การใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

นคร ละลอกน้ำ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2564

USING DATA MINING TECHNIQUES TO SCREEN CHILDREN WITH LEARNING DISABILITIES

NAKHON LALOGNUM

A Thesis Submitted in Partial Fulfllment of the Requirements

for the Degree of Master of Engineering

Department of Computer and Telecommunication Engineering

College of Innovative Technology And Engineering,

Dhurakij Pundit University

Academic Year 2021

ใบรับรองวิทยานิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

หัวข้อวิทยานิพนธ์ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

เสนอโดย นายนคร ละลอกน้ำ

สาขาวิชา วิศวกรรมคอมพิวเตอร์และโทรคมนาคม

อาจารย์ที่ปรึกษาวิทยานิพนธ์ อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบวิทยานิพนธ์แล้ว

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.ลัญฉกร วุฒิสิทธิกุลกิจ)

.....กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์

(อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์)

.....กรรมการ

(อาจารย์ ดร.ธนัญ จารุวิทยโกวิท)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.มัชฌิกา อ่องแตง)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์รับรองแล้ว

.....คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

(อาจารย์ ดร.ชัยพร เขมะภาตะพันธ์)

วันที่ ....1..... เดือน .....พ.ค..... พ.ศ. .....2564.....

หัวข้อวิทยานิพนธ์ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้
ชื่อผู้เขียน นคร ละลอกน้ำ
อาจารย์ที่ปรึกษา ดร.ชัยพร เขมะภาตะพันธ์
สาขาวิชา วิศวกรรมคอมพิวเตอร์และโทรคมนาคม
ปีการศึกษา 2563

บทคัดย่อ

จากปัญหาจากการคัดกรองเด็กบกพร่องทางการเรียนรู้ที่ผ่านมาจะใช้วิธีกรอกแบบสอบถามพฤติกรรมผู้เรียนจำนวนมาก และที่สำคัญผู้คัดกรองต้องรู้ข้อมูลนักเรียน ซึ่งต้องใช้เวลาอย่างน้อย 1 ภาคการศึกษาหรือ 4 เดือน และนักเรียน 1 คน ต้องใช้คนผู้คัดกรอง 2 คน ขึ้นไป ซึ่งเป็นการเพิ่มภาระงานให้ครู จากปัญหาดังกล่าวผู้วิจัยจึงสนใจพัฒนาแบบคัดกรองขึ้นมาใหม่ โดยกำหนดวัตถุประสงค์การวิจัย คือ 1) เพื่อพัฒนาแบบคัดกรองและศึกษาปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ 2) เพื่อสร้างและทดสอบประสิทธิภาพโมเดลคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ 3) เพื่อเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติโดยเก็บข้อมูลนักเรียนชั้นประถมศึกษา ปีที่ 4-6 จำนวน 238 คน เพื่อคัดกรองเด็กความบกพร่องทางการเรียนรู้โดยจำแนกเป็น 2 กลุ่ม คือ กลุ่มเด็กที่บกพร่องทางการเรียนรู้ และ กลุ่มเด็กที่ไม่บกพร่องทางการเรียนรู้ด้วยเทคนิคการทำเหมืองข้อมูลดำเนินการตามกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining) สร้างโมเดลคัดกรองคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้และเปรียบเทียบความแม่นยำของโมเดลจากอัลกอริทึม Decision Tree, Naïve Bayes, และNeural Network

ผลการวิจัย พบว่า 1) ปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ ได้แก่ คะแนนเฉลี่ยวิชาคณิตศาสตร์, ผลสอบด้านการอ่าน, คะแนนเฉลี่ยวิชาภาษาไทย, จำนวนพี่น้อง, ระดับชั้นเรียน 2) อัลกอรริทึม Neural Network และ Decision Tree มีค่าประสิทธิภาพสูงสุด โดยค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy) เท่ากับ 95.82% รองลงมาคืออัลกอรริทึม Naïve Bayes เท่ากับ 92.05% 3) ผลการเปรียบเทียบการแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ มีความถูกต้องด้านบกพร่องฯ เท่ากับ 82.60% ด้านไม่บกพร่องฯ เท่ากับ 0 และในส่วนของการทำนายโมเดลด้วยเทคนิค Decision Tree จากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่มีความถูกต้องด้านบกพร่องฯ เท่ากับ 89.47% ด้านไม่บกพร่องฯ เท่ากับ 100%

Thesis Title	USING DATA MINING TECHNIQUES TO SCREEN CHILDREN WITH LEARNING DISABILITIES
Author	Nakhon Lalognum
Thesis Advisor	Dr. Chaiyaporn Khemapatapan
Department	Computer and Telecommunications Engineering
Academic Year	2020

ABSTRACT

Previous issues with screening children with learning disabilities had suggested that a large number of student behavior questionnaires had to be administered, and that a screener was required to know student information. Clearly, this took at least one semester or four months, and one student normally required two or more screeners, which, in turn, increased teacher workloads. Accordingly, the researcher was interested in developing a new screening approach. The objectives of this research were 1) to develop a screening form and investigate the factors affecting the screening of children with learning disabilities; 2) to create and measure the effectiveness of a screening model of children with learning disabilities; and 3) to compare the predictions of each algorithm model of a newly developed screening model with the Ministry of Education's screening model which conducted a normal screening. The data was collected from 238 grade 4-6 students. This was to screen children with learning disabilities by categorizing them into two groups: children with learning disabilities and children without learning disabilities. With data mining techniques implemented according to a CRISP-DM (Cross-Industry Standard Process for Data Mining) process, a screening model of children with learning disabilities was constructed and accuracy of the model was compared using Decision Tree, Naïve Bayes, and Neural Network.

The results revealed that factors affecting the screening of children with learning disabilities were mean scores of mathematics, reading test scores, average scores of Thai language subjects, the number of siblings, and education levels. Moreover, it was found that Neural Network and Decision Tree had the highest efficiency. The overall accuracy of the all-class model (Accuracy) was 95.82%, followed by Naïve Bayes, which was 92.05%. It was also

discovered that the Ministry of Education's screeding model which conducted a normal screening indicated the accuracy of children with learning disabilities of 82.60% and the accuracy of children without learning difficulties of 0. The model prediction using Decision Tree based on the newly developed screening approach yielded the accuracy of children with learning difficulties of 89.47% and the accuracy of children without learning difficulties of 100%.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาอย่างดียิ่งจาก ดร.ชัยพร เขมะภาตะพันธ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ และขอขอบพระคุณอาจารย์กรรมการสอบวิทยานิพนธ์ทุกท่าน ที่กรุณาให้แนวคิด คำปรึกษา แนะนำช่วยเหลือในการตรวจแก้ไขข้อบกพร่องต่างๆ ผู้วิจัยมีความซาบซึ้งและขอบคุณยิ่ง

ขอกราบขอบพระคุณผู้ทรงคุณวุฒิที่ให้ความอนุเคราะห์ในการพัฒนาเครื่องมือการวิจัย ผู้วิจัยขอกราบขอบพระคุณมา ณ โอกาสนี้

นคร ละลอกน้ำ

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ฆ
บทคัดย่อภาษาอังกฤษ..... ง
กิตติกรรมประกาศ..... ฉ
สารบัญตาราง..... ซ
สารบัญภาพ..... ฌ
บทที่
1 บทนำ..... 1
1.1 ความเป็นมาและความสำคัญของปัญหา..... 1
1.2 วัตถุประสงค์การวิจัย..... 2
1.3 ขอบเขตการวิจัย..... 3
1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 5
1.5 นิยามศัพท์เฉพาะ..... 5
1.6 แผนการดำเนินงาน..... 7
1.7 การตอบรับการนำเสนอผลงานวิจัย..... 7
2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 8
2.1 ข้อมูลพื้นฐานเกี่ยวกับเด็กที่มีความบกพร่องทางการเรียนรู้..... 8
2.2 กระบวนการทำเหมืองข้อมูล..... 15
2.3 เทคนิคที่ใช้ในการทำเหมืองข้อมูล..... 25
2.4 โปรแกรม RapidMiner Studio..... 27
2.5 เอกสารงานวิจัยที่เกี่ยวข้องเกี่ยวกับการทำเหมืองข้อมูล..... 31
3 วิธีดำเนินการวิจัย..... 35
3.1 Business Understanding..... 36
3.2 Data Understanding..... 37
3.3 Data Preparation..... 44

สารบัญ (ต่อ)

บทที่		หน้า
	3.4 Modeling…….	47
	3.5 Evaluation……...	51
	3.6 Deployment…….	54
4	ผลการทดลอง.....	55
5	สรุป อภิปรายผลและข้อเสนอแนะ.....	65
	5.1 อภิปรายผล.....	65
	5.2 ข้อเสนอแนะจากการทำวิจัย.....	66
	5.3 แนวทางในการทำวิจัยต่อไป.....	67
บรรณานุกรม.....		68
ประวัติผู้เขียน.....		72

สารบัญตาราง

ตารางที่ **หน้า**

1.1 แผนดำเนินงาน 7

2.1 ข้อมูลสภาพอากาศย้อนหลัง 10 วัน..... 23

2.2 แสดงตาราง confusion matrix ของข้อมูล weather ซึ่งมี 2 คลาส..... 23

2.3 แสดงข้อมูลแอตทริบิวต์ Play จากเทรนนิงดาต้า (actual) 10 ตัวแรก และ ค่าที่ทำนายได้ (predicted)..... 24

2.4 แสดงตาราง confusion matrix ของข้อมูล weather ซึ่งมี 2 คลาส..... 25

3.1 แสดงขั้นตอนการดำเนินการวิจัย..... 35

3.2 การเลือกแอตทริบิวต์ (Attribute) และการแปลงรูปแบบของข้อมูล (Data transformation)..... 46

4.1 เปรียบเทียบผลการทดลองแต่ละอัลกอริทึม..... 57

4.2 เปรียบเทียบกระบวนการคัดกรองที่พัฒนาขึ้นใหม่กับกระบวนการคัดกรอง รูปแบบเดิม..... 60

4.3 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูลกับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ..... 62

4.4 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่กับการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ..... 63

สารบัญภาพ

ภาพที่ **หน้า**

2.1 การจำแนกแบ่งประเภทข้อมูลในการทำดาต้าไมน์นิง..... 17

2.2 กระบวนการทำเหมืองข้อมูล..... 19

2.3 การแปลงข้อมูลแบบมีโครงสร้าง และแบบไม่มีโครงสร้างเพื่อนำไปวิเคราะห์..... 20

2.4 โครงสร้างต้นไม้ตัดสินใจ..... 26

2.5 โครงสร้าง Layer ของโครงข่ายประสาทเทียมแบบหลายชั้น..... 27

2.6 โปรแกรม RapidMiner Studio 9..... 28

2.7 การเข้าสู่ account โดยการกรอก E-mail และ Password กรณีใช้ฟรีและการใส่ license Key กรณีที่ได้จากการซื้อโปรแกรมใส่เพื่อใช้งาน..... 28

2.8 องค์ประกอบของ RapidMiner Studio 9..... 29

2.9 โอเปอเรเตอร์และส่วนประกอบพารามิเตอร์ของโอเปอเรเตอร์ Read Excel..... 30

3.1 แบบทดสอบด้านการอ่าน..... 39

3.2 แบบทดสอบด้านการเขียน..... 40

3.3 แบบทดสอบด้านการคิดคำนวณ..... 41

3.4 การกรอกข้อมูลพื้นฐานนักเรียน..... 42

3.5 แบบคัดกรองของกระทรวงศึกษาธิการ..... 43

3.6 การเตรียมข้อมูลแบบคัดกรองสร้างใหม่สำหรับวิเคราะห์..... 45

3.7 ขั้นตอนการกำหนดข้อมูลก่อนนำไปวิเคราะห์เพื่อสร้างโมเดล..... 48

3.8 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Decision Tree..... 49

3.9 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Naïve Bayes..... 50

3.10 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Neural Network 51

3.11 เป็นประสิทธิภาพของค่า Precision..... 52

3.12 เป็นประสิทธิภาพของค่า recall 53

3.13 เป็นประสิทธิภาพของค่า Accuracy 54

สารบัญภาพ (ต่อ)

หน้า

ภาพที่

4.1 การแปลงข้อมูลเพื่อรวมคะแนนเทียบกับเกณฑ์เพื่อประเมินผลการคัดกรอง…….. 56

4.2 เปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่…….. 57

4.3 ผลกฎความสัมพันธ์การคัดกรองเด็กบกพร่องทางการเรียนรู้ด้วยเทคนิค Decision Tree โดยใช้แบบคัดกรองที่พัฒนาขึ้นใหม่ 59

4.4 ผลการทำนาย และ ค่า Confidence การทำนายด้วยวิธี Decision Tree……..… 61

4.5 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่กับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ……..... 62

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

รัฐธรรมนูญแห่งราชอาณาจักรไทยพุทธศักราช 2550 ปรับปรุงแก้ไข ฉบับที่ 2 พ.ศ. 2554 มีเจตนารมณ์ในการจัดการศึกษาสำหรับกลุ่มคนที่มีความต้องการพิเศษ โดยกำหนดไว้อย่างชัดเจนและเป็นรูปธรรมมากกว่าที่เคยผ่านมา ทำให้การจัดการศึกษาเอื้อประโยชน์ต่อเด็กที่มีความต้องการพิเศษให้ได้รับสิทธิและความเสมอภาค ทั้งทางด้านการศึกษา การพัฒนาคุณภาพชีวิต การสื่อสาร การเข้าถึงอาคาร สถานที่ สิ่งอำนวยความสะดวก สื่อ บริการและความช่วยเหลืออื่นใดทางการศึกษาตามกฎหมายและกฎกระทรวงที่เกี่ยวข้อง ซึ่งได้กำหนดไว้เช่นเดียวกับนักเรียนทั่วไป โดยรัฐต้องพัฒนาศักยภาพการเรียนรู้ในรูปแบบและวิธีการอันเหมาะสมกับสภาพความต้องการพิเศษทางด้านการศึกษาเป็นรายบุคคลและเป็นการจัดการศึกษา เพื่อมุ่งแก้ไขปัญหาและอุปสรรคอันเนื่องมาจากสภาพความบกพร่องหรือความต้องการพิเศษโดยจัดให้เข้าเรียนตามสภาพความบกพร่องในสภาพแวดล้อมที่มีขีดจำกัดน้อยที่สุด ซึ่งโดยทั่วไปหมายถึงการจัดให้เข้าเรียนร่วมในชั้นเรียนปกติร่วมกับนักเรียนทั่วไป

จากข้อมูลที่ผ่านมาประเทศไทยมีแนวโน้มเด็กที่มีความต้องการพิเศษลักษณะมีความบกพร่องทางการเรียนรู้ (Learning Disabilities : LD) เป็นจำนวนมากส่วนใหญ่จะมีช่วงอายุ 9-12 ปี ซึ่งจะศึกษาอยู่ระดับชั้นประถมศึกษาปีที่ 4-6 ลักษณะของเด็กกลุ่มนี้จะมีสติปัญญาและไอคิวปกติหรืออาจสูงกว่าปกติ แต่อาจมีความบกพร่องในเรื่องบางอย่างโดยเฉพาะเรื่อง โดยเฉพาะเรื่องการอ่านหนังสือ การเขียนหนังสือ และการคิดคำนวณ ซึ่งเราสามารถพัฒนาและช่วยเหลือเขาได้ ถ้าเราทราบแน่ชัดว่านักเรียนคนนั้นมีความบกพร่องทางการเรียนรู้หรือไม่ ที่ผ่านมาการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จะใช้แบบคัดกรองที่สร้างไว้แล้ว ซึ่งต้องคำนวณผลเองเพื่อนำไปเทียบกับเกณฑ์เพื่อตัดสินว่าผู้เรียนมีความเสี่ยงบกพร่องทางการเรียนรู้หรือไม่ ที่ใช้กันอยู่จะมีหลายแบบทุกแบบจะวิเคราะห์และประเมินลักษณะเดียวกัน ทำให้ครูหรือผู้คัดกรองต้องกรอกข้อมูลจำนวนมากโดยเฉพาะการประเมินนักเรียนแต่ละคนว่ามีความเสี่ยงหรือไม่ ต้องใช้ครูหรือผู้คัดกรองไม่น้อยกว่า 2 คน เป็นการเพิ่มภาระงานให้ครู และที่สำคัญการประเมินนักเรียนต่อคนต้องใช้เวลาไม่น้อยกว่า 1 ภาคการศึกษา หรือ 4 เดือน เพราะผู้คัดกรองต้องใช้เวลาในการสังเกตพฤติกรรมผู้เรียนจึงจะ

สามารถประเมินผู้เรียนแต่ละคนได้ ทำให้เสียเวลาในการคัดกรองนักเรียนมาก ส่งผลต่อการจัดแผนการเรียนที่ไม่เหมาะสมกับคุณลักษณะเฉพาะของผู้เรียนที่มีความบกพร่องทางการเรียนรู้ และที่สำคัญการคัดกรองในกระบวนการสุดท้ายครูต้องส่งเด็กไปให้แพทย์รับรองซึ่งต้องได้รับการยินยอมจากผู้ปกครอง ซึ่งเป็นอุปสรรคในทางปฏิบัติของครูผู้สอนเพราะผู้ปกครองไม่ยอมรับว่าบุตรหลานตนเองมีความบกพร่องทางการเรียนรู้ ส่วนใหญ่จะไม่ให้ความร่วมมือ

จากที่กล่าวมาข้างต้นผู้วิจัยจึงสนใจใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ให้สามารถทำได้ง่ายขึ้น มีความแม่นยำซึ่งจะช่วยผู้สอนทราบข้อมูลนักเรียนรายบุคคลเพื่อหาแนวทางในการพัฒนาตัวผู้เรียนที่มีลักษณะจำเพาะรายบุคคล จัดหลักสูตรรายวิชาที่เหมาะกับผู้เรียนรวมทั้งแผนการศึกษาเฉพาะของตนเอง (IEP) และการจัดการแผนการสอนเฉพาะบุคคล (IIP) ผู้สอนสามารถช่วยเหลือได้อย่างเจาะจงและถูกจุดสำหรับผู้เรียนแต่ละคน ซึ่งจะนำไปสู่การทำความเข้าใจในบทเรียนที่รวดเร็วขึ้นและผลการเรียนที่ดียิ่งขึ้น โดยงานวิจัยนี้จะใช้ข้อมูล 2 ส่วน คือ ส่วนที่ 1 ข้อมูลพื้นฐานผู้เรียน ได้แก่ เพศ จำนวน พี่น้อง ช่วงชั้น อายุ ผลการเรียนสะสมภาษาไทย ผลการเรียนสะสมคณิตศาสตร์ ผลการเรียนสะสมรวมทุกวิชา สถานะครอบครัว ผู้ปกครองปัจจุบัน ข้อมูลบิดา–มารดา (อายุ อาชีพ รายได้ ภูมิลำเนา) ส่วนที่ 2 จัดสร้างแบบทดสอบคัดกรอง ด้านการอ่าน ด้านการเขียน และด้านการคิดคำนวณ โดยออกแบบแบบทดสอบผ่าน Google Form จากนั้นนำข้อมูลไปสร้างโมเดลพยากรณ์ผ่านเทคนิคการทำเหมืองข้อมูล 3 เทคนิค เพื่อให้โมเดลมีความหลากหลายและมีความแมนยำมากที่สุด ได้แก่ Decision Tree, Naïve Bayesian และ Neural Network

1.2 วัตถุประสงค์การวิจัย

1.2.1 เพื่อพัฒนาแบบคัดกรองและศึกษาปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

1.2.2 เพื่อสร้างและทดสอบประสิทธิภาพโมเดลคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

1.2.3 เพื่อเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ

1.3 ขอบเขตการวิจัย

1.3.1 ขอบเขตด้านข้อมูล

ผู้วิจัยได้ศึกษาเอกสาร ตำรา งานวิจัยที่เกี่ยวข้องกับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ เพื่อกำหนดเป็นปัจจัยที่ส่งผลต่อความบกพร่องทางการเรียนรู้เพื่อนำมาสร้างแบบทดสอบและข้อมูลพื้นฐานผู้เรียน โดยการเก็บรวบรวมข้อมูลได้ดำเนินการช่วงกรกฎาคม - พฤษศจิกายน 2562 จากครูประจำชั้นและนักเรียนชั้นประถมศึกษาปีที่ 4-6 จำนวนทั้งหมด 8 โรงเรียน 250 คน กำหนดเป็น Attribute จำนวน 20 Attribute ต่อไปนี้

1) เพศ (SEX)
2) จำนวนพี่น้อง (BRETHREN)
3) ช่วงชั้น (GRADE)
4) อายุนักเรียน (AGE STUDENT)
5) ผลการเรียนสะสมภาษาไทย (GPA THAI)
6) ผลการเรียนสะสมคณิตศาสตร์ (GPA MATH)
7) ผลการเรียนสะสมรวมทุกวิชา (GPA)
8) สถานะครอบครัว (STATUS)
9) ผู้ปกครองปัจจุบัน (SPONSOR)
10) อายุบิดา (AGE FA)
11) อาชีพบิดา (CAREER FA)
12) รายได้บิดา (INCOME FA)
13) ภูมิลำเนาบิดา (REGION FA)
14) อายุมารดา (AGE MON)
15) อาชีพมารดา (CAREER MON)
16) รายได้มารดา (INCOME MON)
17) ภูมิลำเนามารดา (REGION MON)
18) แบบทดสอบคัดกรองด้านการอ่าน (READ)
19) แบบทดสอบคัดกรองด้านการเขียน (WRITE)
20) แบบทดสอบคัดกรองด้านคิดคำนวณ (CALCULATION)

1.3.2 ขอบเขตด้านกระบวนการทำเหมืองข้อมูล

วิธีดำเนินการวิจัยครั้งนี้ได้ดำเนินการตามกระบวนการทำเหมืองข้อมูลของ CRISP-DM ได้แก่ 1) Business Understanding 2) Data Understanding 3) Data Preparation 4) Modeling 5) Evaluation และ 6) Deployment

1.3.3 ขอบเขตด้านอัลกอริทึมและการวัดประสิทธิภาพ

เทคนิคที่ใช้ในการวิเคราะห์เพื่อสร้างแบบจำลองเพื่อหากฎความสัมพันธ์ของข้อมูลต่างๆ ได้แก่ เทคนิค Decision Tree เทคนิค Naïve Bayes และ เทคนิค Neural Network

การวัดประสิทธิภาพของผลลัพธ์โดยแบ่งกลุ่มข้อมูลเพื่อทดสอบ (Cross-validation) ประกอบด้วยข้อมูลส่วน Training set และ Testing set โดยใช้วิธี 10-fold cross-validation

1.3.4 ขอบเขตด้านเครื่องมือในการวิจัย

1.3.4.1 แบบทดสอบเด็กบกพร่องทางการเรียนรู้ ประกอบด้วย ด้านการอ่าน ด้านการเขียน และด้านคำนวณ เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก ด้านละ 10 ข้อ รวมทั้งหมด 30 ข้อ โดยให้ทำผ่าน Google Form

1.3.4.2 ขอบเขตทางด้านฮาร์ดแวร์ (Hardware) มีดังนี้

1) คอมพิวเตอร์ส่วนบุคคล (Personal Computer) ใช้สำหรับกรอกข้อมูล และประมวลผลข้อมูล

2) เครื่องบันทึกเสียงใช้สำหรับสร้างแบบทดสอบด้านการอ่าน และด้านการเขียน

1.3.4.3 ซอฟต์แวร์(Software)

1) โปรแกรม RapidMiner Studio9 ซึ่งเป็น Software License ที่ให้ใช้ฟรีในสถานะเป็นสถาบันการศึกษา โดยใช้สำหรับวิเคราะห์และประมวลผลข้อมูลเพื่อสร้างโมเดลคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

2) โปรแกรม Microsoft Excel ซึ่งเป็น Software License ที่ให้ใช้ฟรีในมหาวิทยาลัยบูรพาโดยใช้สำหรับการเลือกแอตทริบิวต์ (Attribute) และการแปลงรูปแบบของข้อมูล (Data transformation)

3) การใช้ Google Form เพื่อสร้างแบบสอบถามและแบบทดสอบ Online ในการเก็บข้อมูลมาวิเคราะห์คัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้แบบคัดกรองใหม่และทราบปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ ทำให้คัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ได้สะดวกและง่ายขึ้น ลดความยุ่งยาก ลดขั้นตอน

1.4.2 ได้โมเดลเพื่อช่วยในการคัดกรองเด็กที่ที่มีความบกพร่องทางการเรียนรู้และทราบอัลกอริทึมที่มีความแม่นยำในการพยากรณ์

1.4.3 ครูผู้สอนนำโมเดลที่ได้ไปใช้เป็นแนวทางในการคัดคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ได้อย่างสะดวก แม่นยำเพื่อให้สามารถจัดแผนการเรียนที่เหมาะสมกับคุณลักษณะเฉพาะของผู้เรียนที่มีความบกพร่องทางการเรียนรู้ได้อย่างมีประสิทธิภาพ

1.5 นิยามศัพท์เฉพาะ

1.5.1 ความบกพร่องทางการเรียนรู้ หรือ LD หมายถึง ความผิดปกติด้านการเขียน ด้านการอ่าน ด้านคณิตศาสตร์ ทำให้บุคคลมีข้อจำกัดในการเรียนรู้

1.5.2 การทำเหมืองข้อมูล หมายถึง การวิเคราะห์ข้อมูลจากฐานข้อมูลขนาดใหญ่ เพื่อให้ทราบความสัมพันธ์ รูปแบบของข้อมูลนั้นๆ ให้สามารถนำไปใช้ประโยชน์ได้ เช่น การตัดสินใจ การวางแผน การทำนายแนวโน้มสิ่งที่จะเกิดขึ้นในอนาคต หรือเป็นการแปรเปลี่ยนข้อมูลไปสู่ความรู้ใหม่ๆ

1.5.3 การจัดจำแนก (Classification) หมายถึง การนำข้อมูลที่มีในอดีตมาสอนระบบเพื่อให้เรียนรู้รูปแบบที่เกิดขึ้นในข้อมูล จากนั้นนำมาสร้างเป็นสมการหรือโมเดลขึ้นมา เพื่อหาคำตอบให้สำหรับข้อมูลใหม่ หรือ “เป็นโมเดลที่ใช้สำหรับนำข้อมูลที่มีอยู่มาทำนายอนาคต” อัลกอริทึมที่ใช้ในการวิเคราะห์ครั้งนี้ ได้แก่ เทคนิค Decision Tree เทคนิค Naïve Bayes และ เทคนิค Neural Network

1.5.4 วิธีการแบ่งข้อมูลเพื่อทำการทดสอบโมเดลด้วย วิธี Cross-validation Test หมายถึงการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) โดยการวิจัยครั้งนี้จะแบ่งเป็น 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กันแล้วทำการทดสอบโดยเปลี่ยนข้อมูลชุดทดสอบตั้งแต่ส่วนที่ 1 เป็นชุดทดสอบ และส่วนที่ 2-9 เป็นชุดเทรนนิ่งไปเรื่อยๆ จนกระทั่งถึงชุดทดสอบที่ส่วนที่ 10 เป็นชุดทดสอบ และส่วนที่ 1-9 เป็นชุดเทรนนิ่ง เพื่อให้ได้ค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy)

1.5.5 “คลาส (class)” หรือ “ลาเบล (label)” หมายถึง ค่าคำตอบจากการวิเคราะห์การวิจัยครั้งนี้ จะกำหนดเป็น label 2 ค่า (Binominal) คือ เด็กที่มีความเสี่ยงบกพร่องทางการเรียนรู้จะใช้แทนด้วย 1 และเด็กที่ไม่มีความเสี่ยงบกพร่องทางการเรียนรู้จะใช้แทนด้วย 0

1.5.6 ตัวที่ใช้วัดประสิทธิภาพของโมเดล หมายถึง ค่าตัวเลขที่สะท้อนถึงคุณภาพของโมเดลมีค่าต่อไปนี้

*Precision คือ ความแม่นยำที่สนใจแค่ส่วนที่ทำนาย หรือ ความสามารถในการทำนายว่าเป็น LD แล้วถูกต้อง ในการวิจัยครั้งนี้ส่วนที่ทำนาย คือ เด็กที่มีความบกพร่องทางการเรียนรู้ หรือเป็น LD (คลาส positive) คำนวณได้จาก Precision = TP / (TP + FP)

