



Chapter 3

Introduction to RapidMiner Studio





Data mining software

- Commercial software
 - SAS Enterprise Miner



- IBM SPSS Modeller



- RapidMiner Studio



- Open source software

- RapidMiner Studio



- Weka



- R





Data mining software

- RapidMiner Studio 7

ExampleSet (661 examples, 2 special attributes, 9 regular attributes) Filter (661 / 661 examples): all

Row No.	User_ID	eReader_A...	Gender	Age	Marital_Sta...	Website_Ac...	Browsed_El...	Bought_Ele...	Bought_Di...
1	9552	Late Majority	M	61	M	Seldom	Yes	Yes	Yes
2	6757	Innovator	M	27	M	Regular	Yes	No	Yes
3	3599	Early Adopter	F	29	M	Seldom	Yes	Yes	Yes
4	6811	Early Adopter	M	49	M	Seldom	Yes	Yes	Yes
5	4104	Late Majority	M	29	S	Seldom	Yes	Yes	Yes
6	7226	Early Majority	M	60	S	Seldom	No	Yes	Yes
7	7814	Early Adopter	F	53	S	Regular	Yes	No	Yes
8	4693	Late Majority	F	63	S	Seldom	Yes	Yes	Yes
9	7266	Innovator	M	46	M	Regular	Yes	Yes	Yes
10	5259	Early Majority	F	25	S	Seldom	Yes	Yes	Yes
11	7417	Innovator	F	40	M	Regular	Yes	Yes	No
12	6795	Early Adopter	M	29	S	Seldom	Yes	Yes	Yes
13	8669	Late Majority	F	49	S	Seldom	Yes	Yes	Yes
14	9888	Innovator	M	41	M	Regular	Yes	Yes	Yes
15	9181	Innovator	F	47	S	Regular	Yes	No	No



Data mining software

- Weka

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The current relation is 'weather.symbolic' with 5 attributes and 14 instances. The 'outlook' attribute is selected, showing a distribution of 5 'sunny', 4 'overcast', and 5 'rainy' instances. A stacked bar chart visualizes the distribution of the 'play' class (Nominal) across the 'outlook' categories.

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) Visualize All

Outlook	play = Yes	play = No
sunny	3	2
overcast	4	0
rainy	2	3



Data mining software

- R

```
R Console
~ Help Search

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
[R.app GUI 1.65 (6784) x86_64-apple-darwin13.1.0]

WARNING: You're using a non-UTF8 locale, therefore only ASCII characters will
work.
Please read R for Mac OS X FAQ (see Help) section 9 and adjust your system
preferences accordingly.
[History restored from /Users/sit/.Rapp.history]

> x <- matrix(nrow=2,ncol=1,c(1,2))
> x
      [,1]
[1,]    1
[2,]    2
>
```



Data mining software

- **RapidMiner**

- ในตอนแรกใช้ชื่อบริษัทว่า Rapid-I ก่อตั้งขึ้นเมื่อปี 2006
- ในช่วงแรกบริษัทตั้งอยู่ที่ประเทศเยอรมนี
- ปี 2013 ได้เปลี่ยนชื่อบริษัทเป็น RapidMiner หลังจากได้รับเงินลงทุนจำนวน 5 ล้านดอลลาร์สหรัฐ
- และย้ายบริษัทมาอยู่ที่บอสตัน ประเทศสหรัฐอเมริกา
- ผลิตภัณฑ์หลักของบริษัทคือ RapidMiner Studio 7
- บริษัทชั้นนำต่างๆ เช่น PayPal ใช้ซอฟต์แวร์ RapidMiner Studio 7



