

การเปรียบเทียบการใช้ อัลกอริทึม อพริเออร์ คาร์มา และซีแควน

สำหรับการวิเคราะห์พฤติกรรมนักลงทุน

A COMPARISON OF USING APRIORI, CARMA AND SEQUENCE
FOR ANALYZING BEHAVIORAL TRADERS



T139357

โดย

ธีระโรจน์ ศรีศาสวัตกุล

TEERAROJ SRISASWATTAKUN

อาจารย์ที่ปรึกษา

พศ.ดร. กัณฑ์พงษ์ วรรัตน์ปัญญา



เลขหมู่.....
เลขทะเบียน..... 139357
วัน,เดือน,ปี..... 30 ต.ค. 2558

รายงานนี้เป็นส่วนหนึ่งของวิชาการศึกษาอิสระ 2

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2556

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**A COMPARISON OF USING APRIORI, CARMA AND SEQUENCE
FOR ANALYZING BEHAVIORAL TRADERS**



**A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS OF THE COURSE
INDEPENDENT STUDY 2
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

2 / 2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2014

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองการศึกษาอิสระ 2 (Independent Study 2)

เรื่อง

การเปรียบเทียบการใช้ อัลกอริทึม อพริออริ คาร์มา และซีแควนซ์
สำหรับการวิเคราะห์พฤติกรรมนักลงทุน

A COMPARISON OF USING APRIORI, CARMA AND SEQUENCE
FOR ANALYZING BEHAVIORAL TRADERS

นายธีระโรจน์ ศรีศาศวัตกุล

รหัสประจำตัว 55660920

ขอรับรองว่ารายงานฉบับนี้ ข้าพเจ้าไม่ได้คัดลอกมาจากที่ใด
รายงานฉบับนี้ได้รับการตรวจสอบและอนุมัติให้เป็นส่วนหนึ่งของการศึกษาวิชาการศึกษาอิสระ 2 หลักสูตรวิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศ) ภาคเรียนที่ 2 ปีการศึกษา 2556

.....อาจารย์ที่ปรึกษา

(ผศ.ดร. กนต์พงษ์ วรรณปัญญา)

๑๖๒ เปรมชัยเชิฐ

.....กรรมการสอบ

(ดร. นล เปรมชัยเชิฐ)

.....กรรมการสอบ

(ดร. ปานวิทย์ ชูระนุติ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การเปรียบเทียบการใช้อัลกอริทึม อพริโอรि คาร์มา และซีเควนสำหรับการวิเคราะห์พฤติกรรมนักลงทุน
นักศึกษา	นายธีระโรจน์ ศรีศาสวัตกุล
รหัสนักศึกษา	55660920
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	เทคโนโลยีสารสนเทศและการจัดการ
ปีการศึกษา	2556
อาจารย์ที่ปรึกษา	ผศ.ดร. กันต์พงษ์ วรรณปัญญา

บทคัดย่อ

เนื่องจากการแข่งขันทางธุรกิจที่สูง และสถานะตลาดหลักทรัพย์มีความผันผวน เพื่อให้ธุรกิจหลักทรัพย์ประสบความสำเร็จ จึงต้องความเข้าใจเกี่ยวกับพฤติกรรมของลูกค้ำ โดยอาศัยเทคโนโลยีดาต้าไมน์นิ่งในการวิเคราะห์พฤติกรรมบางอย่างที่คาดไม่ถึง ซึ่งในโครงการการเปรียบเทียบการใช้ อัลกอริทึม อพริโอริ คาร์มา ซีเควน สำหรับการวิเคราะห์พฤติกรรมนักลงทุน สามารถช่วยวิเคราะห์ข้อมูล ได้อย่างรวดเร็ว แม่นยำ และมีประสิทธิภาพ เพื่อให้เกิดประโยชน์ต่อธุรกิจในการบริหารจัดการหลักทรัพย์ แล้วนำสารสนเทศที่ได้มาใช้ในการตัดสินใจ โดยการวิเคราะห์ความสัมพันธ์ของข้อมูล จาก โปรแกรม IBM SPSS Modeler มาเป็นเครื่องมือช่วยในการดึงข้อมูลจากฐานข้อมูลที่เก็บอยู่ในอดีต และแสดงพฤติกรรมของนักลงทุนในตลาดหลักทรัพย์ให้เกิดองค์ความรู้ใหม่การศึกษาครั้งนี้ยังสามารถนำผลการวิเคราะห์เพื่อใช้ประกอบการตัดสินใจสำหรับกองทุน

Title	A Comparison of Using Apriori,Carmaand Sequene for Analyzing Behavioral Traders
Student	Mr.Teeraroj Srisaswattakun
Student ID	55660920
Degree	Master of Science
Program	Information Technology
Major	Information Technology and Management
Academic Year	2013
Advisor	Asst.Prof.Dr.Kuntpong Woraratpanya

ABSTRACT

Owing to the high business competition and unstable stock exchange condition, understanding customer behavior is the key factor to make the stock exchange successful. Data mining is efficient technology for analyzing unexpected customer behaviors. Thus, in this project a comparison of Apriori, Carma and Sequence algorithms is studied for analyzing customer behaviors. IBM SPSS Modeler program is a tool for pull out information kept in the past to show trader behaviors in stock markets to create the new knowledge for firms. The findings of this study help support in decisionmaking for the funds.

กิตติกรรมประกาศ

การศึกษาอิสระในหัวข้อเรื่อง “การเปรียบเทียบการใช้อัลกอริทึม อพริโอรี คาร์มา ซีควน สำหรับการวิเคราะห์พฤติกรรมนักลงทุน” จะไม่สามารถดำเนินการมาจนแล้วเสร็จได้ หากขาดความช่วยเหลือจาก ศศ.ดร. กนต์พงษ์ วรรณปัญญา ผู้เป็นอาจารย์ที่ปรึกษาโครงการที่กรุณาให้คำปรึกษา และคำแนะนำในด้านต่างๆ ข้าพเจ้ามีความรู้สึกขอบคุณทุกท่านที่มีส่วนช่วยเหลือในด้านต่างๆ ด้วยความจริงใจ หากขาดบุคคลที่กล่าวถึงดังต่อไปนี้ไป ก็จะส่งผลกระทบต่อความสำเร็จของโครงการศึกษาระดับพิเศษฉบับนี้

ขอขอบพระคุณบิดา มารดาที่ให้โอกาส กำลังใจและคอยสนับสนุนในทุกๆ เรื่อง

ข้าพเจ้าหวังเป็นอย่างยิ่งว่าบทความนี้จะเป็แนวคิดในการปฏิบัติงานเพื่อสามารถนำไปประยุกต์กับงานด้านอื่นๆ ได้เป็นอย่างดี



ธีระโรจน์ ศรีสาส์นกุล

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ	
1.1 ความเป็นมา และความสำคัญของพฤติกรรมนักลงทุน.....	1
1.2 วัตถุประสงค์.....	1
1.3 ขอบเขตการดำเนินงาน.....	1
1.4 กรอบแนวคิดในการดำเนินงาน.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง	
2.1 พฤติกรรมของนักลงทุน.....	3
2.2 คาด้าโมนิ่ง.....	7
2.3 เทคนิคของการหาความสัมพันธ์.....	10
2.4 งานวิจัยที่เกี่ยวข้อง.....	18
บทที่ 3 การดำเนินการ	
3.1 แหล่งข้อมูล.....	20
3.2 การกำหนดเป้าหมายของพฤติกรรมนักลงทุน.....	25
3.3 การจัดเตรียมข้อมูล.....	25
3.4 ขั้นตอนการดำเนินการนำข้อมูลเข้าสู่โปรแกรม.....	26
บทที่ 4 การดำเนินงานวิเคราะห์ข้อมูล	
4.1 การนำข้อมูลเข้า.....	28
4.2 ซอฟต์แวร์ที่นำมาวิเคราะห์และการประยุกต์ใช้งานด้วย IBM SPSS Modeler.....	30
4.3 ผลของการวิเคราะห์พฤติกรรมนักลงทุน และอัลกอริทึม.....	33

สารบัญ(ต่อ)

หน้า

บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ	
5.1 สรุปผลการศึกษา.....	36
5.2 ข้อเสนอแนะ	36
บรรณานุกรม.....	37
ประวัติผู้เขียน	39



สารบัญตาราง

ตารางที่

หน้า

3.1 ตาราง tblOrderHistory.....	22
3.2 ตาราง tblCustAcc	23
3.3 ตาราง tblCommSummaryHistory.....	24
3.4 ตาราง tblDealHistory.....	24



สารบัญรูป

รูปที่	หน้า
1.1 กรอบแนวคิดในการดำเนินงาน	2
2.1 ขั้นตอนการทำคาค่าไม้นิ่ง	8
2.2 ขั้นตอนการทำงานของอัลกอริทึมอปริโอริ	11
2.3 คู่มือจากฐานข้อมูลเพื่อค้นหา L1 ของอัลกอริทึมอปริโอริ	12
2.4 กราฟสร้างกลุ่มข้อมูลที่อ้างอิง C2 ของอัลกอริทึมอปริโอริ	12
2.5 การนับค่าความถี่ของข้อมูลที่อ้างอิง C2 ของอัลกอริทึมอปริโอริ	13
2.6 ผลลัพธ์การค้นหาของกลุ่มข้อมูล L2 ของอัลกอริทึมอปริโอริ	13
2.7 การสร้างกลุ่มข้อมูลที่อ้างอิง C3 ของอัลกอริทึมอปริโอริ	14
2.8 การนับค่าความถี่ของกลุ่มข้อมูลที่อ้างอิง C3 ของอัลกอริทึมอปริโอริ	14
2.9 ผลลัพธ์การค้นหาของกลุ่มข้อมูล L2 ของอัลกอริทึมอปริโอริ	14
2.10 ผลลัพธ์การค้นหาของกลุ่มข้อมูลที่ปรากฏร่วมกันของอัลกอริทึมอปริโอริ	15
2.11 ขั้นตอนของคาร์มา	16
2.12 ข้อมูลการทำธุรกรรมของอัลกอริทึมซีเควนซ์	17
2.13 ลำดับความสัมพันธ์ของข้อมูลที่พบจากการค้นหาความสัมพันธ์แบบลำดับ	17
2.14 ผลลัพธ์จากกฎความสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย	18
2.15 ผลลัพธ์จากกฎความสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย อินโดนีเซีย	19
2.16 ผลลัพธ์แบบจำลองการทำนายตลาดหุ้น และกฎความสัมพันธ์	19
3.1 ความสัมพันธ์ตารางข้อมูลที่สำคัญของระบบซื้อขายหลักทรัพย์ย้อนหลัง	21
3.2 ตัวอย่างจากฐานข้อมูลที่นำมาใช้	26
3.3 โปรแกรม SPSS IBM Modeler	27
4.1 การเลือกฐานข้อมูลที่จะนำมาวิเคราะห์พฤติกรรม	28
4.2 การกำหนดชนิดตัวแปรของค่าแอตทริบิวต์	29
4.3 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมอปริโอริ คาร์มา และซีเควนซ์	29
4.4 การตั้งค่าของอัลกอริทึมอปริโอริ	30
4.5 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมอปริโอริ	30
4.6 การตั้งค่าของอัลกอริทึมคาร์มา	31
4.7 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมคาร์มา	31

สารบัญรูป(ต่อ)

รูปที่	หน้า
4.8 การตั้งค่าของอัลกอริทึมซีเควนซ์	32
4.9 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมซีเควนซ์	32
4.10 เวลาที่ใช้ในการประมวลผล	34
4.11 กราฟแสดงหุ้น PTT และหุ้น INTUCH.....	34
4.12 กราฟแสดงหุ้นในกลุ่มธนาคาร	35



บทที่ 1

บทนำ

1.1 ความเป็นมา และความสำคัญของปัญหา

ในปัจจุบันธุรกิจมีการแข่งขันสูงและสภาวะตลาดหลักทรัพย์มีความแปรปรวนค่อนข้างมาก ทำให้ผู้ลงทุนสามารถทำกำไร หรือขาดทุนได้ในเวลาเดียวกัน หากปราศจากความรู้ความเข้าใจพฤติกรรมลูกค้าอาจส่งผลกระทบต่อความอยู่รอด และความสำเร็จในการดำเนินธุรกิจหลักทรัพย์ได้ จึงจำเป็นต้องพึ่งพาเทคนิค และเทคโนโลยีที่สามารถเก็บข้อมูลได้ในปริมาณมาก และประมวลผลได้อย่างรวดเร็ว จึงมีแนวคิดที่จะนำข้อมูลที่มีอยู่มาใช้ในการวิเคราะห์เพื่อให้เกิดประโยชน์ต่อธุรกิจในการบริหารจัดการหลักทรัพย์ โดยการนำเทคนิคดาต้าไมน์นิ่ง มาประยุกต์ใช้เพื่อเพิ่มศักยภาพในการแข่งขันธุรกิจ

สำหรับดาต้าไมน์นิ่งเป็นวิธีการที่มีประสิทธิภาพในการวิเคราะห์ข้อมูลจำนวนมาก เพื่อให้เห็นมุมมองที่แตกต่างจากข้อมูลเดิม อีกทั้งสามารถค้นหาความสัมพันธ์ที่แอบแฝงหรือถูกซ่อนในฐานข้อมูลขนาดใหญ่ การค้นพบความสัมพันธ์ของข้อมูลสามารถนำไปสู่การค้นพบพฤติกรรมบางอย่างที่คาดไม่ถึงแล้วนำสารสนเทศที่ได้มาใช้ในการตัดสินใจของผู้เกี่ยวข้องให้เกิดประสิทธิภาพ

ดังนั้นหากต้องการทราบพฤติกรรมการซื้อขายของนักลงทุน จำเป็นต้องพึ่งกระบวนการทำดาต้าไมน์นิ่ง เพื่อช่วยวิเคราะห์ข้อมูลได้อย่างรวดเร็ว แม่นยำ มีประสิทธิภาพ โดยวิธีการอัลกอริทึมอพริโอรคาร์มา ซีควนซ์ที่เหมาะสมกับการวิเคราะห์ความสัมพันธ์ของข้อมูล และโปรแกรม IBM SPSS Modeler เป็นเครื่องมือช่วยในการดึงข้อมูลจากฐานข้อมูลที่เก็บอยู่ในอดีต และแสดงพฤติกรรมของนักลงทุนในตลาดหลักทรัพย์ให้เกิดองค์ความรู้ใหม่

1.2 วัตถุประสงค์

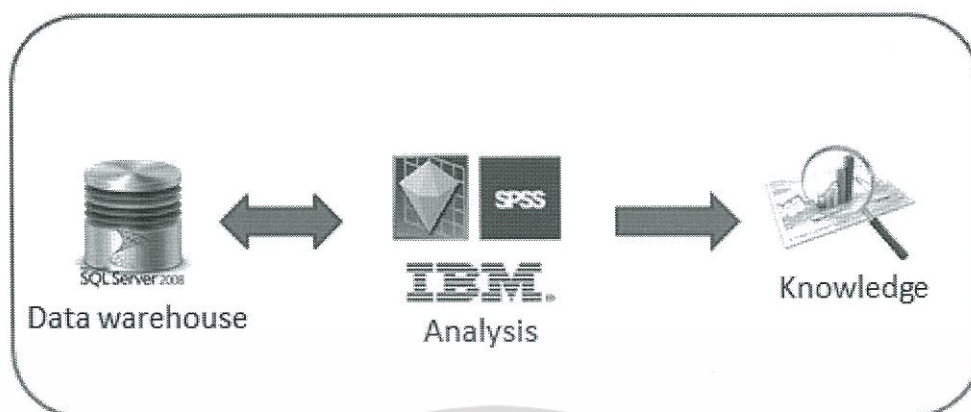
1. เพื่อวิเคราะห์สารสนเทศด้วยเทคนิคของดาต้าไมน์นิ่งในแบบความสัมพันธ์ข้อมูล
2. เพื่อสกัดองค์ความรู้ใหม่ของพฤติกรรมของนักลงทุนในบริษัทหลักทรัพย์

1.3 ขอบเขตการดำเนินงาน

1. ทำการหาความสัมพันธ์ของข้อมูลด้วยอัลกอริทึมอพริโอรคาร์มา ซีควนซ์
2. โปรแกรมที่ใช้ในการทำดาต้าไมน์นิ่งคือ IBM SPSS Modeler
3. ฐานข้อมูล Microsoft SQL Server 2008

