

Volume

1

SCHOOL OF LIFELONG EDUCATION

Chiang Mai University

Intelligence Data Analysis with RapidMiner

CMU LIFELONG EDUCATION

Intelligence Data Analysis with RapidMiner

Assistant Professor Chumphol Bunkhumpornpat, Ph.D.
Department of Computer Science
Faculty of Science
Chiang Mai University

เนื้อหา

บทที่ 1. การติดตั้งโปรแกรม

บทที่ 2. การเตรียมข้อมูลและการแสดงมโนภาพข้อมูล

บทที่ 3. การวิเคราะห์การถดถอย

บทที่ 4. การจัดกลุ่มข้อมูล

บทที่ 5. การจำแนกข้อมูล

บทที่ 6. กฎความสัมพันธ์

การติดตั้งโปรแกรม

How to install software

บทแรกนี้ นำเสนอการติดตั้งโปรแกรม RapidMiner ทีละขั้นตอนแบบ Step by Step
สำหรับ ตั้งแต่การเข้าสู่เว็บไซต์เพื่อดาวน์โหลดโปรแกรม การลงทะเบียนเพื่อเข้าใช้งานโปรแกรม
และ การเข้าสู่โปรแกรมนี้

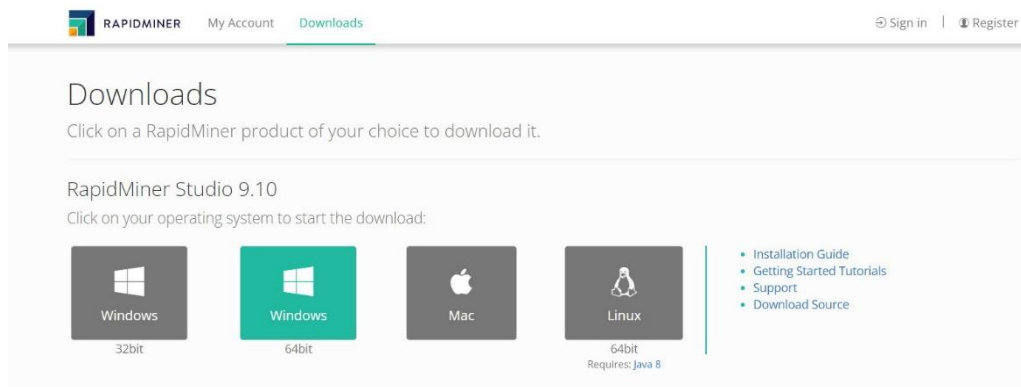


เว็บไซต์สำหรับดาวน์โหลดโปรแกรม

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

การดาวน์โหลดโปรแกรม

เมื่อเข้าสู่เว็บไซต์จาก URL ข้างต้นแล้ว หน้าตาของเว็บไซต์จะเป็นดังรูปที่ 1.1 ให้เลือกไอคอนที่ตรงกับระบบปฏิบัติการที่ใช้ในเครื่องคอมพิวเตอร์ของผู้อบรมเอง เพื่อดาวน์โหลดไฟล์สำหรับติดตั้งโปรแกรมลงในเครื่องคอมพิวเตอร์



รูปที่ 1.1 เว็บไซต์สำหรับการดาวน์โหลดโปรแกรม



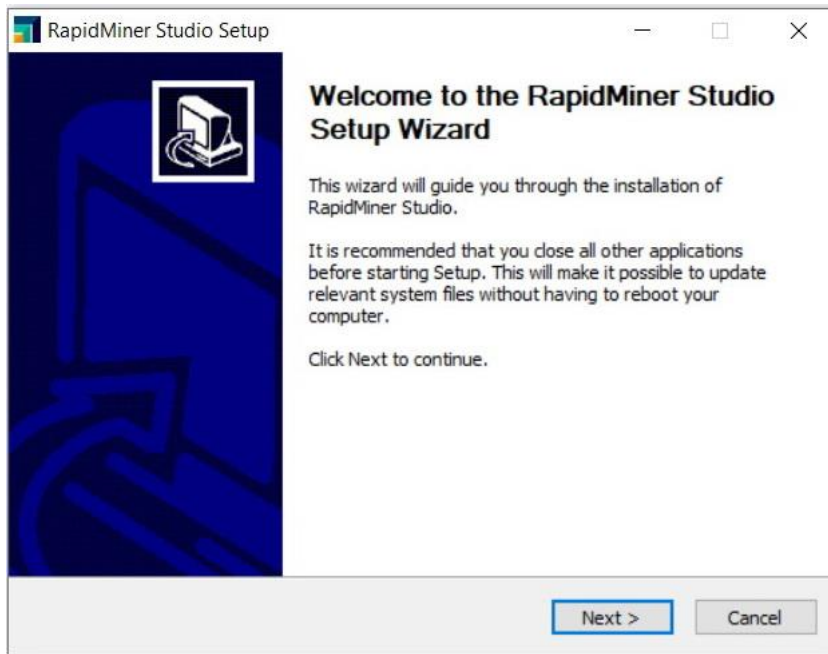
Icon ไฟล์ติดตั้ง

เมื่อดาวน์โหลดไฟล์ลงเครื่องคอมพิวเตอร์ของผู้ใช้แล้ว
ไอคอนของไฟล์ดังกล่าวมีลักษณะเหมือนรูปด้านซ้ายมือนี้
ให้ผู้ใช้ดับเบิลคลิกเพื่อลงโปรแกรมต่อไป

การติดตั้งโปรแกรม

เมื่อดับเบิลคลิกไฟล์สำหรับติดตั้งแล้ว จะขึ้นหน้าต่างที่มีลักษณะดังรูปที่ 1.2 ให้ดำเนินการตามขั้นตอน ดังต่อไปนี้

1. กดปุ่ม **Next>** เพื่อเริ่มติดตั้ง
2. กดปุ่ม **I Agree** เพื่อไปขั้นตอนถัดไป
3. เลือกโฟลเดอร์สำหรับติดตั้ง แล้วกดปุ่ม **Install**
4. สังเกตว่า **Launch RapidMiner Studio** จะมีเครื่องหมายถูกอยู่
5. กดปุ่ม **Finish** เพื่อสิ้นสุดการติดตั้ง