Ingo Mierswa



Ralf Klinkenberg





Data mining software

- ผลการสำรวจจากเว็บไซต์ KDnuggets ในปี 2014 พบว่าผู้ร่วมตอบแบบสำรวจใช้ RapidMiner ในการวิเคราะห์ข้อมูลมากเป็นอันดับ 1
- Gartner ได้จัดให้ RapidMiner อยู่ในกลุ่ม Leaders สำหรับซอฟต์แวร์ในการวิเคราะห์ข้อมูล (2016)
- SAS
- IBM
- KNIME
- RapidMiner





Rapidminer

- download ได้จาก <http://rapidminer.com>

rapidminer

PREDICTIVE ANALYTICS REIMAGINED

Do More Data Science

Prep data, create models & embed in business processes faster than ever before

GET STARTED

คลิกที่เมนูนี้

DOWNLOAD

Sales



Rapidminer

- เลือกดาวน์โหลด version ที่เหมาะสมกับระบบปฏิบัติการของเรา

The screenshot shows the 'Downloads' page of the RapidMiner website. The browser address bar shows 'my.rapidminer.com'. The page has a navigation bar with 'Sign up' and 'Downloads' (highlighted) links. There are input fields for 'Email address' and 'Password', and a 'Log in' button. Below the navigation bar, the main heading is 'Downloads' with the instruction 'Click on a RapidMiner product of your choice to download it.' The main content area is titled 'RapidMiner Studio 7' and says 'Click on your operating system to start the download:'. There are four buttons for operating systems: 'Windows 32bit', 'Windows 64bit', 'Mac' (highlighted in red), and 'Linux'. The 'Mac' button has the text 'Requires: Mac OS 10.8+'. The 'Linux' button has the text 'Requires: Java 7 or Java 8'. To the right of these buttons is a list of links: 'Installation Guide', 'Getting Started Tutorials', 'Support', and 'Download Source'.



Rapidminer

The screenshot displays the RapidMiner Studio Professional 7.0.000 interface on a Mac. The main window is titled "Process" and shows a "Design" view of a workflow. A central window titled "About RapidMiner Studio" is open, displaying the RapidMiner logo and "Studio Version 7.0". Below this, it states: "RapidMiner Studio 7.0.000 (rev: 841a69, platform: OSX) Professional Edition registered to Eakasit Pacharawongsakda Copyright (C) 2001 - 2016 RapidMiner GmbH".

The interface includes several panels:

- Repository:** Contains "Add Data" and a list of data sources: Samples, DB, Local Repository (sit), and Cloud Repository (disconnected).
- Operators:** A search bar contains "decision t". The "Modeling (3)" folder is expanded to show "Predictive (3)", which is further expanded to show "Trees (3)". The "Decision Tree" operator is selected.
- Parameters:** Shows the configuration for the "Decision Tree" operator:
 - Criterion: gain_ra...
 - Maximal depth: 4
 - Apply pruning:
 - Confidence: 0.25
 - Apply prepruning:
 - Minimal gain: 0.1
 - Link: [Show advanced parameters](#)
- Help:** Shows the "Decision Tree" operator details, including "RapidMiner Studio Core" and a "Synopsis" section that begins with "Generates a Decision Tree for".

At the bottom, there is a "Recommended Operators" section with a green checkmark.



Rapidminer Professional Licence





Rapidminer Professional Licence

- พบกับหน้าต่าง Home Screen

The screenshot shows the Rapidminer Professional Home Screen. On the left, there is a vertical navigation menu with four main sections: 'GET STARTED', 'LEARN', 'NEW PROCESS', and 'OPEN PROCESS'. Each section is numbered 1 through 4 in blue circles, with corresponding Thai text annotations in white boxes with blue borders. The main content area on the right is titled 'start from:' and features four project cards: 'Churn Modeling', 'Credit Risk Modeling', 'Direct Marketing', and 'Predictive Maintenance'. Each card includes an icon and a brief description of the model's purpose.

