

การวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต

ANALYTIC OF CREDIT CARD USER'S BEHAVIOR



ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

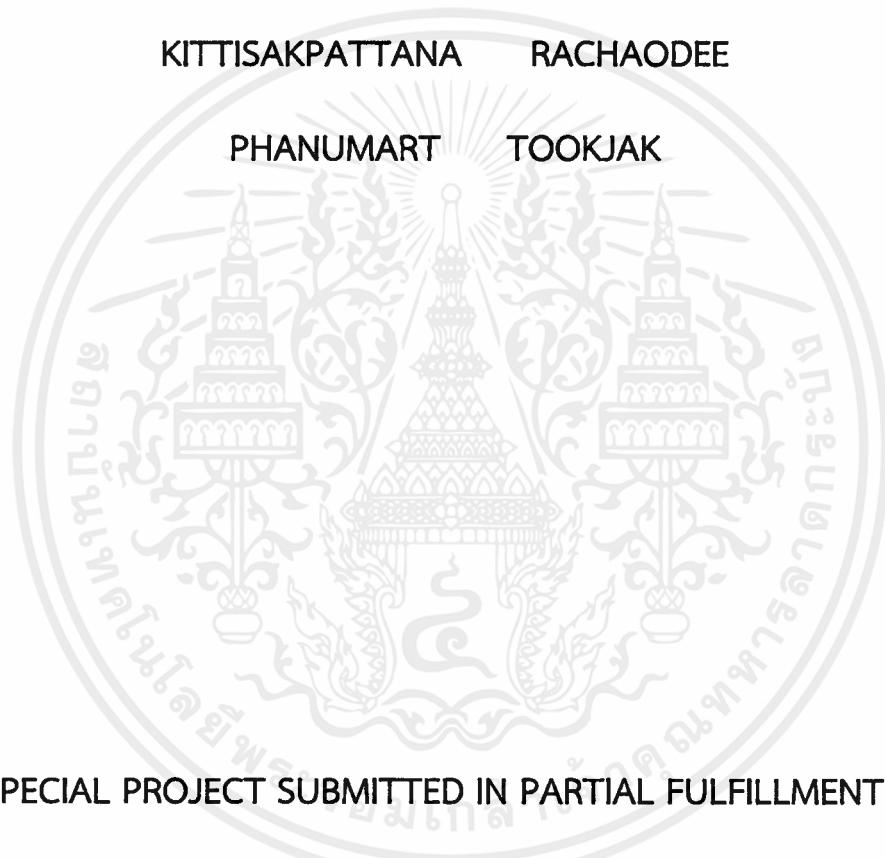
ปีการศึกษา 2561

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ANALYTIC OF CREDIT CARD USER'S BEHAVIOR

KITTISAKPATTANA RACHAODEE

PHANUMART TOOKJAK



A SPECIAL PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ การวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต
Analytic of Credit Card User's Behavior

ชื่อนักศึกษา นายกิตติศักดิ์พัฒนา ราชาสวัสดิ์ รหัสนักศึกษา 58050216
นางสาวภาณุมาศ ทุกข์จาก รหัสนักศึกษา 58050356

ปริญญา วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)

ภาควิชา วิทยาการคอมพิวเตอร์

คณะ วิทยาศาสตร์

มหาวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)

ปีการศึกษา 2561

อาจารย์ที่ปรึกษา ผศ.ดร.อนันตพร ทรรษकुณาคัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2561

คณะกรรมการสอบ	ลายมือชื่อ
ดร.อัคเดช อุดมชัยพร ประธานกรรมการ	อัคเดช
ดร.อินทราพร อรัญยานาค กรรมการ	อินทราพร อรัญยานาค
ผศ.ดร.อนันตพร ทรรษकुณาคัย กรรมการและอาจารย์ที่ปรึกษา	อนันตพร ทรรษकुณาคัย

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต			
ชื่อนักศึกษา	นายกิตติศักดิ์พัฒนา	ราเชาว์ดี	รหัสนักศึกษา	58050216
	นางสาวภานุมาศ	ทุกซ์จาก	รหัสนักศึกษา	58050356
ปริญญา	วิทยาศาสตร์บัณฑิต (วิทยาการคอมพิวเตอร์)			
ภาควิชา	วิทยาการคอมพิวเตอร์			
คณะ	วิทยาศาสตร์			
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)			
ปีการศึกษา	2561			
อาจารย์ที่ปรึกษา	ผศ.ดร.อนันตพร ทรรษकुณาฒย			

บทคัดย่อ

ปัญหาพิเศษนี้ใช้เทคนิคการทำเหมืองข้อมูลเพื่อวิเคราะห์พฤติกรรมของผู้ใช้บัตรเครดิต ในงานวิจัยนี้ใช้ชุดข้อมูลที่มีชื่อว่า Purchase Card (PCard) Fiscal Year 2014 ซึ่งประกอบไปด้วยข้อมูล 427,927 รายการ ในส่วนของการวิเคราะห์ข้อมูลจะใช้หลักการหากฎความสัมพันธ์เพื่อหาความสัมพันธ์ระหว่างหมวดหมู่การซื้อสินค้าและบริการของลูกค้าจากกลุ่มเดียวกันเช่นตามฤดูกาล และประเภทของวัน (วันธรรมดา วันหยุดสุดสัปดาห์ และวันหยุดเทศกาล) นอกจากนี้ยังมีการสร้างโมเดลต้นไม้ตัดสินใจสำหรับแต่ละหมวดหมู่สินค้าและบริการเพื่อทำนายหมวดหมู่ของการใช้จ่ายผ่านบัตรเครดิตของลูกค้า ในส่วนของเว็บแอปพลิเคชันที่พัฒนาขึ้นแบ่งผู้ใช้สองประเภทคือผู้ใช้ทั่วไป และผู้ดูแลระบบ โดยที่ผู้ใช้ทั่วไปสามารถดูกราฟทางสถิติของปริมาณการใช้งานบัตรเครดิตที่แสดงด้วยกราฟวงกลม และกราฟแท่ง นอกจากนี้ผู้ใช้ทั่วไปยังสามารถดูกฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ ได้โดยผู้ใช้งานสามารถเลือกฤดูกาล และประเภทของวันเพื่อดูกฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ และเลือกหมวดหมู่สินค้าหรือบริการเพื่อดูโมเดลต้นไม้ตัดสินใจ ส่วนผู้ดูแลระบบสามารถอัปเดตหรือลบกฎความสัมพันธ์ และโมเดลต้นไม้ตัดสินใจได้

คำสำคัญ : การวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต กฎความสัมพันธ์ ต้นไม้ตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Analytic of Credit Card User's Behavior		
Students	Mr.Kittisakpattana	Rachaodee	Student ID 58050216
	Ms.Phanumart	Tookjak	Student ID 58050356
Degree	Bachelor of Science (Computer Science)		
Department	Computer Science		
Faculty	Science		
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)		
Academic Year	2018		
Advisor	Asst.Prof.Dr.Anantaporn Hanskunatai		

Abstract

This special problem applies data mining techniques to analyze the behavior of credit card users. This work used the dataset called Purchase Card (PCard) Fiscal Year 2014 which consists of 427,927 transactions. For data analytics, association rule mining was used to discover the relationship among the categories in the same transaction grouping by season and day type (weekday, weekend, and holiday). In addition, the decision tree models were constructed for each product and service category in order to predict the type of credit card expense of the customer in the specific time. For the deployment, the web application was develop for two types of user, general users and an administrator. The general users can view statistical graphs of credit card transaction volume represented by pie chart and bar chart. Moreover, they can view the association rules by selecting the time period and choose the product or service category to view the decision tree model. On the part of the administrator, he can upload or delete the association rules and decision tree models.

Keywords : analytic of credit card user's behavior, association rule, decision tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การจัดทำปัญหาพิเศษเรื่องการวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิตนี้สำเร็จลุล่วงไปได้ด้วยดี คณะผู้จัดทำต้องขอขอบพระคุณอาจารย์ที่ปรึกษา ผศ.ดร.อนันตพร ทรราชคุณาภัย ที่ช่วยให้คำปรึกษาและคำแนะนำที่ดีแก่คณะผู้จัดทำในการพัฒนาปรับปรุงปัญหาพิเศษนี้

ขอขอบพระคุณอาจารย์ผู้ควบคุมการสอบปัญหาพิเศษ ดร.อัคเดช อุดมชัยพร และ ดร.อินทราพร อรัญยานาค ที่มีส่วนช่วยในการตรวจสอบ และให้คำแนะนำ ทำให้ปัญหาพิเศษนี้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังที่มอบโอกาสให้ได้เข้าศึกษาในสถาบันแห่งนี้ทำให้ได้พบกับคณาจารย์และบุคลากรที่มีศักยภาพ ในการช่วยพัฒนาทักษะและมอบความรู้ให้แก่นักศึกษา

กิตติศักดิ์พัฒนา ราชาวดี

ภานุมาศ ทุกข์จาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญรูป.....	ช
บทที่ 1 บทนำ.....	1
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	3
2.1 ข้อมูลขนาดใหญ่.....	3
2.2 การทำเหมืองข้อมูล.....	4
2.3 การเรียนรู้ต้นไม้ตัดสินใจ.....	6
2.4 กฎความสัมพันธ์.....	15
บทที่ 3 วิธีการดำเนินงาน.....	22
3.1 หาชุดข้อมูลและทำการศึกษาค้นคว้าข้อมูล.....	23
3.2 การเตรียมข้อมูล.....	24
3.3 การสร้างโมเดล.....	34
3.4 การพัฒนาเว็บแอปพลิเคชัน.....	39
บทที่ 4 ผลการดำเนินงาน.....	41
4.1 ผลการดำเนินงานของการหาความสัมพันธ์ของข้อมูล.....	41
4.2 ผลการดำเนินงานของการสร้างโมเดลต้นไม้ตัดสินใจ.....	46
4.3 การพัฒนาเว็บแอปพลิเคชัน.....	54
4.4 อภิปรายผลการดำเนินงาน.....	59
บทที่ 5 สรุปผลการดำเนินงานและข้อเสนอแนะ.....	62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1 สรุปผลการดำเนินงาน	62
5.2 ข้อเสนอแนะ	63
เอกสารอ้างอิง	64



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางแสดงปัจจัยที่เกี่ยวข้องกับการเล่นเทนนิส	9
2.2 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ OUTLOOK = OVERCAST	11
2.3 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ OUTLOOK = SUNNY	11
2.4 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ HUMIDITY = NORMAL	12
2.5 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ HUMIDITY = HIGH	12
2.6 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ OUTLOOK = RAIN	13
2.7 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ WIND = WEAK.....	13
2.8 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ WIND = STRONG	14
2.9 ข้อมูลการซื้อสินค้าของแต่ละรายการ.....	16
2.10 การคำนวณค่า SUPPORT ของสินค้าแต่ละชนิด.....	16
2.11 FREQUENT ITEMSET เมื่อ $K=1$	17
2.12 ค่า SUPPORT ของรายการสินค้าที่ $K=2$	17
2.13 FREQUENT ITEMSET เมื่อ $K=2$	18
2.14 FREQUENT ITEMSET เมื่อ $K=3$	18
2.15 จำนวน FREQUENT ITEMSET ทั้งหมด	19
2.16 กฎความสัมพันธ์ทั้งหมดที่สร้างได้พร้อมทั้งค่า CONFIDENCE และ LIFT	21
3.1 การกำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำในแต่ละชุดข้อมูล	37
4.1 กฎความสัมพันธ์ของชุดข้อมูลฤดูร้อน	41
4.2 กฎความสัมพันธ์ของชุดข้อมูลฤดูใบไม้ร่วง	42
4.3 กฎความสัมพันธ์ของชุดข้อมูลฤดูหนาว	42
4.4 กฎความสัมพันธ์ของชุดข้อมูลฤดูใบไม้ผลิ.....	42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 กฎความสัมพันธ์ของชุดข้อมูลต้นเดือน	43
4.6 กฎความสัมพันธ์ของชุดข้อมูลกลางเดือน	43
4.7 กฎความสัมพันธ์ของชุดข้อมูลปลายเดือน	43
4.8 กฎความสัมพันธ์ของชุดข้อมูลวันธรรมดา.....	44
4.9 กฎความสัมพันธ์ของชุดข้อมูลวันหยุดสุดสัปดาห์	44
4.10 กฎความสัมพันธ์ของชุดข้อมูลวันหยุดเทศกาล.....	45
4.11 การดำเนินงานของโมเดลต้นไม้ตัดสินใจของทุกชุดข้อมูล.....	46



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ลักษณะของข้อมูลขนาดใหญ่.....	3
2.2 กระบวนการ CRISP-DM.....	4
2.3 รูปแบบของการจัดเตรียมข้อมูล	5
2.4 ของต้นไม้ที่มีคุณลักษณะ OUTLOOK มาเป็นโหนดบน	11
2.5 ต้นไม้ตัดสินใจใช้ทำนายว่าจะมีการเล่นเทนนิสในวันนี้	14
2.6 การหาค่า CONFIDENCE ของกฎ LHS → RHS.....	20
2.7 การหาค่า CONFIDENCE ของกฎ APPLE → CEREAL.....	20
2.8 การหาค่า LIFT ของกฎ LHS → RHS.....	20
2.9 การหาค่า LIFT ของกฎ APPLE → CEREAL.....	20
3.1 ลำดับขั้นตอนการดำเนินงาน.....	22
3.2 ข้อมูลบางส่วนจากชุดข้อมูล PURCHASE CARD (PCARD) FISCAL YEAR 2014	23
3.3 ตัวอย่างข้อมูลในส่วนที่เป็นข้อมูลขาดหาย.....	24
3.4 ตัวอย่างข้อมูลในส่วนที่ทำการลดขนาดของข้อมูล	25
3.5 หมวดหมู่อสังหาริมทรัพย์	26
3.6 หมวดหมู่การศึกษา.....	26
3.7 หมวดหมู่บันเทิง.....	27
3.8 หมวดหมู่แฟชั่น	27
3.9 หมวดหมู่การแพทย์	28
3.10 หมวดหมู่ร้านอาหาร	28
3.11 หมวดหมู่บริการ.....	29
3.12 หมวดหมู่ร้านค้า.....	29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.13 หมวดหมู่เทคโนโลยี 30

3.14 หมวดหมู่การขนส่ง 30

3.15 หมวดหมู่การท่องเที่ยว 31

3.16 หมวดหมู่การบำรุงรักษารถยนต์..... 31

3.17 ตัวอย่างข้อมูลในส่วนที่เป็นคุณลักษณะใหม่..... 32

3.18 ลำดับขั้นตอนการเตรียมข้อมูลสำหรับการสร้างกฎความสัมพันธ์ 34

3.19 ตัวอย่างข้อมูลก่อนเตรียมข้อมูลเพื่อหากฎความสัมพันธ์ 35

3.20 ตัวอย่างข้อมูลหลังการรวมคุณลักษณะหมวดหมู่ที่มีผู้ถือบัตรเดียวกัน..... 36

3.21 ตัวอย่างข้อมูลหลังการเปลี่ยนชนิดของข้อมูลคุณลักษณะหมวดหมู่..... 36

3.22 ลำดับขั้นตอนการเตรียมข้อมูลสำหรับสร้างโมเดลต้นไม้ตัดสินใจ 37

3.23 ตัวอย่างชุดข้อมูลก่อนเตรียมข้อมูล 38

3.24 ตัวอย่างชุดข้อมูลหลังเตรียมข้อมูลเพื่อนำไปสร้างโมเดลต้นไม้ตัดสินใจ..... 39

4.1 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่บริการ..... 47

4.2 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยี 49

4.3 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้า..... 52

4.4 หน้าแสดงผลภาพรวมของข้อมูลด้วยกราฟรูปแบบกราฟวงกลม 54

4.5 หน้าแสดงผลภาพรวมของข้อมูลด้วยกราฟรูปแบบกราฟแท่ง 55

4.6 หน้าแสดงผลกฎความสัมพันธ์..... 56

4.7 หน้าแสดงผลโมเดลต้นไม้ตัดสินใจ 56

4.8 หน้าเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบ 57

4.9 หน้าเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบในกรณีการยืนยันผิดพลาด..... 57

4.10 หน้าการนำเข้ากฎความสัมพันธ์และการนำเข้าโมเดลต้นไม้ตัดสินใจ 58

4.11 โมเดลต้นไม้ตัดสินใจหมวดหมู่การศึกษา..... 61

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันธุรกิจบัตรเครดิตมีแนวโน้มที่อึมทัว ธนาคารต่าง ๆ ต้องเปลี่ยนแผนการตลาด และมุ่งเน้นการกระตุ้นยอดใช้จ่ายผ่านบัตรเครดิต นายโชค ณ ระนอง ผู้ช่วยกรรมการผู้จัดการใหญ่ธนาคารกรุงเทพ จำกัด (มหาชน) กล่าวว่า การลดลงของจำนวนบัตรเครดิตจากตัวเลขของธนาคารแห่งประเทศไทย สะท้อนถึงความอึมทัวของตลาดบัตรเครดิต ส่วนหนึ่งอาจเป็นผลจากสถาบันการเงินไม่เร่งขยายฐานบัตรเครดิตหรือตัดบัตรเครดิตที่ไม่มียอดการใช้จ่ายสม่ำเสมอออกไป ซึ่งจากผลวิจัยพบว่า ลูกค้ำ 1 คนจะถือบัตรเครดิตอย่างน้อย 3 ใบ และอาจจะทำให้ลูกค้ำนั้นไม่เลือกใช้บัตรเครดิตของธนาคาร และไปเลือกใช้บัตรเครดิตของธนาคารอื่น อีกทั้งเมืองไทยเข้าสู่สังคมผู้สูงอายุผู้สูงอายุก็จะลดการใช้จ่ายลง แม้ทุกค่ายจะมีกลยุทธ์เจาะกลุ่มผู้สูงอายุก็ตาม

บัตรเครดิตนั้นถือว่าเป็นสินเชื่อ สินเชื่อหนึ่งที่สามารถปล่อยให้กับบุคคลไหนก็ได้ ไม่จำเป็นต้องเป็นผู้ประกอบการเท่านั้น ดังนั้นลูกค้ำบัตรเครดิตนั้นถือว่ามีความสำคัญต่อธนาคารอย่างมาก ทั้งดอกเบี้ยจากการที่ลูกค้ำใช้จ่ายผ่านบัตรเครดิต ค่าธรรมเนียมบัตรเครดิตรายปีที่ลูกค้ำต้องจ่ายให้กับธนาคาร

ดังนั้นปัญหาพิเศษนี้จึงจัดทำการวิเคราะห์พฤติกรรมการใช้งานบัตรเครดิตเพื่อส่งเสริมการใช้บัตรเครดิตของธนาคาร โดยใช้กฎความสัมพันธ์ แสดงความสัมพันธ์ของพฤติกรรมกาซื้อสินค้าและบริการที่ลูกค้ำบัตรเครดิตมักจะซื้อพร้อมกัน และใช้เทคนิคการจำแนกประเภทของข้อมูล โดยการนำข้อมูลพฤติกรรมกาใช้บัตรเครดิตของลูกค้ำปัจจุบัน มาทำนายเพื่อจำแนกประเภทของข้อมูลลูกค้ำใหม่

1.2 วัตถุประสงค์ของการวิจัย

- 1) ศึกษาและวิเคราะห์พฤติกรรมของลูกค้ำบัตรเครดิตโดยประยุกต์ใช้กฎความสัมพันธ์
- 2) ศึกษาและวิเคราะห์พฤติกรรมกาใช้บัตรเครดิตในการซื้อสินค้าและบริการของลูกค้ำด้วยการประยุกต์ใช้โมเดลต้นไม้ตัดสินใจ
- 3) พัฒนาเว็บแอปพลิเคชันเพื่อใช้สำหรับผู้บริหารเพื่อดูกราฟทางสถิติและวิเคราะห์พฤติกรรมกาใช้บัตรเครดิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตการศึกษา

- 1) ชุดข้อมูลที่ใช้ ชื่อชุดข้อมูลว่า Purchase Card (PCard) Fiscal Year 2014 จากเว็บไซต์ <https://catalog.data.gov/dataset/purchase-card-pcard-fiscal-year-2014>
- 2) ใช้ภาษาไพธอนในการพัฒนาเว็บแอปพลิเคชันเพื่อใช้สำหรับผู้บริหารเพื่อดูกราฟทางสถิติและวิเคราะห์พฤติกรรมการใช้บัตรเครดิต