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 กรอบแนวคิดในการดำเนินงาน



รูปที่ 1.1 กรอบแนวคิดในการดำเนินงาน

ขั้นตอนของกรอบแนวคิด

1. หาแหล่งข้อมูลที่เหมาะสมสำหรับการวิเคราะห์
2. นำข้อมูลจากฐานข้อมูลขนาดใหญ่เข้าสู่กระบวนการค้าไม้หนึ่ง
3. แสดงผลลัพธ์ที่ได้มาใช้ในการตัดสินใจ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นแนวทางในการประยุกต์การใช้งานค้าไม้หนึ่งเพื่อวิเคราะห์พฤติกรรมนักลงทุน
2. ธุรกิจหลักทรัพย์สามารถนำผลการวิเคราะห์เพื่อใช้ประกอบการตัดสินใจสำหรับกองทุน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 พฤติกรรมของนักลงทุน

การลงทุน หมายถึง การนำเงินที่เก็บสะสมไปสร้างผลตอบแทนในระยะเวลาที่ยาวที่ผลตอบแทนที่ได้รับมีมูลค่าสูงกว่าการออม โดยการลงทุนในพันธบัตรรัฐบาล หุ้นกู้ หุ้น กองทุนรวมกองทุนส่วนบุคคล และกองทุนสำรองเลี้ยงชีพ ซึ่งจะมีความเสี่ยงมากหรือน้อยตามประเภท และลักษณะของหลักทรัพย์ที่ลงทุน ซึ่งถือว่ามีความเสี่ยงสูงกว่าการฝากเงิน ผลตอบแทนที่ได้รับอาจเป็นดอกเบี้ย เงินปันผลหรือผลกำไรหรือขาดทุน (ตลาดหลักทรัพย์แห่งประเทศไทย, 2002)

การลงทุนทางการเงิน หมายถึง การที่ผู้ลงทุนนำเงินที่มีอยู่ไปซื้อหลักทรัพย์ต่างๆ ซึ่งหลักทรัพย์ดังกล่าวก่อให้เกิดรายได้กับผู้ลงทุนนั้น การลงทุนทางการเงินโดยทั่วไปมักจะทำผ่านกลไกของตลาดการเงิน เช่น ตลาดหลักทรัพย์แห่งประเทศไทย เพื่อเป็นสื่อกลางในการแลกเปลี่ยนเงิน ในการลงทุนทางการเงินเพื่อที่จะได้รับผลตอบแทนทางการเงิน (Monetary Return) จากการลงทุนในรูปแบบของดอกเบี้ย (Interest) เงินปันผล (Dividend) กำไรจากการซื้อขายหุ้น (Capital Gain) และสิทธิพิเศษอื่น ๆ (ชาติรี โรจนอากาศ, 2010)

นักลงทุนส่วนใหญ่มีพฤติกรรมในการเปรียบเทียบสิ่งๆ ที่ตัวเองมีความต้องการโดยอาศัยประสบการณ์ที่เคยพบเจอในอดีต และมีความพยายามที่จะค้นหาหลักฐานหรือตัวอย่างของเหตุการณ์ที่เคยเกิดขึ้นมา เพื่อเป็นเงื่อนไขในการพยากรณ์เหตุการณ์ที่จะเกิดขึ้นเพื่อนำมาช่วยในการตัดสินใจ เช่น “ความเชื่อที่ว่าเมื่อเดือนที่แล้วได้ลงทุนหุ้นของ PTT ไปแล้วผลกำไรเป็นอย่างดี ดังนั้นแล้วถ้าเดือนปัจจุบันนี้จะซื้อหุ้นของ PTT เพิ่มจากเดือนที่แล้วอาจจะได้ผลกำไรที่มากกว่าเดิม” เป็นต้น อย่างไรก็ตามอย่างไรก็ดีเงื่อนไขง่าย ๆ เหล่านี้ อาจไม่ใช่ปัจจัยบ่งชี้เหตุการณ์ที่จะเกิดตามมาได้ดีเพียงพอ เนื่องจากตัวอย่างที่นำมาเปรียบเทียบอาจไม่ใช่ตัวอย่างที่ดี และปรากฏการณ์ต่างๆ ที่เกิดขึ้นอาจมาจากหลายสาเหตุ หรือสาเหตุเดียวกันก็อาจทำให้เกิดผลที่ตามมาได้หลากหลาย การใช้เงื่อนไขง่าย ๆ เพียงหนึ่งหรือสองเงื่อนไขจึงไม่สามารถพยากรณ์เหตุการณ์ที่จะเกิดตามมาได้อย่างแม่นยำ

ตัวอย่างของ Representativeness ที่มักพบเห็นได้บ่อยในกระบวนการตัดสินใจลงทุนมีดังต่อไปนี้

1. การลงทุนในหุ้นของบริษัทที่ดีจะเป็นการลงทุนที่ให้ผลกำไร บริษัทที่มีความแข็งแกร่ง และผลประกอบการที่ดีอย่างสม่ำเสมอเป็นทางเลือกในการลงทุนที่ดี แต่นักลงทุนก็สามารถ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขาดทุนมหาศาลจากบริษัทที่มีผลประกอบการที่ดีได้ หากเข้าซื้อหุ้นดังกล่าวที่ราคาสูงเกินไป หรือเข้าซื้อหุ้นในขณะที่ผลกำไรของบริษัทปรับตัวสูงสุด (Peak of the Business Cycle) และมีแนวโน้มที่จะแยกลง การลงทุนจึงควรพิจารณาโดยการวิเคราะห์สิ่งที่จะเกิดในอนาคตมากกว่าอดีตที่ผ่านมาแล้ว

2. กลยุทธ์การลงทุนแบบโมเมนตัม (Momentum Strategy) คือ กลยุทธ์การลงทุนที่ไล่ซื้อหุ้นที่ราคากำลังปรับตัวขึ้น โดยคาดว่าราคาหุ้นจะปรับตัวขึ้นไป แม้ว่าการลงทุนแบบโมเมนตัมอาจเป็นทางเลือกหนึ่งที่จะทำกำไรให้นักลงทุนได้ แต่หากการปรับตัวขึ้นของราคาหุ้นไม่ได้เกิดขึ้นจากการเปลี่ยนแปลงของปัจจัยที่เกี่ยวข้องในระยะยาวเป็นเพียงพฤติกรรมแห่ตามกันอย่างขาดเหตุผล การลงทุนแบบโมเมนตัมก็อาจจบลงด้วยผลขาดทุนที่คาดไม่ถึงได้นอกจากนี้งานวิจัยในต่างประเทศก็ยังระบุว่าหุ้นที่มีผลการดำเนินงานดีกว่าตลาด (Outperform Market) ในช่วง 2-3 ปีก่อนหน้า จะมีผลการดำเนินงานที่แย่กว่าตลาดในช่วง 2-3 ปีต่อไป (Underperform Market) จึงไม่อาจรับประกันได้ว่าหุ้นที่ให้ผลตอบแทนที่ดีในอดีตจะให้ผลกำไรอย่างต่อเนื่องในอนาคต

กล่าวโดยสรุป Representativeness เป็นพฤติกรรมที่จะลดทอนประสิทธิภาพการเก็บรวบรวมข้อมูล และการวิเคราะห์โดยใช้เหตุผลของนักลงทุน ทำให้การตัดสินใจลงทุนขาดความรอบคอบ และมักทำให้เกิดการขาดทุนอย่างไม่คาดหมาย การตัดสินใจลงทุนจึงควรวิเคราะห์สาเหตุปัจจัยพื้นฐาน ภาวะตลาด และสภาพแวดล้อมการลงทุน เป็นต้นอย่างรอบคอบเพื่อคาดการณ์ผลลัพธ์ที่จะเกิดตามมามากกว่าการใช้เงื่อนงำง่าย ๆ เป็นเกณฑ์ในการตัดสินใจ

Fear of Regret คือ พฤติกรรมที่นักลงทุนกลัวที่จะต้องยอมรับว่าการตัดสินใจของตนผิดพลาด เป็นเหตุให้ไม่ตัดสินใจขายหุ้นที่ตนลงทุนผิดพลาดออกไป นักลงทุนจึงขาดทุนจากการไม่ยอมรับความจริงนั้นมากขึ้นเรื่อยๆ คำกล่าวที่มักพบบ่อยในกรณีนี้ เช่น “ไม่ขายไม่ขาดทุน” “ขาดทุนอย่าขาย” ในความเป็นจริงหากผลขาดทุนที่เกิดขึ้นนั้นมาจากปัจจัยที่ส่งผลกระทบระยะยาว เช่น ปัจจัยพื้นฐานของบริษัทที่แยกลงและกำลังจะแย่ต่อไปอีก การถือครองหุ้นนั้นต่อไปจะยิ่งทำให้ขาดทุนมากขึ้น การตัดสินใจขายหุ้นนั้นออกไปเพื่อไปลงทุนในหุ้นตัวอื่นที่ให้ผลตอบแทนดีกว่ามาชดเชย หรือขายออกไปก่อนเพื่อรอรับหุ้นนั้นกลับในราคาที่ถูกลงกว่าจึงเป็นทางเลือกที่สมเหตุสมผลกว่า

Seeking of Pride คือ พฤติกรรมที่นักลงทุนรู้สึกยึดมั่นเมื่อรู้ว่าการตัดสินใจของตนถูกต้อง จึงพยายามขายหุ้นเพื่อรับรู้กำไร โดยละเลยการพิจารณาปัจจัยแวดล้อมที่เกี่ยวข้อง เป็นเหตุให้นักลงทุนขายหุ้นที่ดีเร็วเกินไปทำให้เสียโอกาสที่จะทำกำไรจำนวนมากจากหุ้นนั้น Fear of Regret และ Seeking of Pride เป็นพฤติกรรมที่กดดันให้นักลงทุนขายหุ้นที่เข้าขาเกินไป (Hold Loser Too Long) และขายหุ้นที่ดีเร็วเกินไป (Sell Winner Too Early) ซึ่งอาจลดทอนผลกำไรจากการลงทุนอย่างมีนัยสำคัญ นักลงทุนจึงควรตัดความรู้สึกดังกล่าวออกจากกระบวนการตัดสินใจลงทุน โดยยึดมุมมองต่อหุ้นเหล่านั้นในอนาคตเป็นหลักในการตัดสินใจ นักลงทุนพึงระลึกอยู่เสมอว่าหากราคาเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หุ้นที่ถือครองปรับตัวลดลงต่ำกว่าราคาซื้อแปลว่านักลงทุนได้ขาดทุนไปแล้ว แม้ว่านักลงทุนจะยังไม่ขายหุ้นออกไปก็ตาม ดังนั้นหากนักลงทุนเชื่อว่า ราคาหุ้นดังกล่าวอาจปรับตัวลงต่อได้อีก ก็ควรจะขายหุ้นนั้นออกไปไม่ว่าจะซื้อมาที่ราคาเท่าไร (Stop Loss) ในทางกลับกันหากนักลงทุนเชื่อว่าหุ้นดังกล่าวจะปรับตัวสูงขึ้น ควรถือครองหุ้นนั้นๆต่อไปไม่ว่าจะได้กำไรมาแล้วเท่าใดก็ตาม (Let Profit Run) พฤติกรรมการลงทุนเพื่อให้ประสบความสำเร็จ แบ่งออกเป็น 3 ช่วงด้วยกัน ได้แก่

ก่อนการลงทุน

- ศึกษาหาความรู้ความเข้าใจเกี่ยวกับวิธีการซื้อขาย
- การกระจายการลงทุนในทรัพย์สินที่หลากหลาย
- วางแผนแนวทางหรือหลักเกณฑ์การลงทุน
- การกำหนดอัตราขาดทุนที่ยอมรับได้
- ช่องทางการซื้อขายหลักทรัพย์

ช่วงการลงทุน

- ปัจจัยในการพิจารณาประกอบการตัดสินใจ
- การทยอยซื้อหลักทรัพย์
- รูปแบบการซื้อขายหลักทรัพย์
- ความถี่ในการซื้อขายหลักทรัพย์
- ปฏิบัติได้ตามแนวทาง

หลังการลงทุน

- บันทึกรายการซื้อขายหลักทรัพย์
- ความถี่การติดตามข้อมูลบทวิเคราะห์

ปัจจัยที่มีผลกระทบต่อราคาหุ้นผู้ลงทุนต้องเข้าใจถึงปัจจัยที่เป็นตัวกำหนดราคาหุ้น หรือปัจจัยที่ส่งผลกระทบต่อราคาหุ้นซึ่งแนวคิดที่จะช่วยให้ผู้ลงทุนตอบคำถามดังกล่าวได้ และช่วยให้ผู้ลงทุนเข้าใจถึงปัจจัยที่เป็นตัวกำหนดราคาหุ้น คือ การหามูลค่าที่แท้จริง (Intrinsic Value) ของหุ้นสามัญ (ตลาดหลักทรัพย์แห่งประเทศไทย, 2014)

การหามูลค่าที่แท้จริงของหุ้นเกิดจากความเชื่อที่ว่า ผู้ลงทุนที่ต้องการลงทุนในหุ้นวันนี้ ย่อมคาดหวังผลประโยชน์หรือผลตอบแทนที่คาดว่าจะเกิดขึ้นในอนาคตจากการลงทุนในหุ้นดังกล่าว ดังนั้น ราคาหุ้นที่ผู้ลงทุนยอมจ่ายในวันนี้ จึงเป็นราคาสำหรับสิ่งที่คาดว่าจะได้รับในอนาคตที่เปลี่ยนแปลงไป น่าจะเกิดจากการที่ผู้ลงทุนคาดการณ์ว่า สิ่งที่คุณจะได้รับในอนาคต ทั้งในรูปของ เงินปันผล และ ส่วนต่างราคาที่เปลี่ยนแปลงไปจากเดิม เช่น ถ้ามีผู้คาดการณ์ว่าการลงทุนในหุ้น ABC จะได้รับผลตอบแทนในรูปของเงินปันผลมากขึ้น ก็จะทำให้บุคคลนั้นยินดีที่จะจ่ายเงินเพื่อซื้อหุ้น ABC ในราคาที่สูงขึ้น ราคาหุ้น ABC ก็จะปรับตัวสูงขึ้นถ้าผู้ลงทุนคนหนึ่งคาดการณ์ว่าหากลงทุนในหุ้น ABC ณ ปัจจุบัน โดยถือไปอีก 2 – 3 เดือนแล้วขายต่อ อาจจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ราคาขายต่อน้อยกว่าตอนที่ซื้อมา ผู้ลงทุนคนนั้นก็จะไม่ยอมจ่ายเงินเพื่อซื้อหุ้น ABC ณ ราคาปัจจุบัน แต่จะยอมจ่ายในราคาที่ต่ำกว่า เพื่อชดเชยกับความเสี่ยงที่คาดว่าตนเองจะเผชิญในอีก 2 – 3 เดือนข้างหน้า ราคาหุ้น ABC ก็จะปรับตัวลดลง คือ ราคาหุ้นปัจจุบันจะถูก

ผลตอบแทน และความเสี่ยงแนวโน้มผลกำไรในอนาคตของบริษัท ซึ่งเกิดจากผลการดำเนินงานของบริษัท รูปแบบธุรกิจของบริษัท การก่อกำเนิดของบริษัท รวมทั้งแนวโน้มภาวะอุตสาหกรรม ภาวะเศรษฐกิจ และตลาดหุ้นปัจจัยต่างๆ ที่กำหนดราคาหุ้นเหล่านี้ ล้วนเป็น พื้นฐาน สำคัญที่สามารถนำมาจัดกลุ่มเป็น 3 ปัจจัยใหญ่ๆ คือ ภาวะเศรษฐกิจ ภาวะอุตสาหกรรม และผลการดำเนินงานของบริษัท ดังนั้น หากคุณต้องการวิเคราะห์ปัจจัยที่กำหนดราคาหุ้น ก็ต้องวิเคราะห์ทั้ง 3 ปัจจัยนี้ ซึ่งเราเรียกการวิเคราะห์หลักทรัพย์ตามแนวคิดนี้ว่า “การวิเคราะห์ปัจจัยพื้นฐาน” (Fundamental Analysis)

