทริบิวต์ทั้งหมด แอทริบิวต์ A จะเป็นแอทริบิวต์ที่ดีที่สุดสำหรับการจำแนกหรือแบ่งชุดข้อมูลเป็นชุดข้อมูลย่อย เนื่องจากแอทริบิวต์ A ต้องการปริมาณข้อมูลที่น้อยที่สุดที่ใช้ในการจำแนกข้อมูล (Minimal $\text{Info}_A(D)$)

ค่าอัตราส่วนเกน (gain ratio)

จะเป็นตัววัดการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยที่พัฒนามาจากค่าเกนความรู้ เมื่อใช้ค่าเกนความรู้ในการแบ่งชุดข้อมูลจะทำให้เกิดความเอนเอียง เมื่อแอทริบิวต์ที่พิจารณามีค่าที่เกิดขึ้นเป็นจำนวนมาก โดยในการใช้ค่าเกนความรู้ มักจะเลือก แอทริบิวต์ที่มีค่าที่เกิดขึ้นเป็นจำนวนมาก เช่น ลองพิจารณาแอทริบิวต์รายการสินค้าที่มีค่าเป็นรหัสสินค้าต่าง ๆ ถ้าแบ่งชุดข้อมูลตามแอทริบิวต์รายการสินค้า จะทำให้มีชุดข้อมูลย่อยเป็นจำนวนมาก (ชุดข้อมูลย่อยจะมีจำนวนเท่ากับรายการสินค้าทั้งหมดที่มีอยู่ในชุดข้อมูล) โดยแต่ละชุดข้อมูลจะมีข้อมูลเพียง 1 เรคคอร์ดเท่านั้นและจะทำให้ชุดข้อมูลย่อย ๆ นั้นมีชุดข้อมูลที่มีหมวดหมู่ของข้อมูลเหมือนกันหมด ซึ่งจะทำให้ค่า

$Info_{product_ID}(D) = 0$ ที่จะทำให้ค่าเกนความรู้มีค่าสูง $Gain(A) = Info(D) - Info_A(D) = Info(D) - 0 = Info(D)$ ดังนั้น แอทริบิวต์ที่มีลักษณะคล้ายกับแอทริบิวต์รายการสินค้ามักจะถูกเลือกเพื่อใช้ในการแบ่งชุดข้อมูลเสมอ จากความโน้มเอียงที่เกิดขึ้นจากการใช้ค่าเกนความรู้ ได้มีนักวิจัยพยายามที่จะลดความเอนเอียงลง โดยการพัฒนาตัววัดตัวใหม่ที่มีชื่อว่า อัตราส่วนเกน (gain ratio) ที่จะประยุกต์ใช้ในการทำต้นไม้ตัดสินใจ ค่าเกนความรู้ด้วยการใช้ค่า “Split information” ซึ่งสามารถคำนวณได้ดังนี้

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right)$$

โดยค่า $SplitInfo_A(D)$ จะแสดงถึงปริมาณข้อมูลที่ถูกพิจารณาโดยการแบ่งข้อมูลในชุดข้อมูล D ออกเป็น v ชุดข้อมูลย่อยตามค่าแอทริบิวต์ A โดยหลังจากคำนวณหาค่า $SplitInfo_A(D)$ แล้วจะสามารถคำนวณหาค่าอัตราส่วนเกนได้ดังนี้

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

เมื่อกำหนดหาค่าอัตราส่วนเกนของทุก ๆ แอทริบิวต์ที่พิจารณาแล้ว จะเลือกแอทริบิวต์ที่มีค่าอัตราส่วนเกนสูงที่สุดเพื่อเป็นแอทริบิวต์สำหรับแบ่งชุดข้อมูลออกเป็นชุดข้อมูลย่อยต่อไป (ข้อสังเกตในการคำนวณค่า $SplitInfo_A(D)$ มีค่าเท่ากับ 0 จะทำให้การคำนวณค่าอัตราส่วนเกนของแอทริบิวต์ A จะมีค่าไม่แน่นอน)

ดัชนีจีนิ (gini index)

ดัชนีจีนิจะเป็นตัวชี้วัดที่จะพิจารณาความไม่บริสุทธิ์ของชุดข้อมูล D ที่จะมีเซตของเรคคอร์ดที่มีหมวดหมู่ของข้อมูลไม่เหมือนกันอยู่ โดยเริ่มจากการคำนวณหาค่าความไม่บริสุทธิ์ของชุดข้อมูล D ดังนี้

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2$$

เมื่อ p_i คือ ค่าความน่าจะเป็นที่เรคคอร์ดหนึ่ง ๆ จะมีหมวดหมู่ของข้อมูลเป็นหมวดหมู่ C_i ซึ่งสามารถคำนวณได้จาก $\frac{|C_i, D|}{|D|}$ ในการแบ่งชุดข้อมูลโดยใช้ดัชนีจีนิ สามารถแบ่งชุดข้อมูลออกได้เป็น 2 ชุดข้อมูลย่อยเท่านั้น เมื่อพิจารณาแอทริบิวต์ A ใด ๆ ที่มีค่าที่เกิดขึ้นเป็นแบบไม่ต่อเนื่องและมีค่าที่แตกต่างกันทั้งสิ้น v ค่า ($\{a_1, a_2, \dots, a_v\}$) จะต้องค้นหาการแบ่งชุดข้อมูลที่ดีที่สุด ภายใต้การพิจารณาแอทริบิวต์ A ด้วยการพิจารณาทุกสับเซตที่เป็นไปได้ของค่าที่เกิดขึ้นในแอทริบิวต์ A ที่จะถูกจัดอยู่ในเซตย่อย 2 เซตด้วยกัน เช่น แอทริบิวต์รายได้ ที่มีค่าที่เกิดขึ้นในแอทริบิวต์ 3 ค่าด้วยกัน คือ $\{\text{low, medium, high}\}$ ซึ่งทั้งจาก 3 ค่าที่เกิดขึ้นจะสามารถแบ่งค่าทั้ง 3 ออกเป็นเซตย่อย 2 เซตได้ดังต่อไปนี้ $\{\text{low, medium, high}\}, \{\text{low, medum}\}, \{\text{low, high}\}, \{\text{medum, high}\}, \{\text{low}\}, \{\text{medum}\}, \{\text{high}\},$ และ $\{\}$

เมื่อแบ่งค่าที่เกิดขึ้นออกเป็น 2 เซตย่อย จะคำนวณค่าน้ำหนักรวมของค่าความไม่บริสุทธิ์ โดยการพิจารณาแอทริบิวต์ A จะแบ่งชุดข้อมูล D ออกเป็น 2 ชุดข้อมูลย่อย คือ D_1 และ D_2 ดังนั้น จะสามารถหาค่าดัชนีจีนิเมื่อพิจารณาแอทริบิวต์ A ภายใต้การพิจารณา D_1 และ D_2 ได้เป็น

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2)$$

เมื่อคำนวณค่าดัชนีจีนิสำหรับ 2 เซตย่อยใด ๆ ภายใต้แอทริบิวต์ A 2 เซต D_1 และ D_2 ใดที่ให้ค่า $\text{Gini}_A(D)$ น้อยที่สุดจะถูกเลือกเป็นตัวแทนของแอทริบิวต์ A

ในกรณีที่แอทริบิวต์ที่พิจารณามีค่าแบบต่อเนื่อง จะดำเนินการด้วยวิธีการเดียวกันกับการคำนวณหาค่าเอนทาลปี ด้วยการหาจุดแบ่งที่ดีที่สุด โดยเรียงลำดับค่าทั้งหมดจากน้อยไปมาก แล้วหาค่ากลางระหว่าง 2 ค่าใด ๆ จากนั้นนำแต่ละค่ากลางมาพิจารณาเป็นจุดแบ่ง โดยค่าใด ๆ ก็ตาม ที่มีค่าน้อยกว่าหรือเท่ากับค่ากลางจะถูกแบ่งไว้ในชุดข้อมูล

ย่อย D_1 และค่าใดก็ตามที่มีค่ามากกว่าค่ากลางจะถูกแบ่งไว้ในชุดข้อมูลย่อย D_2 ตามลำดับ จากนั้นค่ากลางใดที่ให้ค่า $Gini_A(D)$ น้อยที่สุดจะถูกเลือกเป็นตัวแทนของ แอทริบิวต์ A ขั้นตอนสุดท้ายของการแบ่งชุดข้อมูลด้วยการใช้ค่าดัชนีนี้ จะต้องคำนวณ ส่วนหักลบของความไม่บริสุทธิ์คือ $\Delta Gini(A)$ โดยเมื่อคำนวณค่า $\Delta Gini(A)$ สำหรับทุก แอทริบิวต์แล้ว แอทริบิวต์ใดก็ตามที่มีค่า $\Delta Gini(A)$ มากที่สุด (หรือมีค่า $Gini_A(D)$ น้อย ที่สุด) แอทริบิวต์นั้นจะถูกเลือกเพื่อเป็นแอทริบิวต์สำหรับแบ่งชุดข้อมูลเป็นชุดข้อมูลย่อย ดังนี้

$$\Delta Gini(A) = Gini(D) - Gini_A(A)$$

(Jiawei, Micheline, & Jian, 2012)

ขั้นตอนวิธีต้นไม้ตัดสินใจ (หทัยรัตน์ เกตุมณีชัยรัตน์, 2014)

1. ต้นไม้ตัดสินใจสร้างโดยวิธีแบบเวียนเกิดจากบนลงล่าง Top-down Recursive
2. เริ่มต้นด้วยการนำตัวอย่างการสอนมาสร้างเป็นราก
3. แอทริบิวต์ควรอยู่ในรูปของ Categorical คือ ข้อมูลชนิดกลุ่ม หากเป็นข้อมูลที่อยู่ในรูปค่าต่อเนื่อง (continuous) หรือเป็นตัวเลข (numeric) ควรแบ่งข้อมูลให้เป็นกลุ่มก่อน
4. การสร้างต้นไม้ตัดสินใจมีพื้นฐานมาจากวิธีการเลือกแอทริบิวต์
5. จะหยุดการสร้างต้นไม้ เมื่อทุกข้อมูลใน โหนดนั้นเป็นคลาสเดียวกันและข้อมูลใน โหนดนั้นมีค่าของแอทริบิวต์เดียวกัน

ข้อดีของต้นไม้ตัดสินใจ

1. เป็นเทคนิคที่ให้ผลเร็วเมื่อเทียบกับเทคนิคอื่น
2. ผลลัพธ์ที่ได้สามารถนำไปใช้งานและสามารถแปลงเป็นกฎได้
3. สามารถนำไปประยุกต์ใช้กับการค้นหาข้อมูล SQL
4. ให้ความแม่นยำสูง

ข้อจำกัดของต้นไม้ตัดสินใจ

1. การแบ่งกลุ่มแบบต้นไม้ตัดสินใจกรณีเป็นข้อมูลที่มีค่าต่อเนื่อง เช่น ข้อมูลรายได้ ข้อมูลราคา ต้องแปลงให้อยู่ในช่วงหรือตัดเป็นกลุ่มก่อน

2. เมื่ออัลกอริทึมเลือกจะใช้ค่าไหนเป็นตัวแบ่งกลุ่มแล้ว ก็จะไม่สนใจค่าอื่นที่อาจมีความสำคัญเช่นเดียวกัน
3. การจัดการกับข้อมูลที่ไม่ทราบค่า อาจมีผลกระทบกับผลลัพธ์ของต้นไม้ตัดสินใจ
4. ต้นไม้ที่มีระดับชั้นมากเกินไป จะทำให้ข้อมูลที่ผ่านโหนดแตกออกเป็นชิ้นเล็กชิ้นน้อย ซึ่งข้อมูลเหล่านั้นจะไม่มีประโยชน์ในการนำมาใช้ในการวิเคราะห์
5. ปัญหาเรื่อง Overfitting /Overtraining เกิดจากการที่แบบจำลองได้เรียนรู้เข้าไปถึงรายละเอียดของข้อมูลมากเกินไป จะทำให้เกิดโหนดที่เป็นส่วนเฉพาะเจาะจงกับกลุ่มข้อมูลที่ใช้ในการเรียนรู้ซึ่งจะต้องหาวิธีในการตัดกิ่งนี้ออกไป

เทคนิคการถดถอยโลจิสติก (logistic regression)

คือ การวิเคราะห์สมการถดถอยเมื่อ Y เป็นตัวแปรที่มีค่าเพียง 2 ค่า เรียกอีกอย่างหนึ่งว่า Binary Regression วัตถุประสงค์ของการนำเทคนิคการถดถอยโลจิสติกมาใช้วิเคราะห์ข้อมูล เพื่อทำนายว่าเหตุการณ์หนึ่งจะเกิดขึ้นได้หรือไม่ หรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด โดยกำหนดว่ามีตัวแปรตัวหนึ่งหรือหลายตัวที่ส่งผลต่อการเกิดเหตุการณ์นั้น ๆ ทำให้เราทราบเหตุผลของการเกิดเหตุการณ์นั้นหรือไม่เกิดเหตุการณ์นั้น ยกตัวอย่างเช่น ทำไมบางคนถึงเป็นโรคหัวใจ ในขณะที่คนอื่น ๆ ไม่เป็น ทำไมบางคนถึงไปออกเสียงเลือกตั้ง แต่บางคนไม่ยอมออกไป หรือทำไมธุรกิจบางประเภทถึงประสบความสำเร็จ ในขณะที่ธุรกิจบางประเภทล้มเหลว เป็นต้น เทคนิคการถดถอยโลจิสติก จะทำให้เราทราบได้ว่ามีสาเหตุใดที่บ่งชี้ความสำเร็จที่จะเกิดขึ้นของเหตุการณ์เหล่านี้ ค่าของตัวแปรตามที่ปรากฏเป็นข้อมูลสำหรับวิเคราะห์จะมีเพียง 2 ค่าเท่านั้น กล่าวคือ ใช่ (yes) หรือไม่ใช่ (no) ต่อไปจะแทนด้วยค่า 1 และ 0 ตามลำดับ ดังนั้น การอนุมานทางสถิติ จึงไม่ต้องมีข้อสมมติฐานว่าการแจกแจงความคลาดเคลื่อนมีการแจกแจงแบบปกติ เนื่องจากว่า ค่าของตัวแปรตามที่ถูกวัดเป็นค่า 1 และ 0 ไม่ใช่ค่าต่อเนื่อง แต่ค่าในที่สุดค่าที่ถูกทำนายจากสมการ Logistic Regression จะให้ค่าที่อยู่ระหว่าง 0 และ 1 หมายถึง ค่าของความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (อุไรวรรณ อมรนิมิตร, 2546)

สมการถดถอยโลจิสติกพหุ (multiple logistic regression)

ตัวแบบและวิธีประเมินค่า กรณีเมื่อมีตัวแปรอิสระมากกว่า 1 ตัว เรียกว่าสมการถดถอยโลจิสติกพหุ (มนตรี พิริยกุล, 2545)

$$\text{ให้ } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

$X'_i = [1, X_{1i}, X_{2i}, \dots, X_{pi}] ; i = 1, 2, \dots, n$ คือ ค่าสังเกตในแถวที่ i ของอะเรย์ X

$$\text{ดังนั้น } \beta'X_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_p X_{pi}$$

$$\text{ดังนั้น สมการถดถอยคือ } \begin{aligned} Y_i &= E(Y_i) + u \\ &= P_i + u_i \end{aligned}$$

ดังนั้น

$$\begin{aligned} L(\beta) &= \prod_i^n P_i^{Y_i} (1 - P_i)^{1 - Y_i} \\ &= \sum_i^n Y_i \ln\left(\frac{P_i}{1 - P_i}\right) + \sum \ln(1 - P_i) \end{aligned}$$

แต่

$$P_i = \frac{1}{1 + e^{-\beta'X_i}}$$

$$1 - P_i = \frac{1}{1 + e^{\beta'X_i}}$$

$$\frac{P_i}{1 - P_i} = e^{\beta'X_i} \text{ หรือ } \ln\left(\frac{P_i}{1 - P_i}\right) = \beta'X_i$$

ดังนั้น

$$L(\beta) = \sum Y_i (\beta'X_i) + \sum \ln(1 + e^{\beta'X_i})^{-1}$$

$$= \sum Y_i(\beta'X_i) - \sum \ln(1 + e^{\beta'X_i})$$

จากนี้ให้ทดลองแปรค่า β ไปเรื่อย ๆ เพื่อคำนวณค่า $\ln L(\beta)$ ค่าที่ β ที่เหมาะสมคือ $\hat{\beta}$ ที่มีผลให้ $\ln L(\hat{\beta})$ มีค่าสูงสุด

$$\text{ให้ } \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} \text{ คือ MLE ของ } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

สมการประมาณค่าคือ Fitted Logit Response Function คือ

$$\hat{P}_i = \frac{1}{1 + e^{-\hat{\beta}'X_i}}$$

$$\text{หรือ } E(Y_i) = \frac{1}{1 + e^{-\hat{\beta}'X_i}}$$

และพบว่า odds1 และ odds2 เมื่อ X_{Ai} คือ แถวที่ i ของ X สำหรับกลุ่มที่ A หรือหน่วยสำรวจ/ทดลองที่มีคุณสมบัติตาม A และ X_{Bi} คือ แถวที่ i ของ X สำหรับกลุ่มที่ B หรือหน่วยสำรวจ/ทดลองที่มีคุณสมบัติตาม B

$$\text{odds1} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1Ai} + \hat{\beta}_2 X_{2Ai} + \dots + \hat{\beta}_p X_{pAi}}$$

$$\text{odds2} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1Bi} + \hat{\beta}_2 X_{2Bi} + \dots + \hat{\beta}_p X_{pBi}}$$

$$\widehat{OR} = \frac{\text{odds1}}{\text{odds2}} = e^{\sum_{j=1}^p \hat{\beta}_j (X_{jAi} - X_{jBi})}; i = 1, 2, \dots, c$$

เมื่อ c คือ จำนวนกลุ่มของตัวแปรที่ประกอบกัน ตามค่าของ Exposure Variable และตัวแปรอื่น ๆ

การแปลผลด้วยค่า Odds Ratio นิยมใช้ค่า Odds Ratio (OR) หรือ $\exp(B)$ หรือ e^B ในการแปลผล ถ้าค่า Odds Ratio = 1 แสดงว่า การเปลี่ยนแปลงของ x ไม่มีผลต่อ y (และการทดสอบทางสถิติจะพบว่า ค่า 95% CI ครอบคลุม 1)

ถ้า Odds Ratio > 1 แสดงว่าเมื่อ x เพิ่มขึ้น ทำให้โอกาสของการเกิดเหตุการณ์เพิ่มขึ้น เมื่อ x เปลี่ยนไป 1 หน่วย โอกาสที่จะเกิดเหตุการณ์ที่สนใจจะเพิ่มขึ้น...เท่า เมื่อเทียบกับค่าเดิมของ x หรือ เพิ่มขึ้น ...% (ค่าสัมประสิทธิ์ถดถอย จะมีค่าเป็นบวก)

ถ้า Odds Ratio < 1 แสดงว่าเมื่อ x เพิ่มขึ้น ทำให้โอกาสของการเกิดเหตุการณ์ลดลง เปลี่ยนไป 1 หน่วย โอกาสที่จะเกิดเหตุการณ์ที่สนใจจะลดลง...เท่าเมื่อเทียบกับ

ค่าเดิมของ x หรือลดลง...% (ค่าสัมประสิทธิ์ถดถอย จะมีค่าเป็นลบ)

กรณีค่า $OR < 1$ นิยมอ่านผลเป็น ร้อยละการเปลี่ยนแปลงของค่า Odds จำนวน จาก $(Odds Ratio = 1) * 100$ (อุทัยทิพย์ เจียวิวรรธน์กุล, 2553)

เทคนิคโครงข่ายประสาทเทียม

วิธีการโครงข่ายประสาทเทียม เป็นวิธีการที่เลียนแบบการทำงานของสมองมนุษย์ซึ่งประกอบด้วยหน่วยประสาทหลายหน่วยที่เชื่อมต่อกันเป็นระบบประสาทขึ้นมา (เอกสิทธิ์ พัทธวงษ์ศักดิ์, 2557) โครงสร้างของโครงข่ายประสาทเทียม ประกอบด้วย โหนด (รูปวงกลม) และเส้นเชื่อมโยงแต่ละโหนด เมื่อเราพิจารณาการวางตัวของโหนดแล้วจะแบ่งออกได้เป็น 3 ชั้น คือ

ชั้นที่ 1 ข้อมูลนำเข้า (input layer) ประกอบด้วยโหนดซึ่งแสดงแอตริบิวต์ต่าง ๆ ในข้อมูลเรียนรู้ที่ใช้ในการสอนระบบ (จำนวนโหนดจะเท่ากับจำนวนแอตริบิวต์)

ชั้นที่ 2 ชั้นซ่อน (hidden layer) ประกอบด้วยโหนดต่าง ๆ ซึ่งจำนวนของโหนดในชั้นนี้ โดยทั่วไปแล้วจะเท่ากับจำนวนแอตริบิวต์ทั้งหมดหารด้วย 2

ชั้นที่ 3 ชั้นผลลัพธ์ (output layer) ประกอบด้วยโหนดที่เป็นคลาส ในกรณีที่ต้องการทำนายเป็นตัวเลขจะมีแค่โหนดเดียว

ในระหว่างชั้นจะเห็นว่า มีเส้นเชื่อมโยงระหว่างโหนดต่าง ๆ และมีค่าน้ำหนักปรากฏอยู่ด้วย การสร้างตัวแบบโครงข่ายประสาทเทียมอาศัยการปรับน้ำหนักระหว่างโหนดเหล่านี้จนทำให้ค่าที่ทำนายได้มีความคลาดเคลื่อน (error) น้อยที่สุด ซึ่งจะแสดงให้เห็นในส่วนตัวอย่างการสร้างตัวแบบโครงข่ายประสาทเทียม และแต่ละโหนดที่อยู่ในชั้นซ่อน (hidden layer) และชั้นผลลัพธ์ (output layer) จะประกอบด้วย 2 ส่วนดังแสดงในภาพ 2

ประเภทของโครงข่ายประสาทเทียม

จากแนวคิดของเซลล์ประสาทเทียมที่ต้องการมีการเชื่อมต่อและมีการคำนวณค่าแอกติเวชันฟังก์ชัน โดยมากจะเป็นตัวบ่งบอกโครงสร้างทางสถาปัตยกรรม ซึ่งจะมีกฎกำหนด

วิธีการของโครงข่ายโดยแบ่งออกเป็นประเภทได้ 5 ประเภท (Anderson & McNeill, 1992) คือ (1) การทำนาย (prediction) (2) การจัดลำดับหมวดหมู่ (classification) (3) การเชื่อมโยงข้อมูล (data association) (4) กระบวนการสร้างความคิด (data conceptualization) (5) การกลั่นกรองข้อมูล (data filtering) (ดูตาราง 1)

ตาราง 1

ประเภทของโครงข่ายประสาทเทียม

ชนิดโครงข่าย	โครงข่าย	การใช้โครงข่าย
	Back-propagation	
	Delta Bar Delta	
การทำนาย (prediction)	Extended Delta Bar Delta	ใช้ค่าอินพุตเพื่อ ทำนายเอาต์พุต
	Directed Random Search	
	Higher Order Neural Networks	
	Self-organizing map into Back-propagation	
การจัดหมวดหมู่ (classification)	Learning Vector Quantization	ใช้ค่าอินพุตเพื่อ กำหนดการจัด หมวดหมู่
	Counter-propagation	
	Probabilistic Neural Networks	

ตาราง 1 (ต่อ)

ชนิดโครงข่าย	โครงข่าย	การใช้โครงข่าย
	Hopfield	
การเชื่อมโยงข้อมูล (data association)	Boltzmann Machine	เหมือนกันกับการจัด หมวดหมู่ แต่จะจดจำ
	Hamming Network	ข้อมูลที่มีค่า
	Bidirectional associative Memory	กลาดเคลื่อนด้วย
กระบวนการสร้างความคิด (data conceptualization)	Adaptive Resonance Network	วิเคราะห์อินพุตเพื่อการ จัดกลุ่ม
	Self -Organizing Map	
การกลั่นกรองข้อมูล (data filtering)	Recirculation	ทำให้สัญญาณอินพุต เรียบสม่ำเสมอขึ้น

ที่มา. จาก *Artificial Neural Networks Technology: A DACS State of the Art Report*, by D. Anderson, and G. MaNeill, 1992, Retrieved October 6, 2016, from <https://knn.es/Artificial%20Neural%20Network%20Technologies.pdf>

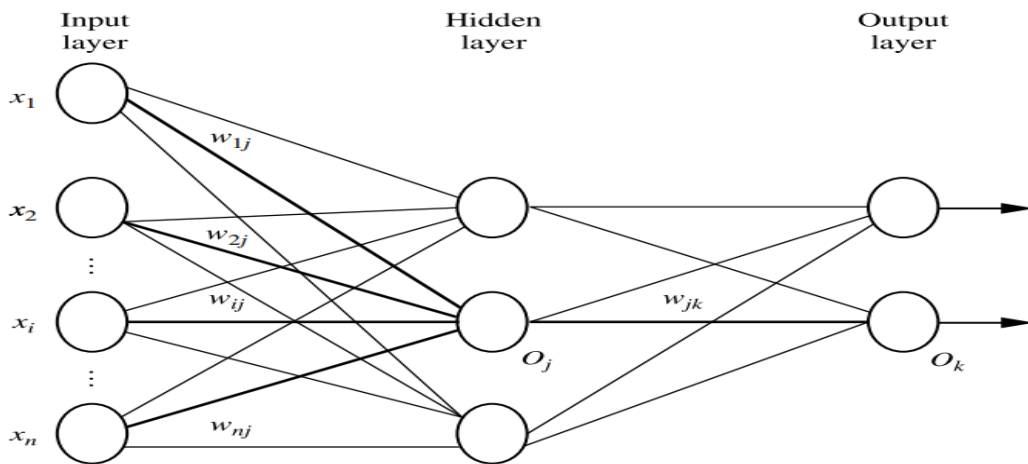
การจำแนกข้อมูลด้วยการส่งค่าย้อนกลับ (classification by backpropagation)

การส่งค่าย้อนกลับ เป็นอัลกอริทึมหนึ่งของโครงข่ายประสาทเทียม ซึ่งเป็นเซตของการเชื่อมต่อกันระหว่างโหนดของตัวนำเข้าและผลลัพธ์ โดยแต่ละการเชื่อมต่อกันจะมีค่าน้ำหนักเกี่ยวเนื่องด้วย เมื่อพิจารณาชุดข้อมูลเรียนรู้ เราจะสร้างเครือข่ายด้วยการปรับค่าน้ำหนักที่เชื่อมต่อกันระหว่างโหนดต่าง ๆ เพื่อจะทำนายหมวดหมู่ของข้อมูลได้อย่างถูกต้อง ในการปรับค่าน้ำหนักจะใช้เวลาค่อนข้างมากเพราะต้องปรับค่าน้ำหนักหลายครั้ง และในการสร้างเครือข่ายจะต้องให้ผู้ใช้งานกำหนดค่าพารามิเตอร์ต่าง ๆ เพื่อกำหนดโครงสร้างของเครือข่ายด้วย ข้อดีของโครงข่ายประสาทเทียม คือ (1) มีความคงทน

ต่อสิ่งรบกวนสูง (2) มีความสามารถในการจำแนกข้อมูลเมื่อยังไม่ได้นำข้อมูลมาฝึกสอน (3) สามารถใช้ในการจำแนกข้อมูลเมื่อทราบ หรือรับรู้ หรือมีความรู้เพียงเล็กน้อยเกี่ยวกับความสัมพันธ์ระหว่างแอทริบิวต์ต่าง ๆ กับหมวดหมู่ต่าง ๆ (4) สามารถทำงานได้ดีกับข้อมูลนำเข้าและผลลัพธ์ที่มีค่าแบบไม่ต่อเนื่อง (5) ถูกประยุกต์ใช้อย่างแพร่หลายในงานหลาย ๆ ด้าน เช่น การรู้จำลายมือ การวิเคราะห์ข้อมูลเกี่ยวกับพยาธิต่าง ๆ และการสอนให้คอมพิวเตอร์ออกเสียงภาษาอังกฤษ และ (6) สามารถนำไปพัฒนาแบบขนานได้โครงข่ายประสาทเทียม ได้ถูกพัฒนาอย่างต่อเนื่องทำให้อัลกอริทึมถูกคิดค้นขึ้นจำนวนมาก แต่อย่างไรก็ตาม มีอัลกอริทึมหนึ่งที่ได้รับคามนิยมเป็นอย่างมากและถูกประยุกต์ใช้อย่างแพร่หลาย นั่นคือ อัลกอริทึมการส่งค่าย้อนกลับ (backpropagation)

โครงข่ายประสาทเทียมแบบหลายลำดับชั้น (multilayer feed-forward neural network)

อัลกอริทึมการส่งค่าย้อนกลับจะเรียนรู้โดยใช้โครงข่ายประสาทเทียมแบบหลายลำดับชั้น ประกอบไปด้วย 1 ลำดับชั้นสำหรับข้อมูลนำเข้า (input layer) 1 ลำดับชั้น หรือมากกว่าสำหรับชั้นซ่อน (hidden layer) และ 1 ลำดับชั้นสำหรับข้อมูลผลลัพธ์ (output layer) ดังแสดงในภาพ 2 จากภาพเราจะสังเกตเห็นได้ว่า แต่ละลำดับชั้นจะประกอบไปด้วย โหนดต่าง ๆ โดยที่ข้อมูลนำเข้าสำหรับโครงข่ายจะสอดคล้องกับแอทริบิวต์ต่าง ๆ ของเรคคอร์ดหนึ่ง ๆ จากชุดข้อมูลเรียนรู้ ข้อมูลนำเข้าจะถูกป้อนเข้าไปยังโหนดต่าง ๆ ในชั้นข้อมูลนำเข้า จากนั้นจะผ่านชั้นนี้ไปยังชั้นซ่อนด้วยค่าน้ำหนักต่าง ๆ โดยที่ผลลัพธ์ของชั้นซ่อนลำดับชั้นหนึ่ง ๆ อาจจะเป็นข้อมูลนำเข้าของชั้นซ่อนชั้นหนึ่ง ๆ ก็ได้ โดยจำนวนลำดับชั้นของชั้นซ่อนจะมีเท่าไรก็ได้ เมื่อผ่านชั้นนี้ไปจะได้ค่าน้ำหนักที่จะใช้เป็นข้อมูลนำเข้าสำหรับชั้นผลลัพธ์ที่จะคืนค่าผลของการจำแนก หรือทำนายหมวดหมู่ของข้อมูลสำหรับเรคคอร์ดหนึ่ง ๆ ที่เป็นข้อมูลนำเข้าเมื่อเราพิจารณาที่โหนดต่าง ๆ ในชั้นผลลัพธ์ จะเห็นว่าแต่ละโหนดจะรวมค่าน้ำหนักทั้งหมดจากลำดับชั้นก่อนหน้า แล้วประยุกต์ใช้ฟังก์ชันกระตุ้นแบบไม่ใช่เส้นตรง (nonlinear (activation) function) เพื่อกำหนดค่าน้ำหนัก ดังแสดงในภาพ 2



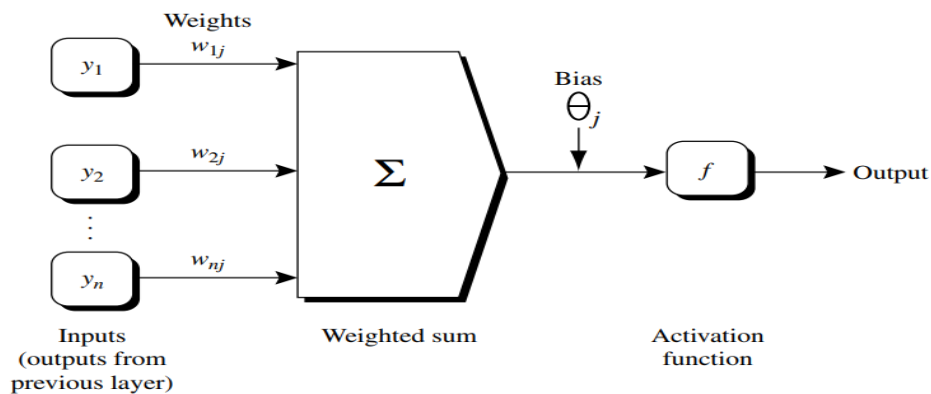
ภาพ 2 แสดงตัวอย่างโครงข่ายประสาทเทียม

ที่มา. จาก *DATA MINING: Concepts and Techniques. Third Edition. Chapter 9 Classification: Advanced Methods* (p. 399), by J. Han, M. Kamber and J. Pei, 2012, New York: Morgan Kaufmann.

