

คู่มือการวิเคราะห์ข้อมูลด้วย

Rapid Miner Studio 9

ในเล่มนี้จะแนะนำให้รู้จักการวิเคราะห์ข้อมูลด้วยวิธี

Data mining ซึ่ง สามารถทำได้ง่าย ๆ จาก Software RapidMiner Studio9

<mark>โดยนางสาวนฤชล โรจนบุรานนท์</mark> @Rodjanaburanon



คู่มือการใช้งาน Rapid Miner Studio 9

หัวข้อต่าง ๆ

1. แนะนำ Data Mining และ RapidMiner Studio	1
 เริ่มต้นใช้งาน RapidMiner Studio 9 	4
 องค์ประกอบของ RapidMiner Studio 9 	.6
4. ตัวอย่างการสร้างโมเดล Decision Tree	10
5.การจัดการข้อมูล Data Manipulation	19
6.การทดสอบการทำนาย	36
7. เอกสารอ้างอิง	

การทำเหมืองข้อมูล (Data Mining)

เนื่องด้วยปัจจุบันเป็นยุคที่ข้อมูลสารและสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสาร ที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการ สื่อสารข้อมูลจำนวนมากให้แก่ผู้ ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยน ความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้จำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรม ของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับ ความต้องการใช้งานหรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในซุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ใน งานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมาย ข้อมูล จากเดิม ที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึง การทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

วิวัฒนาการของการทำเหมืองข้อมูล

- ปี 1960 Data Collection คือ การนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือและ ป้องกันการสูญ หายได้เป็นอย่างดี

- ปี 1980 Data Access คือ การนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ต่อกันในข้อมูลเพื่อ ประโยชน์ใน การนำไปวิเคราะห์ และการตัดสินใจอย่างมีคุณภาพ

- ปี 1990 Data Warehouse & Decision Support คือ การรวบรวมข้อมูลมาจัดเก็บลงไปใน ฐานข้อมูลขนาดใหญ่โดยครอบคลุมทุกด้านของ องค์กร เพื่อช่วยสนับสนุนการตัดสินใจ

- ปี 2000 Data Mining คือ การนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยการ สร้างแบบจำลองและความสัมพันธ์ทางสถิติ

ขั้นตอนการทำเหมืองข้อมูล



ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วย ขั้นตอนดังนี้

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการ นำเสนอเพื่อให้เข้าใจ

RapidMiner Studio 9

ซอฟต์แวร์ RapidMiner Studio 7 แรกเริ่มพัฒนาขึ้นจากบริษัทที่ชื่อว่า Rapid-I ในประเทศ เยอรมนีและเมื่อช่วงปลายปี 2013 ที่ผ่านมาได้รับทุนก้อนโตจากนักลงทุนในประเทศสหรัฐอเมริกาจึง เปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทน และย้ายสำนักงานใหญ่มาอยู่ประเทศสหรัฐอเมริกา เราสามารถดาวน์โหลดซอฟต์แวร์ RapidMiner Studio 9 ซึ่งเป็นเวอร์ชันปัจจุบันได้จากเว็บไซต์ https://rapidminer.com

ข้อดีของซอฟต์แวร์ RapidMiner Studio 9 สรุปได้ดังนี้

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

้เมื่อเริ่มต้นใช้งาน RapidMiner Studio 7 จะแสดงหน้าต่างเริ่มต้นซึ่งประกอบด้วย 4 เมนูหลัก ดังนี้

LEARN เป็นหน้าที่รวบรวมและแสดงวิธีการใช้งานของ RapidMiner Studio 7 ซึ่งทำ Link ไปยัง หน้าเว็บที่แสดงการใช้งานในรูปแบบ Document, VDO และมีบทเรียนฝึกหัดให้ทำตาม 3 บท คือ

- Basic สอนพื้นฐานการใช้งาน
- Data Handling สอนการจัดการข้อมูล
- Modeling, Scoring and Validation โดยสอนเกี่ยวกับการสร้างโมเดลเพื่อทำ Prediction การวิเคราะห์ผลและนำผลลัพธ์ข้อมูลเชิงลึกที่คาดการณ์ได้มาใช้จริงและสอนกระบวนการ ยืนยันความถูกต้องของโมเดล

NEW PROCESS สร้างโปรเซสใหม่เพื่อเริ่มการใช้งาน RapidMiner ซึ่งทุกครั้งที่ต้องการสร้าง งาน ใหม่ที่แตกต่างจะต้องสร้างโปรเซสใหม่

OPEN PROCESS เปิดโปรเซสเก่าที่เคยสร้างไว้เพื่อดูหรือแก้ไข โดยโปรเซสที่สร้างไว้แล้วสามารถ Reuse ได้ หรือ ส่งให้คนอื่นได้

เริ่มต้นใช้งาน RapidMiner Studio 9





หลังจากเข้าสู่หน้าโปรแกรม RapidMiner Studio 9 จะให้เรา ลงทะเบียนเข้าสู่โปรแกรมซึ่ง โปรแกรมนี้มีให้ใช้ทั้งใช้งานฟรี และชำระเงินเพื่อใช้งานบาง function หากใครมี account ของโปรแกรม RapidMiner Studio 9อยู่แล้ว สามารถเข้าใช้งานได้โดยใช้ account เดิมได้ หรือจะใส่ Key เพื่อใช้งานโดย เลือกดที่ I already an account or license key.

Create a RapidMiner acc	count 🕼 rapidminer
You'll use your RapidMiner Account to access: the Community forum free cloud storage free cloud storage product news and updates product license information	Account Type Commercial (e.g., business, evaluation, not-for-profit) Educational (e.g., educator, student) Your first name Your last name Your last name I already have an account or license key

หลังจากกด I already an account or license key. เข้ามาแล้วสามารถ เข้าสู่ account โดยการ กรอก E-mail และ Password และเข้าใชงานได้ทันที หรือเลือก ที่ปุ่ม Manually enter license Key . เพื่อ ใส่ license Key ที่ได้จากการซื้อโปรแกรมใส่เพื่อใช้งานได้เช่นกัน

Login to your Ra	apidMiner Account
	<i>(</i>))
	Login with your RapidMiner account to download and install your licenses.
e	example@domain.com
Le la	Password ✔ Remember my password
	Login and Install
	Create a new RapidMiner account
	<u>Manually enter a license key</u> I forgot my email or password

องค์ประกอบของ RapidMiner Studio 9



 Repository เป็นส่วนสำหรับจดการไฟล์ RapidMiner จะจัดการข้อมูลจาก 3 แหล่ง คือ DB , Local (ในเครื่องคอมพิวเตอร์ที่ใช้อยู่), และ Cloud Repository โดยเก็บไฟล์ Data Set และ Process ต่าง ๆ แยกเก็บไว้คนละโฟลเดอร์ 2.Operators เป็นส่วนที่ใช้เก็บตัวโอเปอเรเตอร์ ที่ใช้ในการท างานทั้งหมด ซึ่งจัดเป็นกลุ่ม ๆ โดยกลุ่มที่ ใช้งานคล้ายคลึงกันจะจัดอยู่ในกลุ่มเดียวกัน มี 8 กลุ่ม คือ



- 2.1 Data Access
- 2.2 Blending
- 2.3 Cleansing
- 2.4 Modeling
- 2.5 Scoring
- 2.6 Validation
- 2.7 Utility
- 2.8 Extensions

โอเปอเรเตอร์แต่ละตัวจะประกอบด้วย

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

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

3. Process เป็นหน้าหลักในการท างานในการสร้างโปรเซสสำหรับทำ Machine Learning ของ ซอฟต์แวร์นี้ โดยจะน าโอเปอเรเตอร์มาประกอบเพื่อสร้างโปรเซสขึ้นตามวัตถุประสงค์ของโจทย์ที่ตั้งไว้ 4. Parameters เป็นส่วนสำหรับแสดงพารามิเตอร์ (parameter) ที่เกี่ยวข้องกับแต่ละโอเปอเรเตอร์ เช่น โอเปอเรเตอร์ Read Excel ที่ใช้เพื่ออ่านไฟล์ประเภท Excel จะมีพารามิเตอร์ที่เกี่ยวข้อง เช่น ชื่อและที่ อยู่ของ ไฟล์ Excel เป็นต้น

Parameters	×	
📩 Read Excel		
🌮 Import C	Configuration Wizard	Œ
excel file		
sheet number	1	ſ
🖌 first row as nam	nes	Œ
date format	Enter value 🔻	12 (
Show advanced	i parameters	

แสดงส่วนประกอบพารามิเตอร์ของโอเปอเรเตอร์ Read Excel

5. Help เป็นส่วนช่วยเหลือ ซึ่งจะแสดงรายละเอียดของตัวโอเปอเรเตอร์ที่เลือกใช้งานอยู่ ส่วน ช่วยเหลือของ RapidMiner จะบอกเพียงหน้าที่และรายละเอียดคราว ๆ ของโอเปอเรเตอร์ หากต้องการดู รายละเอียดมากกว่านั้นต้องไปที่ Jum to Tutorai Process ซึ่งจะลิงก์ไปยังเว็บไซต์ที่มีรายละเอียดที่ เกี่ยวกับโอเปอเรเตอร์ที่ใช้อยู่

นอกจากทั้ง 5 ส่วนใหญ่ ๆ ที่ได้อธิบายแล้วยังมีส่วนเมนูด้านบนเพิ่มเติมดังนี้



เมนูสำหรับสร้าง โพรเซสใหม่



เมนูสำหรับการบันทึก โพรเซส หรือบันทึก โพรเซสเป็นชื่อใหม่



เมนูสำหรับสั่งให้ โพรเซสทำงาน

เมนูสำหรับยกเลิกการทำงาน โพรเซส

Views:	Design	Results	Turbo Prep	Auto Model	
	1	2	3	4	

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

ข้อมูลที่มีอยู่ในตารางต่าง ๆ ที่ประกอบไปด้วยแถวและคอลัมน์ ซึ่งจะเรียกในแถวเป็น ตัวอย่าง (Example) ส่วนคอลัมน์เรียก แอตทริบิวต์(Attribute) มีหน้าที่ 3 แบบ คือ