ภาวะเศรษฐกิจ ถ้าเศรษฐกิจดี ราคาหุ้นมักจะเพิ่มขึ้น เพราะผู้ลงทุนมองว่าโอกาสที่บริษัทจะขาดทุนในช่วงนี้มีน้อย จึงกล้าที่จะลงทุนในหุ้นเพิ่มขึ้น กล่าวคือ ในช่วงที่เศรษฐกิจดี ประชาชนจะจับจ่ายใช้สอยเพิ่มขึ้น ส่งผลให้ยอดขายและผลกำไรของบริษัทเพิ่มขึ้น เมื่อบริษัทมีกำไรเพิ่มขึ้น บริษัทก็จะสามารถจ่ายผลตอบแทนให้แก่ผู้ลงทุนได้มากขึ้นตามไปด้วยในทางตรงกันข้าม ถ้าเศรษฐกิจไม่ดี ราคาหุ้นก็มักจะลดลง เพราะผู้ลงทุนมีความกังวลในผลประโยชน์ของบริษัท กล่าวคือ ช่วงที่เศรษฐกิจไม่ดี กำลังซื้อของประชาชนลดลง ยอดขายของบริษัทจะลดลง กำไรของบริษัทก็จะลดลงหรืออาจจะขาดทุน ท้ายที่สุดผู้ลงทุนก็จะไม่ได้รับการจัดสรรกำไรจากบริษัท เหตุการณ์ดังกล่าวทำให้ผู้ลงทุนรู้สึกถึงความเสี่ยงจากการลงทุน จึงมีการขายหุ้นออกมา เมื่อมีการขายหุ้นออกมามาก ก็จะทำให้ราคาหุ้นปรับตัวลดลงในที่สุด

ภาวะอุตสาหกรรม ถ้าอุตสาหกรรมใดอยู่ในช่วงขาขึ้น ราคาหุ้นที่อยู่ในอุตสาหกรรมนั้นมักจะปรับตัวเพิ่มขึ้น เนื่องจากผู้ลงทุนคาดการณ์ว่าผลประโยชน์ของบริษัทในอุตสาหกรรมนั้นจะปรับเพิ่มขึ้นตามภาวะอุตสาหกรรมในทางตรงกันข้าม ถ้าอุตสาหกรรมใดอยู่ในช่วงขาลง ราคาหุ้นที่อยู่ในอุตสาหกรรมนั้นก็จะมีการปรับตัวลดลง เนื่องจากผู้ลงทุนกังวลถึงความเสี่ยงจึงขายหุ้นในอุตสาหกรรมดังกล่าวออกมา

ผลการดำเนินงานของบริษัท เป็นตัวแปรที่เกี่ยวข้องกับผลการดำเนินงานของบริษัทนั้น โดยตรง ทั้งนี้ ขึ้นอยู่กับประเภทของธุรกิจ กลยุทธ์ของบริษัท และความสามารถของผู้บริหาร กล่าวคือ ในช่วงที่สภาวะเศรษฐกิจ และอุตสาหกรรมไม่ดี หากผู้บริหารของบริษัทมีความสามารถสูง ก็อาจจะทำให้ผลประโยชน์ของบริษัทลดลงไม่มาก และราคาหุ้นก็จะปรับตัวลดลงไม่มาก

2.2 ดาต้าไมน์นิ่ง

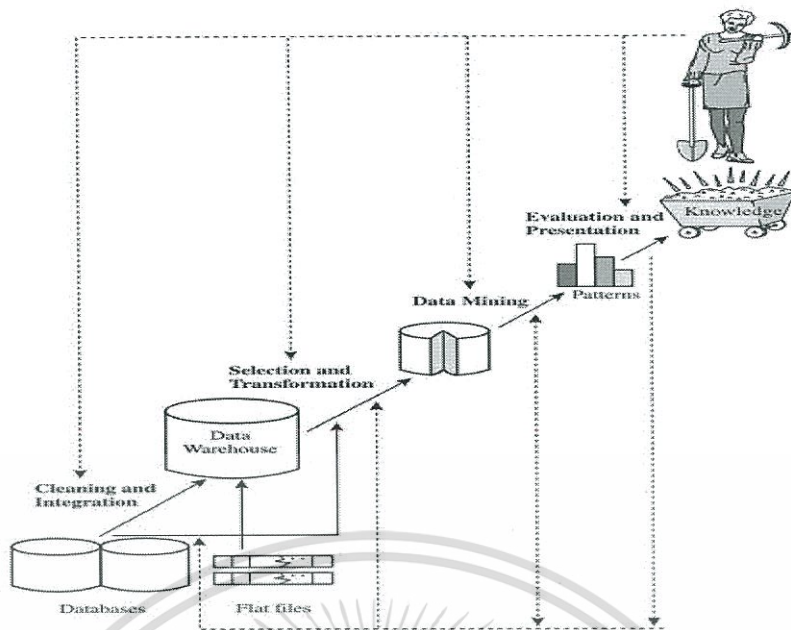
2.2.1 ความหมายและความสำคัญของดาต้าไมน์นิ่ง (Han, Jiawei, Kamber and Micheline, 2001)

ในการรวบรวมข้อมูลเพื่อดำเนินธุรกิจการลงทุนเป็นสิ่งที่จะต้องให้ความสำคัญ เนื่องจากเป็นองค์ประกอบสำคัญที่จะทำธุรกิจการลงทุนประสบความสำเร็จและช่วยสนับสนุนการตัดสินใจซึ่งจำเป็นต้องกลั่นกรองข้อมูลออกมาจากฐานข้อมูลขนาดใหญ่ที่มีอยู่จำนวนมาก

ดาต้าไมน์นิ่งคือสิ่งที่แตกต่างจากการวิเคราะห์ข้อมูลในประเภทต่างๆจะเป็นการมองความสัมพันธ์ในหลายมิติ โดยการตั้งความสัมพันธ์แล้วดูผลว่าเป็นไปตามสิ่งที่ตั้งเป้าหมายไว้ หรือการนำดาต้าไมน์นิ่งมาช่วยในการค้นหาสารสนเทศจากฐานข้อมูล โดยในปัจจุบันการดำเนินการทางธุรกิจด้านการลงทุนต้องใช้ฐานข้อมูลมาใช้ในการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับกิจกรรมต่างๆที่เกิดขึ้น ซึ่งในแต่ละวันข้อมูลจะถูกเพิ่มขึ้น และจากข้อมูลทั้งหมดที่ถูกจัดเก็บอยู่ในฐานข้อมูลอาจจะมีลักษณะบางอย่างที่แฝงอยู่เช่น ความสัมพันธ์ของข้อมูล ซึ่งสารสนเทศนี้มีประโยชน์อย่างมากในการช่วยสนับสนุนการตัดสินใจเชิงธุรกิจ เพื่อให้องค์กรมีโอกาสในการแข่งขันมากยิ่งขึ้นและจากการที่มีข้อมูลจำนวนมากในฐานข้อมูลซึ่งเป็นข้อมูลที่ซับซ้อน ทำให้การวิเคราะห์ข้อมูลให้ถูกต้องนั้นยากที่มนุษย์จะเข้าใจ มนุษย์อาจจะวิเคราะห์ข้อมูลได้ไม่ครอบคลุม หรือมองข้ามตัวแปรบางตัวที่อาจทำให้ผลของการวิเคราะห์คาดเคลื่อน แต่ถ้าใช้หลักการดาต้าไมน์นิ่งในการวิเคราะห์จะสามารถมั่นใจได้ว่าผลลัพธ์ของการวิเคราะห์นั้นถูกต้อง

2.2.2 ขั้นตอนการทำงานของดาต้าไมน์นิ่ง

กระบวนการทำดาต้าไมน์นิ่ง เป็นขั้นตอนการสร้างแบบจำลองโดยจะสร้างแบบจำลองจากกลุ่มข้อมูลเพื่อสร้างความเข้าใจในแนวโน้ม รูปแบบ ความเกี่ยวข้องสัมพันธ์กันของกลุ่มข้อมูล และสามารถทำนายข้อมูลได้ ซึ่งประกอบด้วย 5 ขั้นตอน (กฤษณะ ไวยมัย, 2549)



รูปที่ 2.1 ขั้นตอนการทำค้ำค่าไมน์นึ่ง

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

ขั้นตอนที่ 1 การกำหนดวัตถุประสงค์ทางธุรกิจ (Business Object Determination)

การกำหนดปัญหา และวัตถุประสงค์ให้ชัดเจน โดยการวิเคราะห์ข้อมูลเบื้องต้นว่ามีข้อมูลใดอยู่บ้าง ใช้สารสนเทศจากแหล่งข้อมูลใด และกำหนดว่าเมื่อไรควรใช้ค้ำค่าไมน์นึ่ง ในการแก้ไขปัญหา

ขั้นตอนที่ 2 การเตรียมข้อมูล (Data Gathering and Preparation)

การเตรียมข้อมูลเป็นขั้นตอนที่สำคัญเป็นอย่างมาก และใช้เวลาที่ค่อนข้างนาน เพราะต้องเลือกข้อมูลที่เหมาะสมกับความต้องการที่จะทำให้ธุรกิจประสบความสำเร็จ ประกอบไปด้วย 3 ขั้นตอนดังต่อไปนี้

1. การคัดเลือกข้อมูล (Data Selection) การระบุลักษณะ และการคัดเลือกข้อมูลที่ต้องการ และนำข้อมูลที่ต้องการออกไปซึ่งขึ้นอยู่กับวัตถุประสงค์ของธุรกิจที่กำหนดไว้ และแหล่งที่มาของข้อมูล ความหมายของข้อมูล ประเภทของข้อมูล ค่าที่สามารถเป็นไปได้ของข้อมูล และความทันสมัยของข้อมูล ซึ่งมีการเปลี่ยนแปลงอยู่ตลอดเวลาเนื่องจากการเปลี่ยนแปลงของเวลา และสถานการณ์ภายนอกมีอยู่ตลอดเวลา และมีผลทำให้ประสิทธิภาพของข้อมูลลดลง นอกจากนี้ยังมีประเด็นที่สำคัญที่ต้องพิจารณาอีก 4 ประการคือ

1.1 ระดับของข้อมูลโดยขึ้นอยู่กับวัตถุประสงค์ในการทำว่าเป็นข้อมูลที่สรุปแล้วหรือว่าเป็นข้อมูลระดับรายการ (Item)

1.2 ลักษณะการจัดเก็บของข้อมูล ซึ่งขึ้นอยู่กับภาษาคอมพิวเตอร์ และระบบปฏิบัติการที่ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ความแตกต่างของข้อมูลในแต่ละแหล่งทั้งรูปแบบความหมาย และลักษณะการจัดเก็บ

1.4 ประเภทข้อความซึ่งค่าของข้อมูลที่จัดเก็บอาจเกิดจากความผิดพลาดในขั้นตอนการบันทึกข้อมูล (Data Entry)

2. การกรองข้อมูล (Data Preprocessing) เป็นกระบวนการนำข้อมูลผ่านขั้นตอนการคัดเลือกข้อมูลมาแล้ว มาทำการกลั่นกรองข้อมูล (Data Cleaning) เพื่อคัดกรองข้อมูลที่ไม่เหมาะสม โดยพิจารณาใน 2 ประเด็น คือ Noisy Data คือ ข้อมูลมีลักษณะแตกต่างจากที่คาดการณ์ หรืออาจเกิดจากความผิดพลาดในการบันทึกข้อมูล ในการตรวจสอบแก้ไขให้ถูกต้องนั้นต้องอาศัยวิธีการต่างๆ เข้าช่วยเพื่อทำ Noise Detection เพื่อไม่นำข้อมูลนั้นมาวิเคราะห์

3. การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการใช้งานที่พร้อมจะนำไปวิเคราะห์ตามอัลกอริทึมที่จะนำมาใช้ในการทำค้ำไมน์นิ่ง ซึ่งจะมีลักษณะแตกต่างกันไป

ขั้นตอนที่ 3 การทำ Mining ข้อมูล (Data Mining)

เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่มีการเลือกอัลกอริทึมที่จะนำมาใช้ให้เหมาะสมกับลักษณะปัญหา และลักษณะของข้อมูลที่ได้ดำเนินการไว้ในขั้นตอนที่ผ่านมา แล้วทำการประมวลผลข้อมูลตามอัลกอริทึมที่เลือกไว้ ซึ่งในขั้นตอนนี้อาจต้องย้อนกลับไปทำขั้นตอนที่เกี่ยวข้องใหม่ โดยจะพัฒนากระบวนการในขั้นตอนต่างๆ เพื่อให้ได้ผลที่มีประสิทธิภาพ และมีความถูกต้องมากขึ้น

ขั้นตอนที่ 4 การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Result)

การวิเคราะห์ผลลัพธ์ที่ได้เป็นขั้นตอนที่มีความสำคัญมาก โดยจะทำการแปลความหมาย และประเมินผลลัพธ์ที่ได้จากขั้นตอนการทำค้ำไมน์นิ่งทั้งนี้จะต้องอาศัยทักษะทั้งในส่วนของการวิเคราะห์ข้อมูลและการวิเคราะห์ทางธุรกิจร่วมกัน รวมถึงเครื่องมือด้าน Graphical Visualization ซึ่งช่วยให้การวิเคราะห์ข้อมูลทำได้สะดวก และรวดเร็วขึ้น นอกจากนี้ยังพิจารณาสารสนเทศที่ได้มาว่ามีความสนใจ ความสมบูรณ์ และสามารถนำไปใช้ได้หรือไม่ หากผลที่ได้ออกมาไม่เป็นตามที่คาดการณ์ไว้ อาจต้องมีการกลับไปพิจารณาขั้นตอนต่างๆ ใหม่อีกครั้งเพื่อแก้ไขโมเดลใหม่

ขั้นตอนที่ 5 การนำสารสนเทศที่ได้ไปใช้ประโยชน์ (Assimilation of Knowledge)

การนำสารสนเทศที่ได้ไปใช้ประโยชน์ โดยการรวบรวมความรู้ความเข้าใจทางธุรกิจ ซึ่งเกี่ยวข้องโดยตรงกับวัตถุประสงค์ทางธุรกิจ ทั้งนี้มีหลักที่ต้องพิจารณา 2 ประการ คือ การนำเสนอแนวคิดทางธุรกิจที่ค้นพบใหม่ และหาแนวทางในการที่จะใช้ความรู้ใหม่ที่ค้นพบเพื่อให้เกิดประโยชน์สูงสุด

2.3 เทคนิคของการหาความสัมพันธ์ (Association)

การแสดงความสัมพันธ์ของข้อมูลหรือวัตถุที่เกิดขึ้นพร้อมกัน เช่น การวิเคราะห์ข้อมูลการขายสินค้าโดยเก็บข้อมูลจากระบบณ.จุดขายหรือระบบขายสินค้าออนไลน์ และพิจารณาสินค้าที่อยู่ในตระกร้าเดียวกันหรือผู้ซื้อมักซื้อพร้อมกัน โดยสามารถเห็นถึงพฤติกรรมที่สอดคล้องหรือสิ่งที่ซ่อนเร้นได้

2.3.1 อัลกอริทึมอปริโอรื

อัลกอริทึมอปริโอรืเป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้กฎความสัมพันธ์ เป็นการค้นหาค่าความสัมพันธ์ระหว่างตัวแปรในฐานข้อมูลขนาดใหญ่ ตัวอย่างกฎความสัมพันธ์เช่น รองเท้ากางเกงที่พบใน ข้อมูลการขายในร้านขายเสื้อผ้า กฎระบุว่าลูกค้าที่จะซื้อเสื้อยืดและกางเกงยังมีแนวโน้มที่จะซื้อรองเท้า กฎสัมพันธ์จะใช้ในการแสดงความสัมพันธ์ ระหว่างรายการข้อมูล โดยค่าความสำคัญของกฎทำการวัดโดยใช้ข้อมูลสองตัวด้วยกันคือค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้หรือเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้องและข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่กฎมีความสัมพันธ์กันสามารถนำไปใช้ได้ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกันแต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายการกำหนดค่าความสัมพันธ์มีดังต่อไปนี้

ขั้นตอนที่ 1 เลือก Frequent Item Sets ซึ่งเป็น Item Sets ที่มีความถี่หรือค่าสนับสนุน (Support) มากกว่าหรือเท่ากับค่าความถี่ขั้นต่ำ (Minimum Support) ส่วน Item Sets ที่มีความถี่น้อยกว่าค่าความถี่ขั้นต่ำจะเรียกว่า Rare Item Sets

ขั้นตอนที่ 2 สร้าง Association Rules จาก Frequent Item Sets ที่ได้จากการดำเนินการในขั้นตอนที่ 1 และจะยอมรับ Association Rules ที่สร้างขึ้นมาก็ต่อเมื่อกฎนี้มีค่าความเชื่อมั่น (Confidence) มากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) โดยมีกฎความสัมพันธ์ที่ต้องคำนวณคือค่า สนับสนุน (Support) และค่าความเชื่อมั่น (Confidence)