การกำหนดโครงสร้างของโครงข่าย (defining a network topology)

ก่อนที่จะเรียนรู้โดยการประยุกต์ใช้โครงข่ายประสาทเทียม จะต้องกำหนดโครงสร้างของโครงข่ายประสาทเทียมก่อน โดยจะกำหนดโหนดในชั้นนำเข้า จำนวนโหนดในชั้นซ่อน และจำนวนโหนดในชั้นผลลัพธ์ ตามลำดับ โดยการกำหนดโหนดในชั้นนำเข้าแบ่งออกเป็น 2 กรณี ตามลักษณะของแอทริบิวต์ที่พิจารณา โดยถ้าเป็นแบบค่าต่อเนื่อง (มีค่าในเชิงตัวเลขที่บ่งบอกถึงปริมาณ) ควรนอร์มัลไลซ์ข้อมูลให้อยู่ในช่วง 0.0-1.0 เสียก่อน เพื่อที่จะช่วยให้กระบวนการเรียนรู้นั้นเป็นไปอย่างรวดเร็วและอาจกำหนดโหนดเพียงหนึ่งโหนดสำหรับแอทริบิวต์ที่มีค่าแบบต่อเนื่องเท่านั้น แต่สำหรับแอทริบิวต์ที่มีค่าแบบไม่ต่อเนื่อง (มีค่าแบบหมวดหมู่หรือมีค่าเป็นช่วง) อาจกำหนดให้แต่ละโหนดจะแทนค่าที่ปรากฏขึ้นในแอทริบิวต์นั้น เช่น แอทริบิวต์ A มีค่าแบบไม่ต่อเนื่อง จะมีค่าปรากฏขึ้น 3 ค่าด้วยกันคือ $\{a_0, a_1, a_2\}$ อาจกำหนดโหนด I_0, I_1 และ I_2 สำหรับค่าที่ปรากฏขึ้นทั้ง 3 ค่าโดยกำหนดให้โหนด I_0 จะเป็นโหนดสำหรับค่า a_0 โหนด I_1 จะเป็นโหนดสำหรับค่า a_1 และ โหนด I_2 จะเป็นโหนดสำหรับค่า a_2 ตามลำดับ โดยในขั้นตอนเริ่มต้นทุก ๆ โหนดจะถูกกำหนดให้มีค่าเป็น 0 ถ้า $A = a_0$ แล้ว I_0 คือ เซตถึง 1 และส่วนที่เหลือคือ 0 ถ้า $A = a_1$ แล้ว I_0 คือ เซตถึง 1 และส่วนที่เหลือ คือ 0 เป็นต้น โครงข่ายประสาท

เทียมนั้นสามารถใช้ในการทำนายหมวดหมู่ของข้อมูลและทำนายตัวเลข (ทำนายผลลัพธ์ค่าต่อเนื่อง) สำหรับการจำแนกประเภทข้อมูลผลลัพธ์หนึ่งผลลัพธ์ใช้เป็นตัวแทนของจำนวน 2 กลุ่มได้ (ในขณะที่ค่า 1 เป็นตัวแทนหนึ่งหมวดหมู่ และค่า 0 เป็นตัวแทนของอีกหมวดหมู่) ในส่วนของการกำหนดโหนดในชั้นซ่อนจะไม่มีกฎเกณฑ์ที่ตายตัวว่ากำหนดแบบไหนดีที่สุดที่สุด แต่จำนวนโหนดที่กำหนดอาจจะส่งผลกระทบต่อความถูกต้องในการเรียนรู้ของโครงข่ายประสาทเทียม นอกจากนี้ การกำหนดค่าน้ำหนักเริ่มต้นของแต่ละโหนดก็ส่งผลกระทบต่อความถูกต้องในการเรียนรู้ด้วยเช่นกัน เมื่อโครงข่ายประสาทเทียมได้รับการฝึกสอนค่าความถูกต้องของมันไม่ควรพิจารณายอมรับเลยควรมีการทำกระบวนการซ้ำ ๆ ด้วยโครงสร้างแบบอื่น ๆ ของโครงข่ายประสาทเทียมหรือการกำหนดค่าน้ำหนักเริ่มต้นที่แตกต่างด้วยเทคนิค Cross-validation นำมาใช้ในการประเมินประสิทธิภาพค่าความถูกต้อง สามารถช่วย ในการตัดสินใจว่าจะการยอมรับโครงข่ายที่พบได้หรือไม่ ถ้าไม่สามารถยอมรับได้ สามารถดำเนินการใหม่โดยการปรับเปลี่ยนเทคนิคอื่น ๆ เพื่อความเหมาะสมมากยิ่งขึ้น



ภาพ 3 การเชื่อมต่อระหว่างชั้นนำเข้า ชั้นซ่อน และชั้นผลลัพธ์

ที่มา. จาก *DATA MINING: Concepts and Techniques. Third Edition. Chapter 9*

Classification: Advanced Methods (p. 402) by J. Han, M. Kamber and J. Pei, 2012, New York: Morgan Kaufmann.

อัลกอริทึมการส่งค่าย้อนกลับ (backpropagation)

อัลกอริทึมการส่งค่าย้อนกลับจะเรียนรู้โดยใช้กระบวนการทำซ้ำกับชุดข้อมูลเรียนรู้ แล้วเปรียบเทียบการทำนายของข้อมูลแต่ละเรคคอร์ดด้วยค่าเป้าหมายตามความเป็นจริง ค่าเป้าหมายอาจจะทำให้ทราบหมวดหมู่ของชุดข้อมูลเรียนรู้หรือค่าต่อเนื่อง การพิจารณาสำหรับแต่ละชุดข้อมูลเรียนรู้นั้นจะทำให้ค่าน้ำหนักโหนดหนึ่ง ๆ มีค่าเปลี่ยนแปลงไปที่จะทำให้ค่าความผิดพลาดนั้นน้อยที่สุดเท่าที่จะเป็นไปได้ โดยเป็นไปในทิศทางย้อนกลับซึ่งสามารถอธิบายขั้นตอนได้ดังนี้

การทำงานของอัลกอริทึมการส่งค่าย้อนกลับจะเริ่มจาก การกำหนดค่าน้ำหนักเริ่มต้น (initialize the weight) กำหนดค่าน้ำหนักเริ่มต้นในโครงข่ายด้วยการสุ่มให้มีค่าน้อย ๆ (เช่นในช่วง -1.0 ถึง 1.0 หรือ -0.5 ถึง 0.5) และต้องกำหนดค่าความเอนเอียงของแต่ละโหนดด้วยเช่นกัน แต่ละเรคคอร์ด X คือ กระบวนการดังจะกล่าวต่อไปนี้

1. การพิจารณาตัวนำเข้าด้วยการพิจารณาไปข้างหน้า (propagate the inputs forward) เริ่มต้นด้วยการพิจารณาตัวนำเข้า โดยการเปรียบเทียบกับ โหนดในชั้นนำเข้า (input layer) เพื่อส่งต่อไปยังชั้นซ่อน (hidden layer) นั่นคือ สำหรับตัวนำเข้า j เราจะได้ผลลัพธ์เป็น O_j ที่มีค่าเท่ากับค่านำเข้าค่า I_j จากนั้น คำนวณค่าที่เป็นตัวนำเข้า และผลลัพธ์ของแต่ละโหนดทั้งหมดในชั้นซ่อนหรือในชั้นผลลัพธ์ ที่คำนวณผลรวมด้วยผลรวมเชิงเส้น ดังในภาพ 2 ที่แต่ละโหนดในชั้นซ่อนจะได้รับตัวนำเข้ามาจากโหนดในชั้นข้อมูลนำเข้า จากนั้นคำนวณค่านำเข้าของชั้นซ่อนด้วยการคูณแต่ละค่าผลลัพธ์ O_j ด้วยค่าน้ำหนักที่สอดคล้องกับโหนดนั้น ๆ แล้วรวมไว้

จากนั้นบวกด้วยค่าความเอนเอียง ซึ่งสามารถคำนวณได้ดังนี้

$$I_j = \sum_t w_{ij} O_j + \theta_j$$

เมื่อค่า w_{ij} คือ น้ำหนักของการเชื่อมต่อระหว่างโหนดที่ i ในลำดับชั้นที่กำลังพิจารณากับโหนดที่ j ในลำดับชั้นก่อนหน้า ค่า O_j คือ ค่าผลลัพธ์จากโหนดที่ i ในลำดับชั้นก่อนหน้า และค่า θ_j คือ ค่าความเอนเอียงที่ทำหน้าที่คล้ายค่าขีดแบ่ง (threshold) ที่จะทำหน้าที่เปลี่ยนแปลงกิจกรรมของโหนดที่พิจารณา แต่ละโหนดในชั้นซ่อนและในชั้นผลลัพธ์ (output layer) จะคำนวณค่านำเข้าแล้วประยุกต์ใช้ฟังก์ชันกระตุ้น (activation-

function) กับค่านำเข้าที่คำนวณไว้แล้วดังภาพ 3 แสดงด้วยสัญลักษณ์ f ที่เป็นการกระตุ้นการทำงานของเซลล์ประสาทในส่วนของโหนดนั้น ๆ เมื่อประยุกต์ใช้ฟังก์ชันกระตุ้นแล้วจะได้ผลลัพธ์ของโหนดที่พิจารณา j สามารถคำนวณได้ดังนี้

$$O_j = \frac{1}{1 + e^{-I_j}}$$

เมื่อคำนวณค่าผลลัพธ์สำหรับในแต่ละโหนดในชั้นซ่อนและชั้นผลลัพธ์ที่ให้ค่าการทำงานของโครงข่าย ในทางปฏิบัติ เป็นความคิดที่ดีในการจัดเก็บค่าผลลัพธ์กึ่งกลาง (intermediate output) สำหรับในแต่ละโหนดในภายหลังอีกครั้ง กรณีที่การส่งค่าย้อนกลับเกิดการผิดพลาด เคสึกลับนี้สามารถลดความต้องการในการคำนวณได้อย่างมากมาย

2. การส่งค่าความผิดพลาดย้อนกลับ (back propagate the error) ค่าความผิดพลาดจะถูกส่งย้อนกลับด้วยการปรับปรุ้ค่าน้ำหนัก และค่าความเอนเอียงให้เป็นปัจจุบันที่จะส่งผลต่อการทำนายของโครงข่ายประสาทเทียม สำหรับแต่ละโหนดที่ j ในชั้นผลลัพธ์จะคำนวณค่าความผิดพลาด Err_j ได้ดังนี้

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

เมื่อ O_j คือ ค่าผลลัพธ์จริงที่ได้จากโหนดที่ j และค่า T_j จะเป็นค่าเป้าหมายของเรคคอร์ดในชุดข้อมูลเรียนรู้ที่พิจารณา ในการที่จะคำนวณค่าความผิดพลาดของโหนดที่ j หนึ่ง ๆ ในชั้นซ่อน จะพิจารณาค่าน้ำหนักรวมของค่าความผิดพลาดของโหนดในลำดับชั้นถัดไปที่เชื่อมต่อกับโหนด ที่ค่าความผิดพลาดของโหนดที่ j จะสามารถคำนวณได้จาก

$$Err_j = O_j(1 - O_j) \sum_k Err_k W_{jk}$$

เมื่อ W_{jk} คือ ค่าน้ำหนักของการเชื่อมต่อระหว่างโหนดที่ j ในชั้นซ่อน ที่พิจารณากับโหนดที่ k ในลำดับชั้นถัดไป และค่า Err_k คือ ค่าความผิดพลาดของโหนด ที่ k ค่าน้ำหนักจะถูกปรับปรุ้ให้เป็นปัจจุบันเพื่อแพร่กระจายความผิดพลาด โดยจะถูกปรับปรุ้ดังนี้

$$\Delta W_{ij} = (l) Err_j O_i$$

$$W_{ij} = W_{ij} + \Delta W_{ij}$$

เมื่อ ΔW_{ij} คือ ค่าความเปลี่ยนแปลงที่จะเกิดขึ้นกับค่าน้ำหนัก W_{ij} และค่า l คือ ค่าอัตราการเรียนรู้ที่มีค่าคงที่อยู่ระหว่าง 0.0 ถึง 1.0 ค่าความเอนเอียงจะถูกปรับปรุงให้เป็นปัจจุบันเพื่อแพร่กระจายความผิดพลาดด้วยเช่นกัน โดยจะถูกปรับปรุงดังนี้

$$\Delta \theta_j = (l) \text{Err}_j$$

$$\theta_j = \theta_j + \Delta \theta_j$$

เมื่อ $\Delta \theta_j$ คือ ค่าความเปลี่ยนแปลงที่จะเกิดขึ้นกับค่าความเอนเอียง θ_j

3. การหยุดการทำงาน (terminating condition) กระบวนการเรียนรู้จะหยุดการทำงานก็ต่อเมื่อ

- 3.1 ทุก ๆ ค่า มีค่าน้อยมากซึ่งน้อยกว่าขีดแบ่งที่กำหนด หรือ
- 3.2 ค่าร้อยละในการจำแนกข้อมูลผิดพลาดมีค่าน้อยกว่าขีดแบ่งที่กำหนด หรือ
- 3.3 กระบวนการเรียนรู้ได้ถูกดำเนินการซ้ำจนครบตามจำนวนรอบที่กำหนด

ในทางปฏิบัติแล้ว เราจะกำหนดจำนวนรอบในการดำเนินงาน ซึ่งจะทำให้กระบวนการเรียนรู้หยุดการทำงานก่อนที่ค่าน้ำหนักต่าง ๆ จะมีค่าคงที่

จากกระบวนการทำงานทั้งหมด เราจะสังเกตได้ว่า ประสิทธิภาพของการคำนวณของอัลกอริทึมการส่งค่าย้อนกลับจะขึ้นอยู่กับเวลาที่ใช้ในการกระบวนการเรียนรู้ กำหนดให้ในชุดข้อมูลเรียนรู้ประกอบด้วย $|D|$ เรคคอร์ด และมีค่าน้ำหนักที่ต้องพิจารณาทั้งสิ้น w ค่า ดังนั้น เราจะสามารถคำนวณเวลาที่ใช้ในการเรียนรู้ในแต่ละรอบได้เป็น $O(|D| \times w)$ (Jiawei, Micheline, & Jian, 2012)

ข้อดีของวิธีโครงข่ายประสาทเทียม (हत्यरतनं केतुमणिशरतनं, 2014)

1. ทนทานต่อชุดข้อมูลสอนที่มีค่าผิดพลาด
2. มักจะให้ความแม่นยำในการทำนายสูง
3. ให้ผลลัพธ์เป็นเวกเตอร์คุณลักษณะค่าจริงหรือค่าไม่ต่อเนื่อง
4. ใช้ได้ดีกับปัญหาการรู้จำเสียง ลายมือ หรือรูปภาพ

ข้อด้อยของวิธีโครงข่ายประสาทเทียม

1. ใช้เวลาสอนหรือเรียนรู้นาน
2. ผลลัพธ์หรือค่าน้ำหนักตัวแปรฟังก์ชันที่ได้จากการเรียนรู้ค่อนข้างเข้าใจยาก

ตัววัดประสิทธิภาพของตัวแบบการจำแนกประเภทข้อมูล

การที่จะนำตัวแบบไปใช้งานจริงได้นั้นเราจำเป็นจะต้องทราบประสิทธิภาพของตัวแบบเสียก่อน (เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา, 2557) โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงานต่าง ๆ ได้แก่

Confusion Matrix

True Positive (TP) จำนวนที่ทำนายตรงกับข้อมูลจริงในคลาสที่กำลังพิจารณา หรือเรียกว่า ค่าความอ่อนไหว (Sensitivity) และ **True Negative (TN)** จำนวนที่ทำนายตรงกับข้อมูลจริงในคลาสที่ไม่ได้กำลังพิจารณา หรือเรียกว่า ค่าความจำเพาะ (Specificity)

False Positive (FP) จำนวนที่ทำนายผิดเป็นคลาสที่กำลังพิจารณา และ **False Negative (FN)** จำนวนที่ทำนายผิดเป็นคลาสที่ไม่ได้กำลังพิจารณา

ค่าความระมัดระวัง (precision)

คือ ค่าที่บอกว่าโปรแกรมทำนายว่าจริง ถูกต้องเท่าไร โดยพิจารณาแยกทีละคลาส

Recall หรือค่า Sensitivity

คือ ค่าที่บอกว่าโปรแกรมทำนายได้ว่าจริงเป็นอัตราส่วนเท่าไรของจริงทั้งหมด โดยพิจารณาแยกทีละคลาส

ค่าความจำเพาะ (specificity)

คือ ค่าที่บอกว่าโปรแกรมทำนายได้ว่าไม่จริง เป็นอัตราส่วนเท่าไรของจริงทั้งหมด

F-measure

เป็นการวัดค่า Precision และ Recall พร้อมกันของตัวแบบ โดยพิจารณาแยกทีละคลาส

ค่าความแม่นยำ (accuracy)

เป็นการวัดความถูกต้องของตัวแบบ โดยพิจารณารวมทุกคลาส

ROC Graph (Receiver Operating Characteristics)

แสดงกราฟความสัมพันธ์ระหว่างข้อมูลที่ทำนายถูก (แกน Y) และทำนายผิด (แกน X)

วิธีการคำนวณค่า Precision Recall/Sensitivity Specificity F-measure และ Accuracy

Precision ค่าที่บอกว่าโปรแกรมทำนายว่าจริง ถูกต้องเท่าไร โดยพิจารณาแยกทีละคลาส

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall/Sensitivity ค่าที่บอกว่าโปรแกรมทำนายได้ว่าจริงเป็นอัตราส่วนเท่าไรของจริงทั้งหมด โดยพิจารณาแยกทีละคลาส

$$\text{Recall/Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Specificity ค่าที่บอกว่าโปรแกรมทำนายได้ว่าไม่จริง เป็นอัตราส่วนเท่าไรของจริงทั้งหมด

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของตัวแบบ โดยพิจารณาแยกทีละคลาส

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Accuracy เป็นการวัดความถูกต้องของตัวแบบ โดยพิจารณารวมทุกคลาส คือ จำนวน True Positive ของทุกคลาสรวมกัน

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของตัวแบบการจำแนกประเภทข้อมูล

ในการวัดประสิทธิภาพของตัวแบบนั้น จำเป็นจะต้องแบ่งข้อมูลออกเป็น 2 ส่วน โดยในส่วนของที่ 1 ใช้เพื่อสร้างตัวแบบและส่วนของที่ 2 ให้ตัวแบบทำนายค่าคลาสคำตอบออกมา การแบ่งข้อมูลเพื่อทดสอบมี 3 วิธีการได้แก่วิธี Self Consistency Test วิธี Split Test และวิธี Cross-validation Test (เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา, 2557)

วิธี Self-consistency Test

เป็นวิธีการที่ง่ายที่สุด ข้อมูลที่ใช้ในการสร้างตัวแบบและข้อมูลที่ใช้ในการทดสอบตัวแบบเป็นข้อมูลชุดเดียวกัน กระบวนการนี้เริ่มจากสร้างตัวแบบด้วยข้อมูลชุดเรียนรู้ หลังจากนั้นนำตัวแบบที่สร้างได้มาทำนายข้อมูลชุดเรียนรู้ชุดเดิม การวัดประสิทธิภาพด้วยวิธีนี้จะให้ผลการวัดประสิทธิภาพที่มีค่าสูงมาก (อาจจะเข้าใกล้ร้อยละ 100) เนื่องจากเป็นข้อมูลชุดเดิมที่ระบบได้เรียนรู้มาแล้ว แต่ผลการวัดที่ได้ไม่เหมาะสม ที่จะนำไปรายงานในงานวิจัยต่าง ๆ ซึ่งวิธีการนี้เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของตัวแบบที่สร้างขึ้น ถ้าได้ผลการวัดที่น้อยแสดงว่าตัวแบบไม่เหมาะสมกับข้อมูล จึงไม่ควรนำไปทดสอบด้วยวิธีการแบ่งข้อมูลแบบต่าง ๆ

วิธี Split Test

เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น ร้อยละ 70 ต่อ 30 หรือร้อยละ 80 ต่อ 20 โดยข้อมูลส่วนที่หนึ่ง (ร้อยละ 70 หรือ ร้อยละ 80) ใช้ในการสร้างตัวแบบและข้อมูลส่วนที่สอง (ร้อยละ 30 หรือ ร้อยละ 20) ใช้ในการทดสอบประสิทธิภาพตัวแบบ แต่การทดสอบแบบ Split Test นี้สุ่มข้อมูลเพียงครั้งเดียวซึ่งในบางครั้งถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะคล้ายกับข้อมูลที่ใช้สร้างตัวแบบทำให้ผลการวัดประสิทธิภาพออกมาดี ในทางตรงกันข้าม ถ้าการสุ่มข้อมูลที่ใช้ในการทดสอบที่มีลักษณะแตกต่างจากข้อมูลที่ใช้สร้างตัวแบบมาก ทำให้ผลการวัดประสิทธิภาพได้ออกมาแย่ ดังนั้น จึงควรใช้วิธีนี้หรือสุ่มหลาย ๆ ครั้ง แต่ข้อดีของวิธีนี้ ก็คือใช้เวลาในการสร้างตัวแบบน้อยซึ่งเหมาะกับชุดข้อมูลที่มีขนาดใหญ่

วิธี Cross-validation Test

เป็นวิธีการวัดประสิทธิภาพในการทำนายตัวอย่างของตัวแบบโดยพื้นฐานของเทคนิคนี้ คือ การสุ่มตัวอย่าง (re-sampling) โดยเริ่มจากแบ่งชุดข้อมูลออกเป็น ส่วน ๆ หรือเรียกว่า fold และนำบางส่วนจากชุดข้อมูลนั้นมาทดสอบผลลัพธ์จากการทำนายข้อมูลทดสอบของตัวแบบ กรณีเลือกสุ่มข้อมูลแบบความเที่ยงตรง k กลุ่ม จะแบ่งข้อมูลออกเป็น k ชุดเท่า ๆ กันและคำนวณค่าความแม่นยำจากการทำนาย k รอบ โดยแต่ละรอบจะมีการสร้างตัวแบบจำแนกประเภทหนึ่งตัว จากข้อมูลเรียนรู้ $k-1$ ชุด และใช้ข้อมูลทดสอบ 1 ชุด (ชุดที่ไม่ได้นำมาเรียนรู้)

งานวิจัยที่เกี่ยวข้อง

งานวิจัยในประเทศ

บุปผา จันทจรรัส (2546) ศึกษาเรื่องการพัฒนาแบบประเมินภาวะเสี่ยงต่อการหกล้มในผู้ป่วยออร์โธปิดิกส์ ของโรงพยาบาลมหาราช จังหวัดเชียงใหม่ ที่มีอายุตั้งแต่ 15 ปี ขึ้นไป โดยมีแนวคิดการพัฒนามาจากกระบวนการนำผลการวิจัยมาใช้ตามรูปแบบของ สเตทเลอร์ (1994) การประเมินประกอบด้วย (1) อายุ (2) ระดับความรู้สึกรักตัว (3) ประวัติการพลัดตกหกล้ม (4) การเคลื่อนไหว และ (5) ความต้องการเกี่ยวกับการจับถ้ำย ซึ่งปัจจัยที่นำมาประเมินเป็นปัจจัยที่มีงานวิจัยหลายงานวิจัยให้การสนับสนุนว่า เป็นปัจจัยเสี่ยงต่อการพลัดตกหกล้มของผู้ป่วยในโรงพยาบาล มีค่าคะแนน 0-6 คะแนน และจุดตัดที่ 1 คะแนน ถือว่ามีความเสี่ยงต่อการหกล้ม และได้มีการนำเครื่องมือไปทดลองกับผู้ป่วยจำนวน 10 ราย ได้ค่าความเชื่อมั่นเท่ากับ 0.87 ซึ่งผลจากการนำไปทดลองโดยพยาบาลพบว่า แบบประเมินสะดวกต่อการใช้ ซึ่งใช้เวลาในการประเมินเพียง 1.02 นาที และสามารถประเมินได้อย่างต่อเนื่อง และบ่งชี้ถึงแนวทางการให้กิจกรรมการพยาบาลขั้นต่อไป

จันทร์ศิริ มีดี (2549) ศึกษาเรื่องปัจจัยเกี่ยวข้องต่อการหกล้มในผู้ป่วยโรคสมองเสื่อมตามการรับรู้ของพยาบาลในโรงพยาบาลสวนปรุง จังหวัดเชียงใหม่ โดยกลุ่มตัวอย่างเป็นพยาบาลวิชาชีพ และพยาบาลเทคนิคที่มีประสบการณ์ในการดูแลผู้ป่วยโรคสมองเสื่อม

อย่างน้อย 1 ปี จำนวน 119 คน ผลการศึกษาพบว่า มากกว่าครึ่งหนึ่งของพยาบาลรับรู้ว่ามีปัจจัยด้านพยาธิสภาพของผู้ป่วย ปัจจัยด้านการพยาบาล ปัจจัยด้านสิ่งแวดล้อม เป็นปัจจัยที่เกี่ยวข้องต่อการพลัดตกหกล้มในโรงพยาบาล และร้อยละ 63 รับรู้ว่าการผูกยึดผู้ป่วยไม่เกี่ยวข้องกับการพลัดตกหกล้ม

พิมวรินทร์ ลิ่มสุขสันต์ (2551) ศึกษาเรื่องปัจจัยทำนายการเกิดการพลัดตกหกล้มของผู้สูงอายุ ที่มีอายุตั้งแต่ 60 ปี ขึ้นไป ในอำเภอคลองหลวง จังหวัดปทุมธานี โดยศึกษาปัจจัยส่วนบุคคล ได้แก่ เพศ อาชีพ จำนวนสมาชิกในบ้าน ปัจจัยด้านกายภาพ ได้แก่ ดัชนีมวลกาย แรงบีบมือ พฤติกรรมออกกำลังกาย ปัจจัยด้านจิตใจ ได้แก่ ภาวะซึมเศร้า ภาวะกลัวการพลัดตกหกล้ม และปัจจัยด้านสิ่งแวดล้อม ได้แก่ การสวมรองเท้า ลักษณะรองเท้า ความสูงของส้นรองเท้า ความหนาของพื้นรองเท้า มีความสัมพันธ์กับการเกิดการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) และเมื่อวิเคราะห์การถดถอยแบบโลจิสติกแบบธรรมดาพบว่า ปัจจัยด้านการทำนายการเกิดการพลัดตกหกล้มของผู้สูงอายุอย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) ได้แก่ ลักษณะรองเท้า ความหนาของพื้นรองเท้า ภาวะกลัวการพลัดตกหกล้ม จำนวนสมาชิกในบ้าน และการสวมรองเท้า