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ธนาคารได้ทราบข้อมูลทางสถิติเกี่ยวกับข้อมูลลูกค้าบัตรเครดิตและพฤติกรรมการใช้บัตรเครดิตของลูกค้า
- 2) ธนาคารสามารถนำเสนอโปรโมชั่นบัตรเครดิตเพื่อจูงใจการใช้จ่ายผ่านบัตรเครดิตของธนาคารตามพฤติกรรมของลูกค้า

1.5 ขั้นตอนการดำเนินงาน

- 1) ศึกษาวิธีการทำงานของการทำเหมืองข้อมูล
- 2) เตรียมความพร้อมของข้อมูล
- 3) วิเคราะห์ข้อมูลโดยใช้การสร้างกฎความสัมพันธ์
- 4) วิเคราะห์ข้อมูลโดยใช้การสร้างโมเดลต้นไม้ตัดสินใจ
- 5) สรุปผลการวิเคราะห์ข้อมูล
- 6) พัฒนาเว็บแอปพลิเคชัน
- 7) จัดทำเอกสารประกอบปัญหาพิเศษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

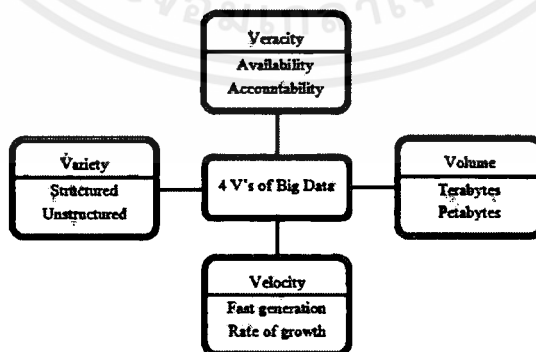
บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ซึ่งจะประกอบไปด้วย ข้อมูลขนาดใหญ่ การทำเหมืองข้อมูล การเรียนรู้ต้นไม้ตัดสินใจ และกฎความสัมพันธ์ โดยที่จะแสดงรายละเอียดเพิ่มเติมในหัวข้อ 2.1 ถึง 2.5

2.1 ข้อมูลขนาดใหญ่

ในปัจจุบันข้อมูลถูกสร้างขึ้นจากแหล่งต่าง ๆ และมีการเปลี่ยนแปลงอย่างรวดเร็วจากเทคโนโลยีดิจิทัล ได้นำไปสู่การเติบโตของข้อมูลขนาดใหญ่ (Big Data) โดยทั่วไปข้อมูลขนาดใหญ่หมายถึงชุดข้อมูลที่มีขนาดใหญ่และชุดข้อมูลที่ซับซ้อนซึ่งยากที่จะประมวลผลโดยใช้เครื่องมือการจัดการฐานข้อมูลแบบเดิมหรือแอปพลิเคชันการประมวลผลข้อมูล ข้อมูลเหล่านี้มีอยู่ในรูปแบบโครงสร้างที่มี โครงสร้าง กึ่งโครงสร้าง และไม่มีโครงสร้าง ตามหลักการก็คือการกำหนดจาก 3Vs เป็น 4Vs โดยที่ 3Vs หมายถึง ปริมาตร (Volume) ความเร็ว (Velocity) และความหลากหลาย (Variety) ปริมาณหมายถึงข้อมูลจำนวนมากที่มีการสร้างขึ้นทุกวัน ในขณะที่ความเร็วคืออัตราการเติบโตและความเร็วในการรวบรวมข้อมูลเพื่อการวิเคราะห์ ความหลากหลายให้ข้อมูลเกี่ยวกับประเภทของข้อมูลเช่นมีโครงสร้าง ไม่มีโครงสร้าง และกึ่งโครงสร้าง เป็นต้น V ตัวที่ 4 หมายถึง ความจริง (Veracity) ที่ประกอบด้วยความพร้อมใช้งานและตรวจสอบได้ แสดงในรูปที่ 2.1



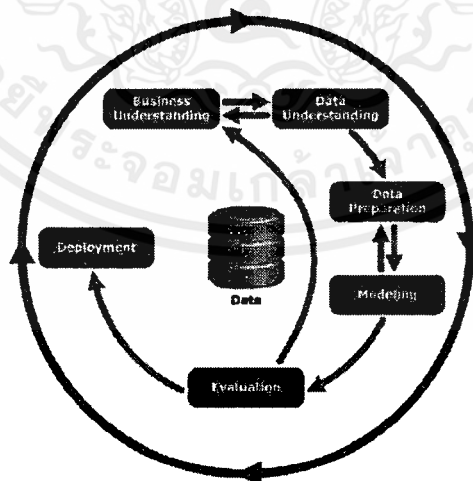
รูปที่ 2.1 ลักษณะของข้อมูลขนาดใหญ่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างของการใช้ข้อมูลขนาดใหญ่ของบริษัทชั้นนำต่าง ๆ เช่น บริษัท Amazon บริษัทยักษ์ใหญ่ทางด้านขายปลีกที่มีข้อมูลลูกค้าให้จัดการมากมาย ทั้งข้อมูลชื่อลูกค้า ที่อยู่ ข้อมูลการชำระเงิน และประวัติการค้นหา โดยที่บริษัท Amazon ได้เอามาใช้ในการพัฒนาความสัมพันธ์กับลูกค้าซึ่งต้องอาศัยข้อมูลขนาดใหญ่เป็นตัวช่วย บริษัท Netflix เป็นบริษัทผู้ให้บริการสตรีมมิ่งรายใหญ่และเป็นบริษัทที่มีข้อมูลมากมายเพื่อใช้ในการวิเคราะห์หาข้อมูลเชิงลึก ในการดูพฤติกรรมการใช้งานของลูกค้าหลายล้านคนทั่วโลก และมีการใช้ข้อมูลวิเคราะห์ที่ได้ในการสร้างโปรแกรมการฉายหนังที่ถูกใจลูกค้า และใช้เพื่อดูว่าภาพยนตร์เรื่องไหนที่ควรซื้อลิขสิทธิ์เอามาลงใน Netflix

2.2 การทำเหมืองข้อมูล

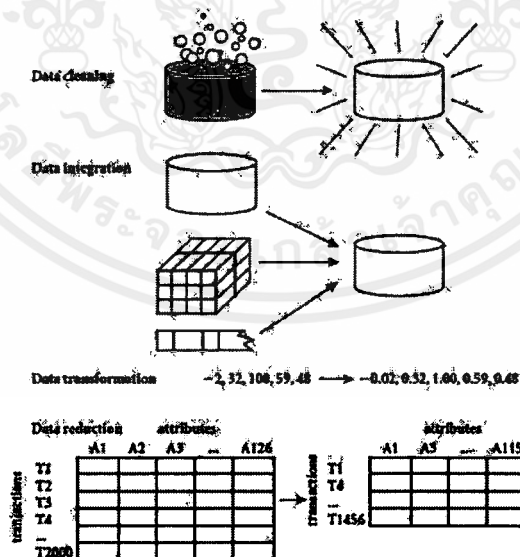
การทำเหมืองข้อมูลเป็นเทคนิควิธีการที่มีเป้าหมายเพื่อค้นหารูปแบบ และความสัมพันธ์ของข้อมูลที่ซ่อนอยู่จากข้อมูลจำนวนมากโดยอัตโนมัติ โดยใช้วิธีจากสถิติ การเรียนรู้ของเครื่อง การรู้จำ และหลักคณิตศาสตร์ โดยที่การทำเหมืองข้อมูลมีกระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านการทำเหมืองข้อมูลที่พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler SPSS และ NCR กระบวนการทำงานนี้เรียกว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อว่า “CRISP-DM” โดยในกระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอน แสดงดังรูปที่ 2.2



รูปที่ 2.2 กระบวนการ CRISP-DM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนในกระบวนการ CRISP-DM มีดังนี้ 1) การทำความเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาที่ได้ให้อยู่ในรูป โจทย์ของการวิเคราะห์ข้อมูลด้านการทำเหมืองข้อมูล พร้อมทั้งวางแผนในการดำเนินการ 2) การทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนการเก็บรวบรวมข้อมูล และตรวจสอบข้อมูลที่ได้ทำการ รวบรวมมา เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูล บางส่วนมาใช้ในการวิเคราะห์ข้อมูล 3) การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ทำการแปลง ข้อมูลที่ได้ทำการเก็บรวบรวมมาให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ได้ โดยขั้นตอนนี้จะเป็นขั้นตอนที่ ใช้เวลามากที่สุดของกระบวนการ CRISP-DM เนื่องจากข้อมูลที่มีในความเป็นจริงเป็นข้อมูลที่ไม่สมบูรณ์ เช่น ค่าของคุณลักษณะขาดหาย ขาดคุณลักษณะที่น่าสนใจ ขาดรายละเอียดของข้อมูล ข้อมูลรบกวน เช่น ข้อมูลที่มีค่าผิดพลาด หรือที่มีค่าผิดปกติ ข้อมูลไม่สอดคล้อง เช่น ข้อมูลเดียวกันแต่ตั้งชื่อต่างกัน หรือ ใช้ค่าแทนข้อมูลที่ ต่างกัน โดยที่การจัดเตรียมข้อมูลมีเทคนิคการเตรียมข้อมูลดังนี้ การทำความสะอาดข้อมูล เป็นการจัดการ ข้อมูลรบกวน เช่น การเติมค่าที่ขาดหาย การระบุให้เป็นค่าที่รบกวน และการแก้ไขความซ้ำซ้อนที่เกิดจากการ รวบรวมข้อมูล การผสานข้อมูล เป็นการรวบรวมข้อมูลจากหลายแห่งที่เชื่อมโยงกัน การแปลงข้อมูล เป็นการ เปลี่ยนขนาดของข้อมูล และการลดรูปข้อมูล เป็นการลดรูปของข้อมูลเพื่อแทนข้อมูลด้วยรูปแบบกะทัดรัด และประหยัดเนื้อที่เก็บข้อมูล แต่ยังสามารถนำไปใช้วิเคราะห์ แล้วยังเกิดผลลัพธ์ที่เท่ากัน หรือเทียบเท่ากัน กับ การที่วิเคราะห์จากข้อมูลทั้งหมด โดยที่รูปแบบของการจัดเตรียมข้อมูลแสดงในรูปที่ 2.3



รูปที่ 2.3 รูปแบบของการจัดเตรียมข้อมูล

4) สร้างแบบจำลอง (Modeling) นำผลที่ได้มาทดลองทำแบบจำลอง เพื่อใช้ทำการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปสารสนเทศเป็นข้อมูลที่น่าไปใช้ประโยชน์ได้จริงในทางธุรกิจ 5) การประเมินผล (Evaluation) ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก การทดสอบประสิทธิภาพของโมเดลที่ได้รับความนิยม คือ Cross-validation test 6) การนำไปใช้ (Deployment) คือการนำเอาข้อมูลที่เป็นผลลัพธ์มาลองปฏิบัติจริงกับธุรกิจ โดยแปลงแนวคิดที่มีให้เกิดเป็นการนำไปใช้ประโยชน์ และติดตามรวบรวมผลที่ได้เพื่อการปรับปรุงต่อไป

2.3 การเรียนรู้ต้นไม้ตัดสินใจ

การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning) จัดเป็นอัลกอริทึมที่ใช้แก้ปัญหาทางด้านการจำแนกประเภทของข้อมูล (Data Classification) ซึ่งเป็นประเภทย่อยของการเรียนรู้แบบมีผู้สอน โดยที่การจำแนกประเภทข้อมูลมีเป้าหมาย คือ การทำนายประเภทของคลาสคำตอบในข้อมูลใหม่ โดยสังเกตจากข้อมูลในอดีต หรือในอีกนิยามหนึ่งของการเรียนรู้แบบมีผู้สอน เป็นการสร้างฟังก์ชันเชื่อมโยงระหว่างคุณสมบัติของข้อมูลกับประเภทของข้อมูลจากตัวอย่างสอน แล้วจึงใช้ฟังก์ชันนี้ทำนายประเภทของข้อมูลที่ไม่เคยพบ โดยการจำแนกประเภทข้อมูลจะสามารถใช้ได้กับข้อมูลประเภท ข้อมูลดิสครีต และข้อมูลนอมินอลเท่านั้น เครื่องมือหรืออัลกอริทึมที่ใช้สำหรับการจำแนกประเภทข้อมูล เช่น โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจ และวิธีการเรียนรู้แบบอย่างง่าย เป็นต้น

การเรียนรู้ต้นไม้ตัดสินใจ เป็นวิธีหนึ่งที่เป็นการประมาณฟังก์ชันที่มีค่าไม่ต่อเนื่อง ด้วยแผนผังต้นไม้ โดยที่อาจจะประกอบด้วยเซตของกฎต่าง ๆ แบบ ถ้า-แล้ว เพื่อที่เราจะสามารถเข้าใจการตัดสินใจของต้นไม้ได้ง่าย โดยที่การเรียนรู้ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของข้อมูลโดยที่จะพิจารณาจากลักษณะของข้อมูล โดยที่โหนดภายในของต้นไม้จะแสดงถึงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร และส่วนโหนดใบจะแสดงประเภทของข้อมูล ซึ่งปัจจุบันในการบริหารธุรกิจมีการนำต้นไม้ตัดสินใจไปใช้ในการบริหารธุรกิจ โดยที่แผนผังต้นไม้ช่วยในการตัดสินใจจะแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและ และผลลัพธ์ที่มีโอกาสเกิดขึ้น นิยมใช้มากในการบริหารความเสี่ยง ต้นไม้ตัดสินใจถือว่าเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ และทฤษฎีกราฟ ซึ่งต้นไม้ตัดสินใจถือว่าเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล ลักษณะของต้นไม้ตัดสินใจถือว่าเป็นการทำกรแบ่งกลุ่ม โดยที่ชุดข้อมูลนำเข้าในแต่ละโหนดของต้นไม้ตัดสินใจ คือตัวแปรต่าง ๆ ของชุดข้อมูล เช่น ถ้าหากว่าต้องการตัดสินใจว่า มีคนไปเล่นกีฬาหรือไม่ ซึ่งจะมีตัวแปรต้นที่จะต้องพิจารณา คือ ทัศนียภาพ ลม ความชื้น และอุณหภูมิ เป็นต้น และตัวแปรตามซึ่งเป็นเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลลัพธ์จากต้นไม้ คือ การตัดสินใจว่ามีคนไปเล่นกีฬาหรือไม่ ซึ่งแต่ละตัวแปรจะมีค่าของตัวแปร เป็นชุดของตัวแปร ค่าของตัวแปร เช่น ทักษะกีฬาเป็นตัวแปร สามารถมีค่าเป็น ฝนตก แดดออก โดยที่การทำนายประเภทด้วยต้นไม้ตัดสินใจ จะเริ่มจากโหนดราก โดยทดสอบค่าตัวแปรของโหนดแล้วจึงตามกิ่งของต้นไม้ที่กำหนดค่าเพื่อไปยังโหนดลูกถัดไป การทดสอบนี้จะกระทำไปจนกระทั่งเจอโหนดใบซึ่งจะแสดงเป็นผลการทำนายประเภท โดยที่การเรียนรู้ต้นไม้ตัดสินใจมีอัลกอริทึมการเรียนรู้ดังอัลกอริทึมต่อไปนี้ ซึ่งจะแสดงในรูปแบบชุดโค้ด