1. ID เป็นแอตทริบิวต์ที่แสดงหมายเลขของข้อมูลหรือ primary key ในฐานข้อมูล

 แอททริบิวต์ทั่วไป (Attribute) เป็นแอตทริบิวต์ปกติที่จะใช้ในการสร้างโมเดลหรือเรียกว่าฟีเจอร์ (feature) หรือตัวแปรต้น(independent variable)

3. Label คือ แอตทริบิวต์ที่เป็นค าตอบที่เราต้องการจะสร้างโมเดลขึ้นมาทำนาย หรือ เรียกว่า คลาส (class) หรือตัวแปรตาม (dependent variable)

Row No.	รหัสแักส์กษา	endgrade	เพศ	กรุ่มเลือด	น้ำหนัก	ส่วนสูง	สถานภาพกา	std_
1	5822041025	2.210	ญ	0	46	159	DRP	ปกติ
2	5122010235	2.490	ญ	0	55	163	END	ปกต์
3	4922010357	2.280	ญ	0	48	162	END	ปกต์
4	5122010225	2.500	ญ	0	45	156	END	ปกต์
5	4922010315	2.820	ญ	0	47	156	END	ปกต์
6	4922010336	2.400	ญ	0	50	170	END	ปกต์
7	5022010174	2.240	ญ	0	49	159	END	ปกต์
8	4922010321	2.310	ญ	0	50	153	END	ปกต์
9	4922010353	2.530	ญ	0	40	153	END	ปกต์
10	5022010208	2.380	ญ	0	48	158	END	ปกติ
11	4922010278	2.370	ญ	0	61	162	END	ปกต์
12	4922010343	2.560	ญ	0	46	147	END	ปกต์
	\longleftrightarrow	\longleftrightarrow	<	Î	Î	Î		\rightarrow
	ID	Label			Attri	bute		

ประเภทของข้อมูลที่เก็บไว้ในแต่ละแอตทริบิวต์ มีดังนี้

- 1. Polynominal เป็นข้อมูลประเภท Category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่าแตกต่างกันมากว่า 2 ค่า
- 2. Binominal เป็นข้อมูลประเภท Category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่าเพียง 2 ค่าเท่านั้น
- 3. Numeric หรือ Integer ข้อมูลประเภทตัวเลข
- 4. Text ข้อมูลประเภทข้อความ

ขั้นตอนการสร้าง Decision Tree

 1. ไปที่ Operators>Data Access>Files>Read ลากโอเปอเรเตอร์ชื่อ Read Excel มาวางที่ Process (สามารถพิมพ์คำว่า Read ในช่อง Search for Operators เพื่อค้นหาโอเปอเรเตอร์ Read Excel ก็ได้) จากนั้น ลากเส้นเชื่อมจากพอร์ตที่ชื่อout (output) ของโอเปอเรเตอร์ Read Excel ไปยังพอร์ตที่ชื่อว่า res(result)

Repository ×	Process
🕒 Import Data 🛛 = 👻	Process
Training Resources (connected)	Process
Samples	Read Excel
Community Samples (connected)	Dinp 🧣 fil 🛃 out
DB 📃 DB	
Local Repository (H61M)	
Cloud Repository (disconnected)	
Operators ×	
Files (16)	
🔻 📇 Read (15)	
tead CSV	
Read Excel	
Read Excel with For	
📩 Read URL	
< Read SPSS ×	
We found "RestAPI Services", "Recommender Extension" and 53 more results in the Marketplace. Show me!	

2. ในส่วนค่าพารามิเตอร์คลิกเลือก Import Configuration Wizard จะปรากฏหน้าต่างดังภาพ ให้ เลือก ข้อมูลซึ่งเป็นไฟล์ประเภท excel จากนั้นคลิก next

	Select the data i	ocation.		
.รับพล			•	← 📑 🋧 🚖 🐸
Bookmarks	File Name	Size	Туре	Last Modified
Last Directory	30-1-2019		File Folder	Jan 30, 2019
	30-1-62		File Folder	Jan 30, 2019
	Jacob 6-2-62		File Folder	Feb 6, 2019
	box dokuwiki-8a269cc015a64b40e4c918699f1e1142		File Folder	Dec 12, 2018
	AllData.xlsx	14 MB	แผ่นงาน Microsoft Excel	Jan 9, 2019
	AllData1.xlsx	575 KB	แผ่นงาน Microsoft Excel	Feb 6, 2019
	AIIData2.xlsx	547 KB	แผ่นงาน Microsoft Excel	Feb 6, 2019
	Student-utf8.xlsx	65 KB	แผ่นงาน Microsoft Excel	Jan 9, 2019
	i weather_nominal.xls	23 KB	แผ่นงาน Microsoft Excel 97	Jan 9, 2019

3. หน้านี้จะเป็นการแสดงการเลือกไฟล์ข้อมูลหรือ Data Set ที่จะนำมาใช้งาน โดยไฟล์ที่เราเลือก Read เป็นไฟล์ excel ดังนั้นไฟล์ที่เราเปิดได้จะเป็นไฟล์ excel

					Sele	ct the cel	lls to imp	ort.					
heet	b ¥	Cell range	A:Q		Select All	🖌 Def	ine header ro	row: 1 +					
	Α	в	С	D	E	F	G	н	1	J	к	L	1
1	รหัสนักศ์	เพศ	กรุ่มเลือด	น้ำหนัก	ส่วนสูง	สถานภา	endgrade	std_edu	แผนก	ๆ ทย	คณิต	วิทย์	1
2	5822041	ល្	0	46	159	DRP	2.21	ปกดิ	คอมพิวเ	4.00	2.35	3.50	
3	5122010	លួ	0	55	163	END	2.49	ปกดิ	การขาย	2.50	1.93	1.83	
4	4922010	លួ	0	48	162	END	2.28	ปกดิ	การขาย	2.50	1.43	2.16	
5	5122010	ญ	0	45	156	END	2.50	ปกดิ	การขาย	2.00	2.37	2.16	
6	4922010	ญ	0	47	156	END	2.82	ปกดิ	การขาย	3.75	1.81	2.33	:
7	4922010	ญ	0	50	170	END	2.40	ปกดิ	การขาย	3.00	1.50	2.50	
8	5022010	ญ	0	49	159	END	2.24	ปกดิ	การขาย	2.25	1.81	2.50	
9	4922010	ญ	0	50	153	END	2.31	ปกดิ	การขาย	2.25	2.00	2.50	·
10	4922010	ญ	0	40	153	END	2.53	ปกดิ	การขาย	3.75	2.18	2.50	:
11	5022010	ญ	0	48	158	END	2.38	ปกดิ	การขาย	2.25	2.75	2.50	:
12	4922010	លួ	0	61	162	END	2.37	บกดิ	การขาย	2.00	1.56	2.66	:
13	4922010	លួ	0	46	147	END	2.56	บกดิ	การขาย	2.50	1.62	2.66	
14	4922010	ญ	0	50	161	END	2.97	ปกดิ	การขาย	2.50	2.50	2.66	
15	5022010	ญ	0	40	159	END	2.64	ปกดิ	การขาย	2.50	2.43	2.75	
16	4922010	សួ	0	51	164	END	2.63	ปกติ	การขาย	3.50	1.50	3.00	
17	4922010	ญ	0	53	160	END	2.50	ปกดิ	การขาย	2.50	1.93	3.00	
18	4922010	ญ	0	52	158	END	2.79	ปกดิ	การขาย	2.75	2.12	3.00	:
19	4922010	ญ	0	42	154	END	2.76	ปกดิ	การขาย	2.75	2.68	3.00	:
20	5022010	กเ	0	47	157	END	2.62	แกล์	การขาย	3.00	2.18	3.25	

4. หน้านี้เป็นหน้าต่างที่เราสามารถเลือกได้ว่าจะนะ Attribute ใดบ้างใช้งานได้บ้างหลังจากเรา เลือก Attribute แล้วก็กด Next

	Format your columns.									
	Replace errors w	ith missing values ①								
	รหัสนักศึกษา 🌣 🔻 real	เพศ์ 🔅 🔻 polynominal	<mark>กรุ๊มเล้อด ‡ ⊽</mark> polynominal	<mark>น้ำหนัก \$</mark> ▼ integer	สำนสูง 🔹 ▼ integer	<mark>สถานภาพกา+‡</mark> ▼ polynominal	endgrade	std_edu a polynominal		
1	5822041025.000	លួ	0	46	159	DRP	2.210	ปกดิ		
2	5122010235.000	លួ	0	55	163	END	2.490	ปกติ		
3	4922010357.000	លួ	0	48	162	END	2.280	ปกติ		
1	5122010225.000	លួ	0	45	156	END	2.500	ปกติ		
5	4922010315.000	លួ	0	47	156	END	2.820	ปกติ		
5	4922010336.000	លួ	0	50	170	END	2.400	ปกติ		
7	5022010174.000	លួ	0	49	159	END	2.240	ปกติ		
3	4922010321.000	ų	0	50	153	END	2.310	ปกดิ		
9	4922010353.000	ų	0	40	153	END	2.530	บกด์		
0	5022010208.000	ญ	0	48	158	END	2.380	ปกดิ		
1	4922010278.000	ญ	0	61	162	END	2.370	ปกติ		
2	4922010343.000	ญ	0	46	147	END	2.560	ปกติ		
3	4922010304.000	លួ	0	50	161	END	2.970	ปกติ		
4	5022010173.000	លួ	0	40	159	END	2.640	ปกติ		
5	4922010318.000	លួ	0	51	164	END	2.630	ปกติ		
6	4922010390.000	លួ	0	53	160	END	2.500	ปกดิ		
7	4922010295.000	លួ	0	52	158	END	2.790	ปกติ		
8	4922010271.000	លួ	0	42	154	END	2.760	ปกติ		
-			•				0.000			