วิภา หาทอน (2556) ศึกษาเรื่องการพัฒนาแนวปฏิบัติการพยาบาลเพื่อป้องกันและดูแลหลังหกล้มในผู้ป่วยสูงอายุที่มีความบกพร่องด้านการมองเห็น หอผู้ป่วยจักษุ โรงพยาบาลขอนแก่น โดยการพัฒนาแนวปฏิบัติการประเมินปัจจัยเสี่ยง การให้ความรู้ การจัดสิ่งแวดล้อม การเฝ้าระวัง การป้องกันตามปัจจัยเสี่ยงเฉพาะ และการดูแลเมื่อหกล้ม ผลการวิจัยพบว่า การพลัดตกหกล้มลดลง ผู้ป่วยสูงอายุและญาติพึงพอใจ พยาบาลสามารถปฏิบัติได้ทุกขั้นตอนและพึงพอใจต่อการใช้แนวทางการปฏิบัติ

อภิฤดี จิระวิโรจน์ (2556) ศึกษาเรื่องการศึกษาความเสี่ยงต่อการพลัดตกหกล้มในผู้ป่วยโรคหลอดเลือดสมองระยะฟื้นฟูสภาพ โดยใช้แบบประเมินความเสี่ยงต่อการหกล้ม (Fall Prediction Index) ของ Nyberg และ Gustafson ผลการศึกษาพบว่า กลุ่มตัวอย่างส่วนใหญ่ มีความเสี่ยงต่อการหกล้มในระดับปานกลางร้อยละ 56.60 รองลงมา มีความเสี่ยงต่อการหกล้มระดับต่ำร้อยละ 33.30 และส่วนหนึ่งมีความเสี่ยงต่อการหกล้มระดับสูงร้อยละ 10.10 และพบว่าปัจจัยที่มีความสำคัญต่อการหกล้ม ได้แก่ ความสามารถในการทำกิจวัตรประจำวัน ความสามารถในการทรงตัว อาการอ่อนแรงของกล้ามเนื้อต่าง ๆ และ

ภาวะกลืนปัสสาวะ โดยพบว่า การกลืนปัสสาวะเป็นปัจจัยสำคัญที่ทำให้ระดับความเสี่ยงต่อการหกล้มในกลุ่มตัวอย่างแตกต่างกัน ส่วนปัจจัยที่มีผลกระทบน้อยคือ การมองเห็นเพียงครึ่งซีกและการใช้ยาต่าง ๆ

ปริญดา พิมพ์ภูลาด (2556) ศึกษาเรื่องผลการนำแบบประเมินปัจจัยเสี่ยงต่อการหกล้มไปทดลองใช้กับผู้สูงอายุ หอผู้ป่วยเวชกรรมฟื้นฟู โรงพยาบาลอุดรธานี โดยศึกษาในผู้สูงอายุที่เข้ารับการรักษาในหอผู้ป่วยเวชกรรมฟื้นฟู โรงพยาบาลอุดรธานี ตั้งแต่เดือนธันวาคม พ.ศ. 2555 ถึงเดือนกุมภาพันธ์ พ.ศ. 2556 ผลการศึกษาพบว่า แบบประเมินปัจจัยเสี่ยงต่อการหกล้มที่พัฒนาขึ้นไปทดลองใช้กับผู้สูงอายุกลุ่มตัวอย่าง 20 คน พบกลุ่มตัวอย่างที่มีความเสี่ยงต่อการหกล้ม จำนวน 10 คน กลุ่มที่มีความเสี่ยงต่ำ จำนวน 8 คน ส่วนกลุ่มที่มีความเสี่ยงสูง จำนวน 2 คน และพบว่าเฉพาะกลุ่มเสี่ยงสูงมีการหกล้ม 1 ครั้ง หลังจากจำหน่ายออกจากโรงพยาบาล โดยมีปัจจัยเสี่ยงที่ทำให้เกิดการพลัดตกหกล้ม ได้แก่ มีประวัติการหกล้ม การใช้ยามากกว่า 4 ชนิด ได้รับการวินิจฉัยภาวะการเจ็บป่วยในครั้งที่ 2

พัชรินทร์ คณะพล (2557) ศึกษาเรื่องความชุกและปัจจัยทำนายการเกิดการหกล้มของผู้ป่วยโรคหลอดเลือดสมองหลังการจำหน่ายจากโรงพยาบาล โดยศึกษาผู้ป่วยโรคหลอดเลือดสมองที่มารับการตรวจติดตามผลการรักษาหลังจำหน่ายที่แผนกผู้ป่วยนอกโรงพยาบาลศรีนครินทร์ คณะแพทยศาสตร์มหาวิทาลัยขอนแก่น ตั้งแต่วันที่ 1 มีนาคม พ.ศ. 2556-31 พฤษภาคม พ.ศ. 2556 จำนวน 205 คน โดยใช้แบบทดสอบ Thai Mini-Mental State Examination (TMSE) แบบประเมิน Modified Barthel Index (MBI) แบบทดสอบ Berg Balance Test (BBT) แบบประเมิน Falls Efficacy Scale, Swedish version: FES(S) แบบวัดภาวะซึมเศร้า The Center for Epidemiologic Studies-Depression scale (CES-D) และแบบทดสอบ Line Bisection Test (LBT) ผลการวิจัยพบว่า มีการประมาณค่าความชุกของการหกล้มของกลุ่มตัวอย่าง จำนวน 70 คน คิดเป็นร้อยละ 34.15 (95%CI = 27.60-40.69) ปัจจัยที่มีความสัมพันธ์กับการเกิดการหกล้มอย่างมีนัยสำคัญทางสถิติ ปัจจัยภายในบุคคลที่เกิดจากพยาธิสภาพของโรคหลอดเลือดสมอง ได้แก่ ความบกพร่องในการทรงตัว การละเลยร่างกายครึ่งซีก การอ่อนแรงของกล้ามเนื้อ ความบกพร่องในการดูแลตนเอง และความรู้คิดบกพร่อง มีความสัมพันธ์กับการเกิดการ

หกล้ม ปัจจัยส่วนบุคคล ได้แก่ ประวัติการเคยหกล้ม อาการข้างเคียงจากยา ภาวะกัวการหกล้ม และภาวะซึมเศร้า ปัจจัยภายนอกบุคคล ได้แก่ สภาพแวดล้อมที่มีสิ่งกีดขวาง และพื้นที่ลื่น ส่วนปัจจัยที่ไม่มีความสัมพันธ์กับการเกิดการหกล้ม ได้แก่ อายุ เพศ และตำแหน่งพยาธิสภาพของโรคหลอดเลือดทางสมอง สภาพแวดล้อมที่มีแสงสว่างไม่เพียงพอ ผู้ดูแล ตัวแปรด้านความรู้คิดบกพร่อง อาการข้างเคียงจากยา ภาวะกัวการหกล้ม ภาวะซึมเศร้า สภาพแวดล้อมที่มีสิ่งกีดขวาง มีความสัมพันธ์ต่อการเกิดการพลัดตกหกล้ม และสามารถทำนายการเกิดการหกล้มของกลุ่มตัวอย่างได้ร้อยละ 44.60

งานวิจัยในต่างประเทศ

Marschollek, et al., (2012) วิจัยเรื่อง Mining geriatric assessment data for in-patient fall prediction models and high-risk subgroups ศึกษาในกลุ่มผู้ป่วยในผู้สูงอายุ จาก Evangelisches Geriatriezentrum Berlin gGmbH (EGZB) แผนกอายุรกรรมผู้สูงอายุ ของมหาวิทยาลัยแพทย์ Charite' ในกรุงเบอร์ลิน ประเทศเยอรมนี ข้อมูลของเดือนกรกฎาคม ค.ศ. 2006 ถึงเดือนธันวาคม ค.ศ. 2007 จำนวน 5,176 ราย โดยมีวัตถุประสงค์เพื่อจำแนกรูปแบบของปัจจัยเสี่ยงที่ได้จากการประเมินเซตข้อมูลขนาดใหญ่ของกลุ่มผู้สูงอายุที่เป็นผู้ป่วยใน และประเมินประสิทธิภาพของตัวทำนายที่ใช้ และเพื่อที่จะระบุกลุ่มย่อยของปัจจัยที่มีความเสี่ยงสูงจากข้อมูลเหล่านี้ โดยใช้ตัวแบบของต้นไม้ตัดสินใจ C4.5 และ Logistic Regression ผลการศึกษาพบว่า ตัวแบบของต้นไม้ตัดสินใจมีความถูกต้องอยู่ที่ร้อยละ 66 และมีค่าความไวอยู่ที่ร้อยละ 55.4 มีค่าความจำเพาะอยู่ที่ร้อยละ 67.1 และพบตัวปัจจัยที่มีความเสี่ยงสูง 5 ปัจจัยได้แก่ (1) อายุที่มากขึ้น (high age) (2) ดัชนีความสามารถในการดำเนินชีวิตประจำวันต่ำ (low barthel index) (3) ภาวะสมองเสื่อม (cognitive impairment) (4) การได้รับยาหลายชนิด (multi-medication) (5) การมีโรคแทรกซ้อน (co-morbidity)

Lee, et al., (2010) วิจัยเรื่อง Application of data mining to the identification of critical factors in patient falls using a web-based reporting system ศึกษาในผู้ป่วยที่พลัดตกหกล้มจากระบบรายงานอุบัติการณ์ที่ศูนย์การแพทย์ของประเทศไต้หวัน จำนวนทั้งหมด

725 ราย มีวัตถุประสงค์เพื่อหาปัจจัยที่มีความเสี่ยงที่มีความสัมพันธ์กับการพลัดตกหกล้มของผู้ป่วย ศึกษาความเป็นไปได้ของโปรแกรมค้ำไม้นิ่งที่สามารถใช้ได้กับข้อมูลในระบบฐานข้อมูลของโรงพยาบาล โดยศึกษาในกระบวนการของค้ำไม้นิ่ง ใช้ตัวคัดเลือกปัจจัย (feature selection) ในขั้นตอนแรก หลังจากนั้นเลือกใช้ 10 ปัจจัยทำนายที่สำคัญเพื่อทำนายตัวแปรตาม (dependent variables) ได้แก่ ได้รับความเจ็บกับไม่ได้รับความเจ็บ (injury versus non-injury) ใช้เทคนิคโครงข่ายประสาทเทียม (ANN) มาสร้างตัวแบบทำนายและวิธีการถดถอยโลจิสติกแบบคัดเลือกลำดับขั้น (multivariate stepwise logistic regression) มาเปรียบเทียบกัน ผลการศึกษาพบว่า ตัวแบบ ANN มีค่าพื้นที่ใต้โค้ง (ROC) อยู่ที่ร้อยละ 77 ค่าความถูกต้อง (accuracy) ค่าทำนายที่เป็นผลบวก (positive predictive value: PPV) อยู่ที่ร้อยละ 68 และค่าทำนายที่เป็นลบ (negative predictive value: NPV) อยู่ที่ร้อยละ 72 ในขณะที่ผลของวิธีถดถอย โลจิสติกแบบคัดเลือกลำดับขั้นระบุเพียง 3 ตัวแปร ได้แก่ การประเมินการพลัดตกหกล้ม ยาด้านการซึมเศร้า และยาขับปัสสาวะ มีค่าตัวทำนายที่สำคัญด้วยค่า ROC ที่ร้อยละ 42 ค่า PPV ที่ร้อยละ 26.24 และค่า NPV ที่ร้อยละ 87.12

Evans, et al., (1998) วิจัยเรื่อง Fall in acute hospital a systematic review พบว่างานวิจัยส่วนใหญ่ระบุว่า มีปัจจัยที่มีความสัมพันธ์ต่อการพลัดตกหกล้มในโรงพยาบาล ได้แก่ อายุ สภาพจิตใจ มีประวัติพลัดตกหกล้ม การได้รับยา ความต้องการในการเข้าห้องน้ำ ความสามารถในการเคลื่อนไหว บริเวณที่พบว่าการพลัดตกหกล้มมากที่สุด คือ ช้างเตียงผู้ป่วย และลักษณะกิจกรรมที่ทำให้เกิดการพลัดตกหกล้มบ่อยที่สุด คือ ขณะเคลื่อนย้ายจากเตียงหรือเก้าอี้ การประเมินคัดกรองผู้ป่วยที่มีความเสี่ยงต่อการพลัดตกหกล้มตั้งแต่แรกเริ่ม ทำให้ประเมินปัญหา และวางแผนทางป้องกันที่รวดเร็ว และพบว่าเครื่องมือที่ใช้ในการประเมินความเสี่ยงต่อการพลัดตกหกล้มส่วนใหญ่ ยังไม่มีการศึกษาผลการนำมาใช้ในทางปฏิบัติที่ชัดเจน

Richardson (2002) วิจัยเรื่อง Factors associated with falls in older patients with diffuse polyneuropathy ศึกษาในผู้ป่วยอายุตั้งแต่ 50 ถึง 85 ปี ที่มีอาการแสดงของความผิดปกติของเส้นประสาทส่วนปลายและผลการตรวจที่ยืนยันภาวะโรคจากการตรวจกล้ามเนื้อและเส้นประสาทด้วยไฟฟ้า พบว่า ปัจจัยที่เป็นสาเหตุที่สัมพันธ์กับการพลัดตกหกล้ม

ของผู้สูงอายุที่มีความผิดปกติของระบบประสาทส่วนปลาย คือ อุบัติเหตุที่เกี่ยวข้องกับ สิ่งแวดล้อม การอ่อนแรงของกล้ามเนื้อ การทรงตัว อาการวิงเวียน หน้ามืดเป็นลม ความดันโลหิตต่ำ ระดับความรู้สึกตัว และการมองเห็นที่ผิดปกติ และได้สรุปว่าการป้องกันการ พัดตกหกล้มที่มีประสิทธิภาพ องค์ประกอบที่สำคัญ คือ การประเมินปัจจัยเสี่ยงต่อการ พัดตกหกล้ม การลดปัจจัยเสี่ยงเฉพาะของผู้ป่วยแต่ละราย โปรแกรมการออกกำลังกาย การจัดการสิ่งแวดล้อม และการให้ความรู้ ทำให้ลดการเกิดการพัดตกหกล้มได้อย่างมี นัยสำคัญทางสถิติ ($p < 0.05$)

Hitcho, et al., (2004) วิจัยเรื่อง Characteristics and circumstances of falls in a hospital setting: a prospective analysis โดยศึกษาผู้ป่วยในที่เกิดการพัดตกหกล้ม จำนวน 183 ราย ในโรงเรียนแพทย์ขนาด 1,300 เตียง พบว่าแผนกที่มีการพัดตกหกล้ม มากที่สุด คือ เวรดิค ร้อยละ 59 และปัจจัยด้านผู้ป่วยที่มีปัญหาการขยับถ่าย อายุมากกว่า 65 ปี และอัตราส่วนพยาบาลกับผู้ป่วยพบว่า มีความสัมพันธ์กับการเกิดการพัดตกหกล้ม อย่างมีนัยสำคัญทางสถิติ ($p < 0.05$)

Shah, et al., (2012) วิจัยเรื่อง A retrospective study of fall risk factors ศึกษาปัจจัย ที่เกี่ยวกับการเกิดการพัดตกหกล้มในโรงพยาบาลชุมชนขนาด 40 เตียง โดยการศึกษา ข้อมูลย้อนหลังในผู้ป่วยที่เกิดพัดตกหกล้มขณะเข้ารับการรักษาในโรงพยาบาล เดือน เมษายนถึงเดือนกันยายน ค.ศ. 2011 จำนวน 78 ราย พบว่ามีผู้ป่วย 4 ราย ที่เกิดการพัดตก หกล้ม ผู้ป่วยร้อยละ 40 อายุ 85-95 ปี และร้อยละ 22 อายุ 75-85 ปี ร้อยละ 78 มีประวัติการ พัดตกหกล้ม เพศที่เกิดพัดตกหกล้มทั้งหญิงและชายไม่มีความแตกต่างกัน ด้านการ เคลื่อนไหวพบว่า ร้อยละ 70 มีการใช้อุปกรณ์ช่วยเดินขณะอยู่บ้าน ห้องผู้ป่วยเป็นจุดที่พบ ว่ามีการพัดตกหกล้มมากที่สุดร้อยละ 61 และในห้องอาบน้ำร้อยละ 21 ที่น่าสังเกตจาก การประเมิน Morse Fall Risk Assessment พบว่าผู้ป่วยร้อยละ 57 ถูกระบุว่ามีความเสี่ยง ระดับสูงต่อการเกิดพัดตกหกล้มเมื่อแรกรับ

จากการทบทวนวรรณกรรมที่เกี่ยวข้องจะเห็นได้ว่า การศึกษาปัจจัยที่เกี่ยวข้องกับ การพัดตกหกล้มในโรงพยาบาลของผู้ป่วยสรุปได้ว่า

1. ปัจจัยส่วนบุคคล ประกอบด้วย อายุ โดยเฉพาะผู้ป่วยสูงอายุ เพศ ประวัติพัดตก หกล้ม

2. ปัจจัยภายใน ประกอบด้วย สภาพจิตใจและอารมณ์ อาการสับสน/มึนงง อาการวิงเวียนบ้านหมุน (vertigo) ความถี่ในการขับถ่ายอุจจาระ/ปัสสาวะที่มากขึ้น ทำให้ผู้ป่วยต้องลุกเข้าห้องน้ำบ่อย ๆ (Lahmann, et al., 2015) โดยเฉพาะในผู้ป่วยที่มีความสามารถในการเคลื่อนไหว (mobility) และการเคลื่อนย้าย (transfer) ที่ลดลง ส่งผลให้ผู้ป่วยเกิดการสะดุดหรือหกล้มได้ง่าย (อภิฤดี จิระวิโรจน์, 2556)

3. ปัจจัยภายนอกประกอบด้วย การมีผู้ดูแลหรือญาติเฝ้าประจำช่วยลดอุบัติเหตุการพลัดตกหกล้มของผู้ป่วยได้ (Krauss, et. al., 2005)

4. ปัจจัยด้านการรักษาพยาบาล ประกอบด้วย การผ่าตัด ผู้ป่วยที่ได้รับการผ่าตัดจะได้รับยาระงับความรู้สึก และการได้รับยา Sedative, Benzodiazepines, PCA/Opiates, Hypnotics ส่งผลให้เกิดอาการง่วงซึมทำให้เสี่ยงต่อการพลัดตกหกล้มได้ การคาสายอุปกรณ์ทางการแพทย์ (medical device) ทำให้ผู้ป่วยเคลื่อนย้ายหรือเดินไม่สะดวก อาจเกิดการสะดุด หรือหกล้มได้ง่าย (Morse, 2009) ระยะเวลานอนโรงพยาบาลที่เกี่ยวกับการพลัดตกหกล้ม เนื่องจากผู้ป่วยที่นอนโรงพยาบาลนาน ๆ ส่วนใหญ่เกิดจากภาวะการเจ็บป่วยและแผนการรักษา ทำให้ผู้ป่วยเกิดความเหนื่อยล้า ส่งผลให้มีความเสี่ยงต่อการพลัดตกหกล้ม (Evans, et al., 1998)

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ แนวทางการดำเนินการวิจัย ประกอบด้วยหัวข้อ ดังนี้

1. รูปแบบงานวิจัย
2. ข้อมูลที่ใช้ในการศึกษา
3. เทคนิคที่ใช้ในการวิเคราะห์
4. ลำดับของงาน
5. การรวบรวมข้อมูล
6. การเตรียมข้อมูล
7. การค้นหาตัวแบบ
8. การทดสอบประสิทธิภาพตัวแบบทำนาย

รูปแบบงานวิจัย

เป็นการวิจัยเชิงปริมาณ โดยเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกประเภทข้อมูล ด้วยวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก เพื่อหาตัวแบบทำนายที่เหมาะสมในการทำนายโอกาสการเกิดพลัดตกหกล้มของผู้ป่วยที่เข้ารับบริการเป็นผู้ป่วยในของโรงพยาบาล

ข้อมูลที่ใช้ในการศึกษา

โดยการศึกษาข้อมูลย้อนหลัง (retrospective study) โดยวิเคราะห์ข้อมูลจากแฟ้มประวัติผู้ป่วย รายงานอุบัติการณ์ และข้อมูลทบทวนเหตุการณ์การพลัดตกหกล้มของผู้ป่วยระหว่างที่เข้ารับการรักษาในโรงพยาบาล ในช่วงเดือนมกราคม พ.ศ. 2557 ถึง เดือนธันวาคม พ.ศ. 2559 รวมทั้งสิ้นจำนวน 400 คน แบ่งข้อมูลด้วยวิธี Split Test ออกเป็น 2

ส่วน ข้อมูลส่วนที่หนึ่งร้อยละ 70 ใช้ในการสร้างตัวแบบคิดเป็นชุดข้อมูลจำนวน 280 คน ใช้วิธีการตรวจสอบแบบไขว้ (k-fold cross-validation) โดยกำหนดให้ $k = 10$ เพื่อทดสอบประสิทธิภาพของตัวแบบ และข้อมูลส่วนที่สองร้อยละ 30 คิดเป็นชุดข้อมูลจำนวน 120 คน ใช้เป็นข้อมูลทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน ที่เข้ารับการรักษาพยาบาล

เทคนิคที่ใช้ในการวิเคราะห์

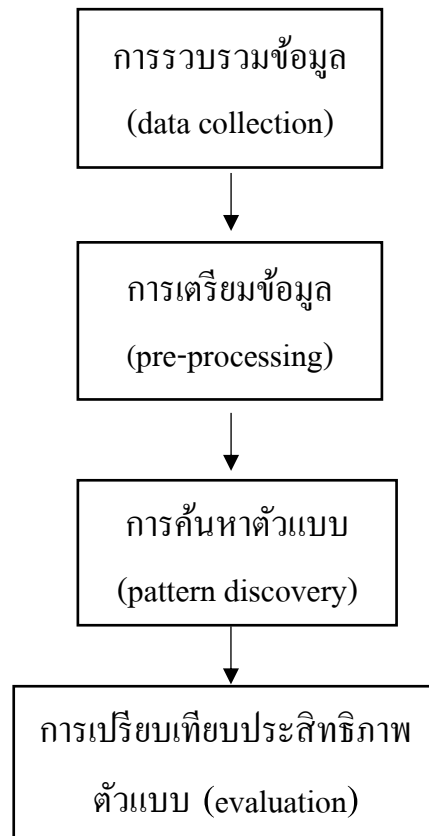
1. เทคนิควิธีถดถอยโลจิสติก โดยจะทดสอบชุดข้อมูลเรียนรู้ ใช้พารามิเตอร์

Solver แบบ Auto

2. เทคนิควิธีต้นไม้ตัดสินใจ โดยจะทดสอบชุดข้อมูลเรียนรู้ใช้พารามิเตอร์ตัวชี้วัดการแบ่งข้อมูลด้วยค่าอัตราส่วนเกน (gain ratio) Maximal Depth 40 และค่าความเชื่อมั่น (confidence) เท่ากับ 0.25

3. เทคนิคโครงข่ายประสาทเทียมเชิงลึก โดยจะทดสอบชุดข้อมูลเรียนรู้ ใช้พารามิเตอร์ฟังก์ชันกระตุ้นแบบตัวเรียงกระแสไฟฟ้า (rectifier)

ลำดับของงาน (proposed framework)



ภาพ 4 แผนภูมิลำดับของการทำงาน

การรวบรวมข้อมูล (data collection)

ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ประชากรที่ศึกษา คือ ผู้ป่วยที่มีอายุมากกว่า 15 ปี ขึ้นไป ที่เข้ารับการรักษาเป็นผู้ป่วยในในโรงพยาบาลเอกชนแห่งหนึ่ง ระหว่างเดือนมกราคม พ.ศ. 2557 ถึงเดือนธันวาคม พ.ศ. 2559 จำนวน 66,000 คน (เฉลี่ยประมาณ 22,000 คนต่อปี)

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ ผู้ป่วยในในโรงพยาบาลเอกชนแห่งหนึ่งที่มีประวัติพลัดตกหกล้มและไม่มีประวัติพลัดตกหกล้ม ที่เข้ารับการรักษาเป็นผู้ป่วยในระหว่างเดือนมกราคม พ.ศ. 2557 ถึงเดือนธันวาคม พ.ศ. 2559

กำหนดขนาดกลุ่มตัวอย่างในการเทียบจากตารางสำเร็จรูปของ ทาโร่ ยามาเน่ (Yamane, อ้างถึงใน บุญใจ ศรีสถิตยัณรากร, 2550) กำหนดระดับความเชื่อมั่นร้อยละ 95 และระดับความคลาดเคลื่อนของการสุ่มร้อยละ 0.05 จากจำนวนประชากร 66,000 คน ได้ขนาดกลุ่มตัวอย่างเท่ากับ 397 คน ผู้วิจัยเลือกกลุ่มตัวอย่างมาจำนวน 400 คน

No.	Fall case	HN	Date	Admission Date	Discharge date	LOS(day)	Age-1	Male Gender	Nationality	Ethnic	Airport to admission	Time Zone	Market segment	Location	Room location (nearly nurse counter)	Language
1	0	0130907	29-01-2014	29-Jan-14	31-Jan-14	NA	7	1	39	1	NA	NA	1	14	0	1
2	0	0072003	10-02-2014	10-Feb-14	11-Feb-14	NA	6	0	39	1	NA	NA	1	14	0	1
3	0	0230944	15-03-2014	15-Mar-14	16-Mar-14	NA	2	0	1	2	5	5	3	13	1	8
4	0	0151046	18-04-2014	18-Apr-14	19-Apr-14	NA	5	0	39	1	NA	NA	1	20	1	1
5	0	0031480	01-05-2014	1-May-14	20-May-14	NA	8	0	39	1	NA	NA	1	9	0	1
6	0	0047568	09-07-2014	9-Jul-14	14-Jul-14	NA	5	1	39	1	NA	NA	1	18	0	1
7	0	0006778	12-08-2014	12-Aug-14	16-Aug-14	NA	3	1	39	1	NA	NA	1	19	0	1
8	0	0236793	14-09-2014	14-Sep-14	18-Sep-14	NA	4	0	16	3	1	2	3	22	1	3
9	0	0039819	12-11-2014	12-Nov-14	21-Nov-14	NA	7	0	39	1	NA	NA	1	2	1	1
10	0	0082897	13-01-2015	13-Jan-15	15-Jan-15	NA	9	0	39	1	NA	NA	1	14	0	1
11	0	0241185	13-02-2015	13-Feb-15	20-Feb-15	NA	4	0	31	3	3	1	3	11	0	3
12	0	0231132	12-03-2015	12-Mar-15	14-Mar-15	NA	5	1	8	1	NA	1	3	5	0	13
13	0	0210336	17-03-2015	17-Mar-15	18-Mar-15	NA	5	1	39	1	NA	NA	1	19	0	1
14	0	0194843	11-04-2015	11-Apr-15	12-Apr-15	NA	4	0	31	3	3	1	3	20	1	3
15	0	0169167	19-04-2015	19-Apr-15	1-May-15	NA	7	1	39	1	NA	NA	1	19	0	1
16	0	1E+08	14-05-2015	14-May-15	17-May-15	NA	4	0	22	3	4	2	3	20	1	3
17	0	0141462	22-06-2015	22-Jun-15	25-Jun-15	NA	3	0	39	1	NA	NA	1	21	0	1
18	0	0237928	23-07-2015	23-Jul-15	27-Jul-15	NA	8	0	16	3	1	2	3	10	0	3
19	0	0160234	29-08-2015	29-Aug-15	1-Sep-15	NA	6	1	8	1	2	1	3	8	0	13
20	0	0038069	23-09-2015	23-Sep-15	24-Sep-15	NA	7	1	39	1	NA	NA	1	15	0	1
21	0	0241378	27-10-2015	27-Oct-15	31-Oct-16	NA	6	0	39	1	NA	NA	1	12	0	1
22	0	0071120	07-12-2015	7-Dec-15	6-Jan-16	NA	8	1	39	1	NA	NA	1	14	0	1
23	0	0101828	31-12-2015	31-Dec-15	1-Jan-16	NA	6	1	39	1	NA	NA	1	20	0	1
24	0	0181172	22-01-2016	22-Jan-16	23-Jan-16	NA	7	1	39	1	NA	NA	1	9	0	1
25	0	0056680	26-02-2016	26-Feb-16	4-Mar-16	NA	4	0	39	1	NA	NA	1	7	1	1
26	0	0203528	12-04-2016	12-Apr-16	13-Apr-16	NA	7	0	39	1	NA	NA	1	14	0	1
27	0	0107410	19-05-2016	19-May-16	20-May-16	NA	4	1	39	1	NA	NA	1	19	1	1
28	0	0249761	07-06-2016	7-Jun-16	8-Jun-16	NA	2	0	2	2	1	1	3	9	0	8
29	0	0258410	06-07-2016	6-Jul-16	11-Jul-16	NA	6	1	8	1	1	1	3	19	0	13
30	0	0259491	31-07-2016	31-Jul-16	2-Aug-16	NA	2	0	39	1	NA	NA	1	22	0	1

ภาพ 5 ข้อมูลในรูปแบบไฟล์เอ็กเซล

การเตรียมข้อมูล (preprocessing)

ขั้นตอนนี้มีหน้าที่ คือ จัดการข้อมูลให้สามารถนำเข้าสู่กระบวนการของคอต้า ไม่นิ่งได้ กระบวนการเตรียมข้อมูลก่อนการประมวลผล ซึ่งเป็นขั้นตอนที่ใช้เวลานานที่สุด แบ่งออกเป็น 3 ขั้นตอนย่อย ได้แก่

การคัดเลือกข้อมูล (data selection)

ผู้วิจัยกำหนดเป้าหมายว่า จะวิเคราะห์กลุ่มของผู้ป่วยในที่เข้ารับบริการในโรงพยาบาล ดังนั้น จึงเลือกใช้เฉพาะแอทริบิวต์ข้อมูลที่เกี่ยวข้องที่จะวิเคราะห์เท่านั้น ดังในภาพ 6 ซึ่งมีจำนวน 17 แอทริบิวต์