หลักการทำงานของอัลกอริทึมอปริโอรื ประกอบด้วยขั้นตอนการทำงานทั้งหมด 2 ขั้นตอนด้วยกัน ขั้นตอนที่ 1 การสร้างกลุ่มข้อมูลที่ซ้ำ และขั้นตอนที่ 2 การทดสอบกลุ่มข้อมูลที่ซ้ำเหล่านั้นว่าเป็นกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยหรือไม่ ดังรูปที่ 2.2 โดยการทำงานของอัลกอริทึมอปริโอรื สามารถอธิบายได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. คู่มุ่มข้อมูลจากฐานข้อมูลครั้งแรกเพื่อนับค่าความถี่ของแต่ละกลุ่มข้อมูลที่ปรากฏทั้งหมดในฐานข้อมูล
2. ตรวจสอบค่าความถี่ของแต่ละกลุ่มข้อมูล เพื่อกำหนดค่าสนับสนุน โดยหากกลุ่มข้อมูล มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำก็จะถือว่าเป็น กลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยที่มีขนาดของกลุ่มข้อมูล 1 ข้อมูล (L_1 : Frequent 1-Itemsets)
3. นำ L_1 ที่ได้มาสร้างกลุ่มข้อมูลทำจริงที่มีขนาดของกลุ่มข้อมูล 2 ข้อมูล (C_2 : Candidate 2-Itemsets)
4. คู่มุ่มข้อมูลจากฐานข้อมูลอีกครั้งเพื่อนับค่าความถี่ของ C_2 และตัด C_2 ที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนขั้นต่ำ หาก C_2 มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำก็จะกลายเป็น L_2
5. ทำในหัวข้อที่ 3) และ 4) ซ้ำจนกว่าไม่สามารถสร้าง C_k จาก L_{k-1} ได้ เมื่อ k คือขนาดของกลุ่มข้อมูลจึงสิ้นสุดการสร้างกลุ่มข้อมูลทำจริง และจบการทำงานของอัลกอริทึมอปริโอริ ทำให้ได้กลุ่มข้อมูลที่ปรากฏร่วมกันบ่อยทั้งหมด

```

1   $L_1 = \{ \text{large 1-itemsets} \};$ 
2  For ( $k=2; L_{k-1} \neq \emptyset; k++$ ) do begin
3     $C_k = \text{apriori-gen}(L_{k-1});$  // New candidates
4    Forall transactions  $t \in D$  do begin
5       $C_t = \text{subset}(C_k, t);$  // Candidates contained in  $t$ 
6      Forall candidates  $c \in C_t$  do
7         $c.\text{count}++;$ 
8      End
9     $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{minsup} \};$ 
10 End
11 Answer =  $\bigcup_k L_k;$ 

```

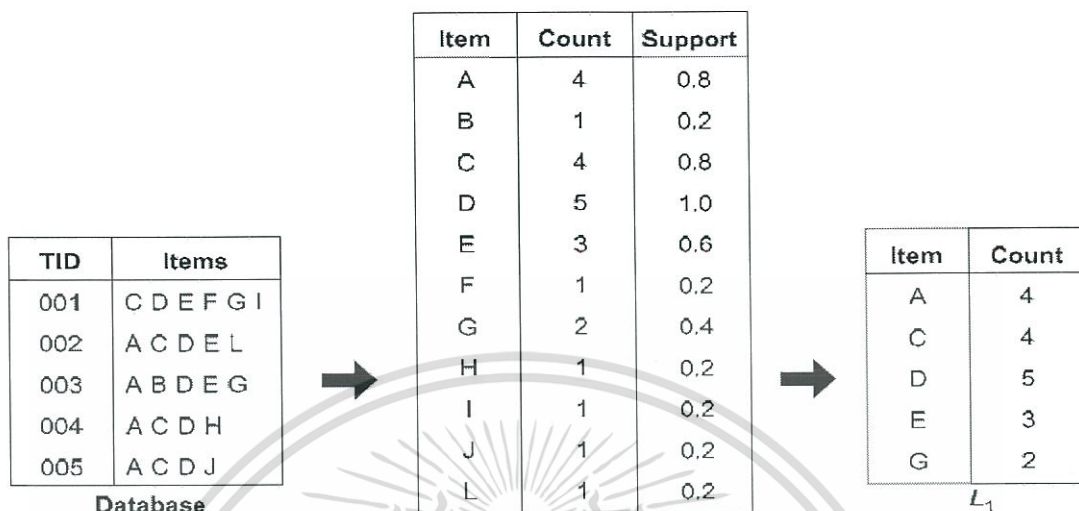
รูปที่ 2.2 การทำงานของอัลกอริทึมอปริโอริ

2.4.1 ตัวอย่างการทำงานของอัลกอริทึมอปริโอริ

การค้นหากลุ่มข้อมูลที่ปรากฏบ่อย โดยกำหนดค่าสนับสนุนขั้นต่ำเป็น 0.4 (มีค่าความถี่อย่างน้อยเท่ากับ 2 รายการข้อมูลปรากฏในฐานข้อมูล) โดยตัวอย่างการทำงานของอัลกอริทึมอปริโอริสามารถอธิบายได้ดังนี้

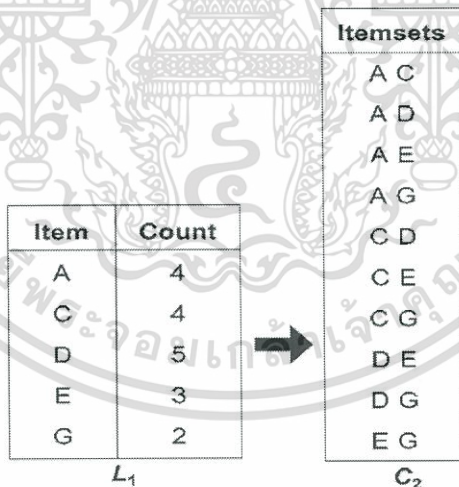
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. คูกลุ่มข้อมูลจากฐานข้อมูลเพื่อนับค่าความถี่ของแต่ละข้อมูล และคำนวณค่าสนับสนุน จากค่าความถี่ที่ได้ พร้อมทั้งตัดคูกลุ่มข้อมูลที่ไม่ถึงค่าสนับสนุนขั้นต่ำที่ได้กำหนดไว้ โดยผลลัพธ์ที่ได้จะเรียกว่า L_1 ดังรูปที่ 2.3



รูปที่ 2.3 คูกข้อมูลจากฐานข้อมูลเพื่อกค้นหา L_1 ของอัลกอริทึมอปริโอริ

2. นำ L_1 ที่ได้ไปสร้างคูกข้อมูลทำซิง C_2 โดยวิธีการสร้าง C_2 นั้นจะจับคู่กัน ของคูกข้อมูลทุกตัวที่เป็น L_1 ในทุกกรณีที่เป็นไปได้ที่ประกอบด้วยคูกข้อมูล 2 ข้อมูล ดังรูปที่ 2.4



รูปที่ 2.4 การสร้างคูกข้อมูลทำซิง C_2 ของอัลกอริทึมอปริโอริ

3. หลังจากสร้างคูกข้อมูลทำซิง C_2 ได้แล้ว ต่อไปคืออ่านคูกข้อมูลจากฐานข้อมูลเพื่อนับค่าความถี่ของ C_2 ดังรูปที่ 2.5

TID	Items
001	C D E F G I
002	A C D E L
003	A B D E G
004	A C D H
005	A C D J

Database

→

Itemsets	Count
AC	3
AD	4
AE	2
AG	1
CD	4
CE	2
CG	1
DE	3
DG	2
EG	2

C₂

รูปที่ 2.5 การนับค่าความถี่ของข้อมูลทำซ้ำ C₂ ของอัลกอริทึมเอพริโอรี

4. นำค่าความถี่ของ C₂ ที่ได้ไปคำนวณหาค่าสนับสนุนเพื่อใช้ในการตัดสินใจให้เป็นกลุ่มข้อมูล L₂ ผลลัพธ์ที่ได้ ดังรูปที่ 2.6

Itemsets	Count	Support
AC	3	0.6
AD	4	0.8
AE	2	0.4
AG	1	0.2
CD	4	0.8
CE	2	0.4
CG	1	0.2
DE	3	0.6
DG	2	0.4
EG	2	0.4

C₂

→

Itemsets	Count
AC	3
AD	4
AE	2
CD	4
CE	2
DE	3
DG	2
EG	2

L₂

รูปที่ 2.6 ผลลัพธ์การค้นหากลุ่มข้อมูล L₂ ของอัลกอริทึมเอพริโอรี

5. นำกลุ่มข้อมูล L₂ ที่ได้ไปสร้างเป็นกลุ่มข้อมูล C₃ โดยวิธีสร้างจะจับคู่ร่วมกันของกลุ่มข้อมูลทุกตัวที่เป็น L₂ ที่มีกลุ่มข้อมูลตัวแรกเหมือนกัน โดยผลลัพธ์ของการจับคู่จะประกอบด้วยกลุ่มข้อมูล 3 ข้อมูลดังรูปที่ 2.7

Itemsets	Count
AC	3
AD	4
AE	2
CD	4
CE	2
DE	3
DG	2
EG	2

L_2

→

Itemsets
ACD
ACE
ADE
CDE
DEC

C_3

รูปที่ 2.7 การสร้างกลุ่มข้อมูลทำชิง C_3 ของอัลกอริทึมอปริโอรี

6. หลังจากสร้างกลุ่มข้อมูลทำชิง C_3 แล้ว ขั้นตอนต่อไป อ่านกลุ่มข้อมูลจากฐานข้อมูลเพื่อ นับค่าความถี่ของ C_3 ดังรูปที่ 2.8

TID	Items
001	CDEFGI
002	ACDEL
003	ABDEG
004	ACDH
005	ACDJ

Database

→

Itemsets	Count
ACD	3
ACE	1
ADE	2
CDE	2
DEC	2

C_3

รูปที่ 2.8 การนับค่าความถี่ของกลุ่มข้อมูลทำชิง C_3 ของอัลกอริทึมอปริโอรี

7. กำหนดค่าสนับสนุนจากค่าความถี่ที่ได้ พร้อมทั้งตัดกลุ่มข้อมูลทำชิง C_3 ที่ไม่ผ่านค่าสนับสนุนขั้นต่ำ ซึ่งผลลัพธ์ที่ได้จะเรียกว่า L_3 ดังรูปที่ 2.9

Itemsets	Count	Support
ACD	3	0.6
ACE	1	0.2
ADE	2	0.4
CDE	2	0.4
DEC	2	0.4

C_3

→

Itemsets	Count
ACD	3
ACE	2
CDE	2
DEC	2

L_3

รูปที่ 2.9 ผลลัพธ์การค้นหากลุ่มข้อมูล L_3 ของอัลกอริทึมอปริโอรี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. นำ L_3 ที่ได้ไปสร้าง C_4 โดยวิธีการสร้าง C_4 นั้นจะจับคู่ร่วมกันของกลุ่มข้อมูลทุกตัวที่เป็น L_3 กับตัวมันเองที่มีกลุ่มข้อมูลสองตัวแรกเหมือนกัน โดยผลลัพธ์ของการจับคู่จะประกอบด้วยกลุ่มข้อมูล 4 ข้อมูล จากรูปที่ 2.9 จะเห็นได้ว่าเมื่อได้ L_3 แล้วไม่สามารถสร้าง C_4 จาก L_3 ได้เนื่องจากไม่มีกลุ่มข้อมูลใดใน L_3 ที่มีข้อมูล 3 ข้อมูลแรกเหมือนกันทำให้จบการทำงาน ผลลัพธ์ที่ได้ทั้งหมดแสดงได้ดังรูปที่ 2.10

Level	Frequent Itemsets
L_1	A:4, C:4, D:5, E:3, G:2
L_2	AC:3, AD:4, AE:2, CD:4, CE:2, DE:3, DG:2, EG:2
L_3	ACD:3, ADE:2, CDE:2, DEG:2

รูปที่ 2.10 ผลลัพธ์การค้นหากลุ่มข้อมูลที่ปรากฏร่วมกันของอัลกอริทึมอพรีโอริ

2.3.2 อัลกอริทึมคาร์มา

คาร์มา (CARMA) หรือ Continuous Association Rule Mining Algorithm เป็นการคำนวณกลุ่มของข้อมูลที่มีขนาดใหญ่ โดยแสดงให้เห็นถึงกฎความสัมพันธ์ โดยสามารถกำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำ ในการสแกนของฐานข้อมูลคาร์มาจะสร้างกลุ่มของข้อมูลในการสแกนครั้งแรกเพื่อให้ข้อมูลมีขนาดเล็กลง และสแกนครั้งที่สองจากข้อมูลสแกนครั้งแรก ซึ่งคล้ายคลึงกับอัลกอริทึมอพรีโอริ แต่อัลกอริทึมคาร์มามีความยืดหยุ่นกว่าดังนี้ (Christian Hidber, 1999)

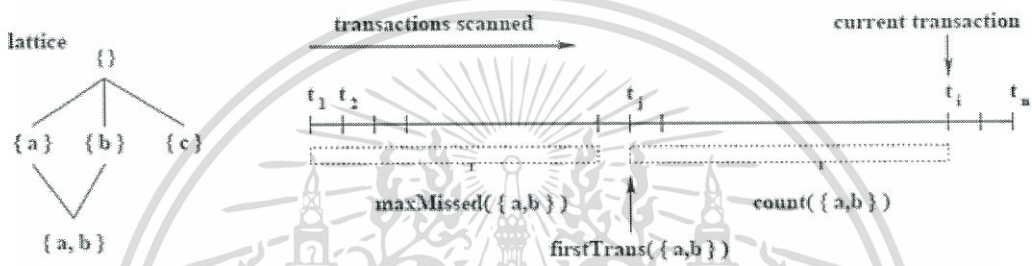
- **Continuous Feedback** ในระหว่างที่มีการสแกนฐานข้อมูลทำให้เกิดความต่อเนื่องของกฎความสัมพันธ์กลุ่มของข้อมูล
- **User Controllable** ผู้ใช้งานสามารถควบคุมพารามิเตอร์นำมาใช้ในการคำนวณกลุ่มของข้อมูลได้ตามความต้องการ เช่น ค่าสนับสนุน และค่าความเชื่อมั่น
- **Deterministic and Accurate Results** หลังจากผ่านการสแกนข้อมูลครั้งที่สอง สามารถสร้างกฎความสัมพันธ์ และได้ค่าสนับสนุนกับค่าความเชื่อมั่นที่แม่นยำ

อัลกอริทึมคาร์มาแบ่งขั้นตอนออกเป็น 2 ขั้นตอน ในช่วงแรกจะดำเนินการสแกนกลุ่มของข้อมูลทั้งหมดที่เกิดขึ้นจากฐานข้อมูลขนาดใหญ่ หลังจากได้ข้อมูลบางส่วนจากการสแกน โดยแต่ละเซตข้อมูลจะถูกเก็บไว้สำหรับขั้นตอนต่อไป ดังรูปที่ 2.11 ในการสแกนข้อมูลในขั้นตอนแรก จะได้เซตข้อมูล $\{a,b\}$ จากการสแกนข้อมูลทั้งหมด ซึ่งอยู่ระหว่างตำแหน่ง t_j และ t_k

ความหมายของตัวแปรในขั้นตอนของคาร์มา มีดังนี้

count(v)	เซตของจำนวนข้อมูลที่ปรากฏจะแทรกอยู่ในตาราง โดยค่าจะอยู่ระหว่าง v_1 ถึง v_2 ของจำนวนข้อมูลทั้งหมดที่มีอยู่
firstTrans(v)	ดัชนีของการทำธุรกรรมของ v จะถูกแทรกอยู่ในตาราง
maxMissed (v)	ขอบเขตบนของจำนวนข้อมูลที่ปรากฏของ v ก่อนที่ v จะถูกแทรกอยู่ในตาราง

ขั้นตอนที่ 1 สร้าง Superset ของเซตของจำนวนข้อมูลขนาดใหญ่ ในขณะที่สแกนข้อมูลลงตารางจะคอยช่วยในเรื่องการสนับสนุน และการจัดลำดับแบบคร่าวๆ โดยการสร้างตารางสำหรับจัดเก็บข้อมูลที่ได้จากการสแกนข้อมูล



รูปที่ 2.11 ขั้นตอนของคาร์มา

ขั้นตอนที่ 2 นำ Superset ของตารางขั้นตอนแรกมาคำนวณค่าสนับสนุน และค่าความเชื่อมั่นตามที่ได้กำหนดข้อมูลไว้ โดยในขั้นตอนที่ 2 โดยสามารถลดขนาดตารางในการสแกนหาค่าความสัมพันธ์