รูปที่ 1.2 การติดตั้งโปรแกรม

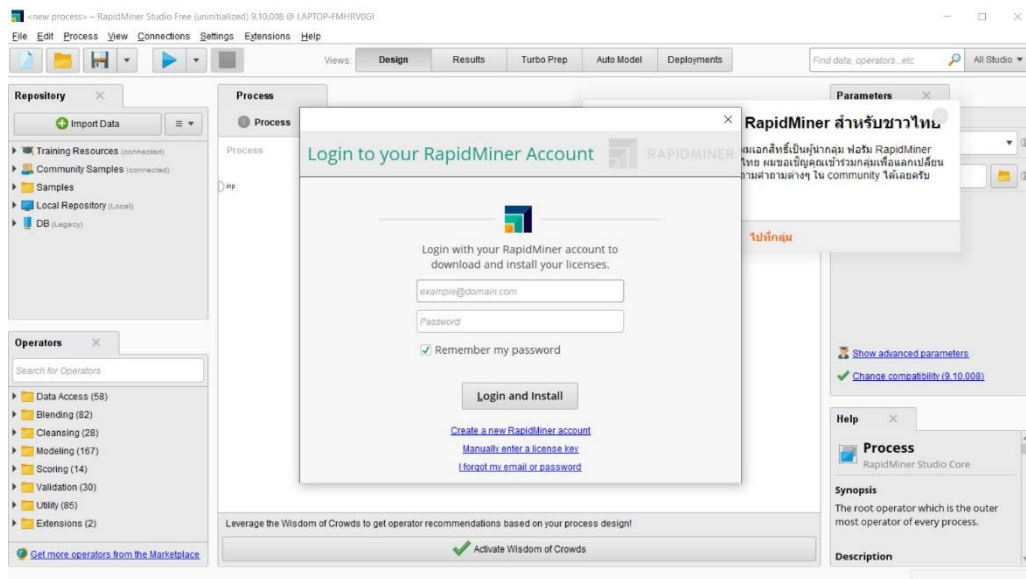


Shortcut โปรแกรม

เมื่อติดตั้งโปรแกรมสำเร็จแล้ว ให้สังเกตว่าบนหน้าจอ Desktop ของผู้ใช้ จะปรากฏ Shortcut ลักษณะทางด้านซ้ายมือ ให้ดับเบิลคลิกเพื่อเข้าสู่โปรแกรม

การลงทะเบียนผู้ใช้งาน

เมื่อดับเบิลคลิก Shortcut เพื่อเข้าสู่โปรแกรมแล้ว ในครั้งแรกจะมีลักษณะตามรูปที่ 1.3 ซึ่งผู้ใช้ จำเป็นต้องมี Account ของ RapidMiner สำหรับ Login ใช้งานโปรแกรม ในกรณีที่ยังไม่มี Account ดังกล่าว สามารถทำการสมัครใหม่ได้ โดยคลิกที่ Create a new RapidMiner account จะเข้ามายัง หน้าจอตามรูปที่ 1.4 ให้เลือก Educational purposes แล้วกรอกข้อมูลของผู้ใช้ เพื่อสร้าง Account



รูปที่ 1.3 การเข้าสู่โปรแกรม

My Account

Downloads

Sign in

Register

Create your RapidMiner Account

This account gives you access to RapidMiner products (trials, licenses, updates, and extensions), training via the Academy, and the RapidMiner Community.

What are you using Rapidminer for?

Commercial purposes (e.g., business, evaluation, non-for-profit)

Educational purposes (e.g., educator, student)

First name:

Last name:

University:

Role:

Please select an option

Email address:

Create a password:

Confirm your password:

รูปที่ 1.4 การลงทะเบียนโปรแกรม

การเตรียมข้อมูล และการแสดงมโนภาพข้อมูล

Data Preparation and Data Visualization

บทที่สองนี้ นำเสนอการนำเข้าเซตข้อมูลไปยังโปรแกรม RapidMiner การแสดงค่าทาง
สำหรับ สถิติของฟีเจอร์ต่าง ๆ ในเซตข้อมูล และ การแสดงมโนภาพของแต่ละคลาสในเซตข้อมูล



เว็บไซต์สำหรับดาวน์โหลดเซตข้อมูล

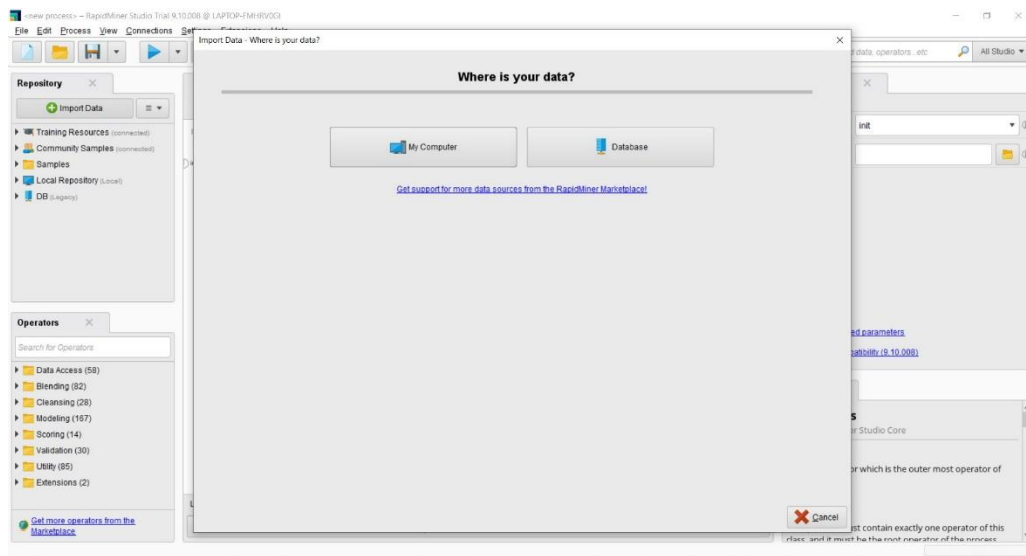
<https://gist.github.com/netj/8836201>

ดาวน์โหลดเซตข้อมูล

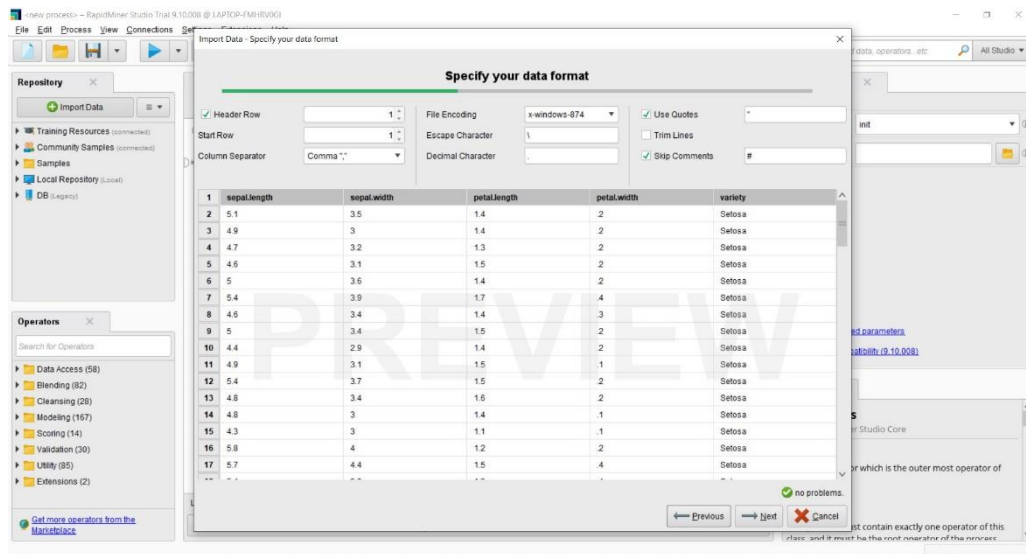
เมื่อเข้าสู่เว็บไซต์จาก URL ข้างต้นแล้ว ให้กดปุ่ม **Download ZIP** เพื่อดาวน์โหลดเซตข้อมูลลงใน
เครื่องคอมพิวเตอร์ของผู้อบรม เนื่องจากไฟล์นี้ถูกบีบอัดมา ให้ทำการแตกไฟล์นี้ออกมาก่อน จะได้
ไฟล์ชื่อ iris