*Recall คือ ความแม่นยำที่สนใจในส่วนของความเป็นจริง หรือ Data หรือความสามารถในการตรวจจับการเป็น LD จากจำนวนข้อมูลที่เป็นจริงทั้งหมด คำนวณได้จาก Recall = TP/(TP+FN)

*Accuracy คือ ความถูกต้องที่ทำนายได้ตรงกับสิ่งที่เกิดขึ้นจริง ในการวิจัยครั้งนี้ จะเป็นการดูว่าโมเดลสามารถทำนายว่าเป็น LD ถูกต้อง เทียบกับจำนวนครั้งที่ทำนายทั้งหมด คำนวณได้จาก Accuracy = (TP + TN) / (TP+TN+FP + FN)

1.6 แผนการดำเนินงาน

ในการดำเนินการวิจัยได้กำหนดแผนดำเนินงาน ดังตารางที่ 1

ตารางที่ 1.1 แผนดำเนินงาน

รายการดำเนินงาน	ระยะเวลา (เดือน)											
	1	2	3	4	5	6	7	8	9	10	11	12
1.ศึกษาเอกสาร ตำรา งานวิจัยที่เกี่ยวข้องกับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้												
2.ทำ Data Preparation เพื่อคัดเลือกข้อมูลให้มีความพร้อมสำหรับการวิเคราะห์												
3.ทำ Modeling โดยนำข้อมูลเข้ามาวิเคราะห์เพื่อสร้างแบบจำลองเพื่อหากฎความสัมพันธ์ของข้อมูลต่างๆ												
4.วัดประสิทธิภาพของผลลัพธ์ (Evaluation)												
5.วิเคราะห์และสรุปผลการทดลอง												
6.จัดทำรูปเล่มฉบับสมบูรณ์												

1.7 การตอบรับการนำเสนอผลงานวิจัย

งานวิจัย เรื่อง “การใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้” ได้รับการตอบรับให้นำเสนอการประชุมวิชาการระดับชาติวิทยาศาสตร์และเทคโนโลยี ครั้งที่ 5 (The 5th National Conference on Science and Technology) วันที่ 15 - 16 มกราคม พ.ศ. 2564

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ ผู้วิจัยศึกษาเอกสารงานวิจัยที่เกี่ยวข้องในประเด็นต่อไปนี้

2.1 ข้อมูลพื้นฐานเกี่ยวกับเด็กที่มีความบกพร่องทางการเรียนรู้

2.2 กระบวนการทำเหมืองข้อมูล

2.3 เทคนิคที่ใช้ในการทำเหมืองข้อมูล

2.4 โปรแกรม RapidMiner Studio

2.5 เอกสารงานวิจัยที่เกี่ยวข้องเกี่ยวกับการทำเหมืองข้อมูล

2.1 ข้อมูลพื้นฐานเกี่ยวกับเด็กที่มีความบกพร่องทางการเรียนรู้

2.1.1 ความหมายความบกพร่องทางการเรียนรู้

นักวิชาการด้านการศึกษาพิเศษ ได้ให้ความหมายบุคคลที่มีความบกพร่องทางการเรียนรู้ ดังนี้

ความบกพร่องทางการเรียนรู้ หมายถึง เด็กที่มีความบกพร่องในขบวนการทางจิตวิทยา ทำให้เด็กมีปัญหาในการใช้ภาษา ทั้งในการฟัง การอ่าน การพูด การเขียนและการสะกดคำ หรือมีปัญหาในการเรียนวิชาคณิตศาสตร์ ปัญหาดังกล่าวมิได้มีสาเหตุมาจากความบกพร่องทางร่างกาย แขน ขา ลำตัว สายตา การได้ยิน อารมณ์ และสภาพแวดล้อมรอบตัวเด็ก

บุคคลที่มีความบกพร่องทางการเรียนรู้ ได้แก่ บุคคลที่มีความผิดปกติในการทำงานของสมองบางส่วนที่แสดงถึงความบกพร่องในกระบวนการเรียนรู้ที่อาจเกิดขึ้นเฉพาะความสามารถด้านใดด้านหนึ่งหรือหลายด้าน คือ การอ่าน การเขียน การคิดคำนวณ ซึ่งไม่สามารถเรียนรู้ในด้านที่บกพร่องได้ทั้งที่มีระดับสติปัญญาปกติ

เด็กที่มีปัญหาทางการเรียนรู้ หมายถึง เด็กที่มีความบกพร่องเกี่ยวกับกระบวนการทางจิตวิทยาซึ่งทำให้เด็กมีปัญหาการเรียนด้านต่างๆ คือ การอ่าน การเขียน การฟัง การคิด หรือการคำนวณ รวมถึงเด็กที่มีความบกพร่องทางการรับรู้จากการได้รับความกระทบกระเทือนทางสมอง

แต่ไม่รวมเด็กที่มีปัญหาอันเกิดจากความบกพร่องทางสายตา การได้ยิน การเคลื่อนไหวทางร่างกาย หรือปัญญาอ่อน ตลอดจนความบกพร่องทางอารมณ์และการเสียเปรียบทางสภาพแวดล้อม

ความบกพร่องทางการเรียนรู้ว่า หมายถึง ความล่าช้า ความผิดปกติ หรือพัฒนาการล่าช้าในกระบวนการอย่างหนึ่งหรือมากกว่านั้นเกี่ยวกับการพูด ภาษา การอ่าน การสะกดคำ การเขียน หรือคณิตศาสตร์ที่เป็นผลมาจากการทำงานผิดปกติเกี่ยวกับสมอง และหรืออารมณ์ พฤติกรรมอยู่ไม่นิ่ง และไม่ใช่มาจากสาเหตุปัญญาอ่อน การสูญเสียประสาทสัมผัสหรือปัจจัยทางวัฒนธรรมหรือการเรียนการสอน

สรุปความหมายความบกพร่องทางการเรียนรู้ หมายถึง ความผิดปกติด้านการเขียน ด้านการอ่าน ด้านคณิตศาสตร์ ทำให้บุคคลมีข้อจำกัดในการเรียนรู้

2.1.2 สาเหตุของความบกพร่องทางการเรียนรู้

สาเหตุของความบกพร่องทางการเรียนรู้สามารถสรุปได้ 3 สาเหตุหลักๆ ดังนี้

1) ปัจจัยด้านพันธุกรรม คาดว่าเป็นปัจจัยหนึ่งที่ส่งผลให้เด็กมีความบกพร่องทางการเรียนรู้ มีงานวิจัยจำานวนมากที่ระบุตรงกันว่าปัญหาทางการเรียนรู้บางอย่างสามารถถ่ายทอดทางพันธุกรรม จากการศึกษาเป็นรายกรณีพบว่า เด็กที่มีปัญหาในการเรียนรู้บางคน อาจมีพี่น้องที่เกิดจากท้องเดียวกันมีปัญหาการเรียนรู้เช่นกัน หรืออาจมีพ่อ แม่ พี่น้อง หรือญาติใกล้ชิดมีปัญหาทางการเรียนรู้

2) การได้รับการกระทบกระเทือนทางสมอง บางทฤษฎีกล่าวว่าเด็กที่มีพัฒนาการสมองผิดปกติ เช่น เด็กที่มีน้ำหนักแรกเกิดต่ำกว่ามาตรฐาน คลอดก่อนกำหนด สมองขาดออกซิเจน หรือได้รับอุบัติเหตุที่กระทบกระเทือนสมอง อาจมีแนวโน้มเกิดภาวะบกพร่องทางการเรียนรู้มากกว่าเด็กทั่วไป

3) ปัจจัยด้านสิ่งแวดล้อม เป็นสิ่งรอบตัวที่เด็กเติบโตขึ้นมา เช่น การสูดดมหรือสัมผัสสารพิษจากสิ่งแวดล้อมเป็นประจำ โภชนาการที่ไม่ดีตั้งแต่เด็ก พฤติกรรมพ่อแม่ การได้รับโอกาสทางการศึกษาที่ไม่เหมาะสม ทั้งหมดนี้อาจส่งผลให้เกิดความพร่องทางการเรียนรู้

2.1.3 ลักษณะของความบกพร่องทางการเรียนรู้

การจำแนกลักษณะของความบกพร่องทางการเรียนรู้สรุปได้ 4 ลักษณะ ดังนี้

1) ปัญหาทางด้านการอ่าน มีลักษณะพฤติกรรม ดังนี้

- การเคลื่อนไหวร่างกายที่แสดงอาการเครียด เช่น อารมณ์เสีย หน้าง้อ
- อ่านหลงบรรทัด อ่านซ้ำคำ
- อ่านตกหล่น อ่านเพิ่มคำ หาคำมาแทนที่ หรืออ่านกลับคำ
- อ่านเรียงลำดับคำผิด สับสนต่ำแหน่งประธาน กริยา กรรม

- อ่านสับสนระหว่างอักษรหรือคำที่คล้ายคลึงกัน
- อ่านช้าและตะกุกตะกัก
- อ่านด้วยความลังเลไม่แน่ใจ
- อ่านเอาเรื่องไม่ได้
- บอกลำดับเรื่องราวไม่ได้
- จำประเด็นสำคัญของเรื่องราวไม่ได้ (อ่านได้แต่อธิบายไม่ได้)
- แยกสระเสียงสั้น – ยาวไม่ได้

2) ปัญหาทางด้านการเขียน อาจมีสาเหตุมาจากความบกพร่องใน 3 ลักษณะ คือ

(1) การประสานสัมพันธ์ระหว่างกล้ามเนื้อมือและสายตาไม่ดี จึงทำให้เด็กลอกตัวอักษรและตัวเลขไม่ถูกต้อง

(2) ความบกพร่องของการจำสิ่งที่มองเห็น จึงทำให้เด็กจำสิ่งที่เห็นไม่ได้

(3) ความบกพร่องในการทำความเข้าใจกฎเกณฑ์และความสัมพันธ์ระหว่างถอยคำในประโยคจึงทำให้เด็กที่มีปัญหาในการรวบรวมหรือจัดระบบความคิดเพื่อการสื่อสารออกมาโดยการเขียนไม่ได้

เด็กที่มีปัญหาในการเขียนจะมีพฤติกรรมปรากฏอย่างใดอย่างหนึ่ง หรือหลายอย่างดังนี้

- มีปัญหาในการจับดินสอ หรือดินสอสี
- เขียนตัวอักษรผิดทิศทางกลับซ้าย – ขวา หน้า – หลัง บน – ล่าง
- เว้นระยะตัวอักษรหรือคำไม่ถูกต้อง
- เขียนหนังสือไม่เป็นตัว ลายมืออ่านได้ยาก
- เขียนหนังสือรูปร่างไม่สม่ำเสมอ
- เขียนหนังสือแบบที่ปรากฏในกระจกเงา
- เขียนตัวอักษรหลายแบบปะปนกัน
- จำทักษะพื้นฐานการเขียนไม่ได้

3) ปัญหาทางด้านการสะกดคำ มีลักษณะพฤติกรรม ดังนี้

- เรียงตัวอักษรในคำผิด
- สลับตัวอักษรและคำ
- มีปัญหาในการเชื่อมโยงเสียงที่ถูกกับตัวอักษร
- สะกดข้ามตัวอักษรหลายตัว
- สร้างการสะกดคำแบบใหม่ของตัวเอง

4) ปัญหาทางด้านคณิตศาสตร์ มีลักษณะพฤติกรรม ดังนี้

- ลืมตำแหน่งบนหน้ากระดาษ
- ทำงานไม่เสร็จในหนึ่งหน้าที่กำหนดให้
- มีปัญหาในการอ่านเลขหลายหลัก
- มีปัญหาในการจำแนกตัวเลขบางตัว เช่น 6 กับ 9, 2 กับ 5, 17 กับ 71
- เขียนตัวเลขจากหลังมาหน้า
- เขียนตัวเลขจากขวามาซ้าย
- มีปัญหาในการลากเส้นตัวเลข
- มีปัญหาในการลอกรูปทรงต่างๆ และลอกโจทย์ปัญหา
- ไม่สามารถจำเงื่อนไขกฎ ข้อเท็จจริงทางคณิตศาสตร์
- มีปัญหาในการเรียนรู้ข้อเท็จบางเรื่อง
- ทำผิดเพราะสะเพร่าบ่อยๆ
- ยอมแพ้ง่ายๆ
- มีปัญหาในการเชื่อมโยงจำนวนกับสัญลักษณ์
- สับสนข้อมูลในแนวตั้งและการเว้นระยะ
- ทำโจทย์ปัญหาไม่ได้
- ไม่มีความเข้าใจเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์

2.1.4 การคัดกรองบุคคลที่มีความบกพร่องทางการเรียนรู้

เครื่องมือสำหรับคัดแยกนักเรียนที่มีความบกพร่องทางการเรียนรู้ที่นิยมใช้ มีดังนี้

1) แบบสำรวจเด็กที่มีป์ญหาในการเรียนรู้ได้รับการพัฒนาขึ้นโดยศาสตราจารย์ ดร.ผดุง อารยะวิญญู ใช้กับเด็กอายุ 6 - 12 ปี ซึ่งศึกษาอยู่ในระดับประถมศึกษา โดยมีครูผู้สอนซึ่งรู้จัก และใกล้ชิดกับเด็กเป็นอย่างดีเป็นผู้ประเมินเด็กแต่ละคน โดยใช้การสังเกตเด็กติดต่อกันเป็นระยะเวลาไม่น้อยกว่า 3 เดือน แบบสำรวจนี้ใช้เพื่อสำรวจป์ญหาของเด็ก 5 ด้านใหญ่ ๆ ได้แก่ ด้านภาษา ด้านคณิตศาสตร์ ด้านเวลาและทิศทาง ด้านการเคลื่อนไหว และด้านพฤติกรรม

2) แบบคัดกรองนักเรียนที่มีภาวะสมาธิสั้น บกพร่องทางการเรียนรู้ และออทิซึม (KUS-SI Rating Scales: ADHD/LD/Autism (PDDs)) เพื่อใช้ในการคัดกรองนักเรียนตั้งแต่ชั้น ป.1- ป.6 อายุระหว่าง 6-13 ปี 11 เดือน ที่มีภาวะสมาธิสั้น บกพร่องทางการเรียนรู้และออทิซึม โดยแบ่งพฤติกรรมออกเป็น 5 ด้าน ดังนี้ (1) พฤติกรรมภาวะสมาธิสั้น (KUS-SI Rating Scale 1: ADHD) มีข้อความ 30 ข้อ ใช้คัดแยกนักเรียนที่มีภาวะสมาธิสั้น (2) พฤติกรรมภาวะบกพร่องทางการเรียนรู้

ด้านการอ่าน (KUS-SI Rating Scale 2: LD-Reading) มีข้อความ 20 ข้อ ใช้คัดกรองนักเรียนที่มีภาวะบกพร่องทางการเรียนรู้ด้านการอ่าน (3) พฤติกรรมภาวะบกพร่องทางการเรียนรู้ด้านการเขียน (KUS-SI Rating Scale 3: LD-Writing) มีข้อความ 20 ข้อ ใช้คัดกรองนักเรียนที่มีภาวะบกพร่องทางการเรียนรู้ด้านการเขียน (4) พฤติกรรมภาวะบกพร่องทางการเรียนรู้ด้านการคิดคำนวณ (KUS-SI Rating Scale 4: LD-Mathematics) มีข้อความ 20 ข้อ ใช้คัดกรองนักเรียนที่มีภาวะบกพร่องทางการเรียนรู้ด้านการคิดคำนวณ (5) พฤติกรรมภาวะออทิซึม (KUS-SI Rating Scale 5: Autism/PDDs) มีข้อความ 40 ข้อ ใช้คัดกรองนักเรียนที่มีภาวะออทิซึม

การคัดกรองสำหรับนักเรียนที่มีความบกพร่องทางการเรียนรู้โดยใช้เครื่องมือนี้ ผู้ตอบแบบคัดกรอง ต้องเป็นผู้ที่รู้จักนักเรียนเป็นอย่างดี ผู้ตอบแบบคัดกรอง คือ ผู้ตอบข้อความที่บ่งบอกพฤติกรรมแต่ละด้านของแบบคัดกรอง ส่วนผู้ประเมิน คือ ผู้รับผิดชอบในการประเมินและแปลผลคะแนน ในบางกรณีผู้ตอบและผู้ประเมินอาจเป็นคน ๆ เดียวกันได้

3) แบบคัดกรองบุคคลที่มีความบกพร่องทางการเรียนรู้ เป็นแบบจำแนกทางการศึกษามีทั้งระดับประถมศึกษา และมัธยมศึกษา โดยผู้ที่ใกล้ชิดกับเด็กมากที่สุดจะทำหน้าที่เป็นผู้ประเมิน แบบคัดกรองนี้แบ่งเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 ต้องตอบว่า ใช่ ทั้ง 3 ข้อ จึงจะทำต่อส่วนที่ 2 ส่วนที่ 2 แบ่งความบกพร่องออกเป็น 3 ด้าน คือ 1) ความบกพร่องทางด้านการอ่าน 2) ความบกพร่องทางด้านการเขียน และ 3) ความบกพร่องทางด้านการคำนวณ 4) ด้านพฤติกรรมทั่วไป ดังตัวอย่าง

- ตัวอย่างแบบคัดกรองบุคคลที่มีความบกพร่องทางการเรียนรู้ของกระทรวงศึกษา

คำชี้แจง

๑ แบบคัดกรองฉบับนี้เป็นแบบคัดกรองเพื่อประโยชน์ในทางการจัดการศึกษาเท่านั้น

๒ วิเคราะห์ลักษณะ/พฤติกรรม ของเด็กซึ่งเป็นลักษณะหรือพฤติกรรม ที่เด็กแสดงออกบ่อยๆ โดยให้ทำเครื่องหมาย/ลงในช่อง“ใช่”หรือ“ไม่ใช่”ที่ตรงกับลักษณะหรือพฤติกรรมนั้นๆ ของเด็กโดยเปรียบเทียบกับเด็กทั่วไปในชั้นเรียน

๓ ผู้ทำการคัดกรองเบื้องต้นต้องผ่านการอบรมวิธีการใช้ และการประเมิน ตามแบบคัดกรองนี้และควรสอบถามข้อมูลเพิ่มเติมจากผู้ที่อยู่ใกล้ชิดเด็กมากที่สุด เช่น ครูผู้สอน ครูประจำชั้น เพื่อให้เกิดความชัดเจน ถูกต้อง

๔ ผู้คัดกรองควรจะมีอย่างน้อย ๒ คนขึ้นไป

ส่วนที่ ๑ การวิเคราะห์เบื้องต้น / ข้อมูลพื้นฐานของบุคคลที่มีความบกพร่องทางการเรียนรู้

ที่	ลักษณะ / พฤติกรรม	ผลการวิเคราะห์	
		ใช่	ไม่ใช่
๑	ดูฉลาดหรือปกติ ในด้านอื่น ๆ นอกจากในด้านการเรียน		
๒	ต้องมีปัญหาทางการเรียน ซึ่งอาจทำไม่ได้เลยหรือทำได้ต่ำกว่า ๒ชั้นเรียน ในด้านใดด้านหนึ่งหรือมากกว่า ๑ด้าน ต่อไปนี้ ๑.๑ ด้านการอ่าน ๑.๒ ด้านการเขียน ๑.๓ ด้านการคำนวณ		
๓	ไม่มีปัญหาทางด้านการเห็น การได้ยิน สติปัญญา หรือออทิสติกหรือจากการถูกละทิ้ง ละเลย หรือความด้อยโอกาสอื่น ๆ		

เกณฑ์การพิจารณา

ถ้าตอบว่าใช่ ๓ ข้อ แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ควรสังเกตในส่วนที่ ๒ ต่อ

ผลการพิจารณาส่วนที่ ๑

☐ พบ ☐ ไม่พบ (ถ้าพบสังเกตในส่วนที่ ๒ ต่อ)

ส่วนที่ ๒ การวิเคราะห์ความบกพร่องทางการเรียนรู้ของเด็กในแต่ละด้าน

ที่	ลักษณะ / พฤติกรรม	ผลการวิเคราะห์	
		ใช่	ไม่ใช่
	๑. ด้านการอ่าน		
๑	อ่านช้าอ่านข้าม อ่านไม่หมด		
๒	จำคำศัพท์คำเดิมไม่ได้ ทั้งๆ ที่เคยผ่านสายตามาแล้วหลายครั้ง		
๓	อ่านเพิ่มคำ ซ้ำคำ อ่านผิดตำแหน่ง		
๔	อ่านสลับตัวอักษรหรือออกเสียงสลับกัน เช่น บก อ่านเป็น กบ		
๕	สับสนในพยัญชนะคล้ายกันเช่น ก ภ ถ, ฦ ฦ ฦ, ค ต ค		
๖	จำศัพท์ใหม่ไม่ค่อยได้		
๗	มีปัญหาในการผสมคำ การอ่านออกเสียงคำ		
๘	สับสนคำที่คล้ายกัน เช่น บาน/ บ้าน		

ที่	ลักษณะ / พฤติกรรม	ผลการวิเคราะห์	
		ใช่	ไม่ใช่
๘	อ่านคำที่ไม่คุ้นเคยไม่ได้		
๑๐	อ่านคำในระดับชั้นของตนเองไม่ได้		
	๒. ด้านการเขียน		
๑	ไม่ชอบและหลีกเลี่ยงการเขียน หรือการลอกคำ		
๒	เขียนไม่สวยไม่เรียบร้อย สกปรก ขีดทิ้ง ลบทิ้ง		
๓	เขียนตัวอักษรและคำที่คล้ายๆ กันผิด		
๔	ลอกคำบนกระดานผิด (ลอกไม่ครบตกหล่น)		
๕	เขียนหนังสือไม่เว้นวรรค ไม่เว้นช่องไฟ ตัวอักษรเบียดกันจนทำให้อ่านยาก		
๖	เขียนสลับตำแหน่งระหว่างพยัญชนะ สระ เช่น ตโ		
๗	เขียนตามคำบอกของคำในระดับชั้นตนเองไม่ได้		
๘	เขียนตัวอักษรหรือตัวเลขกลับด้าน คล้ายมองกระจกเงา เช่น ๖๘,๗๙		
๙	เขียนพยัญชนะหรือตัวเลขที่มีลักษณะคล้ายกันสลับกัน เช่น ม-น, ด-ค, พ-ย, b-d, p-q, ๖-๙		
๑๐	เรียงลำดับตัวอักษรผิด เช่น สถิติ เป็น สติถิ		
	๓. ด้านการคำนวณ		
๑	นับเลขเรียงลำดับ นับเพิ่ม นับลดไม่ได้		
๒	ยากลำบากในการบวก,ลบ จำนวนจริง		
๓	ยากลำบากในการใช้เทคนิคการนับจำนวนเพิ่มทีละ ๒, ๕, ๑๐, ๑๐๐		
๔	ยากลำบากในการประมาณจำนวนค่า		
๕	ยากลำบากในการเปรียบเทียบ มากกว่า น้อยกว่า		
๖	แก้โจทย์ปัญหาง่ายๆ ไม่ได้		
๗	สับสนไม่เข้าใจเรื่องเวลา ทิศทาง		
๘	บอกความหมาย หรือสัญลักษณ์ทางคณิตศาสตร์ ไม่ได้ เช่น +, -, ×, >, <, =		
๙	เปรียบเทียบขนาด รูปทรง ระยะทาง ตำแหน่งไม่ได้		
๑๐	เขียนตัวเลขกลับ เช่น ๕ ร , ๖ ๙		

ที่	ลักษณะ / พฤติกรรม	ผลการวิเคราะห์	
		ใช่	ไม่ใช่
	๔. ด้านพฤติกรรมทั่วไป		
๑	ไม่ทำตามคำสั่ง ทำงานไม่เสร็จ		
๒	มีความยากลำบากในการจัดระบบงาน		
๓	ทำของหายบ่อยๆ เป็นประจำ เช่น ของเล่น ดินสอ หนังสือ อุปกรณ์การเรียน		
๔	ลืมทำกิจกรรมที่เป็นกิจวัตรประจำวัน		
๕	สับสนด้านซ้าย ขวา		
๖	วางสิ่งของเครื่องใช้ไม่เป็นระเบียบ		
๗	เสียสมาธิง่ายมองไปยังทุกสิ่งที่เคลื่อนไหวผ่านสายตา		
๘	มีอาการเครียดขณะอ่าน		
๙	ใจไม่ค่อยจดจ่ออยู่กับงาน มองโน่นมองนี่บ่อยๆ		
๑๐	หลีกเลี่ยง ไม่ชอบ หรือลังเลใจที่จะทำงานหรือการบ้านที่ต้องมีระเบียบและใส่ใจในงาน		

การวิจัยครั้งนี้จะเปรียบเทียบความถูกต้องการทำนายโมเดลจากการใช้แบบคัดกรองของกระทรวงศึกษาธิการกับแบบคัดกรองที่พัฒนาขึ้นใหม่ที่ประกอบด้วยแบบทดสอบด้านการอ่าน ด้านการเขียน ด้านการคำนวณ และข้อมูลพื้นฐานผู้เรียน

2.2 กระบวนการทำเหมืองข้อมูล

2.2.1 ความหมายของเหมืองข้อมูล

การทำเหมืองข้อมูล ได้มีผู้เชี่ยวชาญให้ความหมายไว้อย่างหลากหลาย ดังนี้

เหมืองข้อมูล หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น เพื่อช่วยในการตัดสินใจ บอกแนวโน้มสิ่งที่จะเกิดขึ้นในอนาคต

เหมืองข้อมูล หมายถึง กระบวนการของการกลั่นกรองสารสนเทศที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ เพื่อทำนายแนวโน้มและพฤติกรรม โดยอาศัยข้อมูลในอดีต และเพื่อใช้สารสนเทศเหล่านี้ในการสนับสนุนการตัดสินใจในการทำงานด้านต่างๆ

เหมืองข้อมูล หมายถึง ขบวนการทำงานที่เรียกว่า process ที่สกัดข้อมูลจากฐานข้อมูลขนาดใหญ่เพื่อให้ได้สารสนเทศที่เรายังไม่รู้โดยเป็นสารสนเทศที่มีเหตุผล และสามารถนำไปใช้ได้ ซึ่งเป็นสิ่งสำคัญในการที่จะช่วยการตัดสินใจในการทำธุรกิจ

เหมืองข้อมูล หมายถึง การค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่

สรุปการทำเหมืองข้อมูล หมายถึง การวิเคราะห์ข้อมูลจากฐานข้อมูลขนาดใหญ่ เพื่อให้ทราบความสัมพันธ์ รูปแบบของข้อมูลนั้นๆ ให้สามารถนำไปใช้ประโยชน์ได้ เช่น การตัดสินใจ การวางแผน การทำนายแนวโน้มสิ่งที่จะเกิดขึ้นในอนาคต หรือเป็นการแปรเปลี่ยนข้อมูลไปสู่ความรู้ใหม่ๆ

2.2.2 ความสำคัญของเหมืองข้อมูล

1) สกัด คัดเลือกข้อมูลมาใช้ในส่วนที่ตรงกับความต้องการจริงๆ

2) คาดการณ์สิ่งที่จะเกิดขึ้นล่วงหน้า เพื่อใช้ในการวางแผนพัฒนา

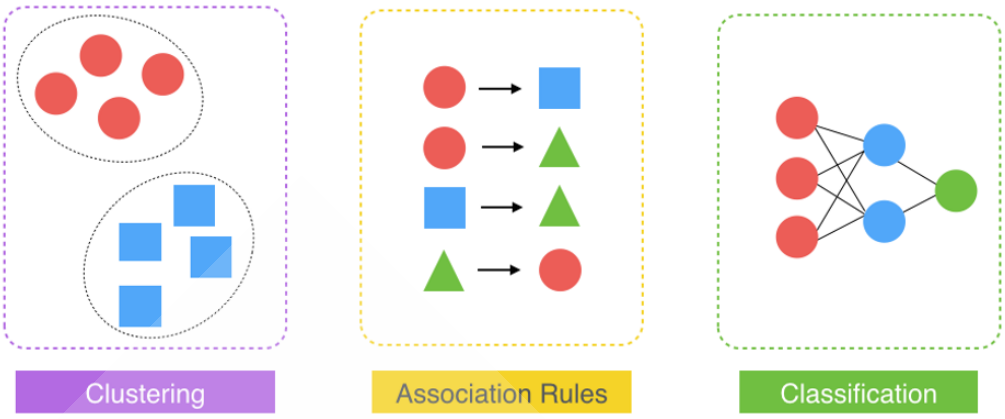
3) ดึงข้อมูลจากฐานข้อมูลขนาดใหญ่หลาย ๆ ฐานข้อมูลนำมาใช้ประโยชน์สูงสุด เพราะได้คำตอบจากข้อมูลที่ซับซ้อนซ่อนเร้นที่อยู่ในข้อมูลขนาดใหญ่หลาย ๆ ฐาน