INPUT: S , where $S = \text{set of classified instances}$

OUTPUT: *Decision Tree*

Require: $S \neq \emptyset$, $\text{num_attributes} > 0$

```

1: procedure BUILD TREE
2:   repeat
3:      $\text{maxGain} \leftarrow 0$ 
4:      $\text{splitA} \leftarrow \text{null}$ 
5:      $e \leftarrow \text{Entropy}(\text{Attributes})$ 
6:     for all Attributes  $a$  in  $S$  do
7:        $\text{gain} \leftarrow \text{InformationGain}(a, e)$ 
8:       if  $\text{gain} > \text{maxGain}$  then
9:          $\text{maxGain} \leftarrow \text{gain}$ 
10:         $\text{splitA} \leftarrow a$ 
11:      end if
12:    end for
13:     $\text{Partition}(S, \text{splitA})$ 
14:  until all partitions processed
15: end procedure

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสร้างต้นไม้ตัดสินใจคือการทำการคัดเลือกคุณลักษณะ (Attribute) ที่มีความสัมพันธ์กับคำตอบ (Label) มากที่สุดขึ้นเป็นโหนดบนสุดของต้นไม้ (root node) หลังจากนั้นก็จะทำการคัดเลือกคุณลักษณะถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของคุณลักษณะใช้ตัววัดที่เรียกว่า เกน (Gain) สามารถคำนวณได้ตามสมการที่ 2.1

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2.1)$$

เมื่อ S คือ ตัวอย่างที่ประกอบด้วยชุดข้อมูลหลาย ๆ กรณี

E คือ เอนโทรปีของตัวอย่าง

A คือ คุณลักษณะที่ต้องการพิจารณา

$value(A)$ คือ เซตของค่าของ A ที่เป็นไปได้

S_v คือ ตัวอย่างที่ A มีค่า v ทั้งหมด

ในการคำนวณหาค่าเกนนั้นจะต้องค่าเอนโทรปี (Entropy) โดยที่ค่าเอนโทรปีสามารถคำนวณได้ตามสมการที่ 2.2

$$E(s) = - \sum_{j=1}^n P_S(j) \log_2 P_S(j) \quad (2.2)$$

เมื่อ S คือ ตัวอย่างที่ประกอบด้วยชุดของข้อมูลหลาย ๆ กรณี

$P_S(j)$ คือ อัตราส่วนของกรณี S ที่ผลลัพธ์ j

ข้อมูลตัวอย่างตามตารางที่ 2.1 ตารางนี้แสดงเกี่ยวกับปัจจัยที่เกี่ยวข้องในการเล่นเทนนิสที่โดยเป็นข้อมูลของ 14 วันที่ผ่านมา

ตารางที่ 2.1 ตารางแสดงปัจจัยที่เกี่ยวข้องกับการเล่นเทนนิส

Day	Outlook	Temp	Humidity	Wind	Label
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

การคำนวณหาค่าเกนนั้นจะต้องค่าเอนโทรปีก่อน โดยที่การตัดสินใจประกอบด้วย 14 ตัวอย่าง (instances) และคลาสคำตอบ (class label) คือ Yes, No โดยที่มีคำตอบ 9 ตัวอย่างระบุว่า Yes และคำตอบ 5 ตัวอย่างระบุว่า No โดยการคำนวณจะเป็นดังนี้

$$\text{Entropy (S)} = - p(\text{Yes}) \log_2 p(\text{Yes}) - p(\text{No}) \log_2 p(\text{No})$$

$$\text{Entropy (S)} = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14) = 0.940$$

การคำนวณเกนของคุณลักษณะ Wind ต่อการตัดสินใจ โดยคุณลักษณะ Wind มีสองคำตอบ คือ Weak กับ Strong

Gain (S, Wind) = Entropy (S) - $\sum [p(S | \text{Wind}) \text{Entropy (S | Wind)}]$ โดยที่การคำนวณนั้นต้องคำนวณ (S | Wind = Weak) และ (S | Wind = Strong) ตามลำดับ โดยที่มี 8 กรณีสำหรับ Wind = Weak

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำตอบของ 2 ตัวอย่างคือ No และ 6 ตัวอย่างคือ Yes และมี 6 กรณีสำหรับ Wind = Strong คำตอบของ 3 ตัวอย่างคือ No และ 3 ตัวอย่างคือ Yes โดยการคำนวณจะเป็นดังนี้

$$\begin{aligned} \text{Entropy (S | Wind = Weak)} &= -p(\text{No}) \log_2 p(\text{No}) - p(\text{Yes}) \log_2 p(\text{Yes}) \\ &= -(2/8) \log_2 (2/8) - (6/8) \log_2 (6/8) = 0.811 \end{aligned}$$

$$\begin{aligned} \text{Entropy (S | Wind = Strong)} &= -p(\text{No}) \log_2 p(\text{No}) - p(\text{Yes}) \log_2 p(\text{Yes}) \\ &= -(3/6) \log_2 (3/6) - (3/6) \log_2 (3/6) = 1 \end{aligned}$$

$$\begin{aligned} \text{Gain (S, Wind)} &= \text{Entropy (S)} - [p(\text{S | Wind = Weak}) \text{Entropy (S | Wind = Weak)}] - [p(\text{S |} \\ \text{Wind = Strong}) \text{Entropy (S | Wind = Strong)}] \\ &= 0.940 - [(8/14) 0.811] - [(6/14) 1] = 0.048 \end{aligned}$$

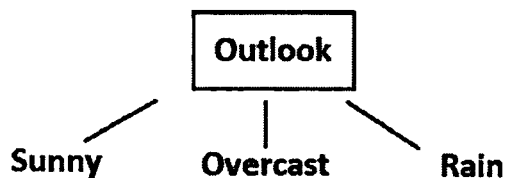
การคำนวณต้องใช้การคำนวณแบบเดียวกันสำหรับคุณลักษณะอื่น ๆ เพื่อค้นหาปัจจัยที่สำคัญที่สุดในคำตอบหรือคุณลักษณะที่มีความสัมพันธ์กับคำตอบมากที่สุด โดยในการคำนวณสำหรับคุณลักษณะอื่น ๆ ได้ค่า Gain ดังนี้

$$\text{Gain (S, Outlook)} = 0.246$$

$$\text{Gain (S, Temp)} = 0.029$$

$$\text{Gain (S, Humidity)} = 0.151$$

จากการคำนวณ Gain ของค่าคุณลักษณะทั้งหมด คุณลักษณะที่มีความสัมพันธ์กับคำตอบมากที่สุด โดยดูจาก Gain ที่มีค่าสูงสุด เป็น Outlook ดังนั้นนำคุณลักษณะ Outlook มาเป็นโหนดบนสุดของต้นไม้ แสดงดังรูปที่ 2.4



รูปที่ 2.4 ของต้นไม้ที่ไม่มีคุณลักษณะ Outlook มาเป็นโหนดบน

คุณลักษณะของ Outlook = Overcast โดยมีทั้งหมด 4 กรณีและคำตอบเป็น Yes ทุกกรณี ดังนั้น Outlook = Overcast จะมีคำตอบเป็น Yes เสมอ แสดงตามตารางที่ 2.2

ตารางที่ 2.2 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Outlook = Overcast

Day	Outlook	Temp	Humidity	Wind	Label
3	Overcast	Hot	High	Weak	Yes
7	Overcast	Cool	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes

ส่วนคุณลักษณะของ Outlook = Sunny โดยมีทั้งหมด 5 กรณีและคำตอบของ 3 ตัวอย่างคือ No และ 2 ตัวอย่างคือ Yes ดังนั้น แสดงตามตารางที่ 2.3

ตารางที่ 2.3 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Outlook = Sunny

Day	Outlook	Temp	Humidity	Wind	Label
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการคำนวณ Gain (Outlook = Sunny) ได้ดังนี้

$$\text{Gain (Outlook = Sunny | Temp)} = 0.570$$

$$\text{Gain (Outlook = Sunny | Humidity)} = 0.970$$

$$\text{Gain (Outlook = Sunny | Wind)} = 0.019$$

จากการคำนวณ Gain ของค่าคุณลักษณะ คุณลักษณะที่มีความสัมพันธ์กับคำตอบมากที่สุดโดยดูจาก Gain ที่มีค่าสูงสุด เป็น Humidity ดังนั้นนำคุณลักษณะ Humidity มาเป็นโหนดต่อจากกิ่งของ Outlook = Sunny

คุณลักษณะของ Humidity = Normal โดยมีทั้งหมด 2 กรณีและคำตอบเป็น Yes ทุกกรณี ดังนั้น Humidity = Normal จะมีคำตอบเป็น Yes เสมอ แสดงตามตารางที่ 2.4

ตารางที่ 2.4 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Humidity = Normal

Day	Outlook	Temp	Humidity	Wind	Label
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

คุณลักษณะของ Humidity = High โดยมีทั้งหมด 2 กรณีและคำตอบเป็น No ทุกกรณี ดังนั้น Humidity = High จะมีคำตอบเป็น No เสมอ แสดงตามตารางที่ 2.5

ตารางที่ 2.5 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Humidity = High

Day	Outlook	Temp	Humidity	Wind	Label
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนคุณลักษณะของ Outlook = Rain โดยมีทั้งหมด 5 กรณีและคำตอบของ 2 ตัวอย่างคือ No และ 3 ตัวอย่างคือ Yes แสดงตามตารางที่ 2.6

ตารางที่ 2.6 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Outlook = Rain

Day	Outlook	Temp	Humidity	Wind	Label
4	Rain	Mild	Normal	Weak	Yes
5	Rain	Cool	High	Weak	Yes
6	Rain	Cool	High	Strong	No
10	Rain	Mild	High	Weak	Yes
14	Rain	Mild	Normal	Strong	No

จากการคำนวณ Gain (Outlook = Rain) คุณลักษณะที่มีความสัมพันธ์กับคำตอบมากที่สุดโดยดูจาก Gain ที่มีค่าสูงสุด เป็น Wind ดังนั้นนำคุณลักษณะ Wind มาเป็นโหนดต่อจากกิ่งของ Outlook = Rain

คุณลักษณะของ Wind = Weak โดยมีทั้งหมด 3 กรณีและคำตอบเป็น Yes ทุกกรณี ดังนั้น Wind = Weak จะมีคำตอบเป็น Yes เสมอ แสดงตามตารางที่ 2.7

ตารางที่ 2.7 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Wind = Weak

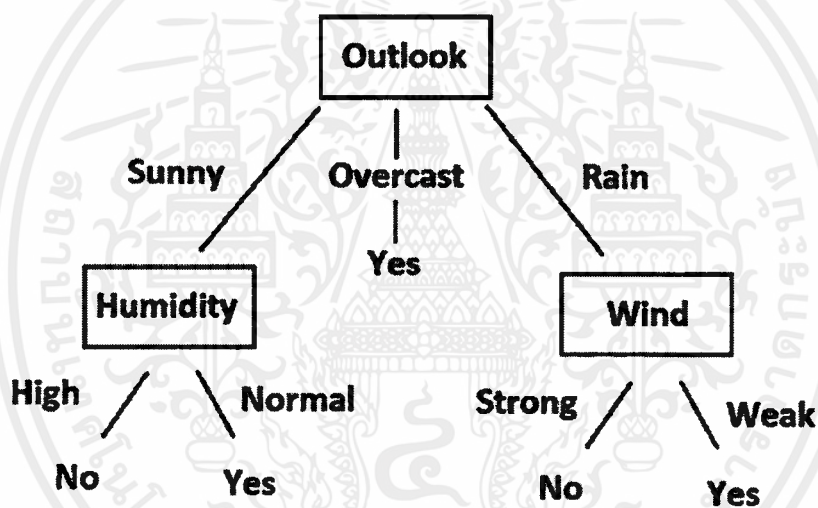
Day	Outlook	Temp	Humidity	Wind	Label
4	Rain	Mild	Normal	Weak	Yes
5	Rain	Cool	High	Weak	Yes
10	Rain	Mild	High	Weak	Yes

คุณลักษณะของ Wind = Strong โดยมีทั้งหมด 2 กรณีและคำตอบเป็น No ทุกกรณี ดังนั้น Wind = Strong จะมีคำตอบเป็น No เสมอ แสดงตามตารางที่ 2.8

ตารางที่ 2.8 ตารางแสดงปัจจัยที่เกี่ยวข้องโดยสนใจเฉพาะ Wind = Strong

Day	Outlook	Temp	Humidity	Wind	Label
6	Rain	Cool	High	Strong	No
14	Rain	Mild	Normal	Strong	No

พอถึงขั้นตอนนี้จะได้ต้นไม้ตัดสินใจดังรูปที่ 2.5 และสามารถใช้ในการทำนายประเภทของข้อมูลที่ใหม่ได้ว่าจะมีการเล่นเทนนิสในวันนี้ โดยดูจากข้อมูล Outlook, Humidity และ Wind ตามต้นไม้ตัดสินใจที่สร้างขึ้นมา



รูปที่ 2.5 ต้นไม้ตัดสินใจใช้ทำนายว่าจะมีการเล่นเทนนิสในวันนี้

2.4 กฎความสัมพันธ์

การสร้างกฎความสัมพันธ์เป็นกระบวนการหนึ่งในการทำเหมืองข้อมูลที่ได้รับความนิยมอย่างมาก ตัวอย่างหนึ่งของกฎความสัมพันธ์ ที่นิยมก็คือ การวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) ที่ใช้ในการหาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อพร้อมกัน เพื่อใช้ในการจัดรายการส่งเสริมการขาย โดยที่กฎความสัมพันธ์นำมาใช้ในธุรกิจขายปลีกอย่างกว้างขวาง เช่น เว็บไซต์ Amazon และเว็บไซต์ขายของออนไลน์ จากทั้งต่างประเทศและในประเทศไทย โดยการสร้างกฎความสัมพันธ์จะค้นหาเหตุการณ์ที่เกิดขึ้นบ่อย ๆ โดยที่ถ้าการพิจารณากฎ $A \rightarrow B$ หมายถึงการซื้อสินค้า A แล้วซื้อสินค้า B จะเกิดขึ้นก็เปอร์เซ็นต์ เหตุการณ์นี้จะมีเชื่อมั่นได้ก็เปอร์เซ็นต์ ผลลัพธ์ที่ได้จะถูกนำไปใช้แนะนำลูกค้าให้ซื้อสินค้าเพิ่ม เพื่อเพิ่มยอดขาย นอกจากนี้ยังมีการประยุกต์ใช้เทคนิคนี้อีกมากมาย เช่น การแนะนำความสัมพันธ์บนเครือข่ายสังคมออนไลน์ที่มีโปรไฟล์ของสมาชิกใกล้เคียงกัน ซึ่งกฎความสัมพันธ์ถือว่าเป็นอัลกอริทึมการเรียนรู้แบบไม่มีผู้สอน หรือแบบที่ไม่มีตัวแปรเป้าหมาย ซึ่งทางธุรกิจการรู้พฤติกรรมของลูกค้า หรือทำนายได้ว่าลูกค้าซื้ออะไรแล้วจะซื้ออะไรต่อ โดยวิเคราะห์จากข้อมูลในอดีตที่เก็บไว้ ผลการวิเคราะห์สามารถนำมาประกอบการวางแผนเชิงกลยุทธ์เพื่อเพิ่มยอดขาย หรือพัฒนากลยุทธ์การขายได้เป็นอย่างดี เช่น กำหนดตำแหน่งวางสินค้า จัดโปรโมชั่นกลุ่มสินค้า และใช้บริหารสินค้าคงคลัง ในปัจจุบันการหาความสัมพันธ์นั้นจะมีอัลกอริทึมการหาหลายวิธีด้วยกัน แต่อัลกอริทึมที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ อัลกอริทึม Apriori

การหาความสัมพันธ์ด้วยวิธี Apriori ซึ่งจะมีอยู่ 2 ขั้นตอนใหญ่ ๆ คือ การหารายการสินค้าที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำ (frequent itemset) เป็นการหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อย ๆ ในฐานข้อมูลหรือมากกว่าค่าสนับสนุน (support) ขั้นต่ำที่ผู้ใช้กำหนด ในขั้นตอนนี้จะแบ่งได้อีกเป็น 2 ขั้นตอนย่อย คือ การสร้างรูปแบบของรายการสินค้าจะใช้รูปแบบของรายการสินค้าที่มีค่ามากกว่า ค่าสนับสนุนขั้นต่ำ (min support) มาทำการสร้างรูปแบบของรายการสินค้าที่มีขนาดยาวมากขึ้นทีละหนึ่งขั้นไปเรื่อย ๆ และการนับค่า support หลังจากที่สร้างรูปแบบของรายการสินค้าได้แล้วขั้นถัดมาจะทำการคำนวณค่า support ที่เกิดขึ้น โดยที่ค่า support คือจำนวนเปอร์เซ็นต์ที่พบรายการสินค้าในฐานข้อมูล และอีกขั้นตอนคือ การสร้างกฎความสัมพันธ์หลังจากที่หา frequent itemset ได้แล้วจะนำรูปแบบที่หาได้มาสร้างเป็นกฎความสัมพันธ์โดย ในตัวอย่างนี้จะใช้ตัวอย่างข้อมูลที่แสดงในตารางที่ 2.9 ซึ่งประกอบด้วยรายการซื้อสินค้าจำนวน 4 รายการ และมีสินค้าที่ลูกค้าซื้อทั้งหมด 5 ประเภทคือ Apple, Beer, Cereal, Diapers และ Eggs

ตารางที่ 2.9 ข้อมูลการซื้อสินค้าของแต่ละรายการ

Transaction ID	Items
1	Apple, Cereal, Diapers
2	Beer, Cereal, Eggs
3	Apple, Beer, Cereal, Egg
4	Beer, Eggs

2.4.1 การสร้าง frequent itemset

การหาความสัมพันธ์โดยกำหนด min support เท่ากับ 50% โดยที่ขั้นตอนการทำงานมีดังต่อไปนี้ เริ่มจากการคำนวณค่า support ของแต่ละชนิด แสดงในตารางที่ 2.10

ตารางที่ 2.10 การคำนวณค่า support ของสินค้าแต่ละชนิด

Items	Transaction ID				Support
	1	2	3	4	
Apple	1	0	1	0	$2/4 = 50\%$
Beer	0	1	1	1	$3/4 = 75\%$
Cereal	1	1	1	0	$3/4 = 75\%$
Diapers	1	0	0	0	$1/4 = 25\%$
Eggs	0	1	1	1	$3/4 = 75\%$

จากค่า support ที่คำนวณได้ในตารางที่ 2.10 จะเห็นว่า Diapers มีค่า support ต่ำกว่า min support ($25\% < 50\%$) ดังนั้น Diapers จะถูกตัดออกและไม่นำไปพิจารณาสร้างเป็นรายการสินค้าที่มีความยาวเท่ากับ 2 ($k=2$) ต่อไป แสดงดังตารางที่ 2.11 และแสดงเซต {Apple, Beer, Cereal และ Eggs} ของ frequent itemset

ตารางที่ 2.15 จำนวน frequent itemset ทั้งหมด

Frequent itemset	Support	Size
{Apple}	$2/4 = 50\%$	1
{Beer}	$3/4 = 75\%$	1
{Cereal}	$3/4 = 75\%$	1
{Eggs}	$3/4 = 75\%$	1
{Apple, Cereal}	$2/4 = 50\%$	2
{Beer, Cereal}	$2/4 = 50\%$	2
{Beer, Eggs}	$3/4 = 75\%$	2
{Cereal, Eggs}	$2/4 = 50\%$	2
{Beer, Cereal, Eggs}	$2/4 = 50\%$	3

2.4.2 การสร้างกฎความสัมพันธ์