การนำเข้าเซตข้อมูล

เมื่อเข้าสู่โปรแกรม RapidMiner แล้ว ให้คลิกปุ่ม **Import Data** จะปรากฏหน้าต่างตามรูปที่ 2.1 ในกรณีที่เซตข้อมูลอยู่ในเครื่องคอมพิวเตอร์ของผู้ใช้ ให้กดปุ่ม **My Computer** เพื่อเลือกไฟล์ดังกล่าว จากนั้นหน้าจอจะแสดงตามรูปที่ 2.2 เพื่อให้ผู้ใช้ระบุรูปแบบข้อมูล ให้กดปุ่ม **Next** ในกรณีที่ไม่ต้องการเปลี่ยนรูปแบบ ให้สังเกตถ้ามีคำว่า **no problems.** แสดงว่าไม่พบปัญหาใด ๆ



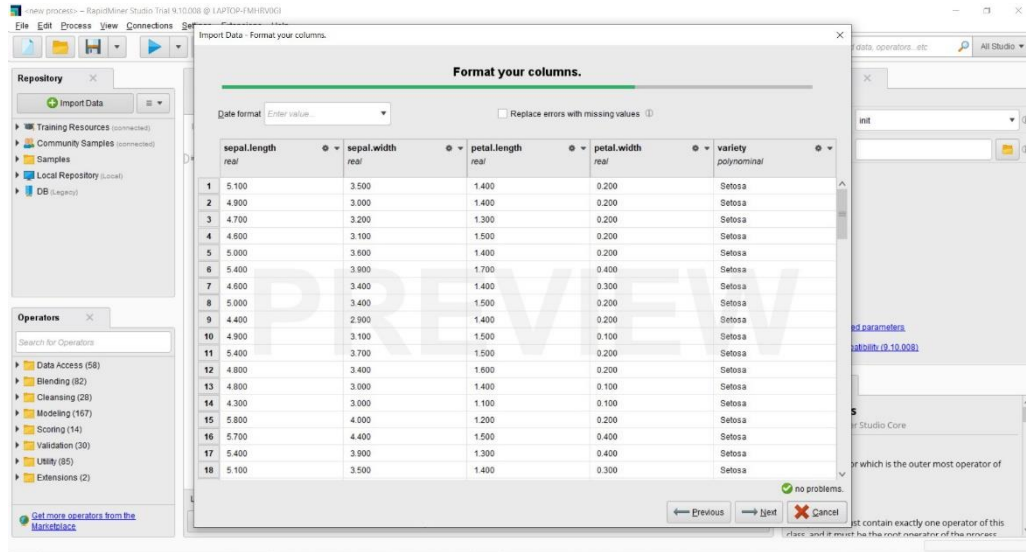
รูปที่ 2.1 การนำเข้าเซตข้อมูล



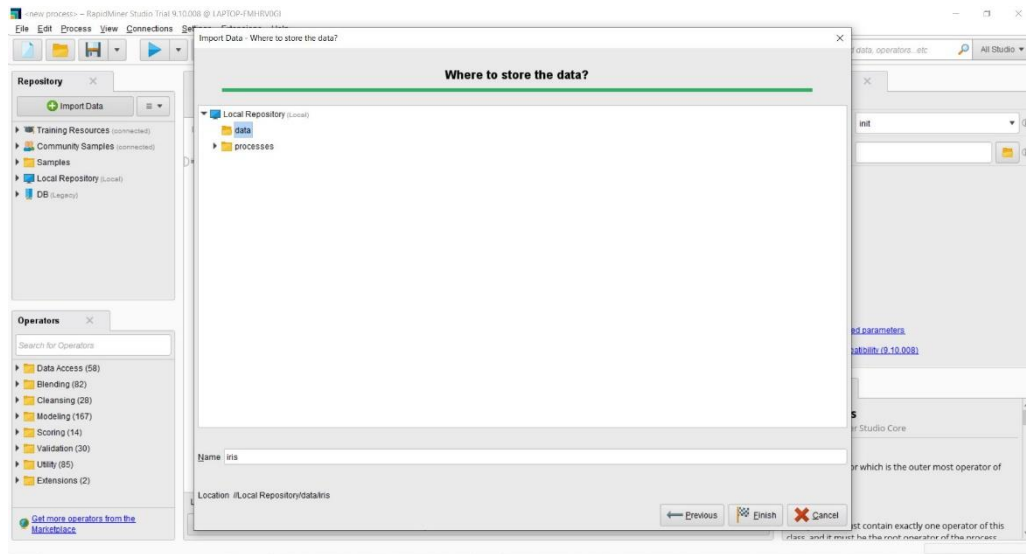
รูปที่ 2.2 การระบุรูปแบบข้อมูล

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER

ต่อไปเป็นการระบุรูปแบบคอลัมน์ แสดงตามรูปที่ 2.3 ได้แก่การกำหนดชนิดข้อมูลในแต่ละคอลัมน์ ในกรณีที่ชนิดข้อมูลถูกต้องและไม่พบข้อผิดพลาดใด ๆ ให้กดปุ่ม **Next** เพื่อเข้าสู่ขั้นตอนถัดไป แสดงตามรูปที่ 2.4 ในส่วนนี้ให้เลือกพื้นที่สำหรับจัดเก็บข้อมูล ปกติแล้วให้เก็บในโฟลเดอร์ชื่อ **data**