อัลกอริทึมคาร์มาสามารถหาความสัมพันธ์ข้อมูลที่มีขนาดใหญ่ โดยสร้างเซตของข้อมูลขนาดใหญ่อย่างต่อเนื่องพร้อมกับการหาค่าสนับสนุนสำหรับแต่ละเซตของข้อมูล ในการใช้อัลกอริทึมนี้จะช่วยให้ผู้ใช้สามารถปรับเปลี่ยนเกณฑ์ที่ใช้ในการสนับสนุนตลอดเวลา ซึ่งในการสแกน 2 ครั้ง การทำงานส่วนใหญ่มักจะ ได้ค่าสนับสนุน และค่าความเชื่อมั่นที่ถูกต้อง

2.3.3 อัลกอริทึมซีเควนซ์

อัลกอริทึมซีเควนซ์เป็นการค้นหาความสัมพันธ์ในช่วงเวลาที่ใกล้เคียงกันของข้อมูลแบบเรียงลำดับในฐานข้อมูลขนาดใหญ่ ซึ่งสามารถทำให้เห็นความถี่ของข้อมูลที่มีความสัมพันธ์กันมากขึ้น โดยกฎความสัมพันธ์ที่ได้มีส่วนช่วยในกระบวนการตัดสินใจทางด้านธุรกิจเช่น นำมาใช้ในการวิเคราะห์รูปแบบการซื้อของลูกค้า (Fisher, 2001)

การหาความถี่ของความสัมพันธ์ในแต่ละข้อมูลจะต้องประกอบด้วยสองพารามิเตอร์ คือ ค่าของข้อมูล (s_j) และเวลา (t_j) สามารถเขียนให้อยู่ในรูปแบบดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$D = \{t1, t2, \dots, ti, \dots, tm\}$ คือ ช่วงเวลาของแต่ละธุรกรรมทั้งหมดในฐานข้อมูล

$ti = ([s1, t1], [s2, t2], \dots, [sj, tj], \dots, [sn, tn])$ เมื่อ $t_{j+1} > t_j$

การเก็บข้อมูลลงบนฐานข้อมูลนั้นเก็บข้อมูลการทำธุรกรรมตั้งแต่ $t1$ ถึง tm สามารถแสดงข้อมูล ดังรูปที่ 2.13

Transaction No.	Item-series
1	$(s_1 s_2 \dots s_j \dots s_n)_1$
2	$(s_1 s_2 \dots s_j \dots s_n)_2$
⋮	⋮
i	$(s_1 s_2 \dots s_j \dots s_n)_i$
⋮	⋮
m	$(s_1 s_2 \dots s_j \dots s_n)_m$

รูปที่ 2.12 ข้อมูลการทำธุรกรรมของอัลกอริทึมซีแควนซ์

หลังจากนั้นจะหาความถี่ของแต่ละ Transaction No. ได้นั้นจะต้องนำข้อมูลออกมาเป็น Itemset ซึ่งเป็นการเรียงลำดับของรายการ โดยจะพิจารณาจาก Transaction ที่เกิดขึ้นพร้อมกัน

สำหรับกฎความสัมพันธ์แบบลำดับจะอยู่ในรูปแบบของ $X \rightarrow Y$ มีค่าความเชื่อมั่น และค่าสนับสนุน ในการค้นหาที่ถูกต้องค่าสนับสนุน และค่าความเชื่อมั่นต้องไม่น้อยกว่าค่าที่กำหนดจากค่าของ minSup และ minConf (Pei, 2004)

ตัวอย่างการหาความสัมพันธ์แบบลำดับ

โดยการพิจารณาจาก minSup = 0.5 และ minConf = 0.5 จะพบข้อมูล ดังรูปที่ 2.14

ID	Rule	Support	Confidence
r1	$\{a, b, c\} \Rightarrow \{e\}$	0.5	1.0
r2	$\{a\} \rightarrow \{c, e, f\}$	0.5	0.66
r3	$\{a, b\} \rightarrow \{e, f\}$	0.5	1.0
r4	$\{b\} \rightarrow \{e, f\}$	0.75	0.75
r5	$\{a\} \rightarrow \{e, f\}$	0.75	1.0
r6	$\{c\} \rightarrow \{f\}$	0.5	1.0
r7	$\{a\} \rightarrow \{b\}$	0.5	0.66
...

รูปที่ 2.13 ลำดับความสัมพันธ์ของข้อมูลที่พบจากการค้นหาความสัมพันธ์แบบลำดับ

อัลกอริทึมซีแควนซ์เป็นที่นิยมมากที่สุด ในเทคนิคการหาความสัมพันธ์ของข้อมูล แต่ อัลกอริทึมซีแควนซ์มีข้อจำกัดสำหรับการคาดการณ์ ตัวอย่างเช่น การพิจารณาจากรูปแบบ $\{x\} \{y\}$ เป็นไปได้ว่าค่า y จะปรากฏหลังจากค่า x แต่ยังมีหลายกรณีเมื่อค่า x ไม่ตามมาด้วยค่า y สำหรับการคาดการณ์ต้องมีการวัดค่าความเชื่อมั่น หากค่า x เกิดขึ้นค่า y จะเกิดขึ้นหลังจากนั้น ในอัลกอริทึมซี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คววนซึ้นนั้เป็นวิธีการหาความสัมพันธ์ซึ่งมีส่วนคล้ายคลึงกับอัลกอริทึมอปริโริ แต่สามารถหาความถี่ที่เกิดขึ้นบ่อยครั้งในแต่ละช่วงเวลาเพื่อให้เกิดประสิทธิภาพสูงมากขึ้น

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 กฎความสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย (Srisawat,2010) งานวิจัยนี้ต้องการทราบถึงแนวโน้มการซื้อขายหุ้นของนักลงทุนในตลาดหลักทรัพย์แห่งประเทศไทยโดยใช้กฎความสัมพันธ์ของข้อมูลเพื่อหาความสัมพันธ์ในแต่ละหุ้น และใช้ข้อมูลย้อนจากบริษัทหลักทรัพย์กิมเิ่งจากการวิเคราะห์ห้ค้นพบว่าหากหุ้น KEST และหุ้น AMATA เพิ่มขึ้นมากกว่า 3 เปอร์เซ็นต์หุ้น ASP ก็จะมีเพิ่มขึ้นมากกว่า 3 เปอร์เซ็นต์เช่นกันตามกฎความสัมพันธ์ของข้อมูลจากผลลัพธ์ที่ได้จะช่วยให้นักลงทุนสามารถตัดสินใจหุ้นที่มีแนวโน้มที่สามารถทำกำไรได้อย่างถูกต้อง ดังรูปที่ 2.14

the percentage of change	minsup (%)	association rules	support (%)	lift
1%	10	1. {dBAY, dTOP, dIRPC} → dPTTAR	11.16	3.84
		2. {dBAY, dTOP, dIRPC, dBMCL} → dPTTAR	10.33	3.84
		3. {dTOP, dESTH, dIRPC} → dPTTAR	10.33	3.84
		4. {dPDI, dPT, dKBANK} → dKIB	10.33	3.72
		5. {dPT, dSTEC, dKIB} → dIBAY	10.33	3.72
2%	5	1. {dQH, dLOXLEY, dSTA} → dIRPC	5.37	8.96
		2. {dTRUE, dTCAP, dSTEC} → dIRPC	5.37	8.96
		3. {dSCC, dAMATA, dHEMRAJ} → dTCAP	5.37	8.07
		4. {dPIPE, dBLAND} → dTCAP	5.37	8.07
		5. {dTHAI, dIRPC} → dTCAP	5.37	8.07
3%	3	1. {dKEST, dAMATA} → dIASP	3.31	9.68
		2. {dPTTAR, dIVL} → dSTA	3.31	8.07
		3. {dCPF, dQH} → dIAP	3.31	7.81
		4. {dLPN, dHEMRAJ} → dISPAI	3.31	7.56
		5. {dHEMRAJ, dGSTEEL} → dGIS	3.31	4.73

รูปที่ 2.14 ผลลัพธ์จากกฎความสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทย

2.4.2 กฎความสัมพันธ์ของหุ้นในตลาดหลักทรัพย์แห่งประเทศไทยอินโดนีเซีย (Widiputra, 2012) เนื่องจากสถานะเศรษฐกิจโลกที่มีความผันผวนที่ค่อนข้างสูง ทำให้นักลงทุนหลายคนขาดทุนเป็นจำนวนมาก ดังนั้นหากทราบกฎความสัมพันธ์ของหุ้นย่อมทำให้นักลงทุนสามารถตัดสินใจได้ถูกต้องมากยิ่งขึ้นงานวิจัยนี้ได้กล่าวถึงหุ้น SGRO ที่มีราคาปิดที่เป็นแบบแผนชัดเจนโดยใช้วิธีการจากกฎความสัมพันธ์สามารถทำนายว่าหุ้น SGRO ในแต่ละวันจะปิดบวกหรือปิดลบซึ่งผลการทดสอบนั้นได้ผลค่อนข้างที่จะแม่นยำถึง 80 เปอร์เซ็นต์ดังรูปที่ 2.15

บทที่ 3

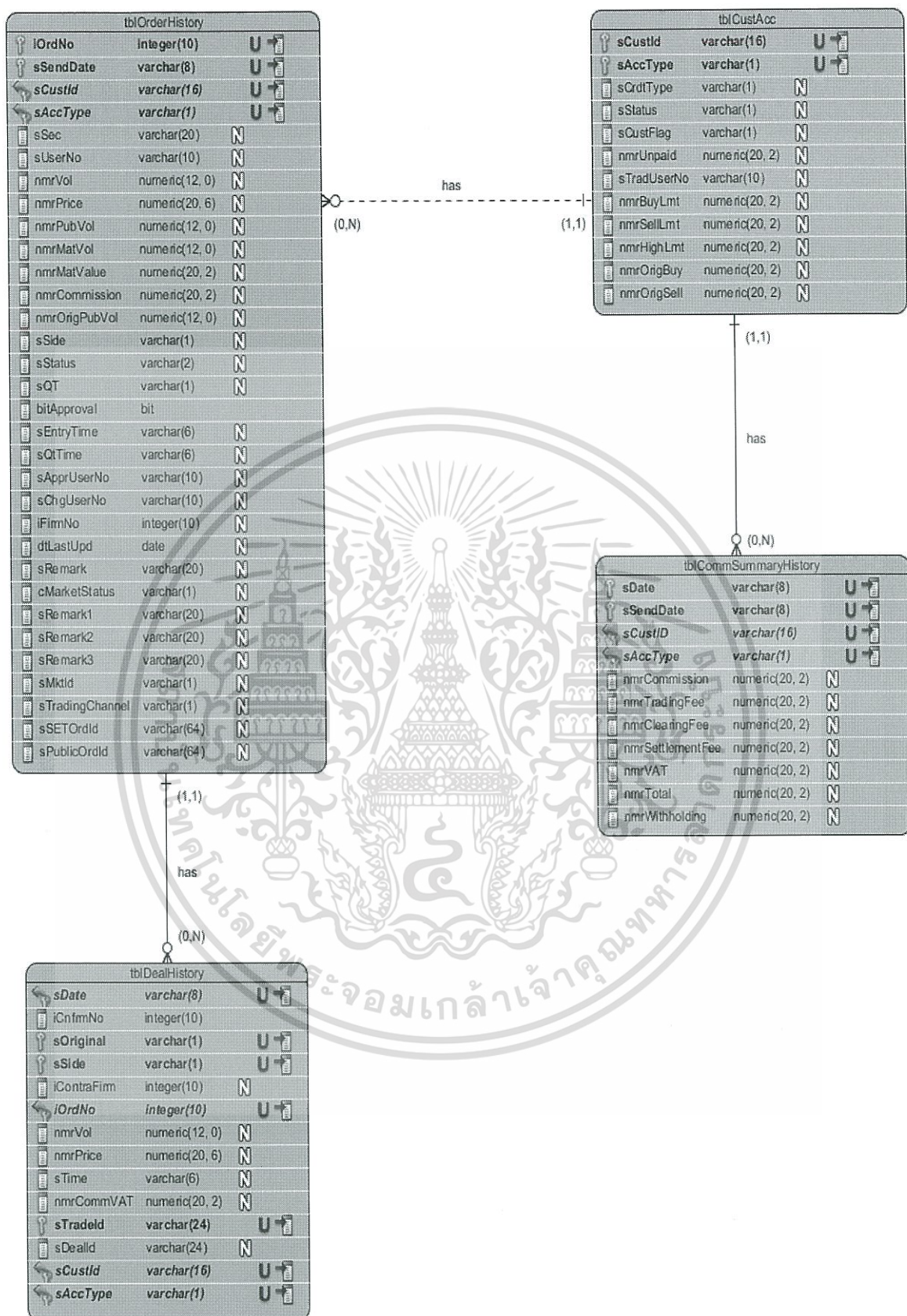
การดำเนินการ

การนำทฤษฎีการสร้างกฎความสัมพันธ์ของข้อมูล (Association Rule) เป็นแนวทางในการวิเคราะห์พฤติกรรมของนักลงทุนในตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้ข้อมูลการซื้อขายหลักทรัพย์ของนักลงทุนในอดีตจนถึงปัจจุบันซึ่งถูกรวบรวมอยู่ในคลังข้อมูลขนาดใหญ่ของบริษัทหลักทรัพย์ตั้งแต่ปี พ.ศ.2551 เพื่อนำองค์ความรู้มาใช้ในกองทุนสนับสนุนการตัดสินใจที่แม่นยำมากยิ่งขึ้น และการศึกษาครั้งนี้ยังสามารถทำให้ทราบถึงพฤติกรรมที่แอบแฝงในการซื้อขายหลักทรัพย์

3.1 แหล่งข้อมูล

การวิเคราะห์พฤติกรรมของนักลงทุนจากการนำข้อมูลมาจากคลังข้อมูลขนาดใหญ่ โดยการเก็บข้อมูลการซื้อขายของนักลงทุนที่มีความสัมพันธ์กันอย่างชัดเจน ทั้งนี้ได้พิจารณาข้อมูลที่มีส่วนเกี่ยวข้องที่ชัดเจน และมีแนวโน้มที่สอดคล้องเพื่อให้ได้ผลตรงตามเป้าหมายของการศึกษาในครั้งนี้

ในโครงสร้างฐานข้อมูลหลักที่ถูกจัดเก็บรวบรวมในระบบซื้อขายหลักทรัพย์ และถูกใช้งานอยู่จริงในปัจจุบันคือ ตาราง tblOrderHistory ซึ่งเก็บข้อมูลการซื้อขายหลักทรัพย์ย้อนหลังของลูกค้าทุกคนที่อยู่ในบริษัทหลักทรัพย์ โดยมีการเก็บรายละเอียดได้แก่ข้อมูลรหัสลูกค้า หุ่นที่ลูกค้าได้ส่งคำสั่งซื้อขาย จำนวนหุ้น ราคา วันเวลาในการสั่งซื้อ และรายละเอียดย่อยต่างๆ ตาราง tblCustAcc เป็นตารางที่เก็บข้อมูลนักลงทุนที่มีอยู่ในบริษัทหลักทรัพย์ โดยข้อมูลหลักจะประกอบไปด้วยรหัสลูกค้าจำนวนเงินที่สามารถสั่งซื้อหุ้นได้ และสถานะการซื้อขายต่างๆ ที่สำคัญ ตาราง tblCommSummaryHistory เป็นตารางที่เก็บค่าคอมมิชชั่นของนักลงทุนทั้งหมดของหลักทรัพย์ โดยมีการเก็บมีการเก็บข้อมูลตั้งแต่จำนวนเงินที่เสียไปในการซื้อหุ้นในแต่ละครั้งตามเงื่อนไขที่ตลาดหลักทรัพย์เป็นผู้กำหนดตั้งแต่อดีตจนถึงปัจจุบัน ตาราง tblDealHistory เป็นตารางที่เก็บรายละเอียดข้อมูลการซื้อขายที่ถูกจับคู่สำเร็จระหว่างนักลงทุนย้อนหลังทั้งหมด โดยรายละเอียดที่เก็บคือ หุ้นที่นักลงทุนได้ซื้อขาย จำนวนหุ้น ราคาในการซื้อขาย และวันเวลาที่ถูกจับคู่ได้สำเร็จ โดยตารางทั้งหมดที่กล่าวมามีความสัมพันธ์กันอย่างชัดเจนเป็นอย่างยิ่ง ดังรูปที่ 3.1