ขั้นตอนถัดมาคือการสร้างกฎความสัมพันธ์จาก frequent itemset ที่หาได้จากขั้นตอนแรก โดยจะพิจารณา frequent itemset ที่ $k \geq 2$ มาสร้างเป็นกฎความสัมพันธ์ เช่น {Apple, Cereal} สามารถสร้างกฎความสัมพันธ์ได้เป็น Apple \rightarrow Cereal เป็นต้น โดยกฎความสัมพันธ์ทั่วไปจะแสดงเป็น LHS \rightarrow RHS โดยที่ LHS (Left Hand Side) แสดงรูปแบบของรายการสินค้าด้านซ้ายของกฎความสัมพันธ์ และ RHS (Right Hand Side) แสดงรูปแบบของรายการสินค้าด้านขวาของกฎความสัมพันธ์ โดยที่การพิจารณาว่ากฎความสัมพันธ์ที่สร้างได้ดีหรือไม่ดี จำเป็นจะต้องมีตัววัดประสิทธิภาพของกฎ โดยที่ confidence และ lift คือค่าที่บ่งบอกว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายและเหตุการณ์ส่วนขวามีความสัมพันธ์มากน้อยแค่ไหน โดย ถ้าค่ามากกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายเป็นการส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าน้อยกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายไม่ได้ส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าเท่ากับ 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายและเหตุการณ์ส่วนขวาไม่มีความสัมพันธ์กัน ซึ่งจะสามารถการเกิดขึ้นคำนวณ confidence ได้ดังรูปที่ 2.6 และรูปที่ 2.7 และสามารถคำนวณ lift ได้ดังรูปที่

2.8 และรูปที่ 2.9 ส่วนตารางที่ 2.10 เป็นตารางที่แสดงกฎความสัมพันธ์ทั้งหมดที่สร้างได้พร้อมทั้งค่า confidence และ lift

$$\text{Confidence (LHS} \rightarrow \text{RHS)} = \frac{\text{support(LHS,RHS)}}{\text{support(LHS)}}$$

รูปที่ 2.6 การหาค่า confidence ของกฎ LHS \rightarrow RHS

$$\begin{aligned} \text{Confidence (Apple} \rightarrow \text{Cereal)} &= \frac{\text{support(Apple,Cereal)}}{\text{support(Apple)}} \\ &= \frac{2/4}{2/4} \times 100 \\ &= 100\% \end{aligned}$$

รูปที่ 2.7 การหาค่า confidence ของกฎ Apple \rightarrow Cereal

$$\text{lift (LHS} \rightarrow \text{RHS)} = \frac{\text{support(LHS,RHS)}}{\text{support(LHS)} \times \text{support(RHS)}}$$

รูปที่ 2.8 การหาค่า lift ของกฎ LHS \rightarrow RHS

$$\begin{aligned} \text{lift (Apple} \rightarrow \text{Cereal)} &= \frac{\text{support(Apple,Cereal)}}{\text{support(Apple)} \times \text{support(Cereal)}} \\ &= \frac{2/4}{2/4 \times 3/4} \\ &= 1.33 \end{aligned}$$

รูปที่ 2.9 การหาค่า lift ของกฎ Apple \rightarrow Cereal

ตารางที่ 2.16 กฎความสัมพันธ์ทั้งหมดที่สร้างได้พร้อมทั้งค่า confidence และ lift

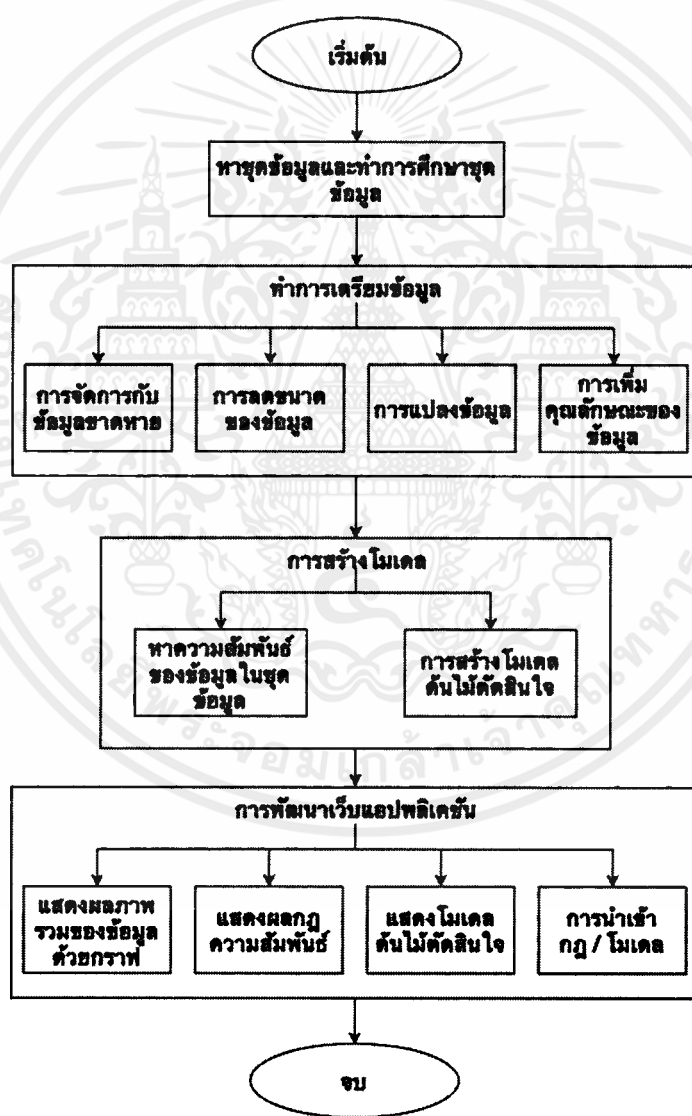
Rule No.	Frequent Itemset	Confidence	Lift
1	Apple → Cereal	100%	1.33
2	Beer → Eggs	100%	1.33
3	Eggs → Beer	100%	1.33
4	Beer, Cereal → Eggs	100%	1.33
5	Cereal, Eggs → Beer	100%	1.33
6	Cereal → Apple	67%	1.33
7	Beer → Cereal, Eggs	67%	1.33
8	Eggs → Beer, Cereal	67%	1.33
9	Beer → Cereal	67%	0.89
10	Cereal → Beer	67%	0.89
11	Cereal → Eggs	67%	0.89
12	Eggs → Cereal	67%	0.89
13	Cereal → Beer, Eggs	67%	0.89
14	Beer, Eggs → Cereal	67%	0.89

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงาน

ในบทนี้จะกล่าวถึงการวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิตเพื่อการส่งเสริมการขายซึ่งแบ่งออกเป็น หาชุดข้อมูลและทำการศึกษาคูหาชุดข้อมูล การเตรียมข้อมูล การสร้างโมเดล และการพัฒนาเว็บแอปพลิเคชัน มีลำดับการดำเนินงานแสดงดังรูปที่ 3.1



รูปที่ 3.1 ลำดับขั้นตอนการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 หาชุดข้อมูลและทำการศึกษาคัดข้อมูล

ชุดข้อมูล Purchase Card (PCard) Fiscal Year 2014 ประกอบด้วยข้อมูลการใช้จ่ายผ่านบัตรเครดิตจำนวน 427,927 รายการ โดยชุดข้อมูลนี้ประกอบด้วยข้อมูลที่เกี่ยวข้องกับการซื้อสินค้าและบริการผ่านบัตรเครดิตของนักศึกษาที่บริหารงานโดยรัฐและสถาบันการศึกษา โดยข้อมูลการซื้อจะได้รับการปรับปรุงเป็นประจำทุกเดือนหลังจากสิ้นเดือน ตัวอย่างเช่น ข้อมูลเดือนกรกฎาคมจะถูกเพิ่มในเดือนสิงหาคมซึ่งข้อมูลนั้นจะประกอบไปด้วยคุณลักษณะต่าง ๆ ต่อไปนี้ ปี-เดือน (Year-Month) หมายเลขหน่วยงาน (Agency Number) ชื่อหน่วยงาน (Agency Name) นามสกุลผู้ถือบัตร (Cardholder Last Name) ชื่อผู้ถือบัตร (Cardholder First Initial) รายละเอียด (Description) ยอดเงิน (Amount) ผู้ขาย (Vendor) วันที่ทำธุรกรรม (Transaction Date) วันที่แสดงข้อมูล (Posted Date) หมวดหมู่สินค้า (Merchant Category Code) โดยรูปที่ 3.2 จะแสดงข้อมูลบางส่วนจากชุดข้อมูล

A	B	C	D	E	F	G	H	I	J	K
Year-Month	Agency Num	Agency Name	Cardholder	Cardholder	Description	Amount	Vendor	Transaction	Posted Date	Merchant C
201307	1000	OKLAHOMA	Mason	C	GENERAL PU	890	NACAS	07/30/2013	07/31/2013	CHARITABU
201307	1000	OKLAHOMA	Mason	C	ROOM CHA	368.96	SHERATON	07/30/2013	07/31/2013	SHERATON
201307	1000	OKLAHOMA	Massey	J	GENERAL PU	165.82	SEARS.COM	07/29/2013	07/31/2013	DIRCT MARI
201307	1000	OKLAHOMA	Massey	T	GENERAL PU	96.39	WAL-MART	07/30/2013	07/31/2013	GROCERY S1
201307	1000	OKLAHOMA	Mauro-Herr	M	HAMMERME	125.96	STAPLES DIR	07/30/2013	07/31/2013	STATIONER
201307	1000	OKLAHOMA	Mauro-Herr	M	GENERAL PU	394.28	KYOCERA D	07/29/2013	07/31/2013	OFFICE, PHK
201307	1000	OKLAHOMA	Mays	B	GENERAL PU	19.87	WAL-MART	07/30/2013	07/31/2013	GROCERY S1
201307	1000	OKLAHOMA	Mays	B	GENERAL PU	10.1	DAYLIGHT D	07/29/2013	07/31/2013	FAST FOOD
201307	1000	OKLAHOMA	McCarthy	D	GENERAL PU	142.6	UFIRST LAU	07/30/2013	07/31/2013	MENS, WOM
201307	1000	OKLAHOMA	McClurg	L	AT&T SERVI	29.99	AT&T DATA	07/30/2013	07/31/2013	TELECOMMI
201307	1000	OKLAHOMA	McGowin	R	GENERAL PU	278.23	RANDY'S FO	07/29/2013	07/31/2013	FREEZER AN
201307	1000	OKLAHOMA	McKee	P	GENERAL PU	127.8	DEARINGER	07/30/2013	07/31/2013	MISCELLANI
201307	1000	OKLAHOMA	McLemore	A	GENERAL PU	572.27	GIDDINGS N	07/30/2013	07/31/2013	COMMERC
201307	1000	OKLAHOMA	McMurry	S	50-LB PLAY S	12.9	LOWES #002	07/30/2013	07/31/2013	HOME SUPP
201307	1000	OKLAHOMA	Meints	K	GENERAL PU	445	OK DEPT OF	07/30/2013	07/31/2013	GOVERNME
201307	1000	OKLAHOMA	Mohiuddin	A	GENERAL PU	452.91	TFS THERMC	07/31/2013	07/31/2013	DENTAL/LAI
201307	1000	OKLAHOMA	Mohiuddin	I	GENERAL PU	2425	WESTERN E	07/29/2013	07/31/2013	HARDWARE
201307	1000	OKLAHOMA	Mohiuddin	I	GENERAL PU	2425	WESTERN E	07/29/2013	07/31/2013	HARDWARE
201307	1000	OKLAHOMA	Mollenhau	R	GENERAL PU	7.63	HOBBY-LOBI	07/29/2013	07/31/2013	HOBBY, TOY
201307	1000	OKLAHOMA	MONTGOM	R	GENERAL PU	18.97	WAL-MART	07/30/2013	07/31/2013	GROCERY S1
201307	1000	OKLAHOMA	Morales	D	ROOM CHA	729.8	WESTIN DER	07/29/2013	07/31/2013	WESTIN HO

รูปที่ 3.2 ข้อมูลบางส่วนจากชุดข้อมูล Purchase Card (PCard) Fiscal Year 2014

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การเตรียมข้อมูล

การเตรียมข้อมูลของชุดข้อมูล Purchase Card (PCard) Fiscal Year 2014 นั้นได้มี การจัดการกับ ข้อมูลขาดหาย การลดขนาดของข้อมูล การแปลงข้อมูล และการเพิ่มคุณลักษณะใหม่ ตามหัวข้อ 3.2.1 ถึง 3.2.4

3.2.1 การจัดการกับข้อมูลขาดหาย

ในชุดข้อมูลการใช้จ่ายผ่านบัตรเครดิตจะมีค่าของคุณลักษณะต่าง ๆ โดยข้อมูลบางส่วนจาก ชุดข้อมูลทั้งหมด ขาดรายละเอียดของข้อมูล ซึ่งข้อมูลที่ขาดหายไปนั้นเราได้ทำการตัดข้อมูลการใช้ จ่ายผ่านบัตรเครดิตนั้นทิ้งไป มีจำนวน 3 รายการ จากข้อมูลทั้งหมด แสดงในรูปที่ 3.3

A	B	C	D	E	F	G	H	I	J	K
Year-Month	Agency Num	Agency Name	Cardholder	Cardholder	Description	Amount	Vendor	Transaction	Posted Date	Merchant Co
201311	18500	CORPORATE	Becker	E	ROOM CHA	351.19	THE ROOSEY	11/08/2013	11/11/2013	
201311	18500	CORPORATE	Becker	E	ROOM CHA	351.19	THE ROOSEY	11/08/2013	11/11/2013	
201310	76000	UNIVERSITY	UNIVERSITY	G	AIR TRAVEL	180.7	UNIVERSITY	10/10/2013	10/14/2013	
201307	1000	OKLAHOMA	Teel	J	GENERAL PU	187.5	SQ SKAR AC	07/25/2013	07/26/2013	ACCOUNTIN
201308	1000	OKLAHOMA	Teel	J	GENERAL PU	150	SQ SKAR AC	08/28/2013	08/29/2013	ACCOUNTIN
201308	1000	OKLAHOMA	Clerk	M	GENERAL PU	204.75	ARTISIGNIN	08/09/2013	08/12/2013	ACCOUNTIN
201308	1000	OKLAHOMA	Toscano	K	GENERAL PU	2798.28	ARTISIGNIN	08/07/2013	08/08/2013	ACCOUNTIN
201309	1000	OKLAHOMA	Payne	S	GENERAL PU	58.5	ARTISIGNIN	09/19/2013	09/16/2013	ACCOUNTIN
201309	1000	OKLAHOMA	MOTEN	D	GENERAL PU	683.97	ARTISIGNIN	09/23/2013	09/24/2013	ACCOUNTIN
201309	1000	OKLAHOMA	MOTEN	D	GENERAL PU	889.96	ARTISIGNIN	09/20/2013	09/23/2013	ACCOUNTIN
201310	1000	OKLAHOMA	Laughlin	J	GENERAL PU	20	SQ THE SAW	10/23/2013	10/24/2013	ACCOUNTIN
201310	1000	OKLAHOMA	Custar	K	GENERAL PU	398.98	ARTISIGNIN	10/16/2013	10/18/2013	ACCOUNTIN
201310	1000	OKLAHOMA	MOTEN	D	GENERAL PU	568.99	ARTISIGNIN	10/09/2013	10/10/2013	ACCOUNTIN
201311	1000	OKLAHOMA	MOTEN	D	GENERAL PU	613.5	ARTISIGNIN	11/13/2013	11/14/2013	ACCOUNTIN
201311	1000	OKLAHOMA	Fancher	L	GENERAL PU	67.25	LABOR LAW	11/06/2013	11/07/2013	ACCOUNTIN
201311	1000	OKLAHOMA	Burch	T	GENERAL PU	90	PAYPAL DJ	11/07/2013	11/08/2013	ACCOUNTIN
201311	1000	OKLAHOMA	Vaughan	W	GENERAL PU	559.8	ARTISIGNIN	11/11/2013	11/12/2013	ACCOUNTIN
201311	1000	OKLAHOMA	Shelton	C	GENERAL PU	96	PAYPAL BO	11/26/2013	11/27/2013	ACCOUNTIN
201401	1000	OKLAHOMA	Hood	K	GENERAL PU	5	PAYPAL MA	01/06/2014	01/07/2014	ACCOUNTIN
201312	9000	OFFICE OF A	McCaulle	J	GENERAL PU	5000	PAYPAL SM	12/02/2013	12/03/2013	ACCOUNTIN
201309	31000	STATE FIRE	Cain	S	GENERAL PU	67.25	LABOR LAW	09/24/2013	09/25/2013	ACCOUNTIN
201310	35000	HISTORICAL	CAFFEY	M	GENERAL PU	285	SQ P & L FIR	10/28/2013	10/29/2013	ACCOUNTIN
201311	43000	DEPARTMENT	HAWKINS	K	GENERAL PU	183.75	LABOR LAW	11/01/2013	11/04/2013	ACCOUNTIN
201307	47700	OKLA. BURE	Gomez	K	GENERAL PU	483.3	FIVE OAKS F	06/28/2013	07/01/2013	ACCOUNTIN
201308	47700	OKLA. BURE	Gomez	K	GENERAL PU	896	FIVE OAKS F	08/05/2013	08/07/2013	ACCOUNTIN
201308	47700	OKLA. BURE	Gomez	K	GENERAL PU	1673.09	FIVE OAKS F	08/16/2013	08/19/2013	ACCOUNTIN
201308	53000	OKLA. PANH	Fesmire	A	ACCOUNTIN	376.93	G FARNEY &	08/17/2013	08/19/2013	ACCOUNTIN

รูปที่ 3.3 ตัวอย่างข้อมูลในส่วนที่เป็นข้อมูลขาดหาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 การลดขนาดของข้อมูล

ในชุดข้อมูลการใช้จ่ายผ่านบัตรเครดิต ในส่วนของข้อมูลยอดเงินที่มีค่าติดลบจำนวน 14,531 รายการจากข้อมูลทั้งหมด โดยข้อมูลยอดเงินที่มีค่าติดลบเป็นข้อมูลที่เป็นการค้าสินค้า และไม่ได้นำมาใช้ในการวิเคราะห์จึงทำการลบแถวหรือรายการ (Transaction) โดยคิดเป็น 3.28 เปอร์เซ็นต์ของข้อมูลทั้งหมด รูปที่ 3.4 แสดงตัวอย่างของข้อมูลในคอลัมน์ยอดเงิน (Amount) ที่มีข้อมูลมีค่าติดลบ

Year-Month	Agency Num	Agency Name	Cardholder	Cardholder	Description	Amount	Vendor	Transaction	Posted Date	Merchant C
201403	2500	OKLAHOMA	MCCALL	L	GENERAL P	-42963.04	SUNSHINE II	02/26/2014	03/04/2014	STATIONERY
201403	76000	UNIVERSITY	MECHAM	S	GENERAL P	-41740	HO RAD NC	03/20/2014	03/21/2014	DENTAL/LA
201307	80000	OKLA. CARE	GOODNER	M	GENERAL P	-38506.87	ORACL OPN	07/06/2013	07/08/2013	COMPUTER
201402	34500	DEPARTMEN	Lewis	G	GENERAL P	-34108	P INTEGRA	02/14/2014	02/17/2014	GENERAL CC
201309	76000	UNIVERSITY	LARSON	B	GENERAL P	-33075.32	RINICK MA	09/09/2013	09/12/2013	LANDSCAPE
201309	34500	DEPARTMEN	Lewis	G	GENERAL P	-30076.45	OBERTS TR	09/27/2013	09/30/2013	CAR AND TR
201401	76000	UNIVERSITY	LARSON	B	GENERAL P	-27864	HERTZ EQUI	01/07/2014	01/09/2014	EQUIP, TO
201406	9000	OFFICE OF N	Newsom	L	GENERAL P	-21000	CONSTRUCT	06/19/2014	06/23/2014	GOVERNME
201307	13100	DEPARTMEN	Bowlin	C	GENERAL P	-20000	EDWOOD T	06/27/2013	07/01/2013	MEDICAL AN
201308	47700	OKLA. BURE	Gomez	K	GENERAL P	-18899	P NETWORK	08/01/2013	08/05/2013	ELECTRONIC
201403	9000	OFFICE OF N	Newsom	L	GENERAL P	-18128.06	CONSTRUCT	03/20/2014	03/24/2014	GOVERNME
201402	40000	OFFICE OF J	Wilborn	D	GENERAL P	-12356.89	CAL UNIFOR	01/31/2014	02/09/2014	CATALOG M
201404	13100	DEPARTMEN	Saltzman	D	CAMERA P	-10140	VW GRAIN	04/07/2014	04/14/2014	INDUSTRIAL