No.	Fall case	HN	Sex	LOS(d)	Age	History of fall	Use IV stand	Assistance walking	Weak	Impaired gait	Mental status/depress	Elimination urine	Elimination Bowel	Dizziness/Vertigo	Personal caregiver	PCA/Opiates	Hypnotics	Sedative	Benzodiazepines
1	0	1E+08	0	4	49	0	1	0	0	0	0	0	0	0	1	0	0	0	0
2	1	1E+08	1	4	68	0	0	0	1	1	0	0	0	0	0	0	0	0	0
3	1	1E+08	1	1	57	0	1	0	0	0	0	0	0	0	1	0	0	0	0
4	0	1E+08	1	14	55	0	1	0	0	1	0	0	0	0	1	0	0	0	0
5	0	1E+08	0	1	83	0	1	0	0	0	0	0	0	0	1	0	0	0	0
6	1	1E+08	0	7	70	1	1	0	0	1	0	0	0	0	1	0	0	0	0
7	1	1E+08	1	12	58	0	1	0	1	1	0	0	0	0	1	0	1	0	1
8	0	1E+08	1	34	54	0	1	0	0	1	0	0	0	0	1	0	0	0	0
9	0	1E+08	0	1	52	0	1	0	0	0	0	0	0	0	1	0	0	0	0

ภาพ 6 ข้อมูลแอทริบิวต์ของชุดข้อมูลเรียนรู้ของผู้ป่วยในของโรงพยาบาลที่คัดเลือก

กลั่นกรองข้อมูล (data cleaning)

ตรวจสอบความถูกต้อง ครบถ้วน ลบข้อมูลที่ซ้ำซ้อน แก้ไขข้อมูลที่ผิดพลาด และขาดหายไป เป็นการทำให้ข้อมูลมีคุณภาพในการนำไปวิเคราะห์ซึ่งมีขั้นตอน คือ ข้อมูลที่ไม่สมบูรณ์ (missing value) จะเติมข้อมูลส่วนที่หายไปให้สมบูรณ์

หลังจากที่ผ่านกระบวนการกลั่นกรองข้อมูลแล้วมีจำนวนผู้ป่วยที่นำมาเป็นข้อมูลในการวิจัยครั้งนี้ จำนวนทั้งสิ้น 400 คน ในการสร้างตัวแบบผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ส่วน โดยข้อมูลผู้ป่วยใน จำนวน 280 คน จะใช้เป็นชุดข้อมูลเรียนรู้ ส่วนข้อมูลของผู้ป่วยใน จำนวน 120 คน จะใช้เป็นชุดทดสอบเพื่อตรวจสอบความถูกต้องของตัวแบบ

ปรับเปลี่ยนรูปแบบของข้อมูล (data transformation)

เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ตามอัลกอริทึมของคาค่าไมนนิ่งที่เลือกใช้ โดยในการวิจัยครั้งนี้เลือกใช้อัลกอริทึมที่รองรับการทำงานของข้อมูลประเภทแบบกลุ่ม Binomial (มีค่าแตกต่างกัน 2 ค่า) และ Polynominal (มีค่าแตกต่างกันมากกว่า 2 ค่า) ดังนั้น จึงต้องปรับค่าต่างๆของแอทริบิวต์ให้อยู่ในรูปแบบที่เหมาะสม

การค้นหาลักษณะ (pattern discovery)

ในการทดสอบตัวแบบในครั้งนี้ผู้วิจัยเลือกใช้ซอฟต์แวร์ RapidMiner Studio 8.1 ซึ่งเป็นโปรแกรม Open Source และผู้วิจัยสามารถที่จะใช้งานใน Educational License ได้โดยจะต้องดาวน์โหลดจากเว็บไซต์

<https://my.rapidminer.com/nexus/account/index.html#downloads> หลังจากดาวน์โหลดโปรแกรมเรียบร้อยแล้ว

เริ่มต้นใช้งานให้ดับเบิลคลิกที่ไอคอน เพื่อเริ่มใช้งานซอฟต์แวร์ หน้าต่างการทำงานของโปรแกรม RapidMiner Studio 8.1 จะแสดงออกมาดังภาพผนวก ก

โดยใช้โปรแกรม RapidMiner Studio 8.1 (auto model) ในขั้นตอนนี้ จะประกอบด้วย 6 ขั้นตอนย่อยได้แก่

1. Select Data เลือกข้อมูลชุดเรียนรู้นำเข้าโปรแกรม
2. Select Task เลือกรูปแบบ Predict เพื่อสร้างตัวแบบทำนายคำตอบ
3. Prepare Target เตรียมคลาสคำตอบที่สนใจคือ โอกาสเกิดการพลัดตกหกล้ม
4. Select Inputs เพื่อคัดเลือกแอทริบิวต์ ทำให้คงเหลือแอทริบิวต์ที่มีความสัมพันธ์กับการที่จะสร้างตัวทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน ที่เข้ารับการรักษายาบาลในโรงพยาบาล จำนวน 17 แอทริบิวต์ ได้แก่ (1) อายุ (age), (2) ระยะเวลาการรักษาตัวในโรงพยาบาล (length of stay), (3) เพศ (sex), (4) ประวัติพลัดตกหกล้ม (history of falls), (5) การคาสายเพื่อให้สารละลายทางหลอดเลือดดำ (use IV stand), (6) การใช้อุปกรณ์ช่วยในการเดิน (assistance walking device), (7) ภาวะความอ่อนแรง (weak), (8) ความสามารถในการเคลื่อนไหวของร่างกายส่วนล่างลดลง (impaired gait), (9) มีความผิดปกติทางด้านอารมณ์และจิตใจ (mental status/depression), (10) อาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม. (elimination urine), (11) อาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม. (elimination bowel), (12) อาการเวียนศีรษะบ้านหมุน (dizziness/vertigo), (13) การดูแลของญาติหรือผู้ดูแลในโรงพยาบาล (personal care giver), (14) ได้รับยาบรรเทาอาการปวด (PCA/opiates), (15) การได้รับยานอนหลับ

(hypnotics), (16) ๒ได้รับยากล่อมประสาท (sedative), (17) การได้รับยากลายเครียด (benzodiazepines)

5. Model Types เลือกวิธีถดถอยโลจิสติก ใช้พารามิเตอร์ Solver แบบ Auto วิธีค้นไม้ตัดสินใจ ใช้พารามิเตอร์ตัวชี้วัดการแบ่งข้อมูลด้วยค่าอัตราส่วนเกิน (gain ratio) Maximal Depth 40 และค่าความเชื่อมั่น (confidence) เท่ากับ 0.25 และวิธีโครงข่ายประสาทเทียมเชิงลึก ใช้พารามิเตอร์ฟังก์ชันกระตุ้นแบบตัวเรียงกระแสไฟฟ้า (rectifier) มาสร้างตัวแบบทำนาย

6. Result ดังภาพแสดงในภาคผนวก ข

การทดสอบประสิทธิภาพตัวแบบทำนาย

การวิจัยนี้มีการวัดประสิทธิภาพและเปรียบเทียบผลการทำนาย โดยใช้การทดสอบวิธี 10-folds Cross-validation ดังภาพแสดงในภาคผนวก ก

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ เป็นการวิจัยเชิงปริมาณ โดยการเปรียบเทียบประสิทธิภาพอัลกอริทึม คาด้าไมน์นิ่งของวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียม เชิงลึก ข้อมูลที่ใช้ในการพัฒนาตัวแบบเป็นข้อมูลของผู้ป่วยในที่ได้รับการรักษาในโรงพยาบาลเอกชนแห่งหนึ่ง ในช่วงเดือนมกราคม 2557 ถึง เดือนธันวาคม 2559 จำนวน 400 คน แบ่งข้อมูลวิธี Split Test แบ่งข้อมูลออกเป็น 2 ส่วน โดยข้อมูลส่วนที่หนึ่งร้อยละ 70 ใช้ในการสร้างตัวแบบคิดเป็นชุดข้อมูลจำนวน 280 คน ใช้วิธี 10-fold Cross-validation ในการทดสอบประสิทธิภาพของตัวแบบ และข้อมูลส่วนที่สองร้อยละ 30 คิดเป็นชุดข้อมูลจำนวน 120 คน ใช้เป็นข้อมูลทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยในที่ได้รับการรักษาพยาบาล ตัวแบบที่มีประสิทธิภาพสูงสุดวัดจาก ค่าความถูกต้อง ค่าความผิดพลาดของการจำแนกประเภทข้อมูล ค่าพื้นที่ใต้กราฟ ROC ค่าความแม่นยำ ค่าความระลึก ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำและค่าความระลึก ค่าความไว และค่าความจำเพาะ โดยใช้โปรแกรม RapidMiner Studio เวอร์ชัน 8.1 มีตัวแปรในการทำนาย 17 ปัจจัย ดังนี้

ผลการวิเคราะห์การคัดเลือกปัจจัย

sum age

Variable	Obs	Mean	Std. Dev.	Min	Max
age	280	53.46429	17.80154	15	95

ภาพ 7 กลุ่มตัวอย่างจำแนกตามอายุ (age)

กลุ่มตัวอย่างชุดข้อมูลเรียนรู้ มีอายุเฉลี่ย (mean) ที่ 53.46429 ปี มีค่าเบี่ยงเบนมาตรฐาน (SD) อยู่ที่ 17.80154 อายุต่ำที่สุด (min) อยู่ที่ 15 ปี และมีอายุมากที่สุด (max) อยู่ที่ 95 ปี

sum losday

Variable	Obs	Mean	Std. Dev.	Min	Max
losday	280	34.68214	71.03482	0	450

ภาพ 8 กลุ่มตัวอย่างจำแนกตามระยะเวลาอนรักษาตัวในโรงพยาบาล (length of stay)

กลุ่มตัวอย่างชุดข้อมูลเรียนรู้ มีระยะเวลาอนรักษาตัวในโรงพยาบาลเฉลี่ย (mean) ที่ 34.68214 วัน มีค่าเบี่ยงเบนมาตรฐาน (SD) อยู่ที่ 71.03482 ระยะเวลาอนรักษาตัวในโรงพยาบาลต่ำที่สุด (min) อยู่ที่ 0 วัน และระยะเวลาอนรักษาตัวในโรงพยาบาลมากที่สุด (max) อยู่ที่ 450 วัน

ตาราง 2

กลุ่มตัวอย่างจำแนกตามเพศ (sex)

เพศ	ร้อยละ
ชาย	53.21
หญิง	46.78
รวม	100.00

จำแนกตามเพศ พบว่า กลุ่มที่มากที่สุด คือ เพศชาย มีร้อยละ 53.21 รองลงมา คือ เพศหญิง มีร้อยละ 46.78

ตาราง 3

กลุ่มตัวอย่างจำแนกตามประวัติพลัดตกหกล้ม (*history of fall*)

ประวัติพลัดตกหกล้ม	ร้อยละ
ใช่	6.07
ไม่ใช่	93.92
รวม	100.00

ประวัติพลัดตกหกล้ม กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่เคยมีประวัติพลัดตกหกล้ม มีร้อยละ 93.92 รองลงมา คือ กลุ่มที่เคยมีประวัติพลัดตกหกล้มมีร้อยละ 6.07

ตาราง 4

กลุ่มตัวอย่างจำแนกตามการได้รับการคาสายเพื่อให้สารละลายทางหลอดเลือดดำ

(*use IV stand*)

ได้รับการคาสายสวนทางเส้นเลือดดำ	ร้อยละ
ใช่	93.21
ไม่ใช่	6.78
รวม	100.00

การได้รับการคาสายเพื่อให้สารละลายทางหลอดเลือดดำพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่มีการคาสายเพื่อให้สารละลายทางหลอดเลือดดำ มีร้อยละ 93.21 รองลงมา คือ กลุ่มที่ไม่ได้รับการคาสายเพื่อให้สารละลายทางหลอดเลือดดำ มีร้อยละ 6.78

ตาราง 5

กลุ่มตัวอย่างจำแนกตามการใช้อุปกรณ์ในการช่วยเดิน (*assistance walking device*)

การใช้อุปกรณ์ในการช่วยเดิน	ร้อยละ
ใช่	6.42
ไม่ใช่	93.57
รวม	100.00

การใช้อุปกรณ์ในการช่วยเดินพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่ได้ใช้อุปกรณ์ในการช่วยเดินมีร้อยละ 93.57 รองลงมา คือ กลุ่มที่ใช้อุปกรณ์ในการช่วยเดินมีร้อยละ 6.42

ตาราง 6

กลุ่มตัวอย่างจำแนกตามภาวะอ่อนแรง (*weak*)

ภาวะอ่อนแรง	ร้อยละ
ใช่	19.28
ไม่ใช่	80.71
รวม	100.00

ภาวะอ่อนแรงพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่มีภาวะอ่อนแรงมีร้อยละ 80.71 รองลงมา คือ กลุ่มที่มีภาวะอ่อนแรงมีร้อยละ 19.28

ตาราง 7

กลุ่มตัวอย่างจำแนกตามความสามารถในการเคลื่อนไหวของรยางค์ส่วนล่างลดลง
(*impaired gait*)

ความสามารถในการเคลื่อนไหวของรยางค์ส่วนล่าง ลดลง	ร้อยละ
ใช่	35.00
ไม่ใช่	65.00
รวม	100.00

ความสามารถในการเคลื่อนไหวของรยางค์ส่วนล่างลดลงพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ความสามารถในการเคลื่อนไหวของรยางค์ส่วนล่างปกติ ร้อยละ 65 รองลงมา คือ กลุ่มที่ความสามารถในการเคลื่อนไหวของรยางค์ส่วนล่างลดลง ร้อยละ 35

ตาราง 8

กลุ่มตัวอย่างจำแนกตามภาวะความผิดปกติทางด้านอารมณ์และจิตใจ
(*mental status/depression*)

ภาวะความผิดปกติทางด้านอารมณ์และจิตใจ	ร้อยละ
ใช่	8.57
ไม่ใช่	91.43
รวม	100.00

ภาวะความผิดปกติทางด้านอารมณ์และจิตใจพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่มีปัญหาทางภาวะความผิดปกติทางด้านอารมณ์และจิตใจมีร้อยละ 91.43 รองลงมา คือ กลุ่มที่มีปัญหาทางภาวะความผิดปกติทางด้านอารมณ์และจิตใจมีร้อยละ 8.57

ตาราง 9

กลุ่มตัวอย่างจำแนกตามอาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม. (elimination urine)

อาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม.	ร้อยละ
ใช่	13.93
ไม่ใช่	86.07
รวม	100.00

อาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม. พบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่มีอาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม. ร้อยละ 86.07 รองลงมา คือ กลุ่มที่มีอาการปัสสาวะบ่อยหรือได้รับยาขับปัสสาวะในระยะเวลา 24 ชม. ร้อยละ 13.93

ตาราง 10

กลุ่มตัวอย่างจำแนกตามอาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม. (elimination bowel)

อาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม.	ร้อยละ
ใช่	12.14
ไม่ใช่	87.86
รวม	100.00

อาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม. พบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่มีอาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม. ร้อยละ 87.86 รองลงมาคือ กลุ่มที่มีอาการท้องเสียหรือได้รับยาระบายในระยะเวลา 24 ชม. ร้อยละ 12.14

ตาราง 11

กลุ่มตัวอย่างจำแนกตามอาการเวียนศีรษะบ้านหมุน (*dizziness/ vertigo*)

อาการเวียนศีรษะบ้านหมุน	ร้อยละ
ใช่	8.57
ไม่ใช่	91.43
รวม	100.00

อาการเวียนศีรษะบ้านหมุนพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่มีอาการเวียนศีรษะบ้านหมุนมีร้อยละ 91.43 รองลงมา คือ กลุ่มที่มีอาการเวียนศีรษะบ้านหมุนมีร้อยละ 8.57

ตาราง 12

กลุ่มตัวอย่างจำแนกตามการดูแลของญาติหรือผู้ดูแลในโรงพยาบาล

(*personal care giver*)

การดูแลของญาติหรือผู้ดูแลในโรงพยาบาล	ร้อยละ
ใช่	93.93
ไม่ใช่	6.07
รวม	100.00

การดูแลของญาติหรือผู้ดูแลในโรงพยาบาล พบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่มีการดูแลของญาติหรือผู้ดูแลในโรงพยาบาล มีร้อยละ 93.93 รองลงมา คือ กลุ่มที่ไม่มีการดูแลของญาติหรือผู้ดูแลในโรงพยาบาล มีร้อยละ 6.07

ตาราง 13

กลุ่มตัวอย่างจำแนกตามการได้รับยาบรรเทาอาการปวด (PCA/opiates)

	ได้รับยาบรรเทาอาการปวด	ร้อยละ
ใช่		8.21
ไม่ใช่		91.79
	รวม	100.00

การได้รับยาบรรเทาอาการปวดพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่ได้รับยาบรรเทาอาการปวด มีร้อยละ 91.79 รองลงมา คือ กลุ่มที่ได้รับยาบรรเทาอาการปวดมีร้อยละ 8.21

ตาราง 14

กลุ่มตัวอย่างจำแนกตามการได้รับยานอนหลับ (hypnotics)

	การได้รับยานอนหลับ	ร้อยละ
ใช่		22.50
ไม่ใช่		77.50
	รวม	100.00

การได้รับยานอนหลับพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่การได้รับยานอนหลับ ร้อยละ 77.50 รองลงมา คือ กลุ่มที่การได้รับยานอนหลับร้อยละ 22.50

ตาราง 15

กลุ่มตัวอย่างจำแนกตามการได้รับยากล่อมประสาท (*sedative*)

การได้รับยากล่อมประสาท	ร้อยละ
ใช่	7.50
ไม่ใช่	92.50
รวม	100.00

การได้รับยากล่อมประสาทพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่ได้รับยากล่อมประสาทมีร้อยละ 92.50 รองลงมา คือ กลุ่มที่ได้รับยากล่อมประสาทมีร้อยละ 7.50

ตาราง 16

กลุ่มตัวอย่างจำแนกตามการได้รับยาคลายเครียด (*benzodiazepines*)

การได้รับยาคลายเครียด	ร้อยละ
ใช่	22.50
ไม่ใช่	77.50
รวม	100.00

การได้รับยาคลายเครียดพบว่า กลุ่มที่มากที่สุด คือ กลุ่มที่ไม่ได้รับยาคลายเครียดร้อยละ 77.50 รองลงมา คือ กลุ่มที่ได้รับยาคลายเครียดร้อยละ 22.50

ตาราง 17

กลุ่มตัวอย่างจำแนกตามสถานะการพลัดตกหกล้ม (*falls*)

สถานะการพลัดตกหกล้ม	ร้อยละ
ใช่	49.29
ไม่ใช่	50.71
รวม	100.00

สถานการณ์ปลัดตกหกล้ม กลุ่มตัวอย่างชุดการเรียนรู้ที่นำมาทดสอบในจำนวน 280 คน มีผู้ป่วยที่มีการปลัดตกหกล้ม คิดเป็นร้อยละ 49.29 และมีผู้ป่วยที่ไม่มีการปลัดตกหกล้มคิดเป็นร้อยละ 50.71

ผลการวิเคราะห์และเปรียบเทียบประสิทธิภาพตัวแบบทำนาย

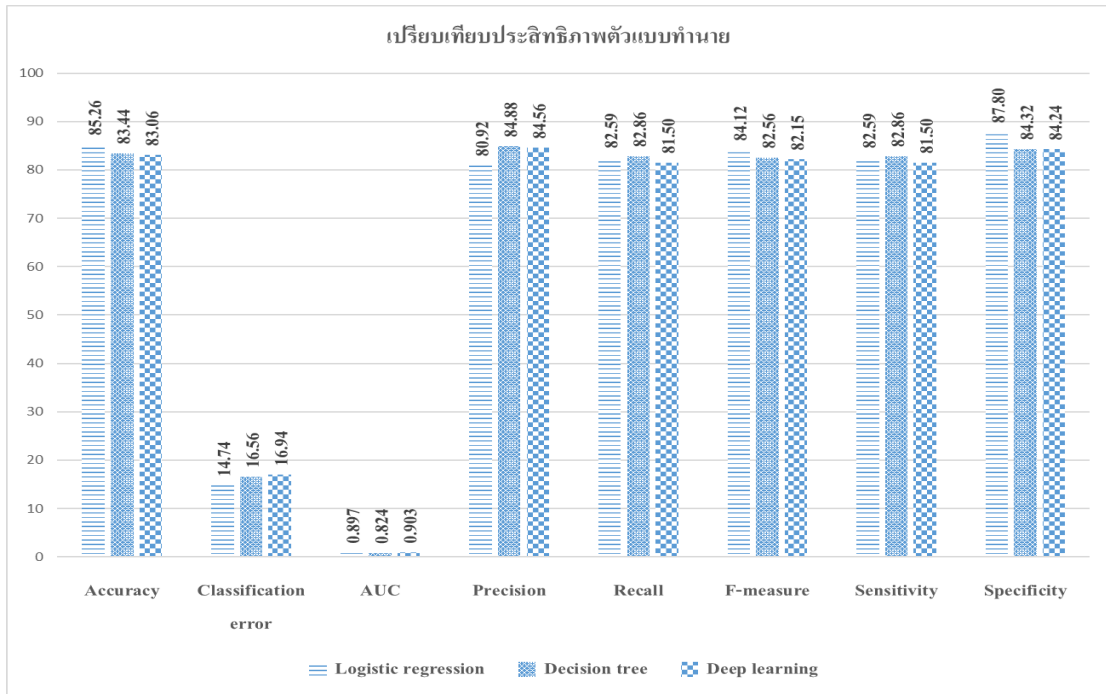
หลังจากผ่านกระบวนการค้นหารูปแบบโดยใช้โปรแกรม RapidMiner Studio 8.1 Auto Model โดยเลือกประเภทเป็นการทำนาย เลือกตัวแบบ วิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก จากนั้นจึงรันตัวโปรแกรมเพื่อหาตัวแบบที่มีประสิทธิภาพมากที่สุด ผลการวิจัยพบว่า

วิธีถดถอยโลจิสติก มีค่าความถูกต้อง (accuracy) ร้อยละ 85.26 ค่าความผิดพลาดของการจำแนกประเภทข้อมูล (classification error) ร้อยละ 14.74 ค่าพื้นที่ใต้กราฟ ROC เท่ากับ 0.897 ค่าความแม่นยำ (precision) ร้อยละ 87.01 ค่าความระลึก (recall) ร้อยละ 82.59 ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำและค่าความระลึก (f-measure) ร้อยละ 84.12 ค่าความไว (sensitivity) ร้อยละ 82.59 และค่าความจำเพาะ (specificity) ร้อยละ 87.80

วิธีต้นไม้ตัดสินใจ มีค่าความถูกต้อง ร้อยละ 83.44 ความผิดพลาดของการจำแนกประเภทข้อมูล ร้อยละ 16.56 ค่าพื้นที่ใต้กราฟ ROC เท่ากับ 0.824 ค่าความแม่นยำร้อยละ 84.88 ค่าความระลึกร้อยละ 82.86 ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำและค่าความระลึกร้อยละ 82.56 ค่าความไวร้อยละ 82.86 และค่าความจำเพาะร้อยละ 84.32

วิธีโครงข่ายประสาทเทียมเชิงลึก มีค่าความถูกต้องร้อยละ 83.06 ค่าความผิดพลาดของการจำแนกประเภทข้อมูลร้อยละ 16.94 ค่าพื้นที่ใต้กราฟ ROC เท่ากับ 0.903 ค่าความแม่นยำร้อยละ 84.56 ค่าความระลึกร้อยละ 81.50 ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำและค่าความระลึกร้อยละ 82.15 ค่าความไวร้อยละ 81.50 และค่าความจำเพาะร้อยละ 84.24

ภาพแสดงหน้าต่างรายงานผลการวิเคราะห์จากโปรแกรม RapidMiner Studio 8.1
 ดังแสดงในภาคผนวก ข ภาพ 46 ถึงภาพ 72



ภาพ 9 ค่าเปรียบเทียบประสิทธิภาพตัวทำนายวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ
 และวิธีโครงข่ายประสาทเทียมเชิงลึก

ผลการทดสอบประสิทธิภาพตัวแบบทำนาย

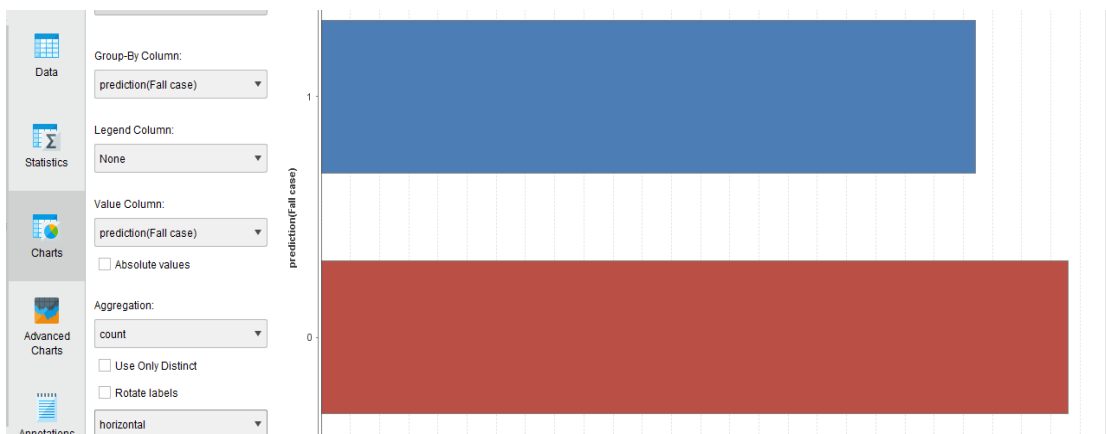
โดยภาพรวมทั้ง 3 วิธีมีประสิทธิภาพสูงที่ใกล้เคียงกัน โดยพิจารณาประสิทธิภาพ
 จากค่าความถูกต้อง ค่าพื้นที่ใต้กราฟ ROC ค่าความแม่นยำ ค่าวัดประสิทธิภาพมวลรวม
 ของค่าความแม่นยำและค่าความระลึก ค่าความไว พบว่าตัวแบบวิธีถดถอยโลจิสติกมีค่า
 สูงที่สุดและค่าความผิดพลาดของการจำแนกประเภทข้อมูลที่น้อยที่สุด ดังนั้นวิธีถดถอย
 โลจิสติกจึงเป็นตัวแบบที่มีค่าประสิทธิภาพที่ดีกว่าวิธีต้นไม้ตัดสินใจและโครงข่ายประ
 สาทเทียมเชิงลึกตามลำดับ จึงได้นำตัวแบบทำนายของวิธีถดถอยโลจิสติก มาใช้เพื่อทำนาย
 โอกาสเกิดการพลัดตกหกล้มของผู้ป่วยในด้วยชุดข้อมูลทดสอบ ซึ่งมีจำนวน 120 เรคคอร์ด
 รูปแบบของกราฟแสดงผลดังภาพ 10 - ภาพ 12 และตัวแปรทำนายตัวแปรทำนายที่มีความ
 สัมพันธ์กับการพลัดตกหกล้มดังภาพ 13

Row...	HN	Fall ca...	Age	LOS(day)	Sex	History of fall	Use IV stand	Assistance ...	Weak	Unsteady gait	Mental statu...	Elimination ...
1	1017...	?	29	39	1	0	1	0	1	0	0	1
2	1006...	?	66	1	1	0	1	0	0	0	0	0
3	1007...	?	31	2	0	0	1	0	0	0	0	0
4	1018...	?	73	1	1	0	1	0	0	1	0	0
5	1025...	?	59	11	1	0	1	0	0	1	0	0
6	1002...	?	66	2	1	0	1	0	0	0	0	0
7	1002...	?	45	1	1	0	1	1	0	1	0	0
8	1019...	?	73	50	1	0	1	0	1	1	0	0
9	1002...	?	83	1	0	0	1	0	0	0	0	1
10	1025...	?	73	209	0	0	1	0	0	0	0	0
11	1011...	?	72	2	1	1	1	0	0	0	0	0
12	1025...	?	55	83	1	0	1	0	0	0	0	0
13	1007...	?	52	3	0	0	0	0	0	0	0	0
14	1025...	?	69	25	0	0	1	0	0	0	0	1
15	1014...	?	71	19	1	0	1	0	0	0	0	1
16	1024...	?	59	2	1	0	1	0	0	0	0	0

ภาพ 10 แสดงหน้าต่างแสดงผล Example Set ของชุดข้อมูลทดสอบกับการทดสอบตัวแบบวิธีถดถอยโลจิสติก

Name	Type	Missing	Statistics	Filter (22 / 22 attributes)
id HN	Integer	0	Min: 100032723, Max: 102634002, Average: 101857723.817	
Label Fall case	Binominal	120	Least, Most, Values	
Prediction prediction(Fall case)	Nominal	0	Least: 1 (56), Most: 0 (64), Values: 0 (64), 1 (56)	
Confidence_0 confidence(0)	Real	0	Min: 0.000, Max: 0.981, Average: 0.496	
Confidence_1 confidence(1)	Real	0	Min: 0.019, Max: 1.000, Average: 0.504	

ภาพ 11 แสดงหน้าต่างแสดงผลสถิติการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้รับบริการรายใหม่ที่น่าเข้าไปทดสอบตัวแบบวิธีถดถอยโลจิสติก



ภาพ 12 แสดงหน้าต่างแสดงผลรูปแบบแผนภูมิของการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้รับบริการรายใหม่ที่น่าเข้าไปทดสอบตัวแบบวิธีถดถอยโลจิสติก

จะเห็นว่า เมื่อนำเข้าข้อมูลชุดทดสอบที่มีจำนวน 120 คนเข้าไปทดสอบในตัวแบบวิธีถดถอยโลจิสติกนั้น จะมีการทำนายโอกาสเกิดการพลัดตกหกล้ม (falls = 1) จำนวน 56 คน และทำนายมีโอกาสไม่เกิดการพลัดตกหกล้ม (falls = 0) จำนวน 64 คน