รูปที่ 3.1 ความสัมพันธ์ของตารางข้อมูลที่สำคัญของระบบซื้อขายหลักทรัพย์ย้อนหลัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการจัดเก็บรายละเอียดข้อมูลของการซื้อขายหลักทรัพย์ตั้งแต่อดีตถึงปัจจุบันจะประกอบด้วย 4 ตารางสำคัญ คือตาราง tblOrderHistory ที่เก็บรายละเอียดซื้อขายดังตารางที่ 3.1 ตาราง tblCustAcc เก็บรายละเอียดของลูกค้าดังตารางที่ 3.2 ตาราง tblCommSummaryHistory เก็บรายละเอียดค่าคอมมิชชั่นของนักกลงทุนดังตารางที่ 3.3 ตาราง tblDealHistory เก็บรายละเอียดข้อมูลการซื้อขายที่ถูกจับคู่สำเร็จดังตารางที่ 3.4 ตามลำดับ

ตารางที่ 3.1 รายละเอียดตาราง tblOrderHistory

Table Name: tblOrderHistory			
Name	Code	Type	Key
ลำดับคำสั่ง	iOrdNo	integer(10)	Y
วันที่ส่งคำสั่ง	sSendDate	varchar(8)	Y
รหัสบัญชี	sCustId	varchar(16)	Y
ชนิดบัญชี	sAccType	varchar(1)	Y
รหัสหุ้น	sSec	varchar(20)	N
ผู้ใช้ที่คีย์คำสั่ง	sUserNo	varchar(10)	N
จำนวนหุ้นที่สั่ง	nmrVol	numeric(12, 0)	N
ราคาหุ้น	nmrPrice	numeric(20, 6)	N
จำนวนหุ้นที่เหลือจากการสั่ง	nmrPubVol	numeric(12, 0)	N
จำนวนหุ้นที่ได้	nmrMatVol	numeric(12, 0)	N
ราคาที่ได้	nmrMatValue	numeric(20, 2)	N
ค่าคอมมิชชั่น	nmrCommission	numeric(20, 2)	N
จำนวนหุ้นที่สั่งเดิม	nmrOrigPubVol	numeric(12, 0)	N
ประเภทซื้อหรือขาย	sSide	varchar(1)	N
สถานะคำสั่ง	sStatus	varchar(2)	N
ค้ายืนยันคำสั่ง	sQT	varchar(1)	N
ค่าการอนุมัติ	bitApproval	bit	N
เวลาที่ส่งคำสั่ง	sEntryTime	varchar(6)	N
ค่าเวลายืนยันคำสั่ง	sQtTime	varchar(6)	N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 รายละเอียดตาราง tblOrderHistory (ต่อ)

Table Name: tblOrderHistory			
Name	Code	Type	Key
ผู้ใช้ที่อนุมัติ	sApprUserNo	varchar(10)	N
ผู้ใช้ที่ปรับปรุง	sChgUserNo	varchar(10)	N
รหัสหลักทรัพย์	iFirmNo	integer(10)	N
เวลาปรับปรุงข้อมูล	dtLastUpd	date	N
คำสำรอง	sRemark	varchar(20)	N
สถานะตลาด	cMarketStatus	varchar(1)	N
คำสำรอง1	sRemark1	varchar(20)	N
คำสำรอง2	sRemark2	varchar(20)	N
คำสำรอง3	sRemark3	varchar(20)	N
สถานะตลาด	sMktId	varchar(1)	N
ช่องทางซื้อขาย	sTradingChannel	varchar(1)	N
คำสั่งยืนยันตลาด	sSETOrdId	varchar(64)	N
คำสั่งยืนยันส่วนรวม	sPublicOrdId	varchar(64)	N

ตารางที่ 3.2 รายละเอียดตาราง tblCustAcc

Table Name: tblCustAcc			
Name	Code	Type	Key
รหัสบัญชี	sCustId	varchar(16)	Y
ชนิดบัญชี	sAccType	varchar(1)	Y
ชนิดบัญชีเครดิต	sCrdtType	varchar(1)	N
สถานะบัญชี	sStatus	varchar(1)	N
สถานะการซื้อขาย	sCustFlag	varchar(1)	N
ยอดเงินคงค้าง	nmrUnpaid	numeric(20, 2)	N
รหัสพนักงาน การตลาด	sTradUserNo	varchar(10)	N
วงเงินในการซื้อ	nmrBuyLmt	numeric(20, 2)	N
วงเงินในการขาย	nmrSellLmt	numeric(20, 2)	N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 รายละเอียดตาราง tblCustAcc (ต่อ)

Table Name: tblCustAcc			
Name	Code	Type	Key
การซื้อสูงสุด	nmrHighLmt	numeric(20, 2)	N
วงเงินในการซื้อเดิม	nmrOrigBuy	numeric(20, 2)	N
วงเงินในการขายเดิม	nmrOrigSell	numeric(20, 2)	N

ตารางที่ 3.3 รายละเอียดตาราง tblCommSummaryHistory

Table Name: tblCommSummaryHistory			
Name	Code	Type	Key
วันที่ส่งคำสั่ง	sDate	varchar(8)	Y
วันที่ส่งคำสั่ง	sSendDate	varchar(8)	Y
รหัสบัญชี	sCustID	varchar(16)	Y
ชนิดบัญชี	sAccType	varchar(1)	Y
ค่าคอมมิชชั่น	nmrCommission	numeric(20, 2)	N
ค่าคำสั่งซื้อขาย	nmrTradingFee	numeric(20, 2)	N
ค่าบริการ	nmrClearingFee	numeric(20, 2)	N
ค่าชำระรายการ	nmrSettlementFee	numeric(20, 2)	N
ค่าภาษี	nmrVAT	numeric(20, 2)	N
ผลรวมทั้งหมด	nmrTotal	numeric(20, 2)	N
ภาษี ณ ที่จ่าย	nmrWithholding	numeric(20, 2)	N

ตารางที่ 3.4 รายละเอียดตาราง tblDealHistory

Table Name: tblDealHistory			
Name	Code	Type	Key
วันที่ส่งคำสั่ง	sDate	varchar(8)	Y
ค่าจำนวนการยืนยัน	iCnfirmNo	integer(10)	N
ค่าเริ่มต้น	sOriginal	varchar(1)	Y
ประเภทซื้อหรือขาย	sSide	varchar(1)	Y
ค่าจำนวนสัญญา	iContraFirm	integer(10)	N
ลำดับคำสั่ง	iOrdNo	integer(10)	Y

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 รายละเอียดตาราง tblDealHistory (ต่อ)

Table Name: tblDealHistory			
Name	Code	Type	Key
จำนวนหุ้นที่สั่ง	nmrVol	numeric(12, 0)	N
ราคาที่คำสั่ง	nmrPrice	numeric(20, 6)	N
เวลาส่งคำสั่ง	sTime	varchar(6)	N
ค่าคอมมิชชั่น	nmrCommVAT	numeric(20, 2)	N
คำสั่งรายละเอียด	sTradeId	varchar(24)	Y
คำสั่งยืนยันตลาด	sDealId	varchar(24)	N
รหัสบัญชี	sCustId	varchar(16)	Y
ชนิดบัญชี	sAccType	varchar(1)	Y

3.2 การกำหนดเป้าหมายของพฤติกรรมนักลงทุน

จากที่บริษัทหลักทรัพย์ต้องการสร้างกลยุทธ์ในการลงทุนเพื่อเพิ่มศักยภาพให้กับการบริหารจัดการกองทุน โดยอาศัยข้อมูลที่มีอยู่มาใช้ให้เกิดประโยชน์และใช้เทคนิคการค้าไม้นิ่งมาวิเคราะห์ โดยพยายามวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุน ตัวอย่างเช่น หากนักกลงทุนซื้อหุ้นอย่างหนึ่งแล้ว มักจะซื้อหุ้นอีกตัวคู่กัน ย่อมสามารถทำให้กองทุนสามารถทำกำไรได้มากยิ่งขึ้น โดยหากรู้เงื่อนไขบางอย่างแล้วย่อมสามารถตัดสินใจได้ดีมากยิ่งขึ้น

3.3 การจัดเตรียมข้อมูล

การนำข้อมูลมาใช้เพื่อวิเคราะห์เป็นเรื่องที่สำคัญมากในการทำค้าไม้นิ่ง โดยการหาพฤติกรรมของนักลงทุนซื้อขายหลักทรัพย์จะเน้นที่ ตาราง tblOrderHistory

การวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุนได้เป้าหมายดังกล่าวที่ตั้งไว้ โดยในฐานข้อมูลมีข้อมูลบางแอตทริบิวต์ที่น่าสนใจคือ แอตทริบิวต์ sDate วันที่ส่งคำสั่งซื้อขาย แอตทริบิวต์sCustId รหัสบัญชีแอตทริบิวต์ sSide การซื้อหรือการขายในแต่ละครั้ง แอตทริบิวต์ sSec หลักทรัพย์ที่นักลงทุนส่งคำสั่งในแต่ละครั้ง

การจัดข้อมูลที่เป็นขยะในแต่ละแอตทริบิวต์ทำให้เกิดความถูกต้องในการค้นหาความสัมพันธ์ที่ซ่อนอยู่ในฐานข้อมูลและเป็นหัวใจหลักของการทำค้าไม้นิ่ง ซึ่งข้อมูลทั้งหมดเป็นคีย์และด้วยโปรแกรมในการส่งคำสั่งจะมีการตรวจสอบว่าค่าทั้งสองต้องมีค่าเท่านั้นถึงจะส่งคำสั่งได้ ทำให้ข้อมูลมีความถูกต้องอยู่แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแปลงความหมายข้อมูลให้เหมาะกับการวิเคราะห์ต้องอาศัยเงื่อนไขความสอดคล้อง โดยการใช้งานซึ่งทั้ง 4 แดตทริบิวต์ที่เก็บอยู่ในฐานข้อมูล แต่เนื่องจากข้อมูลที่เก็บนั้นมีความหมาย และรูปแบบที่ชัดเจนจึงสามารถนำเข้าสู่กระบวนการค้ำไมน์นึ่ง ทำให้สามารถนำข้อมูลมาวิเคราะห์ได้อย่างถูกต้องและมีประสิทธิภาพต่อไป

3.4 ขั้นตอนการดำเนินการนำข้อมูลเข้าสู่โปรแกรม

การดำเนินการศึกษาพฤติกรรมนักลงทุน โดยอาศัยข้อมูลมาวิเคราะห์โดยการนำเอาข้อมูลในอดีตมาใช้เพื่อสกัดความรู้ที่สำคัญต่างๆ โดยการนำ อัลกอริทึมความสัมพันธ์ทั้ง อพริโอรि คาร์มา ซีควนซ์ด้วยโปรแกรม IBM SPSS Modeler ในการค้นหาพฤติกรรม โดยมีกระบวนการดำเนินการดังต่อไปนี้

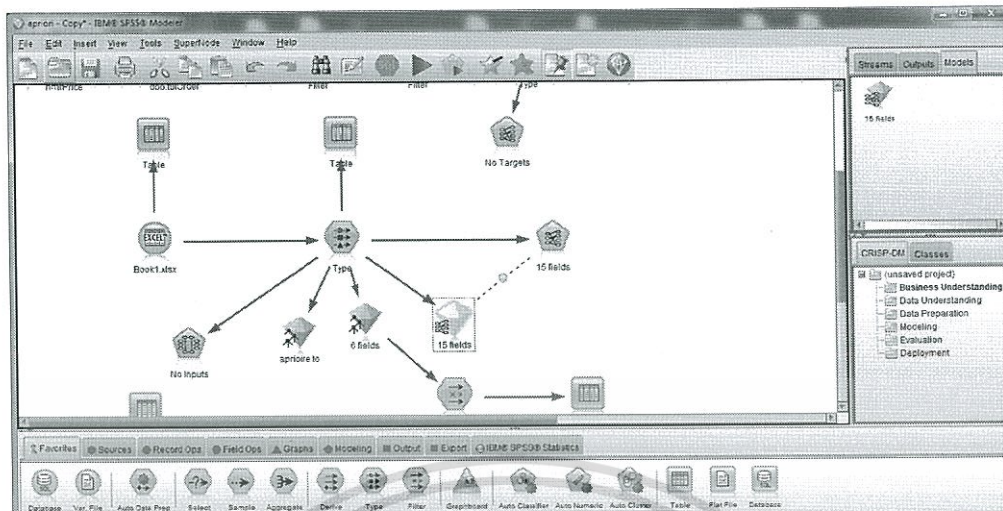
1. ขั้นตอนจัดเตรียมข้อมูล โดยขั้นตอนนี้จะเป็นการปรับปรุงข้อมูลจากกระบวนการที่กล่าวไว้ก่อนหน้านี้เพื่อให้เหมาะสมกับ Input และ Output ของโปรแกรม IBM SPSS Modeler ดังรูปที่ 3.2
2. ทำการคัดแยกข้อมูล
3. ทำการแปลงข้อมูลให้อยู่ในรูปแบบของอัลกอริทึมอพริโอริ คาร์มา และซีควนซ์ เพื่อทราบถึงพฤติกรรมของนักลงทุน
4. นำข้อมูลที่ได้จากการแปลงข้อมูล และเข้าสู่โปรแกรม IBM SPSS Modeler เพื่อรู้พฤติกรรมของนักลงทุน

	ICrdNo	ISendDate	sCustId	sAccType	sSec	sUser	nmrVol	nmrPrice	nmrPubVol	nmrMatVol	nmrMatValue	nmrCommission	nmrOrigPubVol	nm
1	1	20130628	1196359	C	CPF	4216	20000	27.000	20000	0	0.000	0.000	20000	
2	2	20130628	1116401	C	PTTGC	1139	50000	67.500	0	0	0.000	0.000	50000	
3	3	20130628	1121758	C	N-PARK	2895	8000000	0.080	0	0	0.000	0.000	8000000	
4	4	20130628	1196359	C	SRICHA	4216	10000	43.000	0	0	0.000	0.000	10000	
5	5	20130628	1116452	C	SLC	1139	100000	1.350	0	100000	135000.000	0.000	100000	
6	6	20130628	1192884	C	TOP	1594	5000	67.000	5000	0	0.000	0.000	5000	
7	7	20130628	2314320	C	SUSCO	3202	20000	4.520	0	0	0.000	0.000	20000	
8	8	20130628	1189131	C	N-PARK	3863	10000000	0.080	10000000	0	0.000	0.000	10000000	
9	9	20130628	1121189	C	CPALL	1139	40000	41.000	40000	0	0.000	0.000	40000	
10	10	20130628	1135457	C	UAC	4216	5000	8.000	5000	0	0.000	0.000	5000	
11	11	20130628	1121758	C	N-PARK	2895	10000000	0.090	0	0	0.000	0.000	10000000	
12	12	20130628	1222732	C	JAS	7691	80000	7.900	0	80000	632000.000	0.000	80000	
13	13	20130628	1132032	C	ADVANC	1594	2000	290.000	2000	0	0.000	0.000	2000	
14	14	20130628	1231502	C	UV	0290	20000	8.500	20000	0	0.000	0.000	20000	
15	15	20130628	2314320	C	SUSCO	3202	15000	4.540	0	15000	68100.000	0.000	15000	
16	16	20130628	1189131	C	N-PARK	3863	10000000	0.080	10000000	0	0.000	0.000	10000000	
17	17	20130628	1135457	C	UAC	4216	5000	8.500	5000	0	0.000	0.000	5000	
18	18	20130628	1222732	C	BLAND	7691	400000	1.470	0	400000	592000.000	0.000	400000	
19	19	20130628	1159445	C	NOK	1594	10000	28.250	0	0	0.000	0.000	10000	
20	20	20130628	5800307	C	INTUCH	3863	20000	85.000	0	20000	1710000.000	0.000	20000	
21	21	20130628	1157868	C	IEC	D135	2000000	0.030	2000000	0	0.000	0.000	2000000	
22	22	20130628	2314320	C	SUSCO	3202	15000	4.550	0	15000	68400.000	0.000	15000	
23	23	20130628	1196359	C	KBANK	4216	2000	159.000	2000	0	0.000	0.000	2000	
24	24	20130628	5800307	C	CPN	3863	40000	48.000	40000	0	0.000	0.000	40000	
25	25	20130628	1159445	C	NOK	1594	10000	28.250	0	0	0.000	0.000	10000	
26	26	20130628	1222732	C	JAS	7691	150000	8.000	150000	0	0.000	0.000	150000	
27	27	20130628	1196359	C	KBANK	4216	2000	165.000	2000	0	0.000	0.000	2000	
28	28	20130628	2314320	C	PAP	3202	3000	5.500	0	3000	16800.000	0.000	3000	
29	29	20130628	5804230H	C	IEC	3863	1200000	0.040	1200000	0	0.000	0.000	1200000	