4) ลดความเสี่ยงในการทำงานขององค์กรต่างๆ เพราะมีฐานความรู้ที่ได้จากการทำดาต้าไมน์นิงมาใช้ในการสร้างกรอบการทำงานวางแผนกลยุทธ์ทันกับการเปลี่ยนแปลงตรงกับพฤติกรรมผู้รับบริการ

5) วิเคราะห์ข้อมูลแบบเรียลไทม์ (real time) ให้ผลย้อนกลับทันที ปรับปรุงแก้ไขปัญหาได้ทันเหตุการณ์

2.2.3 เราใช้เหมืองข้อมูล ทำอะไรได้บ้าง

ในทางปฏิบัติที่มีการนำเหมืองข้อมูลมาใช้อย่างแพร่หลายเป็นรูปธรรม จะเป็นการนำไปใช้แก้ไขปัญหาที่เกี่ยวข้องกับธุรกิจ ไม่ว่าจะเป็นการวิเคราะห์กลุ่มลูกค้า การเสนอหรือการให้คำแนะนำเพิ่มเติมแก่ลูกค้า การค้นหาความผิดปกติของชุดข้อมูลโดยเฉพาะที่เกี่ยวกับการเงิน หรือนำมาใช้แก้ปัญหาด้านการศึกษา ไม่ว่าจะเป็นการวิเคราะห์พฤติกรรมการเรียน การเสนอหรือการให้คำแนะนำเพิ่มเติมแก่ผู้เรียนเป็นรายบุคคล การเลือกเรียนในสาขาที่เหมาะสมกับความสามารถของตนเอง การพยากรณ์หรือคาดคะเนแนวโน้มที่จะเกิดขึ้น เช่น จะสอบผ่าน/ไม่ผ่านในรายวิชาที่ลงทะเบียนเรียน ในการทำดาต้าไมน์นิ่งต้องมีการจัดกลุ่มประเภทลักษณะของข้อมูลให้ชัดเจน ซึ่งสามารถแบ่งประเภทข้อมูลได้หลายแบบ ดังภาพที่ 2.1



ภาพที่ 2.1 การจำแนกแบ่งประเภทข้อมูลในการทำดาต้าไมน์นิง

ที่มา: https://dataminingtrend.com

แต่ละประเภทมีรายละเอียดสรุปได้ ดังนี้

1) การจัดกลุ่ม (Clustering)

เป็นการจัดข้อมูลออกเป็นกลุ่มย่อยตามลักษณะความคล้ายคลึงของตัวข้อมูลเอง โดยต้องการให้ข้อมูลที่อยู่ในกลุ่มเดียวกันมีความคล้ายคลึงกันมากที่สุด และข้อมูลที่อยู่ต่างกลุ่มกันมีความต่างกันมากที่สุด หรือ “การจัดกลุ่มในพวกเดียวกัน คล้ายคลึงกัน” เช่น การแบ่งสินค้าตามรูปร่าง แบ่งสินค้าตามราคา หรือแบ่งสินค้าตามชนิดการใช้งาน สิ่งที่เห็นจากการใช้ข้อมูลประเภทนี้ในทางการตลาดที่แพร่หลาย เช่น การแบ่งกลุ่มลูกค้าที่มีลักษณะความต้องการที่คล้ายคลึงกัน แล้วจัดโปรโมชั่นให้ตรงกับความต้องการของกลุ่มนั้นๆ ได้ตรงใจกลุ่มเป้าหมายแต่ละกลุ่ม ข้อมูลประเภทนี้เป็นการวิเคราะห์ดาต้าไมน์นิงโดยใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยอัลกอริทึมที่นิยมใช้ในการวิเคราะห์ คือ K-means และ Hierarchical ; Agglomerative Clustering ; Density base ตัวอย่างที่ได้จากการวิเคราะห์ประเภทนี้ คือ (1) การจัดกลุ่มลูกค้า (Segmentation) (2) การจัดกลุ่มดอกไม้ (3) การจัดกลุ่มของพื้นที่จากข้อมูลภาพถ่ายดาวเทียม ฯลฯ

2) การจัดความสัมพันธ์ (Association Rule)

เป็นการเก็บข้อมูลพฤติกรรมการรับบริการด้านต่างๆของผู้บริโภค/ลูกค้า เพื่อนำมาวิเคราะห์หากฎความสัมพันธ์ (Association Rule) หรือ “เป็นการหาความสัมพันธ์ของข้อมูลที่เกิดร่วมกัน” เช่น ในทางธุรกิจจะใช้ค้นหาสินค้าที่มีการซื้อร่วมกันบ่อย ๆ แล้วนำเสนอให้ลูกค้ารายใหม่

เช่น เมื่อลูกค้าซื้อขนมปังจะซื้อแยมด้วย หรือ เมื่อลูกค้าซื้อเบียร์จะซื้อผ้าอ้อมด้วย ข้อมูลประเภทนี้เป็นการวิเคราะห์ดาต้าไมน์นิงโดยใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยอัลกอริทึมที่นิยมใช้ในการวิเคราะห์ คือ Apriori algorithm และ Frequent Pattern Growth (FP) ตัวอย่างที่ได้จากการวิเคราะห์ประเภทนี้ คือ การค้นหาพฤติกรรมการซื้อสินค้าข้ามสายผลิตภัณฑ์จากข้อมูลการซื้อในอดีตเพื่อแนะนำให้ลูกค้าใหม่ได้ทราบเป็นทางเลือก

3) การจัดจำแนก (Classification)

เป็นการนำข้อมูลที่มีในอดีตมาสอนระบบเพื่อให้เรียนรู้รูปแบบที่เกิดขึ้นในข้อมูล จากนั้นนำมาสร้างเป็นสมการหรือโมเดลขึ้นมา เพื่อหาคำตอบให้สำหรับข้อมูลใหม่ หรือ “เป็นโมเดลที่ใช้สำหรับนำข้อมูลที่มีอยู่มาทำนายอนาคต” เช่น การจำแนกอีเมล์ออกเป็นแบบสแปมหรือแบบปกติซึ่งต้องมีการนำข้อมูลของอีเมล์แบบสแปมกับแบบปกติมาให้คอมพิวเตอร์ทำการเรียนรู้เสียก่อน หลังจากนั้นจึงสร้างโมเดลการจำแนกประเภทของอีเมล์และใช้การจำแนกอีเมลที่เข้ามาใหม่ว่าเป็นแบบสแปมหรือแบบปกติ (คำตอบจากการวิเคราะห์จะเรียกว่า “คลาส” หรือ “ลาเบล”) ถือได้ว่าข้อมูลประเภทนี้เป็นการวิเคราะห์ดาต้าไมน์นิ่งโดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (supervised learning) โดยอัลกอริทึมที่นิยมใช้ในการวิเคราะห์ เช่น decision trees, rule induction, k-nearest neighbors, naïve Bayesian, artificial neural networks, และ support vector machines. ตัวอย่างที่ได้จากการวิเคราะห์ประเภทนี้ เช่น (1) ให้ผู้เลือกตั้ง Vote เลือก ส.ส. แล้วสร้างโมเดลทำนาย (2) ให้ลูกค้าตอบแบบสอบถาม แล้วสร้างโมเดลทำนายประเภทลูกค้า (3) การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำ ระดับกลาง และระดับสูง ของความเสี่ยงที่จะได้รับ ฯลฯ

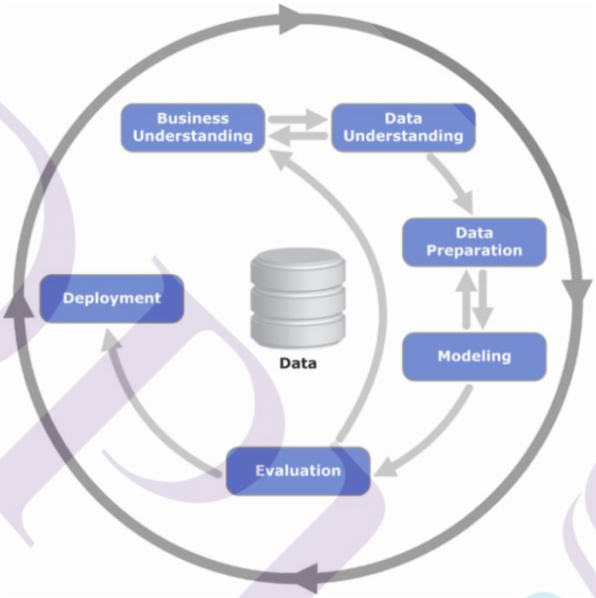
4) สมการถดถอย (Regression) หรือการประมาณค่าข้อมูล

จะมีลักษณะคล้ายกับการจัดจำแนก (Classification) แต่มีข้อแตกต่างกันที่คำตอบที่ต้องการทำนายซึ่ง Classification จะทำนายข้อมูลที่เป็นลักษณะมาตรวัดนามบัญญัติ (Nominal Scale) คือ เป็นแค่การกำหนดสัญลักษณ์หรือตัวเลขขึ้นมาเพื่อจำแนกประเภทสิ่งของหรือคุณลักษณะต่างๆ ออกเป็นกลุ่ม แต่จะไม่ได้แสดงถึงปริมาณ (มากหรือน้อย) หรือความสูง-ต่ำ ซึ่งไม่สามารถจัดลำดับก่อน-หลังได้ เช่น เพศ ก็แยกได้แค่เพศชายกับ เพศหญิงเท่านั้น ส่วนการประมาณค่าข้อมูล หรือ Regression จะใช้กับคำตอบที่เป็นเชิงปริมาณ หรือ จำนวนตัวเลขเป็นหลัก เช่น คาดการณ์ว่าพรุ่งนี้จะมีอุณหภูมิเท่าไหร่ ก็ต้องไปเก็บข้อมูลก่อนหน้านี้ ซึ่งจะสนใจเก็บข้อมูลที่เป็นตัวเลข ข้อมูลประเภทนี้เป็นการวิเคราะห์ดาต้าไมน์นิงโดยใช้เทคนิคการเรียนรู้แบบมีผู้สอน (supervised learning) โดยอัลกอริทึมที่นิยมใช้ในการวิเคราะห์ คือ สมการถดถอยเชิงเส้น (Multiple Linear Regression) และการถดถอยแบบ *โลจิสติกส์* (Logistic Regression) ตัวอย่างที่ได้จากการ

วิเคราะห์ประเภทนี้ เช่น ทำนายยอดขายสำหรับปีต่อๆ ปี โดยใช้จำนวนลูกค้าและงบการตลาดเป็นตัวแปรอิสระ หรือทำนายรายได้รวมต่อครอบครัว

2.2.4 กระบวนการทำเหมืองข้อมูล

กระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านดาต้าไมน์นิงได้พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler SPSS และNCR กระบวนการทำงานนี้เรียกว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อว่า “CRISP-DM” ดังภาพ 2.2



ภาพที่ 2.2 กระบวนการทำเหมืองข้อมูล

ที่มา: https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

แต่ละขั้นตอนของ CRISP-DM มีรายละเอียด ดังนี้

(1) Business Understanding เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิง พร้อมทั้งวางแผนในการดำเนินการคร่าวๆ ตัวอย่างเช่น การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติหรือ การใช้เทคนิคการทำเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนสารสนเทศ การใช้เทคนิคดาต้าไมน์นิงเพื่อพัฒนาคุณภาพ

การศึกษาคณะวิศวกรรมศาสตร์ หรือ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพัฒนาโปรแกรมการเรียนรู้เฉพาะบุคคล สำหรับเด็กที่มีความต้องการพิเศษระดับปฐมวัย

(2) Data Understanding ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับขั้นที่ 1 แล้วตรวจสอบความถูกต้องและความน่าเชื่อถือของข้อมูล และพิจารณาปริมาณข้อมูลที่เพียงพอในการนำไปวิเคราะห์

(3) Data Preparation เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมาอาจอยู่ในรูปแบบข้อความ ตัวอักษร ภาพ เสียง ฯลฯ ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ (แปลงให้ถูกต้อง เป็นระเบียบ จัดกลุ่ม หมวดหมู่ให้ง่ายในการวิเคราะห์ ใส่รหัสเพื่อให้โปรแกรมคอมพิวเตอร์มองออก) โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไปดังภาพที่ 2.3

ขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM จึงมีการพัฒนาเครื่องมือหรือซอฟต์แวร์สำหรับงานนี้ขึ้นมาใช้โดยเฉพาะ เช่น RapidMiner, Weka, R, SPSS

Stu_code	Sex	Address	SchoolGPA	…	GPA
37058063	male	Bangkok	2.5	….	2.3
37058167	male	Songkla	3.4	….	3.2
……	….	……	……	….	….

Stu_code	Sub_code	Section	Term	Year	Grade
37058063	204111	2	1	2537	C+
37058063	403111	6	1	2537	D
37058063	208111	1	1	2537	B+
……	……..	……	…….	……	…

เอกสาร	ข้อความในเอกสาร
1	ชาวนาหลายจังหวัดอีสานปลื้มทหาร หลังได้รับเงินจำนำข้าว
2	ชาวนาโคราช แห่รับเงินจำนำข้าวจาก ทหาร-ธ.ก.ส. พร้อมขอบคุณ คสช.
3	เกษตรกร เฮ ผู้ว่าฯปราจีนบุรี นำเงินจำนำข้าวคืนชาวนาครบทุกราย

เอกสาร	ชาวนา	ปลื้ม	รับเงิน	จำนำข้าว	ทหาร	ขอบคุณ
1	Y	Y	Y	Y	Y	N
2	Y	N	Y	Y	Y	Y
3	Y	N	N	Y	N	N

ภาพที่ 2.3 การแปลงข้อมูลแบบมีโครงสร้าง และแบบไม่มีโครงสร้างเพื่อนำไปวิเคราะห์

ที่มา: http://dataminingtrend.com

(4) Modeling เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิง ซึ่งต้องเลือกเทคนิคการวิเคราะห์ที่เหมาะสมกับข้อมูลที่ออกแบบไว้ในขั้นตอนที่ 3 โดยเทคนิคการวิเคราะห์ข้อมูลจะแบ่งหลักๆ คือ 1) การจัดกลุ่ม (Clustering) โดยอัลกอริทึมที่นิยมใช้ เช่น K-means และ Hierarchical ; Agglomerative Clustering ; Density base 2) การจัดความสัมพันธ์ (Association

Rule) โดยอัลกอริทึมที่นิยมใช้ เช่น Apriori algorithm และ Frequent Pattern Growth (FP) 3) การจัดจำแนก (Classification) โดยอัลกอริทึมที่นิยมใช้ เช่น decision trees, rule induction, k-nearest neighbors, naïve Bayesian, artificial neural networks, และ support vector machines.

(5) Evaluation เป็นการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใดก่อนนำไปใช้จริง กรณีที่มีการสร้างโมเดลด้วยเทคนิค Classification การวัดประสิทธิภาพของโมเดลจำเป็นต้องแบ่งข้อมูลออกเป็น 2 ส่วน โดยส่วนที่ 1 ใช้เพื่อสร้างโมเดล หรือ เรียกว่า ส่วน Training data ส่วนที่ 2 ให้โมเดลทำนายคลาสออกมา หรือ เรียกว่าส่วน Testing data การแบ่งข้อมูลเพื่อทำการทดสอบนี้มี 3 วิธีการใหญ่ๆ คือ 1) วิธี Self-Consistency Test 2) วิธี Split Test 3) วิธี Cross-validation Test

(6) Deployment เป็นการนำองค์ความรู้จากการทำดาต้าไมน์นิงไปใช้จริงในองค์กรหรือบริษัท เช่น การสร้างรายงานเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจได้ง่ายและสามารถนำไปออกโปรโมชั่นที่ตรงกับกลุ่มเป้าหมาย หรือการทำโฆษณาใน google หรือ amazon, lazada, youtube ระบบเล่นอัตโนมัติ เป็นต้น

2.2.5 การทดสอบประสิทธิภาพของโมเดลจากการทำเหมืองข้อมูล

การสร้างโมเดลขึ้นทุกครั้งต้องมีการทดสอบประสิทธิภาพของโมเดลที่สร้างได้ ซึ่งจะแบ่งเป็น 2 ส่วน คือ 1) วิธีการแบ่งข้อมูลเพื่อทำการทดสอบโมเดล 2) ตัวที่ใช้วัดประสิทธิภาพโมเดล

2.2.5.1 วิธีการแบ่งข้อมูลเพื่อทำการทดสอบโมเดล สามารถแบ่งข้อมูลเพื่อทำการทดสอบนี้มี 3 วิธีการ คือ

(1) วิธี Self Consistency Test หรือบางครั้งเรียกว่า Use Training Set เป็นวิธีการที่ง่ายที่สุดนั่นคือข้อมูลที่ใช้ในการสร้างโมเดล และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน การวัดประสิทธิภาพด้วยวิธีนี้จะให้ผลการวัดประสิทธิภาพที่มีค่าสูงมาก (อาจจะเข้าใกล้ 100%) เนื่องจากเป็นข้อมูลชุดเดิมที่ระบบได้ทำการเรียนรู้มาแล้ว แต่ผลการวัดที่ได้ไม่เหมาะที่จะนำไปรายงานในงานวิจัยต่างๆ ซึ่งวิธีการนี้เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น

(2) วิธี Split Test เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่ง (70% หรือ 80%) ใช้ในการสร้างโมเดลและข้อมูลส่วนที่สอง (30% หรือ 20%) ใช้ใน การทดสอบประสิทธิภาพของโมเดล ข้อดีของวิธีการนี้คือใช้เวลาในการสร้างโมเดลน้อยซึ่งเหมาะกับชุดข้อมูลที่มีขนาดใหญ่

(3) วิธี Cross-validation Test เป็นวิธีที่นิยมในการทำงานวิจัย เพื่อใช้ในการทดสอบประสิทธิภาพของโมเดลเนื่องจากผลที่ได้มีความน่าเชื่อถือ โดยวิธีนี้นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน (มักจะแสดงด้วยค่า k) เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หรือ 10-fold cross-validation คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้

2.2.5.2 ตัวที่ใช้วัดประสิทธิภาพของโมเดล โดยทั่วไปแล้วจะมีตัววัดประสิทธิภาพที่นิยมใช้ คือ

- Precision เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส
- Recall เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส
- F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส
- Accuracy เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส

วิธีการคำนวณ Precision, Recall, F-measure, Accuracy มีรายละเอียด ดังนี้

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

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

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

Accuracy = จำนวน True Positive ของทุกคลาสรวมกัน โดยสมการในการทดสอบ คือ

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

เพื่อให้เห็นภาพตัววัดประสิทธิภาพของโมเดลแต่ละตัวแสดงดังตัวอย่างต่อไปนี้

ตารางที่ 2.1 ข้อมูลสภาพอากาศย้อนหลัง 10 วัน

No.	Outlook	Temperature	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	mild	normal	false	Yes
10	rainy	mild	normal	false	yes

จากตารางที่ 2.1 มีคลาสคำตอบอยู่ 2 ค่า คือ yes และ no นำมาสร้างตาราง confusion matrix ได้เป็นตารางขนาด 2x2 ดังในตารางที่ 2 โดยข้อมูลด้านคอลัมน์ คือ คลาสที่อยู่ในข้อมูลเทรนนิงดาต้า (actual) และข้อมูลในแนวแถว คือ คลาสที่โมเดลทำนายมาได้ (predicted)

ตารางที่ 2.2 แสดงตาราง confusion matrix ของข้อมูล weather ซึ่งมี 2 คลาส

คลาสที่โมเดลทำนายมาได้ (predicted)	ข้อมูลเทรนนิงดาต้า (actual)	
	yes	**no**
yes	TP	FP
no	FN	TN

จากในตารางที่ 2.2 ค่าที่แสดงในช่องต่างๆ ของตารางมีความหมาย คือ

True Positive (TP) คือ สิ่งที่โปรแกรมทำนายว่าจริง และเป็นข้อมูลจริง

True Negative (TN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง และเป็นข้อมูลไม่จริง

False Positive (FP) คือ สิ่งที่โปรแกรมทำนายว่าจริง แต่เป็นข้อมูลไม่จริง

False Negative (FN) คือ สิ่งที่โปรแกรมทำนายว่าไม่จริง แต่เป็นข้อมูลจริง

ตารางที่ 2.3 แสดงข้อมูลแอตทริบิวต์ Play จากเทรนนิงดาต้า (actual) 10 ตัวแรก และค่าที่ทำนายได้ (predicted)

No.	Actual	Predicted
1	no	no
2	no	no
3	yes	no
4	yes	yes
5	yes	no
6	no	yes
7	yes	yes
8	no	no
9	yes	no
10	yes	yes

จากตารางที่ 2.3 โดยที่กำลังพิจารณาคลาส Play = yes จะสามารถสรุปได้ว่า

- True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาส Play = yes
  - มีจำนวน 3 ตัว (แถวที่ 4, 7 และ 10)
- True Negative (TN) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาส Play = no
  - มีจำนวน 3 ตัว (แถวที่ 1, 2 และ 8)
- False Positive (FP) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาส Play = yes
  - มีจำนวน 1 ตัว (แถวที่ 6)
- False Negative (FN) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาส Play = no
  - มีจำนวน 3 ตัว (แถวที่ 3, 5 และ 9)

ตารางที่ 2.4 แสดงตาราง confusion matrix ของข้อมูล weather ซึ่งมี 2 คลาส

คลาสที่โมเดลทำนายมาได้ (predicted)	ข้อมูลเทรนนิงดาต้า (actual)	
	yes	**no**
yes	3	1
no	3	3

จากตารางที่ 2.4 นำไปคำนวณค่า Precision, Recall, F-measure และ Accuracy ตามวิธีการคำนวณข้างต้น

2.3 เทคนิคที่ใช้ในการทำเหมืองข้อมูล

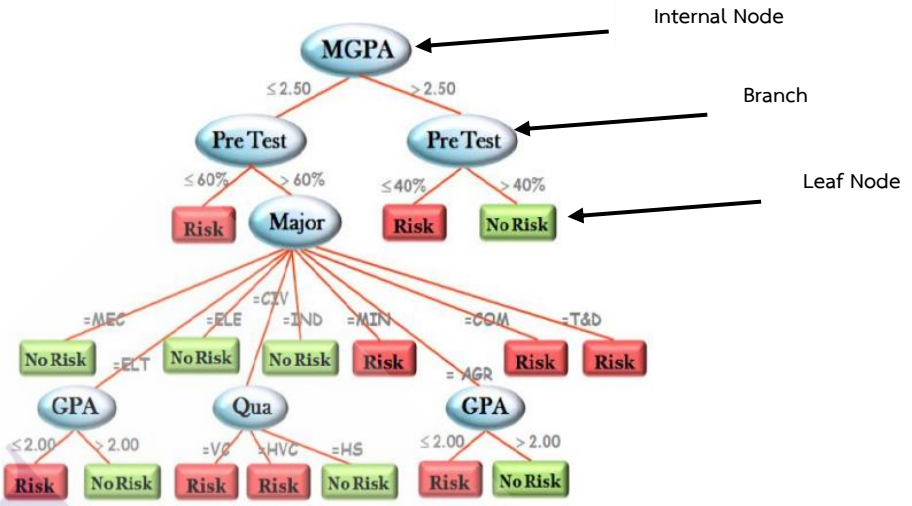
เทคนิคที่ใช้ในการทำเหมืองข้อมูลเพื่อการวิจัยครั้งนี้ประกอบด้วย

2.3.1 ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคที่นิยมใช้ในการจัดหมวดหมู่ข้อมูลมักใช้ในการตรวจสอบข้อมูลเพื่อพยากรณ์ เทคนิคนี้จะมีลักษณะคล้ายโครงสร้างต้นไม้ โดยการแตกแขนงไปตามเงื่อนไขหรือ เส้นทางของกิ่งไม้ และข้อมูลที่คาดคะเนไว้ว่าจะเกิดขึ้น ซึ่งจะใช้กฎในรูปแบบ “ถ้า (เงื่อนไข) แล้ว (ผลลัพธ์)” (If-then Rule) มาประกอบการสร้างโครงสร้างต้นไม้ตัดสินใจ สำหรับโครงสร้างต้นไม้ตัดสินใจประกอบด้วย ดังนี้

• โหนดภายใน (Internal Node) คือโหนดที่แสดงถึงคุณลักษณะ (Feature) ที่นำมาใช้ในการแบ่งกลุ่มของข้อมูลซึ่งมีโหนดราก (Root Node) อยู่บนสุดของโครงสร้าง ซึ่งเป็นโหนดที่มีอิทธิพลต่อการจำแนกกลุ่มมากที่สุด

• กิ่ง (Branch) เป็นตัวเชื่อมระหว่างโหนดที่ใช้เป็นเงื่อนไขหรือทางเลือกของการกระทำซึ่งมาจากผลลัพธ์แต่ละตัวของทุกตัวทำนาย (Predictor) หรือคุณสมบัติ(Feature)

• โหนดใบ (Leaf Node) เป็นโหนดที่แสดงผลลัพธ์ของเงื่อนไข หรือการกระทำตามเงื่อนไขที่เกิดขึ้น ดังภาพที่ 2.4



ภาพที่ 2.4 โครงสร้างต้นไม้ตัดสินใจ

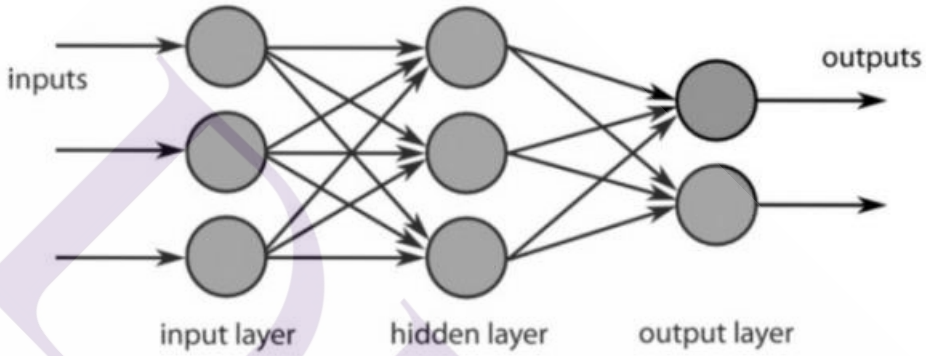
ที่มา: สุวิมล สิทธิชาติ. (2560)

สำหรับการสร้าง Decision Tree เริ่มต้นที่การคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุด ขึ้นมาเป็นโหนดบนสุดของต้นไม้ (Root Node) หลังจากนั้นจะทำการแตกกิ่งแอตทริบิวต์ออกไปเรื่อยๆ จนสามารถแบ่งข้อมูลออกเป็นคลาสได้ชัดเจน

2.3.2 นาอีฟเบย์ (Naïve Bayes) เป็นเทคนิคการทำเหมืองข้อมูลอย่างง่ายโดยนำโมเดลมาใช้ในการคิดแยกประเภทข้อมูลผ่านหลักความน่าจะเป็นที่อยู่บนพื้นฐานของทฤษฎี Bayes และสมมติฐานของการเกิดของเหตุการณ์เป็นอิสระต่อกัน เทคนิคนี้จะไม่มีการหมุนวนที่ซับซ้อนส่งผลให้สามารถทำงานได้ดีและมีประโยชน์กับชุดข้อมูลขนาดใหญ่อย่างมาก ทำให้เทคนิคนี้ถูกใช้อย่างแพร่หลาย เทคนิคนี้จะคำนวณจากทฤษฎีโดยสันนิษฐานว่า ผลลัพธ์หรือค่าที่เกิดจากตัวที่ใช้ทำนาย (predictor) เป็นอิสระต่อกันโดยเขียนเป็นสมการ $P(c \mid x) = P(x \mid c) P(c) / P(x)$

โดย $P(c \mid x)$ คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น x จะมีคลาส c; $P(x \mid c)$ คือ ค่าความน่าจะเป็นที่ข้อมูลในชุดข้อมูลสอนที่มีคลาส c และมีแอตทริบิวต์ x โดยที่ $x = x1 \cap x2 \ldots \cap x M$ โดยที่ M คือ จำนวน แอตทริบิวต์; P(c) คือ ค่าความน่าจะเป็นของคลาส C; และ P(x) คือ ค่าความน่าจะเป็นของแอตทริบิวต์ x 7 [21]