Annotation Number	Annotation Text (Thai)	Card Title	Card Description
1	แนะนำการใช้งานเบื้องต้น	Churn Modeling	Predict which of your customers will churn and why with an optimized decision tree.
2	ดูตัวอย่างที่เตรียมไว้	Credit Risk Modeling	Model credit default risk by training an optimized Support Vector Machine (SVM) model.
3	สร้าง process ใหม่	Direct Marketing	Predict response to campaigns and increase the conversion rate of your campaign.
4	เปิด process ที่มีอยู่	Predictive Maintenance	Model equipment failures to schedule maintenance pre-emptively.



Rapid Miner

- พบกับหน้าต่าง Home Screen

The screenshot shows the Rapid Miner Home Screen interface. On the left, there is a vertical navigation menu with four main sections: GET STARTED, LEARN, NEW PROCESS, and OPEN PROCESS. Each section is numbered 1 through 4 in blue circles. The main content area on the right is titled 'start from:' and contains several project templates. The first template is 'Churn Modeling', which includes an icon of a radio tower and a description: 'Predict which of your customers will churn and why with an optimized decision tree.' Below this is 'Credit Risk Modeling', with an icon of a hand holding a coin and a description: 'Model credit default risk by training an optimized Support Vector Machine (SVM) model.' At the bottom, there are two more templates: 'Direct Marketing' (with an envelope icon) and 'Market Basket Analysis' (with a shopping cart icon). The 'Predictive Maintenance' template is partially visible at the bottom right, with an icon of a robotic arm and a description: 'Model equipment failures to schedule maintenance pre-emptively.'

1. แนะนำการใช้งานเบื้องต้น

2. ดูตัวอย่างที่เตรียมไว้

3. สร้าง process ใหม่

4. เปิด process ที่มีอยู่

start from:

Churn Modeling
Predict which of your customers will churn and why with an optimized decision tree.

Credit Risk Modeling
Model credit default risk by training an optimized Support Vector Machine (SVM) model.

Direct Marketing
Predict response to campaigns and increase the conversion rate of your campaign.

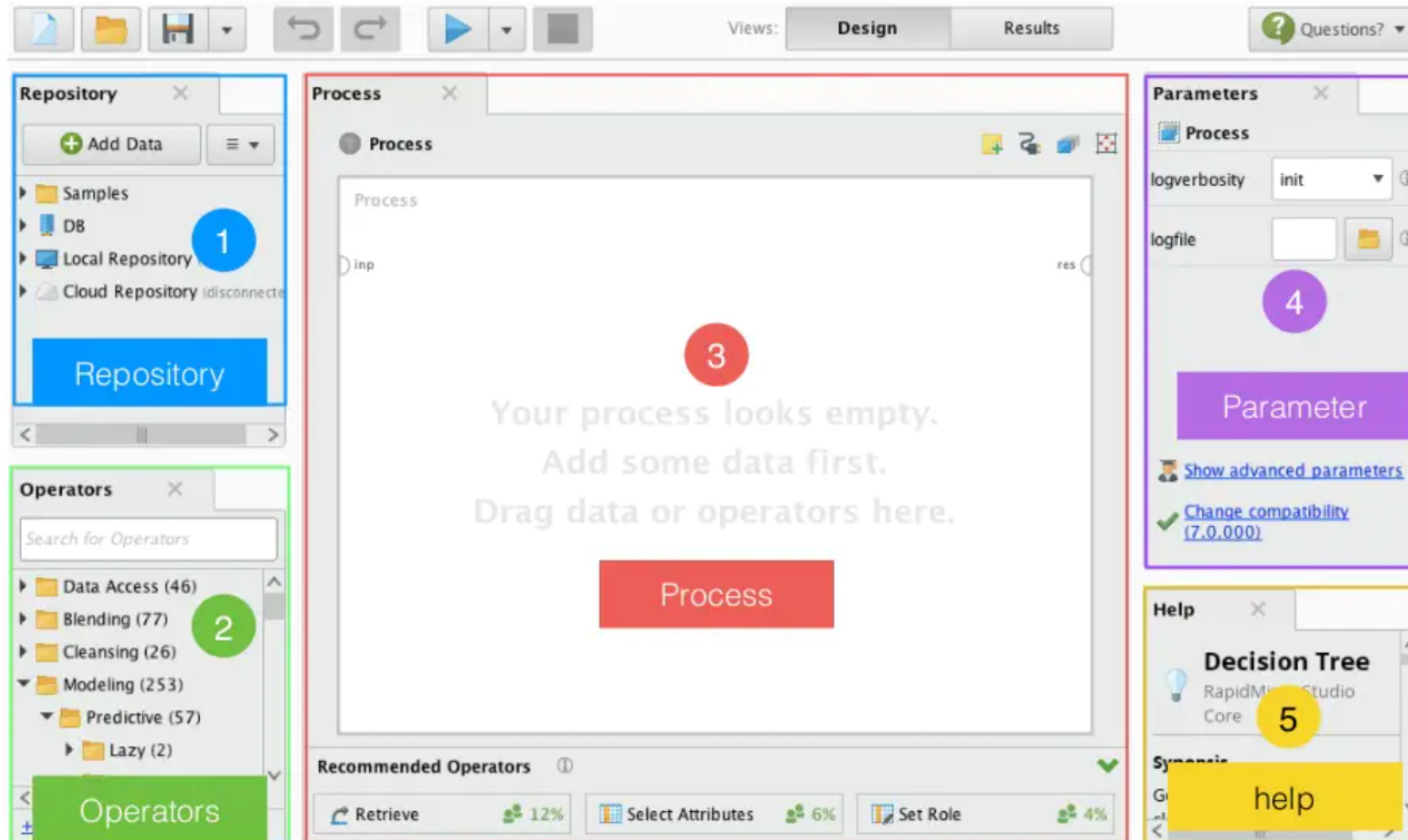
Market Basket Analysis
Find products frequently purchased together and turn them into rules for recommendations.