รูปที่ 3.2 ตัวอย่างจากฐานข้อมูลที่นำมาใช้

5. สร้างกฎความสัมพันธ์ของพฤติกรรมนักลงทุนในตลาดหลักทรัพย์ โดยใช้ อัลกอริทึมอพริโอริ คาร์มา ซีควนซ์เพื่อมาใช้ตัดสินใจในกองทุนต่อไปดังรูปที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 โปรแกรม SPSS IBM Modeler



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

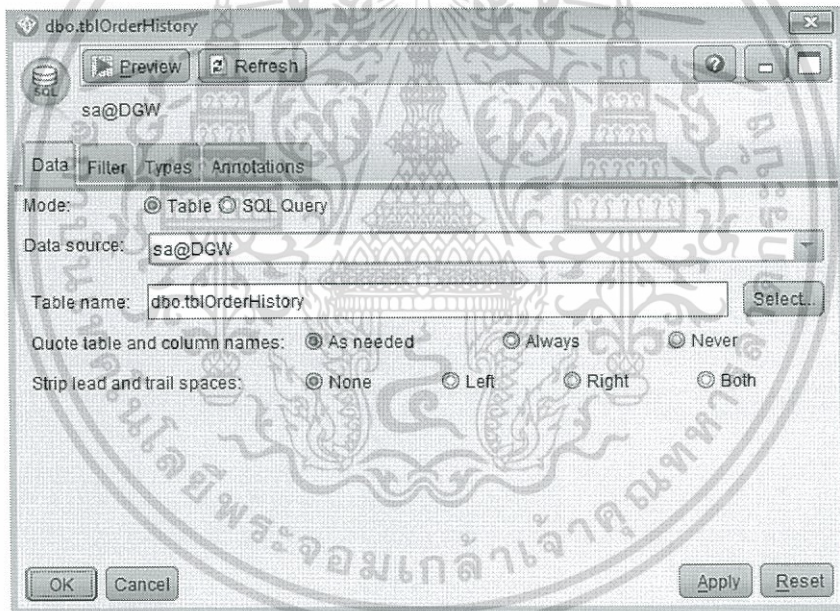
บทที่ 4

การดำเนินงานวิเคราะห์ข้อมูล

หลังจากได้กำหนดเป้าหมายและความต้องการในการวิเคราะห์ข้อมูล โดยเริ่มจากการคัดกรองข้อมูลเพื่อแปลงข้อมูลที่มีความหมายที่ถูกซ่อนเร้นแปลงข้อมูลในรูปแบบต่างๆเป็นในแบบที่ต้องการ หลังจากนั้นดำเนินการตามเป้าหมายที่ได้กำหนดไว้ก่อนหน้านี้ เพื่อนำข้อมูลมาวิเคราะห์แล้วใช้ซอฟต์แวร์ค่าใดมาหนึ่งในการหาสิ่งที่ซ่อนเร้นต่อไป

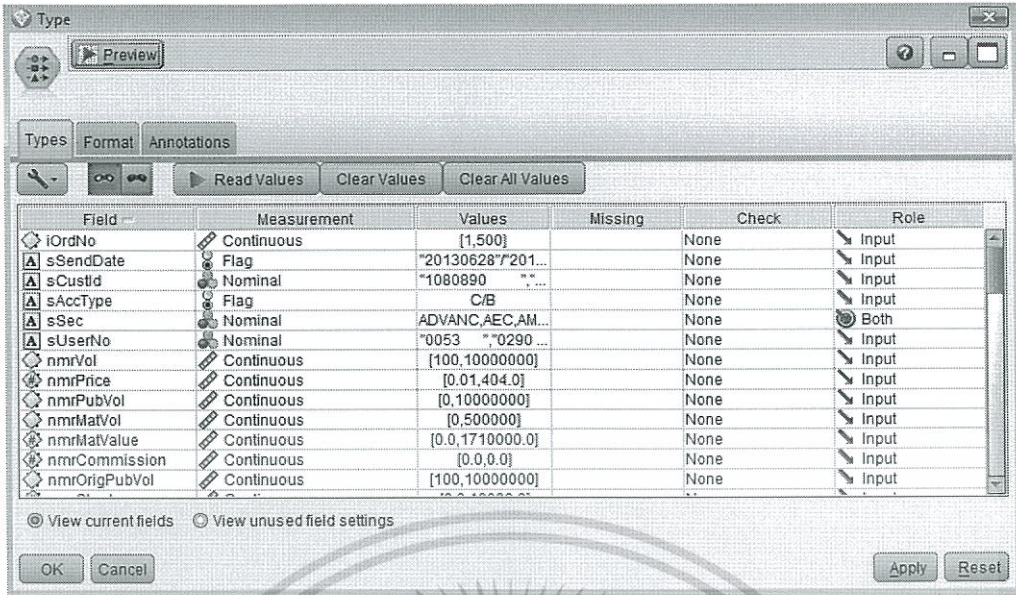
4.1 การนำข้อมูลเข้า

การนำข้อมูลจากฐานข้อมูลสามารถใช้โปรแกรม IBM SPSS Modeler เพื่อดึงข้อมูล และเลือกชื่อตารางที่ต้องการมาวิเคราะห์พฤติกรรมคือตาราง tblOrderHistory ดังรูปที่ 4.1



รูปที่ 4.1 การเลือกฐานข้อมูลที่จะนำมาวิเคราะห์พฤติกรรม

หลังจากนั้นการกำหนดชนิดข้อมูลเพื่อสามารถนำข้อมูลเข้าอัลกอริทึมทั้ง 3 อัลกอริทึมได้กำหนดเป้าหมายในบทที่ 3 ดังรูปที่ 4.2



รูปที่ 4.2 การกำหนดชนิดตัวแปรของค่าแอตทริบิวต์

การวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุน เช่น นักลงทุนส่วนใหญ่เมื่อซื้อหุ้นหนึ่งแล้วมักจะซื้อหุ้นอีกตัวคู่กันด้วยอัลกอริทึมอพริโอริคาร์มา และซีแควนซ์โดยใช้ข้อมูลตัวอย่างที่แสดงรายละเอียดดังรูปที่ 4.3

iOrdNo	sSendDate	sCustId	sAccType	sSec	sUser	nmrVol	nmrPrice	nmrPubVol	nmrMatVol	nmrMatValue	nmrCommission	nmrOrigPubVol
20296	10014998	20130628	4202899	C	TGPRO X119	500000	0.800	0	500000	400000.000	0.000	500000
20297	10014998	20130628	3517730M	B	DEMCO X119	10000	9.050	0	10000	90500.000	0.000	10000
20298	10015000	20130628	1221809	C	TRU X119	1800	10.800	1800	0	0.000	0.000	1800
20299	10015001	20130628	2225735	C	DTACO X119	200000	0.660	0	200000	130000.000	0.000	200000
20300	10015002	20130628	5702372	C	ADVANC X119	100	282.000	0	100	28200.000	0.000	100
20301	10015003	20130628	1145193H	C	LIVE X119	100000	0.610	0	100000	61000.000	0.000	100000
20302	10015004	20130628	1179489	C	CCET X119	40000	2.760	0	0	0.000	0.000	40000
20303	10015005	20130628	5419735	C	TGPRO X119	35000	0.800	0	35000	28000.000	0.000	35000
20304	10015006	20130628	2514923	C	BANPU X119	50	239.000	0	50	11951.000	0.000	50
20305	10015007	20130628	1231464	C	AJ X119	6100	16.100	0	6100	98210.000	0.000	6100
20306	10015008	20130628	4212126	C	THCOM X119	5000	33.000	5000	0	0.000	0.000	5000
20307	10015009	20130628	1226975	C	INTUO X119	125000	2.020	0	125000	252500.000	0.000	125000
20308	10015010	20130628	1145823	C	MLINK X119	150000	4.880	150000	0	0.000	0.000	150000
20309	10015011	20130628	3411044	C	NMG X119	200000	1.520	0	200000	304000.000	0.000	200000
20310	10015012	20130628	5114576H	C	TIFCO X119	2000	9.950	0	2000	19900.000	0.000	2000
20311	10015013	20130628	1234420H	C	MLINK X119	5700	4.720	0	0	0.000	0.000	5700
20312	10015014	20130628	1222490	C	BLAND X119	40000	1.530	40000	0	0.000	0.000	40000
20313	10015015	20130628	5602106	C	AOT X119	300	167.000	0	300	50100.000	0.000	300
20314	10015016	20130628	5508754	C	PPS X119	20000	1.250	20000	0	0.000	0.000	20000
20315	10015017	20130628	1131559	C	CK X119	1000	21.300	0	1000	21300.000	0.000	1000
20316	10015018	20130628	2539640M	B	CK X119	10000	21.400	10000	0	0.000	0.000	10000
20317	10015019	20130628	2319667	C	BTC X119	50000	0.970	0	0	0.000	0.000	50000
20318	10015020	20130628	1231014	C	BBL06 X119	20000	0.840	0	20000	16800.000	0.000	20000
20319	10015021	20130628	2415518H	C	ESSO0 X119	50000	0.020	44700	5300	106.000	0.000	50000
20320	10015022	20130628	1128078	C	BLAND X119	4000	1.520	0	4000	6080.000	0.000	4000
20321	10015023	20130628	5413892H	C	RCP23 X119	20000	0.170	0	20000	3400.000	0.000	20000

รูปที่ 4.3 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมอพริโอริคาร์มา และซีแควนซ์

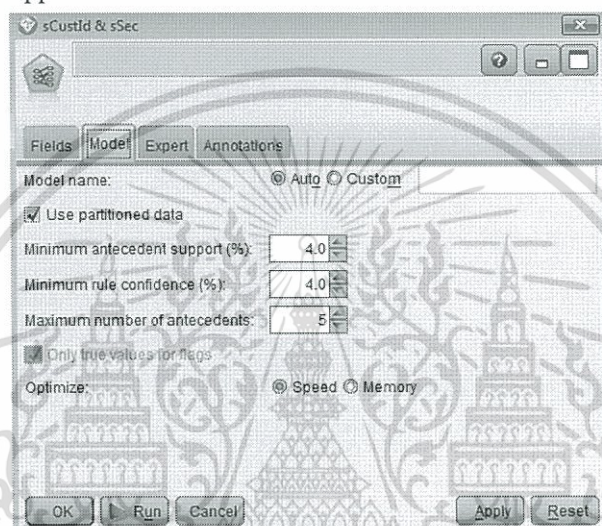
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ซอฟต์แวร์ที่นำมาวิเคราะห์ และการประยุกต์ใช้งานด้วย IBM SPSS Modeler

หลังจากที่ได้จัดเตรียมข้อมูลตามหลักของคาค่าไมนึ่งแล้ว ข้อมูลที่จะมาวิเคราะห์นั้น สามารถนำเข้าไปอัลกอริทึมแต่ละอย่างเพื่อได้ค่าข้อมูลที่แตกต่างในแต่ละกรณี

4.2.1 การวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุน เช่นนักลงทุนส่วนใหญ่เมื่อซื้อหุ้นหนึ่งแล้วมักจะซื้อหุ้นอีกตัวคู่กันด้วยอฟรีโอริอัลกอริทึม

1. ในการวิเคราะห์พฤติกรรมของนักลงทุนด้วยความสัมพันธ์ของข้อมูลซึ่งการกำหนดค่า Minimum antecedent support เท่ากับ 4% และ Minimum rule confidence เท่ากับ 4% ดังรูปที่ 4.4



รูปที่ 4.4 การตั้งค่าของอัลกอริทึมอฟรีโอริ

2. ผลจากอัลกอริทึมอฟรีโอริที่แสดงค่ากฎความสัมพันธ์ดังรูปที่ 4.5

Consequent	Antecedent	Support %	Confidence %
PTT = T	INTUCH = T	4.036	33.333
SCB = T	TOP = T	4.036	33.333
KTB = T	TOP = T	4.036	33.333
BBL = T	KBANK = T	5.381	25.0
SCB = T	KBANK = T	5.381	25.0
KTB = T	KBANK = T	5.381	25.0
DTAC = T	INTUCH = T	4.036	22.222
DTAC = T	TOP = T	4.036	22.222
BROOK = T	TOP = T	4.036	22.222
PTT = T	TOP = T	4.036	22.222
PITTEP = T	INTUCH = T	4.036	22.222
ADVANC = T	INTUCH = T	4.036	22.222
ADVANC = T	TOP = T	4.036	22.222
TOP = T	INTUCH = T	4.036	22.222
INTUCH = T	TOP = T	4.036	22.222
PTTGC = T	INTUCH = T	4.036	22.222
SCB = T	INTUCH = T	4.036	22.222

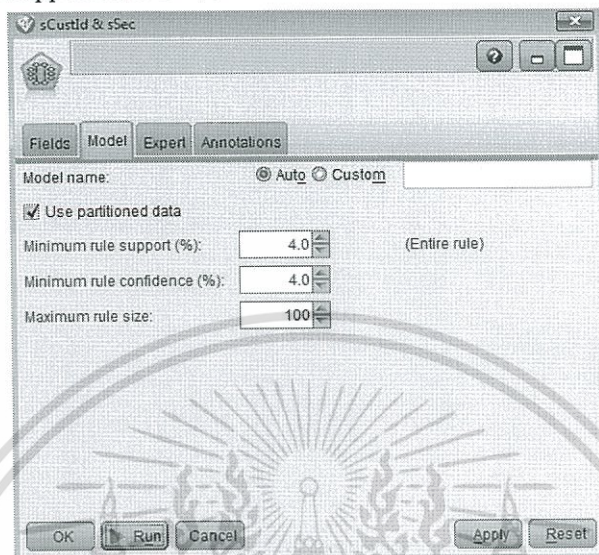
รูปที่ 4.5 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมอฟรีโอริ

จากการประมวลผลจะเห็นว่าหุ้นธนาคารมีโอกาสซื้อขายต่อเนื่องเป็น 25% จากหุ้นแต่ละกลุ่มทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 การวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุน เช่นนักลงทุนส่วนใหญ่เมื่อซื้อหุ้นหนึ่งแล้วมักจะซื้อหุ้นอีกตัวคู่กันด้วยคาร์มาอัลกอริทึม

1. ในการวิเคราะห์พฤติกรรมของนักลงทุนด้วยความสัมพันธ์ของข้อมูลซึ่งการกำหนดค่า Minimum Antecedent Support เท่ากับ 4% และ Minimum Rule Confidence เท่ากับ 4% ดังรูปที่ 4.6



รูปที่ 4.6 การตั้งค่าของอัลกอริทึมคาร์มา

2. ผลจากอัลกอริทึมคาร์มาที่แสดงค่าความสัมพันธ์ ดังรูปที่ 4.7

Consequent	Antecedent	Support %	Confidence %
PTT	PTTGC	8.28	53.845
PTT	INTUCH	8.917	50.0
INTUCH	PTT	8.917	50.0
PTTGC	PTT	8.917	50.0

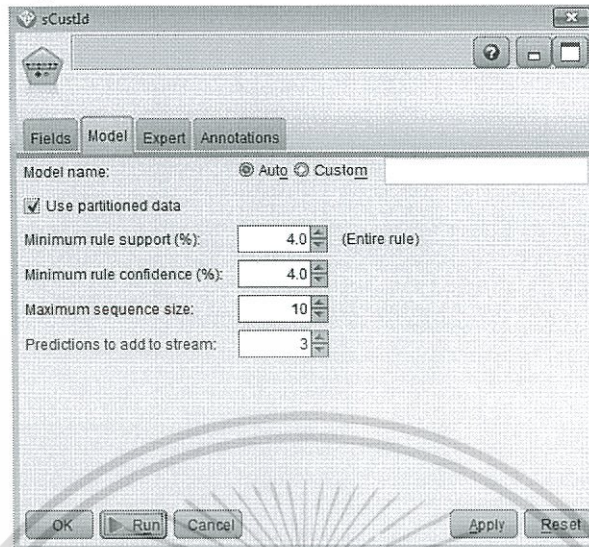
รูปที่ 4.7 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมคาร์มา

จากการประมวลผลจะเห็นว่าหุ้นพลังงาน สื่อสารมีโอกาสซื้อต่อเนื่องเป็น 50% จากหุ้นแต่ละกลุ่ม การซื้อขายทั้งหมด