2.3.3 โครงข่ายประสาท (Neural Network) เป็นการใช้โมเดลทางคณิตศาสตร์มาประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) โดยได้แนวคิดมาจากการจำลองการทำงานของเซลล์สมองมนุษย์ที่แต่ละเซลล์ประสาทจะประกอบไปด้วยเดนไดรท์ (Dendrite) หรือปลายในรับกระแสประสาท ซึ่งเป็นตัว input ของเซลล์และแอกซอน (Axon) เป็นเสมือน output ของเซลล์ที่จะส่งกระแสประสาทไปยังเซลล์ตัวอื่นถ้าสมองได้รับกระแสไฟฟ้ามากพอ จะทำให้เซลล์ส่งต่อกระแสประสาทไปเรื่อยๆ ดังภาพที่ 2.5



ภาพที่ 2.5 โครงสร้าง Layer ของโครงข่ายประสาทเทียมแบบหลายชั้น

ที่มา: https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetworkBigger_english.png

สำหรับโครงสร้างของประสาทเทียมจะประกอบด้วย input และ output เช่นกัน โดยแบ่งเป็นชั้นหรือ layer ซึ่งจะมีชั้นคั่นตรงกลางคือ hidden layer โดยโครงสร้างประสาทเทียมจะมีหน่วยย่อย เรียกว่า perceptron ซึ่งเทียบเท่าได้กับเซลล์สมองของมนุษย์หนึ่งเซลล์

2.4 โปรแกรม RapidMiner Studio

ซอฟต์แวร์ RapidMiner Studio 7 แรก เริ่มพัฒนาขึ้นจากบริษัทที่ชื่อว่า Rapid-1 ในประเทศเยอรมนีและเมื่อช่วงปลายปี 2013 ได้รับทุนก้อนโตจากนักลงทุนในประเทศสหรัฐอเมริกา จึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทน และย้ายสำนักงานใหญ่มาอยู่ประเทศสหรัฐอเมริกา โปรแกรมนี้มีให้ใช้ทั้งใช้งานฟรี และชำระเงินเพื่อใช้งานบาง function เราสามารถดาวน์โหลดซอฟต์แวร์ RapidMiner Studio9 ซึ่งเป็นเวอร์ชั่นปัจจุบันได้จากเว็บไซต์ https://rapidminer.com ดังภาพที่ 2.6

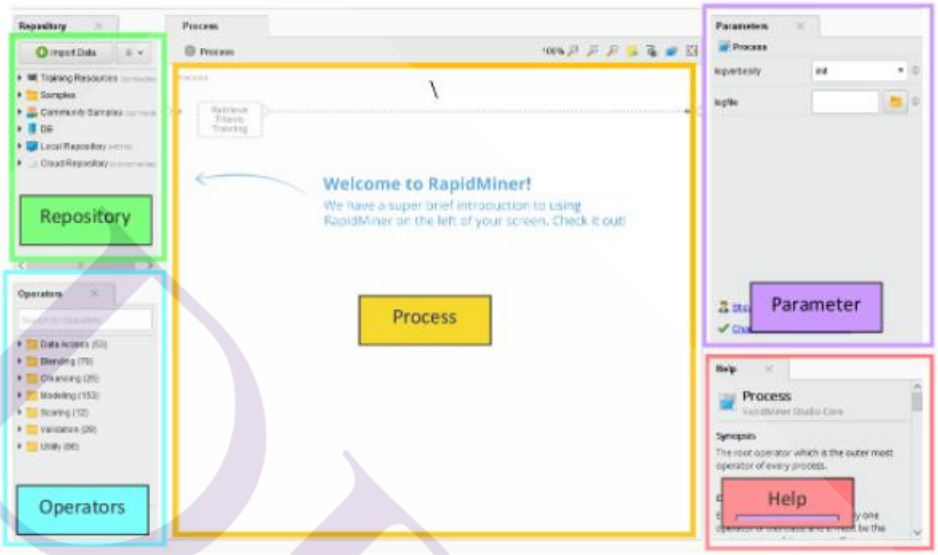
ภาพที่ 2.6 โปรแกรม RapidMiner Studio 9

หลังจากกด I already an account or license key. เข้ามาแล้วสามารถเข้าสู่ account โดยการกรอก E-mail และ Password และเข้าใชงานได้ทันที หรือเลือกที่ปุ่ม Manually enter license Key . เพื่อใส่ license Key ที่ได้จากการซื้อโปรแกรมใส่เพื่อใช้งานได้เช่นกัน ภาพที่ 2.7

ภาพที่ 2.7 การเข้าสู่ account โดยการกรอก E-mail และ Password กรณีใช้ฟรีและการใส่ license Key กรณีที่ได้จากการซื้อโปรแกรมใส่เพื่อใช้งาน

องค์ประกอบของ RapidMiner Studio 9 ประกอบด้วย

1) Repository เป็นส่วนสำหรับจัดการไฟล์ RapidMiner จะจัดการข้อมูลจาก 3 แหล่ง คือ DB ,Local (ในเครื่องคอมพิวเตอร์ที่ใช้อยู่), และ Cloud Reposilory โดยเก็บไฟล์ Data Set และ Process ต่างๆ แยกเก็บไว้คนละโฟลเดอร์ดังภาพที่ 2.8



ภาพที่ 2.8 องค์ประกอบของ RapidMiner Studio 9

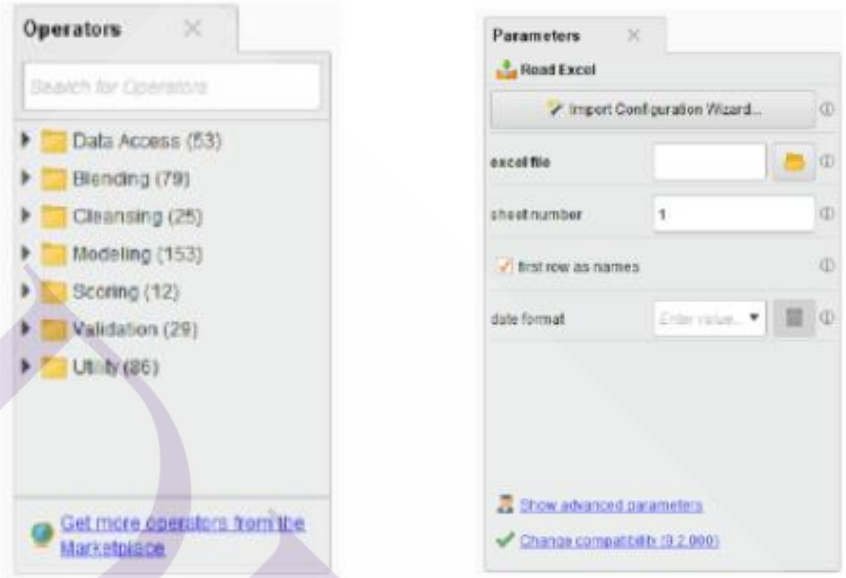
2) Operators เป็นส่วนที่ใช้เก็บตัวโอเปอเรเตอร์ที่ใช้ในการทำงานทั้งหมด ซึ่งจัดเป็นกลุ่ม ๆ โดยกลุ่มที่ใช้งานคล้ายคลึงกันจะจัดอยู่ในกลุ่มเดียวกัน มี 8 กลุ่ม คือ 1) Data Access 2) Blending 3) Cleansing 4) Modeling 5) Scoring 6) Validation 7) Utility 8) Extensions ดังภาพที่ 2.9 โอเปอเรเตอร์แต่ละตัวจะประกอบด้วย

- ชื่อของโอเปอเรเตอร์
- อินพุท พอร์ต (Input port)เป็นส่วนรับข้อมูลเข้ามาประมวลผล
- เอาท์พุต พอร์ต (Output por!) เป็นส่วนส่งผลลัพธ์ที่ประมวลผลได้

โดยชื่ออินพุท พอร์ตและชื่อเอาท์พุต พอร์ต จะแสดงด้วยตัวอักษร 3 ตัวแรกของชื่อพอร์ต เช่น exa ย่อมาจาก example set เป็นต้น

3) Process เป็นหน้าหลักในการทำงานในการสร้างโปรเซส สำหรับทำ Machine Learning ของซอฟต์แวร์นี้ โดยจะนำโอเปอเรเตอร์มาประกอบเพื่อสร้างโปรเซสขึ้นตามวัตถุประสงค์ของโจทย์ที่ตั้งไว้

4) Parameters เป็นส่วนสำหรับแสดงพารามิเตอร์ (parameter) ที่เกี่ยวข้องกับแต่ละโอเปอเรเตอร์ เช่น โอเปอเรเตอร์ Read Excel ที่ใช้เพื่ออ่านไฟล์ประเภท Excel จะมีพารามิเตอร์ที่เกี่ยวข้อง เช่น ชื่อและที่อยู่ของไฟล์ Excel ดังภาพที่ 2.9



ภาพที่ 2.9 โอเปอเรเตอร์และส่วนประกอบพารามิเตอร์ของโอเปอเรเตอร์ Read Excel

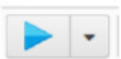
5) Help เป็นส่วนช่วยเหลือ ซึ่งจะแสดงรายละเอียดของตัวโอเปอเรตอร์ที่เลือกใช้งานอยู่ ส่วนช่วยเหลือของ RapidMiner จะบอกเพียงหน้าที่และรายละเอียดคร่าว ๆ ของโอเปอเรเตอร์ หากต้องการดูรายละเอียดมากกว่านั้นต้องไปที่ Jum to Tutoral Process ซึ่งจะลิงก์ไปยังเว็บไซค์ที่มีรายละเอียดที่เกี่ยวกับโอเปอเรเตอร์ที่ใช้อยู่

นอกจากทั้ง 5 ส่วนใหญ่ ๆ ที่ได้อธิบายแล้วยังมีส่วนมนูด้านบนเพิ่มเติมดังนี้

เมนูสำหรับสร้าง Process ใหม่

เมนูสำหรับการโหลดไฟล์ต่าง ๆ จาก repository

เมนูสำหรับการบันทึก Process หรือบันทึก Process เป็นชื่อใหม่


เมนูสำหรับสั่งให้ Process ทำงาน

เมนูสำหรับยกเลิกการทำงาน Process

1. แสดงหน้าจอการออกแบบ Design
2. แสดงหน้าจอผลลัพธ์การทำงาน Results
3. แสดงหน้าต่าง ออกแบบมาเพื่อให้การเตรียมข้อมูลง่ายขึ้น
4. แสดงหน้าต่างการสร้างโมเดล อัตโนมัติ Auto Model

ข้อดีของซอฟต์แวร์ RapidMiner Studio 9 สรุปได้ ดังนี้

- รองรับการใช้งานไฟส์ได้หลายประเภท เช่น ไฟส์ Excel 2007
- สามารถแสดงข้อมูลได้หลายรูปแบบ เช่น scatter plot 3D
- สามารถแสดงผลโมเดลที่สวยงามและแก้ไขการแสดงผลให้สามารถอ่านได้ง่ายขึ้น
- สามารถบันทึกไฟล์โมเดลออกเป็นไฟล์ภาพประเภทต่างๆ เช่น PNG, JPG หรือ PDF
- มีวิธีการเตรียมข้อมูล (preprocess) และการวิเคราะห์ได้หลากหลายรูปแบบ

2.5 เอกสารงานวิจัยที่เกี่ยวข้องเกี่ยวกับการทำเหมืองข้อมูล

งานวิจัยเรื่อง การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก ใช้เทคนิค Naïve Bayes และNeural Networks สำหรับจำแนกผู้ใช้ไฟฟ้าออกตามประเภทกิจการที่ได้กำหนดไว้จากผลการทดสอบ พบว่า Naïve Bayes ให้ความแม่นยำที่ร้อยละ 81.25 และNeural Networks ให้ความแม่นยำที่ร้อยละ 87.5

งานวิจัยเรื่อง การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัยของผู้ประกอบอาชีพด้านคอมพิวเตอร์ ใช้เทคนิค Neural Networks, Naive Bayes และ Decision Tree เพื่อค้นหาลักษณะนิสัยของผู้ประกอบอาชีพด้านคอมพิวเตอร์ตามทฤษฎีการเลือกอาชีพของฮอลแลนด์ พบว่า ตัวแบบที่มีประสิทธิภาพมากที่สุด คือ ตัวแบบที่ได้จากเทคนิค Naive Bayes มีค่า Accuracy เท่ากับ 79.64 เทคนิค Neural Network มีค่าเท่ากับ 77.82% และเทคนิค Decision Tree มีค่าเท่ากับ 54.69% ตามลำดับ

งานวิจัยเรื่อง การวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูล ใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจในการจำแนกประเภทข้อมูลและคัดเลือกคุณลักษณะที่สำคัญด้วย Filter Ranker Method โดยคำนวณค่าน้ำหนักด้วย Chi-Square และGain Ratio จากผลการศึกษา พบว่า ความแม่นยำในการจำแนกข้อมูลจาก 50 คุณลักษณะนั้น วิธีANN มีค่า 71.52% และ Decision Trees - J48 มีค่า 66.23% หลังจากทำการคัดเลือกคุณลักษณะแสดงให้

เห็นว่าการจำแนกข้อมูลด้วยวิธีANN จาก 5 คุณลักษณะแรกที่ได้จากการคัดเลือกโดยวิธี Filter Ranker Method ที่คำนวณค่าน้ำหนักด้วย Chi-Square ทำให้ได้ค่าความถูกต้องมีค่าสูงสุด คือ 80.13% และการจำแนกข้อมูลด้วย Decision Tree ก็ให้ผลไปในทางเดียวกัน โดยมีค่าความถูกต้องสูงสุดที่ 75.83%

งานวิจัยเรื่อง เทคนิคเหมืองข้อมูลสำหรับการพยากรณ์ผลสัมฤทธิ์ทางการศึกษาด้วยวิธีกรจัดการเรียนการสอนแบบผสมผสาน ข้อมูลที่ใช้สำหรับสร้างแบบจำลองได้จากชุดข้อมูลคะแนนการทดสอบของผู้เรียนในระหว่างการศึกษามีทั้งสิ้น 20 ตัวแปร การวิจัยได้แบ่งข้อมูลออกเป็น 2 ชุดข้อมูลโดยใช้วิธีการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธี information Gain และGain Ratio จากนั้นลดมิติของข้อมูลเพื่อนำไปวัดค่าประสิทธิภาพของแบบจำลองการพยากรณ์โดยใช้วิธี 10-fold Cross Validation ผลการวิจัยพบว่า แบบจำลองสำหรับการพยากรณ์แบบเคเนียเรสเนเบอร์ (KNN) มีค่าความถูกต้องสูงสุดเท่ากับ 86.13% ต้นไม้ตัดสินใจ (Decision Tree) เท่ากับ 81.74% กฎพื้นฐาน (Rule-Base) เท่ากับ 81.67% และนาอีฟเบย์ (Naive Bays) เท่ากับ 55.05%

งานวิจัยเรื่อง การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทสำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม งานวิจัยนี้นำเสนอการสร้างตัวจำแนกประเภทสำหรับคดักรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม โดยประยุกต์ใช้เทคนิคทางเหมืองข้อมูล ประกอบด้วย เทคนิคการจัดกลุ่ม โดยใชอัลักอริทึมเคมีนเพื่อ จัดกลุ่มผู้เรียน และสร้างคลาสคำตอบเพื่อหาจำนวนกลุ่มที่เหมาะสม เทคนิคกฎความสัมพันธ์โดยใช้อัลกอริทึมเอไพรออริ เพื่อหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ของนักเรียน และเทคนิคการจำแนกประเภทโดยใช้โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับเพื่อสร้างโมเดลจำแนกประเภทผู้เรียนโดยมีกระบวนการ ดังนี้ 1) เตรียมข้อมูล 2) คัดเลือกคุณลักษณะ 3) จัดกลุ่มข้อมูล 4) การหากฎความสัมพันธ์ 5) สร้างโมเดลจำแนกประเภท 6) ทดสอบประสิทธิภาพ และ7) การนำไปใช้จากการทดลองจดักลุ่มผู้เรียนสามารถจัดกลุ่มข้อมูลได้ 2 กลุ่ม คือ กลุ่มปกติและกลุ่มเสี่ยง นำผลลัพธ์ที่ได้ไปหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้โดยใช้อัลกอริทึมเอไพรออริโดยเลือกเฉพาะกลุ่มเสี่ยงเพื่อหาคุณลักษณะที่ส่งผลต่อความบกพร่องทางการเรียนรู้ และสร้างโมเดลจำแนกประเภทข้อมูลโดยใช้โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับ จากการทดลองสร้างโมเดลจำแนกประเภทและทดสอบโมเดลด้วยวิธี K-fold Cross Validation พบว่า โครงข่ายประสาทเทียมเหมาะสมสำหรับจำแนกประเภทผู้เรียนโดยให้ค่าความถูกต้อง 98.30% ค่าความคลาดเคลื่อน RMSE = 0.069 จากนั้นจึงพัฒนาเป็นระบบต้น แบบสำหรับคัดกรองเด็กบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม

โดยใช้ภาษาพีเอชพีร่วมกับเวก้าเอฟไอ จากผลการประเมินระดับความคิดเห็นของผู้ใช้งานมีค่าเฉลี่ย 4.34 ซึ่งอยู่ในเกณฑ์มาก

งานวิจัยเรื่อง การศึกษาเชิงจิตประสาทวิทยาและการพัฒนาแบบคัดกรองเน้นกระบวนการทางปัญญาในเด็กที่มีความเสี่ยงต่อภาวะความบกพร่องทางการเรียนรู้ด้านคณิตศาสตร์ การวิจัยนี้ประกอบไปด้วยช่วงศึกษานำร่องและทดสอบเครื่องมือจำนวน 2 ครั้ง และสุดท้ายได้ ชุดแบบคัดกรองฯ ที่ประกอบไปด้วยแบบคัดกรองย่อยจำนวน 6 แบบ แบ่งเป็น 9 แบบย่อย ดังต่อไปนี้ เปรียบเทียบจำนวนจุดและตัวเลข เปรียบเทียบตัวเลขหนึ่งและสองหลัก เส้นจำนวน เปรียบเทียบค่าและขนาดตัวเลขหนึ่ง และสองหลัก ตัวเลขสลับสีและบวกลบตัวเลขในใจ โดยห้าแบบคัดกรองย่อยแรกสะท้อนระบบตัวเลขโดยตรง ส่วนอีกสี่แบบคัดกรองย่อยที่เหลือสะท้อนระบบช่วยเหลือ นอกจากนี้แบบคัดกรองมาตรฐานผลรวมชุดตัวเลข ถูกนำมาใช้ทดสอบความตรงของแบบคัดกรองฯ ที่พัฒนาขึ้นและใช้จำแนกเด็กที่มีความเสี่ยงฯ กลุ่มตัวอย่างนักเรียนจำนวน 500 คน ความเที่ยงของแบบคัดกรองฯที่พัฒนาขึ้นวิเคราะห์ด้วยความสอดคล้องภายใน ส่วนความตรงของแบบคัดกรองฯ วิเคราะห์ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันและโมเดลสมการโครงสร้างกลุ่มพหุ และ MANOVA ใช้วิเคราะห์ความแตกต่างของคะแนนระหว่างกลุ่มและภายในกลุ่ม

ผลการวิจัย พบว่า แบบคัดกรองที่พัฒนาขึ้นมีค่าความเที่ยงอยู่ในระดับดีมากจนถึงดีเยี่ยม และความตรงอยู่ในระดับยอมรับได้ นอกจากนี้ผลการเปรียบเทียบคะแนนจากแบบคัดกรองระหว่างนักเรียนสองระดับชั้น ครอบคลุมทั้งระบบตัวเลขโดยตรงและระบบช่วยเหลือ พบว่า นักเรียนระดับชั้นอนุบาลปีที่ 3 ได้คะแนนต่ำกว่านักเรียนระดับชั้นประถมศึกษาปีที่ 1 ทุกแบบคัดกรองย่อยอย่างมีนัยสำคัญทางสถิติ ผลการเปรียบเทียบระหว่างนักเรียนชั้นอนุบาลปีที่ 3 กลุ่มเสี่ยงฯ และปกติ พบว่า นักเรียนกลุ่มเสี่ยงได้คะแนนต่ำกว่านักเรียนปกติแทบทุกแบบคัดกรองย่อยยกเว้นเส้นจำนวน ซึ่งสามารถตีความถึงพัฒนาการล่าช้าของทั้งสองระบบในนักเรียนกลุ่มเสี่ยงนี้เช่นเดียวกัน นักเรียนระดับชั้นประถมศึกษาปีที่ 1 กลุ่มเสี่ยงมีคะแนนต่ำกว่านักเรียนปกติอย่างมีนัยสำคัญทางสถิติในทุกด้านของแบบคัดกรองย่อยของระบบตัวเลขโดยตรง และเพียงแบบคัดกรองย่อยเปรียบเทียบค่าและขนาดตัวเลขหนึ่งและสองหลักของระบบช่วยเหลือ สะท้อนถึงพัฒนาการล่าช้าในระบบตัวเลขโดยตรงเป็นหลัก และบางส่วนของระบบช่วยเหลือ

สรุปจากผลการวิจัยนี้แสดงให้เห็นว่า เทคนิคเหมืองข้อมูลสามารถนำมาประยุกต์ใช้ทางด้านการศึกษาในการวิเคราะห์ข้อมูลเพื่อแก้ปัญหาเกี่ยวกับการเรียนการสอนได้ และเทคนิคที่ใช้จำแนกประเภทข้อมูลโดยใช้เทคนิค Classification ด้วยวิธี Decision Tree , Naive Bayes และ Neural Network และการนำเทคนิค Data Mining สามารถประยุกต์ใช้เพื่อการเรียนการสอนได้อย่างหลากหลาย เช่น ช่วยพัฒนาผลการศึกษาและพัฒนาการของผู้เรียน ช่วยในการแบ่งกลุ่มผู้เรียนใน

การทำรายงานหรือแบ่งกลุ่มเรียนตามลักษณะของผู้เรียนที่มีพฤติกรรมและรูปแบบการเรียนที่ใกล้เคียงกันหรือมีความชอบคล้ายคลึงกันมาอยู่กลุ่มเดียว เพื่อให้เกิดประสิทธิภาพในการเรียนสูงสุด ช่วยในการสร้าง/จัดโปรแกรมการเรียน-การสอนที่มีประสิทธิภาพ นักศึกษาจะมีโอกาสในการสร้างแผนการเรียนที่จะช่วยพัฒนาทักษะของตัวเองและเข้ากับตัวเองมากที่สุด ได้เรียนในวิชาที่เป็นความถนัดและได้ทำงานในส่วนที่สนใจ พัฒนาและเรียนรู้ผ่านประสบการณ์แบบเรียลไทม์ ทราบข้อมูลเชิงลึกของผู้เรียนแต่ละคนว่าเรียนเป็นอย่างไรบ้างในแต่ละระดับ ซึ่งผู้เรียนแต่ละคนก็จะมีรูปแบบการเรียนที่แตกต่างกันออกไป และวิธีการเรียนที่แตกต่างกันนั้นเองที่มีผลต่อผลการเรียนในรายวิชานั้นๆ และลดการถอน/เพิ่มผลลัพธ์ สามารถทำนายล่วงหน้านี้จะสามารถช่วยให้ทางสถาบันปรับเปลี่ยนโปรแกรมการสอนได้ทันหากผลการคาดการณ์ออกมาในแง่ลบปรับเปลี่ยนได้ทันท่วงที

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ได้ได้ทำการเก็บข้อมูล 3 ส่วน คือ ส่วนที่ 1ได้ทำการเก็บข้อมูลพื้นฐานที่ผู้วิจัยกำหนดเป็นปัจจัยที่ส่งผลต่อความบกพร่องทางการเรียนรู้ของนักเรียนชั้นประถมศึกษาปีที่ 4-6 โดยครูประจำชั้นเป็นผู้บันทึกผ่าน google form ส่วนที่ 2 ได้จากการทำแบบทดสอบด้านการอ่าน ด้านการเขียน และด้านการคำนวณที่ผู้วิจัยพัฒนาขึ้นมาใหม่ผ่าน google form และนำไปทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 4-6 จำนวน 8 โรงเรียน 238 คน ส่วนที่ 3 เป็นการใช้แบบคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ของกระทรวงศึกษาธิการเพื่อใช้เปรียบเทียบกับแบบคัดกรองฯ ที่ผู้วิจัยพัฒนาขึ้น วิธีดำเนินการวิจัยครั้งนี้ได้ดำเนินการตามกระบวนการทำเหมืองข้อมูลของ CRISP-DM ดังตาราง 3.1

ตารางที่ 3.1 แสดงขั้นตอนการดำเนินการวิจัย

ขั้นตอน	การดำเนินการ
(1) Business Understanding	-เด็กที่มีความบกพร่องทางการเรียนรู้ (LD) จำนวนเพิ่มขึ้น -การคัดกรองเด็ก LD จะทำได้อย่างลำบาก -การจัดแผนการเรียนที่ไม่เหมาะสมกับคุณลักษณะเฉพาะของผู้เรียนที่เป็น LD (แบบ IEP และ แบบ บุคคล IIP)
(2) Data Understanding	-ศึกษาเอกสาร ตำรา งานวิจัยที่เกี่ยวข้องกับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ - เก็บข้อมูลพื้นฐานผู้เรียน ได้แก่ เพศ จำนวน พี่น้อง ช่วงชั้น อายุ ผลการเรียนสะสมภาษาไทย ผลการเรียนสะสมคณิตศาสตร์ ผลการเรียนสะสมรวมทุกวิชา สถานะครอบครัว ผู้ปกครองปัจจุบัน ข้อมูลบิดา–มารดา (อายุ อาชีพ รายได้ ภูมิลำเนา) -จัดสร้างแบบทดสอบคัดกรอง ด้านการอ่าน ด้านการเขียน และด้านการคิดคำนวณผ่าน Google Form -สร้างแบบคัดกรองเด็กบกพร่องทางการเรียนรู้ของกระทรวงศึกษาธิการผ่าน Google Form

ตารางที่ 3.1 (ต่อ)

ขั้นตอน	การดำเนินการ
(3) Data Preparation	-ทำการคัดเลือกข้อมูให้มีความพร้อมสำหรับการวิเคราะห์ -ทำการกลั่นกรอง (Data Cleaning) ตรวจสอบความซ้ำซ้อนของข้อมูล -ปรับปรุงในส่วนของค่าข้อมูลที่ขาดหาย (Missing value) -การกำจัดข้อมูลว่าง (Null value) -ทำการแปลงรูปแบบของข้อมูล (Data transformation)
(4) Modeling	- นำข้อมูลเข้ามาวิเคราะห์เพื่อสร้างแบบจำลองเพื่อหากฎความสัมพันธ์ของข้อมูลต่างๆ ด้วยเทคนิคDecision Tree, Naïve Bayes, และNeural Network ด้วยโปรแกรม RapidMiner Studio
(5) Evaluation	-แบ่งกลุ่มข้อมูลเพื่อ Training dataและ Testing data โดยวิธี 10-fold cross-validation - เปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการ
(6) Deployment	-ครูผู้สอนนำโมเดลที่ได้ไปใช้เป็นแนวทางในการคัดคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้

แต่ละขั้นตอนมีรายละเอียด ดังนี้

3.1 Business Understanding

ปัญหาคือประเทศไทยมีแนวโน้มเด็กที่มีความต้องการพิเศษลักษณะมีความบกพร่องทางการเรียนรู้ (Learning Disabilities : LD) เป็นจำนวนมากส่วนใหญ่จะมีช่วงอายุ 9-12 ปี ซึ่งจะศึกษาอยู่ระดับชั้นประถมศึกษาปีที่ 4-6 ลักษณะของเด็กกลุ่มนี้จะมีสติปัญญาและไอคิวปกติหรืออาจสูงกว่าปกติ แต่อาจมีความบกพร่องในเรื่องบางอย่างโดยเฉพาะเรื่อง โดยเฉพาะเรื่องการอ่านหนังสือ การเขียนหนังสือ และการคิดคำนวณ ซึ่งเราสามารถพัฒนาและช่วยเหลือเขาได้ถ้าเราทราบแน่ชัดว่านักเรียนคนนั้นมีความบกพร่องทาง การเรียนรู้หรือไม่ ที่ผ่านมาการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จะทำได้อย่างลำบาก ครูผู้สอนต้องรู้ข้อมูลผู้เรียนโดยใช้เวลาไม่น้อยกว่า 1 ภาคการศึกษา หรือ 4 เดือน เพื่อทำแบบสังเกตพฤติกรรมของกระทรวงศึกษาธิการ ซึ่งต้องกรอกข้อมูลจำนวนมากทำให้เสียเวลาทั้งการกรอกและการวิเคราะห์ผลเป็นรายบุคลซึ่งเด็ก 1 คน ต้องใช้ครูกรอกมากกว่า 2 คน ในการคัดกรองเพื่อให้เกิดความถูกต้อง แม่นยำ จากปัญหาดังกล่าวทำให้