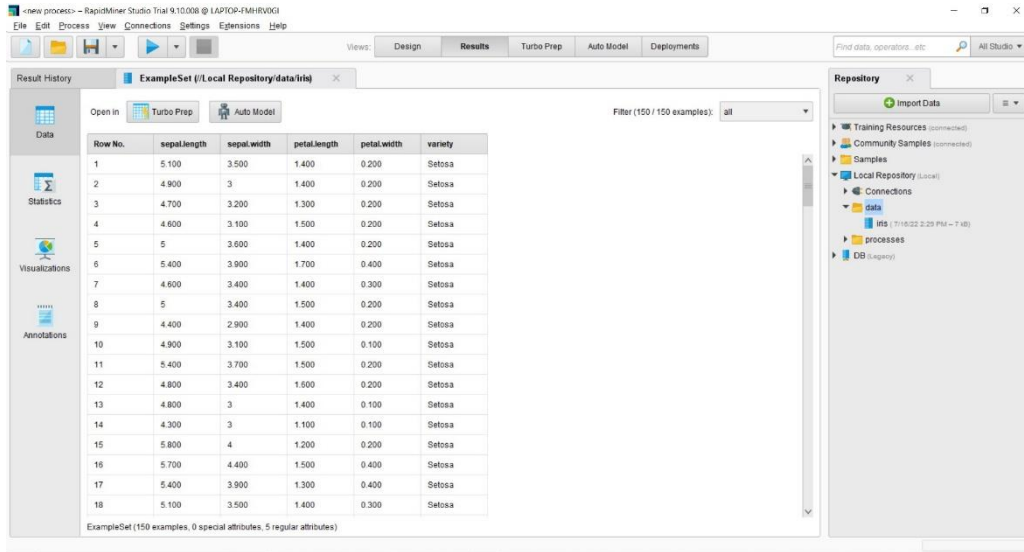
รูปที่ 2.3 การระบุรูปแบบคอลัมน์



รูปที่ 2.4 การเลือกพื้นที่จัดเก็บเซตข้อมูล

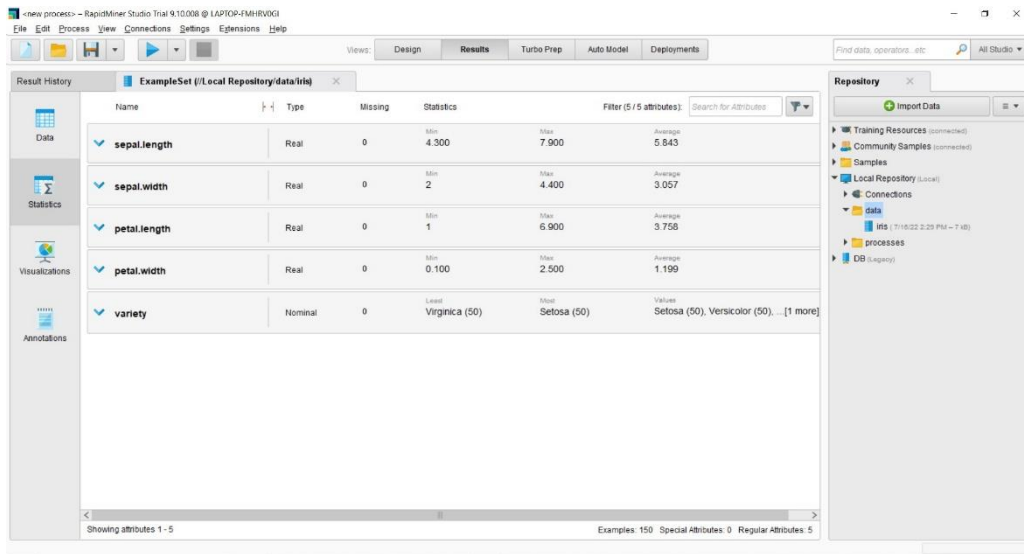
INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER

เมื่อนำเข้าข้อมูลสำเร็จแล้ว หน้าจอจะแสดงตามรูปที่ 2.5 เป็นการแสดงค่าแต่ละแถวในเซตข้อมูล ให้ผู้ใช้กดปุ่ม **Statistics** จะแสดงค่าทางสถิติของแต่ละฟีเจอร์ ได้แก่ ค่าน้อยที่สุด ค่ามากที่สุด และค่าเฉลี่ย



Row No.	sepal.length	sepal.width	petal.length	petal.width	variety
1	5.100	3.500	1.400	0.200	Setosa
2	4.900	3	1.400	0.200	Setosa
3	4.700	3.200	1.300	0.200	Setosa
4	4.600	3.100	1.500	0.200	Setosa
5	5	3.600	1.400	0.200	Setosa
6	5.400	3.900	1.700	0.400	Setosa
7	4.600	3.400	1.400	0.300	Setosa
8	5	3.400	1.500	0.200	Setosa
9	4.400	2.900	1.400	0.200	Setosa
10	4.900	3.100	1.500	0.100	Setosa
11	5.400	3.700	1.500	0.200	Setosa
12	4.800	3.400	1.600	0.200	Setosa
13	4.800	3	1.400	0.100	Setosa
14	4.300	3	1.100	0.100	Setosa
15	5.800	4	1.200	0.200	Setosa
16	5.700	4.400	1.500	0.400	Setosa
17	5.400	3.900	1.300	0.400	Setosa
18	5.100	3.500	1.400	0.300	Setosa

รูปที่ 2.5 การแสดงค่าในเซตข้อมูล

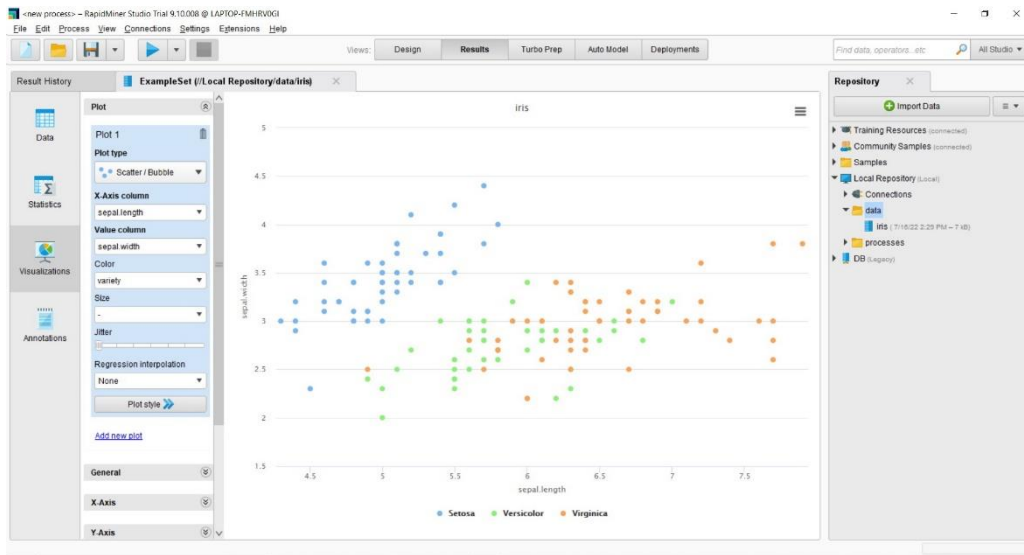


Name	Type	Missing	Statistics	Filter (5 / 5 attributes)
sepal.length	Real	0	Min: 4.300, Max: 7.900, Average: 5.843	
sepal.width	Real	0	Min: 2, Max: 4.400, Average: 3.057	
petal.length	Real	0	Min: 1, Max: 6.900, Average: 3.758	
petal.width	Real	0	Min: 0.100, Max: 2.500, Average: 1.199	
variety	Nominal	0	Level: Virginica (50), Most: Setosa (50), Values: Setosa (50), Versicolour (50), ... [1 more]	

รูปที่ 2.5 การแสดงค่าทางสถิติของฟีเจอร์ต่าง ๆ ในเซตข้อมูล

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER

ให้ผู้ใช้งานปุ่ม **Visualizations** จะเป็นการแสดงมโนภาพของเซตข้อมูล สำหรับ iris ประกอบไปด้วย 3 คลาส ได้แก่ Setosa Versicolor และ virginica แพนด้วยสีต่าง ๆ ในระนาบสองแกน



รูปที่ 2.6 การแสดงมโนภาพของแต่ละคลาสในเซตข้อมูล

การวิเคราะห์การถดถอย

Regression Analysis

บทที่สามนี้ นำเสนอการเตรียมข้อมูลก่อนการประมวลผล การสร้างตัวแบบการวิเคราะห์
สำหรับ การถดถอย การประเมินค่าประสิทธิภาพของตัวแบบ และ การนำตัวแบบไปประมาณค่า
ในเซตข้อมูลจริง