4.2.3 การวิเคราะห์พฤติกรรมการซื้อขายหลักทรัพย์ของนักลงทุน เช่นนักลงทุนส่วนใหญ่เมื่อซื้อหุ้นหนึ่งแล้วมักจะซื้อหุ้นอีกตัวคู่กันด้วยซีเควนซ์อัลกอริทึม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ในการวิเคราะห์พฤติกรรมของนักลงทุนด้วยความสัมพันธ์ของข้อมูลซึ่งการกำหนดค่า Minimum Antecedent Support เท่ากับ 4% และ Minimum Rule Confidence เท่ากับ 4% ดังรูปที่ 4.8



รูปที่ 4.8 การตั้งค่าของอัลกอริทึมซีเควนซ์

2. ผลจากอัลกอริทึมซีเควนซ์ที่แสดงค่าภูควมสัมพันธ์ดังรูปที่ 4.9

Antecedent	Consequent	Support %	Confidence %
N-PARK	N-PARK	10.526	100.0

รูปที่ 4.9 ข้อมูลที่ใช้วิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึมซีเควนซ์

จากการประมวลผลจะเห็นว่าหุ้นก่อสร้าง มีโอกาสซื้อต่อเนื่องเป็น 100% จากหุ้นแต่ละกลุ่มการซื้อขายทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลของการวิเคราะห์พฤติกรรมนักลงทุนด้วยอัลกอริทึม

จากการวิเคราะห์พฤติกรรมที่ได้มานั้นทั้ง 3 อัลกอริทึมอพริโอริ คาร์มา และซีเวนซ์มีผลดังนี้

4.3.1 อัลกอริทึมอพริโอริ

เมื่อนำข้อมูลในแต่ละไตรมาสของแต่ละปีค้นพบว่าการทำงานใช้เวลาที่สูง โดยยังมี Itemset สูงหรือข้อมูลที่มากจะยังมีความสัมพันธ์ที่ทวีคูณ อัลกอริทึมอพริโอริจะมีการประมวลผลข้อมูลโดยใช้เวลาประมาณ 24 นาที ซึ่งอาจจะไม่เหมาะในแง่ความเร็วถ้าเทียบกับอัลกอริทึมคาร์มา และซีเวนซ์เพราะมีข้อมูลจำนวนมากดังรูปที่ 4.10 นอกจากนี้ อัลกอริทึมอพริโอรินั้นยังค้นพบกฎที่ว่าเมื่อซื้อหุ้น PTT แล้วนักลงทุนยังมีโอกาสซื้อหุ้น INTUCH เพิ่มถึง 33.33% ในผลลัพธ์นั้นคล้ายกับกราฟ ดังรูปที่ 4.11 ในหุ้นดังกล่าว 2 หุ้นนั้นมีความเกี่ยวข้องในเรื่องของผู้ถือหุ้นรายใหญ่ได้ถือหุ้น 2 ตัวนี้อยู่พร้อมกัน จึงทำให้นักลงทุนส่วนใหญ่เชื่อว่าหุ้นทั้ง 2 จะมีแนวโน้มขึ้น และลงพร้อมกันทำให้เกิดพฤติกรรมดังกล่าวนอกจากนั้นอัลกอริทึม อพริโอริยังพบกฎที่ว่าเมื่อซื้อหุ้นกลุ่มธนาคารแล้วยังมีโอกาสซื้อหุ้นกลุ่มธนาคารอีกถึงถึง 25% เช่น BBL กับ KBANK เนื่องมาจากช่วงปลายปีทุกธนาคารจะมีการประกาศรายได้กำไรขาดทุนนักลงทุนหลายรายจึงมีการเก็งกำไร ถ้าหากธนาคารแรกที่ประกาศนั้นมีกำไร จะมีโอกาสที่นักลงทุนจะไปซื้อหุ้นธนาคารอีกตัวเพื่อเก็งกำไรตาม ดังรูปที่ 4.12

4.3.2 อัลกอริทึมคาร์มา

เมื่อนำข้อมูลในแต่ละไตรมาสของแต่ละปีค้นพบว่าการทำงานใช้เวลาที่สั้นลงกว่าอัลกอริทึมอพริโอริและซีเวนซ์ โดยเวลาประมาณ 14 นาทีโดยประมาณเพราะว่าในโครงสร้างอัลกอริทึมคาร์มาจะมีการกรองข้อมูลขั้นตอนแรกเพื่อตัด Itemset ที่ไม่เข้าเงื่อนไขทำให้มีข้อมูลในวงจำกัด และในขั้นตอนที่ 2 นั้นการกรองข้อมูลจะเหลือข้อมูลที่น้อยลงข้อมูลที่แสดงจะเป็นข้อมูลที่มีค่าสนับสนุนและค่าความเชื่อมั่นตรงตามเป้าหมาย โดยผลลัพธ์ไม่ต่างกับ อัลกอริทึมอพริโอริ เนื่องจากอัลกอริทึมคาร์มา นั้นมีกั้นกรองถึง 2 ครั้งจึงทำให้ได้ผลลัพธ์ที่ถูกต้องมากยิ่งขึ้น นอกจากนี้ อัลกอริทึมคาร์มานั้นยังค้นพบกฎที่ว่าเมื่อซื้อหุ้น PTT ยังมีโอกาสซื้อหุ้น PTTGC ตามถึง 53.84% เนื่องจากหุ้น 2 ตัวนั้นมีความสัมพันธ์ทางธุรกิจร่วมกันจึงทำให้นักลงทุนซื้อตามกันเมื่อหุ้นใดหุ้นหนึ่งราคาขึ้น

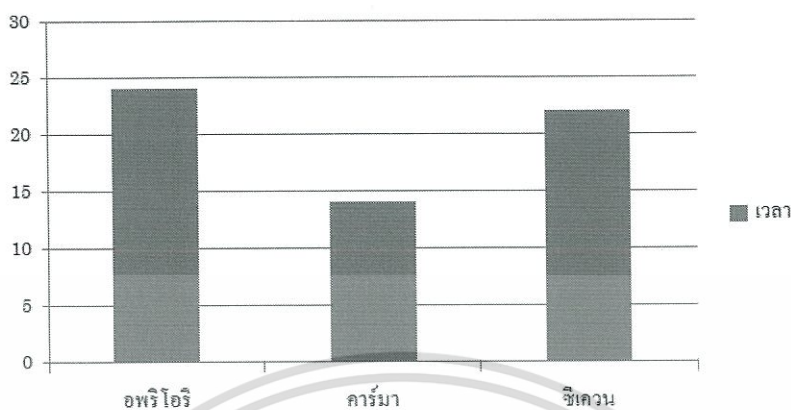
4.3.3 อัลกอริทึมซีเวนซ์

เมื่อนำข้อมูลในแต่ละไตรมาสของแต่ละปีค้นพบว่าการทำงานค่อนข้างช้า ประมาณ 22 นาที เนื่องจากอัลกอริทึมซีเวนซ์มีเงื่อนไขของความถี่หรือเวลาเข้ามาเกี่ยวข้องในการหาความสัมพันธ์ในการกำหนดค่าสนับสนุนและค่าความเชื่อมั่นที่น้อย ทำให้ความข้อมูลที่คัดกรองนั้นไม่มีความน่าเชื่อถือส่วนการใช้ ซีเวนซ์นั้นไม่เหมาะนำมาวิเคราะห์หุ้นเนื่องจากมีความหลากหลาย ข้อมูลเยอะเกินไปทำให้ความถี่นั้นมีน้อยจึงไม่สามารถได้ผลลัพธ์ที่ต้องการได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนั้นอัลกอริทึมซีเควนซ์สามารถเห็นกฎที่ว่าเมื่อซื้อหุ้น N-PARK มีโอกาสซื้อหุ้น N-PARK อีก 100% เนื่องจากหุ้นเป็นหุ้นตัวเล็กทำให้สามารถซื้อได้ง่ายเมื่อราคามีการขยับเพิ่มขึ้น

เวลาการทำงาน (นาที)

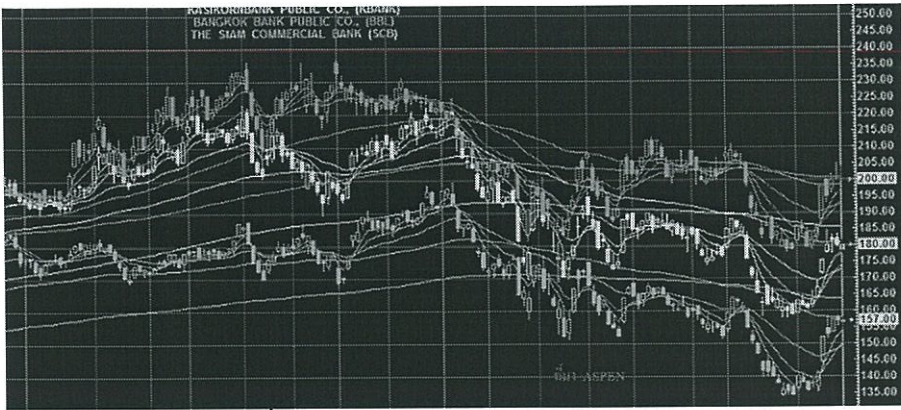


รูปที่ 4.10 เวลาที่ใช้ในการประมวลผล



รูปที่ 4.11 กราฟแสดงหุ้น PTT และหุ้น INTUCH

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 กราฟแสดงหุ้นในกลุ่มธนาคาร



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สรุปผลการศึกษา และข้อเสนอแนะ

5.1 สรุปผลการศึกษา

ในการศึกษาค้นคว้าครั้งนี้ได้นำความรู้และแนวคิดที่มีคุณค่าสำหรับการวิเคราะห์พฤติกรรมนักลงทุน โดยใช้ฐานข้อมูลที่มีอยู่มาใช้ให้เกิดกลยุทธ์ใหม่ๆ ในการดำเนินธุรกิจกองทุนภายในบริษัทหลักทรัพย์ โดยการนำข้อมูลการวิเคราะห์พฤติกรรมนั้น ได้มีการนำฐานข้อมูลในอดีตตั้งแต่ปีพ.ศ. 2551 และมีจำนวนถึง 12 ล้านกว่าเรคคอร์ด โดยในแต่ละเรคคอร์ดมีการเก็บรายละเอียดการซื้อขาย ชื่อหุ้น ประเภทการซื้อขาย หรือการขาย วันที่ส่งคำสั่งเป็นเรื่องง่ายและตรงตามเป้าหมายในการศึกษาโครงการ

ฐานข้อมูล SQL Server 2008 เป็นระบบการจัดการฐานข้อมูลที่มีความยืดหยุ่นในการใช้งานควบคู่กับโปรแกรม IBM SPSS Modeler ที่ใช้ในการวิเคราะห์พฤติกรรมนักลงทุนด้วยเทคนิคดาต้าไมน์นิ่ง โดยในกระบวนการแรกนั้นคือการดึงข้อมูลจากโปรแกรม IBM SPSS Modeler เพื่อนำเอาข้อมูลทั้งหมดในฐานข้อมูลมาเตรียมพร้อมใช้งาน คัดเลือกข้อมูลที่จะนำมาวิเคราะห์ ปรับเปลี่ยนข้อมูลบางอย่างที่ความหมายไม่ถูกต้อง สุดท้ายตรวจสอบความถูกต้องเพื่อนำข้อมูลเข้าสู่อัลกอริทึมในหลักการของดาต้าไมน์นิ่ง

การวิเคราะห์ด้วยเทคนิคดาต้าไมน์นิ่งในการศึกษาค้นคว้าครั้งนี้สามารถได้อรรถประโยชน์ที่แตกต่างจากองค์ความรู้เดิมที่มีอยู่ เป็นการช่วยเสริมในการตัดสินใจของกองทุนให้สามารถทำกำไรในการลงทุนในธุรกิจหลักทรัพย์ได้มากขึ้น ในเรื่องการทราบพฤติกรรมนักลงทุนในหุ้นต่างๆ

5.2 ข้อเสนอแนะ

ในการนำเสนอรูปแบบการพัฒนาที่วิเคราะห์พฤติกรรมของนักลงทุนนั้นจำเป็นต้องมีการเก็บข้อมูลให้ถูกต้องให้มากที่สุด เพื่อเพิ่มความน่าเชื่อถือให้กับผลลัพธ์ที่ได้จากการทำดาต้าไมน์นิ่งโดยหากมีการเพิ่มเติมการทำงานดาต้าไมน์นิ่งโดยมีข้อเสนอแนะดังต่อไปนี้

1. หากมีการเก็บข้อมูลของลูกค้าย่อยๆ ก็สามารหหาความสัมพันธ์ที่เพิ่มขึ้นได้มากขึ้นตามข้อมูลที่เพิ่มขึ้นได้
2. ปรับปรุงกระบวนการทำดาต้าไมน์นิ่งให้มีประสิทธิภาพและช่วยลดระยะเวลาการทำงานให้เร็วขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กฤษณะ ไวยมัย.2549.คลังข้อมูลและการทำเหมืองข้อมูล. เอกสารคำสอน วิชา 214554. ภาควิชา
วิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
ชาติรี โรจนอาภา.2010.การลงทุนทางการเงิน.[ออนไลน์].เข้าถึงได้จาก

http://www.moneymartthai.com/knowledge/index.php?cat=dd5c07036f2975ff4bce568b6511d3bc&know_id=40

ตลาดหลักทรัพย์แห่งประเทศไทย.2002.พฤติกรรมของนักลงทุน.[ออนไลน์].เข้าถึงได้จาก

http://www.thaimutualfund.com/AIMC/aimc_basicKnowledge.jsp?pg=4

ตลาดหลักทรัพย์แห่งประเทศไทย.2014.ปัจจัยที่มีผลต่อราคาหุ้น.[ออนไลน์].เข้าถึงได้จาก

http://www.tsi-thailand.org/index.php?option=com_content&task=view&id=1808

Anantaporn Srisawat, “An Application of Association Rule Mining Based on Stock Market,”

IEEE International Conference, pp. 259 –262 ,October 2010.

Christian Hidber, “Online Association Rule Mining”, Proceedings ACM SIGMOD

International Conference on Management of Data, pp.145-156, June 1999.

Douglas Fisher. “An Efficient Algorithm for Mining Frequent Sequences”, Machine

Learning, Vol 42, pp. 31–60, 2001.

Han, Jiawei, Kamber and Micheline. “Data Mining: Concepts and Techniques”, San

Francisco, CA:Morgan Kaufmann Publishers, 2001.

Harya Widiputra, “Inter-transaction Association Rule Mining in the Indonesia Stock

Exchange Market,” IEEE International Conference, vol. 51 , pp. 144 – 150 , September 2013.

Jian Pei and Member, “Mining Sequential Patterns by Pattern-Growth:

The PrefixSpan Approach,” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Vol 16, 2004.

Parinya Jalenga, “An association analysis of the MAI datasets by using cluster analysis and

association rule mining,” IEEE International Conference, vol. 1 , pp. 158 – 163 , October 2005.

Sheikh Shaugat Abdullah, “Stock market prediction model using TPWS and association rules

mining,” IEEE International Conference, pp. 390 – 393 , December 2012.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Yo-Ping Huang, "A prefix tree-based model for mining association rules from quantitative temporal data," IEEE International Conference, vol. 1 , pp. 158 – 163 , October 2005.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายธีระโรจน์ ศรีศาสวัตกุล
วันเดือนปีเกิด	19 มีนาคม 2531
สถานที่เกิด	โรงพยาบาลรามาริบัติ กรุงเทพมหานคร
สถานที่อยู่ปัจจุบัน	บ้านเลขที่446 ซอย 4/4 ถนนร่มเกล้า 19/1 แขวงคลองสามประเวศ เขตลาดกระบัง กรุงเทพมหานคร10520
ตำแหน่งหน้าที่การงาน	รองผู้จัดการเทคโนโลยีสารสนเทศ
สถานที่ทำงานปัจจุบัน	บริษัทหลักทรัพย์ เคเคเทรด จำกัด 500 อาคารอัมรินทร์ทาวเวอร์ ชั้น 7 ถนนเพลินจิต แขวงลุมพินี เขตประทุมวัน กรุงเทพมหานคร 10330
ประวัติการศึกษา	<p>พ.ศ.2543-2548 ระดับมัธยมศึกษาจาก โรงเรียนเตรียมอุดมศึกษาน้อมเกล้า รามคำแหง แขวงสะพานสูง เขตสะพานสูง กรุงเทพมหานคร 10240</p> <p>พ.ศ.2549-2552 ระดับปริญญาตรี วิทยาศาสตร์บัณฑิตจากมหาวิทยาลัยเทคโนโลยี มหานครหนองจอก กรุงเทพมหานคร</p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้