5. หน้าถัดมาให้ทำการเปลี่ยนที่ชื่อ Attribute ให้เป็น label เพื่อใช้ในการทำนายผลของเรา ในที่นี้ เรากำหนด endgrade เป็น label และกำหนด รหัสนักศึกษา เป็น ID เพื่อเป็น Key ในข้อมูล จากนั้นคลิก Finish

	รหลังนักถึกษา ‡ ▼ real id	₩ের্ল \$ ▼ polynominal	<mark>กรุีมเลือด \$ ▼</mark> polynominal	<mark>น้ำหนัก 🔅 ▼</mark> integer	ส่วนสูง 🌲 🔻 integer	สถานภาพกา 🏶 🔻 polynominal	endgrade ✿ ▼ real label	std_edu \$
1	5822041025.000	លួ	0	46	159	DRP	2.210	ปกติ '
2	5122010235.000	សួ	0	55	163	END	2.490	ปกติ
3	4922010357.000	សួ	0	48	162	END	2.280	ปกติ
4	5122010225.000	សួ	0	45	156	END	2.500	ปกติ
5	4922010315.000	សួ	0	47	156	END	2.820	ปกติ
6	4922010336.000	សួ	0	50	170	END	2.400	ปกติ
7	5022010174.000	សួ	0	49	159	END	2.240	ปกติ
8	4922010321.000	ญ	0	50	153	END	2.310	ปกติ
9	4922010353.000	ญ	0	40	153	END	2.530	ปกติ
10	5022010208.000	ญ	0	48	158	END	2.380	ปกติ
11	4922010278.000	លួ	0	61	162	END	2.370	ปกติ
12	4922010343.000	លួ	0	46	147	END	2.560	ปกติ
**			<u>^</u>	50		515	0.070	
	\longleftrightarrow						\longleftrightarrow	



Label

6. หลังจากเลือไฟล์และกด Finish จะขึ้นหน้าต่างการทำงาน หลังจากนั้นลากเส้นเชื่อมการทำงาน ให้ได้ดังรูป



7. หลังจากเราลากเส้นเชื่อมการทำงานแล้ว ให้เรา คลิก Run Process จะปรากฏหน้าต่าง ผลลัพธ์ของการทำงาน

🧶 <new process*=""> – RapidMiner Studio Educational 9.2.000 @ H61M-PC</new>					
<u>File Edit Process View connections coud Settings Extensions Help</u>					
	Views:	Design	Results	Turbo Prep	Auto Mod
Repository X Process					
Import Data E Process					100%
Training Resources (connected) Samples Community Samples (connected) DB Local Repository (H81M) Cloud Repository (disconnected) Operators re					

8. เมื่อเราสั่งโปรแกรมทำงานแล้ว มีแอตทริบิวต์พิเศษจำนวน 1 แอตทริบิวต์ (ตัวที่เราเลือกเป็น labelจะมีแถบสีเป็นสีเขียว) และแอตทริบิวต์ทั่ว ๆ ไป ที่ใช้สำหรับสร้างโมเดล ดังภาพ

	Open in 📑	Turbo Prep	Auto Model		B	Filter (5,4	22 / 5,422 examples):	all	
Data	Row No.	รหสันกลึกษา	endgrade	เพศ	กรุ่มเลือด	น้ำหนัก	ส่วนสูง	สถานภาพกา	st
	1	5822041025	2.210	លួ	0	46	159	DRP	IJ٢
Σ	2 C	5122010235	2.490	លួ	0	55	163	END	ปร
statistics	3	4922010357	2.280	ญ	0	48	162	END	IJf
isualizations	4	5122010225	2.500	លួ	0	45	156	END	IJf
	5	4922010315	2.820	ญ	0	47	156	END	IJf
	6	4922010336	2.400	ญ	0	50	170	END	IJf
	7	5022010174	2.240	ญ	0	49	159	END	IJf
	8	4922010321	2.310	ญ	0	50	153	END	IJf
	9	4922010353	2.530	ญ	0	40	153	END	IJf
notations	10	5022010208	2.380	លួ	0	48	158	END	IJf
	11	4922010278	2.370	ญ	0	61	162	END	IJf
	12	4922010343	2.560	លួ	0	46	147	END	IJf
	13	4922010304	2.970	ญ	0	50	161	END	IJf
	14	5022010173	2.640	លួ	0	40	159	END	IJf
	15	4922010318	2.630	លួ	0	51	164	END	IJf
	16	4922010390	2.500	លួ	0	53	160	END	IJf
	17	4922010295	2 790	กเ	0	52	158	END	115

A แสดงจำนวนตัวอย่างและแอตทริบิวต์ที่ปรากฏในข้อมูลซึ่งในไฟล์ตัวอย่างนี้มีจำนวน 5,422 ตัวอย่าง 1 แอตทริบิวต์ประเภทลาเบล และ 1 แอตทริบิวต์ประเภท ไอดี และ 15 แอตทริบิวต์ทั่วไป B ส่วนของการกรองข้อมูล (filter) ซึ่งมีให้เลือกได้ว่าจะดูข้อมูลทั้งหมดหรือข้อมูลที่มีความ

ผิดพลาด (missing_attributes) อยู่

C ในส่วนของตารางเราสามารถคลิกเลือกที่ชื่อแอตทริบิวต์เพื่อทำการเรียงลำดับข้อมูลได้ โดย ตารางข้อมูลจะแบ่งแอตทริบิวต์ออกเป็น 3 แบบคือ

- แอตทริบิวต์ที่เป็นลาเบลแสดงด้วยคอลัมน์สีเขียว
- แอตทริบิวต์ที่เป็นไอดีแสดงคอลัมน์สีฟ้า
- แอตทริบิวต์ทั่วไปแสดงด้วยคอลัมน์ที่เป็นสีเทา

Result History	ExampleSet (Read Excel)	×						
	Name	⊦• Туре	Missing	. Filter (17 / 17 attributes):	Search for Attributes			
Data	^{เส} รหัสนักศึกษา	Real	0	Min 4722010012	^{Мах} лах			
Statistics	Label endgrade	Real	0	Min 1.280	Max 4			
	🗸 เพศ	Polynominal	0	Least ಶ (1227)	^{Most} ญ (4195)			
Visualizations	💙 กรุ์มเลือด	Polynominal	27	Least - (64)	^{Most} B (2088)			
	💙 น้ำหนัก	Integer	0	Min 1	Max 701			
Annotations	💙 ส่วนสูง	Integer	0	Min 43	Мах 990			
	✓ สถานภาพการเรียน	Polynominal	143	DRP (5)	Most END (5209)			
	❤ std_edu	Polynominal	0	_{Least} เทียบโอน (1)	_{Most} ปกดิ (4920)			
	<	Polynominal	0	Least เทควีนวีลย []ทัศน์ (1)	Most การบับเช้≀(1157)			
	Showing attributes 1 - 17			Examples: 5,422 Special Attributes: 2 Regular Attributes: 15				

9. คลิกเลือก Statistics ด้านซ้ายมือ เพื่อแสดงค่าสรุปทางสถิติของแอตทริบิวต์ต่าง ๆ โดยจะแสดง ชื่อประเภทของข้อมูลที่เก็บอยู่ กราฟแสดงค่าความถี่ของข้อมูลในแต่ละแอตทริบิวต์ดังภาพ

ile Edit Proce:	ss <u>View Connections</u> Clou <u>d</u> <u>S</u> ettings	Eglensi	ons <u>H</u> elp	Views: Des	ign Results Tur	bo Prep Auto Mode	1		Find data, operationsetc	🔎 All Studio
Result History	ExampleSet (Read Excel)	×						Repository		
-	Name	Н	Туре	Missing	. Filter (17 / 17 attributes):	Search for Attributes	7 -		G Import Data	=
Data	💙 รห์สหักถึกษา		Real	0	4722010012	5827011042	^	 Training Resou Samples 		
Σ	v Label endgrade		Real	o	^{Min} 1.280			Community Sar DB Local Repositor	ry (retrut)	
Statistics	🗸 DAU		Polynominal	0	Luat s' (1227)	^{Most} กู: (4195)		Cloud Reposito	NY (disconnected)	
A lisualizations	💙 กรุ๊บเลือด		Polynominal	27	- (64)	Most B (2088)				
					Min	Max 204				

10. คลิกเลือกมุมมอง Design ต่อไปเราจะทำการสร้างโมเดล Decision Tree โดยการเลือก โอเปอเรเตอร์ Decision Tree จากส่วนของ Operators โดยการพิมพ์ตรงช่องค้นหา โดยพิมพ์คำว่า Decision กดปุ่ม Enter ก็จะปรากฏโอเปอเรเตอร์ Decision Tree ขึ้นมา หรือจะทำการเลือกจากหมวด Modeling >> Classification and Regression >> Tree Induction



11. ลากโอเปอเรเตอร์ Decision Tree มาวางในส่วนของ Process ตรงเส้นที่เชื่อมต่อเดิมที่ โอเปอเรเตอร์ Read Excel ลากไว้ (โปรแกรมจะทำการเชื่อมโอเปอเรเตอร์ทั้งสองตัวทันทีจากพอร์ต out ของโอเปอเรเตอร์ Read Excel ไปยังพอร์ต tra (training) ของโอเปอเรเตอร์ Decision Tree เพื่อเป็นการส่ง ข้อมูลไปสร้างโมเดล



12. ลากเส้นเชื่อมจากพอร์ต mod (model) และพอร์ต exa (example) ของโอเปอเรเตอร์ Decision Tree ไปยังพอร์ต res (result) ทั้งสองพอร์ต เพื่อไปแสดงในส่วนของหน้าจอผลลัพธ์โดยพอร์ต mod จะทำการส่งโมเดล Decision Tree ที่สร้างออกไปแสดงในรูปต้นไม้ และพอร์ต exa จะส่งข้อมูลที่ import เข้ามาไปแสดงในรูปแบบตาราง