201401	76000	UNIVERSITY	LARSON	B	GENERAL P	-9853.4	MORRISON	12/31/2013	01/09/2014	PLUMBING /
201309	98000	GRAND RIVE	Cardholder	e	GENERAL P	-9000	MOBLE ENG	09/10/2013	09/12/2013	COMMERC
201311	94500	DEPARTMEN	Gadberry	P	GENERAL P	-7860.22	LOGAN COL	11/07/2013	11/11/2013	GENERAL CC
201403	76000	UNIVERSITY	MECHAM	S	GENERAL P	-7880	HO RAD NC	03/20/2014	03/21/2014	DENTAL/LA
201310	77000	UNIV. OF O	STRAWN	J	GENERAL P	-7188.61	GENVILLES	10/25/2013	10/28/2013	DENTAL/LA
201404	63500	COMM. ON	Killingswor	S	GENERAL P	-6938.22	DEPT OF CO	03/31/2014	04/02/2014	GOVERNME
201404	1000	OKLAHOMA	Ropers	A	GENERAL P	-6266.53	FREEMAN A	04/13/2014	04/14/2014	SPECIAL TRA
201401	34500	DEPARTMEN	Wise	V	GENERAL P	-5222.1	ENNAMET	01/07/2014	01/08/2014	INDUSTRIAL
201307	56600	DEPARTMEN	Everett	T	GENERAL P	-5134.83	EN E. KEIT	07/09/2013	07/04/2013	FREEZER AN
201308	76000	UNIVERSITY	ZACCARELLI	C	GENERAL P	-5000	TOTAL COR	08/01/2013	08/05/2013	OFFICE, PHC
201311	26900	OK. COMM.	Bennett	J	ROOM CHA	-5000	SHERATON	11/14/2013	11/18/2013	SHERATON
201406	1000	OKLAHOMA	Mulder	P	GENERAL P	-4999	OFFIN SYST	05/30/2014	06/04/2014	INDUSTRIAL
201308	1000	OKLAHOMA	Davis	C	GENERAL P	-4999	YAMAHA OF	08/02/2013	08/05/2013	MOTORCYCL
201308	1000	OKLAHOMA	Elmore	D	GENERAL P	-4999	YAMAHA OF	08/02/2013	08/05/2013	MOTORCYCL
201406	30800	STATE BURE	McFarland	N	GENERAL P	-4987.04	WHEATON I	06/17/2014	06/19/2014	INDUSTRIAL

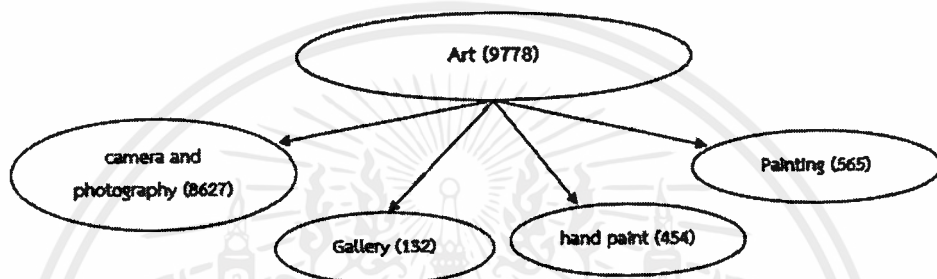
รูปที่ 3.4 ตัวอย่างข้อมูลในส่วนที่ทำการลดขนาดของข้อมูล

3.2.3 การแปลงข้อมูล

ในชุดข้อมูลการใช้จ่ายผ่านบัตรเครดิต ส่วนที่เป็นคุณลักษณะ Merchant Category Code มีข้อมูลที่ยังไม่สามารถนำมาใช้เป็นหมวดหมู่สินค้าได้จริง เช่น หมวดหมู่ย่อยสายการบิน (Airline) เกิดจากการรวมคุณลักษณะ Merchant Category Code เช่น air berlin air canada air china air

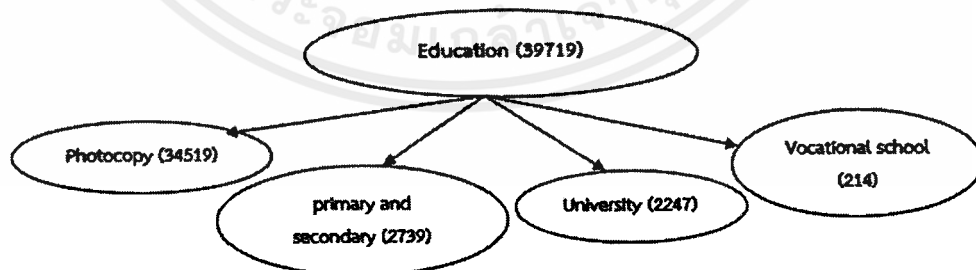
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

france air india และ air new zealand ltd. International เป็นต้น เข้าด้วยกันเพื่อจัดหมวดหมู่สินค้าให้เป็นหมวดหมู่เดียวกัน จึงต้องทำการแปลงข้อมูลในคุณลักษณะ Merchant Category Code โดยเริ่มจากที่นำข้อมูลที่มีมาทำการแยกออกเป็นหมวดหมู่ใหญ่ต่าง ๆ 12 หมวดหมู่ดังนี้ ศิลปะ (Art) การศึกษา (Education) บันเทิง (Entertainment) แฟชั่น (Fashion) การแพทย์ (Medical) ร้านอาหาร (Restaurant) บริการ (Service) ร้านค้า (Store) เทคโนโลยี (Technology) การขนส่ง (Transportation) การท่องเที่ยว (Travel) และการบำรุงรักษารถยนต์ (Car maintenance) แสดงในรูปที่ 3.5 ถึง รูปที่ 3.16



รูปที่ 3.5 หมวดหมู่ศิลปะ

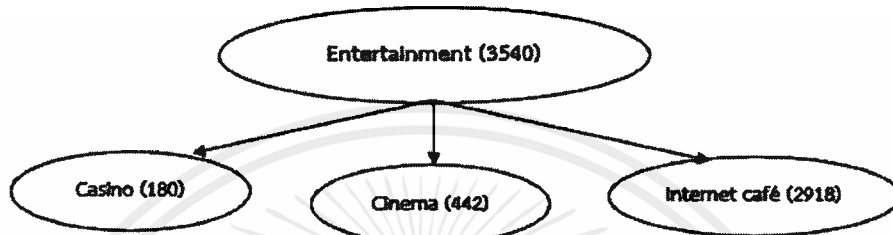
หมวดหมู่ศิลปะ ประกอบด้วยหมวดหมู่ย่อย 4 หมวด คือ กล้องและการถ่ายภาพ (camera and photography) จำนวน 8,627 รายการ แกลเลอรี (gallery) จำนวน 132 รายการ งานฝีมือ (hand paint) จำนวน 454 รายการ และจิตรกรรม (painting) จำนวน 565 รายการ



รูปที่ 3.6 หมวดหมู่การศึกษา

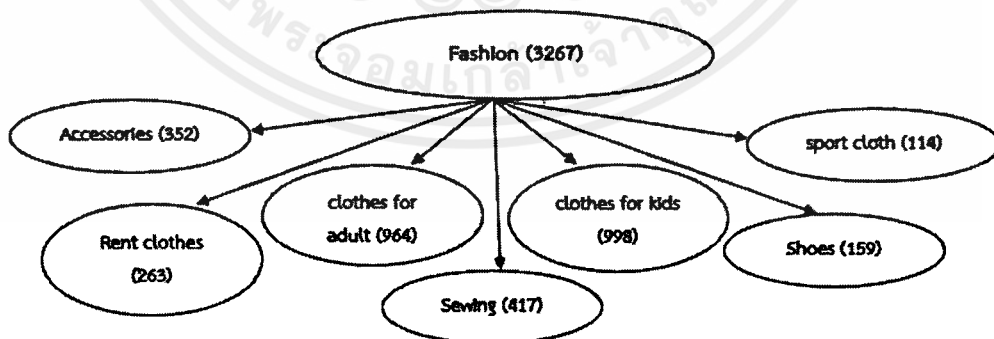
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่การศึกษา ประกอบด้วยหมวดหมู่ย่อย 4 หมวด คือ ถ่ายเอกสาร (photocopy) จำนวน 34,519 รายการ โรงเรียนประถมศึกษาและมัธยมศึกษา (primary and secondary schools) จำนวน 2,739 รายการ มหาวิทยาลัย (university) จำนวน 2,247 รายการ และอาชีวศึกษา (vocational school) จำนวน 214 รายการ



รูปที่ 3.7 หมวดหมู่อบันเทิง

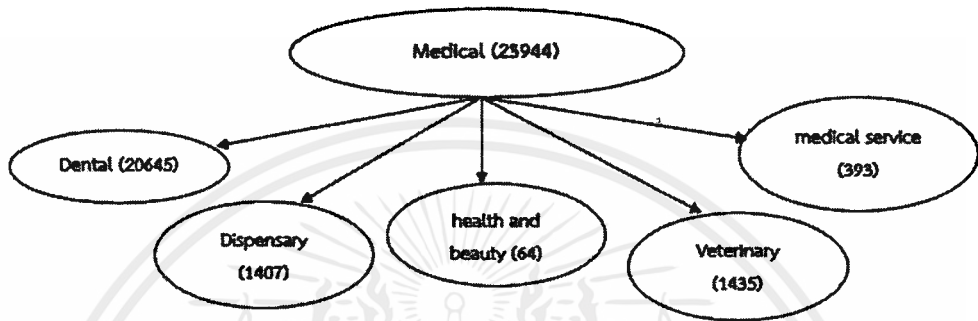
หมวดหมู่อบันเทิง ประกอบด้วยหมวดหมู่ย่อย 3 หมวด คือ คาสิโน (casino) จำนวน 180 รายการ โรงภาพยนตร์ (cinema) จำนวน 442 รายการ และอินเทอร์เน็ตคาเฟ่ (internet café) จำนวน 2,918 รายการ



รูปที่ 3.8 หมวดหมู่แฟชั่น

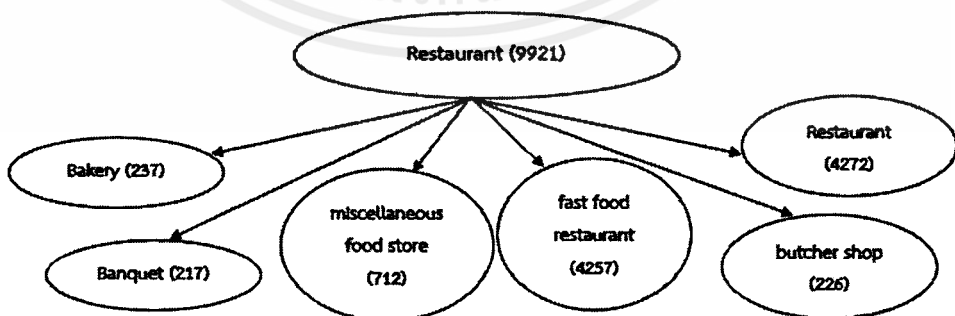
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่แฟชั่น ประกอบด้วยหมวดหมู่ย่อย 7 หมวด คือ เครื่องประดับ (accessories) จำนวน 352 รายการ เสื้อผ้าสำหรับผู้ใหญ่ (clothes for adult) จำนวน 964 รายการ เสื้อผ้าสำหรับเด็ก (clothes for kids) จำนวน 998 รายการ ค่าเช่าเสื้อผ้า (rent clothes) จำนวน 263 รายการ การเย็บผ้า (sewing) จำนวน 417 รายการ รองเท้า (shoes) จำนวน 159 รายการ และเสื้อผ้ากีฬา (sport cloth) จำนวน 114 รายการ



รูปที่ 3.9 หมวดหมู่การแพทย์

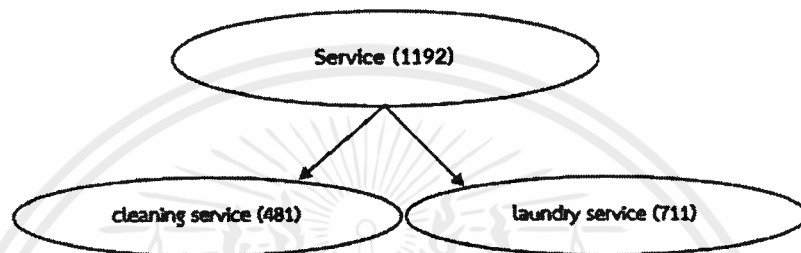
หมวดหมู่การแพทย์ ประกอบด้วยหมวดหมู่ย่อย 5 หมวด คือ ทันตกรรม (dental) จำนวน 20,645 รายการ ร้านขายยา (dispensary) จำนวน 1,407 รายการ สุขภาพและความงาม (health and beauty) จำนวน 64 รายการ บริการทางการแพทย์ (medical service) จำนวน 393 รายการ และคำสัตวแพทย์ (veterinary) จำนวน 1,435 รายการ



รูปที่ 3.10 หมวดหมู่ร้านอาหาร

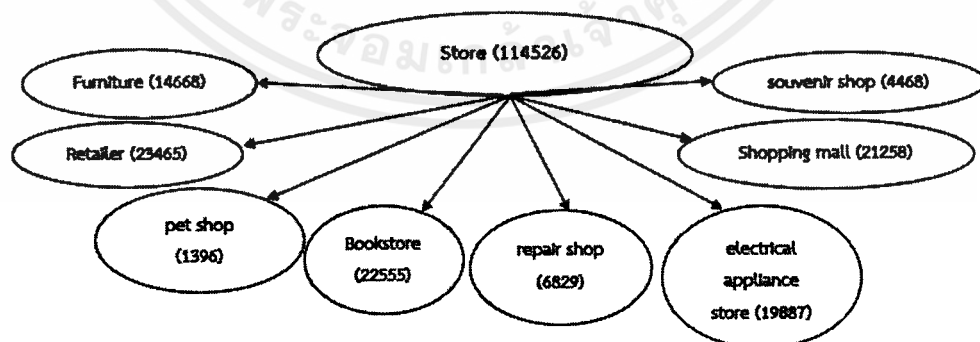
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่ร้านอาหาร ประกอบด้วยหมวดหมู่ย่อย 6 หมวด คือ ร้านเบเกอรี่ (bakery) จำนวน 237 รายการ ค่างานเลี้ยง (banquet) จำนวน 217 รายการ ร้านขายเนื้อ (butcher shop) จำนวน 226 รายการ ร้านอาหารจานด่วน (fast food restaurant) จำนวน 4,257 รายการ ร้านอาหารทั่วไป (miscellaneous food store) จำนวน 712 รายการ และร้านอาหาร (restaurant) จำนวน 4,272 รายการ



รูปที่ 3.11 หมวดหมู่บริการ

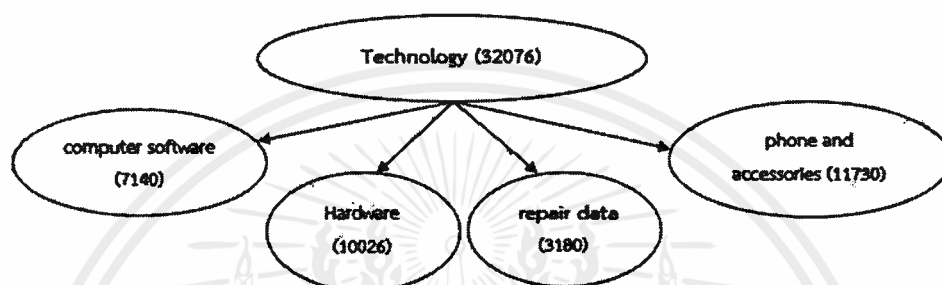
หมวดหมู่บริการ ประกอบด้วยหมวดหมู่ย่อย 2 หมวด คือ บริการทำความสะอาด (cleaning service) จำนวน 481 รายการ และบริการซักกรีด (laundry service) จำนวน 711 รายการ



รูปที่ 3.12 หมวดหมู่ร้านค้า

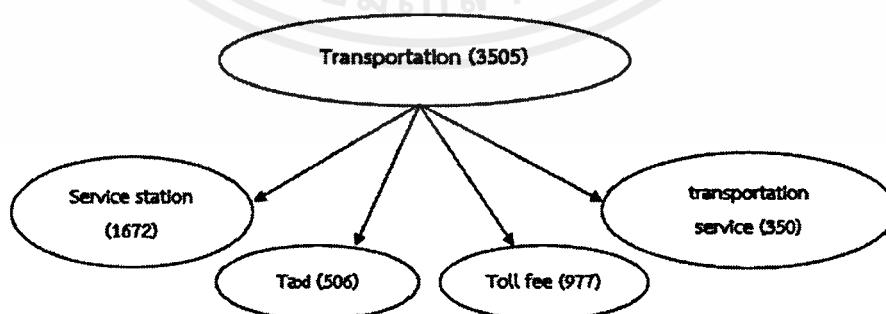
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่ร้านค้า ประกอบด้วยหมวดหมู่ย่อย 8 หมวด คือ ร้านหนังสือ (bookstore) จำนวน 22,555 รายการ ร้านขายเครื่องใช้ไฟฟ้า (electrical appliance store) จำนวน 19,887 รายการ เฟอร์นิเจอร์ (furniture) จำนวน 14,668 รายการ ร้านขายสัตว์เลี้ยง (pet shop) จำนวน 1,396 รายการ ร้านซ่อม (repair shop) จำนวน 6,829 รายการ ร้านค้าปลีก (retailer) จำนวน 23,465 รายการ ห้างสรรพสินค้า (shopping mall) จำนวน 21,258 รายการ และร้านขายของที่ระลึก (souvenir shop) จำนวน 4,468 รายการ



รูปที่ 3.13 หมวดหมู่เทคโนโลยี

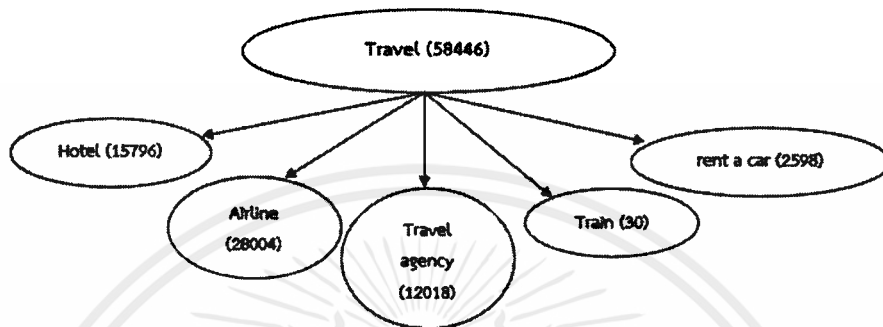
หมวดหมู่เทคโนโลยี ประกอบด้วยหมวดหมู่ย่อย 4 หมวด คือ ซอฟต์แวร์คอมพิวเตอร์ (computer software) จำนวน 7,140 รายการ ฮาร์ดแวร์ (hardware) จำนวน 10,026 รายการ การซ่อมแซมข้อมูล (repair data) จำนวน 3,180 รายการ และโทรศัพท์และอุปกรณ์เสริม (phone and accessories) จำนวน 11,730 รายการ



รูปที่ 3.14 หมวดหมู่การขนส่ง

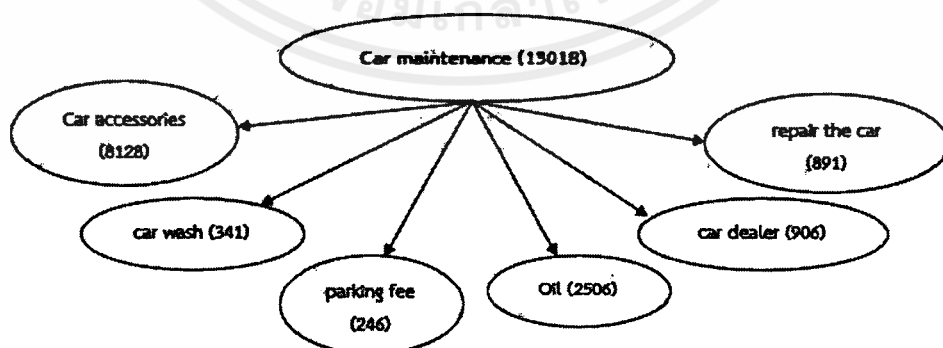
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่การขนส่ง ประกอบด้วยหมวดหมู่ย่อย 4 หมวด คือ สถานีบริการ (service station) จำนวน 1,672 รายการ รถแท็กซี่ (taxi) จำนวน 506 รายการ ค่าทางด่วน (toll fee) จำนวน 977 รายการ และบริการขนส่ง (transportation service) จำนวน 350 รายการ



รูปที่ 3.15 หมวดหมู่การท่องเที่ยว

หมวดหมู่การท่องเที่ยว ประกอบด้วยหมวดหมู่ย่อย 5 หมวด คือ สายการบิน (airline) จำนวน 28,004 รายการ โรงแรม (hotel) จำนวน 15,796 รายการ ค่าเช่ารถ (rent a car) จำนวน 2,598 รายการ รถไฟ (train) จำนวน 30 รายการ และตัวแทนการท่องเที่ยว (travel agency) จำนวน 12,018 รายการ



รูปที่ 3.16 หมวดหมู่การบำรุงรักษารถยนต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมวดหมู่การบำรุงรักษารถยนต์ ประกอบด้วยหมวดหมู่ย่อย 6 หมวด คือ อุปกรณ์ตกแต่งรถยนต์ (car accessories) จำนวน 8,128 รายการ ตัวแทนจำหน่ายรถยนต์ (car dealer) จำนวน 906 รายการ ล้างรถ (car wash) จำนวน 341 รายการ น้ำมัน (oil) จำนวน 2,506 รายการ ค่าจอดรถ (parking fee) จำนวน 246 รายการ และซ่อมรถ (repair the car) จำนวน 891 รายการ

3.2.4 การเพิ่มคุณลักษณะใหม่จากข้อมูล

การเพิ่มคุณลักษณะใหม่จากคุณลักษณะเดิมที่มีอยู่ โดยมีคุณลักษณะใหม่ ดังต่อไปนี้ คุณลักษณะวันที่ (Date) คุณลักษณะเดือน (Month) คุณลักษณะปี (Year) คุณลักษณะวัน (Day) คุณลักษณะชื่อเดือน (MonthName) คุณลักษณะฤดู (Season) คุณลักษณะช่วงของเดือน (SplitMonth) คุณลักษณะประเภทของวัน (DayType) แสดงดังรูปที่ 3.17

การเพิ่มคุณลักษณะใหม่จากคุณลักษณะเดิมที่มีอยู่เพื่อเพิ่มมิติของชุดข้อมูล เพื่อนำไปใช้ในการหาความสัมพันธ์ และนำไปใช้ในการสร้างโมเดลต้นไม้ตัดสินใจ

Date	Month	Year	Day	MonthName	Season	SplitMonth	DayType
10	4	2014	Thursday	April	Spring	Earlymonth	Weekdays
13	5	2014	Tuesday	May	Spring	Midmonth	Weekdays
21	10	2013	Monday	October	Fall	Latemonth	Weekdays
21	10	2013	Monday	October	Fall	Latemonth	Weekdays
15	11	2013	Friday	November	Fall	Midmonth	Weekdays
13	12	2013	Friday	December	Winter	Midmonth	Weekdays
24	2	2014	Monday	February	Winter	Latemonth	Weekdays
21	5	2014	Wednesday	May	Spring	Latemonth	Weekdays
2	1	2014	Thursday	January	Winter	Earlymonth	Weekdays
25	2	2014	Tuesday	February	Winter	Latemonth	Weekdays
21	10	2013	Monday	October	Fall	Latemonth	Weekdays
28	1	2014	Tuesday	January	Winter	Latemonth	Weekdays
3	9	2013	Tuesday	September	Fall	Earlymonth	Weekdays
21	2	2014	Friday	February	Winter	Latemonth	Weekdays
18	3	2014	Tuesday	March	Spring	Midmonth	Weekdays
14	2	2014	Friday	February	Winter	Midmonth	Weekdays
23	6	2014	Monday	June	Summer	Latemonth	Weekdays
1	11	2013	Friday	November	Fall	Earlymonth	Weekdays
30	10	2013	Wednesday	October	Fall	Latemonth	Weekdays
4	11	2013	Monday	November	Fall	Earlymonth	Weekdays
10	6	2014	Tuesday	June	Summer	Earlymonth	Weekdays
14	3	2014	Friday	March	Spring	Midmonth	Weekdays
9	7	2013	Tuesday	July	Summer	Earlymonth	Weekdays
10	2	2014	Monday	February	Winter	Earlymonth	Weekdays
30	6	2013	Sunday	June	Summer	Latemonth	Weekends
17	1	2014	Friday	January	Winter	Midmonth	Weekdays
29	5	2014	Thursday	May	Spring	Latemonth	Weekdays
25	6	2014	Wednesday	June	Summer	Latemonth	Weekdays

รูปที่ 3.17 ตัวอย่างข้อมูลในส่วนที่เป็นคุณลักษณะใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะวันที่ (Date) คุณลักษณะเดือน (Month) และคุณลักษณะปี (Year) ได้มาจากคุณลักษณะวันที่ทำธุรกรรมซึ่งทำการตัดค่าของคุณลักษณะในส่วนที่เป็นวันที่แยกออกเป็นคุณลักษณะวันที่ ส่วนการตัดค่าของคุณลักษณะในส่วนที่เป็นเดือนแยกออกเป็นคุณลักษณะเดือน และส่วนการตัดค่าของคุณลักษณะในส่วนที่เป็นปีแยกออกเป็นคุณลักษณะปี ยกตัวอย่างเช่น คุณลักษณะวันที่ทำธุรกรรม 25/6/2014 โดยที่ 25 จะแยกออกเป็นคุณลักษณะวันที่ โดยที่ 6 จะแยกออกเป็นคุณลักษณะเดือน และ โดยที่ 2014 จะแยกออกเป็นคุณลักษณะปี

คุณลักษณะวัน (Day) นั้นได้มาจากการที่นำคุณลักษณะวันที่ คุณลักษณะเดือน และคุณลักษณะปี ทั้ง 3 คุณลักษณะมาเทียบกับอาเรียที่เก็บค่าวันของวันที่-เดือน-ปีต่าง ๆ ว่าวันที่-เดือน-ปีนี้มีค่าวันเป็นอะไรแล้วนำมาใส่คุณลักษณะวัน

คุณลักษณะชื่อเดือนได้มาจากการนำคุณลักษณะเดือน มาเปลี่ยนเป็นชื่อเดือน เดือนที่ 1 เป็นเดือนมกราคม (January) เดือนที่ 2 เดือนกุมภาพันธ์ (February) เดือนที่ 3 เดือนมีนาคม (March) เดือนที่ 4 เดือนเมษายน (April) เดือนที่ 5 เดือนพฤษภาคม (May) เดือนที่ 6 เดือนมิถุนายน (June) เดือนที่ 7 เดือนกรกฎาคม (July) เดือนที่ 8 เดือนสิงหาคม (August) เดือนที่ 9 เดือนกันยายน (September) เดือนที่ 10 เดือนตุลาคม (October) เดือนที่ 11 เดือนพฤศจิกายน (November) เดือนที่ 12 เดือนธันวาคม (December)

คุณลักษณะฤดูนั้นได้มาจากการที่นำค่าคุณลักษณะเดือนมาเข้าเงื่อนไขโดยที่ ฤดูร้อน (Summer) จะอยู่ในช่วงเดือนมิถุนายน-สิงหาคม ฤดูใบไม้ร่วง (Fall) จะอยู่ในช่วงเดือนกันยายน-พฤศจิกายน ฤดูหนาว (Winter) จะอยู่ในช่วงเดือนธันวาคม-กุมภาพันธ์ ฤดูใบไม้ผลิ (Spring) จะอยู่ในช่วงเดือนมีนาคม-พฤษภาคม

คุณลักษณะช่วงของเดือนนั้นได้มาจากการนำค่าของคุณลักษณะวันที่มาเข้าเงื่อนไขโดยที่ ต้นเดือน (Earlymonth) จะอยู่ในช่วงวันที่ 1-10 กลางเดือน (Midmonth) จะอยู่ในช่วงวันที่ 11-20 ปลายเดือน (Latemonth) จะอยู่ในช่วงวันที่ 21-31

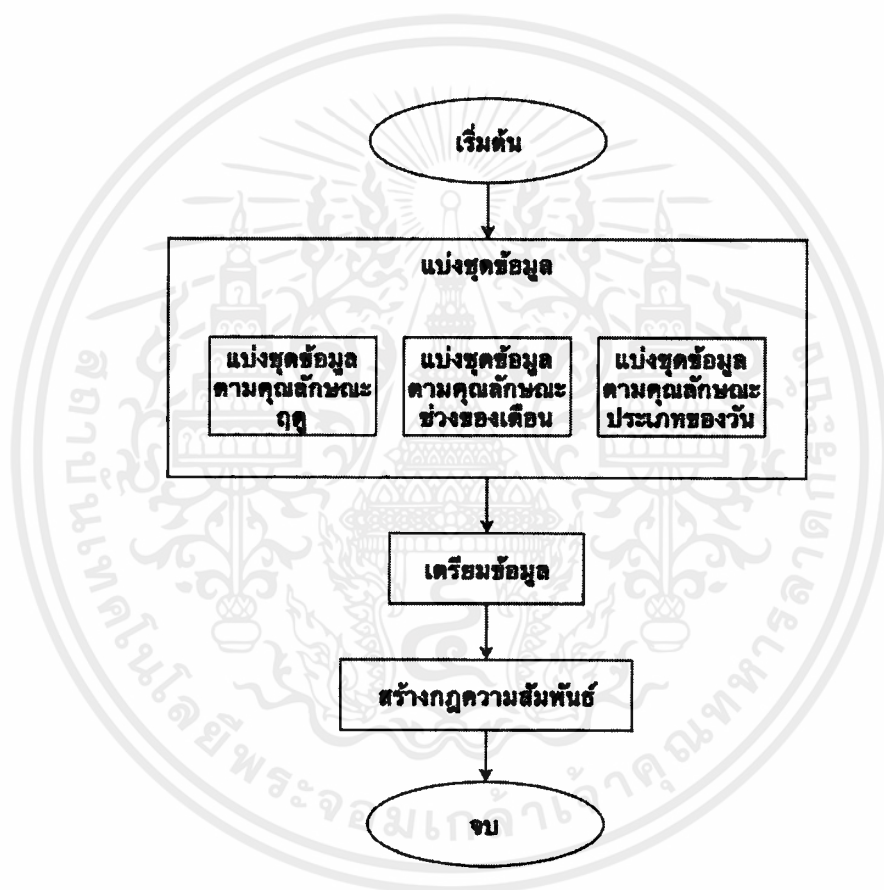
คุณลักษณะประเภทของวัน ได้ทำประเภทจัดของวันออกเป็น 3 ประเภทได้แก่ วันหยุดเทศกาล (Holidays) วันหยุดสุดสัปดาห์ (Weekends) วันธรรมดา (Weekdays) โดยวันหยุดเทศกาลได้แก่ วันคริสต์มาส วันที่ 25 ธันวาคม วันคริสต์มาสอีฟ วันที่ 24 ธันวาคม วันขอบคุณพระเจ้า วันที่ 22 พฤศจิกายน วันชาติอเมริกา วันที่ 4 กรกฎาคม และวันหยุดสุดสัปดาห์ได้แก่ วันเสาร์ วันอาทิตย์ และวันธรรมดาคือวันที่ไม่ใช่วันหยุดสุดสัปดาห์และวันหยุดเทศกาล