เสียเวลามาก ไม่มีความสะดวก มีขั้นตอนซับซ้อน ส่งผลต่อการจัดแผนการเรียนที่ไม่ทันเวลาและไม่เหมาะสมกับคุณลักษณะเฉพาะของผู้เรียนที่มีความบกพร่องทางการเรียนรู้

3.2 Data Understanding

ศึกษาเอกสาร ตำรา งานวิจัยที่เกี่ยวข้องกับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ พบว่ามีหลายปัจจัยสามารถแยกเก็บข้อมูลออก 3 ส่วน คือ

ส่วนที่ 1 จัดสร้างแบบทดสอบคัดกรอง ด้านการอ่านจำนวน 10 ข้อ ด้านการเขียนจำนวน 10 ข้อ และด้านการคิดคำนวณจำนวน 10 ข้อ รวมทุกด้านจำนวน 30 ข้อ โดยผู้เรียนเป็นคนทำแบบทดสอบผ่าน Google Form ที่ผู้วิจัยพัฒนาขึ้น รายละเอียดดังภาพที่ 3.1-3.3

ส่วนที่ 2 ข้อมูลพื้นฐานผู้เรียน ได้แก่ เพศ จำนวน พี่น้อง ช่วงชั้น อายุ ผลการเรียนสะสมภาษาไทย ผลการเรียนสะสมคณิตศาสตร์ ผลการเรียนสะสมรวมทุกวิชา สถานะครอบครัว ผู้ปกครองปัจจุบัน ข้อมูลบิดา–มารดา (อายุ อาชีพ รายได้ ภูมิลำเนา) โดยผู้ที่ใกล้ชิดกับเด็กมากที่สุด (ส่วนใหญ่จะเป็นครูประจำชั้นและครูประจำวิชาที่เคยสอนเด็กมาแล้ว) จะทำหน้าที่เป็นผู้ประเมินบันทึกผ่าน Google Form รายละเอียดดังภาพที่ 3.4

ส่วนที่ 3 แบบคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ของกระทรวงศึกษาธิการ โดยต้องเป็นครูผู้สอนที่เคยสอนเด็กมาแล้ว และครูประจำชั้น ซึ่งจะทำหน้าที่เป็นผู้กรอกแบบคัดกรอง จะแบ่งเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 ต้องตอบว่า ใช่ ทั้ง 3 ข้อ จึงจะทำต่อส่วนที่ 2 ต่อไป และส่วนที่ 2 จะแบ่งความบกพร่องออกเป็น 3 ด้าน คือ 1) ความบกพร่องทางด้านการอ่านให้ประเมิน ใช่,ไม่ใช่ มีจำนวนน 10 ข้อ 2) ความบกพร่องทางด้านการเขียนให้ประเมิน ใช่,ไม่ใช่ มีจำนวนน 10 ข้อ และ 3) ความบกพร่องทางด้านการคำนวณให้ประเมิน ใช่,ไม่ใช่ มีจำนวนน 10 ข้อ และ4) ด้านพฤติกรรมทั่วไปให้ประเมิน ใช่,ไม่ใช่ มีจำนวน 10 ข้อ รวมทุกด้าน จำนวน 40 ข้อ [15] ทั้งหมดให้ประเมินบันทึกผ่าน Google Form รายละเอียด ดังภาพที่ 3.5

การสร้างและหาคุณภาพแบบคัดกรองเด็กบกพร่องทางการเรียนรู้ มีรายละเอียดดังนี้

1. ศึกษาแนวคิด หลักการของแบบคัดกรองเด็กบกพร่องทางการเรียนรู้เพื่อนำไปกำหนดเกณฑ์การคัดกรองให้ครอบคลุมลักษณะของเด็กบกพร่องทางการเรียนรู้ที่ผู้วิจัยพัฒนาขึ้น

2. วิเคราะห์ปัจจัยที่คาดว่าจะส่งผลต่อการบกพร่องทางการเรียนรู้ เพื่อนำมากำหนดขอบข่าย และเขียนข้อคำถามในการสร้างแบบคัดกรองให้ครอบคลุมลักษณะของเด็กที่มีความบกพร่องทางการเรียนรู้

3. ร่างแบบคัดกรองเสนออาจารย์ที่ปรึกษา เพื่อตรวจสอบเนื้อหาและการใช้ภาษา แล้วนำข้อเสนอแนะจากอาจารย์ที่ปรึกษามาปรับปรุงแก้ไขข้อคำถามให้ถูกต้องชัดเจน

4. นำแบบคัดกรองที่ปรับปรุงแก้ไข แล้วให้ผู้เชี่ยวชาญจำนวน 3 คน ตรวจสอบความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหา (Index of item objective congruence : IOC) โดยกำหนดเกณฑ์ในการพิจารณาค่า IOC ดังนี้

4.1 ข้อคำถามที่มีค่า IOC ตั้งแต่ 0.6-1.00 คัดเลือกไว้ใช้ได้

4.2 ข้อคำถามที่มีค่า IOC ต่ำกว่า 0.6 ควรพิจารณาปรับปรุงหรือตัดทิ้ง

5. การตรวจสอบความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหา (Index of item objective congruence : IOC) ใช้เกณฑ์ในการตรวจพิจารณาข้อคำถามดังนี้

กำหนดคะแนนเป็น	+1	มีความเห็นว่า	สอดคล้อง
กำหนดคะแนนเป็น	0	มีความเห็นว่า	ไม่แน่ใจ
กำหนดคะแนนเป็น	–1	มีความเห็นว่า	ไม่สอดคล้อง

จากนั้นนำมาแทนค่าในสูตรหาดัชนีความสอดคล้อง

$IOC = \sum R/N$

เมื่อ IOC หมายถึง ดัชนีความสอดคล้อง

$\sum R$ หมายถึง ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญ

N หมายถึง จำนวนผู้เชี่ยวชาญ

เกณฑ์การคัดเลือกผู้เชี่ยวชาญ มีดังนี้

1. มีคุณวุฒิไม่ต่ำกว่าปริญญาโท สาขาเทคโนโลยีการศึกษา/การศึกษาพิเศษ

2. มีความรู้และประสบการณ์เกี่ยวกับการการสอนภาษาไทย/คณิตศาสตร์/การสอนเด็กบกพร่องทางการเรียนรู้ไม่ต่ำกว่า 2 ปี

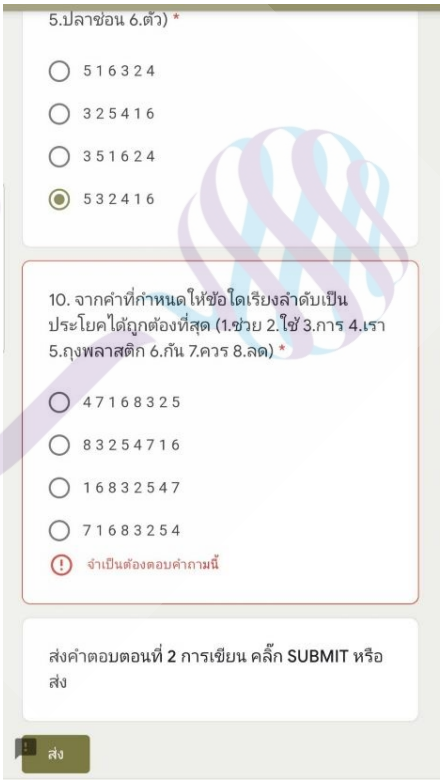
รายชื่อผู้เชี่ยวชาญตรวจแบบทดสอบคัดกรองเด็ก LD

1. รองศาสตราจารย์ ดร.ฉลอง ทับศรี	อาจารย์พิเศษภาควิชานวัตกรรมและเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
2. อาจารย์ ดร.พลธาวิน วัชรทรธำรงค์	ผู้อำนวยการโรงเรียนบ้านสวนอุดมวิทยา สำนักงานเขตพื้นที่การศึกษาประถมศึกษาชลบุรี เขต 1
3. อาจารย์ภัทราดา เอี่ยมบุญญฤทธิ์	อาจารย์กลุ่มสาระภาษาไทย ฝ่ายประถมศึกษา โรงเรียนสาธิต "พิบูลบำเพ็ญ" มหาวิทยาลัยบูรพา

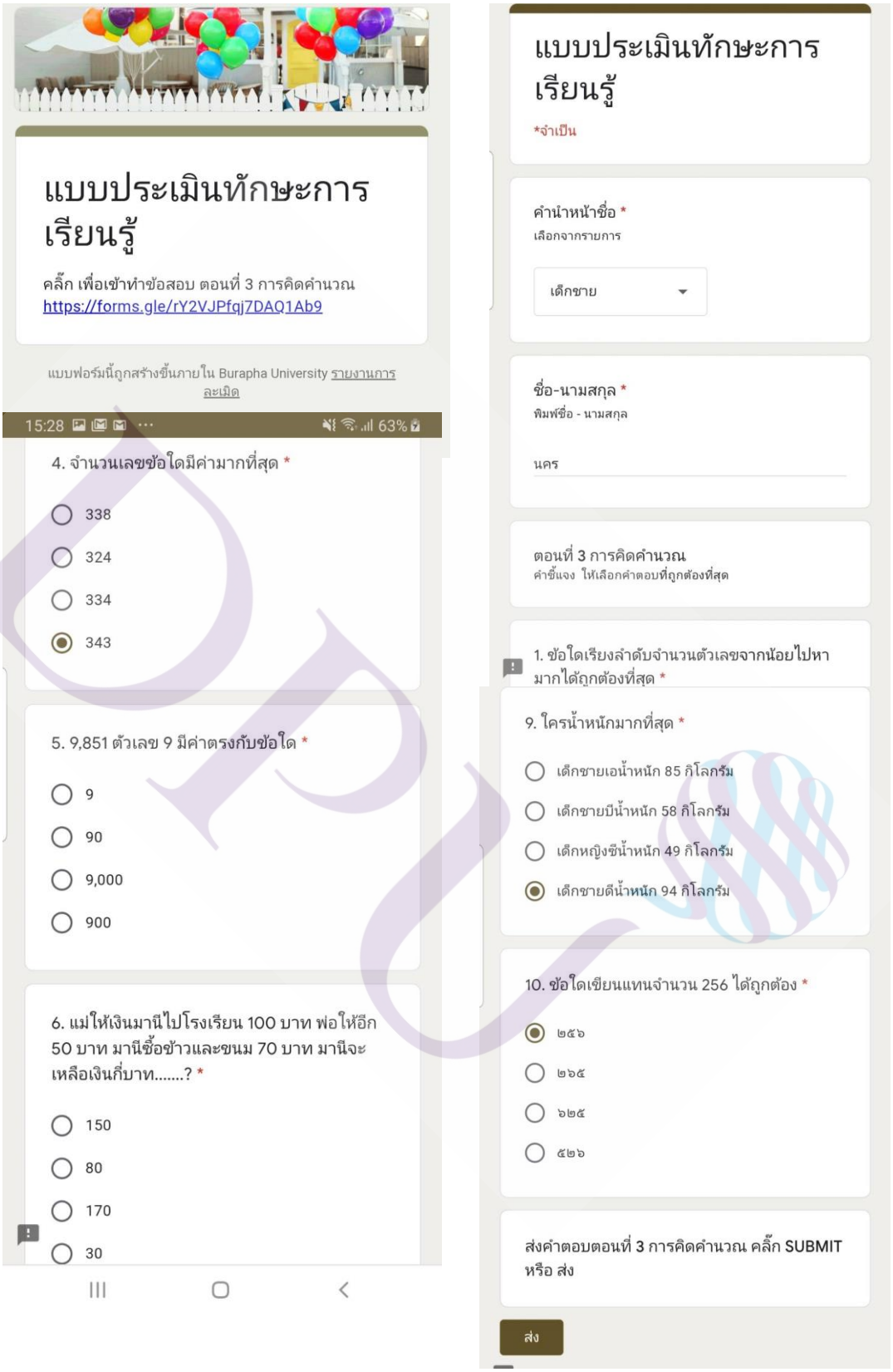
ได้ผลการหาค่าความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์หรือเนื้อหา (Index of item objective congruence : IOC) อยู่ระหว่าง .96-1.00 แสดงว่ามีความสอดคล้อง ถือว่าอยู่ในเกณฑ์ใช้ได้ไม่ต้องทำการปรับปรุง

การเก็บรวบรวมข้อมูลได้ดำเนินการช่วง กรกฎาคม - พฤษศจิกายน 2562 จากครูประจำชั้นและนักเรียนชั้นประถมศึกษาปี่ที่ 4-6 จำนวนทั้งหมด 8 โรงเรียน 250 คน มีรายละเอียดตามลิงค์แบบประเมินทักษะการเรียนรู้ คือ https://forms.gle/HyhZayrj2YSXqRjJA และลิงค์การกรอกข้อมูลพื้นฐานนักเรียน (สำหรับครู) คือ https://forms.gle/3KRgRYiXcDS77A1JA และลิงค์แบบคัดกรองของกระทรวงศึกษาธิการ คือ https://forms.gle/AGntYhz9UGwj62av8

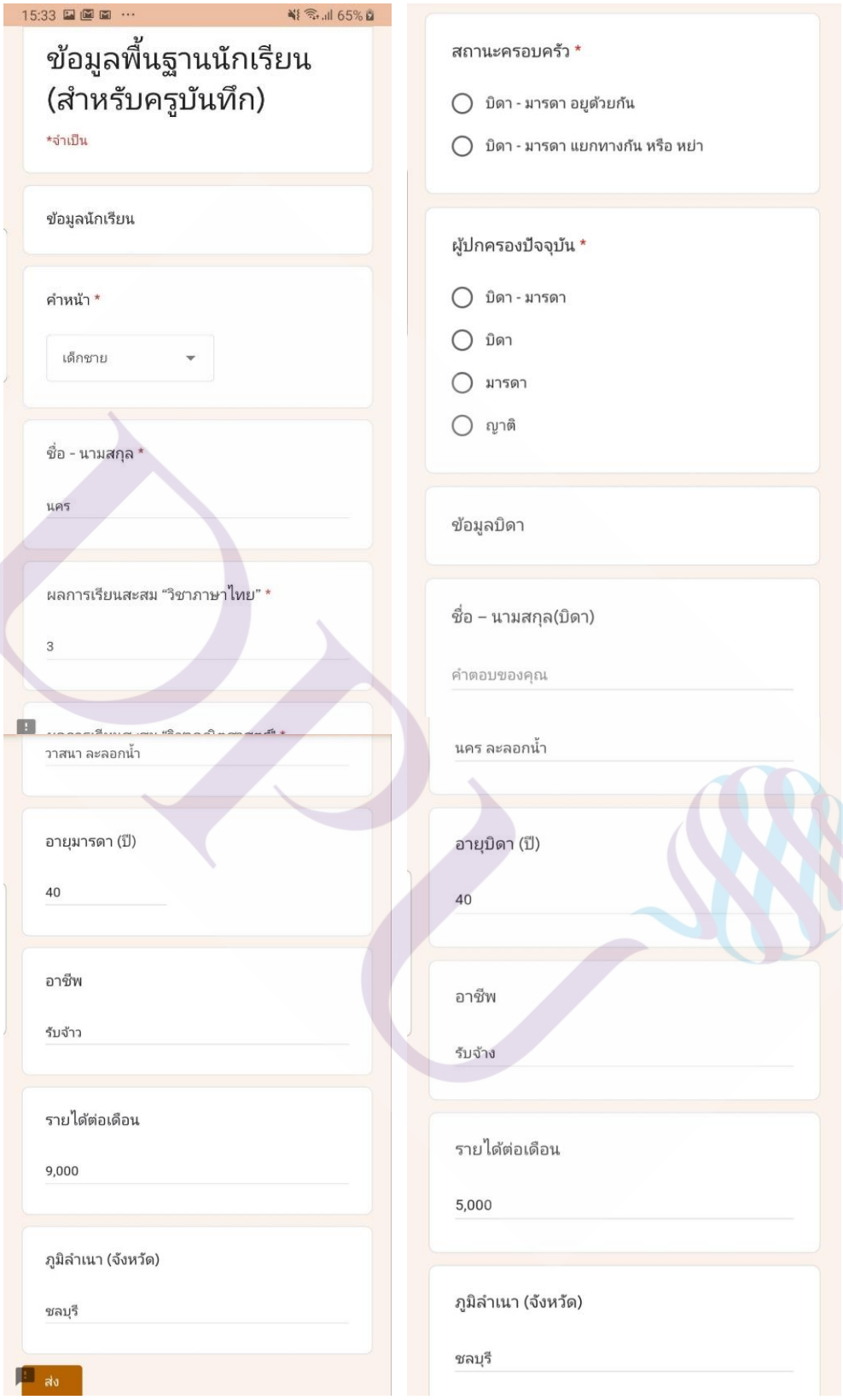
ภาพที่ 3.1 แบบทดสอบด้านการอ่าน



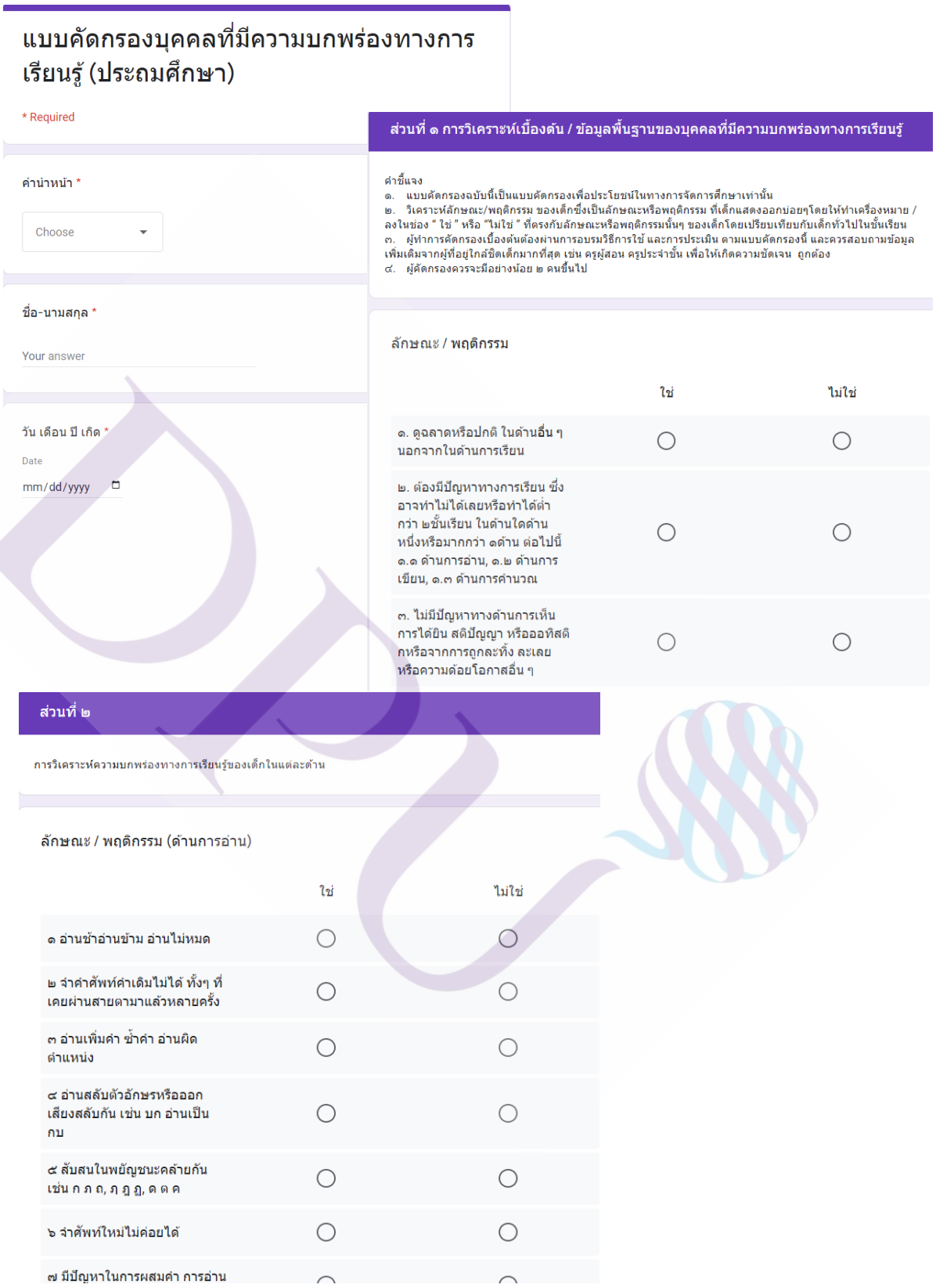
ภาพที่ 3.2 แบบทดสอบด้านการเขียน



ภาพที่ 3.3 แบบทดสอบด้านการคิดคำนวณ



ภาพที่ 3.4 การกรอกข้อมูลพื้นฐานนักเรียน



แบบคัดกรองบุคคลที่มีความบกพร่องทางการเรียนรู้ (ประถมศึกษา)

* Required

ส่วนที่ ๑ การวิเคราะห์เบื้องต้น / ข้อมูลพื้นฐานของบุคคลที่มีความบกพร่องทางการเรียนรู้

คำชี้แจง
๑. แบบคัดกรองฉบับนี้เป็นแบบคัดกรองเพื่อประโยชน์ในทางการจัดการศึกษาเท่านั้น
๒. วิเคราะห์ลักษณะ/พฤติกรรม ของเด็กซึ่งเป็นลักษณะหรือพฤติกรรม ที่เด็กแสดงออกบ่อยๆโดยให้ทำเครื่องหมาย / ลงในช่อง " ใช่ " หรือ "ไม่ใช่ " ที่ตรงกับลักษณะหรือพฤติกรรมนั้นๆ ของเด็กโดยเปรียบเทียบกับเด็กทั่วไปในชั้นเรียน
๓. ผู้ทำการคัดกรองเบื้องต้นต้องผ่านการอบรมวิธีการใช้ และการประเมิน ตามแบบคัดกรองนี้ และควรสอบถามข้อมูลเพิ่มเติมจากผู้ที่อยู่ใกล้ชิดเด็กมากที่สุด เช่น ครูผู้สอน ครูประจำชั้น เพื่อให้เกิดความชัดเจน ถูกต้อง
๔. ผู้คัดกรองควรจะมีอย่างน้อย ๒ คนขึ้นไป

คำนำหน้า *

Choose

ชื่อ-นามสกุล *

Your answer

วัน เดือน ปี เกิด *

Date

mm/dd/yyyy

ลักษณะ / พฤติกรรม

	ใช่	ไม่ใช่
๑. ดูฉลาดหรือปกติ ในด้านอื่น ๆ นอกจากในด้านการเรียน	○	○
๒. ต้องมีปัญหาทางการเรียน ซึ่งอาจทำไม่ได้เลยหรือทำได้ต่ำกว่า ๒ชั้นเรียน ในด้านใดด้านหนึ่งหรือมากกว่า ๑ด้าน ต่อไปนี้ ๑.๑ ด้านการอ่าน, ๑.๒ ด้านการเขียน, ๑.๓ ด้านการคำนวณ	○	○
๓. ไม่มีปัญหาทางด้านการเห็น การได้ยิน สติปัญญา หรือออทิสติกหรือจากการถูกละทิ้ง ละเลย หรือความด้อยโอกาสอื่น ๆ	○	○

ส่วนที่ ๒

การวิเคราะห์ความบกพร่องทางการเรียนรู้ของเด็กในแต่ละด้าน

ลักษณะ / พฤติกรรม (ด้านการอ่าน)

	ใช่	ไม่ใช่
๑ อ่านช้าอ่านข้าม อ่านไม่หมด	○	○
๒ จำคำศัพท์คำเดิมไม่ได้ ทั้งๆ ที่เคยผ่านสายตามาแล้วหลายครั้ง	○	○
๓ อ่านเพิ่มคำ ซ้ำคำ อ่านผิดตำแหน่ง	○	○
๔ อ่านสลับตัวอักษรหรือออกเสียงสลับกัน เช่น บก อ่านเป็น กบ	○	○
๕ สับสนในพยัญชนะคล้ายกัน เช่น ก ภ ถ, ภ ฎ ฏ, ด ต ค	○	○
๖ จำศัพท์ใหม่ไม่ค่อยได้	○	○
๗ มีปัญหาในการผสมคำ การอ่าน	○	○

ภาพที่ 3.5 แบบคัดกรองของกระทรวงศึกษาธิการ

3.3 Data Preparation

ทำการคัดเลือกข้อมูลให้มีความพร้อมสำหรับการวิเคราะห์โดยนำข้อมูลจากการทำแบบทดสอบของนักเรียนและข้อมูลพื้นฐานของผู้เรียนมีจำนวน 20 Attribute ได้แก่ เพศ (SEX) จำนวนพี่น้อง (BRETHREN) ช่วงชั้น (GRADE) อายุนักเรียน (AGE STUDENT) ผลการเรียนสะสมภาษาไทย(GPA THAI) ผลการเรียนสะสมคณิตศาสตร์ (GPA MATH) ผลการเรียนสะสมรวมทุกวิชา (GPA) สถานะครอบครัว (STATUS) ผู้ปกครองปัจจุบัน (SPONSOR) อายุบิดา (AGE FA) อาชีพบิดา (CAREER FA) รายได้บิดา (INCOME FA) ภูมิลำเนาบิดา (REGION FA) อายุมารดา (AGE MON) อาชีพมารดา (CAREER MON) รายได้มารดา (INCOME MON) ภูมิลำเนามารดา (REGION MON) แบบทดสอบคัดกรองด้านการอ่าน (READ) แบบทดสอบคัดกรองด้านการเขียน (WRITE) และแบบทดสอบคัดกรองด้านคิดคำนวณ (CALCULATION)

ในส่วนแบบคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ของกระทรวงศึกษาธิการจะนำผลที่ได้ไปเทียบกับเกณฑ์ที่กำหนดไว้ มีรายละเอียดดังนี้

1) ด้านการอ่าน ถ้าตอบว่าใช่ 7 ข้อ ขึ้นไป แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ ด้านการอ่าน

2) ด้านการเขียน ถ้าตอบว่าใช่ 7 ข้อ ขึ้นไป แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ด้านการเขียน