@Rodjanaburanon Nam

13. จากนั้นคลิก Run Process จะได้รูปโมเดลต้นไม้ ซึ่งโมเดลต้นไม้ที่สร้างได้มีส่วนประกอบ สำคัญ 3 ส่วน คือ

- ในโมเดล Decision Tree จะมีโหนดต่าง ๆ 2 ประเภท คือ
 - ๑ โหนดที่เป็นแอตทริบิวต์แสดงด้วยรูปสี่เหลี่ยมที่มีมุมโค้ง
 - O โหนดลาเบลแสดงด้วยรูปสี่เหลี่ยมที่มีกราฟแสดงสีต่าง ๆ อยู่ด้วย ในตัวอย่าง

นี้มีหลาย label แต่หากมีการกำหนดเกณฑ์มาตรฐานของคะแนนในตัวอย่างไล่เป็นระดับตามเกรด ตัว label ก็จะมีคำตอบตามลำกับที่เราตั้งค่าไว้ จะมีกราฟสีน้ำเงิน

- ส่วนของ Zoom ใช้สำหรับย่อขยายรูปโมเดล
- ส่วนของ Mode จะใช้สำหรับปรับโหมดของการใช้งานเมาส์



14. ในหน้าต่าง Description จะเป็นโค๊ดข้อความที่เราสามารถนำมาเขียนโปรแกรมเพื่อใช้ในการทำนายได้

Result History	🚦 ExampleSet (Read Excel) 🛛 🛛 🖓 RegressionTree (Decision Tree) 🛛 🗙
Graph	RegressionTree ด่างประเทศ > 3.015 กษ์ต > 2.645
Description	(ต่างประเทศ > 3.505 สุขศึกษา > 3.125 ต่างประเทศ > 3.775 "ทย > 3.580 "ถณิต > 3.840 ต่างประเทศ > 3.940
Annotations	可付新済前者のコントン5124510043.500: 3.976 {count=10} 可付新済前者のコントン5124510043.500: 3.976 {count=10} 可介加第二の第二の第二の第二の第二の第二の第二の第二の第二の第二の第二の第二の第二の第
	חารงาน > 3.085: 3.575 (count=6) การงาน ≤ 3.085: 3.785 (count=46) "ทบุ≤ 3.580 ต่างประเทศ > 3.980 пารงาน > 3.160

การจัดการข้อมูล (Data Manipulation)

หลังจากที่เราลองทำ Decision Tree แล้วเราจำนำโมเลที่เราสร้างมาทำการทำนาย แต่เราจะทำนายไม่ได้หากยังมี Missng รวมถึงเราต้องเลือกใช้เพียงแอตทริบิวต์บางตัวที่จำมาใช่ในการทำนายเท่านั้น ในที่นี้ยังมี แอตทริบิวต์บางตัว ที่ติด Missng อยู่ดังนั้นเราจะนำตัวที่ติด Missng ออกเนื่องจากกว่าตัวที่เรานำออกนั้นไม่ได้มีผลต่อการทำนาย แต่หากว่าแอตทริบิวต์ที่ ติด Missng มีผลต่อการทำนายเราอาจะใช้เป็นค่าเฉลี่ย หรือการแทนค่าเข้าไปแทน

	Name	۲٦	Type	missing	Statistics	
~	Label endgrade		Real	0	Min 1.280	Max 4
~	รหัสนักศึกษา		Real	0	Min 4722010012	Max 5827011042
~	เพศ		Polynominal	0	Least ປ (1227)	^{Most} ญ (4195)
~	กรุ่มเลือด		Polynominal	27	Least - (64)	^{Most} B (2088)
~	น้ำหนัก		Integer	0	Min 1	^{Max} 701
~	ส่วนสูง		Integer	0	Min 43	Max 990
~	สถานภาพการเรียน		Polynominal	143	Least DRP (5)	Most END (5209)
~	std_edu		Polynominal	0	Least เทียบโอน (1)	^{Most} ปกติ (4920)
~	แผนก		Polynominal	0	_{Least} เทคโนโลย [] ทัศน์ (1)	^{Most} การบัญชี (1157)
~	ๆ ทย		Real	0	Min O	Мах 4

 พิมพ์ค้นหาคำว่า Select Attributes แล้วลาก Operators มาวางที่หน่างการทำงาน และลากเส้นเชื่อม ระหว่าง Out ของ Excel มาที่ exa ของ Select Attributes

Repository ×	Proces	s		
🕒 Import Data 🛛 🗉	• Pro	cess >		
Training Resources (connected)	Process			
Samples		Read Excel	Select Attributes	
Community Samples (connected)	Dinp	🚽 fil 🛃 out 📃	exa exa)
🕨 📒 DB	Dinp	✓	ori)
🕨 🌉 Local Repository (H61M)				
Cloud Repository (disconnected)				
att	×			
Blending (9)	^			
🕶 📇 Attributes (9)				
🔻 📇 Names & Roles (2)				
📝 Rename				
🚺 Set Role				
Selection (4)				
Select Attributes				
📗 Remove Attribute Range	• ~			
<	>			
 We found "Web Automation Extension", "Weka Extension" and emore results in the Marketplace. St 	6 now			

2. เลือกที่ Parameters ของ Select Attributes แล้วเลือก Attributes filter type เป็น subset หลังจากนั้น เลือก Select Attributes

liews: Design	Results	Turbo Prep	Auto Model			Find data, operatorsel	c 🔎 All Studio	•
			100% 🔎	e بر م	i 💣 🖸	Parameters ×		
Select Attributes	`				res	attribute filter type attributes	subset	1
ori					res	invert selection		1
							outes	0

3.หลังจากที่เราเลือก Select Attributes จะมีหน้าต่างขึ้นมาให้เราเลือก Attributes ที่เราต้องการที่จำใช้ใน การทำนาย เมื่อเลือกข้อมูลที่ต้องการได้กด Apply และ สั่งรันโปรแกรมให้โปรแกรมทำงาน



4. หลังจากรันการทำงานของโปรแกรมแล้ว จะแสดงเฉพาะ Attributes ที่เราเลือกเท่านั้นเพื่อนำมาใช้งาน หากว่าเราต้องการทำนายแล้วยังไม่ได้กำหนด label สามารถกำหนด label โดยกาพิมพ์ค้นหา Set Row ได้ในช่อง Operation

Result History	E	xampleSet (Sele	ect Attributes)	×									
	Open in Turbo Prep Auto Model Filter (5,422 / 5,422 examples)											all 🔻	
Data	Row No.	endgrade	รหสันกัดศึกษา	แผนก	ใ ทย	คณิต	วิทย์	สังคม	สุขศึกษา	ศึลปะ	การงาน	ด่างประเทศ	
	1	2.210	5822041025	คอมพิวเตอร์ธุ	4	2.350	3.500	3.500	1	4	0	2.040	
Σ	2	2.490	5122010235	การขาย	2.500	1.930	1.830	2.500	2.250	3.330	0	2.540	
Statistics	3	2.280	4922010357	การขาย	2.500	1.430	2.160	1	2.500	3.500	0	2.170	
S	4	2.500	5122010225	การขาย	2	2.370	2.160	2.750	2.500	3.500	0	2.460	
	5	2.820	4922010315	การขาย	3.750	1.810	2.330	2	3	4	0	2.850	
Visualizations	6	2.400	4922010336	การขาย	3	1.500	2.500	1	2.500	3.330	0	2.360	
	7	2.240	5022010174	การขาย	2.250	1.810	2.500	1.330	1.750	3.160	0	2.320	
	8	2.310	4922010321	การขาย	2.250	2	2.500	1	2	3.160	0	2.270	
	9	2.530	4922010353	การขาย	3.750	2.180	2.500	2	1.500	3.500	0	2.360	
Annotations	10	2.380	5022010208	การขาย	2.250	2.750	2.500	2.160	1.250	3.500	0	2.310	
	11	2.370	4922010278	การขาย	2	1.560	2.660	2.750	2	3.830	0	2.230	
	12	2.560	4922010343	การขาย	2.500	1.620	2.660	1.750	2.250	3.500	0	2.560	
	13	2.970	4922010304	การขาย	2.500	2.500	2.660	3.500	3.250	3.830	0	3.050	
	14	2.640	5022010173	การขาย	2.500	2.430	2.750	1.660	1.250	3.500	0	2.780	
	15	2.630	4922010318	การขาย	3.500	1.500	3	1	2	3.330	0	2.620	
	16	2.500	4922010390	การขาย	2.500	1.930	3	1	1.750	3.660	0	2.530	
	17	2 790	4922010295	การขาย	2 750	2 120	3	2 500	2	3 500	0	2,800	

- 5. หลังจากได้ข้อมูลเบื้อต้นแล้ว ให้กำหนด ค่าของตัวเลขเป็นข้อความแบ่งระดับ โดยจะแบ่งระดับ เป็น
 - ดีมาก = 4 - ปานกลาง = 3 - ต่ำ =2

การตั้งค่าเพื่อให้ง่ายต่อการใช้งานโดยการ คลิกขวาเลือก Insert Operator >> Cleansing >> Binning >> Discretize by User Specification

Process Process	Parameters ×	Parameters × Discretize (Discretize by User Specification)		
- Tocess		attribute filter type	subset 🔻 🗊	
Read Excel Select Attributes Discretize	res	attributes	🖹 Select Attributes 🛈	
on pre		Invent selection	Û	
		include special attr	ibutes ①	
		classes	Edit List (2)	
		Show advanced pa	arameters	
		🗸 Change compatibi	lity (9.2.000)	

6. เมื่อได้ Operator แล้วเลือกตัว Operator แล้วเลือก Attributes filter type เป็น subset หลังจากนั้น เลือก Select Attributes เพื่อเลือก Attributes ที่ต้องการแทนที่ค่า เมื่อเลือก Attributes ที่ต้องการเสร็จแล้วกด apply