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value ↑
LOS(day)	0.092	6.518	0.026	3.494	0.000
Weak .1	3.285	3.285	0.963	3.412	0.001
History of fall.1	3.402	3.402	1.232	2.761	0.006
Mental status/depression.1	1.747	1.747	0.794	2.200	0.028
Personal care giver.0	1.877	1.877	0.867	2.166	0.030

ภาพ 13 หน้าต่างตัวแปรทำนายด้วยวิธีถดถอยโลจิสติกที่มีความสัมพันธ์กับการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

พบว่า มี 5 ตัวแปรทำนายที่มีความสัมพันธ์กับการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ ระยะเวลานอนโรงพยาบาล (length of stay) มีภาวะอ่อนแรง (weak) มีประวัติพลัดตกหกล้ม (history of falls) มีภาวะความผิดปกติทางด้านอารมณ์ และจิตใจ (mental status/depression) และการไม่มีญาติหรือผู้ดูแล (personal care giver)

บทที่ 5

สรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะ

สรุปผลการวิจัย

เปรียบเทียบประสิทธิภาพอัลกอริทึมเหมืองข้อมูลวิธีถดถอยโลจิสติกกับวิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก โดยภาพรวมทั้ง 3 วิธี มีประสิทธิภาพสูงที่ใกล้เคียงกัน โดยพิจารณาประสิทธิภาพจากค่าความถูกต้อง ค่าพื้นที่ใต้กราฟ ROC ค่าความแม่นยำ ค่าวัดประสิทธิภาพมวลรวมของค่าความแม่นยำและค่าความระลึกลับ ค่าความไวพบว่า ตัวแบบวิธีถดถอยโลจิสติกที่มีค่าสูงที่สุดและมีค่าความผิดพลาดของการจำแนกประเภทข้อมูลที่น้อยที่สุด ดังนั้น วิธีถดถอยโลจิสติกจึงเป็นตัวแทนทำนายที่มีค่าประสิทธิภาพที่ดีกว่าวิธีต้นไม้ตัดสินใจ และโครงข่ายประสาทเทียมเชิงลึก ตามลำดับ

ผู้วิจัยจึงสรุปได้ว่า วิธีถดถอยโลจิสติกเป็นวิธีที่ดีที่สุด สามารถที่จะนำแบบจำลองของวิธีถดถอยโลจิสติกไปใช้เพื่อการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยในเพื่อเป็นแนวทางในการวิเคราะห์ข้อมูลทางการแพทย์ และวิเคราะห์ปัจจัยที่มีผลต่อการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยที่นอนพักรักษาตัวเป็นผู้ป่วยในของโรงพยาบาล โดยจำนวนของชุดข้อมูลเรียนรู้ที่มีขนาดใหญ่ขึ้น จะมีผลต่อการสร้างตัวแบบทำนายที่มีประสิทธิภาพมากยิ่งขึ้น ถึงแม้ว่าอัลกอริทึมใดก็ตามจะสมารถนำมาใช้ในวงการทางการแพทย์กันอย่างแพร่หลาย ความรู้ที่ได้สามารถนำไปต่อยอดสร้างนวัตกรรมต่าง ๆ เพื่อการดูแลรักษาพยาบาลผู้ป่วยได้อย่างมีประสิทธิภาพได้ แต่ความรู้ที่ได้มานี้จะเป็นเพียงข้อมูลคำแนะนำเพื่อประกอบการตัดสินใจเท่านั้น ไม่ได้มาแทนที่การตัดสินใจที่เกิดจากดุลพินิจของแพทย์ผู้เชี่ยวชาญ

การอภิปรายผล

ผลการศึกษาบางส่วนมีความคล้ายคลึงกับการศึกษาของคนอื่น ๆ ตัวอย่างเช่น การศึกษาของ รัตนาพร ทามิ พวงทิพย์ ชัยพิบาลสถิตย์ และปานตา อภิรักษ์ภานนท์ (2561) ที่พบว่าปัจจัยส่วนบุคคล (อายุ, เพศ, ประวัติพลัดตกหกล้ม, การใช้ล้ามในการสื่อสาร และระยะห่างของประเทศที่อาศัยตามเส้นแบ่งเวลา) ปัจจัยภายใน (ภาวะจิตใจ/อารมณ์ อาการสับสน/มึนงง, อาการเวียนบ้านหมุน, ความถี่ในการขับถ่ายอุจจาระ/ปัสสาวะ, การได้รับยากลุ่ม Hypnotic/Benzodiazepine Anticonvulsants และความบกพร่องในการเคลื่อนไหวทรงตัว) และปัจจัยภายนอก (การมีญาติหรือผู้ดูแล, ระยะเวลานอนโรงพยาบาล และระดับความเสี่ยงต่อการพลัดตกหกล้ม มีความสัมพันธ์กับการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่การศึกษานี้ พบว่า มี 5 ตัวแปร ที่เป็นปัจจัยที่มีความสัมพันธ์กับการพลัดตกหกล้มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ ระยะเวลานอนโรงพยาบาล, มีภาวะอ่อนแรง, มีประวัติพลัดตกหกล้ม, มีภาวะความผิดปกติทางด้านอารมณ์และจิตใจ และการไม่มีญาติหรือผู้ดูแล ซึ่งแตกต่างจากการศึกษาของดีและคณะ (2011) ในประเทศไต้หวัน ที่ใช้เทคนิคดาต้าไมน์นิง ANN ใช้โปรแกรม SPSS, STATISTICA 8.0 วิเคราะห์งานวิจัยแอปพลิเคชันของดาต้าไมน์นิงในการระบุตัวแปรสำคัญที่เป็นสาเหตุของการพลัดตกหกล้มของผู้ป่วย โดยใช้ระบบการรายงานในโปรแกรมประยุกต์บนเว็บ ระบุว่ามีเพียง 3 ตัวแปร ได้แก่ การประเมินการพลัดตกหกล้ม การได้รับยาต่อต้านโรคจิต และยาขับปัสสาวะ เป็นตัวทำนายที่สำคัญด้วยค่าพื้นที่ใต้กราฟ ROC มีค่าร้อยละ 42 ค่าทำนายบวกร้อยละ 26.24 และค่าทำนายลบร้อยละ 87.12 และแตกต่างจากงานวิจัยของ Marschollek et.al. (2012) รายงานว่าในประเทศเยอรมนี ได้ศึกษาการใช้ดาต้าไมน์นิงทำนายกลุ่มเสี่ยงของผู้สูงอายุ ที่จะเกิดการพลัดตกหกล้มในโรงพยาบาล โดยใช้ชุดข้อมูล จำนวน 5,176 คน ใช้อัลกอริทึม C4.5 ของวิธีต้นไม้ตัดสินใจเปรียบเทียบกับวิธีถดถอยโลจิสติก ผลการศึกษาพบว่า วิธีต้นไม้ตัดสินใจมีความถูกต้องร้อยละ 66 ค่าความไวร้อยละ 55.4 ค่าความจำเพาะร้อยละ 67.1 ค่าทำนายบวกร้อยละ 15 และค่าทำนายลบร้อยละ 93.5 มี 5 ตัวแปร ที่มีความสำคัญกับการทำนายโอกาสเกิดการพลัดตก

หกล้มในผู้สูงอายุ ได้แก่ High Age, Low Barthel Index, Cognitive Impairment, Multi-medication และ Co-morbidity

การวิจัยนี้ ผู้วิจัยใช้อัลกอริทึมดาต้า ไมน์นิ่งการจำแนกประเภทข้อมูลวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก มาเปรียบเทียบประสิทธิภาพกันว่าอัลกอริทึมตัวใด จะมีประสิทธิภาพในการทำนายสูงที่สุด ซึ่งพบว่า ทั้ง 3 อัลกอริทึมนี้มีค่าประสิทธิภาพสูงใกล้เคียงกัน แต่วิธีถดถอยโลจิสติก ให้ค่าความถูกต้องสูงที่สุด คือ ร้อยละ 85.26 ค่าความผิดพลาดของการจำแนกประเภทข้อมูลน้อยที่สุด คือ ร้อยละ 14.74 ค่าพื้นที่ใต้กราฟ ROC สูงที่สุด คือ เท่ากับ 0.897 ค่าความแม่นยำสูงที่สุด คือ ร้อยละ 87.01 ค่าวัดประสิทธิภาพรวมของค่าความแม่นยำและค่าความระลึกลูกสูงที่สุด คือ ร้อยละ 84.12 และมีค่าความจำเพาะสูงที่สุด คือ ร้อยละ 87.80 ดังนั้นอัลกอริทึมดาต้า ไมน์นิ่งวิธีถดถอยโลจิสติกจึงเป็นตัวแบบทำนายที่มีประสิทธิภาพสูงที่สุดในการทำนายโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน ตัวแปรทำนายที่มีความสอดคล้อง และไม่สอดคล้องกับงานวิจัยอื่น ๆ นั้น ขึ้นอยู่กับ ขนาดของข้อมูล เครื่องมือ และเทคนิคของดาต้า ไมน์นิ่งที่เลือกใช้ที่มีความแตกต่างกัน

ข้อจำกัดของการวิจัย

ข้อจำกัดของการวิจัยครั้งนี้ คือ จำนวนชุดข้อมูลที่นำมาศึกษาวิเคราะห์นั้น ยังมีค่อนข้างน้อยเมื่อเทียบกับการศึกษาอื่น ๆ ที่นำข้อมูลไปวิเคราะห์กับอัลกอริทึมดาต้า ไมน์นิ่งการจำแนกประเภทข้อมูล เนื่องจากโรงพยาบาลที่มีระบบการป้องกันการเข้าถึงข้อมูลในรายที่มีความเสี่ยงสูง จึงมีข้อจำกัดในการเข้าถึงข้อมูลเพิ่มประวัติของผู้ป่วยบางอย่าง เช่น การบันทึกอาการของผู้ป่วย ณ เวลาที่เกิดเหตุการณ์พลัดตกหกล้ม และการดูแลรักษา หลังเกิดเหตุการณ์ ที่ข้อมูลดังกล่าวจะได้จากการรายงานอุบัติการณ์จากเจ้าหน้าที่พยาบาล เท่านั้น ทำให้ข้อมูลบางส่วนขาดความสมบูรณ์ เครื่องมือในการศึกษาครั้งนี้ คือ โปรแกรม RapidMiner Studio เวอร์ชัน 8.1 ซึ่งมีความทันสมัย มีรูปแบบกราฟสีที่สวยงาม แต่มักจะมีราคาแพงและใช้งานยากซับซ้อนสำหรับคนที่ยังไม่เคยฝึกปฏิบัติ การใช้งานโปรแกรมการแปลผลลัพธ์ของผู้ใช้งาน โปรแกรม อาจจะยังเข้าใจได้ยาก ผลลัพธ์ (output) จะขึ้นอยู่กับตัวแปรข้อมูลนำเข้า (input) ที่นำเข้าไปศึกษาว่ามีความสมบูรณ์มากน้อยแค่ไหนจะ

ต้องแปลงข้อมูลนำเข้าให้เหมาะสมกับอัลกอริทึมที่เลือกใช้ และยิ่งถ้าข้อมูลชุดเรียนรู้มีปริมาณมาก สมบูรณ์มาก ค่าประสิทธิภาพของตัวแบบนั้นจะยิ่งมีค่าทำนายที่สูง อีกทั้งเครื่องมือยังมีความใหม่และข้อมูลการเรียนรู้ในการพัฒนาตัวแบบ ที่เป็นที่แพร่หลายในประเทศไทยยังมีไม่มาก งานวิจัยอัลกอริทึมค้ำไม้นิ่ง ที่เกี่ยวข้องกับการคัดเลือกปัจจัยในการทำนายโอกาสเกิดพลัดตกหกล้มของผู้ป่วยในที่พักรักษาตัวในโรงพยาบาล ก็ยังมีปริมาณน้อย

ข้อเสนอแนะ


ควรมีชุดข้อมูลการเรียนรู้ของข้อมูลผู้ป่วยที่เข้ารับบริการเป็นผู้ป่วยใน ที่มีการพลัดตกหกล้ม และยังไม่มีการพลัดตกหกล้ม ที่มีปริมาณของข้อมูลที่มีขนาดใหญ่กว่างานวิจัยฉบับนี้ เพื่อมาสร้างตัวแบบทำนาย และวัดประสิทธิภาพของโมเดลทำนายในโมเดลอื่น ๆ ด้วย

ข้อเสนอแนะเพื่อนำไปใช้

ควรมีการนำแบบจำลองการทำนายโอกาสเกิดการพลัดตกหกล้ม ที่มีประสิทธิภาพสูงที่สุด มาพัฒนาเป็นระบบตรวจประเมินโอกาสเกิดการพลัดตกหกล้มของผู้ป่วยใน ที่เป็นรูปแบบของเว็บไซต์ภายในโรงพยาบาล ที่พยาบาลวิชาชีพและบุคลากรสุขภาพอื่น ๆ สามารถที่จะเข้าไปใช้งานได้อย่างง่ายผ่านทางโทรศัพท์มือถือ เพื่อความทันสมัยและทันเวลา และเพื่อเป็นการพัฒนาทักษะของพยาบาลวิชาชีพในการประเมินความเสี่ยง และการดูแลผู้ป่วยให้เกิดความปลอดภัยจากการพลัดตกหกล้ม และลดความรุนแรงของการพลัดตกหกล้มของผู้ป่วยในของโรงพยาบาลเอกชนให้ลดน้อยลง

ภาคผนวก ก

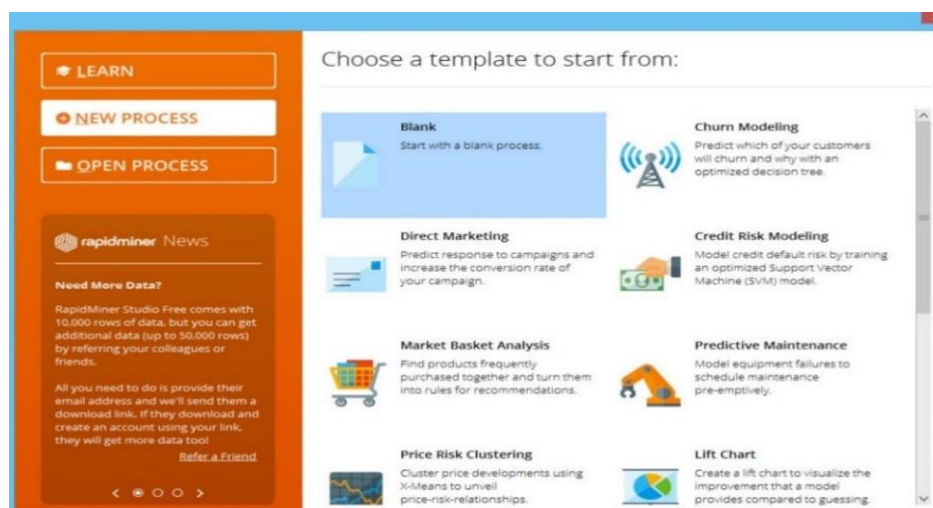
คู่มือการใช้งานโปรแกรม RapidMiner Studio 8.1

1. หลังจากดาวน์โหลดโปรแกรม RapidMiner Studio 8.1 มาเป็นที่เรียบร้อยแล้ว เริ่มต้นใช้งานให้ดับเบิลคลิก ที่ไอคอน  เพื่อเริ่มใช้งานซอฟต์แวร์ หน้าต่างการทำงานของโปรแกรม จะแสดงออกมา



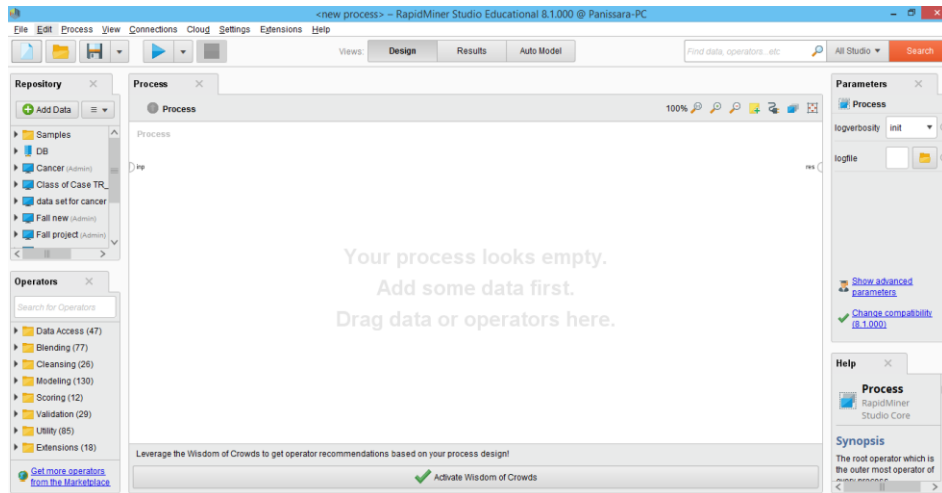
ภาพ 14 หน้าต่าง Welcome ของ RapidMiner Studio 8.1

2. ไปที่ New Process แล้วเลือก Blank เพื่อทำการสร้าง Process ขึ้นมาใหม่




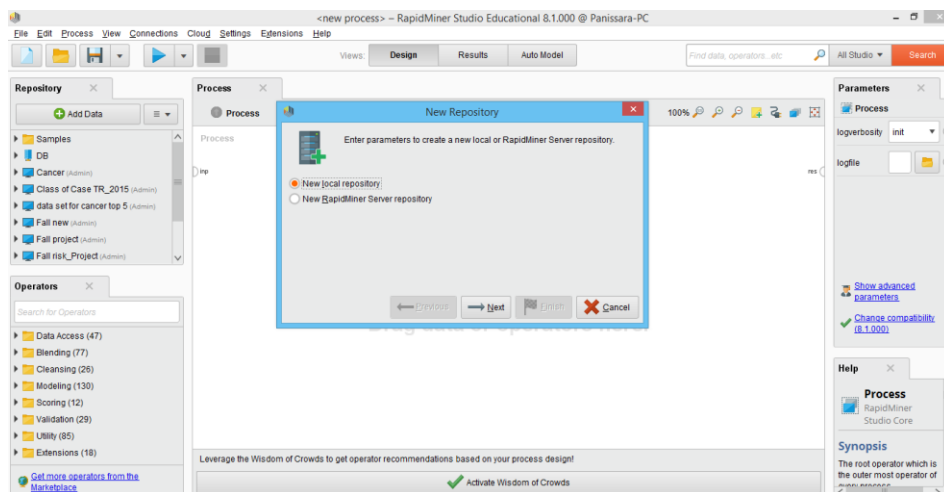
ภาพ 15 หน้าต่างเริ่มต้นการทำงานของโปรแกรม

3. หลังจากเลือกไปที่ Blank จะได้หน้าต่างดังภาพ 16



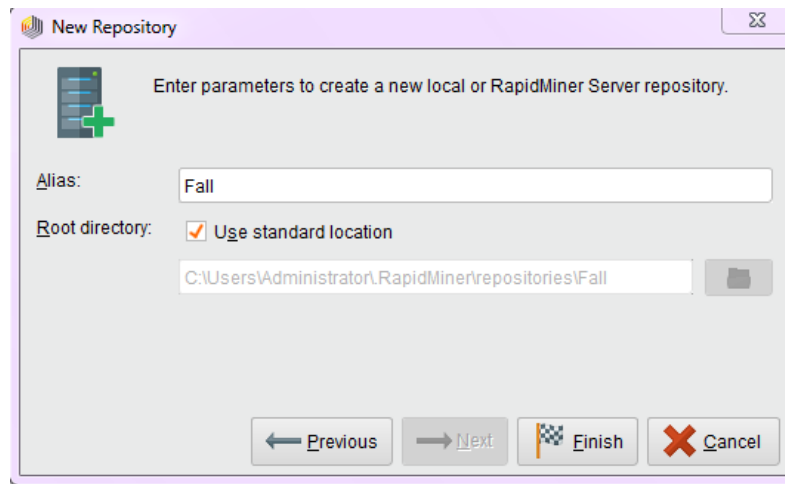
ภาพ 16 หน้าต่างหน้าจอการทำงานในโปรแกรม

4. โหลดข้อมูลเข้าไปเก็บไว้ใน Repository โดยจะสร้าง Repository ขึ้นมา
5. คลิกที่ไอคอน  จากนั้นเลือกเมนู Create Repository และ New Local Repository กดปุ่ม Next



ภาพ 17 หน้าต่างการสร้าง Repository ใหม่ในโปรแกรม

6. เปลี่ยนชื่อ Alias เป็น Fall และกดปุ่ม Finish

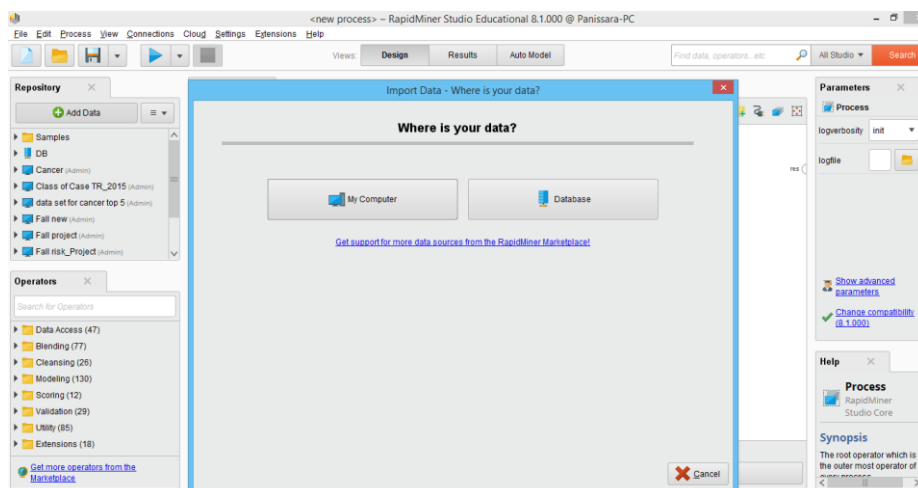
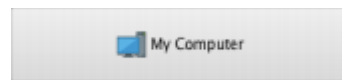


ภาพ 18 หน้าต่างการตั้งชื่อใหม่ใน Repository

7. ในส่วนของ Repository คลิกที่ปุ่ม

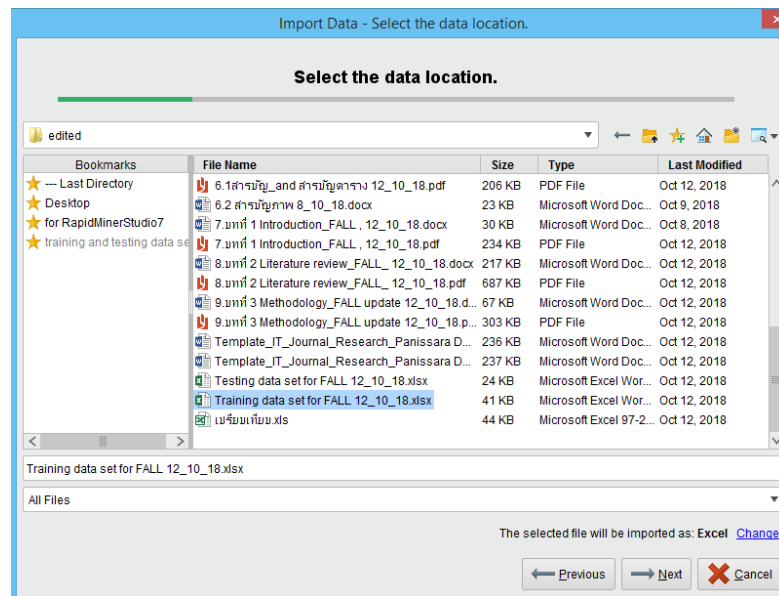


8. คลิกที่ปุ่ม



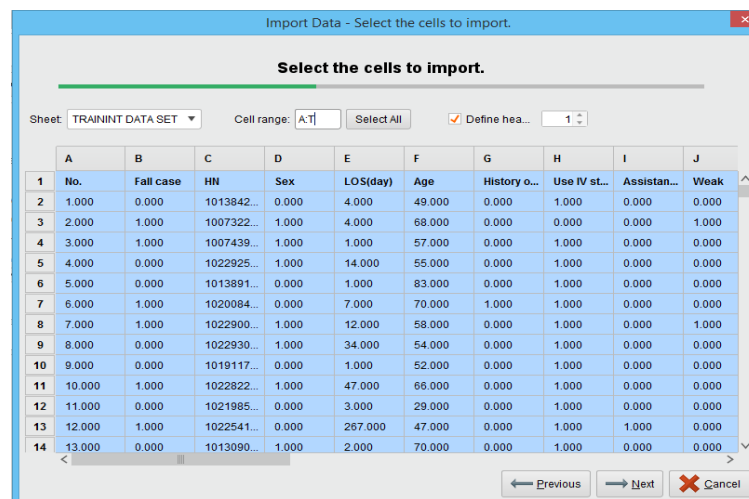
ภาพ 19 หน้าต่างการนำข้อมูลเข้าในโปรแกรม

9. เลือก Training Data และกดปุ่ม Next



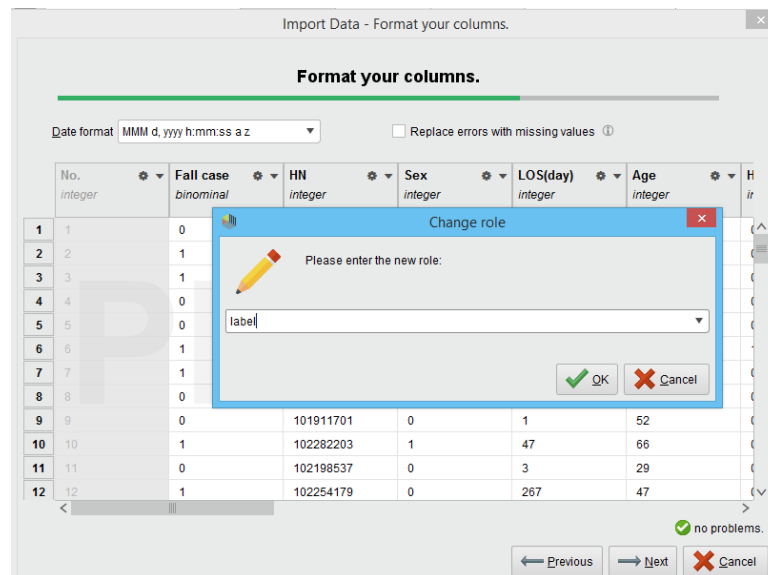
ภาพ 20 หน้าต่างการเลือกไฟล์ที่จะนำเข้าไปในโปรแกรม

10. เลือก Worksheet ของไฟล์ Excel และ Cell ที่ต้องการข้อมูลเข้าไป แล้วกดปุ่ม Next



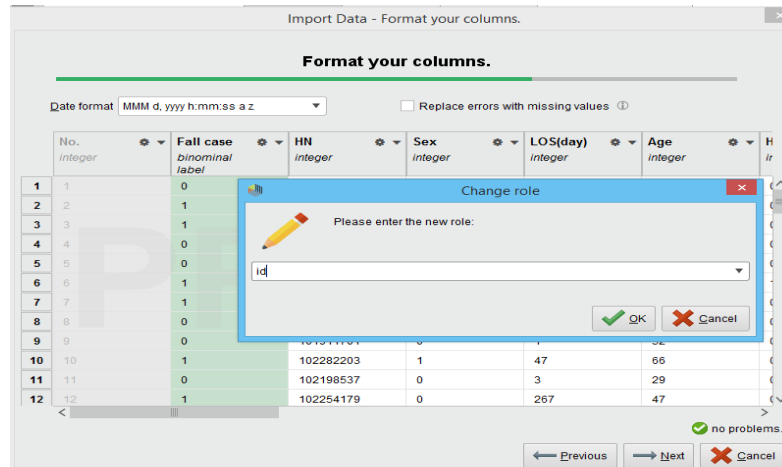
ภาพ 21 หน้าต่างข้อมูลที่โหลดเข้าไปในโปรแกรม

11. หน้าจอนี้จะแสดงประเภทของข้อมูลและหน้าที่ของแต่ละแอทริบิวต์ โดยใน Training Data นี้จะทำการแก้ไขในส่วนของ Fall Case โดยเลือกเมนู Change Role เป็น Label กด OK



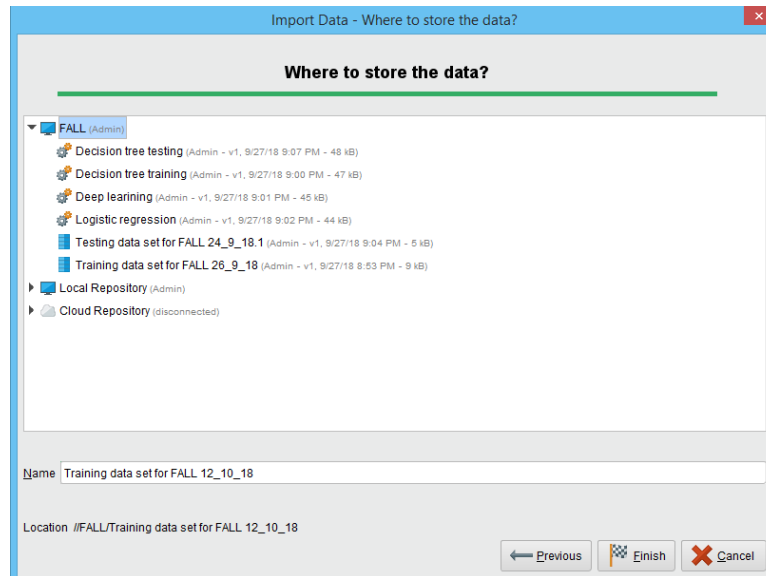
ภาพ 22 หน้าต่างการเปลี่ยนหน้าที่เป็น คลาสคำตอบ (label)