3.3 การสร้างโมเดล

ในส่วนของการสร้างโมเดลนั้นแบ่งออกเป็น การหาทฤษฎีความสัมพันธ์ และการสร้างโมเดลต้นไม้ตัดสินใจ ดังหัวข้อที่ 3.3.1 และ 3.3.2 ตามลำดับ

3.3.1 การหาทฤษฎีความสัมพันธ์ของข้อมูล

ลำดับขั้นตอนในการเตรียมข้อมูลของการหาทฤษฎีความสัมพันธ์ แสดงดังรูป 3.18



รูปที่ 3.18 ลำดับขั้นตอนการเตรียมข้อมูลสำหรับการสร้างทฤษฎีความสัมพันธ์

ในการหาทฤษฎีความสัมพันธ์ของข้อมูล เริ่มจากการแบ่งชุดข้อมูลเพื่อการหาทฤษฎีความสัมพันธ์ของข้อมูล แบ่งชุดข้อมูลทั้งหมดแบ่งเป็นชุดข้อมูลที่แบ่งตามคุณลักษณะฤดู 4 ชุด ได้แก่ ฤดูร้อน ฤดูใบไม้ร่วง ฤดูหนาว และฤดูใบไม้ผลิ ชุดข้อมูลที่แบ่งตามคุณลักษณะช่วงของเดือน 3 ชุด ได้แก่ ต้น

เดือน กลางเดือน และปลายเดือน ชุดข้อมูลที่แบ่งตามคุณลักษณะประเภทของวัน 3 ชุด ได้แก่ วันธรรมดา วันหยุดสุดสัปดาห์ และหยุดเทศกาล รวมทั้งหมด 10 ชุด แสดงดังรูปที่ 3.19

Agency #	Agency #	Cardhold	Cardhold	Description	Amount	Amount	Category	Category	Date	Month	Year	Day	Month	Season	SpInMon	DayType
1000	OKLAHOM	Hiadik	J	Ultimate	8	Very Low	art	gallery	10	4	2014	Thursday	April	Spring	Earlymor	Weekday
56400	DEPARTM	Bemo	R	Fried chi	8	Very Low	art	gallery	13	5	2014	Tuesday	May	Spring	Midmon	Weekday
1000	OKLAHOM	Hiadik	J	Apple a c	8	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Weight s	13	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Bowl of f	13	Very Low	art	gallery	13	11	2013	Friday	November	Fall	Midmon	Weekday
1000	OKLAHOM	Hiadik	J	Man smd	13	Very Low	art	gallery	13	12	2013	Friday	December	Winter	Midmon	Weekday
1000	OKLAHOM	Hiadik	J	Woman	14	Very Low	art	gallery	24	2	2014	Monday	February	Winter	Latemor	Weekday
76000	UNIVERS	GARDNEI	J	GENERAL	14.41	Very Low	art	gallery	23	5	2014	Wednesday	May	Spring	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Upset Of	15	Very Low	art	gallery	2	1	2014	Thursday	January	Winter	Earlymor	Weekday
76000	UNIVERS	MCKINHT	J	GENERAL	17	Very Low	art	gallery	25	2	2014	Tuesday	February	Winter	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Lifeguard	19	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemor	Weekday
56400	DEPARTM	Bemo	R	red light	19	Very Low	art	gallery	28	1	2014	Tuesday	January	Winter	Latemor	Weekday
60500	REGENTS	Ferguson	K	Teacher	19	Very Low	art	gallery	9	9	2013	Tuesday	September	Fall	Earlymor	Weekday
56400	DEPARTM	Bemo	R	Three int	19	Very Low	art	gallery	21	2	2014	Friday	February	Winter	Latemor	Weekday
56400	DEPARTM	Bemo	R	Grass NW	19	Very Low	art	gallery	18	3	2014	Tuesday	March	Spring	Midmon	Weekday
76000	UNIVERS	ASHLEY	J	Prepaid	19.99	Very Low	art	gallery	14	2	2014	Friday	February	Winter	Midmon	Weekday
1000	OKLAHOM	Cunning	A	Prepaid	19.99	Very Low	art	gallery	23	6	2014	Monday	June	Summer	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Healthy	21	Very Low	art	gallery	1	11	2013	Friday	November	Fall	Earlymor	Weekday
56400	DEPARTM	Bemo	R	Autumn	21	Very Low	art	gallery	30	10	2013	Wednesday	October	Fall	Latemor	Weekday
56400	DEPARTM	Bemo	R	Cure girl	21	Very Low	art	gallery	4	11	2013	Monday	November	Fall	Earlymor	Weekday
76000	UNIVERS	CHAPPEL	J	GENERAL	21.37	Very Low	art	gallery	10	6	2014	Tuesday	June	Summer	Earlymor	Weekday
28900	OK COM	Bennett	J	School st	27	Very Low	art	gallery	14	3	2014	Friday	March	Spring	Midmon	Weekday
1000	OKLAHOM	Yang	M	GENERAL	27.99	Very Low	art	gallery	9	7	2013	Tuesday	July	Summer	Earlymor	Weekday
1000	OKLAHOM	Hiadik	J	Saving N	28	Very Low	art	gallery	10	2	2014	Monday	February	Winter	Earlymor	Weekday
1000	OKLAHOM	Williams	J	GENERAL	30	Very Low	art	gallery	20	6	2013	Sunday	June	Summer	Latemor	Weekend
56400	DEPARTM	Bemo	R	Cheerful	30	Very Low	art	gallery	17	1	2014	Friday	January	Winter	Midmon	Weekday
1000	OKLAHOM	Hiadik	J	Beautiful	30	Very Low	art	gallery	29	5	2014	Thursday	May	Spring	Latemor	Weekday
1000	OKLAHOM	Hiadik	J	Losing W	30	Very Low	art	gallery	25	6	2014	Wednesday	June	Summer	Latemor	Weekday

รูปที่ 3.19 ตัวอย่างข้อมูลก่อนเตรียมข้อมูลเพื่อหาความสัมพันธ์

หลังจากการแบ่งชุดข้อมูลแล้ว ต่อไปเป็นการเตรียมข้อมูลโดยการเตรียมข้อมูลแบ่งออกเป็น การนำข้อมูลในธุรกรรมแต่ละรายการของผู้ถือบัตรเดียวกันในส่วนคุณลักษณะหมวดหมู่มารวมกัน โดยที่แยกจากการเช็คที่ชื่อผู้ถือบัตร (Cardholder First Initial) นามสกุลผู้ถือบัตร (Cardholder Last Name) และหน่วยงานของผู้ถือบัตรตรงกัน (Agency Number) แสดงดังรูปที่ 3.20

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1000	Adams	A	shopping	furniture									
2	1000	Adkins	A	photocopy	bookstore	shopping	retailer							
3	77000	ADMIRE	A	bookstore	shopping	furniture	photocopy							
4	1000	Akem	A	electrical	retailer	electrical	hardware	car accessories						
5	1000	Alexander	A	bookstore										
6	1000	Allen	A	bookstore	shopping mall									
7	34500	Aranda	A	furniture	electrical	service st	car accessories							
8	1000	Arena	A	internet c	medical st	bookstore	electrical	retailer	computer hardware	phone an	photocopy			
9	1000	Barclay	A	bookstore	souvenir	shopping	furniture	retailer	photocopy					
10	1000	Barnes	A	retailer	camera an	photocopy								
11	78000	BARNES	A	internet c	shopping	phone an	photocopy							
12	1000	Barredo	A	travel ag	travel agency									
13	53000	Batenhorst	A	hand pain	primary st	clothes fo	clothes fo	bookstore	retailer	shopping	retailer	computer hardware	photocopy	
14	30000	Beats	A	hotel										
15	1000	Begosh	A	shopping	retailer	car accessories								
16	4000	Bender	A	primary st	restaurant	miscellani	bookstore	shopping	retailer	service st	hotel	photocopy		
17	1000	Benn	A	shopping	electrical	hardware	hotel	photocopy						
18	1000	Bhat	A	bookstore	bookstore									
19	1000	Bigl	A	electrical	hardware									
20	63000	Bilby	A	furniture	camera an	phone an	photocopy							

รูปที่ 3.20 ตัวอย่างข้อมูลหลังการรวมคุณลักษณะหมวดหมู่ที่มีผู้ถือบัตรเดียวกัน

หลังจากที่เปลี่ยนข้อมูลเป็นข้อมูลธุรกรรมที่มีผู้ถือบัตรเดียวกันรวมหมวดหมู่กันแล้ว ต่อไปเป็นการเปลี่ยนชนิดของข้อมูลคุณลักษณะหมวดหมู่ที่เป็นชนิดข้อมูลนอมินอล (Nominal) ให้เป็นชนิดข้อมูลไบนารี (Binary) โดยจะแสดงข้อมูลก่อนเตรียมข้อมูลเพื่อหากฎความสัมพันธ์ ข้อมูลหลังการรวมคุณลักษณะหมวดหมู่ที่มีผู้ถือบัตรเดียวกัน และข้อมูลหลังการเปลี่ยนชนิดของข้อมูลคุณลักษณะหมวดหมู่ แสดงดังรูปที่ 3.21

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Agency N	Cardholder	Cardholder	accessories	airline	bakery	banquet	bookstore	butcher st	camera an	car access	car dealer	car wash	casino
2	1000	Adams	A	0	0	0	0	0	0	0	0	0	0	0
3	1000	Adkins	A	0	0	0	0	1	0	0	0	0	0	0
4	77000	ADMIRE	A	0	0	0	0	1	0	0	0	0	0	0
5	1000	Akem	A	0	0	0	0	0	0	0	1	0	0	0
6	1000	Alexander	A	0	0	0	0	0	0	0	0	0	0	0
7	1000	Allen	A	0	0	0	0	1	0	0	0	0	0	0
8	1000	Apblett	A	0	0	0	0	1	0	0	0	0	0	0
9	34500	Aranda	A	0	0	0	0	0	0	0	1	0	0	0
10	1000	Arena	A	0	0	0	0	1	0	0	0	0	0	0
11	1000	Arshakian	A	0	0	0	0	0	0	0	0	0	0	0
12	1000	Barclay	A	0	0	0	0	0	0	0	0	0	0	0
13	1000	Barnes	A	0	0	0	0	0	0	1	0	0	0	0
14	78000	BARNES	A	0	0	0	0	0	0	0	0	0	0	0
15	53000	Batenhorst	A	0	0	0	0	1	0	0	0	0	0	0
16	1000	Benn	A	0	0	0	0	1	0	0	0	0	0	0
17	1000	Begosh	A	0	0	0	0	0	0	0	1	0	0	0
18	1000	Beker	A	0	0	0	0	0	0	0	0	0	0	0
19	77000	BENBROOK	A	0	0	0	0	0	0	0	0	0	0	0
20	4000	Bender	A	0	0	0	0	0	0	0	0	0	0	0

รูปที่ 3.21 ตัวอย่างข้อมูลหลังการเปลี่ยนชนิดของข้อมูลคุณลักษณะหมวดหมู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

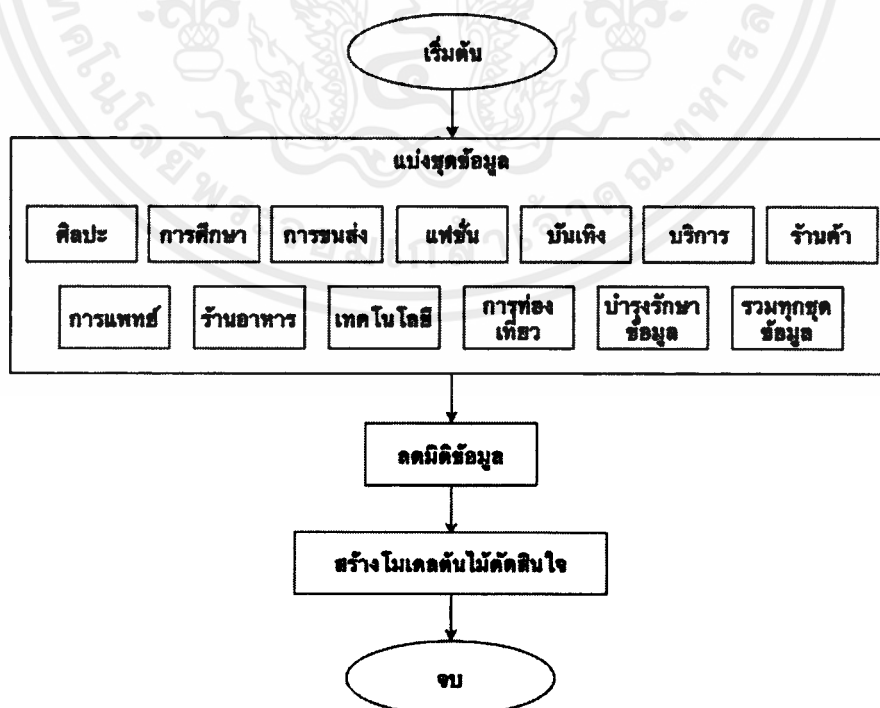
หลังจากข้อมูลการแบ่งชุดข้อมูลแล้ว และเปลี่ยนข้อมูลเป็นข้อมูลธุรกรรมที่ผู้ถือบัตรเดียวกัน รวมหมวดหมู่กันแล้ว ต่อไปคือนำข้อมูลหากฎความสัมพันธ์ของข้อมูลโดยแต่ละชุดข้อมูลมีการ กำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำแสดงดังตารางที่ 3.1

ตารางที่ 3.1 การกำหนดค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำในแต่ละชุดข้อมูล

Dataset	min support	min confidence
summer	0.03	0.80
fall	0.03	0.80
winter	0.03	0.80
spring	0.05	0.80
earlymonth	0.04	0.80
midmonth	0.06	0.80
latemonth	0.04	0.80
weekdays	0.90	0.90
weekends	0.08	0.80
holidays	0.02	0.80

3.3.2 การสร้างโมเดลต้นไม้ตัดสินใจ

การสร้างโมเดลต้นไม้ตัดสินใจ มีลำดับขั้นตอนการเตรียมข้อมูลสำหรับสร้างโมเดลต้นไม้ตัดสินใจ แสดงดังรูปที่ 3.22



รูปที่ 3.22 ลำดับขั้นตอนการเตรียมข้อมูลสำหรับสร้างโมเดลต้นไม้ตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการสร้างโมเดลต้นไม้ตัดสินใจ เริ่มจากการเตรียมชุดข้อมูลเพื่อการสร้างโมเดลต้นไม้ตัดสินใจ โดยชุดข้อมูลก่อนการเตรียมข้อมูล แสดงดังรูปที่ 3.23 และชุดข้อมูลหลังเตรียมที่พร้อมสำหรับการสร้างโมเดลต้นไม้ตัดสินใจ แสดงดังรูปที่ 3.24 โดยจะทำการเตรียมชุดข้อมูลทั้งหมด แบ่งออกเป็น 13 ชุด ได้แก่ ชุดข้อมูลที่แบ่งตามหมวดหมู่ศิลปะ ชุดข้อมูลที่แบ่งตามหมวดหมู่การศึกษา ชุดข้อมูลที่แบ่งตามหมวดหมู่แฟชั่น ชุดข้อมูลที่แบ่งตามหมวดหมู่บันเทิง ชุดข้อมูลที่แบ่งตามหมวดหมู่การแพทย์ ชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านอาหาร ชุดข้อมูลที่แบ่งตามหมวดหมู่บริการ ชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้า ชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยี ชุดข้อมูลที่แบ่งตามหมวดหมู่การขนส่ง ชุดข้อมูลที่แบ่งตามหมวดหมู่การท่องเที่ยว ชุดข้อมูลที่แบ่งตามหมวดหมู่การบำรุงรักษารถยนต์ และชุดข้อมูลที่แบ่งรวมทุกหมวดหมู่

การเตรียมข้อมูลจะทำการลดมิติของข้อมูล จากคุณลักษณะทั้งหมดเหลือ 6 คุณลักษณะดังนี้ คุณลักษณะวัน คุณลักษณะชื่อเดือน คุณลักษณะฤดู คุณลักษณะช่วงของเดือน คุณลักษณะประเภทของวัน และคุณลักษณะหมวดหมู่ย่อย

Agency N	Agency N	Cardhold	Cardhold	Descripti	Amount	Amount2	Category	Category	Date	Month	Year	Day	MonthN	Season	SplitMon	DayType
1000	OKLAHO	Hladik	J	Ultimate	6	Very Low	art	gallery	10	4	2014	Thursday	April	Spring	Earlymor	Weekday
56600	DEPARTN	Bemo	R	Fried chi	6	Very Low	art	gallery	13	5	2014	Tuesday	May	Spring	Midmon	Weekday
1000	OKLAHO	Hladik	J	Apple a d	8	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemon	Weekday
1000	OKLAHO	Hladik	J	Weight s	13	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemon	Weekday
1000	OKLAHO	Hladik	J	Bowl of P	13	Very Low	art	gallery	15	11	2013	Friday	Novemb	Fall	Midmon	Weekday
1000	OKLAHO	Hladik	J	Man smc	13	Very Low	art	gallery	13	12	2013	Friday	Decembe	Winter	Midmon	Weekday
1000	OKLAHO	Hladik	J	Woman I	14	Very Low	art	gallery	24	2	2014	Monday	February	Winter	Latemon	Weekday
76000	UNIVERS	GARDNEI	J	GENERAL	14.41	Very Low	art	gallery	21	5	2014	Wednes	May	Spring	Latemon	Weekday
1000	OKLAHO	Hladik	J	Upset Of	15	Very Low	art	gallery	2	1	2014	Thursday	January	Winter	Earlymor	Weekday
76000	UNIVERS	MCKNIG	T	GENERAL	17	Very Low	art	gallery	25	2	2014	Tuesday	February	Winter	Latemon	Weekday
1000	OKLAHO	Hladik	J	Lifeguard	19	Very Low	art	gallery	21	10	2013	Monday	October	Fall	Latemon	Weekday
56600	DEPARTN	Bemo	R	red lights	19	Very Low	art	gallery	28	1	2014	Tuesday	January	Winter	Latemon	Weekday
60500	REGENTS	Ferguson	K	Teacher (19	Very Low	art	gallery	3	9	2013	Tuesday	Septemb	Fall	Earlymor	Weekday
56600	DEPARTN	Bemo	R	Three lee	19	Very Low	art	gallery	21	2	2014	Friday	February	Winter	Latemon	Weekday
56600	DEPARTN	Bemo	R	Grass NM	19	Very Low	art	gallery	18	3	2014	Tuesday	March	Spring	Midmon	Weekday
76000	UNIVERS	ASHLEY	T	Prepaid:	19.99	Very Low	art	gallery	14	2	2014	Friday	February	Winter	Midmon	Weekday