เว็บไซต์สำหรับดาวน์โหลดเซตข้อมูล

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

การเตรียมข้อมูลก่อนการประมวลผล

เมื่อเข้าสู่เว็บไซต์จาก URL ข้างต้น ซึ่งเป็นเซตข้อมูล Wine Quality ให้เข้าไปยัง Data Folder และ
บันทึกไฟล์ชื่อ `winequality-red.csv` ลงในเครื่องคอมพิวเตอร์ของผู้เข้าอบรม ทำการนำเข้าเซต
ข้อมูลทั้งสองลงในโปรแกรม RapidMiner ให้เรียบร้อย

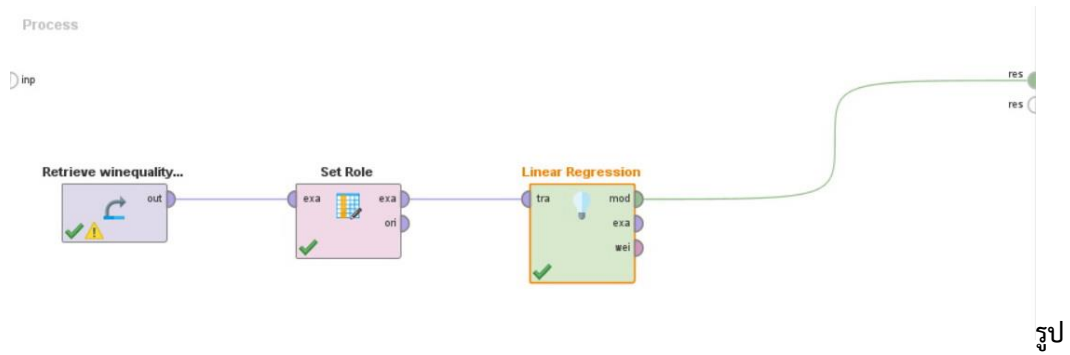
การสร้างตัวแบบ

ในหน้า Design ให้เชื่อมต่อกล่องดังต่อไปนี้

1. data → winequality-red
2. Set Role กำหนด quality เป็น Label
3. Linear Regression

หลังจากเชื่อมต่อกล่องทั้งหมดเข้าด้วยกันแล้วตามรูปที่ 3.1 ให้กดปุ่ม ► เพื่อประมวลผลตัวแบบ กดปุ่ม Results เพื่อดูสมการเชิงเส้นที่ได้ สำหรับเซตข้อมูลนี้จะได้ผลลัพธ์ดังนี้

$0.009 * \text{fixed acidity}$
 $- 1.090 * \text{volatile acidity}$
 $- 0.175 * \text{citric acid}$
 $- 1.890 * \text{chlorides}$
 $+ 0.005 * \text{free sulfur dioxide}$
 $- 0.003 * \text{total sulfur dioxide}$
 $- 0.502 * \text{pH}$
 $+ 0.887 * \text{sulphates}$
 $+ 0.294 * \text{alcohol}$
 $+ 4.442$



ที่ 3.1 ตัวแบบ Linear Regression

การประเมินค่าประสิทธิภาพของตัวแบบ

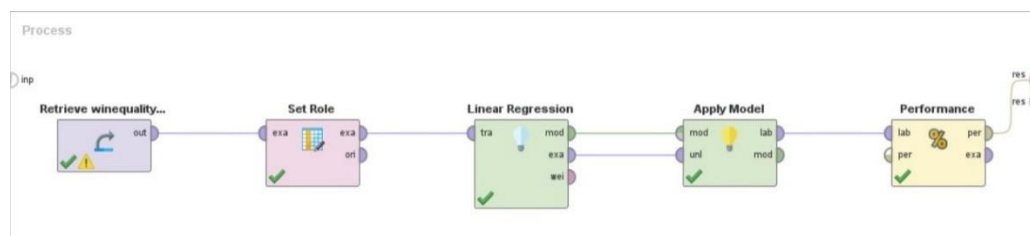
ในหน้า Design ให้ทำการเชื่อมต่อกล่องอีก 2 กล่อง ตามรูปที่ 3.2

1. Apply Model
2. Performance

เมื่อกดปุ่ม ► จะได้ผลลัพธ์ ดังนี้

root_mean_squared_error

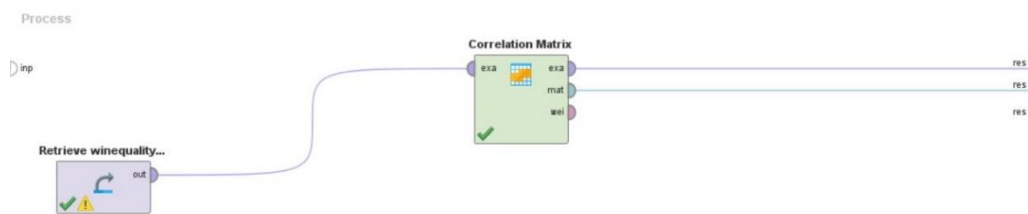
root_mean_squared_error: 0.646 +/- 0.000



รูปที่ 3.2 RMSE

การดูลักษณะความสัมพันธ์ของข้อมูล

ให้ทำการเชื่อมต่อกล่องตามรูปที่ 3.3 โดยเพิ่มกล่อง Correlation Matrix เข้ามา จากนั้นดูผลลัพธ์ใน หน้า Results จะเป็นการแสดงค่า สหสัมพันธ์ (Correlation) ของแต่ละคู่ของตัวแปร ยังมีสีม่วงเข้ม มากเท่าไร แสดงว่าตัวแปรสองตัวมีความสัมพันธ์กันมากตามไปด้วย แสดงตามรูปที่ 3.4