7. หลังจากนั้นกำหนดเกณฑ์ที่ต้องการเทียบกับคะแนน เพื่อแปลงเป็นระดับที่เราตั้งไว้ ในช่อง Classes เลือก Edit List

Process		Parameters X
Process >	100% 🔑 🔎 🔎 🛃 🧰 🔝	Discretize (Discretize by User Specification)
Process		attribute filter type subset 💌 🛈
P. fi to find the second secon	res (attributes 📄 🕄
up pre		invert selection
		include special attributes
		classes 🛛 🔂 Edit List (2) 🛈
		Show advanced parameters
		Change compatibility (9.2.000)

8. เมื่อกดเข้าไปจะมีหน้าต่าง ให้เรากำหนดเกณฑ์ เมื่อเรากำหนดเกณฑ์แล้วกด Apply

- P	
class names	upper limit
มาก	4.0
านกลาง	3.0
h	2.0

9. หลังจากกด apply แล้ว กด play ให้โปรแกรมทำงาน หลังจากทำงานแล้วเกรดแต่ละตัวที่เป็นตัวเลขจะ โดนแปลงเป็นข้อความยกเว้นซ่อง endgrade เพราะเรากำหนดเป็น label เราต้องแก้ไขทีหลัง

Row No.	endgrade	ใ ทย	กณ์ด	วิทย์	ส์งิคม	สุขศึกษา	ศึลปะ	การงาน	ด่างประเทศ	รหัสแักศึกษา	แผนก
1	2.210	ดีมาก	ปานกลาง	ดีมาก	ดีมาก	ด่า	ดีมาก	ดำ	ปานกลาง	5822041025	คอมพิวเดอร์ธุ
2	2.490	ปานกลาง	ดำ	ดำ	ปานกลาง	ปานกลาง	ดีมาก	ด่า	ปานกลาง	5122010235	การขาย
3	2.280	ปานกลาง	ด่า	ปานกลาง	ดำ	ปานกลาง	ดึมาก	ด่า	ปานกลาง	4922010357	การขาย
4	2.500	ดำ	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดำ	ปานกลาง	5122010225	การขาย
5	2.820	ดีมาก	ด่า	ปานกลาง	ด่า	ปานกลาง	ดึมาก	ด่า	ปานกลาง	4922010315	การขาย
6	2.400	ปานกลาง	ด่า	ปานกลาง	ดำ	ปานกลาง	ดึมาก	ด่า	ปานกลาง	4922010336	การขาย
7	2.240	ปานกลาง	ด่า	ปานกลาง	ด่า	ดำ	ดีมาก	ด่า	ปานกลาง	5022010174	การขาย
8	2.310	ปานกลาง	ดำ	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง	4922010321	การขาย
9	2.530	ดีมาก	ปานกลาง	ปานกลาง	ดำ	ดำ	ดึมาก	ด่า	ปานกลาง	4922010353	การขาย
10	2.380	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ด่า	ปานกลาง	5022010208	การขาย
11	2.370	ต่ำ	ด่า	ปานกลาง	ปานกลาง	ด่า	ด์มาก	ด่า	ปานกลาง	4922010278	การขาย
12	2.560	ปานกลาง	ดำ	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	ปานกลาง	4922010343	การขาย
13	2.970	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดีมาก	ดีมาก	ด่า	ดีมาก	4922010304	การขาย
14	2.640	ปานกลาง	ปานกลาง	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง	5022010173	การขาย
15	2.630	ดีมาก	ด่า	ปานกลาง	ดำ	ดำ	ดึมาก	ด่า	ปานกลาง	4922010318	การขาย
16	2.500	ปานกลาง	ดำ	ปานกลาง	ด่า	ด่า	ดีมาก	ดำ	ปานกลาง	4922010390	การขาย
17	2.790	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ด่า	ด์มาก	ด่า	ปานกลาง	4922010295	การขาย

10. หลังจากนั้นเซฟไฟล์ เป็น CSV. เพื่อที่จะสามารถนำมาใช้แล้วนำมาปรับแก้ได้ โดยพิมพ์ค้นหาที่ operator ว่า write Csv.

Operators ×
write Csv X
🕶 🛅 Data Access (1)
🔻 📩 Files (1)
🕶 🔚 Write (1)
🍰 Write CSV

11. น้ำ Operator write Csv. มาวางที่หน้าจอการทำงาน แล้วลากเส้นเชื่อมกัน ตามรูป

	Process		Parameters ×	
	Process >	100% 🔑 🔑 📴 🔁	Process	
)	Process		logverbosity	init 🔻 🗊
) 9) Di	inp fil	dExcel Select Attributes Discretize Write CSV	logfile	D

12. หลังจากลากเส้นแล้วเลือก Operator write Csv. แล้วเลือกที่ Save ไฟล์

	Parameters ×
100% 🔎 🔎 📮 🥃 💣 ⊡	😩 Write CSV
	csv file
Write CSV	A Please set this parameter
inp thr res	column separator ; D
	✓ write attribute names ①
	✓ quote nominal values
`S;	append to file
	Show advanced parameters
	 Change compatibility (9.2.000)

13. เมื่อกดเข้ามาที่ Csv file แล้วจะขึ้นหน้าต่างให้เราเลือก Save file ลงที่ตำแหน่งที่เราต้องการ หลังจาก นั้นตั้งชื่อไฟล์และจากนั้นกด Open เมื่อเสร็จแล้วกด Run โปรแกรมให้โปรแกรมทำงาน ไฟล์ Csv. ก็จะถูกบันทึก

📃 Desktop			▼ ←	i 🛧 🚖 📑 🗔
Bookmarks	File Name	Size	Туре	Last Modifi
🛨 Last Directory	10081231		File Folder	Jan 4, 2019
	🐌 10190101		File Folder	Jan 4, 2019
	10390210		File Folder	Feb 10, 2019
	10790216		File Folder	Feb 16, 2019
	1 Computer		File Folder	
	DATA - Shortcut		File Folder	Sep 13, 2017
	H61M		File Folder	Feb 25, 2019
	🥽 Libraries		File Folder	
	👽 Network		File Folder	
	🐌 New folder		File Folder	Feb 24, 2019
	New folder444		File Folder	Sep 5, 2018
	PNP Nakonbpathom-20190203T11102		File Folder	Feb 3, 2019
	1 SAO		File Folder	Dec 22, 2018

14. หลังจากนั้น เปิด file ขึ้นมาใหม่ ในช่อง Operater พิมพ์ read Csv. เพื่อดึงไฟล์ Csv. ขึ้นมาใช้งาน ดับเบิลคลิกที่ตัว Read Csv.

itopository	1100000	
🕒 Import Data 🛛 🗉 👻	Process	
Training Resources (connected)	Process	
• 🔁 Samples		Read CSV
 Community Samples (connected)) inp	🖌 fil 🛃 out
DB		
Local Repository (H61M)		
Cloud Repository (disconnected)		
<		
Operators ×		
read Csv 🗙		
r 🛅 Data Access (1)		
🔻 📇 Files (1)		
🔻 🔚 Read (1)		
📩 Read CSV		

15. โปรแกรมจะแสดงหน้าต่างให้เราเลือกไฟล์ Csv. ที่เราบันทึไว้ก่อนหน้านี้เพื่อนำมาใช้งาน

ktop				
			•	← 85 <mark>☆</mark> ☆ 8
Bookmarks File M	lame	Size	Туре	Last Modified
ast Directory 🔋 🔒 SA	0		File Folder	Dec 22, 2018
🔊 So	ny - Shortcut		File Folder	Aug 14, 2018
🍑 Ur	titled Export		File Folder	Jan 27, 2019
🧾 W/	4V1		File Folder	Jan 2, 2019
🛃 ครู	รราช -เว็บ		File Folder	Oct 26, 2017
🚺 งา:	แใหม่		File Folder	Dec 20, 2018
길 ดร	มรรมด		File Folder	Jan 28, 2018
길 นิย	าย PDF		File Folder	Dec 22, 2018
길 ນ.ເ	าร์		File Folder	Feb 1, 2019
🗋 ປ.ເ	าร์ - Shortcut		File Folder	Aug 5, 2018
🍶 ນດ	รีเทอม 2		File Folder	Jan 27, 2019
Ja 52	ນນ ວ. ຮາ້ອງ		File Folder	Dec 15, 2018
🔊 ຽນ	น้ำ - Shortcut		File Folder	Oct 3, 2017
🍌 ກໍລັ	b		File Folder	Dec 15, 2018
🌙 สต์	กเกอร์ใลน์		File Folder	Nov 11, 2018
🍌 ส่ง	งาน		File Folder	Feb 10, 2019
) ии	ังสือ		File Folder	Dec 19, 2018
길 ອ.ເ	ล่าย		File Folder	Jan 10, 2019
🍶 ວ."	โอภาส		File Folder	Jan 10, 2019
a IN	ลง - Shortcut		File Folder	Oct 3, 2017
🍶 ເວົ້າ	แช้ญชาน		File Folder	Dec 5, 2017
🔒 ໃນ	ลา		File Folder	Nov 24, 2018
3 55	5.csv	532 KB	แพ้ม Microsoft Excel ที่ใช้จุล	Feb 27, 2019