12. ในส่วนของ HN Change Role เป็น ID กด OK



ภาพ 23 หน้าต่างการเปลี่ยนหน้าที่เป็น ID ของ Primary Key

13. เก็บข้อมูลใน Repository ชุด Training ที่สร้างไว้แล้ว กดปุ่ม Finish



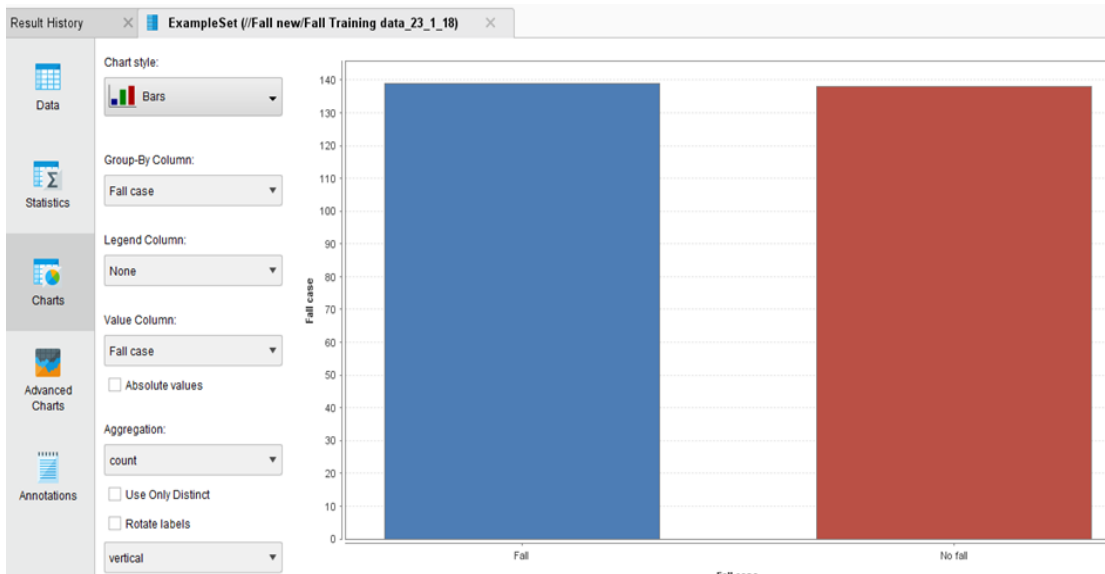
ภาพ 24 หน้าต่างการจัดเก็บข้อมูลใน Repository

14. จะ ได้หน้าต่าข้อมูลทีโหลดเข้าไปแสดงในรูปแบบของตาราง

ExampleSet (280 examples, 2 special attributes, 17 regular attributes) Filter (280 / 280 examples): all

Row No.	HN	Fall ...	Age	LOS...	Sex	Histo...	Use L...	Assi...	Weak	Unstead...	Menta...	Eliminat...	Elimi...	Dizzin...	Pers...	PCAL...	Hypnot...	Seda
1	10138...	0	49	4	0	0	1	0	0	0	0	0	0	0	1	0	0	0
2	10073...	1	68	4	1	0	0	0	1	1	0	0	0	0	0	0	0	0
3	10074...	1	57	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0
4	10229...	0	55	14	1	0	1	0	0	1	0	0	0	0	1	0	0	0
5	10138...	0	83	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0
6	10200...	1	70	7	0	1	1	0	0	1	0	0	0	0	1	0	0	0
7	10229...	1	58	12	1	0	1	0	1	1	0	0	0	0	1	0	1	0
8	10229...	0	54	34	1	0	1	0	0	1	0	0	0	0	1	0	0	0
9	10191...	0	52	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0
10	10228...	1	66	47	1	0	1	0	0	1	0	0	0	0	1	0	0	0
11	10219...	0	29	3	0	0	1	0	0	0	0	0	0	0	1	0	0	0
12	10225...	1	47	267	0	0	1	1	0	1	0	0	0	0	1	0	0	0
13	10130...	0	70	2	1	0	1	0	0	0	0	0	0	0	1	0	0	0

ภาพ 25 หน้าต่าข้อมูลทีโหลดเข้าไปในโปรแกรมในรูปแบบตารางแสดงผลข้อมูล ทีโหลดเข้าไปแสดงในรูปแบบของตาราง



ภาพ 26 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรมในรูปแบบกราฟ

Name	Type	Missing	Statistics	Filter (19 / 19 attributes)
id	Integer	0	Min: 100041460, Max: 102526721, Average: 101779049.171	
Fall case	Binominal	0	Least: 1 (136), Most: 0 (142), Values: 0 (142), 1 (136)	
Age	Integer	0	Min: 15, Max: 95, Average: 53.464	
LOS(day)	Integer	0	Min: 0, Max: 450, Average: 34.682	
Sex	Binominal	0	Least: 0 (131), Most: 1 (149), Values: 1 (149), 0 (131)	
History of fall	Binominal	0	Least: 1 (17), Most: 0 (263), Values: 0 (263), 1 (17)	
Use IV stand	Binominal	0	Least: 0 (19), Most: 1 (261), Values: 1 (261), 0 (19)	

ภาพ 27 หน้าต่างข้อมูลที่ไหลคเข้าไปในโปรแกรมในรูปแบบค่าสถิติ

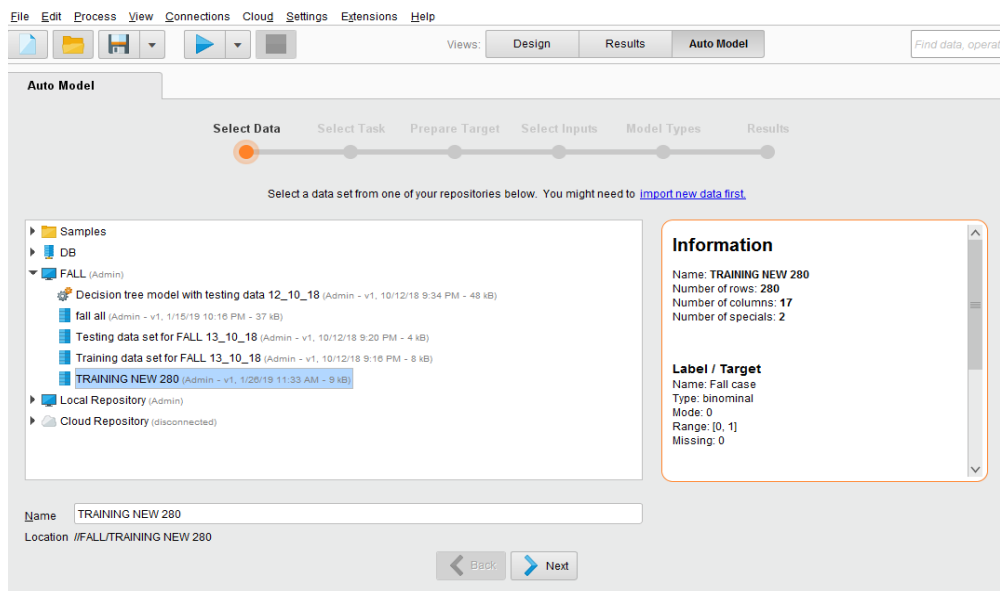
การค้นหารูปแบบ (pattern discovery)

เมื่อผ่านขั้นตอนกระบวนการเตรียมข้อมูลแล้วต่อไปจะเป็นการคัดเลือกแอทริบิวต์ โดยผ่านกระบวนการขั้นตอนการสร้างตัวแบบในโปรแกรม RapidMiner Studio 8.1 ด้วย Auto Model

ขั้นตอนที่ 1 Select Data

จะเป็นขั้นตอนการเลือกไฟล์ที่ต้องการนำมาสร้างตัวแบบที่เป็นข้อมูลชุดเรียนรู้ ซึ่งมีแอทริบิวต์ 2 ประเภท คือ

- 1) แอทริบิวต์ทั่วไป คือ ตัวแปรต่าง ๆ ที่จะใช้ในการสร้างตัวแบบ เปรียบเทียบได้กับตัวแปร X ในทางสถิติ
- 2) แอทริบิวต์ลาเบล คือ ตัวแปรที่เป็นค่าคำตอบที่ต้องการทำนายการพลัดตกหกล้มของผู้ป่วยที่เข้ารับบริการเป็นผู้ป่วยในของโรงพยาบาล หรือเปรียบเทียบกับตัวแปร Y ในทางสถิติ



ภาพ 28 หน้าต่างการเลือกไฟล์ที่ต้องการนำมาสร้างตัวแบบ

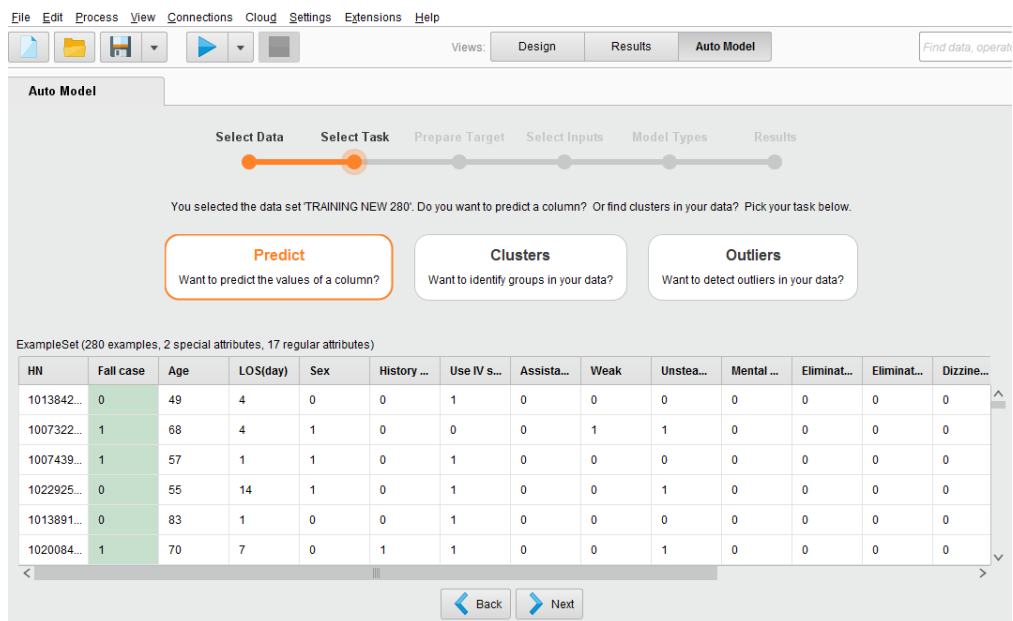
หลังจากที่เลือกข้อมูลแล้วจะแสดงรายละเอียดด้านขวามือว่าข้อมูลที่นำเข้ามามีกี่แถวก็คอลัมน์ (แอทริบิวต์)

ขั้นตอนที่ 2 Select Task

หลังจากเลือกข้อมูลเข้ามาใน โปรแกรมแล้ว จะเป็นขั้นตอนการที่จะเลือกวิเคราะห์ข้อมูลว่าจะเป็นแบบใด ซึ่งใน โปรแกรมนี้มีให้เลือก 3 แบบ ได้แก่

- 1) การทำนาย (predict) เป็นการสร้างตัวแบบเพื่อทำนายคำตอบ
- 2) การจัดกลุ่ม (cluster) เป็นการจัดกลุ่มข้อมูลที่มีลักษณะคล้าย ๆ กันไว้ด้วยกัน
- 3) การหาข้อมูลที่ผิดปกติ (outliers) เป็นการหาข้อมูลที่ผิดปกติ เมื่อเปรียบเทียบกับข้อมูลทั่วไป

ตามภาพด้านล่างจะเป็นการเลือกงานแบบทำนาย (predict) ดังนั้น จึงทำการเลือกที่ Predict และเลือกแอทริบิวต์ที่ต้องการทำนายที่แอทริบิวต์ Fall Case เพื่อเป็นลาเบลสำหรับการสร้างตัวแบบ

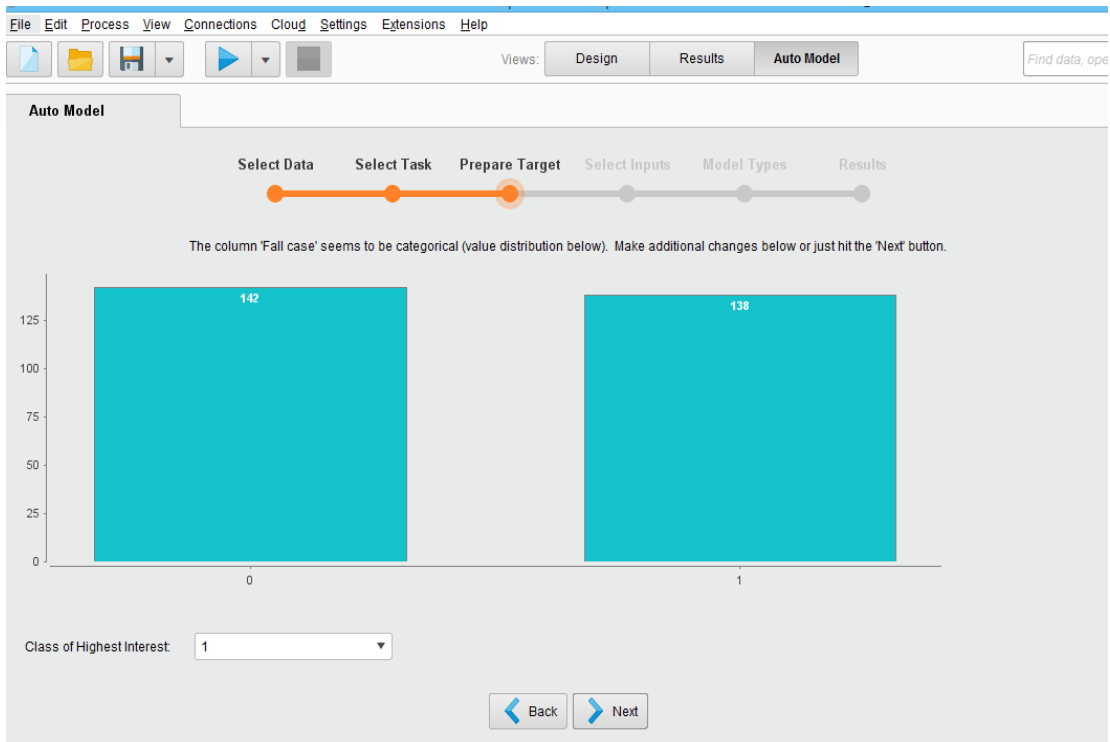


ภาพ 29 หน้าต่างการเลือกวิธีการทำนายและเลือกลาเบล Fall Case

ขั้นตอนที่ 3 Prepare Target

ระบบจะแสดงสัดส่วนของข้อมูลให้เห็น จากภาพจะเป็น 2 ค่า คือ

- 1) Fall แสดงข้อมูลผู้ป่วยในที่มีการพลัดตกหกล้ม
- 2) No Fall แสดงข้อมูลของผู้ป่วยในที่ไม่มีการพลัดตกหกล้ม

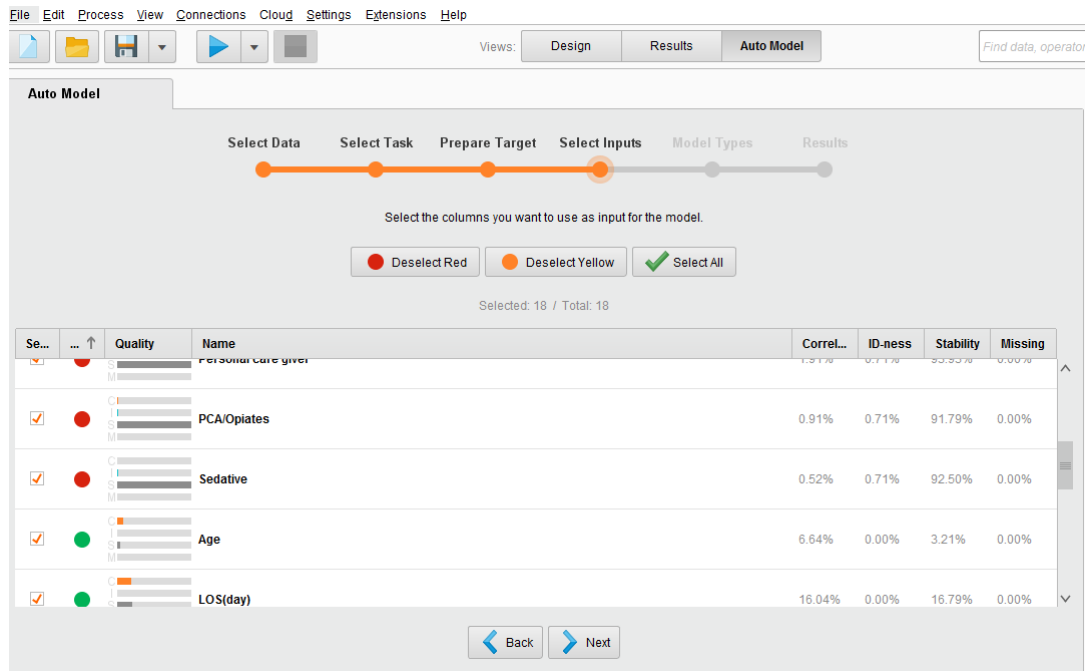


ภาพ 30 หน้าต่างสัดส่วนของข้อมูล Fall และ No Fall

จากรูปจะเห็นว่าคำตอบที่เราสนใจคือการพลัดตกหกล้ม (fall) และระบบเลือกไว้แล้วในส่วนของ Class of Highest Interest

ขั้นตอนที่ 4 Select Inputs

โดยในขั้นตอนนี้จะเป็นการคำนวณความสำคัญของแอทริบิวต์ต่าง ๆ ที่นำเข้ามา และแสดงด้วยสี 3 สี คือ แดง ส้ม และเขียว โดยสีเขียว คือ แอทริบิวต์ที่น่าจะนำไปสร้างตัวแบบได้ดีที่สุด โดยทั่วไปจะตัดแอทริบิวต์ที่มีคุณภาพต่ำ (สถานะสีแดง) ออกไปจากชุดข้อมูล

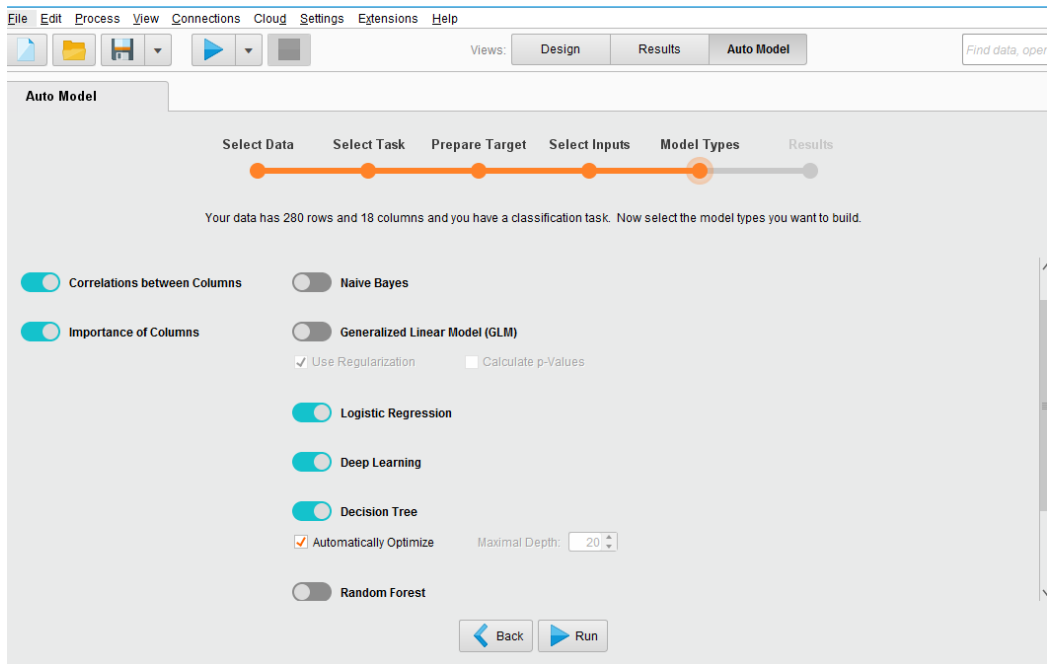


ภาพ 31 หน้าต่างการคำนวณความสำคัญของแอทริบิวต์ที่จะนำไปสร้างตัวแบบ

จากกระบวนการในขั้นตอนนี้ผู้วิจัยได้ทำการเลือกตัวแปรนำเข้าทุกตัวแปร เนื่องจากทุกตัวแปรที่นำเข้าได้รับการวิจัยยืนยันแล้วว่า มีผลต่อการปลดตกหกล้มของผู้ป่วยอย่างมีนัยสำคัญ โดยทำการเลือกที่ปุ่ม Select All ทำให้จำนวนแอทริบิวต์ที่นำเข้ามา มีจำนวน 17 แอทริบิวต์

ขั้นตอนที่ 5 Model Types

ขั้นตอนนี้จะเลือกว่าต้องการที่จะสร้างตัวแบบด้วยเทคนิคอะไร ซึ่งในงานวิจัยนี้ผู้วิจัยได้เลือกวิธีถดถอยโลจิสติก วิธีต้นไม้ตัดสินใจ และวิธีโครงข่ายประสาทเทียมเชิงลึก



ภาพ 32 หน้าต่างขั้นตอนการเลือกตัวแบบที่จะทำการสร้าง

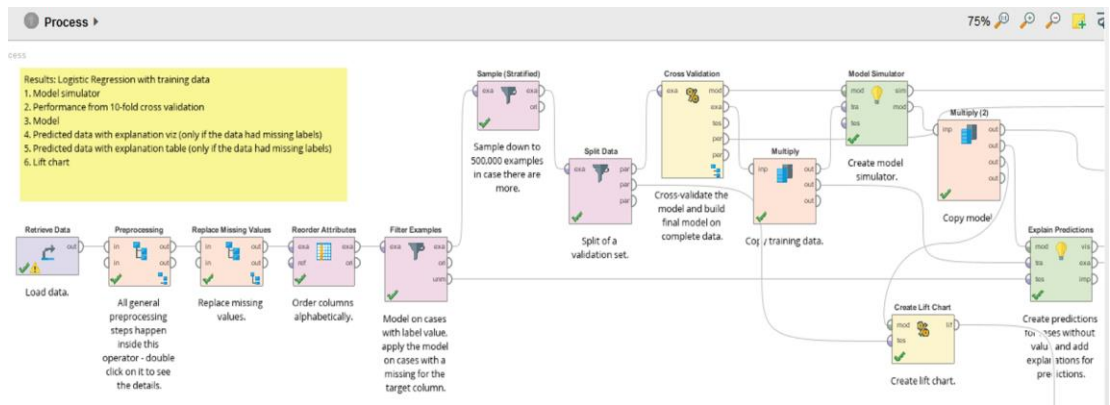
ขั้นตอนที่ 6 Results

เป็นขั้นตอนการแสดงผลการทำงานของตัวแบบชุดข้อมูลเรียนรู้ ที่นำเข้าไปสร้างตัวแบบ เมื่อดำเนินการแล้วเสร็จจะแสดงผลการทำงานเปรียบเทียบในเมนู Overview ในส่วนของ Comparison

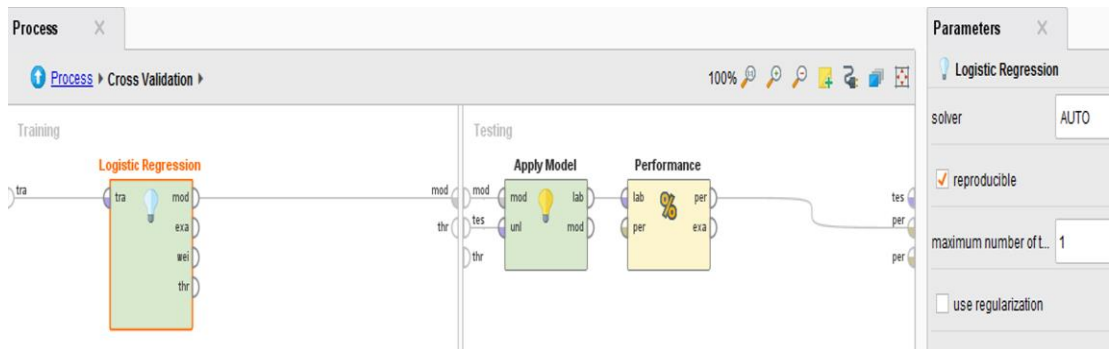
Ro...	HN	Fall...	Age	Assista...	Benzodi...	Dizzine...	Eliminat...	Eliminat...	History ...	Hypnot...	LOS(day)	M
1	101...	0	49	0	0	0	0	0	0	0	4	0
2	100...	1	68	0	0	0	0	0	0	0	4	0
3	100...	1	57	0	0	0	0	0	0	0	1	0
4	102...	0	55	0	0	0	0	0	0	0	14	0
5	101...	0	83	0	0	0	0	0	0	0	1	0
6	102...	1	70	0	0	0	0	0	1	0	7	0
7	102...	1	58	0	1	0	0	0	0	1	12	0
8	102...	0	54	0	0	0	0	0	0	0	34	0
9	101...	0	52	0	0	0	0	0	0	0	1	0
10	102...	1	66	0	0	0	0	0	0	0	47	0

ภาพ 33 หน้าต่างรายงานในส่วนการแสดงผลข้อมูลในรูปแบบตาราง

ในกรณีที่ต้องการแก้ไขในส่วนของ Process สามารถทำได้โดยการคลิกที่ปุ่ม Open Process จะหน้าต่างได้ดังภาพ 34-40 จากนั้น Open Process จากหน้าจอ Auto Model ของการทดสอบตัวแบบวิธีถดถอยโลจิสติก

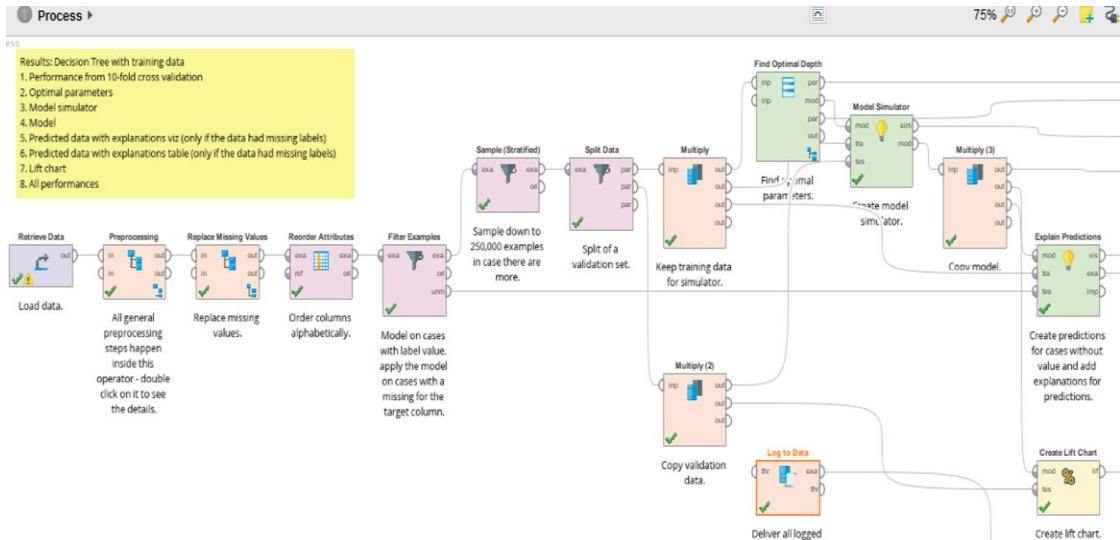


ภาพ 34 หน้าต่างการเชื่อมต่อพอร์ตของโอเปอเรเตอร์ Logistic Regression หน้าจอหลัก จากนั้นกดคดดับเบิลคลิกที่โอเปอเรเตอร์ Cross Validation จะได้ตามภาพ 35



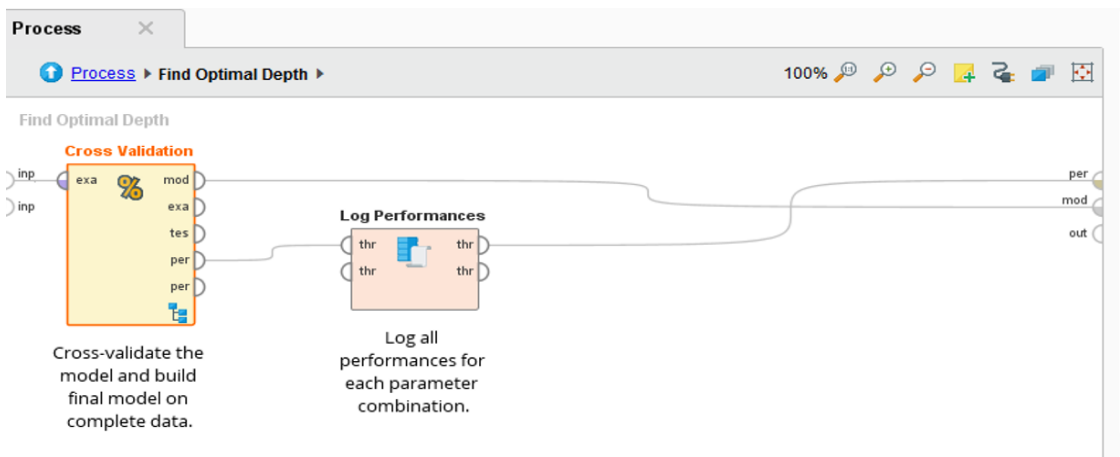
ภาพ 35 หน้าต่างการเชื่อมต่อพอร์ตของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธี Logistic Regression