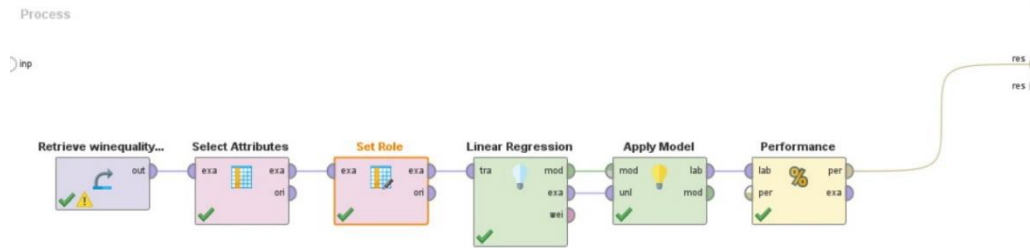
รูปที่ 3.3 Correlation Matrix

Attribut...	fixed ac...	volatile ...	citric ac...	residual...	chlorides	free sul...	total sul...	density	pH	sulphat...	alcohol	quality
fixed aci...	1	-0.256	0.672	0.115	0.094	-0.154	-0.113	0.668	-0.683	0.183	-0.062	0.124
volatile a...	-0.256	1	-0.552	0.002	0.061	-0.011	0.076	0.022	0.235	-0.261	-0.202	-0.391
citric acid	0.672	-0.552	1	0.144	0.204	-0.061	0.036	0.365	-0.542	0.313	0.110	0.226
residual ...	0.115	0.002	0.144	1	0.056	0.187	0.203	0.355	-0.086	0.006	0.042	0.014
chlorides	0.094	0.061	0.204	0.056	1	0.006	0.047	0.201	-0.265	0.371	-0.221	-0.129
free sulf...	-0.154	-0.011	-0.061	0.187	0.006	1	0.668	-0.022	0.070	0.052	-0.069	-0.051
total sulf...	-0.113	0.076	0.036	0.203	0.047	0.668	1	0.071	-0.067	0.043	-0.206	-0.185
density	0.668	0.022	0.365	0.355	0.201	-0.022	0.071	1	-0.342	0.149	-0.496	-0.175
pH	-0.683	0.235	-0.542	-0.086	-0.265	0.070	-0.067	-0.342	1	-0.197	0.206	-0.058
sulphates	0.183	-0.261	0.313	0.006	0.371	0.052	0.043	0.149	-0.197	1	0.094	0.251
alcohol	-0.062	-0.202	0.110	0.042	-0.221	-0.069	-0.206	-0.496	0.206	0.094	1	0.476
quality	0.124	-0.391	0.226	0.014	-0.129	-0.051	-0.185	-0.175	-0.058	0.251	0.476	1

รูปที่ 3.4 Correlarion

ให้ทำการเชื่อมต่อกล่องตามรูปที่ 3.5 โดยเพิ่มกล่อง Select Attribure เข้ามา ให้เลือก subset เป็นกลุ่มของตัวแปรที่ต้องการ จากนั้นสังเกตความแตกต่างของค่า RMSE

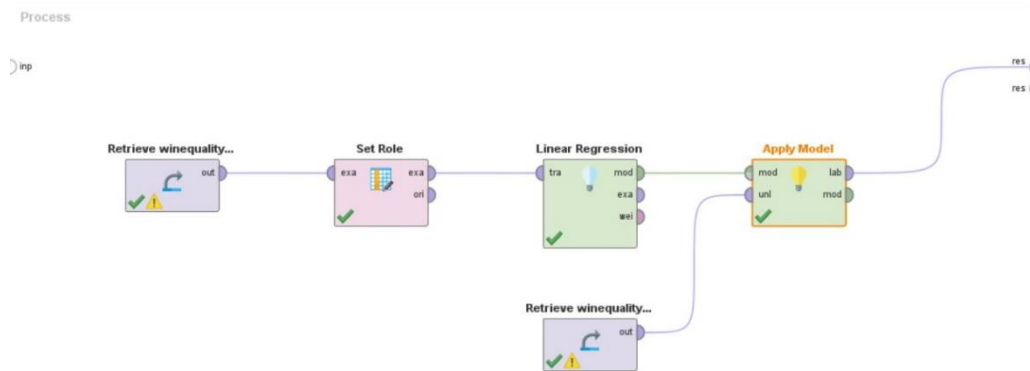
INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER



รูปที่ 3.5 Select Attribute

การประมาณค่าข้อมูล

ให้ทำการเชื่อมต่อกล่องตามรูปที่ 3.6 โดยเพิ่มเซตข้อมูลใหม่ที่ต้องการประมาณค่าลงไป หลังจากกด Run ให้สังเกตค่า prediction เป็นผลลัพธ์ที่ได้จากตัวแบบ



รูปที่ 3.6 Prediction

การจัดกลุ่มข้อมูล

Data Clustering, k-Means, Hierarchical clustering

บทที่สี่นี้ อธิบายแนวคิดของการจัดกลุ่มข้อมูล การจัดกลุ่มข้อมูลโดยใช้เทคนิคเคมีนส์
สำหรับ การจัดกลุ่มข้อมูลแบบลำดับขั้น การประเมินค่าประสิทธิภาพของตัวแบบ และการนำตัว
แบบไปจัดกลุ่มในเซตข้อมูลจริง



เว็บไซต์สำหรับดาวน์โหลดเซตข้อมูล

<https://archive.ics.uci.edu/ml/datasets/Facebook+Live+Sellers+in+Thailand>

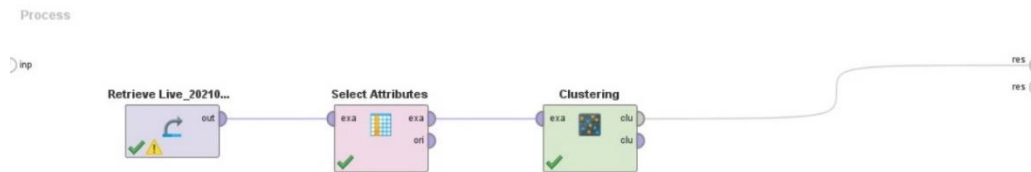
การเตรียมข้อมูลก่อนการประมวลผล

เมื่อเข้าสู่เว็บไซต์จาก URL ข้างต้น ซึ่งเป็นเซตข้อมูล Facebook Live Sellers in Thailand ให้เข้าไป
ยัง Data Folder และบันทึกไฟล์ชื่อ Live_20210128.csv ลงในเครื่องคอมพิวเตอร์ของผู้เข้าอบรม
ทำการนำเข้าเซตข้อมูลทั้งสองลงในโปรแกรม RapidMiner ให้เรียบร้อย

การสร้างตัวแบบ

ในหน้า Design ให้เชื่อมต่อกล่องตามรูปที่ 4.1 ดังนี้ สำหรับกล่อง Select Attributes ให้เลือกข้อมูลที่เป็นตัวเลขทั้งหมด เนื้

1. data → Live_20210128
2. Select Attribute เลือกข้อมูลที่มีชนิดข้อมูลเป็น Numeric ทั้งหมด
3. พิมพ์ k-Means จะเจอกล่องชื่อ Clustering



รูปที่ 4.1 K-means

หลังจากกด run แล้ว จะได้ผลลัพธ์ดังนี้

Cluster Model

Cluster 0: 6148 items

Cluster 1: 4 items

Cluster 2: 392 items

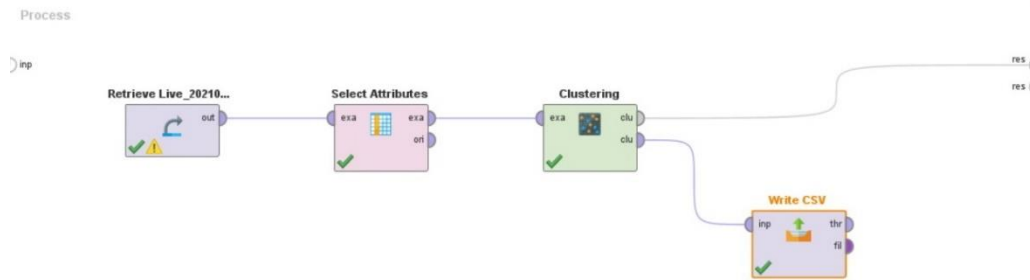
Cluster 3: 101 items

Cluster 4: 405 items

Total number of items: 7050

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER

ต่อไปเพิ่มกล่อง Write CSV ตามรูปที่ 4.2 เพื่อทำการบันทึกผลลัพธ์ลงในเครื่องคอมพิวเตอร์ของตนเอง โดยในช่อง csv file ให้เลือกพื้นที่การจัดเก็บและชื่อไฟล์ที่ต้องการ จากนั้นให้ทำการนำเข้าไฟล์ดังกล่าวเพื่อทำการกระจายตัวของข้อมูล ตามรูปที่ 4.3