1000	OKLAHO	Cunning	A	Prepaid:	19.99	Very Low	art	gallery	23	6	2014	Monday	June	Summer	Latemon	Weekday

รูปที่ 3.23 ตัวอย่างชุดข้อมูลก่อนเตรียมข้อมูล

	A	B	C	D	E	F
1	Day	Month	Season	SplitMonth	DayType	Category2
2	Friday	May	Spring	Earlymonth	Weekdays	camera an
3	Tuesday	February	Winter	Midmonth	Weekdays	camera an
4	Friday	March	Spring	Earlymonth	Weekdays	camera an
5	Friday	January	Winter	Midmonth	Weekdays	camera an
6	Wednesday	February	Winter	Midmonth	Weekdays	camera an
7	Monday	December	Winter	Midmonth	Weekdays	camera an
8	Thursday	March	Spring	Latemonth	Weekdays	camera an
9	Sunday	April	Spring	Latemonth	Weekend	camera an
10	Tuesday	May	Spring	Latemonth	Weekdays	camera an
11	Friday	June	Summer	Latemonth	Weekdays	camera an
12	Monday	March	Spring	Earlymonth	Weekdays	camera an
13	Wednesday	June	Summer	Latemonth	Weekdays	camera an
14	Monday	December	Winter	Midmonth	Weekdays	camera an
15	Tuesday	February	Winter	Midmonth	Weekdays	camera an

รูปที่ 3.24 ตัวอย่างชุดข้อมูลหลังเตรียมข้อมูลเพื่อนำไปสร้างโมเดลต้นไม้ตัดสินใจ

3.4 การพัฒนาเว็บแอปพลิเคชัน

ในส่วนของการพัฒนาเว็บแอปพลิเคชันแบ่งออกเป็น การแสดงผลภาพรวมของข้อมูลด้วยกราฟ การแสดงผลกฎความสัมพันธ์ การแสดงโมเดลต้นไม้ตัดสินใจ และการนำเข้ากฎความสัมพันธ์และโมเดลต้นไม้ตัดสินใจ ซึ่งแสดงรายละเอียดในหัวข้อ 3.4.1 ถึง 3.4.4 ตามลำดับ

3.4.1 การแสดงผลภาพรวมของข้อมูลด้วยกราฟ

ในส่วนของการแสดงผลภาพรวมของข้อมูลด้วยกราฟ จะแสดงข้อมูลเชิงสถิติในรูปแบบกราฟวงกลม โดยที่สามารถเลือกคุณลักษณะที่จะมาแสดงข้อมูล โดยที่มีคุณลักษณะดังนี้ คุณลักษณะฤดู (Season) คุณลักษณะประเภทของวัน (DayType) คุณลักษณะช่วงของเดือน (SplitMonth) คุณลักษณะวัน (Day) คุณลักษณะเดือน (Month) และคุณลักษณะหมวดหมู่ (Category) และรูปแบบกราฟแท่ง โดยที่สามารถเลือกคุณลักษณะส่วนที่จะแสดงข้อมูลได้ 2 คุณลักษณะ โดยที่มีคุณลักษณะที่ 1 ดังนี้ คุณลักษณะฤดู (Season) คุณลักษณะประเภทของวัน (DayType) คุณลักษณะช่วงของเดือน (SplitMonth) และคุณลักษณะที่ 2 ดังนี้ คุณลักษณะวัน (Day) และคุณลักษณะหมวดหมู่ (Category)

3.4.2 การแสดงผลกฎความสัมพันธ์

ในส่วนของหน้าแสดงผลกฎความสัมพันธ์ ผู้ใช้สามารถเลือกดูกฎความสัมพันธ์ของชุดข้อมูลที่สนใจ โดยที่แบ่งออกเป็นชุดข้อมูลดังนี้ ชุดข้อมูลที่แบ่งตามคุณลักษณะฤดู 4 ชุด ได้แก่ ฤดูร้อน ฤดูใบไม้ร่วง ฤดูหนาว และฤดูใบไม้ผลิ ชุดข้อมูลที่แบ่งตามคุณลักษณะช่วงของเดือน 3 ชุด ได้แก่ ต้นเดือน กลางเดือน และปลายเดือน และชุดข้อมูลที่แบ่งตามคุณลักษณะประเภทของวัน 3 ชุด ได้แก่ วันธรรมดา วันหยุดสุดสัปดาห์ และวันหยุดเทศกาล

3.4.3 การแสดงโมเดลต้นไม้ตัดสินใจ

ในส่วนของหน้าแสดงโมเดลต้นไม้ตัดสินใจ ผู้ใช้สามารถเลือกดูโมเดลต้นไม้ตัดสินใจตามหมวดหมู่ที่ต้องการได้ โดยที่แบ่งออกเป็นหมวดต่าง ๆ ดังนี้ ศิลปะ การศึกษา แพ้ชั้น การแพทย์ ร้านอาหาร บริการ ร้านค้า เทคโนโลยี การขนส่ง การท่องเที่ยว การบำรุงรักษารถยนต์ และชุดข้อมูลที่รวมทุกหมวดหมู่

3.4.4 การนำเข้ากฎความสัมพันธ์และโมเดลต้นไม้ตัดสินใจ

ในส่วนของหน้าการนำเข้ากฎความสัมพันธ์และการนำเข้าโมเดลต้นไม้ตัดสินใจ นั้นผู้ใช้จำเป็นต้องเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบก่อนที่จะทำการนำเข้ากฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ โดยหลังจากผู้ดูแลระบบเข้าสู่ระบบแล้ว ผู้ดูแลระบบจำเป็นต้องเลือกว่าจะนำเข้ากฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ จากนั้นเลือกชุดข้อมูลที่จะนำเข้า และกดยืนยันการนำเข้าเพื่อเป็นอันเสร็จสมบูรณ์

ตารางที่ 4.2 กฎความสัมพันธ์ของชุดข้อมูลดูใบไม้ร่วง

No.	Antecedent	Consequent	Support	Confidence	Lift
1	photocopy, electrical appliance store, computer software	retailer	0.031	0.820	2.069
2	photocopy, furniture, electrical appliance store	retailer	0.037	0.809	2.043

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะฤดู กฎความสัมพันธ์ของชุดข้อมูลฤดูหนาว โดยที่ กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.03 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.3

ตารางที่ 4.3 กฎความสัมพันธ์ของชุดข้อมูลฤดูหนาว

No.	Antecedent	Consequent	Support	Confidence	Lift
1	shopping mall, repair data	bookstore	0.035	0.821	3.060
2	retailer, repair data	bookstore	0.042	0.820	3.059
3	photocopy, bookstore, furniture	retailer	0.031	0.814	2.108
4	photocopy, bookstore, electrical appliance store	retailer	0.038	0.810	2.096

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะฤดู กฎความสัมพันธ์ของชุดข้อมูลดูใบไม้ผลิ โดยที่ กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.05 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.4

ตารางที่ 4.4 กฎความสัมพันธ์ของชุดข้อมูลดูใบไม้ผลิ

No.	Antecedent	Consequent	Support	Confidence	Lift
1	shopping mall, repair data	bookstore	0.053	0.850	3.165
2	retailer, repair data	bookstore	0.060	0.833	3.102
3	photocopy, repair data	bookstore	0.054	0.813	3.027

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะช่วงของเดือน กฎความสัมพันธ์ของชุดข้อมูลต้นเดือน โดยที่ กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.04 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.5

ตารางที่ 4.5 กฎความสัมพันธ์ของชุดข้อมูลต้นเดือน

No.	Antecedent	Consequent	Support	Confidence	Lift
1	retailer, repair data	bookstore	0.051	0.818	2.938
2	photocopy, repair data	bookstore	0.049	0.810	2.906
3	photocopy, furniture, electrical appliance store	retailer	0.047	0.821	1.966
4	photocopy, bookstore, furniture	retailer	0.043	0.819	1.962
5	photocopy, hardware	retailer	0.047	0.803	1.923

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะช่วงของเดือน กฎความสัมพันธ์ของชุดข้อมูลกลางเดือน โดยที่กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.06 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.6

ตารางที่ 4.6 กฎความสัมพันธ์ของชุดข้อมูลกลางเดือน

No.	Antecedent	Consequent	Support	Confidence	Lift
1	retailer, repair data	bookstore	0.061	0.860	2.900
2	shopping mall, furniture, electrical appliance store	retailer	0.054	0.809	1.909
3	shopping mall, photocopy, electrical appliance store	retailer	0.058	0.805	1.900

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะช่วงของเดือน กฎความสัมพันธ์ของชุดข้อมูลปลายเดือน โดยที่กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.04 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.7

ตารางที่ 4.7 กฎความสัมพันธ์ของชุดข้อมูลปลายเดือน

No.	Antecedent	Consequent	Support	Confidence	Lift
1	photocopy, repair data	bookstore	0.050	0.846	3.021
2	retailer, repair data	bookstore	0.053	0.839	2.998
3	shopping mall, repair data	bookstore	0.044	0.839	2.998
4	photocopy, furniture, electrical appliance store	retailer	0.049	0.861	2.112

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะประเภทของวัน กฎความสัมพันธ์ของชุดข้อมูลวันธรรมดา โดยที่กำหนดค่านับสนับสนุนขั้นต่ำเท่ากับ 0.9 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.9 แสดงตามตารางที่ 4.8

ตารางที่ 4.8 กฎความสัมพันธ์ของชุดข้อมูลวันธรรมดา

No.	Antecedent	Consequent	Support	Confidence	Lift
1	retailer, repair data	bookstore	0.233	0.901	1.834
2	shopping mall, photocopy, electrical appliance store, furniture	retailer	0.235	0.971	1.306
3	photocopy, electrical appliance store, furniture	retailer	0.285	0.960	1.291
4	photocopy, hardware	retailer	0.230	0.955	1.285
5	photocopy, furniture, bookstore	retailer	0.243	0.953	1.282

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะประเภทของวัน กฎความสัมพันธ์ของชุดข้อมูลวันหยุดสุดสัปดาห์ โดยที่กำหนดค่านับสนับสนุนขั้นต่ำเท่ากับ 0.08 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.9

ตารางที่ 4.9 กฎความสัมพันธ์ของชุดข้อมูลวันหยุดสุดสัปดาห์

No.	Antecedent	Consequent	Support	Confidence	Lift
1	hotel, restaurant	fast food restaurant	0.026	0.833	13.066
2	hotel, fast food restaurant	restaurant	0.026	0.876	12.501
3	computer software, repair data	bookstore	0.017	0.891	3.163
4	retailer, repair data	bookstore	0.017	0.877	3.114
5	photocopy, repair data	bookstore	0.020	0.815	2.894

กฎความสัมพันธ์ของชุดข้อมูลที่แบ่งตามคุณลักษณะประเภทของวัน กฎความสัมพันธ์ของชุดข้อมูลวันหยุดเทศกาล โดยที่กำหนดค่านับสนับสนุนขั้นต่ำเท่ากับ 0.02 และค่าความเชื่อมั่นขั้นต่ำเท่ากับ 0.8 แสดงตามตารางที่ 4.10

ตารางที่ 4.10 กฎความสัมพันธ์ของชุดข้อมูลวันหยุดเทศกาล

No.	Antecedent	Consequent	Support	Confidence	Lift
1	clothes for kids	retailer	0.005	0.800	9.118

จากตารางแสดงกฎความสัมพันธ์ข้างต้น No. เป็นลำดับของกฎ Antecedent เป็นเหตุการณ์ส่วนซ้ายของกฎ Consequent เป็นเหตุการณ์ส่วนขวาของกฎ Support เป็นค่าที่บ่งบอกว่าเหตุการณ์ส่วนซ้ายกับเหตุการณ์ส่วนขวามีความถี่ในการเกิดขึ้นมากน้อยแค่ไหน Confidence เป็นค่าที่บ่งบอกว่าเมื่อเกิดเหตุการณ์ส่วนขวาแล้วมีโอกาสที่จะเกิดเหตุการณ์ส่วนซ้ายมากน้อยแค่ไหน Lift เป็นค่าที่บ่งบอกว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายและเหตุการณ์ส่วนขวามีความสัมพันธ์มากน้อยแค่ไหนโดย ถ้าค่ามากกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายเป็นการส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าน้อยกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายไม่ได้ส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าเท่ากับ 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนซ้ายและเหตุการณ์ส่วนขวาไม่มีความสัมพันธ์กัน

4.2 ผลการดำเนินงานของการสร้างโมเดลต้นไม้ตัดสินใจ

ผลการดำเนินงานของการสร้างโมเดลต้นไม้ตัดสินใจ มีชุดข้อมูลทั้งหมด 13 ชุด ได้แก่ ชุดข้อมูลที่แบ่งตามหมวดหมู่ศิลปะ ชุดข้อมูลที่แบ่งตามหมวดหมู่การศึกษา ชุดข้อมูลที่แบ่งตามหมวดหมู่แฟชั่น ชุดข้อมูลที่แบ่งตามหมวดหมู่บันเทิง ชุดข้อมูลที่แบ่งตามหมวดหมู่การแพทย์ ชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านอาหาร ชุดข้อมูลที่แบ่งตามหมวดหมู่บริการ ชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้า ชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยี ชุดข้อมูลที่แบ่งตามหมวดหมู่การขนส่ง ชุดข้อมูลที่แบ่งตามหมวดหมู่การท่องเที่ยว ชุดข้อมูลที่แบ่งตามหมวดหมู่การบำรุงรักษารถยนต์ และชุดข้อมูลที่แบ่งรวมทุกหมวดหมู่ โดยผลการดำเนินงานของโมเดลต้นไม้ตัดสินใจของทุกชุดข้อมูล แสดงตามตารางที่ 4.11

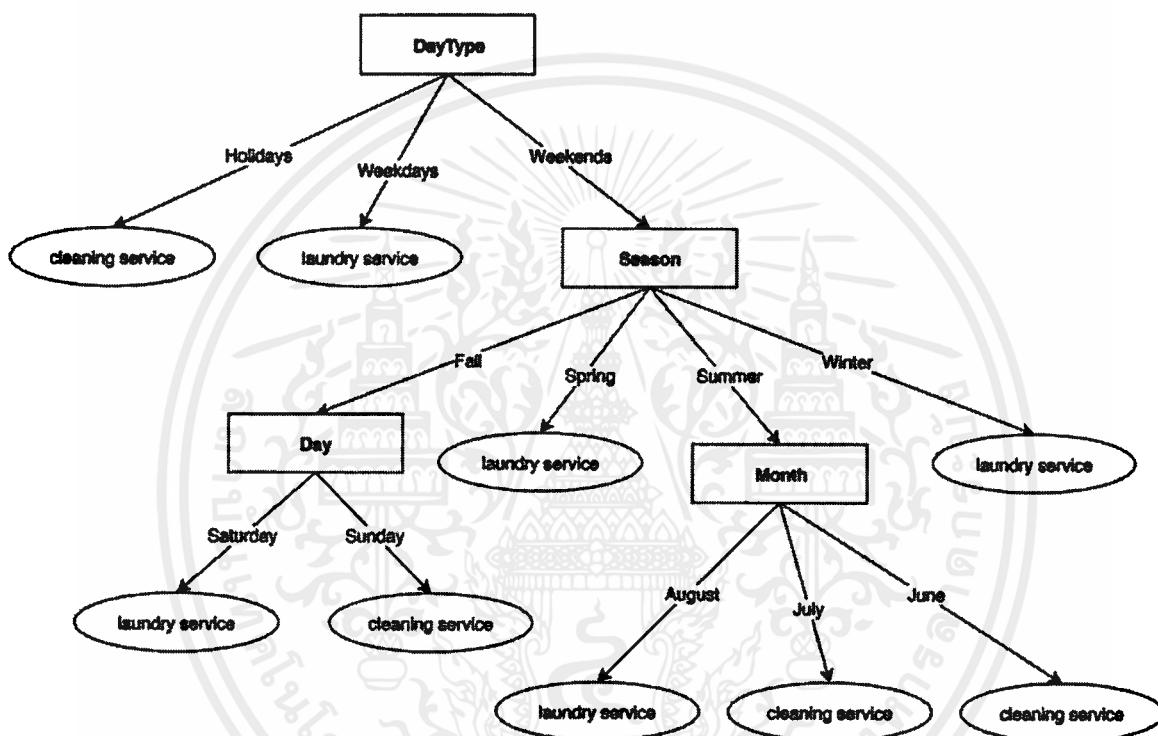
ตารางที่ 4.11 การดำเนินงานของโมเดลต้นไม้ตัดสินใจของทุกชุดข้อมูล

Dataset	#Class	Class proportion	Accuracy	Precision	Recall	F-Measure
all	12	10:40:4:3:24:10:1:114:32:4:60:13	36.60%	10.34%	8.40%	0.070
art	4	86:1:5:6	88.24%	39.53%	25.76%	0.250
education	4	345:27:22:2	86.91%	21.73%	25.00%	0.233
entertainment	3	2:04:29	82.43%	27.48%	33.33%	0.301
fashion	7	4:3:10:4:10:2:1	30.16%	19.19%	16.40%	0.155
medical	5	2000:140:6:143:40	86.22%	17.24%	20.00%	0.185
restaurant	6	2:2:7:43:2:43	43.93%	15.57%	17.01%	0.138
service	2	5:07	58.72%	39.09%	49.70%	0.393
store	8	15:23:1:23:7:20:21:4	22.56%	7.90%	13.84%	0.081
technology	4	7:10:3:12	36.92%	36.01%	28.90%	0.196
transportation	5	17:5:10:4	56.59%	42.70%	39.39%	0.390
travel	5	1580:2800:1200:3:260	53.77%	32.09%	24.91%	0.216
car maintenance	6	81:3:2:25:9:9	63.16%	24.72%	18.24%	0.159

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 ประกอบไปด้วย 7 คอลัมน์ดังนี้ คอลัมน์ชุดข้อมูล (Dataset) คอลัมน์จำนวนคลาส (#Class) คอลัมน์อัตราส่วนของคลาส (Class proportion) คอลัมน์ค่าความถูกต้อง (Accuracy) คอลัมน์ค่าความแม่นยำ (Precision) คอลัมน์ค่ารู้จำ (Recall) และคอลัมน์ประสิทธิภาพ (F-Measure)

ตัวอย่างโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่บริการ โดยที่คลาสคำตอบในชุดข้อมูลนี้เป็นที่เป็นได้มี 2 คลาสคำตอบ คือ บริการทำความสะอาด (cleaning service) และบริการซักรีด (laundry service) แสดงดังรูปที่ 4.1



รูปที่ 4.1 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่บริการ

จากโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่บริการสามารถนำมาเขียนในรูปแบบเงื่อนไขได้

ดังนี้