16. หน้าต่างนี้จะแสดง Attributes ที่เราสามารถเลือกที่จะนำมาใช้ได้

					Specify	your data	format				
🗸 н	eader Row			1 ‡	File Encoding	x-wir	ndows-874	▼ Vse	Quotes		
tart F	Row			1.1	Escape Character			Trim	Lines		
olun	nn Separator	Semi	colon ";"	•	Decimal Characte	ır .		🗸 Skip	Comments	#	
1	ใทย	คณ์ด	วิทย์	สังคม	สุขศึกษา	ศึลปะ	การงาน	ต่างประเทศ	รหัสนักศึก	แผนก	endgrade
2	ดีมาก	ปานกลาง	ดีมาก	ดีมาก	ด่ำ	ดีมาก	ต่ำ	ปานกลาง	5.8220410	คอมพิวเตอ	2.21
3	ปานกลาง	ด่า	ด่า	ปานกลา	าง ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	5.1220102	การขาย	2.49
4	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.28
5	ด่า	ปานกลาง	ปานกลาง	ปานกลา	าง ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	5.1220102	การขาย	2.5
6	ดีมาก	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.82
7	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.4
8	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ต่ำ	ดีมาก	ต่า	ปานกลาง	5.0220101	การขาย	2.24
9	ปานกลาง	ด่า	ปาน <mark>กลา</mark> ง	ต่า	ต่ำ	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.31
10	ดีมาก	ปานกลาง	ปานกลาง	ด่า	ต่ำ	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.53
11	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลา	าง ต่ำ	ดีมาก	ต่า	ปานกลาง	5.0220102	การขาย	2.38
12	ด่า	ด่า	ปานกลาง	ปานกลา	าง ต่ำ	ดีมาก	ต่า	ปานกลาง	4.9220102	การขาย	2.37
13	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.56
14	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดีมาก	ดีมาก	ต่ำ	ดีมาก	4.9220103	การขาย	2.97
15	ปานกลาง	ปานกลาง	ปานกลาง	ต่ำ	ด่า	ดีมาก	ต่ำ	ปานกลาง	5.0220101	การขาย	2.64
16	ดีมาก	ต่ำ	ปานกลาง	ต่ำ	ด่า	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.63
17	ปานกลาง	ต่า	ปานกลาง	ต่ำ	ต่ำ	ดีมาก	ต่ำ	ปานกลาง	4.9220103	การขาย	2.5
18	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลา	าง ต่ำ	ดีมาก	ด่า	ปานกลาง	4.9220102	การขาย	2.79

17. หน้าต่างนี้จะเป็นหน้าต่างคุณสมบัติของ Attributes แต่เรายังไม่ต้องกำหนดเพราะเราต้องการเปลี่ยน ค่าของ endgrade ให้เป็นไปตามเกณฑ์ที่เรากำหนดไว้ หลังจากนั้นกด finish

Format your columns.								
Date format Enter value.								
	ใทย 🔅 🕶 polynominal	ពณ៌ព 🔅 🔻 polynominal	วิทย์ 🔅 🕶 polynominal	ส์งิคม \$ ▼ polynominal	สุขศึกษา ¢ ⊽ polynominal	ที่ลปะ 💠 🔻	<mark>การงาน</mark> ¢ ▼ polynominal	<mark>ด่างประเทศ</mark> polynominal
1	ดีมาก	ปานกลาง	ดีมาก	ดีมาก	ด่า	ดีมาก	ดำ	ปานกลาง
2	ปานกลาง	ดำ	ดำ	ปานกลาง	ปานกลาง	ดีมาก	ด่ำ	ปานกลาง
3	ปานกลาง	ด่า	ปานกลาง	ด่ำ	ปานกลาง	ดีมาก	ด่า	ปานกลาง
4	ด่ำ	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ด่า	ปานกลาง
5	ดีมาก	ด่า	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	ปานกลาง
5	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ด่า	ปานกลาง
7	ปานกลาง	ด่า	ปานกล้าง	ต่ำ	ด่า	ดีมาก	ดำ	ปานกลาง
3	ปานกลาง	ด่า	ปานกลาง	ด่า	ดำ	ดีมาก	ดำ	ปานกลาง
,	ดีมาก	ปานกลาง	ปานกลาง	ต่ำ	ต่ำ	ดีมาก	ดำ	ปานกลาง
0	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ดำ	ปานกลาง
1	ต่ำ	ด่า	ปานกลาง	ปานกลาง	ด่า	ดีมาก	ดำ	ปานกลาง
2	ปานกลาง	ด่า	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	ปานกลาง
3	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดีมาก	ดีมาก	ดำ	ดีมาก
4	ปานกลาง	ปานกลาง	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง
5	ดีมาก	ด่า	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง
6	ปานกลาง	ดำ	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง
7	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ด่า	ดีมาก	ด่า	ปานกลาง
	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ด่า	ปานกลาง

18. เมื่อเข้าสู้หน้าต่าง ลากเส้นเชื่อมเพียงเส้นเดียวดังรูป แล้วโปรแกรมลองกด Run Program หน้าต่างจะ แสดงรายการเช่นเดียวกับไฟล์ excel ที่เราบันทึกมาก่อนน้ำน้านี้

👌 «new process*» – RapidMiner Stur	dio Educational 9.2.000 @ H61M-PC								-		×
Eile Edit Process View Conne	ections Cloug Settings Extensions	s <u>H</u> elp								-	
🔰 🧮 🔹 🚺	•	Views:	Design	Results	Turbo Prep	Auto Model		Find data, operators	ele 👂	All Studio	*
Repository ×	Process							Parameters	×		
😋 Import Data 🛛 = 👻	Process					100% 🔎 .	🗵 电 🗲 📮 🔍 🔍	å Read C SV			
Training Resources (connected	Process							> Import Cor	nfiguration Wizard	. 0	D
Samples Samples Community Samples (connects	at Dimp The Market CSV	d J					***	csv file	esitop\555.csv		D
 DB Local Repository (HS1M) 	-						ms (column separators		d	D
Cloud Repository idisconnected								v se quotes		d	D
								quotes character	-	d	D
								📝 skip comments		d	D
Operators X								comment characters		a	D
								The show advanced	parameters		
read Cav X								Channe compati	hility (9.2.000)		

19. หน้าต่างโปรแกรม หลังจากกด Run โปรแกรม

Fait Floce	ss Yiew Con	nections Cloug	Settings Extensio	ns Help												
		*			Views: Design	Result	s Turbo Pre	p Auto Model	8		Find data, opera	torsetc	2	All St		
sult History	. E	xampleSet (Rea	d CSV) ×											sit		
	Open in	Turbo Prep	Auto Model	Filter (5.422 / 5.422 examples): all									*			
Data	Row No.	180	กณ์ต	ำหม์	สมิตส	สุขศึกษา	fa#	การงาน	ต่างประเทศ	รหสินกิศักษา		endgrade				
	1	ด้มาก	บานกลาง	ด้มาก	ดีมาก	ต่า	ดีมาก	sh	บานกลาง	5822041025	ຄອມເຈົ້າເອອົ້າຮູ	2.210	^	1		
Σ	2	ปานกลาง	ดำ	ด่า	มาแกลาง	ชานกลาง	ด้มาก	ดำ	ปานกลาง	5122010235	การขาย	2.490				
Statistics	3	ปานกลาง	ล่า	ปาหกลาง	sh.	สานกลาง	ด์มาก	ล่า	ปาษกลาง	4922010357	การขาย	2.280		3		
	4	สำ	ปานกลาง	ปานกลาง	ปาแก่สาง	สานกลาง	ด์มาก	ต่ำ	มาษกลาง	5122010225	การขาม	2.500				
	5	ดีมาก	ต่ำ	บานกลาง	dh	มานกลาง	ดีมาก	dh .	บานกลาง	4922010315	การชาย	2.820				
sualizations	6	<u>ปานก</u> ลาง	ดำ	ปานกลาง	ด่า	ชานกลาง	ด้มาก	ดำ	บานกลาง	4922010336	การชาย	2.400				
	7	ปานกลาง	ล่า	ปาษทสาง	ส่า	ต่า	ต้มาก	ล่า	ปาษกลาง	5022010174	การขาย	2.240				
	8	มานคลาง	สำ	ปานกลาง	ต่า	ต่ำ	ด์มาก	ล่า	ปานกลาง	4922010321	การขาม	2.310				
	9	ดีมาก	บานกลาง	มานกลาง	dh	dh	ดีมาก	ต่ำ	บานกลาง	4922010353	การขาย	2.530				
Annotations	10	ปานกลาง	ปานกลาง	ปานกลาง	มาแก่ดาง	ล่า	ดีมาก	ดำ	ปานกลาง	5022010208	การชาย	2.380				
	11	ล่า	สำ	ปาษกลาง	ปาแกลาง	ต่ำ	ที่มาก	สำ	ปาษกลาง	4922010278	การขาย	2.370				
	12	ปานกลาง	ล่า	ปานกลาง	sh	มาแกลาง	ดีมาก	ล่า	ปานกลาง	4922010343	การชาย	2.560				
	13	มานกลาง	มานกลาง	มานกลาง	ดีมาก	ดีมาก	ดีมาก	dh .	ดีมาก	4922010304	การขาย	2.970				
	14	ปานกลาง	ปานกลาง	ปานกลาง	ต่า	ดำ	ด์มาก	ล่า	ปานกลาง	5022010173	การขาม	2.640				
	15	ที่มาก	ล่า	ปานกลาง	ต่า	ต่ำ	ด์มาก	ต่ำ	มาษกลาง	4922010318	การขาย	2.630				
	16	ปานกลาง	ล่า	ปาหกลาง	ต่า	ต่า	ดีมาก	สำ	ม่าพกลาง	4922010390	การชาย	2.500				
	17	บานกลาง	บานกลาง	บานกลาง	มาแกลาง	dh	ดีมาก	<i>i</i> h	มานกลาง	4922010295	การชาย	2.790				

20.หลังจากได้ข้อมูลเบื้อต้นแล้วกลับมาที่หน้าต่าง design เพื่อทำให้ endgrade เปลี่ยนค่าของตัวเลขเป็น ข้อความแบ่งระดับ โดยจะแบ่งระดับ เป็น

> - ดีมาก = 4 - ปานกลาง = 3 - ต่ำ =2

การตั้งค่าเพื่อให้ง่ายต่อการใช้งานโดยการ คลิกขวาเลือก Insert Operator >> Cleansing >> Binning >> Discretize by User Specification