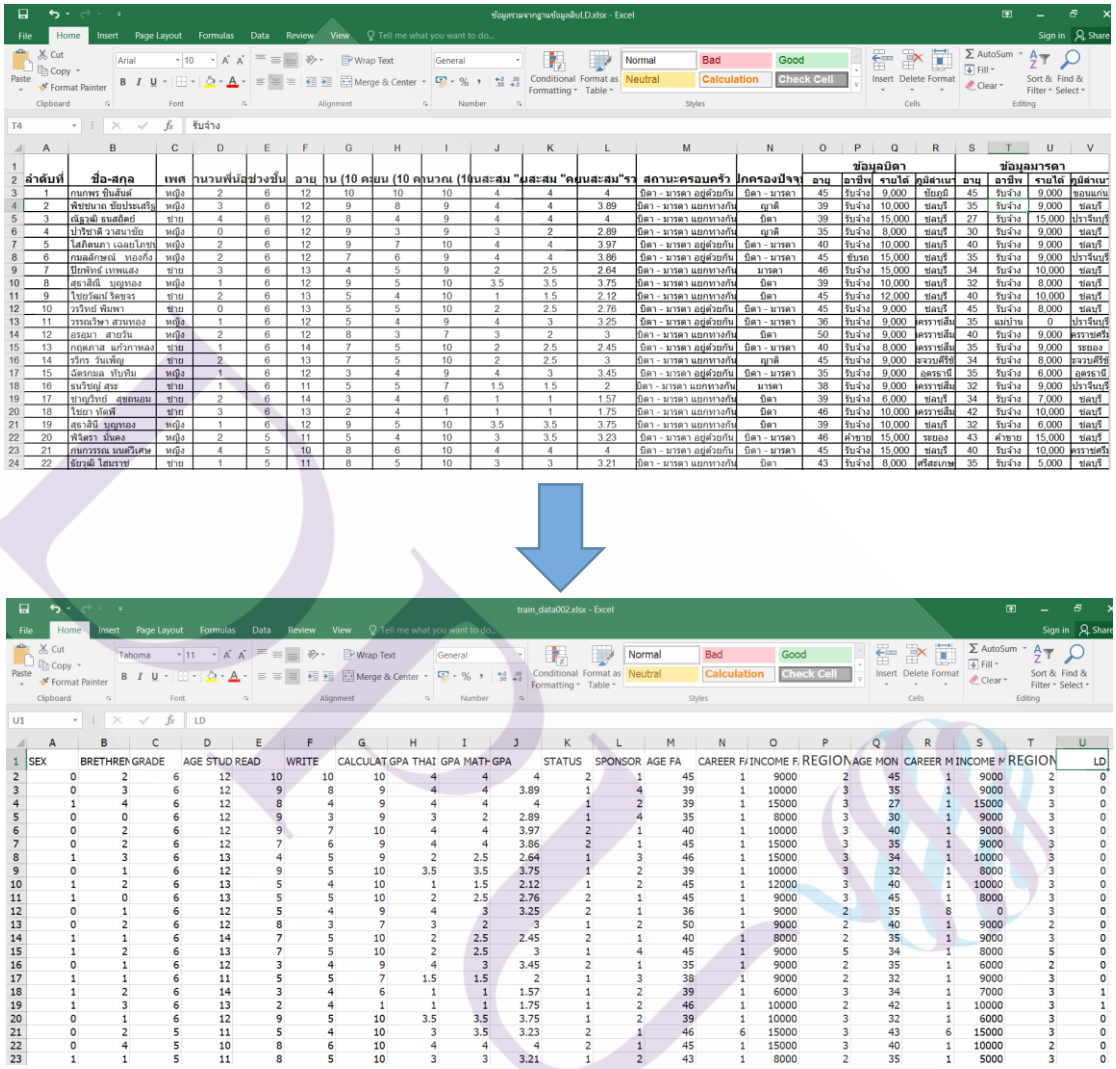
3) ด้านการคำนวณ ถ้าตอบว่าใช่ 6 ข้อ ขึ้นไป แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ด้านการคำนวณ

4) ด้านพฤติกรรมทั่วไป ถ้าตอบว่าใช่ 4 ข้อขึ้นไป แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้

หากพบว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ด้านใดด้านหนึ่งหรือหลายด้านให้จัดบริการการช่วยเหลือตามความต้องการจำเป็นพิเศษ และส่งต่อให้แพทย์ตรวจวินิจฉัยต่อไป

นำข้อมูลที่คัดเลือกที่ได้จากแบบคัดกรองที่ผู้วิจัยพัฒนาขึ้นใหม่มาทำการกลั่นกรอง (Data Cleaning) ตรวจสอบความซ้ำซ้อนของข้อมูล ปรับปรุงในส่วนของค่าข้อมูลที่ขาดหาย (Missing value) และ การกำจัดข้อมูลว่าง (Null value) โดยใช้วิธีการตัดข้อมูลที่ไม่สมบูรณ์ทิ้งทั้งหมด ได้ข้อมูลที่สมบูรณ์ จำนวน 238 คน ลักษณะข้อมูลที่ได้จะมีทั้งตัวเลข และตัวอักษร จึงทำการแปลงรูปแบบของข้อมูล (Data transformation) โดยแทนค่าข้อมูลให้อยู่ในรูปแบบที่สามารถวิเคราะห์ได้ตามอัลกอริทึมDecision Tree, Naïve Bayesian และNeural Network

การแปลงรูปแบบของข้อมูลที่ได้จากการเก็บข้อมูลจากแบบคัดกรองที่ผู้วิจัยสร้างขึ้นเองมีจำนวน 20 Attribute มีรายละเอียดดังภาพที่ 3.6 และสรุปดังตารางที 3.2



ภาพที่ 3.6 การเตรียมข้อมูลแบบคัดกรองสร้างใหม่สำหรับวิเคราะห์

ตารางที่ 3.2 การเลือกแอตทริบิวต์ (Attribute) และการแปลงรูปแบบของข้อมูล (Data transformation)

ลำดับที่	แอตทริบิวต์	การแปลงรูปแบบของข้อมูล
1	เพศ (SEX)	เพศหญิง =0, เพศชาย=1
2	จำนวนพี่น้อง (BRETHREN)	0-7 คน
3	ช่วงชั้น (GRADE)	ประถมศึกษาปีที่ 4-6
4	อายุนักเรียน (AGE STUDENT)	อายุ 9-14 ปี
5	ผลการเรียนสะสมภาษาไทย(GPA THAI)	เกรด 1-4
6	ผลการเรียนสะสมคณิตศาสตร์ (GPA MATH)	เกรด 1-4
7	ผลการเรียนสะสมรวมทุกวิชา (GPA)	สะสมรวม 1.35-4.00
8	สถานะครอบครัว (STATUS)	แยกกันอยู่ =1,อยู่ด้วยกัน =2,
9	ผู้ปกครองปัจจุบัน (SPONSOR)	บิดา-มารดา=1,บิดา=2,มารดา=3,ญาติ=4
10	อายุบิดา (AGE FA)	อายุ 25-79 ปี
11	อาชีพบิดา (CAREER FA)	รับจ้าง=1,ค้าขาย=2, พนักงานบริษัท=3,ข้าราชการ=4, ธุรกิจส่วนตัว=5,เกษตรกร=6,ภิกษุ=7,
12	รายได้บิดา (INCOME FA)	รายได้บิดา 0-186,000 บาท
13	ภูมิลำเนาบิดา (REGION FA)	ภาคเหนือ=1,ภาคตะวันออกเฉียงเหนือ=2,ภาคตะวันออก=3, ภาคตะวันตก=4,ภาคกลาง=5,ภาคใต้=6,ต่างประเทศ=7,
14	อายุมารดา (AGE MON)	อายุ 24-70 ปี
15	อาชีพมารดา (CAREER MON)	รับจ้าง=1,ค้าขาย=2, พนักงานบริษัท=3,ข้าราชการ=4, ธุรกิจส่วนตัว=5,เกษตรกร=6,
16	รายได้มารดา (INCOME MON)	รายได้บิดา 0-150,000 บาท
17	ภูมิลำเนามารดา (REGION MON)	ภาคเหนือ=1,ภาคตะวันออกเฉียงเหนือ=2,ภาคตะวันออก=3, ภาคตะวันตก=4,ภาคกลาง=5,ภาคใต้=6,ต่างประเทศ=7,
18	แบบทดสอบคัดกรองด้านการอ่าน (READ)	ผลสอบคะแนนการอ่าน 1-10
19	แบบทดสอบคัดกรองด้านการเขียน (WRITE)	ผลสอบคะแนนการเขียน 0-10
20	แบบทดสอบคัดกรองด้านคิดคำนวณ (CALCULATION)	ผลสอบคะแนนการคิดคำนวณ 1-10

จากตารางที่ 3.2 แสดงรูปแบบของข้อมูลในแต่ละ Attribute พบว่า ลักษณะข้อมูลมี 3 ลักษณะ คือข้อมูที่มี 2 ค่า (Binary number) คือ 0 กับ 1 ได้แก่ เพศ (SEX) และข้อมูลที่เป็นจำนวนจริง ได้แก่ จำนวนพี่น้อง (BRETHREN) ช่วงชั้น (GRADE) อายุนักเรียน (AGE STUDENT) ผลการเรียนสะสมภาษาไทย(GPA THAI) ผลการเรียนสะสมคณิตศาสตร์ (GPA MATH)ผลการเรียนสะสม

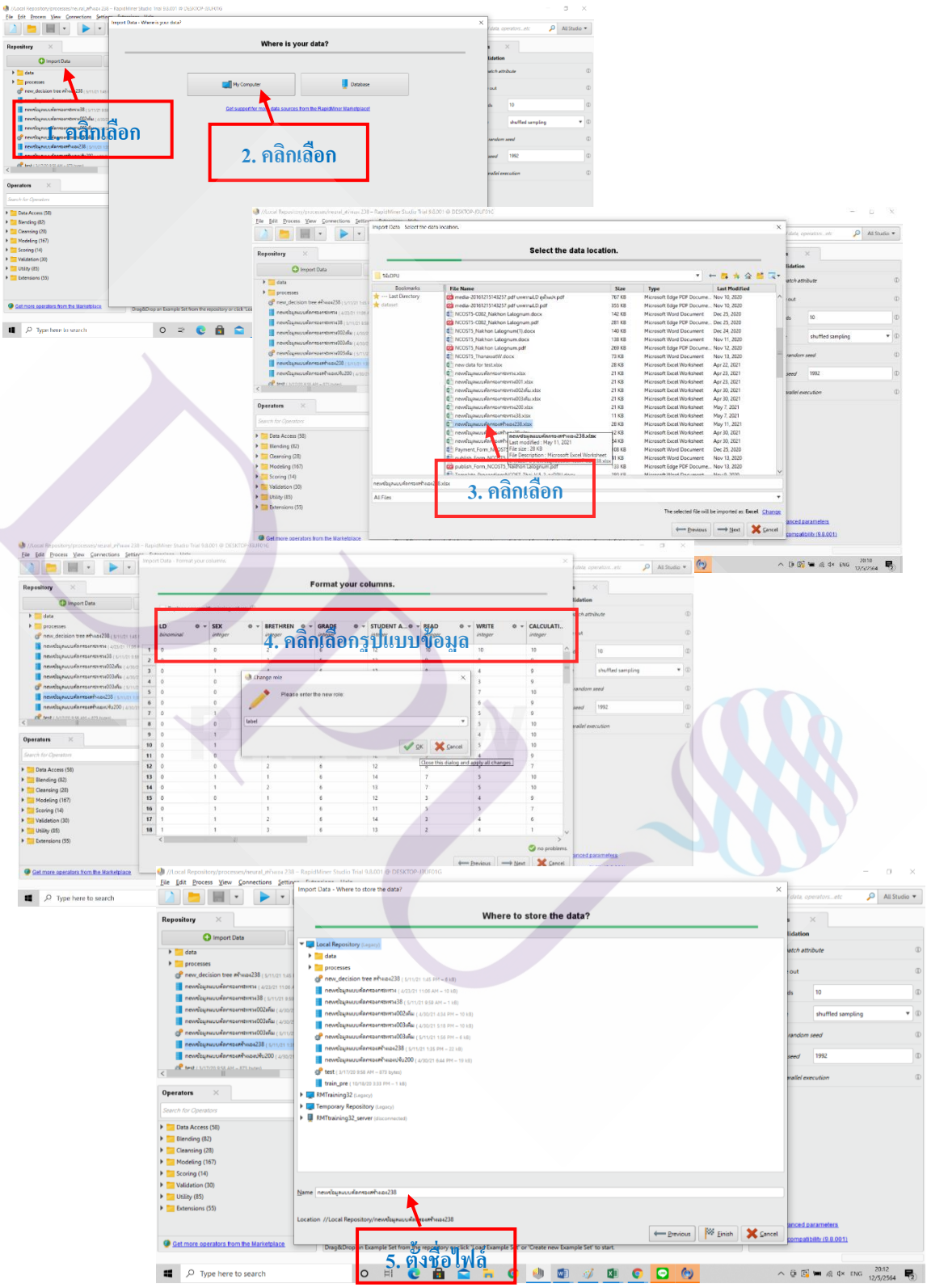
รวมทุกวิชา (GPA) อายุบิดา (AGE FA) รายได้บิดา (INCOME FA)อายุมารดา (AGE MON)รายได้มารดา (INCOME MON) แบบทดสอบคัดกรองด้านการอ่าน (READ) แบบทดสอบคัดกรองด้านการเขียน (WRITE) แบบทดสอบคัดกรองด้านคิดคำนวณ (CALCULATION) และข้อมูลที่ต้องแปลงเป็นรหัส ได้แก่ สถานะครอบครัว (STATUS)ผู้ปกครองปัจจุบัน (SPONSOR) อาชีพบิดา (CAREER FA) ภูมิลำเนาบิดา (REGION FA)อาชีพมารดา (CAREER MON) ภูมิลำเนามารดา (REGION MON) จากการกำหนดรูปแบบข้อมูลดังกล่าวจึงมีความเหมาะสมต่อการนำเข้าโมเดลเพื่อทำนายผลการเรียนด้วยวิธี Decision Tree, Naive Bayes, และ Neural network

3.4 Modeling

นำข้อมูลเข้าจากแบบคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ที่ผู้วิจัยพัฒนาขึ้นใหม่มาวิเคราะห์ เพื่อสร้างแบบจำลองเพื่อหากฎความสัมพันธ์ของข้อมูลต่างๆ ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree), นาอีฟเบย์ (Naïve Bayes), และโครงข่ายประสาท (Neural Network) มีรายละเอียด ดังนี้

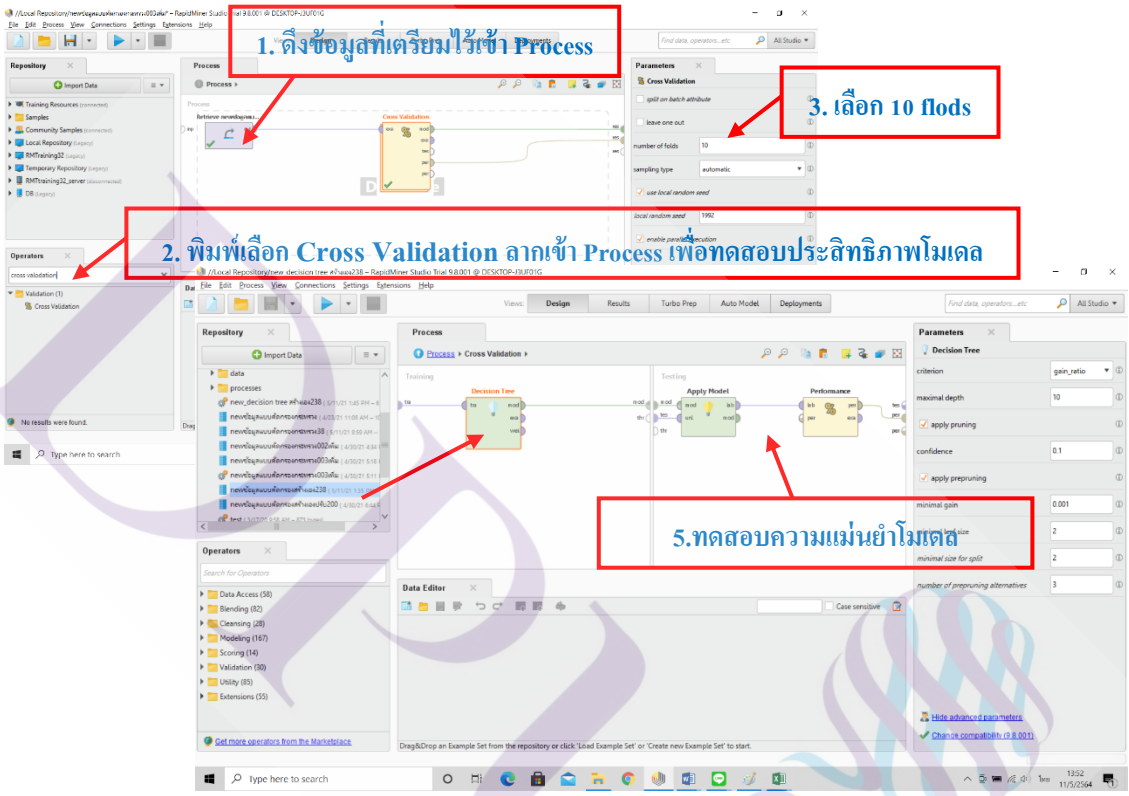
ผลการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่ดำเนินการโดยใช้โปรแกรม RapidMiner Studio 9.8 ดำเนินการดังนี้

1) เปิดโปรแกรม RapidMiner Studio 9.8 คลิกเลือก Import Data คลิกเลือก My computer เลือกไฟล์ข้อมูลที่จะนำมาวิเคราะห์ กำหนดรูปแบบข้อมูลให้ตรงกับอัลกอริทึมที่จะวิเคราะห์ ตั้งชื่อไฟล์หลังกำหนดค่าเรียนบร้อยแล้ว ขั้นตอนตามภาพที่ 3.7



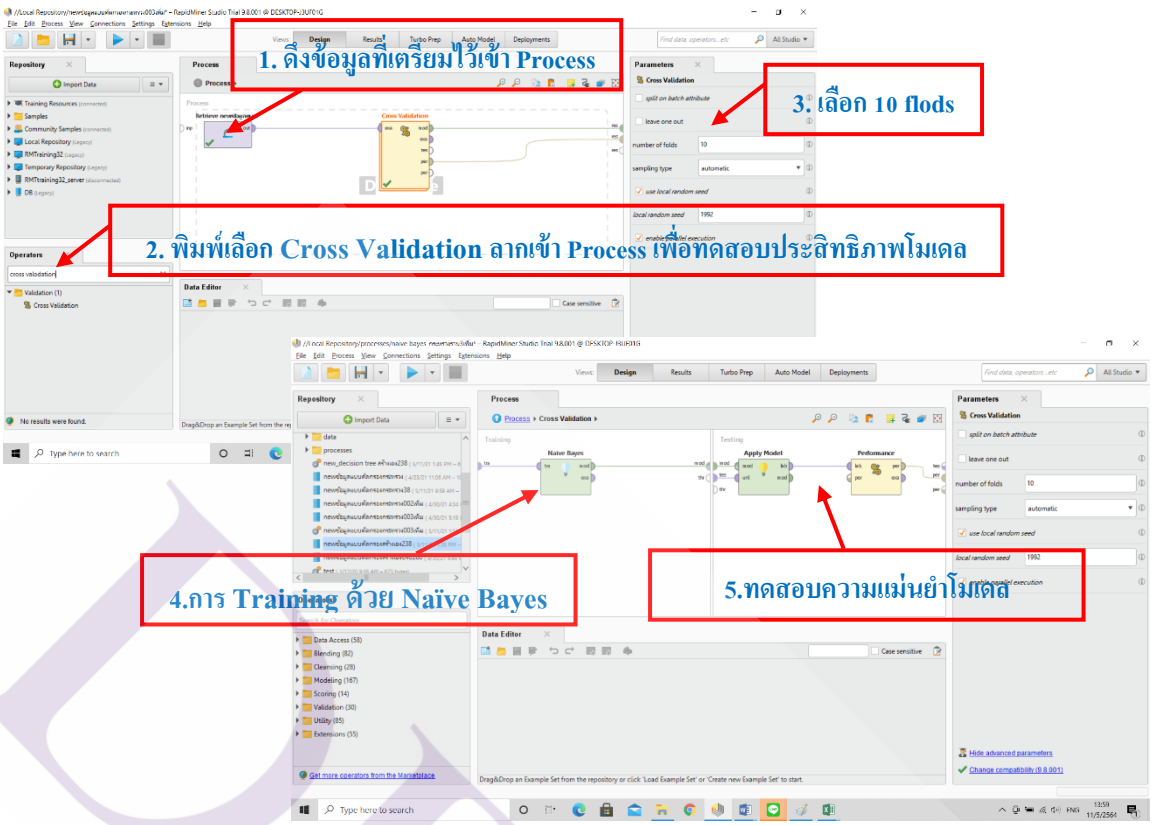
ภาพที่ 3.7 ขั้นตอนการกำหนดข้อมูลก่อนนำไปวิเคราะห์เพื่อสร้างโมเดล

2) สร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Decision Tree โดยการดึงข้อมูลเข้า Process ไปที่ Operator เลือก Cross Validation กำหนดแบบ 10 folds จากนั้นสร้างโมเดลกำหนดกลุ่ม Training กับ Testing และเลือก Operator Decision Tree และเลือก Apply Model และ Perfomance กำหนดค่ารายละเอียดต่างเพื่อให้แสดงผลตามต้องการขั้นตอนตามภาพที่ 3.8



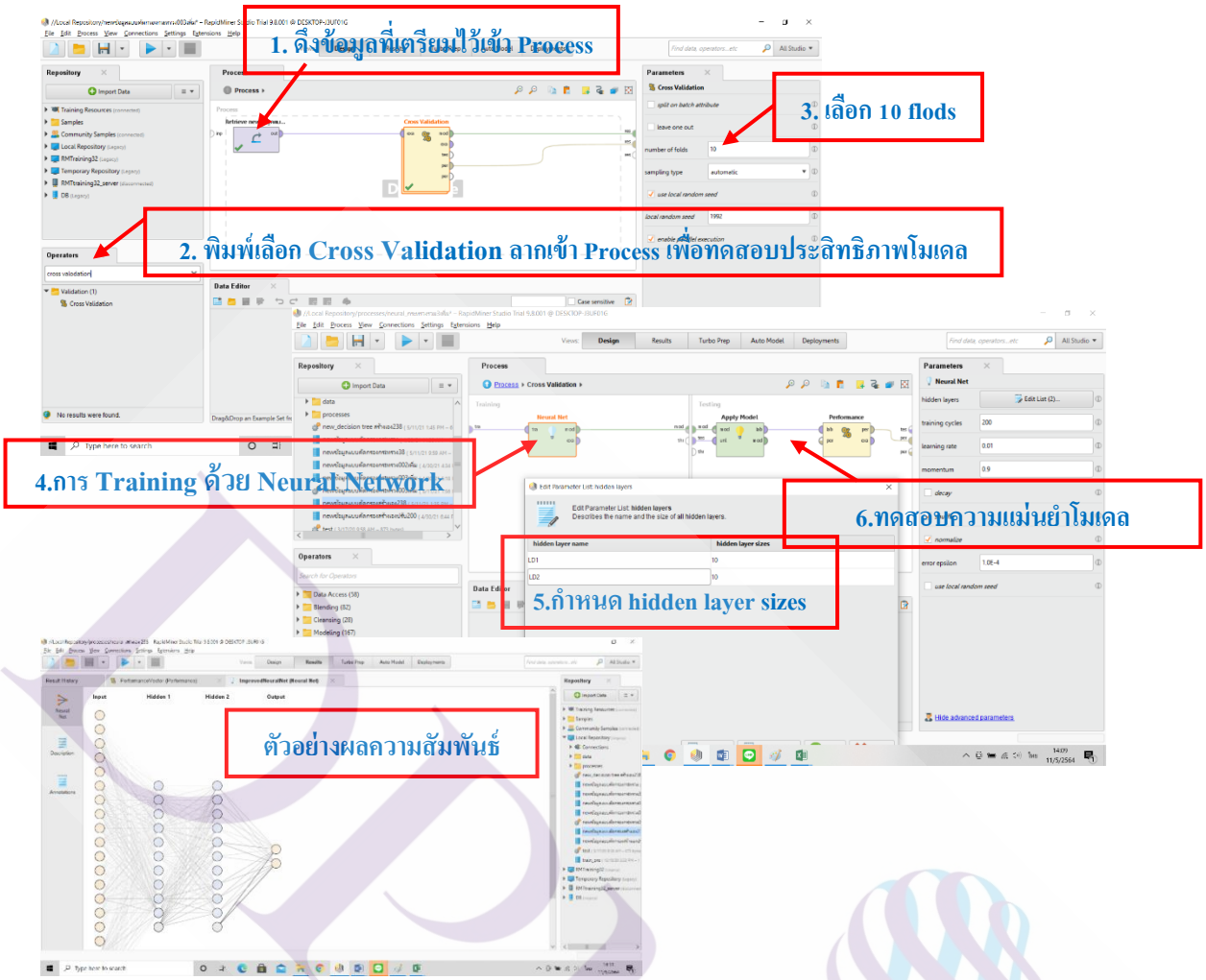
ภาพที่ 3.8 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Decision Tree

3) สร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Naïve Bayes โดยการดึงข้อมูลเข้า Process ไปที่ Operator เลือก Cross Validation กำหนดแบบ 10 folds จากนั้นสร้างโมเดลกำหนดกลุ่ม Training กับ Testing และเลือก Operator Naïve Bayes และเลือก Apply Model และ Perfomance กำหนดค่ารายละเอียดต่างเพื่อให้แสดงผลตามต้องการขั้นตอนตามภาพที่ 3.9



ภาพที่ 3.9 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Naïve Bayes

4) สร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Neural Network โดยการดึงข้อมูลเข้า Process ไปที่ Operator เลือก Cross Validation กำหนดแบบ 10 folds จากนั้นสร้างโมเดลกำหนดกลุ่ม Training กับ Testing และเลือก Operator Neural Network และเลือก Apply Model และ Perfomance กำหนดค่ารายละเอียดต่างเพื่อให้แสดงผลตามต้องการขั้นตอนตามภาพที่ 3.10



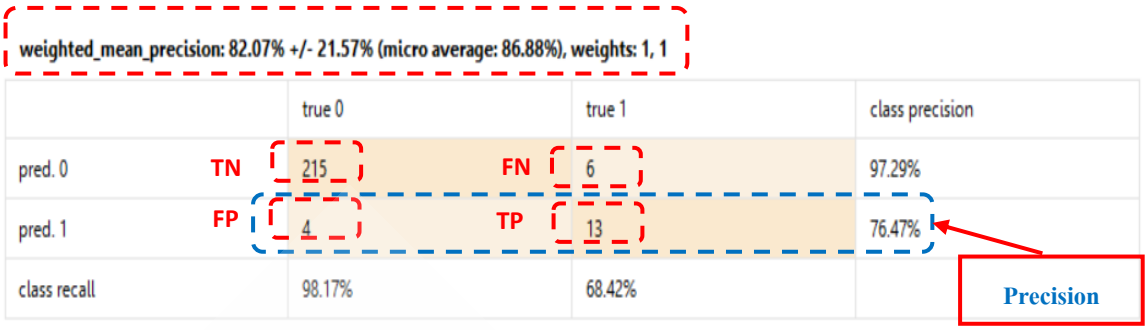
ภาพที่ 3.10 ขั้นตอนการสร้างโมเดลและทดสอบประสิทธิภาพด้วยอัลกอริทึม Neural Network

3.5 Evaluation

วัดประสิทธิภาพของผลลัพธ์โดยแบ่งกลุ่มข้อมูลเพื่อทดสอบ (Cross-validation) ประกอบด้วยข้อมูลส่วน Training set และ Testing set โดยใช้วิธี 10-fold cross-validation

โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กันแล้วทำการทดสอบโดยเปลี่ยนข้อมูลชุดทดสอบตั้งแต่ส่วนที่ 1 เป็นชุดทดสอบ และส่วนที่ 2-9 เป็นชุดเทรนนิ่งไปเรื่อยๆ จนกระทั่งถึงชุดทดสอบที่ส่วนที่ 10 เป็นชุดทดสอบ และส่วนที่ 1-9 เป็นชุดเทรนนิ่ง เพื่อให้ได้ค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy) มีรายละเอียดดังนี้

3.5.1 ตัวอย่างผลประสิทธิภาพอัลกอริทึม Decision Tree จากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่ แสดงผลประสิทธิภาพของค่า Precision มีรายละเอียดดังภาพที่ 3.11



ภาพที่ 3.11 เป็นประสิทธิภาพของค่า Precision

จากภาพที่ 3.11 เป็นค่า Precision (pred.1) สามารถแปลผลประสิทธิภาพได้ว่า เราส่งเด็กที่มีความบกพร่องทางการเรียนรู้ (pred.1) เข้าไปทำนาย 13 แถว แต่ตอบผิดว่าไม่มีความบกพร่องทางการเรียนรู้ (pred.0)ไป 4 แถวจากจำนวน 17 แถว และ ค่า Precision (pred.0) สรุปได้ว่าเราส่งเด็กที่ไม่มีความบกพร่องทางการเรียนรู้ (pred.0) เข้าไปทำนาย 215 แถว แต่ตอบผิดว่ามีความบกพร่องทางการเรียนรู้ (pred.1) ไป 6 แถว จากจำนวน 221 แถว โดยสามารถคำนวณได้ดังนี้

สิ่งที่สนใจทำนายว่ามีความบกพร่องทางการเรียนรู้ (pred.1) Precision = TP/(TP+FP)

แทนค่า Precision = 13 / (13+4)