เชื่อมต่อพอร์ตให้ครบแล้วกด จากนั้นคลิกปุ่ม Run เพื่อตั้งให้ Process ทำงานจะได้หน้าจอตามภาพ 36



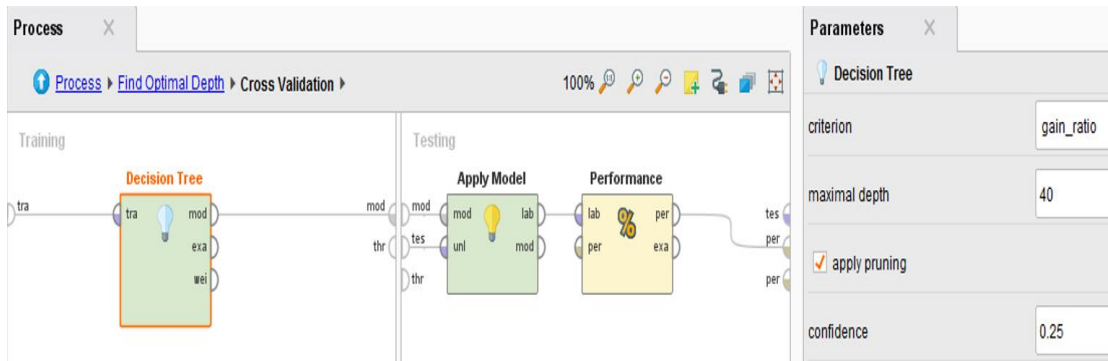
ภาพ 36 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในการหาตัวแบบวิธีต้นไม้ตัดสินใจ หน้าจอหลัก

กดดับเบิลคลิกที่โอเปอเรเตอร์ Find Optimal Depth จะได้ตามภาพ 37




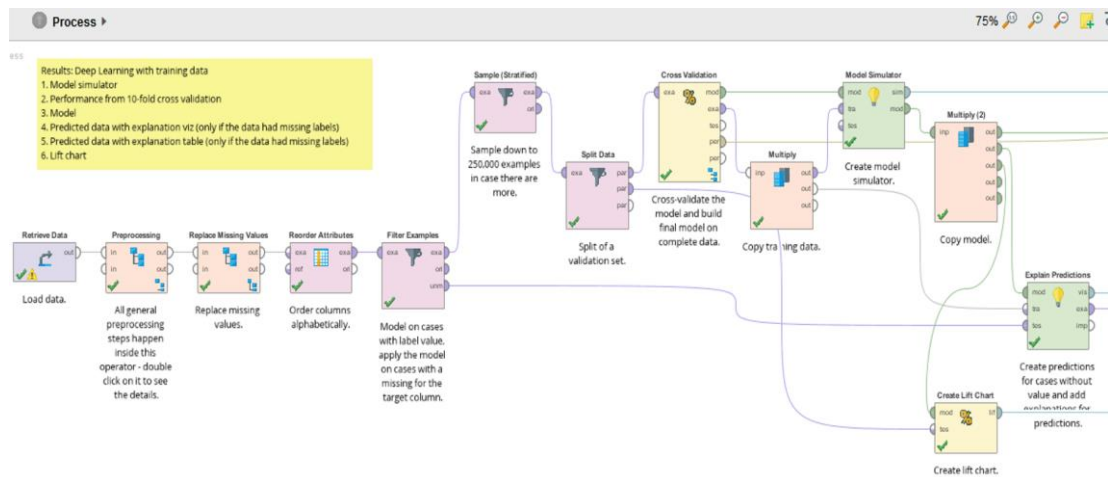
ภาพ 37 หน้าต่างการเชื่อมพอร์ตของโอเปอเรเตอร์ในส่วน Find Optimal Depth ในการทดลองหาตัวแบบวิธีต้นไม้ตัดสินใจ

จากนั้นกดดับเบิลคลิกที่โอเปอเรเตอร์ Cross Validation จะได้ตามภาพ 38



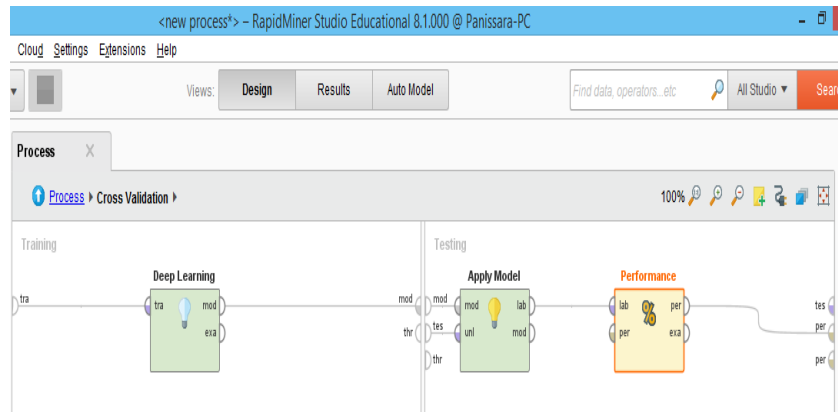
ภาพ 38 หน้าต่างการเชื่อมต่อของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธีต้นไม้ตัดสินใจ

เชื่อมต่อพอร์ตให้ครบแล้วกด จากนั้นคลิกปุ่ม Run  เพื่อสั่งให้ Process ทำงาน



ภาพ 39 หน้าต่างการเชื่อมต่อของโอเปอเรเตอร์ ในการหาตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึกหน้าจอหลัก

กดดับเบิ้ลคลิกที่โอเปอเรเตอร์ Cross Validation จะได้ตามภาพ 40



ภาพ 40 หน้าต่างการเชื่อมต่อพอร์ตของโอเปอเรเตอร์ในส่วน Process ในการทดลองหาตัวแบบวิธี โครงข่ายประสาทเทียมเชิงลึก

เชื่อมต่อพอร์ตให้ครบ จากนั้นคลิกปุ่ม Run  เพื่อสั่งให้ Process ทำงาน

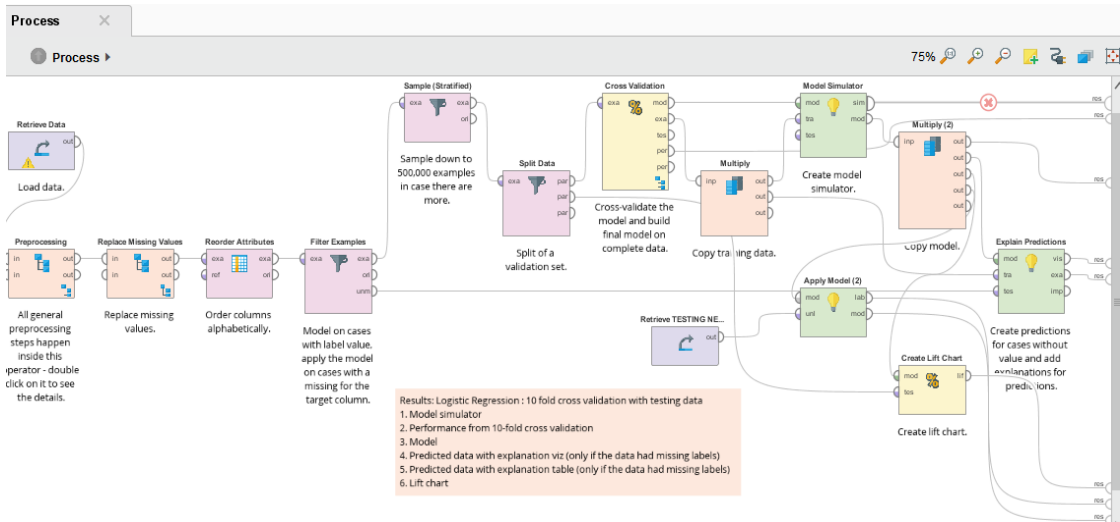
การทดสอบประสิทธิภาพตัวแบบทำนาย

1. นำเข้าสู่ชุดข้อมูลทดสอบของข้อมูลผู้ป่วยใน (inpatient testing data set)

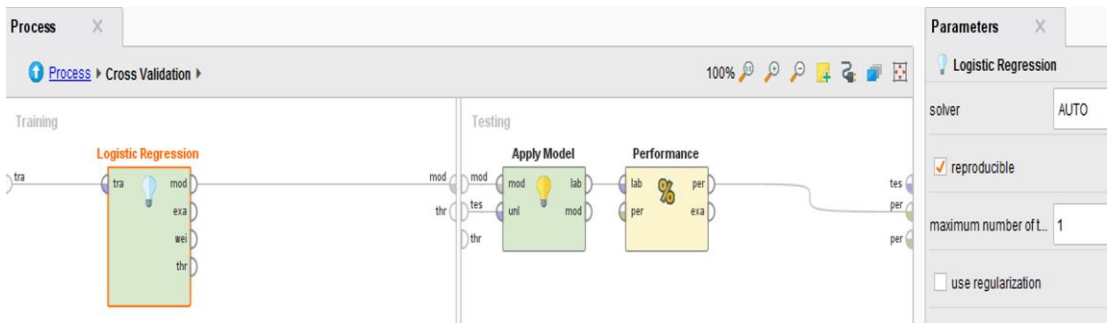
Row...	HN	Fall ca...	Age	LOS(day)	Sex	History of fall	Use IV stand	Assistance ...	Weak	Unsteady gait	Mental statu...	Elimination ...
1	1017...	?	29	39	1	0	1	0	1	0	0	1
2	1006...	?	66	1	1	0	1	0	0	0	0	0
3	1007...	?	31	2	0	0	1	0	0	0	0	0
4	1018...	?	73	1	1	0	1	0	0	1	0	0
5	1025...	?	59	11	1	0	1	0	0	1	0	0
6	1002...	?	66	2	1	0	1	0	0	0	0	0
7	1002...	?	45	1	1	0	1	1	0	1	0	0
8	1019...	?	73	50	1	0	1	0	1	1	0	0
9	1002...	?	83	1	0	0	1	0	0	0	0	1
10	1025...	?	73	209	0	0	1	0	0	0	0	0
11	1011...	?	72	2	1	1	1	0	0	0	0	0
12	1025...	?	55	83	1	0	1	0	0	0	0	0
13	1007...	?	52	3	0	0	0	0	0	0	0	0
14	1025...	?	69	25	0	0	1	0	0	0	0	1
15	1014...	?	71	19	1	0	1	0	0	0	0	1
16	1024...	?	59	2	1	0	1	0	0	0	0	0

ภาพ 41 หน้าต่าง Example Set ของชุดข้อมูลทดสอบที่ยังไม่ทราบคลาสที่นำเข้ามา

2. จากนั้นใช้โอเปอเรเตอร์ Retrieve ดึงข้อมูลชุดทดสอบและนำโอเปอเรเตอร์ Apply Model มาวางไว้ที่บริเวณ Process และต่อพอร์ต Output จากโอเปอเรเตอร์ Multiply (2) มาที่ mod ของ Apply Model (2) และต่อออกพอร์ต Result ตามภาพ 42



ภาพ 42 หน้าต่างการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบ



ภาพ 43 หน้าต่างการทดสอบตัวแบบด้วยชุดข้อมูลทดสอบ

3. กำหนดในส่วนของ Parameter ของโอเปอเรเตอร์ Performance โดยเลือกค่าใน Parameter เป็น Accuracy, Classification Error, AUC, Precision, Recall, F-measure, Sensitivity และ Specificity

จากนั้นคลิกปุ่ม Run  เพื่อสั่งให้ Process ทำงาน

ภาคผนวก ข

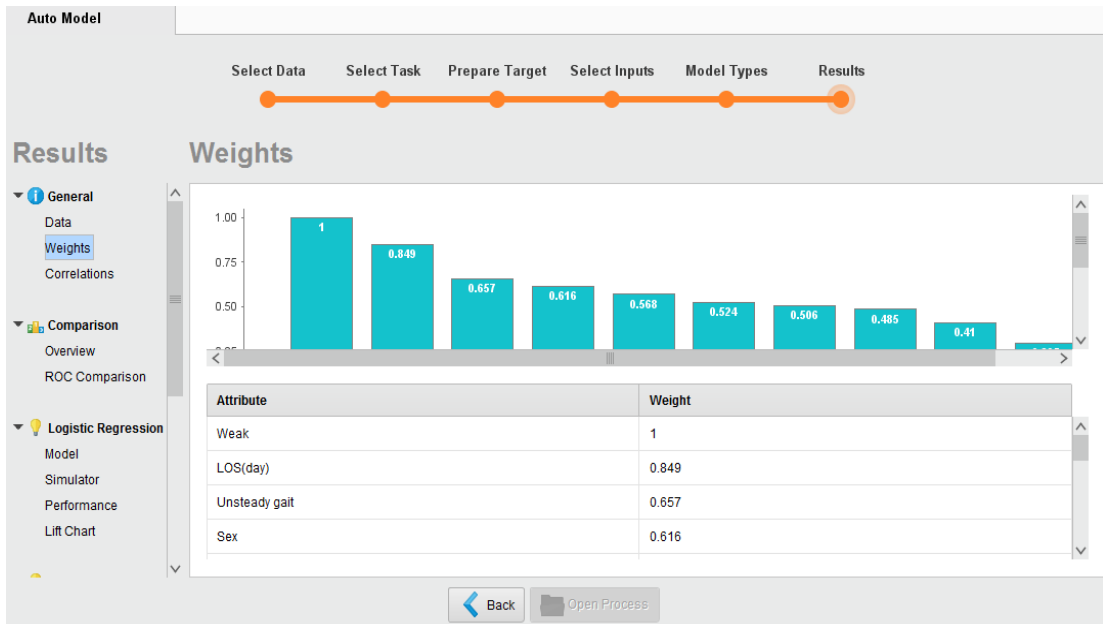
ผลการวิเคราะห์ข้อมูลจากโปรแกรม RapidMiner Studio 8.1

ผลการวิเคราะห์รูปแบบ (pattern analysis)

การวิจัยนี้ทำการทดสอบตัวแบบโดยใช้โปรแกรมประยุกต์ RapidMiner Studio 8.1 ชุดข้อมูลที่เรียนรู้ที่นำมาใช้ในการสร้างตัวแบบทดสอบมีจำนวน 280 เรคคอร์ด ซึ่งในการทดสอบตัวแบบนั้น ได้ใช้วิธีถดถอยโลจิสติก (logistic regression) วิธีต้นไม้ตัดสินใจ (decision tree) และวิธีโครงข่ายประสาทเทียมเชิงลึก (deep learning) และใช้วิธีการทดสอบประสิทธิภาพด้วยวิธี 10-fold Cross Validation Test โดยมีเกณฑ์ที่ใช้ในการพิจารณาคุณภาพของตัวแบบที่ได้จากแต่ละอัลกอริทึม จะพิจารณาจากเปอร์เซ็นต์ของ ค่า Accuracy, Classification Error , AUC, Precision, Recall, F-measure, Sensitivity, Specificity หลังจากทำการทดสอบประสิทธิภาพของตัวแบบจะได้ผลรายงานมีรายละเอียดดังนี้

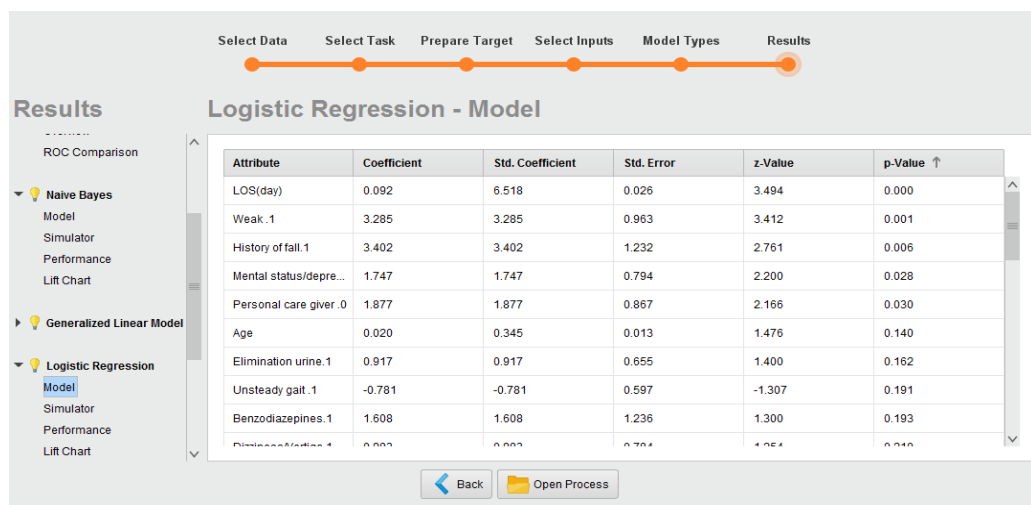
Ro...	HN	Fall...	Age	Assista...	Benzodi...	Dizzine...	Eliminat...	Eliminat...	History ...	Hypnoti...	LOS(day)	M
1	101...	0	49	0	0	0	0	0	0	0	4	0
2	100...	1	68	0	0	0	0	0	0	0	4	0
3	100...	1	57	0	0	0	0	0	0	0	1	0
4	102...	0	55	0	0	0	0	0	0	0	14	0
5	101...	0	83	0	0	0	0	0	0	0	1	0
6	102...	1	70	0	0	0	0	0	1	0	7	0
7	102...	1	58	0	1	0	0	0	0	1	12	0
8	102...	0	54	0	0	0	0	0	0	0	34	0
9	101...	0	52	0	0	0	0	0	0	0	1	0
10	102...	1	66	0	0	0	0	0	0	0	47	0

ภาพ 44 หน้าต่างรายงานผลชุดข้อมูลเรียนรู้ที่นำเข้าไปวิเคราะห์ในส่วนของ Data ที่ได้รับการเปลี่ยน (transformed) สำหรับเพื่อสร้างตัวแบบ

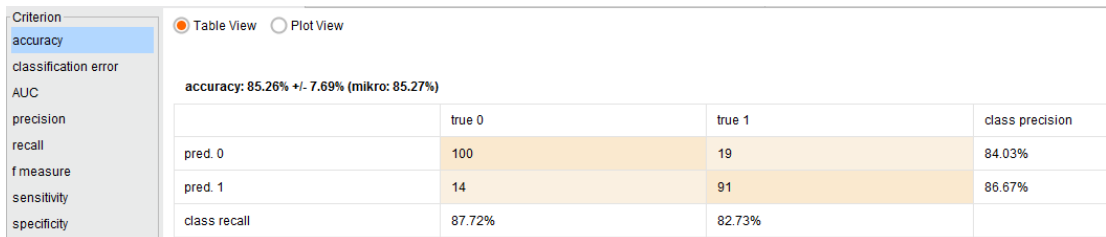


ภาพ 45 หน้าต่างรายงานผลค่าน้ำหนัก (weight) ของตัวแปรนำเข้า (input) ของชุดข้อมูลเรียนรู้ที่นำเข้าไปวิเคราะห์

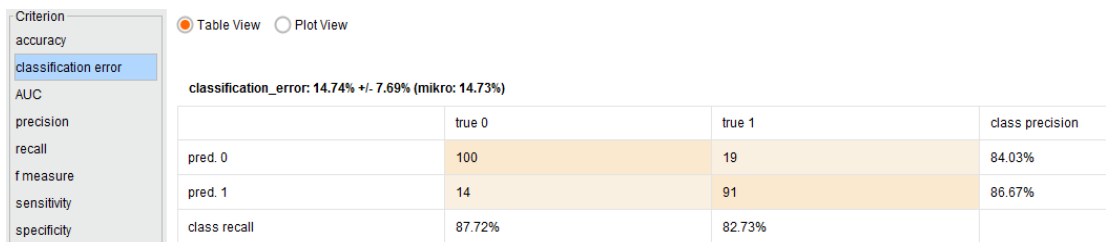
รายงานผลการวิเคราะห์ตัวแบบวิธีถดถอยโลจิสติกมีรายละเอียดของรายงานผลของตัวแบบได้ตามภาพ 46 - ภาพ 54



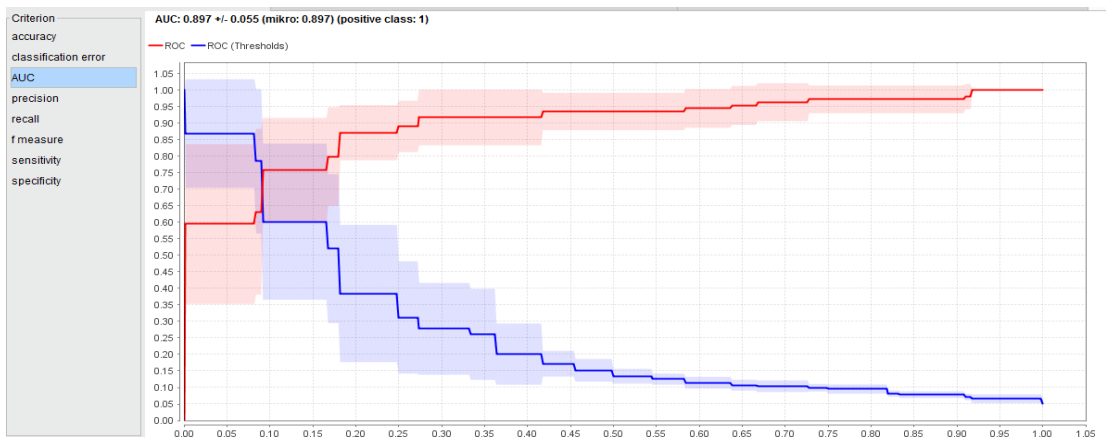
ภาพ 46 หน้าต่างแสดงหน้าต่างรายงานตัวแบบของวิธีถดถอยโลจิสติก



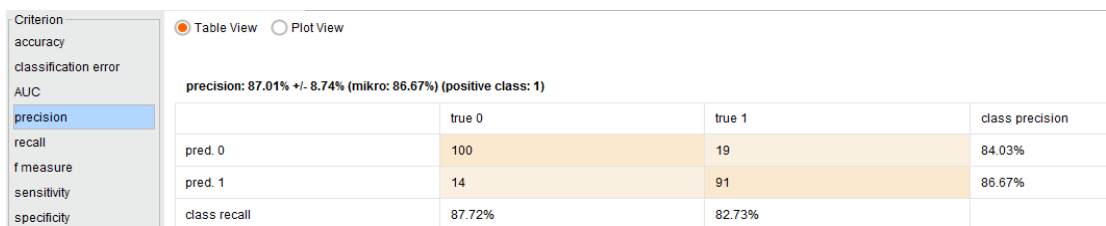
ภาพ 47 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบวิธีถดถอยโลจิสติก



ภาพ 48 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบวิธีถดถอยโลจิสติก



ภาพ 49 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบวิธีถดถอยโลจิสติก



ภาพ 50 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบวิธีถดถอยโลจิสติก

Criterion: Table View Plot View

accuracy
classification error
AUC
precision
recall
f measure
sensitivity
specificity

recall: 82.59% +/- 13.17% (mikro: 82.73%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	100	19	84.03%
pred. 1	14	91	86.67%
class recall	87.72%	82.73%	

ภาพ 51 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบวิธีถดถอยโลจิสติก

Criterion: Table View Plot View

accuracy
classification error
AUC
precision
recall
f measure
sensitivity
specificity

f_measure: 84.12% +/- 8.71% (mikro: 84.65%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	100	19	84.03%
pred. 1	14	91	86.67%
class recall	87.72%	82.73%	

ภาพ 52 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบวิธีถดถอยโลจิสติก

Criterion: Table View Plot View

accuracy
classification error
AUC
precision
recall
f measure
sensitivity
specificity

sensitivity: 82.59% +/- 13.17% (mikro: 82.73%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	100	19	84.03%
pred. 1	14	91	86.67%
class recall	87.72%	82.73%	

ภาพ 53 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบวิธีถดถอยโลจิสติก

Criterion: Table View Plot View

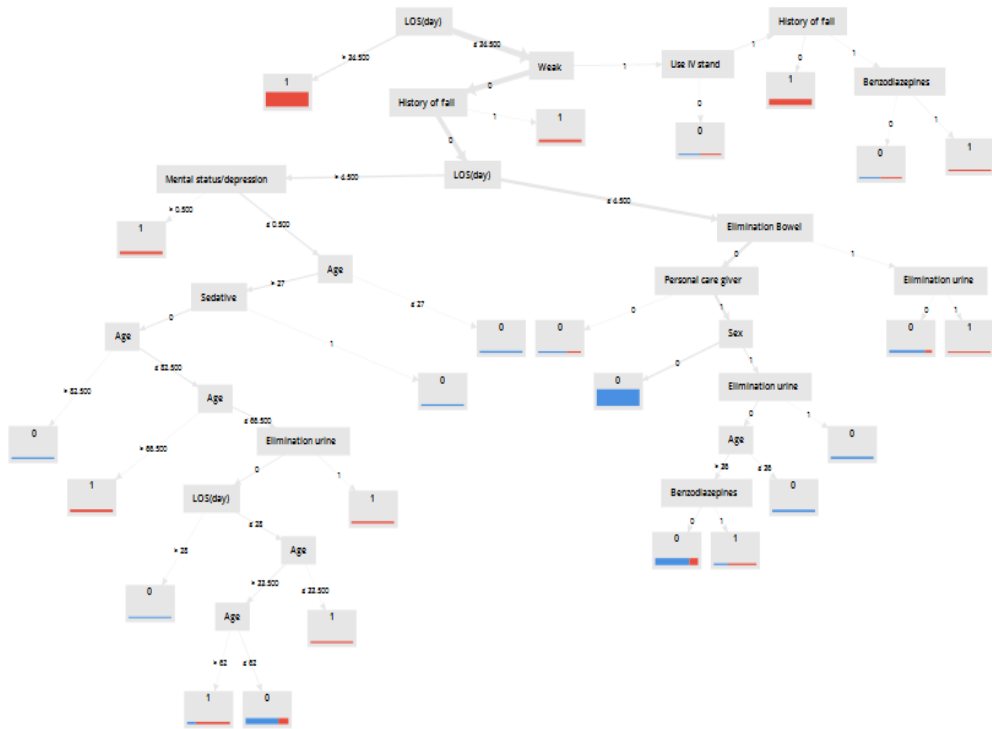
accuracy
classification error
AUC
precision
recall
f measure
sensitivity
specificity

specificity: 87.80% +/- 8.76% (mikro: 87.72%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	100	19	84.03%
pred. 1	14	91	86.67%
class recall	87.72%	82.73%	

ภาพ 54 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบวิธีถดถอยโลจิสติก

รายงานผลการวิเคราะห์ตัวแบบวิธีอื่นไม่ได้ตัดสินใจ มีรายละเอียดของรายงานผลของตัวแบบได้ตามภาพ 55 - ภาพ 63



ภาพ 55 หน้าต่างตัวแบบของวิธีต้นไม้ตัดสินใจในรูปแบบต้นไม้

Criterion: accuracy (selected), classification error, AUC, precision, recall, f measure, sensitivity, specificity

Table View (selected) Plot View

accuracy: 83.44% +/- 6.08% (mikro: 83.48%)

	true 0	true 1	class precision
pred. 0	96	19	83.48%
pred. 1	18	91	83.49%
class recall	84.21%	82.73%	

ภาพ 56 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบวิธีต้นไม้ตัดสินใจ

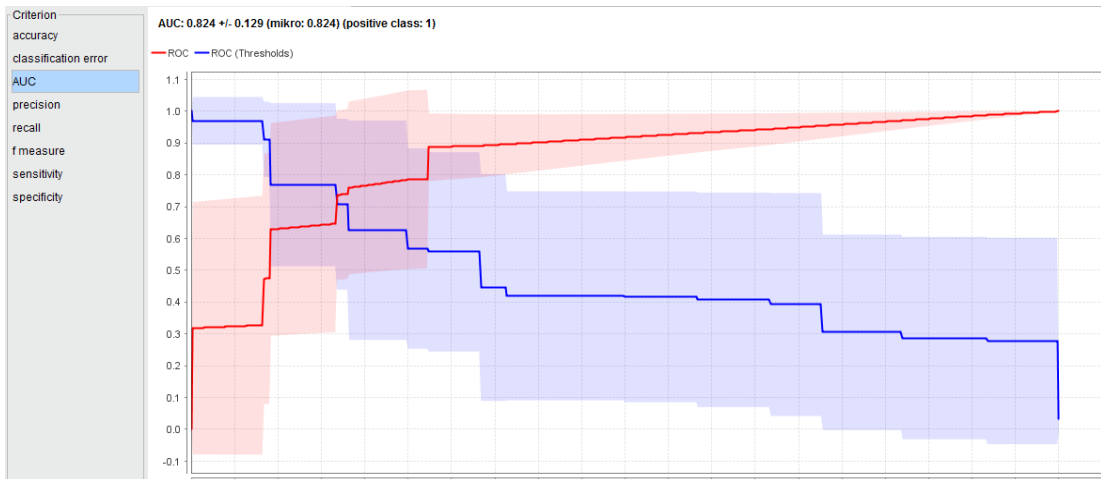
Criterion: accuracy, classification error (selected), AUC, precision, recall, f measure, sensitivity, specificity

Table View (selected) Plot View

classification_error: 16.56% +/- 6.08% (mikro: 16.52%)

	true 0	true 1	class precision
pred. 0	96	19	83.48%
pred. 1	18	91	83.49%
class recall	84.21%	82.73%	

ภาพ 57 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบวิธีต้นไม้ตัดสินใจ



ภาพ 58 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบวิธีต้นไม้ตัดสินใจ

Criterion: accuracy, classification error, AUC, precision, recall, f measure, sensitivity, specificity

Table View Plot View

precision: 84.88% +/- 9.67% (mikro: 83.49%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	96	19	83.48%
pred. 1	18	91	83.49%
class recall	84.21%	82.73%	