รูปที่ 4.2 Write CSV

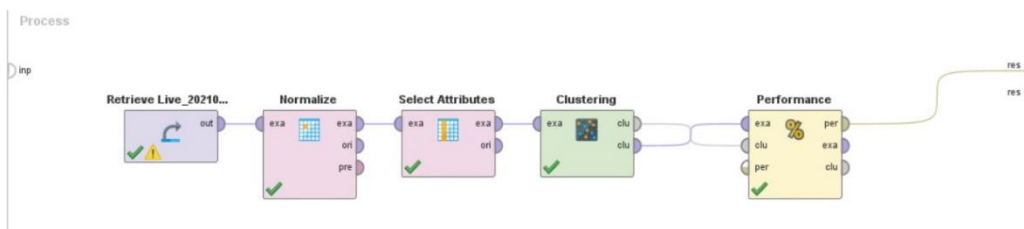


รูปที่ 4.3 Live_20210128 เมื่อ k = 5

ให้เพิ่มกล่อง Cluster Distance Performance ตามรูปที่ 4.5 จะแสดงประสิทธิภาพของกลุ่มข้อมูลตามข้อความด้านล่างนี้

Avg. within centroid distance

Avg. within centroid distance: -212799.306



รูปที่ 4.4 Cluster Distance Performance

สำหรับการจัดกลุ่มข้อมูลด้วยวิธีอื่น ๆ อีกได้แก่ DBSCAN และ Agglomerative Clustering

การจำแนกข้อมูล

Data Classification, kNN, Naïve Bayes, Decision Tree, Random Forest, Perceptron

บทที่ห้านี้ อธิบายแนวคิดของการจำแนกข้อมูล ตัวจำแนกเพื่อนบ้านใกล้ที่สุดเคตตัว นา
สำหรับ อีฟเบส ต้นไม้การตัดสินใจ ป่าสุ่ม เพอร์เซ็ปตรอน การประเมินค่าประสิทธิภาพของตัว
แบบ และ การนำตัวแบบไปจำแนกในเซตข้อมูลจริง

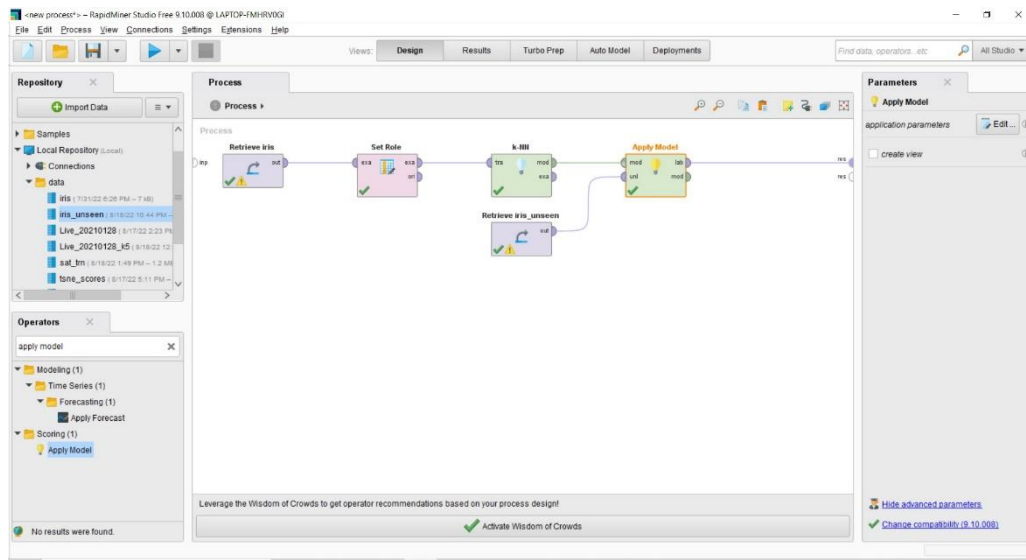
การสร้างตัวแบบและการจำแนกแบบ

ในหน้า Design ให้เชื่อมต่อกล่องตามรูปที่ 5.1 ดังต่อไปนี้

1. data → iris
2. Set Role กำหนด varirty เป็น Label
3. K-NN กำหนดค่า k ที่ต้องการได้
4. Apply Model
5. data → iris_unseen คือข้อมูลที่ต้องการทำนาย

เมื่อกด run จะได้ผลของการทำนายตามรูปที่ 5.2 เป็นการจำแนกว่าข้อมูลแต่ละตัวควรอยู่ใน class
ใด

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER



รูปที่ 5.1 KNN

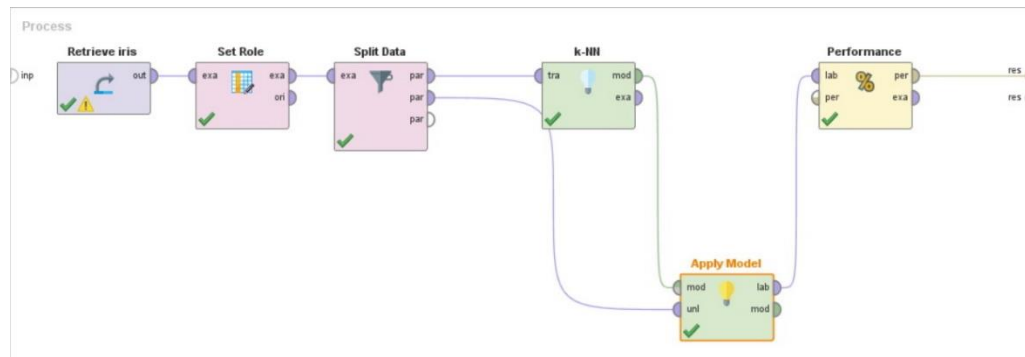
Row No.	prediction(v...	confidence(...	confidence(...	confidence(...	sepal.length	sepal.width	petal.length	petal.width	variety
1	Setosa	1	0	0	5.100	3.500	1.400	0.200	?
2	Setosa	1	0	0	4.900	3	1.400	0.200	?
3	Setosa	1	0	0	4.700	3.200	1.300	0.200	?
4	Setosa	1	0	0	4.600	3.100	1.500	0.200	?
5	Setosa	1	0	0	5	3.600	1.400	0.200	?
6	Versicolor	0	1	0	7	3.200	4.700	1.400	?
7	Versicolor	0	1	0	6.400	3.200	4.500	1.500	?
8	Versicolor	0	1	0	6.900	3.100	4.900	1.500	?
9	Versicolor	0	1	0	5.500	2.300	4	1.300	?
10	Versicolor	0	1	0	6.500	2.800	4.600	1.500	?
11	Virginica	0	0	1	6.300	3.300	6	2.500	?
12	Virginica	0	0	1	5.800	2.700	5.100	1.900	?
13	Virginica	0	0	1	7.100	3	5.900	2.100	?
14	Virginica	0	0	1	6.300	2.900	5.600	1.800	?
15	Virginica	0	0	1	6.500	3	5.800	2.200	?