Predictive Maintenance
Model equipment failures to schedule maintenance pre-emptively.



Rapid Miner

- องค์ประกอบของ RapidMiner Studio 7



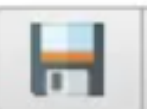






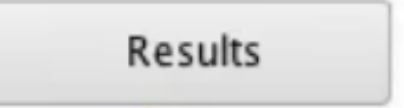
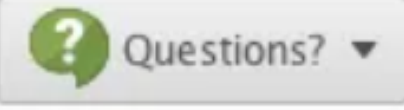




Rapid Miner

- เมนูใน RapidMiner Studio 7



-  ใช้สร้าง process ใหม่
-  โหลด process เดิม
-  บันทึก process
-  undo หรือ redo
-  สั่งให้ process ทำงาน (run)
-  ยกเลิก process (stop)


-  Design
 -  Results
 -  ? Questions? ▼
- เมนูสำหรับปรับเปลี่ยนมุมมองของ RapidMiner Studio 7
-  a
 -  b
- หน้า Design แสดงหน้าสำหรับการสร้างโปรเซส
- หน้า Results แสดงหน้าผลลัพธ์การทำงาน
- เมนูสำหรับแสดงวิธีการใช้งานในรูปแบบต่างๆ





Rapid Miner

- เมนูใน RapidMiner Studio 7



 ใช้สร้าง process ใหม่


 โหลด process เดิม

 บันทึก process

 undo หรือ redo


 สั่งให้ process ทำงาน (run)

 ยกเลิก process (stop)

 Design Results
เมนูสำหรับปรับเปลี่ยนมุมมองของ RapidMiner Studio 7

 a

 b

 a หน้า Design แสดงหน้าสำหรับการสร้างโปรเซส

 b หน้า Results แสดงหน้าผลลัพธ์การทำงาน

 Questions? เมนูสำหรับแสดงวิธีการใช้งานในรูปแบบต่างๆ