ภาพ 59 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบวิธีต้นไม้ตัดสินใจ

Criterion: accuracy, classification error, AUC, precision, recall, f measure, sensitivity, specificity

Table View Plot View

recall: 82.86% +/- 14.73% (mikro: 82.73%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	96	19	83.48%
pred. 1	18	91	83.49%
class recall	84.21%	82.73%	

ภาพ 60 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบวิธีต้นไม้ตัดสินใจ

Criterion: accuracy, classification error, AUC, precision, recall, f measure, sensitivity, specificity

Table View Plot View

f_measure: 82.56% +/- 7.41% (mikro: 83.11%) (positive class: 1)

	true 0	true 1	class precision
pred. 0	96	19	83.48%
pred. 1	18	91	83.49%
class recall	84.21%	82.73%	

ภาพ 61 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบวิธีต้นไม้ตัดสินใจ

Criterion		<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
accuracy	sensitivity: 82.86% +/- 14.73% (mikro: 82.73%) (positive class: 1)			
classification error				
AUC				
precision				
recall				
f measure				
sensitivity		true 0	true 1	class precision
specificity				
	pred. 0	96	19	83.48%
	pred. 1	18	91	83.49%
	class recall	84.21%	82.73%	

ภาพ 62 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบวิธีต้นไม้ตัดสินใจ

Criterion		<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
accuracy	specificity: 84.32% +/- 11.70% (mikro: 84.21%) (positive class: 1)			
classification error				
AUC				
precision				
recall				
f measure				
sensitivity		true 0	true 1	class precision
specificity				
	pred. 0	96	19	83.48%
	pred. 1	18	91	83.49%
	class recall	84.21%	82.73%	

ภาพ 63 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบวิธีต้นไม้ตัดสินใจ

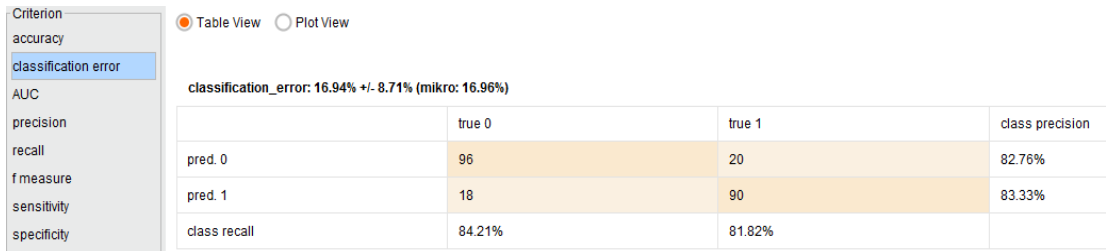
รายงานผลการวิเคราะห์ตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึก มีรายละเอียดของรายงานผลของตัวแบบได้ตามภาพ 64 - ภาพ 72

Description		DeepLearning																																																																																																																																													
Annotations		Model Metrics Type: Binomial Description: Metrics reported on full training frame model id: rm-h2o-model-deep_learning-133427 frame id: rm-h2o-frame-deep_learning-750630 MSE: 0.09251916 R ² : 0.6298053 AUC: 0.94170654 logloss: 0.3095317 CM: Confusion Matrix (vertical: actual; across: predicted): 0 1 Error Rate 0 106 8 0.0702 = 8 / 114 1 13 97 0.1182 = 13 / 110 Totals 119 105 0.0938 = 21 / 224 Gains/Lift Table (Avg response rate: 49.11 %): <table border="1"> <thead> <tr> <th>Group</th> <th>Cumulative Data Fraction</th> <th>Lower Threshold</th> <th>Lift</th> <th>Cumulative Lift</th> <th>Response Rate</th> <th>Cumulative Response Rate</th> <th>Capture Rate</th> <th>Cumulative Capture Rate</th> <th>Gain</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.01339286</td><td>0.999999</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.027273</td><td>0.027273</td><td>103.63636</td></tr> <tr><td>2</td><td>0.02232143</td><td>0.999997</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.018182</td><td>0.045455</td><td>103.63636</td></tr> <tr><td>3</td><td>0.03125000</td><td>0.999993</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.018182</td><td>0.063636</td><td>103.63636</td></tr> <tr><td>4</td><td>0.04017857</td><td>0.999989</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.018182</td><td>0.081818</td><td>103.63636</td></tr> <tr><td>5</td><td>0.05357143</td><td>0.999949</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.027273</td><td>0.109091</td><td>103.63636</td></tr> <tr><td>6</td><td>0.10267857</td><td>0.999079</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.100000</td><td>0.209091</td><td>103.63636</td></tr> <tr><td>7</td><td>0.15178571</td><td>0.981260</td><td>2.036364</td><td>2.036364</td><td>1.000000</td><td>1.000000</td><td>0.100000</td><td>0.309091</td><td>103.63636</td></tr> <tr><td>8</td><td>0.20089286</td><td>0.868955</td><td>1.851240</td><td>1.991111</td><td>0.909091</td><td>0.977778</td><td>0.090909</td><td>0.400000</td><td>85.12396</td></tr> <tr><td>9</td><td>0.29910714</td><td>0.934326</td><td>2.036364</td><td>2.005970</td><td>1.000000</td><td>0.985075</td><td>0.200000</td><td>0.600000</td><td>103.63636</td></tr> <tr><td>10</td><td>0.40178571</td><td>0.768846</td><td>1.770751</td><td>1.945859</td><td>0.869565</td><td>0.955556</td><td>0.181818</td><td>0.781818</td><td>77.07503</td></tr> <tr><td>11</td><td>0.50000000</td><td>0.513251</td><td>1.110744</td><td>1.781818</td><td>0.545455</td><td>0.875000</td><td>0.109091</td><td>0.890909</td><td>11.07438</td></tr> <tr><td>12</td><td>0.59821429</td><td>0.301102</td><td>0.370248</td><td>1.550068</td><td>0.181818</td><td>0.761194</td><td>0.036364</td><td>0.927273</td><td>-62.97520</td></tr> </tbody> </table>												Group	Cumulative Data Fraction	Lower Threshold	Lift	Cumulative Lift	Response Rate	Cumulative Response Rate	Capture Rate	Cumulative Capture Rate	Gain	1	0.01339286	0.999999	2.036364	2.036364	1.000000	1.000000	0.027273	0.027273	103.63636	2	0.02232143	0.999997	2.036364	2.036364	1.000000	1.000000	0.018182	0.045455	103.63636	3	0.03125000	0.999993	2.036364	2.036364	1.000000	1.000000	0.018182	0.063636	103.63636	4	0.04017857	0.999989	2.036364	2.036364	1.000000	1.000000	0.018182	0.081818	103.63636	5	0.05357143	0.999949	2.036364	2.036364	1.000000	1.000000	0.027273	0.109091	103.63636	6	0.10267857	0.999079	2.036364	2.036364	1.000000	1.000000	0.100000	0.209091	103.63636	7	0.15178571	0.981260	2.036364	2.036364	1.000000	1.000000	0.100000	0.309091	103.63636	8	0.20089286	0.868955	1.851240	1.991111	0.909091	0.977778	0.090909	0.400000	85.12396	9	0.29910714	0.934326	2.036364	2.005970	1.000000	0.985075	0.200000	0.600000	103.63636	10	0.40178571	0.768846	1.770751	1.945859	0.869565	0.955556	0.181818	0.781818	77.07503	11	0.50000000	0.513251	1.110744	1.781818	0.545455	0.875000	0.109091	0.890909	11.07438	12	0.59821429	0.301102	0.370248	1.550068	0.181818	0.761194	0.036364	0.927273	-62.97520
Group	Cumulative Data Fraction	Lower Threshold	Lift	Cumulative Lift	Response Rate	Cumulative Response Rate	Capture Rate	Cumulative Capture Rate	Gain																																																																																																																																						
1	0.01339286	0.999999	2.036364	2.036364	1.000000	1.000000	0.027273	0.027273	103.63636																																																																																																																																						
2	0.02232143	0.999997	2.036364	2.036364	1.000000	1.000000	0.018182	0.045455	103.63636																																																																																																																																						
3	0.03125000	0.999993	2.036364	2.036364	1.000000	1.000000	0.018182	0.063636	103.63636																																																																																																																																						
4	0.04017857	0.999989	2.036364	2.036364	1.000000	1.000000	0.018182	0.081818	103.63636																																																																																																																																						
5	0.05357143	0.999949	2.036364	2.036364	1.000000	1.000000	0.027273	0.109091	103.63636																																																																																																																																						
6	0.10267857	0.999079	2.036364	2.036364	1.000000	1.000000	0.100000	0.209091	103.63636																																																																																																																																						
7	0.15178571	0.981260	2.036364	2.036364	1.000000	1.000000	0.100000	0.309091	103.63636																																																																																																																																						
8	0.20089286	0.868955	1.851240	1.991111	0.909091	0.977778	0.090909	0.400000	85.12396																																																																																																																																						
9	0.29910714	0.934326	2.036364	2.005970	1.000000	0.985075	0.200000	0.600000	103.63636																																																																																																																																						
10	0.40178571	0.768846	1.770751	1.945859	0.869565	0.955556	0.181818	0.781818	77.07503																																																																																																																																						
11	0.50000000	0.513251	1.110744	1.781818	0.545455	0.875000	0.109091	0.890909	11.07438																																																																																																																																						
12	0.59821429	0.301102	0.370248	1.550068	0.181818	0.761194	0.036364	0.927273	-62.97520																																																																																																																																						

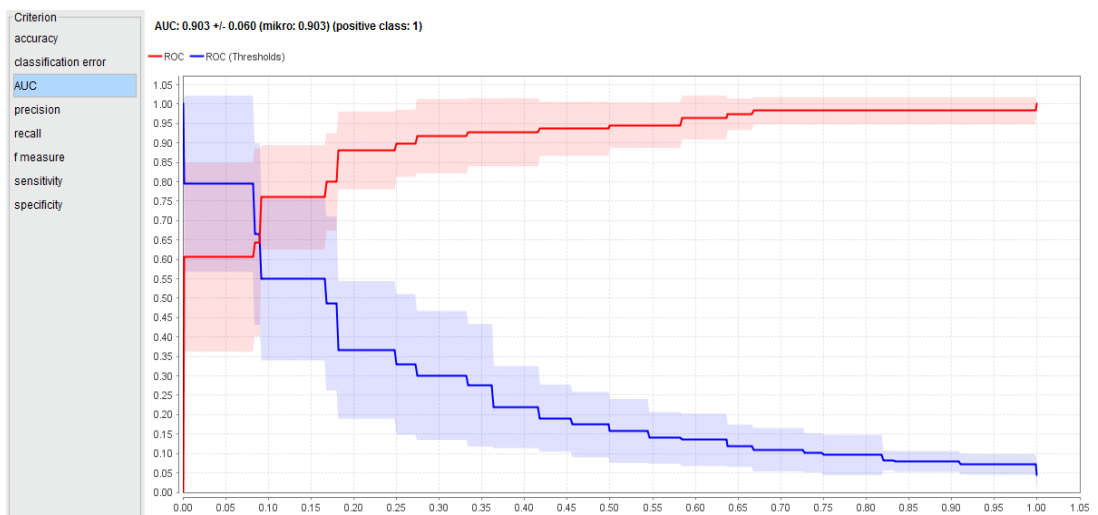
ภาพ 64 หน้าต่างรายงานรายละเอียดของตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึก

Criterion		<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
accuracy	accuracy: 83.06% +/- 8.71% (mikro: 83.04%)			
classification error				
AUC				
precision				
recall				
f measure				
sensitivity		true 0	true 1	class precision
specificity				
	pred. 0	96	20	82.76%
	pred. 1	18	90	83.33%
	class recall	84.21%	81.82%	

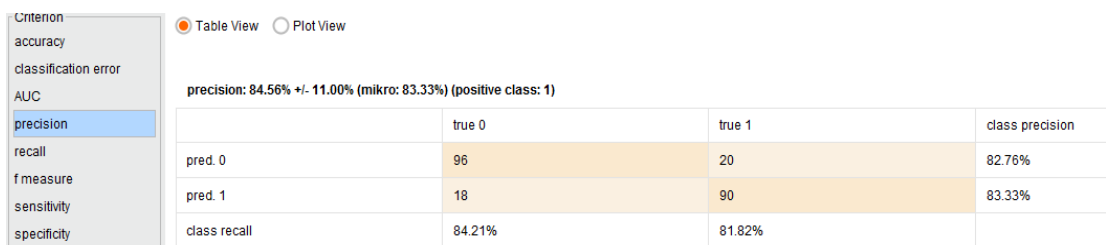
ภาพ 65 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Accuracy ของตัวแบบวิธีโครงข่ายประสาทเทียมเชิงลึก



ภาพ 66 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Classification Error ของตัวแบบวิธี
โครงข่ายประสาทเทียมเชิงลึก



ภาพ 67 หน้าต่างรายงานประสิทธิภาพค่า AUC ของตัวแบบวิธีโครงข่ายประสาทเทียม
เชิงลึก



ภาพ 68 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Precision ของตัวแบบวิธีโครงข่าย
ประสาทเทียมเชิงลึก

Criterion			
accuracy	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
classification error	recall: 81.50% +/- 13.77% (mikro: 81.82%) (positive class: 1)		
AUC	true 0	true 1	class precision
precision	pred. 0	96	20
recall	pred. 1	18	90
f measure	class recall	84.21%	81.82%
sensitivity			82.76%
specificity			83.33%

ภาพ 69 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Recall ของตัวแบบวิธี โครงข่ายประสาทเทียมเชิงลึก

Criterion			
accuracy	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
classification error	f_measure: 82.15% +/- 9.24% (mikro: 82.57%) (positive class: 1)		
AUC	true 0	true 1	class precision
precision	pred. 0	96	20
recall	pred. 1	18	90
f measure	class recall	84.21%	81.82%
sensitivity			82.76%
specificity			83.33%

ภาพ 70 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ F-measure ของตัวแบบวิธี โครงข่ายประสาทเทียมเชิงลึก

Criterion			
accuracy	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
classification error	sensitivity: 81.50% +/- 13.77% (mikro: 81.82%) (positive class: 1)		
AUC	true 0	true 1	class precision
precision	pred. 0	96	20
recall	pred. 1	18	90
f measure	class recall	84.21%	81.82%
sensitivity			82.76%
specificity			83.33%

ภาพ 71 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Sensitivity ของตัวแบบวิธี โครงข่ายประสาทเทียมเชิงลึก

Criterion			
accuracy	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
classification error	specificity: 84.24% +/- 12.58% (mikro: 84.21%) (positive class: 1)		
AUC	true 0	true 1	class precision
precision	pred. 0	96	20
recall	pred. 1	18	90
f measure	class recall	84.21%	81.82%
sensitivity			82.76%
specificity			83.33%

ภาพ 72 หน้าต่างรายงานประสิทธิภาพค่าร้อยละ Specificity ของตัวแบบวิธี โครงข่ายประสาทเทียมเชิงลึก

บรรณานุกรม

- จันทร์ศิริ มีดี. (2549). ปัจจัยเกี่ยวข้องต่อการหกล้มในผู้ป่วยโรคสมองเสื่อมตามการรับรู้ของพยาบาลในโรงพยาบาลสวนปรุง จังหวัดเชียงใหม่. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยเชียงใหม่.
- ธารทิพย์ คนตรง, นงนุช แยมวงษ์, นฤมล คงสาคร และสายรุ้ง ดีนก. (2553). การพัฒนาแนวทางการประเมินภาวะเสี่ยงเพื่อป้องกันการพลัดตกหกล้มในหอผู้ป่วย. กรุงเทพมหานคร: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- นิพา ศรีช้าง และลลิตรา กำวิ. (2559). การพยากรณ์การพลัดตกหกล้มของผู้สูงอายุ (อายุ 60 ปีขึ้นไป) ในประเทศไทย ปี พ.ศ. 2560-2564. นนทบุรี: กระทรวงสาธารณสุข, กรมควบคุมโรค, สำนักโรคไม่ติดต่อ.
- บุปผา จันทร์จรัส. (2546). การพัฒนาแบบประเมินภาวะเสี่ยงต่อการหกล้มในผู้ป่วยออร์โธปิดิกส์. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยเชียงใหม่.
- ปณณวิชญ์ วงศ์วิวัฒนานนท์. (2551). การทำเหมืองข้อมูลกับสารสนเทศทางการแพทย์ [สื่ออิเล็กทรอนิกส์]. วารสารการจัดการข้อมูลและชีวิตวิถี, 4(2), 12-16.
- ปรียดา พิมพ์ภูลาด. (2556). ผลการนำแบบประเมินปัจจัยเสี่ยงต่อการหกล้มไปทดลองใช้ในผู้สูงอายุ หอผู้ป่วยเวชกรรมฟื้นฟู โรงพยาบาลอุดรธานี. การศึกษาอิสระ-พยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยขอนแก่น.
- พัชรินทร์ คณะพล. (2557). ความชุกและปัจจัยทำนายการเกิดการหกล้มของผู้ป่วยโรคหลอดเลือดสมองหลังการจำหน่ายจากโรงพยาบาล. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยขอนแก่น.
- พิมพ์วิรินทร์ ลีสุขสันต์. (2551). ปัจจัยทำนายการเกิดการพลัดตกหกล้มของผู้สูงอายุ. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์.
- มนตรี พิริยะกุล. (2545). เทคนิคการวิเคราะห์สมการถดถอย (Regression Analysis) (ฉบับปรับปรุง). กรุงเทพมหานคร: ผู้แต่ง.

- รัตนานพร ทามิ, พวงทิพย์ ชัยพิบาลสถิตย์ และปานตา อภิรักษ์นภานนท์. (2561). ความสัมพันธ์ระหว่างปัจจัยคัดสรรกับการปลัดตกหกล้มของผู้ป่วยในโรงพยาบาลเอกชน. *วารสารการพยาบาล*, 20(1), 12-26.
- ลัดดา เถียมวงศ์ และเรวดี เพชรศิริสัณฑ์. (2552). ปัจจัยเสี่ยงต่อการหกล้มในผู้สูงอายุไทยที่อาศัยอยู่ในบ้านพักคนชรา. *วารสารสภาการพยาบาล*, 24(1), 77-87.
- วิภา หาทอน. (2556). การพัฒนาแนวปฏิบัติการพยาบาลเพื่อป้องกันและดูแลหลังหกล้มในผู้ป่วยสูงอายุที่มีความบกพร่องด้านการมองเห็น หอผู้ป่วยจักษุ โรงพยาบาลขอนแก่น. การศึกษาอิสระพยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยขอนแก่น.
- สุชาดา กิระนันท์. (2545). “คำถามคำตอบเกี่ยวกับการทำเหมืองข้อมูลเบื้องต้น”. คัดจากหนังสือ “สัมมนาหนึ่งทศวรรษไอทีจุฬา: การเพิ่มศักยภาพระบบการจัดการด้วยเทคโนโลยีสารสนเทศ”. ค้นเมื่อ 19 มิถุนายน 2560, จาก https://home.kku.ac.th/wichuda/Knowledge/6DataMining/Datamining_Suchada.pdf
- ศุรวีชร ศรีเปารยะ และสายชล สีนสมบุญธทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง: กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย [สื่ออิเล็กทรอนิกส์]. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(5), 839-853.
- หทัยรัตน์ เกตุมณีชัยรัตน์. (2014). การจำแนกประเภทข้อมูล 1 (*Decision tree and rule base classification*). ค้นเมื่อ 19 มิถุนายน 2560, จาก <https://www.slideshare.net/NontawatB/05-classification-1-decision-tree-and-rule-based-classification>
- หทัยรัตน์ เกตุมณีชัยรัตน์. (2014). การจำแนกประเภทข้อมูล 3 (*Neural Network*). ค้นเมื่อ 1 มิถุนายน 2560, จาก <https://www.slideshare.net/NontawatB/07-classification-3-neural-network>
- อภิฤดี จิวะโรจน์. (2556). การศึกษาความเสี่ยงต่อการปลัดตกหกล้มในผู้ป่วยโรคหลอดเลือดสมองระยะฟื้นฟูสภาพ. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต, มหาวิทยาลัยขอนแก่น.
- อุทัยทิพย์ เจียวิวรรธน์กุล. (2553). การวิเคราะห์การถดถอยพหุโลจิสติกส์ (*Multiple logistic regression*). ค้นเมื่อ 8 ตุลาคม 2560, จาก http://rlc.nrct.go.th/download/Research_Zone/phase14_4.pdf

- อุไรวรรณ อมรนิมิตร. (2546). การวิเคราะห์ข้อมูลโดยใช้ Logistic Regression: ทางเลือกของการวิเคราะห์ความเสี่ยง [สื่ออิเล็กทรอนิกส์]. *วารสารวิชาการ มหาวิทยาลัย-หอการค้าไทย*, 23(2), 23-27.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคคิต้า ไม่นิ่งเบื้องต้น*. กรุงเทพมหานคร: ผู้แต่ง.
- Agency for Healthcare Research and Quality. (2013). *Preventing falls in hospitals a toolkit for improving quality of care*. Retrieved January 1, 2016, from <http://www.ahrq.gov/sites/default/files/publications/files/fallpctoolkit.pdf>
- Anderson, D., & MaNeill, G. (1992). “*Artificial Neural Networks Technology*”. Retrieved October 6, 2016, from <https://knn.es/Artificial%20Neural%20Network%20Technologies.pdf>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification Regression Trees*. New York: Chapman and Hall.
- Canadian Institute for Health Information. (2014). *Seniors’s Falls in Canada: Second Report*. Public Health Agency of Canada. Retrieved December 2, 2018, from <https://www.canada.ca/en/public-health/services/health-promotion/aging-seniors/publications/publications-generalpublic/seniors-falls-canada-second-report.html>
- Conley, D., Schultz, A., & Selvin, R. (1999). The challenge of predicting patients at risk for falling: development of the Conley Scale. [Electronic version]. *Medsurg Nursing*, 8(6), 348-354.

- Evan, D., Hodgkinson, B., Lambert, L., Wood, J., & Kowanko, I. (1998). Fall in acute hospital a systematic review. *The Joanna Briggs Institute for Evidence Base Nursing and Midwifery*. Retrieved October 2, 2016, from https://www.researchgate.net/profile/Brent_Hodgkinson/publication/253874155_Falls_in_Acute_Hospitals_ASystematic_Review/links/552b63350vf29b22c9c1a8d0.pdf
- Gales, B. J., & Menard, S. M. (1995). Relationship Between the Administration of Selected Medications and Falls in Hospitalized Elderly Patients. *Annals of Pharmacotherapy*, 29(4), 354-358.
- Hendrich, A. L., Bender, P. S., & Nyhuis, A. (2003). Validation of the Hendrich II Fall Risk Model: A Large Concurrent Case/Control Study of Hospitalized Patients. [Electronic version]. *Applied Nursing Research*, 16(1), 9-21.
- Hendrich, A. (2007). Predicting Patient Falls: Using the Hendrich II Fall Risk Model in clinical practice. [Electronic version]. *American Journal of Nursing*, 107(11), 50-58.
- Hitcho, E., Krauss, M., Birge, S., Funagan, W. C., Fisher, I., & Johnson, S. et al. (2004). Characteristics and circumstances of falls in a hospital setting: a prospective analysis. [Electronic version]. *Journal of General International Medicine*, 19(7), 732-739.
- Jiawei, H., Micheline, K., & Jian, P. (2012). "Chapter 8 Classification: Basic Concepts." DATA MINING: Concepts and Techniques (3rd ed.), Published by Morgan Kaufmann Publishers, ISBN 978-0-12-381479-1, 332-341.

- Jiawei, H., Micheline, K., & Jian, P. (2012). "Chapter 9 Classification: Advanced Methods." DATA MINING: Concepts and Techniques (3rd ed.), *Published by Morgan Kaufmann Publishers*, ISBN 978-0-12-381479-1, 398-404.
- Kangwanariyakul, Y., Nantasenamat, C., Tantimongcolwat, T., & Naenna, T. (2010). Data mining of magnetocardiograms for prediction of ischemic heart disease. *EXCLI journal*, 9, 82-95.
- Krauss, M., Evanoff, B., Hitcho, E., Ngugi, K., Funagan, W. C., & Fisher, I. et al. (2005). A case-control study of patient, medication, and care-related risk factors for inpatient falls. [Electronic version]. *Journal of General International Medicine*, 20(2), 116-122.
- Lahmann, N., Tannet, A., Kuntz, S., Raeder, K., Schmitz, G., & Dassen, T. et al. (2015). Mobility is the key! Trends and associations of common care problems in German long-term care facilities from 2008 to 2012. [Electronic version]. *International Journal of Nursing Studies*, 52(1), 167-174.
- 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. [Electronic version]. *International Journal of Medical Informatics*, 80(2), 141-150.
- Marschollek, M., Gövercin, M., Rust, S., Gietzelt, M., Schulze, M., Wolf, K.-H. et al. (2012). Mining geriatric assessment data for in-patient fall prediction models and high-risk subgroups. *BMC Medical Informatics and Decision Making*, 12, 19. Retrieved October 2, 2017, from <http://doi.org/10.1186/1472-6947-12-19>

- Meijers, J., Halfens, R., Neyens, J., Luiking, Y., & Verlaan, G. (2012). Predicting falls in elderly receiving home care: The role of malnutrition and impaired mobility. [Electronic version]. *The Journal of Nutrition, Health Aging, 16*(7), 645-658.
- Miceli, D. G. (2008). Prevention Falls in Acute Care. In Capezuti, E., Zwicker, D., Mezey, M., & Fulmer, T. *Evidence-Based Geriatric Nursing Protocol for Best Practice* (3rd ed., pp. 161-194). New York: Springer.
- Mion, L., Gregor, S., Buettner, M., Chwirschak, D., Lee, O., & Paras, W. (1989). Falls in the rehabilitation setting: incidence and characteristics. [Electronic version]. *Rehabilitaion Nursing, 14*(1), 7-22.
- Morse, J. M. (2009). *Preventing patient falls: Establishing a fall intervention program* (2nd ed.). New York: Springer.
- Nytröm, A., & Hellström, K. (2013). Fall risk six weeks from onset of stroke and the ability of the Prediction of Falls in Rehabilitation Settings Tool and motor function to predict falls. [Electronic version]. *Clinical Rehabilitation, 27*(5), 473-479.
- Oliver, D., Britton, M., Seed, P., Martin, F. C., & Hopper, A. H. (1997). Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. *BMJ (Clinical research ed.), 315*(7115), 1049-1053.
- Oliver, D., Daly, F., Martin, F., & McMurdo, M. E. (2004). Risk factors and risk assessment tools for falls in hospital in-patients: A systematic review. [Electronic version]. *Age and Ageing, 33*, 122-130.

- Passaro, A., Volpato, S., Romagnoni, F., Manzoli, N., Zuliani, G., & Fellin, R. (2000). Benzodiazepines with different half-life and falling in a hospitalized population: the GIFA study. [Electronic version]. *Journal of Clinical Epidemiology*, 53(12), 1222-1229.
- Perell, K., Nelson, A., Goldman, R., Luther, S., Lewis, N. P., & Rubenstein, L. (2001). Fall Risk Assessment Measures: An Analytic Review. [Electronic version]. *The Journal of Gerontology: Series A*, 56(12), 761-766.
- Quigley, P., & Miceli, D. (2011). Assessment Diagnosis and Intervention Strategies, In Boltz, M., Capezuti, E. Fulmer, T., & Zwicker, D. *Evidence-Based Geriatric Nursing Protocols for Best Practice* (4th ed., pp. 268-293). New York: Springer.
- Quinlan, J. R. (1986). Introduction of decision trees. *Machine learning*, 1(1), 81-106.
- Rapp, K., Ravindren, J., Becker, C., Lindemann, U., Jaensch, A., & Llenk, J. (2016). *Fall risk as a function of time after admission to sub-acute geriatric hospital units*. Biocentral Geriatrics, Retrieved October 10, 2016, from <https://pdfs.semanticscholar.org/8d35/93d6ddaedf3e5960830456fec909flc70bfl.pdf>
- Richardson, J. K. (2002). Factors associated with falls in older patients with diffuse polyneuropathy. [Electronic version]. *Journal American Geriatrics Society*, 50(11), 1767-1773.
- Shah, C., Williams, G., Joshi, J., & Aziz, R. (2012). A retrospective study of fall risk factors. [Electronic version]. *Journal of Community Nursing*, 26(5), 34-39.

The Joint Commission International. (2015). *Sentinel Event Alert. A*

Complimentary publication of The Joint Commission Issue 55.

Retrieved June 6, 2016, from http://www.jointcommission.org/assets/1/18/SEA_55.pdf

World Health Organization. (2018). *Fact sheet: Fall*. Retrieved December 2,

2018, from <https://www.who.int/news-room/fact-sheets/detail/falls>

Yamane, T. (1973). *Statistics: An introductory analysis*. London: John Weather Hill.

ประวัติผู้เขียน

ชื่อ ชื่อสกุล	นางปานิสรา อ่อนละมุล
วัน เดือน ปีเกิด	5 กันยายน 2521
สถานที่เกิด	จังหวัดนครพนม
วุฒิการศึกษา	สำเร็จการศึกษาระดับมัธยมศึกษาตอนปลาย จากโรงเรียนปิยะมหาราชาลัย จังหวัดนครพนม ปีการศึกษา 2539 สำเร็จปริญญาตรีพยาบาลศาสตรบัณฑิต จากวิทยาลัยพยาบาลทหารอากาศ ปีการศึกษา 2544
ตำแหน่งหน้าที่ การงานปัจจุบัน	Senior Specialist Tumor Registrar แผนก Research and Development โรงพยาบาลบำรุงราษฎร์ อินเตอร์เนชั่นแนล แขวงคลองเตยเหนือ เขตวัฒนา กรุงเทพมหานคร