21. เมื่อได้ Operator แล้วเลือกตัว Operator แล้วเลือก Attributes filter type เป็น subset หลังจากนั้น เลือก Select Attributes เพื่อเลือก Attributes ที่ต้องการแทนที่ค่า เมื่อเลือก Attributes ที่ต้องการเสร็จแล้วกด apply

Select Attributes: attributes The attribute which should be ch	iosen.		
Attributes	Selec	cted Attributes	
Search	X Sea	irch	G 🕽
# รหัสแก๊กดึกษา	# ¢	endgrade	
			Apply

22. หลังจากนั้นกำหนดเกณฑ์ที่ต้องการเทียบกับคะแนน เพื่อแปลงเป็นระดับที่เราตั้งไว้ ในช่อง Classes เลือก Edit List

Process		Parameters	X
Process >	100% 🔎 🔎 📮 🥃 💣 🔯	🍅 Discretize (Disc	cretize by User Specification)
Process		attribute filter type	subset 🔻 🗊
P Read Excel Select Attributes Discretize	res	attributes	Select Attributes
P on pre		invert selection	٤.
		include special a	attributes II .
		classes	Defit List (2)
		Show advanced	parameters
		Change compatient	<u>ibility (9.2.000)</u>

23. เมื่อกดเข้าไปจะมีหน้าต่าง ให้เรากำหนดเกณฑ์ เมื่อเรากำหนดเกณฑ์แล้วกด Apply

class names	upper limit
ม าก	4.0
านกลาง	3.0
h	2.0

24. หลังจากตั้งค่าเสร็จ กด Apply แล้วสั่งโปรแกรม Run จะได้ผลตามภาพ endgrade เปลี่ยนเป็น

ข้อความ

Eile Edit Proce	ess <u>V</u> iew <u>C</u> on	nections Cloud	Settings Extension	ons <u>H</u> elp		-		-					
	-	•	1		Views: Desig	n Resu	ilts Turbo P	rep Auto Mo	del		Find data, opera	itorsetc	P
Result History	E E	xampleSet (Disc	retize) ×										
	Open in	🚦 Turbo Prep	Auto Model]						Filter (5,422 / 5,42	22 examples): all		Ŧ
Data	Row No.	endgrade	ใ ทย	กณ์ด	วิทย์	สังคม	สุขศึกษา	ค ลปะ	การงาน	ต่างประเทศ	รหัสนักศึกษา	แผนก	
	1	ปานกลาง	ดีมาก	ปานกลาง	ดีมาก	ดีมาก	ดำ	ดีมาก	ดำ	ปานกลาง	5822041025	คอมพิวเตอร์ธุ.	-
Σ	2	ปานกลาง	ปานกลาง	ต่ำ	ต่ำ	ปานกลาง	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	5122010235	การขาม	
Statistics	3	ปานกลาง	ปานกลาง	ดำ	ปานกลาง	ด่า	ปานกลาง	ต์มาก	ดำ	ปานกลาง	4922010357	การขาย	
	4	ปานกลาง	ด่า	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดำ	ปานกลาง	5122010225	การขาย	
S	5	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ต่ำ	ปานกลาง	4922010315	การขาม	
Visualizations	6	ปานกลาง	ปานกลาง	ดำ	ปานกลาง	ด่า	ปานกลาง	ต์มาก	ดำ	ปานกลาง	4922010336	การขาย	
	7	ปานกลาง	ปานกลาง	ดำ	บานกลาง	ด่า	ดำ	ดีมาก	ต่ำ	ปานกลาง	5022010174	การขาย	
	8	ปานกลาง	ปานกลาง	ต่ำ	ปานกลาง	ด่า	ต่ำ	ดีมาก	ต่ำ	ปานกลาง	4922010321	การขาย	
	9	ปานกลาง	ดีมาก	ปานกลาง	ปานกลาง	ด่า	ด่า	ดีมาก	ดำ	ปานกลาง	4922010353	การขาย	
Annotations	10	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ดำ	ปานกลาง	5022010208	การขาย	
	11	ปานกลาง	ดำ	ต่ำ	ปานกลาง	ปานกลาง	ต่ำ	ดีมาก	ต่ำ	ปานกลาง	4922010278	การขาย	
	12	ปานกลาง	ปานกลาง	ดำ	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ดำ	ปานกลาง	4922010343	การขาย	
	13	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดีมาก	ดีมาก	ด่า	ดีมาก	4922010304	การขาย	
	14	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ด่า	ดำ	ดีมาก	ต่ำ	ปานกลาง	5022010173	การขาย	
	15	ปานกลาง	ดีมาก	ด่า	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง	4922010318	การขาย	
	16	ปานกลาง	ปานกลาง	ด่า	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	ปานกลาง	4922010390	การขาย	
	17	ปานกลาง	ปานกลาง	มานกลาง	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ต่ำ	ปานกลาง	4922010295	การขาย	

25. หลังจากนั้นเราจะต้องตั้งค่า endgrade ให้เป็น label เพื่อใช้ในการทำนาย โดยพิมพ์ค้นหาที่ operator ว่า Set Role และลาก Operator Set Row มาว่างที่หน้าต่างการทำงาน และลากเส้นให้เชื่อมต่อ กัน ดังรูป



26. หลังจากนั้น คลิกที่ operator Set Role แล้วเลือก attribute name เป็น endgrade เลือก target role เป็น label หลังจากนั้นลากเส้นเชื่อมเส้นสุดท้าย และ กด Run โปรแกรม ตัว endgrade ก็จะ เป็นเณฑ์เดียวกับ Attributes อื่น ๆ

Parameters ×	
🚺 Set Role	
attribute name	endgrade 🔻 🗊
target role	label 🔻 🛈
set additional roles	Edit List (1)

E:	xampleSet (Set I	Role) $ imes$									
Open in 🕎 Turbo Prep 👫 Auto Model Filter (5,422 / 5,422 examples): all											
Row No.	endgrade	ใ ทย	กณ์ด	วิทย์	สังคม	สุขศึกษา	ศ์ ลปะ	การงาน	ต่างประเทศ	รหัสนักศึกษา	แผนก
1	บานกลาง	ดีมาก	ปานกลาง	ดีมาก	ดีมาก	ด่า	ดีมาก	ด่า	ปานกลาง	5822041025	คอมพิวเตอร์ธุ
2	บานกลาง	ปานกลาง	ด่า	ด่า	ปานกลาง	ปานกลาง	ดีมาก	ด่า	ปานกลาง	5122010235	การขาย
3	บานกลาง	ปานกลาง	ด่า	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	ปานกลาง	4922010357	การขาย
4	บานกลาง	ด่า	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ด่า	ปานกลาง	5122010225	การขาย
5	บานกลาง	ดีมาก	ด่า	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	ปานกลาง	4922010315	การขาย
6	บานกลาง	ปานกลาง	ด่า	ปานกลาง	ด่า	ปานกลาง	ดีมาก	ด่า	บานกลาง	4922010336	การขาย
7	บานกลาง	ปานกลาง	ด่า	ปานกลาง	ด่า	ด่า	ดีมาก	ด่า	บานกลาง	5022010174	การขาย
8	บานกลาง	ปานกลาง	ด่า	ปานกลาง	ด่า	ดำ	ดีมาก	ด่า	ปานกลาง	4922010321	การขาย
9	บานกลาง	ดีมาก	ปานกลาง	ปานกลาง	ด่า	ดำ	ดีมาก	ด่า	ปานกลาง	4922010353	การขาย
10	บานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ด่า	ปานกลาง	5022010208	การขาย
11	ปานกลาง	ด่า	ด่า	ปานกลาง	ปานกลาง	ดำ	ดีมาก	ด่า	ปานกลาง	4922010278	การขาย
12	ปานกลาง	ปานกลาง	ด่า	ปานกลาง	ต่ำ	ปานกลาง	ดีมาก	ด่า	ปานกลาง	4922010343	การขาย
13	ปานกลาง	ปานกลาง	ปานกลาง	ปานกลาง	ดีมาก	ดีมาก	ดีมาก	ด่า	ดีมาก	4922010304	การขาย

27.หลังจากนั้น เลือกมุมมอง Design ต่อไปเราจะทำการสร้างโมเดล Decision Tree โดยการเลือก โอเปอเรเตอร์ Decision Tree จากส่วนของ Operators โดยการพิมพ์ตรงช่องค้นหา โดยพิมพ์คำว่า Decision กด ปุ่ม Enter ก็จะปรากฏโอเปอเรเตอร์ Decision Tree ขึ้นมา หรือจะทำการเลือกจากหมวด Modeling >> Classification and Regression >> Tree Induction

Op	erators ×							
De	ecision	×						
	🕶 🛅 Predictive (8)	^						
	🔻 📇 Trees (8)							
	💡 Decision	Tree						
	💡 Random	Forest						
	Gradient Boosted Ti							
	🖓 ID3							
	💡 Decision	Stump						
<		>						
We found "Weka Extension" and "WhiBo" in the Marketplace. <u>Show mel</u>								

11. ลากโอเปอเรเตอร์ Decision Tree มาวางในส่วนของ Process ตรงเส้นที่เชื่อมต่อเดิมที่ โอเปอเรเตอร์ Read Excel ลากไว้ (โปรแกรมจะทำการเชื่อมโอเปอเรเตอร์ทั้งสองตัวทันทีจากพอร์ต out ของโอเปอเรเตอร์ Read Excel ไปยังพอร์ต tra (training) ของโอเปอเรเตอร์ Decision Tree เพื่อเป็นการส่ง ข้อมูลไปสร้างโมเดล



12. ลากเส้นเชื่อมจากพอร์ต mod (model) และพอร์ต exa (example) ของโอเปอเรเตอร์ Decision Tree ไปยังพอร์ต res (result) ทั้งสองพอร์ต เพื่อไปแสดงในส่วนของหน้าจอผลลัพธ์โดยพอร์ต mod จะทำการส่งโมเดล Decision Tree ที่สร้างออกไปแสดงในรูปต้นไม้ และพอร์ต exa จะส่งข้อมูลที่ import เข้ามาไปแสดงในรูปแบบตาราง