รูปที่ 5.2 Unseen Data

การประเมินค่าประสิทธิภาพของตัวแบบโดย Split Data

ในหน้า Design ให้เชื่อมต่อกล่องตามรูปที่ 5.3 โดยการเพิ่มกล่อง Split Data กำหนด ratio ที่ต้องการจาก partitions เมื่อกด run จะได้ผลลัพธ์ตามรูปที่ 5.4

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER



รูปที่ 5.3 Split Data

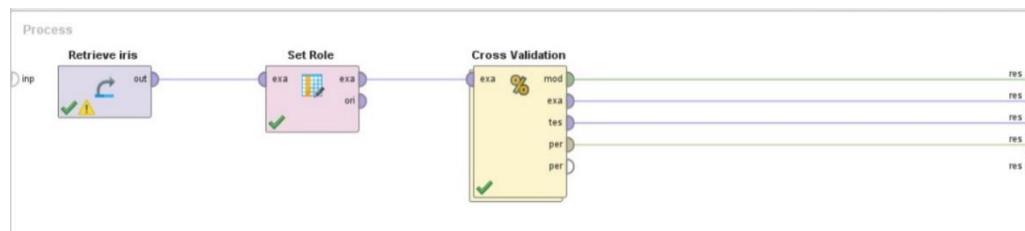
accuracy: 97.78%

	true Setosa	true Versicolor	true Virginica	class precision
pred. Setosa	15	0	0	100.00%
pred. Versicolor	0	15	1	93.75%
pred. Virginica	0	0	14	100.00%
class recall	100.00%	100.00%	93.33%	

รูปที่ 5.4 Performane

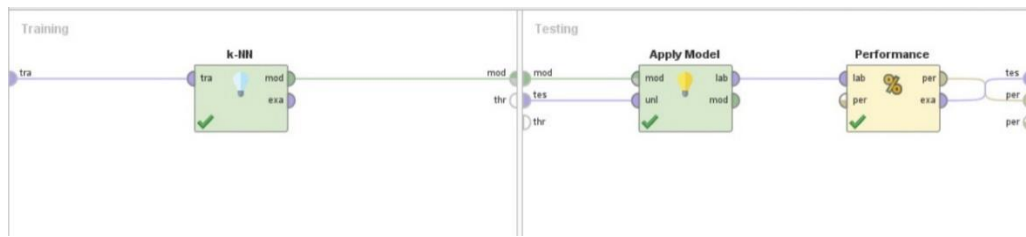
การประเมินค่าประสิทธิภาพของตัวแบบโดย Cross Validation

ในหน้า Design ให้เชื่อมต่อกล่องตามรูปที่ 5.5 โดยการเพิ่มกล่อง Cross Validation กำหนด fold ที่ต้องการจาก ให้ดับเบิลคลิกที่กล่องนี้ แล้วเชื่อมต่อกล่องตามรูป ที่ 5.6 จะได้ผลลัพธ์



รูปที่ 5.5 Cross Validation

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER



รูปที่ 5.6 Cross Validation (inside)

กฎความสัมพันธ์

Association Rules

บทสุดท้ายนี้ อธิบายแนวคิดของกฎความสัมพันธ์ เซตข้อมูลแบบไบนารี เซตข้อมูล
สำหรับ ตระกร้าสินค้าในซูเปอร์มาร์เก็ต การนำตัวแบบไปค้นหาความสัมพันธ์ในเซตข้อมูล
จริง

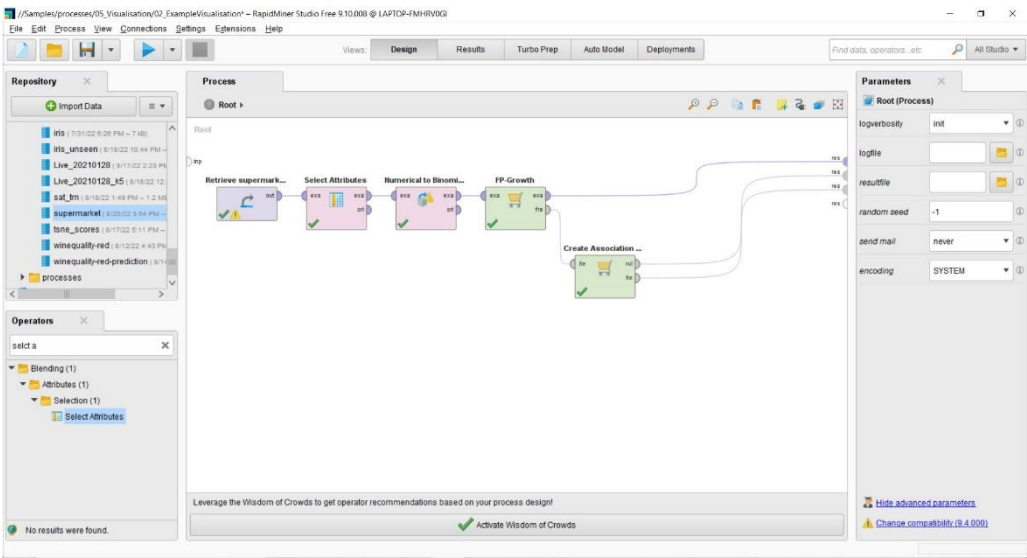
การสร้างตัวแบบและการจำแนกแบบ

เพิ่มเซตข้อมูล supermarket ในหน้า Design ให้เชื่อมต่อกล่องตามรูปที่ 6.1 ดังต่อไปนี้

1. data → supermarket
2. Select Attribute ให้เลือก attribute ทั้งหมด ยกเว้น tid
3. Numerical to Binominal เพื่อแปลงชนิดข้อมูล
4. FP-Growth
5. Create Association Rules เพื่อสร้างกฎความสัมพันธ์

เมื่อกด run จะได้ผลของการทำนายตามรูปที่ 6.2

INTELLIGENCE DATA ANALYSIS WITH RAPIDMINER



รูปที่ 6.1 Association Rules

Size	Support	Item 1	Item 2
2	0.274	Cucumber	Apples
2	0.274	Tomatoes	Cucumber
2	0.262	Cucumber	Sour cream
2	0.260	Cucumber	Celery
2	0.280	Cucumber	Cheese
2	0.263	Cucumber	Oranges
2	0.256	Cucumber	Alaskan King Crab Legs
2	0.259	Cucumber	Eggs
2	0.250	Cucumber	att20
2	0.273	Cucumber	Bananas
2	0.256	Cucumber	Beef
2	0.255	Cucumber	Lettuce
2	0.251	Cucumber	Milk
2	0.267	Cucumber	Squash
2	0.255	Cucumber	Yogurt
2	0.265	Cucumber	Wild Salmon
2	0.254	Cucumber	Mushrooms
2	0.246	Cucumber	Cottage cheese
2	0.280	Apples	Tomatoes

รูปที่ 6.2 Itemsets