= 13 / 17

= 0.7647 หรือ 76.47 %

สิ่งไม่สนใจทำนายว่าไม่มีความบกพร่องทางการเรียนรู้ (pred.0) Precision = TN/(TN+FN

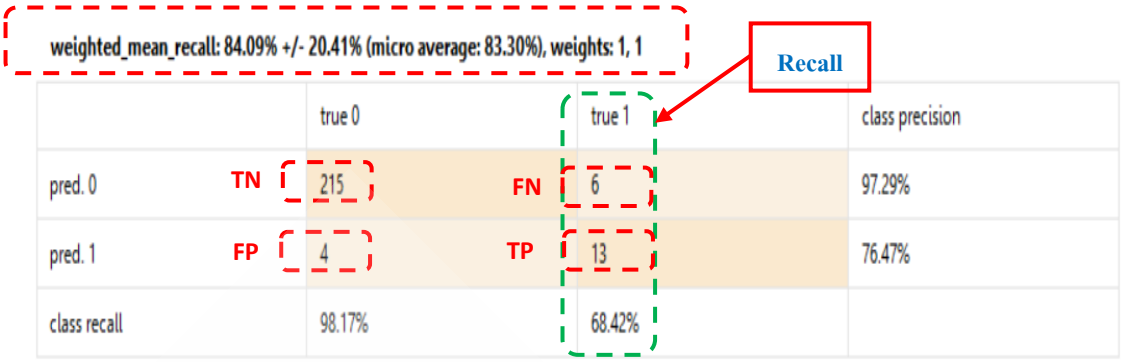
แทนค่า Precision = 215 / (215+6)

= 215 / 221

= 0.9728 หรือ 97.28 %

และประสิทธิภาพของค่า Mean Precision เท่ากับ 82.07 %

แสดงผลประสิทธิภาพของค่า recall มีรายละเอียดดังภาพที่ 3.12



ภาพที่ 3.12 เป็นประสิทธิภาพของค่า recall

จากภาพที่ 3.12 เป็นค่า Recall (pred.1) สามารถแปลผลประสิทธิภาพได้ว่า เราส่งเด็กที่มีความบกพร่องทางการเรียนรู้ (pred.1) เข้า model เพื่อทำนาย 19 แถว (เป็น LD จริงๆ 19) แต่ได้คำตอบที่ตรงกับข้อมูลจริง 13 แถว (ตอบถูกว่าเป็น LD จริงๆ 13 แถว) และตอบผิดกลับมา 6 แถว (เด็กเป็น LD แต่ตอบกลับมาว่าไม่เป็น LD 6 แถว) และ ค่า Recall (pred.0) สรุปได้ว่า เราส่งเด็กที่ไม่มีความบกพร่องทางการเรียนรู้ (pred.0) เข้า model เพื่อทำนาย 219 แถว (ไม่เป็น LD 219) แต่ได้คำตอบที่ตรงกับข้อมูลจริง 215 แถว (ตอบถูกว่าไม่เป็น LD 215แถว) และตอบผิดกลับมา 4 แถว (เด็กไม่เป็น LD แต่ตอบกลับมาว่าเป็น LD 4 แถว) โดยสามารถคำนวณได้ดังนี้

สิ่งที่สนใจทำนายว่ามีความบกพร่องทางการเรียนรู้ (pred.1) Recall = TP/(TP+FN)

แทนค่า Recall = 13 / (13+6)

= 13 / 19

= 0.6842 หรือ 68.42 %

สิ่งไม่สนใจทำนายว่าไม่มีความบกพร่องทางการเรียนรู้ (pred.0) Recall = TN/(TN+FP

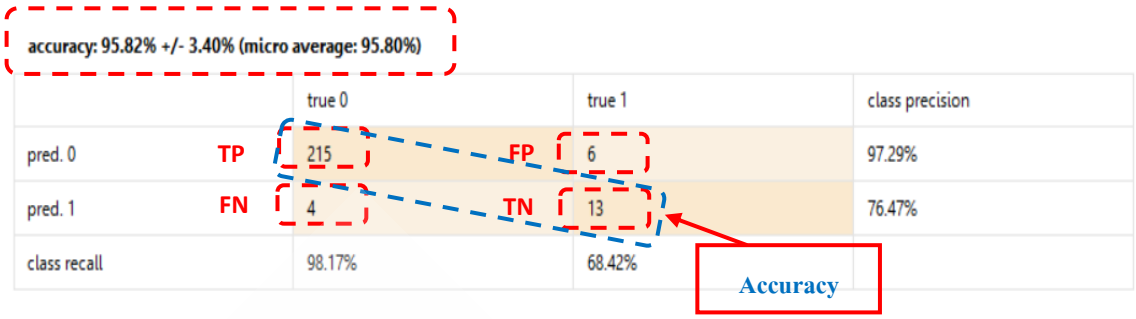
แทนค่า Recall = 215 / (215+4)

= 215 / 219

= 0.9817 หรือ 98.17 %

และประสิทธิภาพของค่า Mean Recall เท่ากับ 84.09 %

แสดงผลประสิทธิภาพของค่า Accuracy มีรายละเอียดดังภาพที่ 3.13



ภาพที่ 3.13 เป็นประสิทธิภาพของค่า Accuracy

จากภาพที่ 3.13 เป็นค่า Accuracy ที่คำนวณจากจำนวนคำตอบที่ถูกต้องเทียบกับจำนวนคำตอบทั้งหมดที่นำไปให้ Model ทำการตอบหรือทำนาย จากภาพมีการส่งจำนวนแถวให้ Model ตอบทั้งหมด 238 แถว โดยมี Class เด็กที่มีความบกพร่องทางการเรียนรู้ (pred.1) ตอบถูก 13 แถว และ Class เด็กที่ไม่มีความบกพร่องทางการเรียนรู้ (pred.0) ตอบถูก 215 แถว รวมจำนวนที่ตอบถูกทั้งหมด 228 แถว โดยสามารถคำนวณได้ดังนี้

Accuracy (ความถูกต้องรวม) = (TP + TN) / (TP+TN+FP + FN)

แทนค่า Accuracy = (215+13) / (215+13+6+4)

= 228 / 238

= 0.95798 หรือ 95.80 %

สรุปประสิทธิภาพของค่า Accuracy (ความถูกต้องรวม) เท่ากับ 95.82 %

3.6 Deployment

ครูผู้สอนนำโมเดลที่ได้ไปใช้เป็นแนวทางในการคัดคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ได้อย่างสะดวก แม่นยำเพื่อให้สามารถจัดแผนการเรียนที่เหมาะสมกับคุณลักษณะเฉพาะของผู้เรียนที่มีความบกพร่องทางการเรียนรู้ได้อย่างมีประสิทธิภาพ

รายละเอียดผลเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติผู้วิจัยนำเสนอในบทที่ 4 ต่อไป

บทที่ 4

ผลการทดลอง

การวิจัยการใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ครั้งนี้นำเสนอผลการทดลอง แยกเป็น 3 ตอนดังนี้

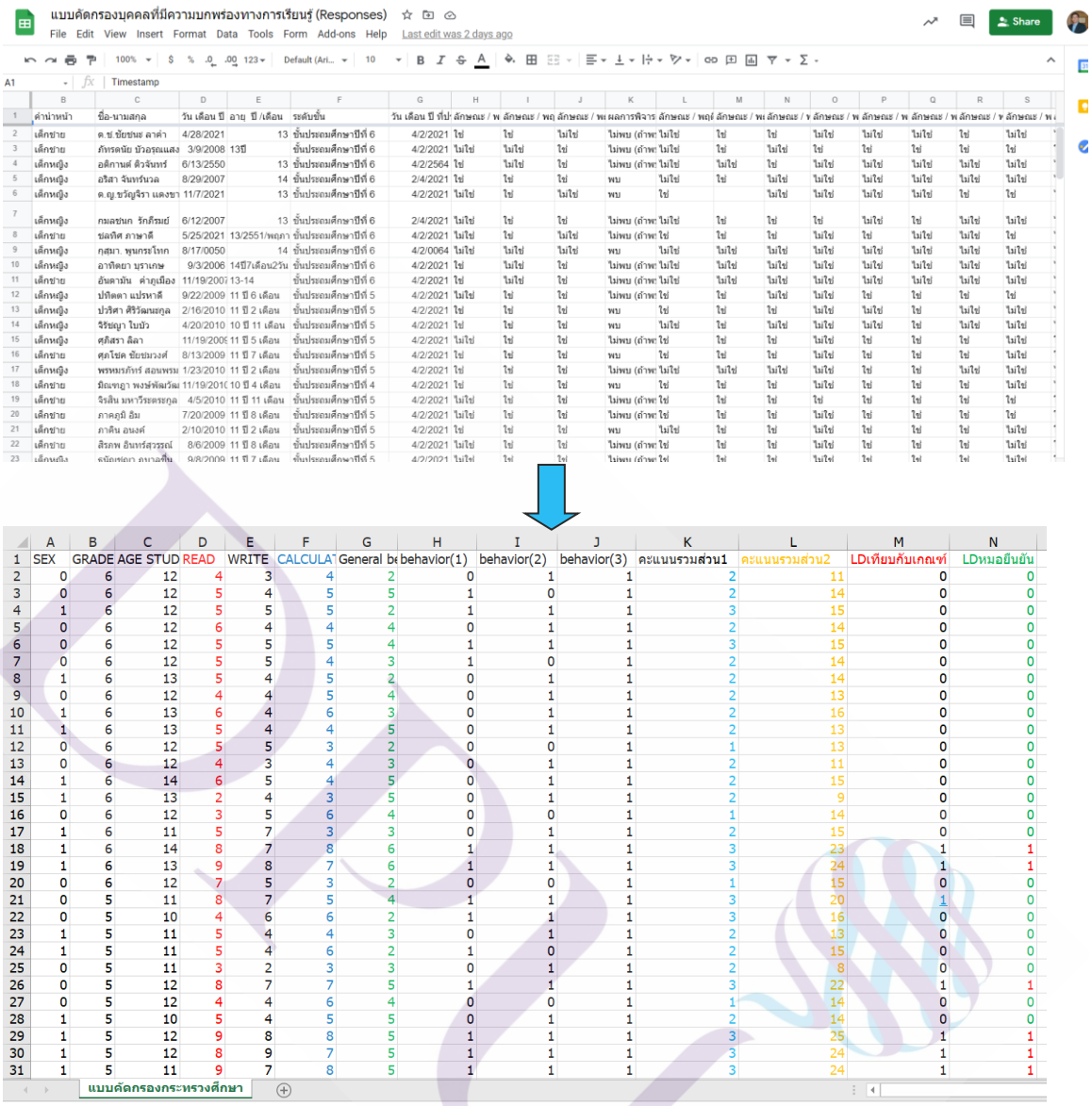
ตอนที่ 1 ผลการคัดกรองจากการใช้แบบคัดกรองของระทรวงศึกษาธิการ ดำเนินการคัดกรองแบบปกติ

ตอนที่ 2 ผลการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่

ตอนที่ 3 ผลการเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ

ตอนที่ 1 ผลการคัดกรองจากการใช้แบบคัดกรองของระทรวงศึกษาธิการ ดำเนินการคัดกรองแบบปกติ

การเก็บรวบรวมข้อมูลนักเรียนชั้นประถมศึกษาปีที่ 4-6 ทั้งหมด 8 โรงเรียน จำนวน 238 คน โดยครูประจำชั้นเป็นผู้กรอกข้อมูลผ่าน Google Form ลิงค์ ttps://forms.gle/AGnt-Yhz9UGwj62av8 นำข้อมูลที่ได้มาเปลงและรวมคะแนนเทียบกับเกณฑ์เพื่อประเมินผลการคัดกรองโดยเกณฑ์กำหนดไว้ คือ พฤติกรรมด้านการอ่าน ด้านการ ด้านการคำนวณ รวมแล้วต้องแสดงออกไม่น้อยกว่า 20 ลักษณะ และด้านพฤติกรรมทั่วไปแสดงออกไม่น้อยกว่า 4 ลักษณะ แสดงว่ามีแนวโน้มที่จะเป็นบุคคลที่มีความบกพร่องทางการเรียนรู้ดังภาพที่ 4.1



ภาพที่ 4.1 การแปลงข้อมูลเพื่อรวมคะแนนเทียบกับเกณฑ์เพื่อประเมินผลการคัดกรอง

จากภาพที่ 4.1 สรุปได้ว่า กลุ่มตัวอย่างทั้งหมด 238 คน ครูประจำชั้นและครูผู้สอนนำผลรวมคะแนนผู้เรียนไปเทียบกับเกณฑ์ที่กำหนดสรุปว่าผู้เรียนที่มีความเสี่ยงบกพร่องทางการเรียนรู้จำนวน 23 คน

จากนั้นครูประจำชั้นส่งนักเรียนทั้ง 23 คน ส่งให้แพทย์วินิจฉัยเพื่อลงความเห็นยืนยันว่าบกพร่องทางการเรียนรู้หรือไม่ พบว่า แพทย์วินิจฉัยสรุปความเห็นว่านักเรียนมีความบกพร่องทางการเรียนรู้ทั้งหมด 19 คน ที่เหลือ 4 คน ไม่มีความบกพร่องทางการเรียนรู้ สรุปได้ว่าความ

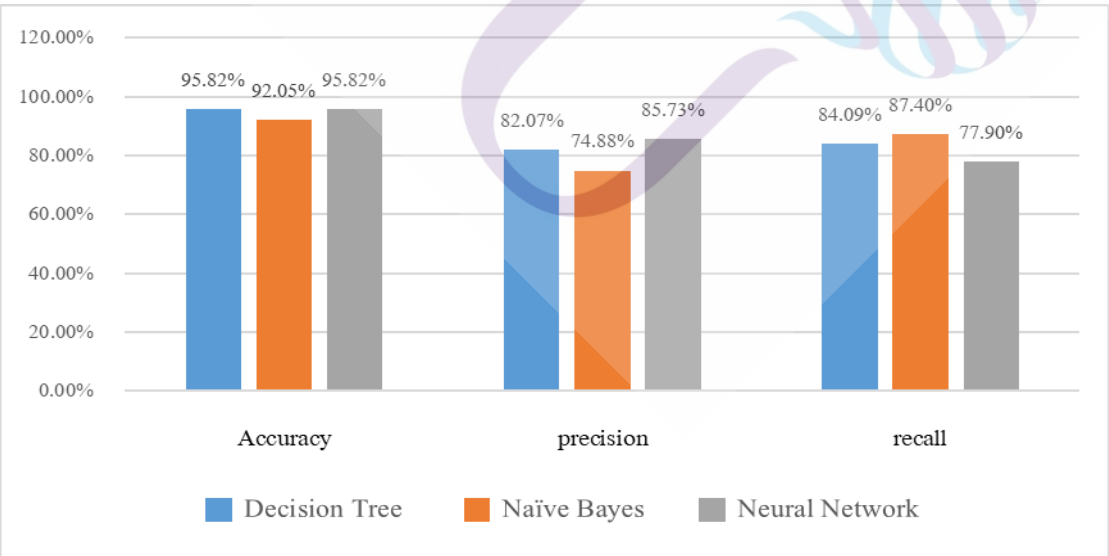
ถูกต้องจากการใช้แบบคัดกรองของระทรวงศึกษาธิการของครู เท่ากับ 82.61 % ((19/23)*100) พิจารณาความเคลื่อนจะมีค่าเท่ากับ เท่ากับ 17.39% ((4/19)*100)

ตอนที่ 2 ผลการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่

การวิจัยครั้งนี้ใช้จำนวน data set ทั้งหมด 238 คน จากนักเรียนระดับชั้นประถมศึกษาปีที่ 4-6 การนำข้อมูลเข้ามาวิเคราะห์เพื่อสร้างแบบจำลอง โดยใช้ตัววัดประสิทธิภาพโมเดลของการจำแนกข้อมูลจากการคำนวณการประเมินผลลัพธ์การทำนาย (confusion matrix) จากค่า precision, recall, accuracy สำหรับการทดลองครั้งนี้ จะแบ่งระดับคะแนนออก เป็น 2 class หรือกำหนดเป็น Label 2 ค่า (Binominal) คือ ไม่บกพร่องทางการเรียนรู้= 0 และบกพร่องทางการเรียนรู้ = 1 ดำเนินการโดยใช้โปรแกรม RapidMiner Studio 9.8 ได้ผลการทำนายโมเดลด้วยอัลกอริทึม Decision Tree, Naïve Bayes และ Neural Network ผลการทดลอง ดังตารางที่ 4.1 กับภาพที่ 4.2

ตารางที่ 4.1 เปรียบเทียบผลการทดลองแต่ละอัลกอริทึม

Algorithm	Accuracy	precision	recall
Decision Tree	95.82%	82.07%	84.09%
Naïve Bayes	92.05%	74.88%	87.40%
Neural Network	95.82%	85.73%	77.90%



ภาพที่ 4.2 เปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่

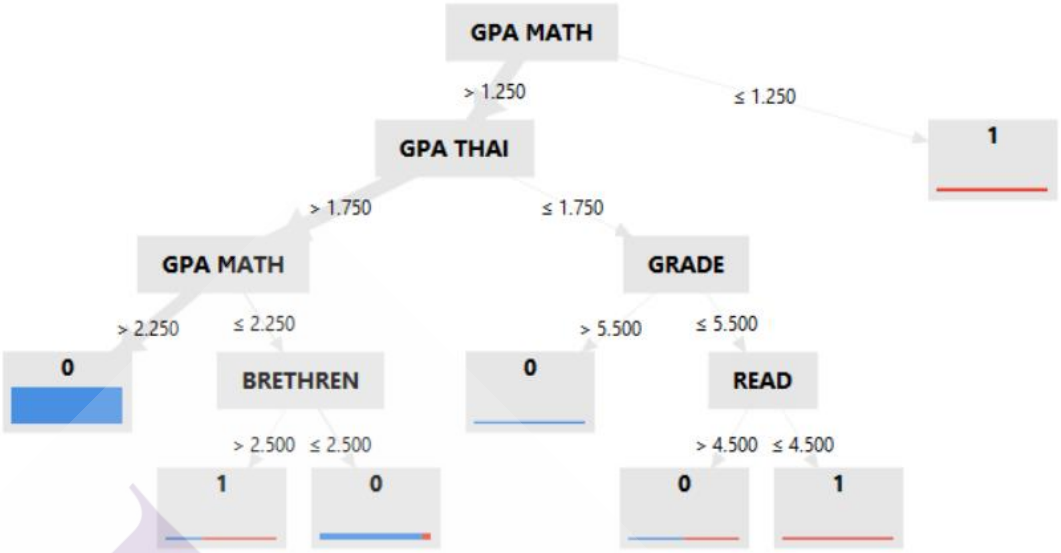
จากตารางที่ 4.1 และภาพที่ 4.2 แสดงให้เห็นว่า อัลกอรริทึม Neural Network และ Decision Tree มีค่าประสิทธิภาพสูงสุดและมีค่าเท่ากัน โดยค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy) เท่ากับ 95.82% รองลงมาคืออัลกอรริทึม Naïve Bayes เท่ากับ 92.05% ตามลำดับ

อัลกอรริทึม Neural Network มีค่าความแม่นยำของโมเดลโดยพิจารณาแยกทีละคลาส (Precision) มากที่สุด เท่ากับ 85.735% รองลงมาคืออัลกอรริทึม Decision Tree เท่ากับ 82.07% และอัลกอรริทึม Naive Bayes เท่ากับ 74.88% ตามลำดับ

อัลกอรริทึม Naive Bayes มีค่าความถูกต้องของโมเดลโดยพิจารณาแยกทีละคลาส (Recall) เท่ากับ 87.40% รองลงมาคืออัลกอรริทึม Decision Tree เท่ากับ 84.09% และ อัลกอรริทึม Neural Network เท่ากับ 77.90% ตามลำดับ

เมื่อพิจารณาผลการจำแนกข้อมูลด้วย Decision Tree โหนดเริ่มต้นหรือ Attribute ที่สัมพันธ์กับ class หรือ Label มากที่สุด คือ เกรดเฉลี่ยวิชาคณิตศาสตร์ หากอยู่ในระดับน้อยกว่า 1.25 จะอยู่ในกลุ่มมีความบกพร่องทางการเรียนรู้ (LD) แต่ถ้าเกรดเฉลี่ยวิชาคณิตศาสตร์ มากกว่า 1.25 และเกรดเฉลี่ยภาษาไทย มากกว่า 1.75 และเกรดเฉลี่ยวิชาคณิตศาสตร์ มากกว่า 2.25 จะอยู่ในกลุ่มไม่มีความบกพร่องทางการเรียนรู้ (LD) แต่ถ้าเกรดเฉลี่ยวิชาคณิตศาสตร์ น้อยกว่าหรือเท่ากับ 2.25 และจำนวนพี่น้องน้อยกว่าหรือเท่ากับ 2.5 จะอยู่ในกลุ่มไม่มีความบกพร่องทางการเรียนรู้ (LD) แต่ถ้าจำนวนพี่น้องมากกว่า 2.5 จะอยู่ในกลุ่มมีความบกพร่องทางการเรียนรู้ (LD)

กรณีเกรดเฉลี่ยวิชาคณิตศาสตร์มากกว่า 1.25 และเกรดเฉลี่ยภาษาไทยน้อยกว่าหรือเท่ากับ 1.75 และระดับชั้นสูงกว่า 5.5 จะอยู่ในกลุ่มไม่มีความบกพร่องทางการเรียนรู้ (LD) แต่ถ้าระดับชั้นต่ำกว่าหรือเท่ากับ 5.5 และผลทดสอบด้านการอ่านมากกว่า 4.5 คะแนน จะอยู่ในกลุ่มไม่มีความบกพร่องทางการเรียนรู้ (LD) แต่ถ้าผลทดสอบด้านการอ่านน้อยกว่าหรือเท่ากับ 4.5 คะแนน จะอยู่ในกลุ่มมีความบกพร่องทางการเรียนรู้ (LD) รายละเอียดดังภาพที่ 4.3

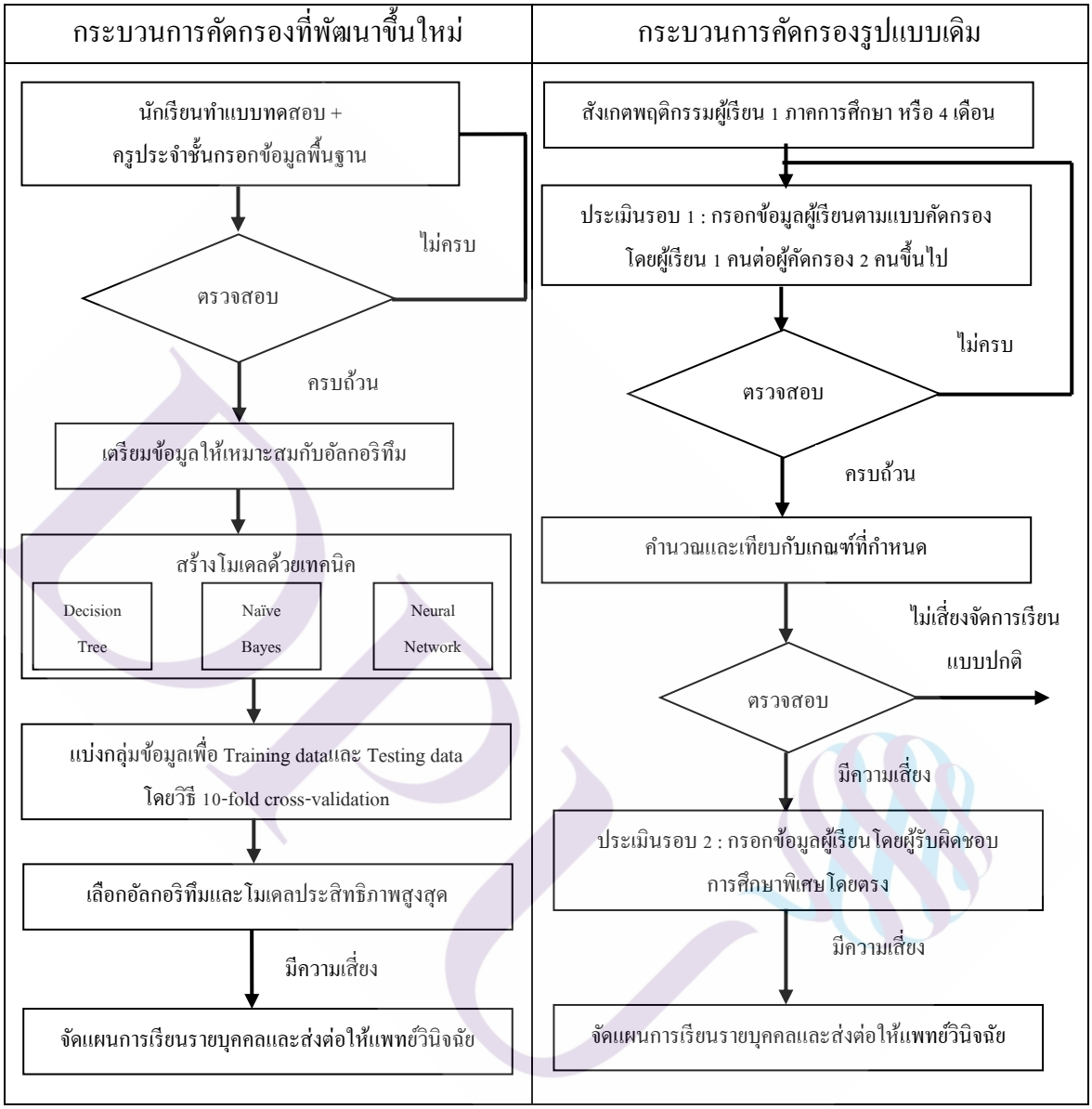


ภาพที่ 4.3 ผลกฎความสัมพันธ์การคัดกรองเด็กบกพร่องทางการเรียนรู้ด้วยเทคนิค Decision Tree โดยใช้แบบคัดกรองที่พัฒนาขึ้นใหม่

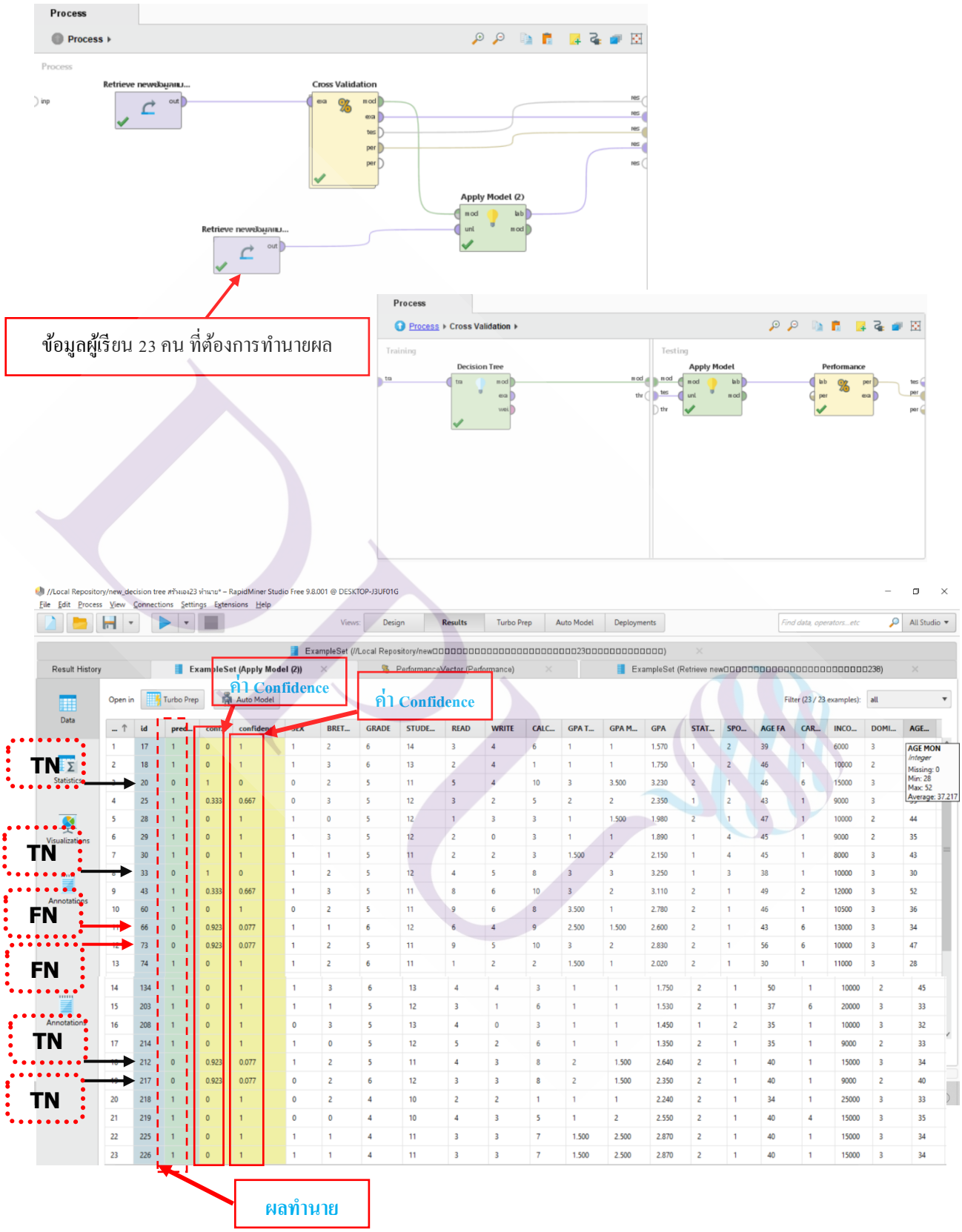
ตอนที่ 3 ผลการเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ

การทำเหมืองข้อมูลครั้งนี้ ดำเนินการโดยใช้โปรแกรม RapidMiner Studio 9.8 ได้ผลการเปรียบเทียบการทำนายโมเดลแต่ละอัลกอริทึมจากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่กับแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ ผู้วิจัยขอเสนอกระบวนการดำเนินงานทั้ง 2 วิธี ดังตารางที่ 4.2

ตารางที่ 4.2 เปรียบเทียบกระบวนการคัดกรองที่พัฒนาขึ้นใหม่กับกระบวนการคัดกรองรูปแบบเดิม



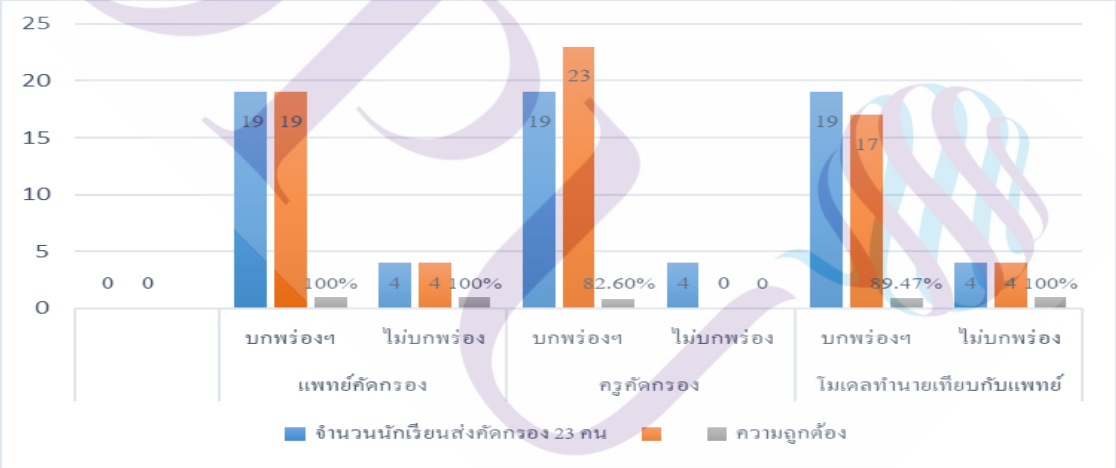
เพื่อให้เห็นภาพการเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูลกับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ จากการวิเคราะห์ด้วยวิธี Decision Tree และพิจารณาเป็นรายบุคคล 23 คนที่ส่งให้แพทย์วินิจฉัย พบว่า โมเดลสามารถทำนายเด็กที่บกพร่องทางการเรียนรู้ได้ตรงกับแพทย์วินิจฉัย 17 คน (แพทย์วินิจฉัยไว้ 19 คนจาก 23 คน) คิดเป็นความถูกต้องเท่ากับ 89.47% ((17/19)*100) และทำนายเด็กที่ไม่บกพร่องทางการเรียนรู้ได้ตรงกับแพทย์วินิจฉัย 4 (แพทย์วินิจฉัยไว้ 4 คน จาก 23 คน) คิดเป็นความถูกต้องเท่ากับ 100 ((4/4)*100) คนรายละเอียดดังภาพที่ 4.3 ภาพที่ 4.4 และตารางที่ 4.3



ภาพที่ 4.4 ผลการทำนาย และ ค่า Confidence การทำนายด้วยวิธี Decision Tree

ตารางที่ 4.3 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูลกับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ

ผู้คัดกรอง		จำนวนนักเรียนส่งคัดกรอง 23 คน		ความถูกต้อง
		ข้อมูลจริง	ผลการวิเคราะห์	
แพทย์คัดกรอง	บกพร่องฯ	19	19	100%
	ไม่บกพร่อง	4	4	100%
ครูคัดกรอง	บกพร่องฯ	19	23	82.60%
	ไม่บกพร่อง	4	0	0
โมเดลทำนายเทียบกับแพทย์	บกพร่องฯ	19	17	89.47%
	ไม่บกพร่อง	4	4	100%



ภาพที่ 4.5 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่กับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ

จากตารางที่ 4.3 และภาพที่ 4.5 สรุปได้ว่า ผลการคัดกรองจะยืนยันตามการวินิจฉัยของแพทย์ คือ จากที่ครูส่งผลการคัดกรองเด็กที่มีความเสี่ยงบกพร่องทางการเรียนรู้ 23 คน แต่แพทย์ยืนยันบกพร่องทางการเรียนรู้ทั้งหมด 19 คน และไม่บกพร่องทางการเรียนรู้ทั้งหมด 4 คน เมื่อพิจารณาผลการคัดกรองของครูระหว่างเด็กบกพร่องฯกับเด็กไม่พกพร่องฯ แบบรายบุคคล พบว่ามีความถูกต้องด้านบกพร่องฯ เท่ากับ 82.60% (ครูคัดกรองว่าเป็น LD 23 คน แต่จริงๆแล้วไม่เป็นLD 4 คน) ด้านไม่บกพร่อง เท่ากับ 0 ด้านและในส่วนของการ คัดกรองโดยโมเดลด้วยเทคนิค Decision Tree พบว่า ความถูกต้องด้านบกพร่องฯ เท่ากับ 89.47% (ทำนายว่าเป็นไม่เป็น LD 2 คน แต่จริงๆแล้วเป็นLD) ด้านไม่บกพร่อง เท่ากับ 100% (ทำนายได้ตรงกับที่แพทย์ยืนยันทุกคน)

เพื่อให้เห็นภาพการเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูลกับรูปแบบการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ รายละเอียดดังตารางที่ 4.4

ตารางที่ 4.4 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่กับการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ

ลำดับ	รายการ	แบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูล	การคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ
1	ความถูกต้อง	ความถูกต้อง เท่ากับ 89.47 % ((17/19)*100)	ความถูกต้อง เท่ากับ 82.60 % ((19/23)*100)
2	ระยะเวลา	สามารถเก็บข้อมูลและประมวลผลได้ทันที	ผู้คัดกรองต้องใช้เวลา 1 ภาคการศึกษาหรือ 4 เดือน เพื่อสังเกตพฤติกรรมผู้เรียน
3	แหล่งที่มาข้อมูล	เก็บข้อมูลจากผู้เรียนโดยตรง	ข้อมูลนักเรียนได้มาจากผู้คัดกรอง อาจทำให้ได้ข้อมูลคลาดเคลื่อนจากความเป็นจริง
4	ความสะดวก	ลดภาระงานครู (ขั้นตอนไม่ซับซ้อนข้อมูลอยู่ในฐานอยู่แล้วและทดสอบกับผู้เรียนโดยตรง)	เพิ่มภาระงานครู (ต้องสังเกตผู้เรียนและกรอกข้อมูลจำนวนมาก)

ตารางที่ 4.4 การเปรียบเทียบระหว่างการคัดกรองโดยแบบคัดกรองที่พัฒนาขึ้นใหม่กับการคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ (ต่อ)

ลำดับ	รายการ	แบบคัดกรองที่พัฒนาขึ้นใหม่ด้วยรูปแบบการใช้เทคนิคเหมืองข้อมูล	การคัดกรองรูปแบบเดิมด้วยแบบคัดกรองกระทรวงศึกษาธิการ
5	วิธีเก็บข้อมูล	เก็บข้อมูลจากผู้เรียนโดยตรง (แบบทดสอบ+ข้อมูลพื้นฐาน)	นักเรียน 1 คน ต่อผู้คัดกรอง 2 คน ขึ้นไป
6	รูปแบบการประเมินผล	ทำนายผลแบบอัตโนมัติด้วย Machine Learning	ใช้การคำนวณเองเทียบกับเกณฑ์ที่กำหนดไว้
7	ปริมาณผู้เรียนเพื่อคัดกรอง	คัดกรองได้จำนวนมากและรวดเร็ว	ถ้ามีนักเรียนจำนวนมากต้องใช้ผู้คัดกรองจำนวนมาก
8	จำนวนข้อคำถามในการคัดกรอง	มีจำนวน 30 ข้อ	มีจำนวน 40 ข้อ

สรุปได้ว่า แบบคัดกรองที่พัฒนาใหม่จะมีค่าความน่าเชื่อถือสามารถใช้คัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้แทนแบบคัดกรองของกระทรวงศึกษาธิการได้อย่างมีประสิทธิภาพ

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการใช้เทคนิคการทำเหมืองข้อมูลเพื่อคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้สรุปผลการวิจัย ดังนี้

1. ปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ ได้แก่ คะแนนเฉลี่ยวิชาคณิตศาสตร์, ผลสอบด้านการอ่าน, คะแนนเฉลี่ยวิชาภาษาไทย, จำนวนพี่น้อง, ระดับชั้นเรียน

2. อัลกอรริทึม Neural Network และ Decision Tree มีค่าประสิทธิภาพสูงสุด โดยค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy) เท่ากับ 95.82% รองลงมาคืออัลกอรริทึม Naïve Bayes เท่ากับ 92.05%

3. ผลการเปรียบเทียบการแบบคัดกรองของกระทรวงศึกษาธิการที่ดำเนินการคัดกรองแบบปกติ มีความถูกต้องด้านบกพร่องฯ เท่ากับ 82.60% ด้านไม่บกพร่องฯ เท่ากับ 0 และในส่วนของการทำนายโมเดลด้วยเทคนิค Decision Tree จากการใช้แบบคัดกรองที่พัฒนาขึ้นใหม่ มีความถูกต้องด้านบกพร่องฯ เท่ากับ 89.47% ด้านไม่บกพร่องฯ เท่ากับ 100%

อภิปรายผล

1. ปัจจัยที่ส่งผลต่อการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้ ได้แก่ คะแนนเฉลี่ยวิชาคณิตศาสตร์, ผลสอบด้านการอ่าน, คะแนนเฉลี่ยวิชาภาษาไทย, จำนวนพี่น้อง, ระดับชั้นเรียน เพราะผู้วิจัยได้ศึกษาตำรา เอกสารงานวิจัยที่เกี่ยวข้องกับความบกพร่องทางการเรียนรู้รวมถึงการสัมภาษณ์ครูผู้สอนแล้วนำมาสรุปเป็นแอตทริบิวต์ (Attribute) เพื่อนำมาสร้างเครื่องมือในการเก็บรวบรวมข้อมูล ซึ่งเครื่องมือที่ใช้ในการวิจัยครั้งนี้ประกอบด้วยแบบบันทึกข้อมูลพื้นฐานของนักเรียน ได้แก่ เพศ จำนวนพี่น้อง ช่วงชั้น อายุ ผลการเรียนสะสมภาษาไทย ผลการเรียนสะสมคณิตศาสตร์ ผลการเรียนสะสมรวมทุกวิชา สถานะครอบครัว ผู้ปกครองปัจจุบัน ข้อมูลบิดา–มารดา (อายุ อาชีพ รายได้ ภูมิลำเนา) และแบบทดสอบคัดกรองด้านการอ่าน ด้านการเขียน และด้านการคิดคำนวณ สอดคล้องกับ [1] [3][4] สรุปว่า การบกพร่องทางการเรียนรู้มักมีปัญหาด้าน การเรียนตั้งแต่ชั้นประถมต้น ถ้ามีความรุนแรงจะพบปัญหาการเรียนร่วมกับปัญหาพฤติกรรม อารมณ์ พบในเพศ

ชายมากกว่าเพศหญิง มีการถ่ายทอดโรคในครอบครัว ความบกพร่องที่พบคือขาดทักษะการอ่านหนังสือ เขียนหนังสือแลการคำนวณ และสอดคล้องกับ [3] กล่าวว่า เด็กที่มีความบกพร่องในการเรียนรู้ เป็นเด็กทั่วไปที่มีสติปัญญาและไอคิวปกติหรืออาจสูงกว่าปกติ แต่มีความบกพร่องในเรื่องบางอย่างโดยเฉพาะเรื่องการอ่านหนังสือ การเขียนหนังสือ และการคิดคำนวณ

2. อัลกอรริทึม Neural Network และ Decision Tree มีค่าประสิทธิภาพสูงสุด โดยค่าความถูกต้องโดยรวมของโมเดลรวมทุกคลาส (Accuracy) เท่ากับ 95.82% รองลงมาคืออัลกอรริทึม Naïve Bayes เท่ากับ 92.05% เพราะโครงข่ายประสาทเทียมเป็นแบบจำลองที่ไม่ใช้พารามิเตอร์แต่จะะใช้หลักการเรียนรู้ของแบบจำลองในการสร้างความสัมพันธ์ของค่าน้ำหนักและโครงสร้างภายในโครงข่ายซึ่งมีความยืดหยุ่นและมีประสิทธิภาพในการสร้างความสัมพันธ์ของข้อมูลที่มีความผันผวนมาก ซึ่งยากต่อการหาความสัมพันธ์ทางสถิติและสร้างค่าพยากรณ์นอกกลุ่มตัวอย่างได้ดีกว่าแบบจำลองอื่น สอดคล้องกับ [23] ทำการวิจัยเรื่อง การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก พบว่า อัลกอริทึม Neural Network ให้ความแม่นยำ 87.5 สูงกว่า Naïve Bayes ให้ความแม่นยำ 81.25 แต่อัลกอริทึมนิวรัลเน็ทเวิร์กมีโครงสร้างการทำงานซับซ้อนทำให้ใช้เวลาในการประมวลผลมากกว่า

ข้อเสนอแนะจากการทำวิจัย

การทำเหมืองข้อมูลขั้นตอนที่ใช้เวลามากคือ ขั้นตอน Data Understanding ต้องสร้างเครื่องมือที่สอดคล้องกับปัญหาการวิจัย และเหมาะสมกับกลุ่มเป้าหมายที่จะไปเก็บเพราะการขอข้อมูลกับ บางหน่วยงานจะต้องประสานหลายส่วนงาน ข้อมูลบางอย่างไม่สามารถนำมาใช้ได้เพราะเป็นความลับ ข้อมูลบางอย่างต้องใช้เวลาในการเก็บ เป็นผลทำให้ได้ข้อมูลไม่ครบถ้วน และไม่เพียงพอในการนำไปวิเคราะห์ และขั้นตอน Data Preparation ก็ถือว่าสำคัญเพราะต้องแปลง จัดกลุ่ม จัดประเภท ตรวจสอบความ ถูกต้อง ความสมบูรณ์ของข้อมูลให้เหมาะสมกับอัลกอริทึมที่จะนำมาสร้างโมเดล

จำนวนของข้อมูลจะมีผลต่อการเรียนรู้ของ Machine learning เพราะจากการวิจัยครั้งนี้มีกลุ่มตัวอย่างทั้งหมด 238 คน และจากลักษณะข้อมูลจะมีเด็กกลุ่มที่เสี่ยงเป็น LD แค่ 19 คน ที่เหลือ 215 คน จะเป็นเด็กที่ไม่มีความเสี่ยงเป็น LD ทำให้ผลการเรียนรู้ของ Machine learning ในส่วนที่โมเดลทำนายว่าเด็กไม่มีความเสี่ยงเป็น LD จะถูกต้องมากกว่าการที่โมเดลทำนายเด็กที่มีความเสี่ยงเป็น LD ดังนั้นการใช้เทคนิคการเรียนรู้แบบมีผู้สอน (supervised learning) ดังนั้นควรมีข้อมูลแต่ละคลาส (เป็น LD คลาส 1 กับ ไม่เป็นLD คลาส 0) เพื่อการจัดจำแนก (Classification) ให้มากเพียงพอเพื่อความถูกต้องของโมเดลในการทำนาย

แนวทางในการทำวิจัยต่อไป

จากการวิจัยครั้งนี้ได้พัฒนาโมเดลคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้โดยใช้อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree), นาอีฟเบย์ (Naïve Bayes), และโครงข่ายประสาท (Neural Network) ผลความแม่นยำ ความถูกต้องอยู่ในระดับที่สูงเป็นที่น่าพอใจ ทั้ง 3 อัลกอริทึม เนื่องจากผู้วิจัยได้ศึกษาเอกสาร งานวิจัยเกี่ยวกับเด็กที่มีความบกพร่องทางการเรียนรู้เพื่อสรุปเป็นปัจจัยที่ส่งผลต่อการคัดกรอง และได้สร้างแบบทดสอบคัดกรองแยกเป็นด้านการอ่าน การเขียน และการคำนวณ ตรวจสอบคุณภาพโดยผู้เชี่ยวชาญและทดลองใช้เบื้องต้น (try out) ก่อนนำไปเก็บข้อมูลกับกลุ่มตัวอย่างจริง อย่างไรก็ตามถ้ามีการเพิ่มขนาดกลุ่มตัวอย่างที่มากขึ้นโดยเฉพาะเพิ่มกลุ่มเด็กที่มีความบกพร่องทางการเรียนรู้เพื่อได้ข้อมูลใน Training data ให้โปรแกรมเรียนรู้มากขึ้น เนื่องจากครั้งนี้มีสัดส่วนเด็กที่มีความบกพร่องทางการเรียนรู้ยังไม่มากเท่าที่ควร และควรมีการวิจัยต่อยอดในเรื่องการพัฒนาโปรแกรมบทเรียนอัจฉริยะสำหรับเด็กบกพร่องทางการเรียนรู้ที่สามารถเก็บข้อมูลการเรียนด้วยระบบเรียลไทม์ (Real-Time System) เพื่อให้สามารถจัดโปรแกรมบทเรียนตามความสารถผู้เรียนรายบุคคล ตอบสนองผลการเรียนได้แบบทันที เป็นการกระตุ้นและพัฒนาการเรียนรู้ของเด็กบกพร่องทางการเรียนรู้ได้อย่างมีประสิทธิภาพ

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

กระทรวงศึกษาธิการ. (2550). *กฎกระทรวงกำหนดหลักเกณฑและวิธีการใหคนพิการมีสิทธิได้รับสิ่งอำนวยความสะดวก สื่อ บริการ และการให้ความช่วยเหลืออื่นใดทางการศึกษา พ.ศ. 2550.* ราชกิจจานุเบกษา, เล่มที่ 124, (ตอนพิเศษ 71ก), น. 1-11.

กระทรวงศึกษาธิการ. (2543). *คู่มือการคัดแยกและส่งต่อเพื่อการศึกษา.* กรุงเทพฯ:โรงพิมพ์องค์การรับส่งสินค้าและพัสดุภัณฑ์. (อัดสำเนา).

จริยา สงวนรัตน์ และเพ็ญณี แนรอท. (2554). การสังเคราะห์งนวิจัยเกี่ยวกับเด็กที่มีความบกพร่องทางการเรียนรู้ในประเทศไทย. *วารสารศึกษาศาสตร์*, ฉบับวิจัยบัณฑิตศึกษา. มหาวิทยาลัยขอนแก่น, 5(1), 16-21.

จิราภา เลาหะวรนันท์ และคณะ. (2558). การใช้เทคนิคการท้าเหมืองข้อมูลในการจำแนกและคัดเลือกแขนงวิชาสำหรับนักศึกษาคณะเทคโนโลยีสารสนเทศ. *วารสารเทคโนโลยีสารสนเทศ. ลาดกระบัง,* 4(2), 70-79.

ชนวัฒน์ ศรีสอ้าน. (2551). *ฐานข้อมูล คลังข้อมูล และเหมืองข้อมูล.* กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยรังสิต.

ดารณี อุทัยรัตนกิจ, ชาญวิทย์ พรนภดล และคณะ. (2550). *แบบคัดกรองนักเรียนที่มีภาวะสมาธิสั้นบกพร่องทางการเรียนรู้และออทิซึม.* กรุงเทพฯ: ศูนย์วิจัยการศึกษาเพื่อเด็กที่มีความต้องการความช่วยเหลือพิเศษ. กรุงเทพฯ: บริษัทพัฒนาคุณภาพวิชาการ (พว.)จำกัด.

นฤชล โรจนบุรานนท์. (2562). *คู่มือการใช้งาน RapidMiner Studio 9 เบื้องต้น.* สืบค้นเมื่อ 9 พฤษภาคม 2562, จาก https://anyflip.com/wwphz/yqdr

เบญจพร ปัญญายง. (2545). *คู่มือ "ช่วยเหลือเด็กบกพร่องด้านการเรียนรู้".* กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

ปณิธิ แก้วสวัสดิ์. (2553). *เหมืองข้อมูล : การค้นหาความรู้และการขุดข้อมูลคอมพิวเตอร์และเทคโนโลยีขั้นสูง.*

ประกาศกระทรวงศึกษาธิการ. (2552). *กำหนดประเภทและหลักเกณฑ์ของคนพิการทางการศึกษา พ.ศ. 2552.* ราชกิจจานุเบกษา, เล่มที่ 126, (ตอนพิเศษ 80ง), 45-47.

ประเทือง วงษ์ทอง, วาฤทธิ์ กันแก้ว, อภิชชญา ขวัญแก้ว, และยุวดี โฉมแดง. (2563). *การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัยของผู้ประกอบอาชีพด้านคอมพิวเตอร์.* คณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ.

บรรณานุกรม (ต่อ)

ผดุง อารยะวิญญู. (2542). *การศึกษาสำหรับเด็กที่มีความต้องการพิเศษ.* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: แว่นแก้ว.

พรนภา ชุมเชื้อ. (2562). *การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทสำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม.* (วิทยานิพนธ์ปริญญามหาบัณฑิต). สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยทักษิณ.

พีร วงศ์อุปราช. (2561). *รายงานวิจัยฉบับสมบูรณ์ โครงการการศึกษาเชิงจิตประสาทวิทยาและการพัฒนาแบบคัดกรอง เน้นกระบวนการทางปัญญาในเด็กที่มีความเสี่ยงต่อภาวะความบกพร่องทางการเรียนรู้ด้านคณิตศาสตร์.* โครงการวิจัยประเภทงบประมาณเงินรายได้จากเงินอุดหนุนรัฐบาล (งบประมาณแผ่นดิน) ประจำปีงบประมาณ พ.ศ. 2560. (รายงานผลการวิจัย). วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา.

รัฐธรรมนูญแห่งราชอาณาจักรไทย พุทธศักราช 2550. (2550). ราชกิจจานุเบกษา, เล่ม 124 (ตอนที่ 47 ก), 15.

วัชรินทร์ ฮ่ำรัตนาพร และคณะ. (2551). *สร้างด้วยใจเพื่อเด็กแอลดี* ส่วนส่งเสริมการผลิตสื่อการศึกษาเพื่อคนพิการศูนย์เทคโนโลยีทางการศึกษา สำนักงานส่งเสริมการศึกษานอกระบบและการศึกษาตามอัธยาศัยสำนักงานปลัดกระทรวงศึกษาธิการ, กระทรวงศึกษาธิการ.

วินัดดา ปิยะศิลป์ และคณะ. (2558). ความบกพร่องทางการเรียนรู้และโรคที่พบร่วม. *วารสารสมาคมจิตแพทย์แห่งประเทศไทย, 60(4),* 287-296.

วีระยุทธ พิมพาภรณ์ และพยุง มีสัจ. (2555). เทคนิคเหมืองข้อมูลสำหรับการพยากรณ์ผลสัมฤทธิ์ทางการศึกษาด้วยวิธีกรจัดการเรียนการสอนแบบผสมผสาน. *วารสารศรีปทุมปริทัศน์.* ฉบับวิทยาศาสตร์และเทคโนโลยี.

วุฒิชัย กำจรกิตติคุณ. (2561). *การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก.* (วิทยานิพนธ์ปริญญามหาบัณฑิต). สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม, วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์, มหาวิทยาลัยธุรกิจบัณฑิตย์.

ศรียา นิยมธรรม. (2549). *ปัญหายุ่งยากทางการเรียนรู้.* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: แว่นแก้ว.

ศรียา นิยมธรรม. (2537). *รายงานการวิจัยการสร้างแบบคัดแยกเด็กที่มีปัญหาทางการเรียนรู้.* กรุงเทพฯ : ภาควิชาการศึกษาพิเศษ คณะศึกษาศาสตร มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.

บรรณานุกรม (ต่อ)

สายชล สินสมบูรณ์ทอง. (2558). *การทำเหมืองข้อมูล Data Mining.* กรุงเทพฯ : จามจุรีโปรดักท์.

สุวิมล สิทธิชาติ. (2560). การวิเคราะห์คุณลักษณะพื้นฐานทางการศึกษาด้วยเทคนิคเหมืองข้อมูล. *วารสารเทคโนโลยีสารสนเทศ. 13(2),* 20-28.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2556). *เอกสารประกอบการจัดอบรม หน่วยที่ 3 การใช้แบบคัดกรองคนพิการทางการศึกษา 9 ประเภท ตามประกาศคณะกรรมการพิจารณาให้คนพิการได้รับสิทธิช่วยเหลือทางการศึกษา เรื่อง กำหนดหลักเกณฑ์และวิธีการ การรับรองบุคคลของสถานศึกษาว่าเป็นคนพิการ พ.ศ. 2556.* กรุงเทพฯ: สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (อัดสำเนา).

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิงเบื้องต้น.* กรุงเทพฯ : เอเชียดิจิตอลการพิมพ์.

ภาษาต่างประเทศ

Allen K. Eileen. ;& Schwartz, llene S. (2001) .The Exceptional Child Inclusion in Early Childhood Education . United States of America.

Boston Children's Hospital. (2012). Learning Disorder. Retrieved November 18, 2012, from http://www.childrenshospital.org/az/Site1229/mainpageS1229P1.html

C. Shearer. (2000). *The CRISP-DM model : The new blueprint for data mining.* Journal of Data Warehousing, *5(4),* 13–22.

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายนคร ละลอกน้ำ
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต (วศ.บ.) สาขาวิชาวิศวกรรมไฟฟ้ากำลัง ม.ภาคตะวันออกเฉียงเหนือ จ.ขอนแก่น พ.ศ.2540 การศึกษามหาบัณฑิต (กศ.ม.) สาขาวิชาเทคโนโลยีการศึกษา ม.บูรพา จ.ชลบุรี พ.ศ.2546 ปรัชญาดุษฎีบัณฑิต (ปร.ด.) สาขาวิชาเทคโนโลยีการศึกษา ม.บูรพา จ.ชลบุรี พ.ศ.2553 ศิลปศาสตรมหาบัณฑิต (ศศ.ม.) สาขาวิชาเทคโนโลยีการฝึกอบรมเพื่อการพัฒนาทรัพยากรมนุษย์ ม.บูรพา จ.ชลบุรี พ.ศ.2558 Master II of Educational Sciences (Professionals of Evaluation, and Management) University of Montpellier III, France พ.ศ.2559 วิศวกรรมศาสตรมหาบัณฑิต (วศ.ม.) สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม ม.ธุรกิจบัณฑิตย์ พ.ศ.2564
ตำแหน่งและสถานที่ทำงานปัจจุบัน	พ.ศ. 2548-ปัจจุบัน อาจารย์ประจำภาควิชานวัตกรรมและเทคโนโลยี การศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา จ.ชลบุรี พ.ศ. 2560-ปัจจุบัน หัวหน้าภาควิชานวัตกรรมและเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา จ.ชลบุรี พ.ศ. 2562-ปัจจุบัน ผู้ช่วยศาสตราจารย์ สาขาเทคโนโลยีการศึกษา