13. จากนั้นคลิก Run Process จะได้รูปโมเดลต้นไม้ ซึ่งโมเดลต้นไม้ที่สร้างได้มีส่วนประกอบ สำคัญ 3 ส่วน คือ

- ในโมเดล Decision Tree จะมีโหนดต่าง ๆ 2 ประเภท คือ
 - ๑ โหนดที่เป็นแอตทริบิวต์แสดงด้วยรูปสี่เหลี่ยมที่มีมุมโค้ง
 - O โหนดลาเบลแสดงด้วยรูปสี่เหลี่ยมที่มีกราฟแสดงสีต่าง ๆ อยู่ด้วย ในตัวอย่าง

นี้มี 2 ลาเบล คือ ดีมาก และ ปานกลาง

- ส่วนของ Zoom ใช้สำหรับย่อขยายรูปโมเดล
- ส่วนของ Mode จะใช้สำหรับปรับโหมดของการใช้งานเมาส์

			ดีมาก	
การงาน				
ดีมาก ดำ ปานกลาง				ดีมาก
ปานกลาง ดีมาก	ก	ารงาน		
	ดีมาก	ต่ำ บานกลาง		ดีมาก ต่ำ
	ใทย	ปานกลาง ปานกลาง	ปานกลาง 🖌 ปานกลาง 🍐	
	ต่ำ บานกลาง			
	ปานกลาง ดีมาก			วิทย์
				ดีมาก บานกล
				ดีมาก ปา

14. ในหน้าต่าง Description จะเป็นโค๊ดข้อความที่เราสามารถนำมาเขียนโปรแกรมเพื่อใช้ในการทำนายได้

Result History	♀ Tree (Decision Tree) ×
Graph	Tree ต่างประเทศ = ดีมาก ศิลปะ = ดีมาก
Description	คณิต = ดีมาก การงาน = ดีมาก วิทย์ = ดีมาก: ดีมาก {ด่ำ=0, ปานกลาง=0, ดีมาก=171} วิทย์ = ปานกลาง วิทย์ = ปานกลาง สุขศึกษา = ดีมาก: ดีมาก {ด่ำ=0, ปานกลาง=0, ดีมาก=8}
Annotations	
	<pre>1 ที่มียุ่ะ ติมาก: ติมาก (ตา=0, บานกลาง=1, ติมาก=3) 1 วิทย์ = ปานกลาง: ปานกลาง (ต่ำ=0, ปานกลาง=3, ดีมาก=0) 1 คณิต = ต่ำ 1 สุขศึกษา = ตีมาก 1 สุขศึกษา = ตีมาก 1 การงาน = ตีมาก: ตีมาก {ต่ำ=0, ปานกลาง=0, ตีมาก=4} 1 การงาน = ต่ำ: ปานกลาง {ต่ำ=0, ปานกลาง=2, ตีมาก=1} 1 การงาน = ปานกลาง: ตีมาก {ต่ำ=0, ปานกลาง=1, ตีมาก=4} 1 สุขศึกษา = ต่ำ 1 สุขศึกษา = ต่ำ 1 สุขศึกษา = ต่ำ 1 สุขศึกษา = ต่ำ 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 การงาน = ตีมาก 1 1 การงาน = ตีมาก 1 1 การงาน = ตีมาก 1 1 การงาน = ปานกลาง: ตีมาก {ต่ำ=0, ปานกลาง=1, ตีมาก=1} 1 1 การงาน = บานกลาง: ตีมาก {ต่ำ=0, ปานกลาง=1, ตีมาก=2} 1 1 การงาน = บานกลาง: ตีมาก (ต่ำ=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: ตีมาก (ต่ำ=0, ปานกลาง=1, ตีมาก=1) 1 1 การงาน = บานกลาง: ตีมาก (ต่ำ=0, ปานกลาง=1, ตีมาก=1) 1 1 การงาน = บานกลาง: ตีมาก (ต่ำ=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = ต่า: ปานกลาง: ตีมาก (ต่า=0, ปานกลาง=1, ตีมาก=1) 1 1 การงาน = บานกลาง: ตีมาก (ต่า=0, ปานกลาง=0, ตีมาก=2) 1 1 การงาน = บานกลาง: ตีมาก (ต่า=0, ปานกลาง=0, ตีมาก=2) 1 1 การงาน = บานกลาง: ตีมาก (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: ตีมาก (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: ตีมาก (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 1 การงาน = บานกลาง: (ต่า=0, ปานกลาง=1, ตีมาก=2) 1 </pre>

การทดสอบการทำนาย

1.การทดสอบการทำนายโดยใช้ Cross Validation ใช้งานโดย คลิกขวาเลือก Insert Operator
 > Validation >> Cross Validation หลังจากนั้นดับเบิลคลิกเข้าไปที่ Cross Validation จะแสดง
 หน้าต่างดังรูป

Process		
Process Cross Validation	100% 🔑 🔑 📴 🛛	•
Training	Testing	
tra mod) mod to	25
thr	tes p	er 🐔
	U thr	

2. ทางด้านซ้ายของช่องให้นำ โมเดล Decision Tree มาวาง และลากเส้นเชื่อม ทางด้านขวา ลาก Apply Model และ Performance มาวาง และลากเส้นดังรูป

Process					
① Process ► Cross V	/alidation ▶			100% 🔎 🎾	🗵 🖷 🍒 📮 🍳
Training		Test	ting		
tra tra	mod exa wei	thr thr	Apply Model	Performance	tes per per

3. ที่หน้าต่าง design ก็จะมีการใช้ Operator ต่าง ๆ ดังนี้



4.ผลกาตรวจสอบคุณภาพของการทำนาย หาก% ผลการทำนายยิ่งมากความถูกต้องก็ยิ่งมากขึ้น

ตามไปด้วย

eVector (Performance) 🛛 🗙	ExampleSet (Set Role) ×	💡 Tree (Decision Tree (2))	×	
Table View O Plot View				
accuracy: 88.56% +/- 1.17% (mic	ro average: 88.57%)			
	true ดำ	true ปานกลาง	true ดีมาก	class precision
pred. ดำ	18	28	1	38.30%
pred. ปานกลาง	42	2806	349	87.77%
pred. ดีมาก	1	199	1978	90.82%
class recall	29.51%	92.52%	84.97%	

% Performance	PerformanceVector: accuracy: 88.56% +/- 1.17% (micro average: 88.57%)
Description	ConfusionMatrix: True: ต่ำ ปานกลาง ดึมาก ต่ำ: 18 28 1 ปานกลาง: 42 2806 349 ดึมาก: 1 199 1978 kappa: 0.770 +/- 0.023 (micro average: 0.770)
Annotations	<pre>ConfusionMatrix: True: ต่ำ ปานกลาง ดีมาก ด่ำ: 18 28 1 ปานกลาง: 42 2806 349 ดีมาก: 1 199 1978</pre>

การทดสอบความถูกต้องของผลการทำนายอีกแบบหนึ่งคือการใช้ Split โดยการพิมพ์ค้นหาที่ Operator ว่า Split Validation แล้วเลือกคลิกมาว่างที่หน้าต่างการทำงานแล้วลากเส้นเชื่อมต่อดังภาพ หลังจากนั้นดับเบิลคลิกเข้าไปจะเจอหน้าต่างการทำงาน



หลังจากดับเบิลคลิกเข้ามาที่หน้าต่างแล้วจะเจอหน้าต่างดังภาพ

Process					Parameters ×			
Process > Validation >			100% 🔎 🔎 📮 🖥	i 🖉 🖬	% Validation (Split Validation)			
Training		Testing			split	relative	•	
tra	mod	mod		ave 🖣	split ratio	0.7	٦	
	thr (tes) thr		ave 🧲 ave 🧲	sampling type	automatic	•	
					-			

หลังจากเข้ามาแล้วให้พิมพ์ค้นหาในช่อง Operator และนำ Operator ทางด้านซ้ายของช่องให้นำ โมเดล Decision Tree มาวาง และลากเส้นเชื่อม ทางด้านขวา ลากApply Model และ Performance มา วาง และลากเส้นดังรูป

Process			
O Process ► Valida	tion 🕨		100% 🔎 🔑 🟳 🛃 🦉 🕅
Training		Testing	
	Decision Tree	Apply Model P	Performance
fra	tra mod mod	tes uni mod lab thr	b per ave exa ave

เมื่อกดรันผลการทำนายจะเป็นดังรูป

Table View Plot View									
accuracy: 88.50%									
	true ดำ	true ปานกลาง	true ดีมาก	class precision					
pred. ต่ำ	3	7	0	30.00%					
pred. ปานกลาง	14	842	104	87.71%					
pred. ดีมาก	1	61	594	90.55%					
class recall	16.67%	92.53%	85.10%						

Result History	% PerformanceVector (Performance)	×	Exan
%	PerformanceVector		
Penormance	PerformanceVector: accuracy: 88.50%		
Description	ConfusionMatrix: True: ด่ำ ปานกลาง ดีมาก ด่ำ: 3 70 ปานกลาง:14 842 104 ดีมาก: 1 61 594 kappa: 0.768		
Annotations	ConfusionMatrix: True: ด่ำ ปานกลาง ดีมาก ด่ำ: 3 70 ปานกลาง:14 842104 ดีมาก: 1 61 594		

@Rodjanaburanon Nam



http://dataminingtrend.com/2014/wp-content/uploads/2014/02/chapter1.pdf (25-2-62)

http://dataminingtrend.com/2014/wp-content/uploads/2014/02/RM7_chapter1.pdf (25-2-62)

https://behavior.lbl.gov/?q=node/11 (25-2-62)

http://compcenter.bu.ac.th/news-information/data-mining(25-2-62)