```

if (DayType = Holidays) {
    then class = cleaning service
}
else if (DayType = Weekdays) {

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

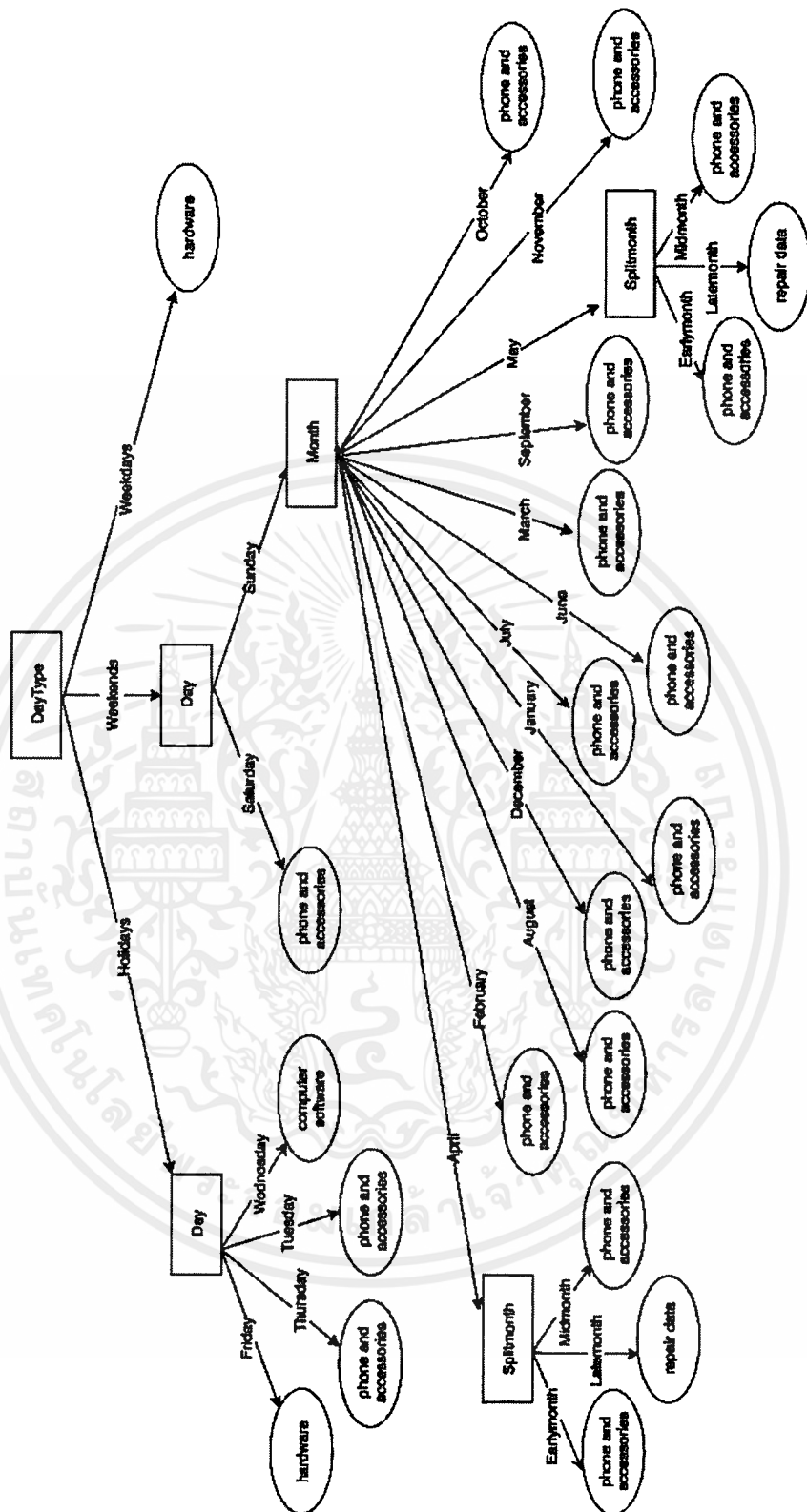
```

    then class = laundry service
}
else if (DayType = Weekends) {
    if (Season = Fall) {
        if (Day = Saturday) {
            then class = laundry service
        }
        else if (Day = Sunday) {
            then class = cleaning service
        }
    }
    else if (Season = Spring) {
        then class = laundry service
    }
    else if (Season = Summer) {
        if (Month = August) {
            then class = laundry service
        }
        else if (Month = July) {
            then class = cleaning service
        }
        else if (Month = June) {
            then class = cleaning service
        }
    }
}
else if (Season = Winter) {
    then class = laundry service
}
}

```

ตัวอย่างโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยี โดยที่คลาสคำตอบในชุดข้อมูลนี้เป็นที่เป็นได้มี 4 คลาสคำตอบ คือ ซอฟต์แวร์คอมพิวเตอร์ (computer software) ฮาร์ดแวร์ (hardware) การซ่อมแซมข้อมูล (repair data) และโทรศัพท์และอุปกรณ์เสริม (phone and accessories) แสดงดังรูปที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่เทคโนโลยีสามารถนำมาเขียนในรูปแบบเงื่อนไข
ได้ดังนี้

```

if (DayType = Holidays) {
    if (Day = Friday) {
        then class = hardware
    }
    else if (Day = Thursday) {
        then class = phone and accessories
    }
    else if (Day = Tuesday) {
        then class = phone and accessories
    }
    else if (Day = Wednesday) {
        then class = computer software
    }
}
else if (DayType = Weekdays) {
    then class = hardware
}
else if (DayType = Weekends) {
    if (Day = Saturday) {
        then class = phone and accessories
    }
    else if (Day = Sunday) {
        if (Month = April) {
            if (SplitMonth = Earlymonth) {
                then class = phone and accessories
            }
            else if (SplitMonth = Latemonth) {
                then class = repair data
            }
            else if (SplitMonth = Midmonth) {
                then class = phone and accessories
            }
        }
    }
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    }
}
else if (Month = August) {
    then class = phone and accessories
}
else if (Month = December) {
    then class = phone and accessories
}
else if (Month = February) {
    then class = phone and accessories
}
else if (Month = January) {
    then class = phone and accessories
}
else if (Month = July) {
    then class = phone and accessories
}
else if (Month = June) {
    then class = phone and accessories
}
else if (Month = March) {
    then class = phone and accessories
}
else if (Month = May) {
    if (SplitMonth = Earlymonth) {
        then class = phone and accessories
    }
    else if (SplitMonth = Latemonth) {
        then class = repair data
    }
    else if (SplitMonth = Midmonth) {
        then class = phone and accessories
    }
}

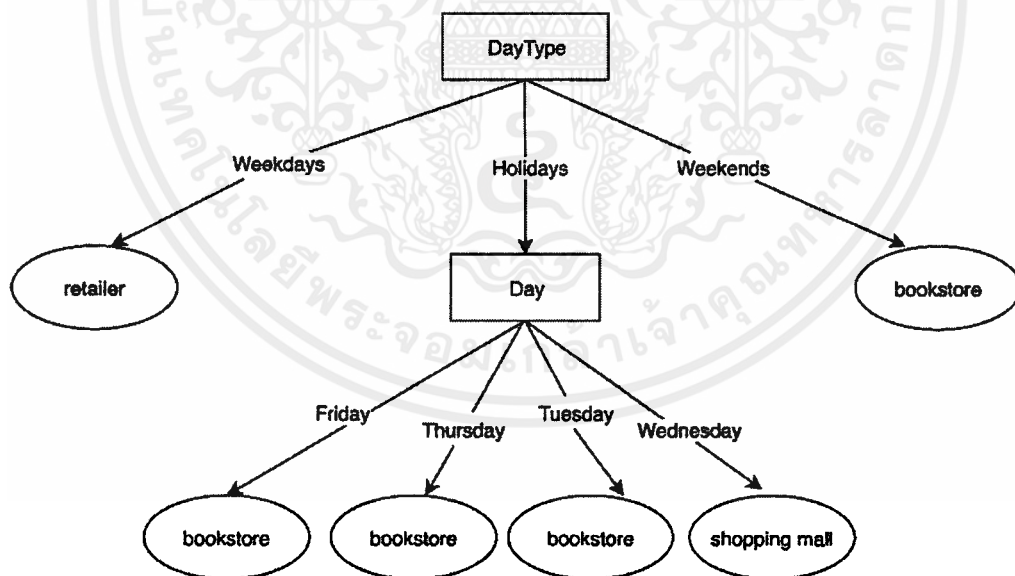
```

```

}
else if (Month = November) {
    then class = phone and accessories
}
else if (Month = October) {
    then class = phone and accessories
}
else if (Month = September) {
    then class = phone and accessories
}
}
}

```

ตัวอย่างโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้า โดยที่คลาสคำตอบในชุดข้อมูลนี้เป็นที่เป็นได้มี 8 คลาสคำตอบ คือ ร้านหนังสือ (bookstore) ร้านขายเครื่องใช้ไฟฟ้า (electrical appliance store) เฟอร์นิเจอร์ (furniture) ร้านขายสัตว์เลี้ยง (pet shop) ร้านซ่อม (repair shop) ร้านค้าปลีก (retailer) ห้างสรรพสินค้า (shopping mall) และร้านขายของที่ระลึก (souvenir shop) แสดงดังรูปที่ 4.3



รูปที่ 4.3 โมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้า

จากโมเดลต้นไม้ตัดสินใจชุดข้อมูลที่แบ่งตามหมวดหมู่ร้านค้าสามารถนำมาเขียนในรูปแบบเงื่อนไขได้
ดังนี้

```

if (DayType = Holidays) {
    if (Day = Friday) {
        then class = bookstore
    }
    else if (Day = Thursday) {
        then class = bookstore
    }
    else if (Day = Tuesday) {
        then class = bookstore
    }
    else if (Day = Wednesday) {
        then class = shopping mall
    }
}
else if (DayType = Weekdays) {
    then class = retailer
}
else if (DayType = Weekends) {
    then class = bookstore
}

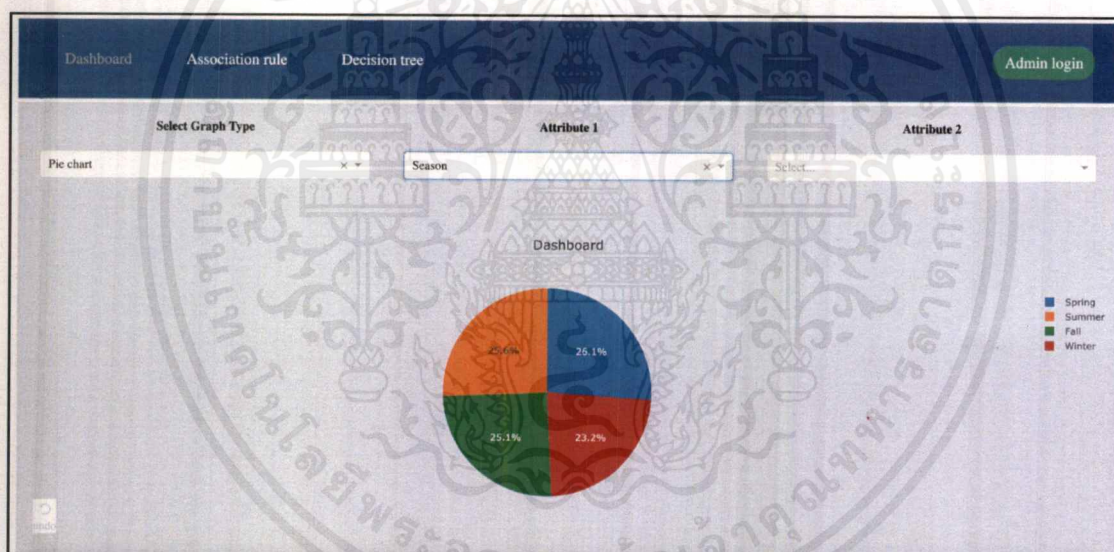
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การพัฒนาเว็บแอปพลิเคชัน

ในส่วนของเว็บแอปพลิเคชัน แบ่งผู้ใช้งานเว็บแอปพลิเคชันออกเป็น ผู้ใช้ทั่วไป และผู้ดูแลระบบ โดยที่ผู้ใช้ทั่วไปมีสิทธิในการใช้งานเว็บแอปพลิเคชันในส่วนของการแสดงผลภาพรวมของข้อมูลด้วยกราฟ การแสดงผลกฎความสัมพันธ์ การแสดงโมเดลต้นไม้ตัดสินใจ ส่วนผู้ดูแลระบบมีสิทธิการใช้งานเว็บแอปพลิเคชันในส่วน การนำเข้ากฎความสัมพันธ์และโมเดลต้นไม้ตัดสินใจ รวมทั้งการแสดงผลข้อมูลด้วยกราฟ กฎความสัมพันธ์ และโมเดลต้นไม้ตัดสินใจ

การแสดงผลภาพรวมของข้อมูลด้วยกราฟ จะแสดงเป็นข้อมูลเชิงสถิติในรูปแบบกราฟวงกลม โดยที่ผู้ใช้สามารถเลือกคุณลักษณะที่จะมาแสดงข้อมูล โดยที่มีคุณลักษณะดังนี้ คุณลักษณะฤดู (Season) คุณลักษณะประเภทของวัน (DayType) คุณลักษณะช่วงของเดือน (SplitMonth) คุณลักษณะวัน (Day) คุณลักษณะเดือน (Month) และคุณลักษณะหมวดหมู่ (Category) แสดงดังรูปที่ 4.4

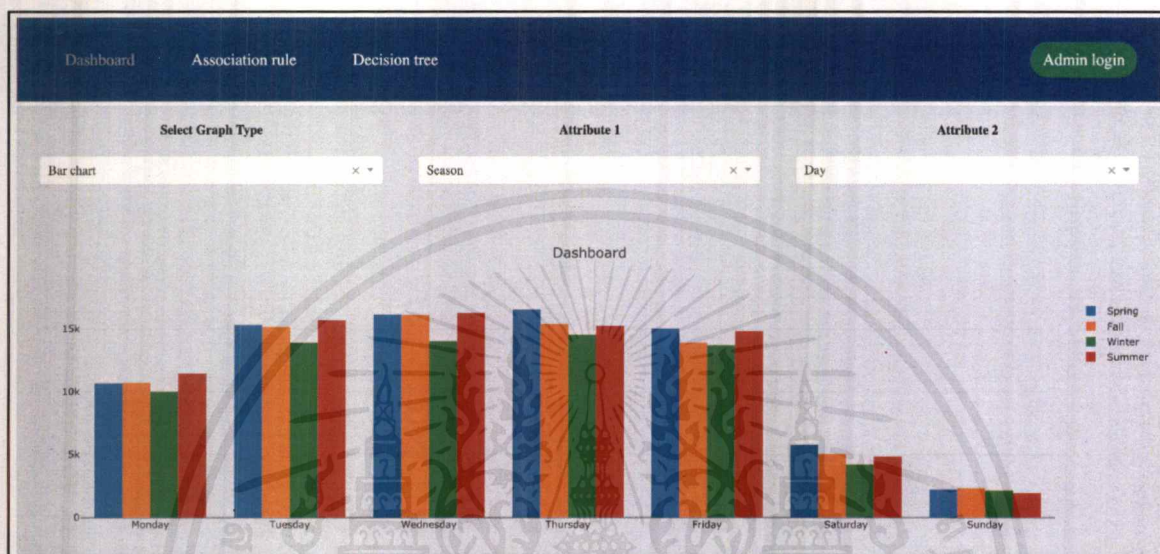


รูปที่ 4.4 หน้าแสดงผลภาพรวมของข้อมูลด้วยกราฟรูปแบบกราฟวงกลม

จากรูปคือการเลือกรูปแบบของกราฟ (Graph Type) เป็น กราฟวงกลม (Pie chart) แล้วเลือกคุณลักษณะ (Attribute) ที่ 1 เป็นคุณลักษณะฤดูกาล (Season) เพื่อแสดงผลภาพรวมของข้อมูลในรูปแบบกราฟวงกลม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแสดงผลแบบกราฟแท่ง ผู้ใช้สามารถเลือกคุณลักษณะส่วนที่จะแสดงข้อมูลได้ 2 คุณลักษณะ ซึ่งสามารถเลือกคุณลักษณะที่ 1 คือ คุณลักษณะฤดู (Season) คุณลักษณะประเภทของวัน (DayType) คุณลักษณะช่วงของเดือน (SplitMonth) และคุณลักษณะที่ 2 คือ คุณลักษณะวัน (Day) และคุณลักษณะหมวดหมู่ (Category) แสดงดังรูปที่ 4.5



รูปที่ 4.5 หน้าแสดงผลภาพรวมของข้อมูลด้วยกราฟรูปแบบกราฟแท่ง

จากรูปคือการเลือกรูปแบบของกราฟ (Graph Type) เป็น กราฟแท่ง (Bar chart) แล้วเลือกคุณลักษณะ (Attribute) ที่ 1 เป็นคุณลักษณะฤดูกาล (Season) และคุณลักษณะ (Attribute) ที่ 2 เป็นคุณลักษณะวัน (Day) เพื่อแสดงผลภาพรวมของข้อมูลในรูปแบบกราฟแท่ง

การแสดงผลจะแสดงกฎความสัมพันธ์ ผู้ใช้สามารถเลือกชุดข้อมูลกฎความสัมพันธ์ได้ ซึ่งประกอบด้วยกฎความสัมพันธ์ในชุดข้อมูลฤดูร้อน กฎความสัมพันธ์ในชุดข้อมูลฤดูใบไม้ร่วง กฎความสัมพันธ์ในชุดข้อมูลฤดูหนาว กฎความสัมพันธ์ในชุดข้อมูลฤดูใบไม้ผลิ กฎความสัมพันธ์ในชุดข้อมูลต้นเดือน กฎความสัมพันธ์ในชุดข้อมูลกลางเดือน กฎความสัมพันธ์ในชุดข้อมูลปลายเดือน กฎความสัมพันธ์ในชุดข้อมูลวันธรรมดา กฎความสัมพันธ์ในชุดข้อมูลวันหยุดสุดสัปดาห์ และกฎความสัมพันธ์ในชุดข้อมูลวันหยุดเทศกาล ตัวอย่างแสดงดังรูปที่ 4.6

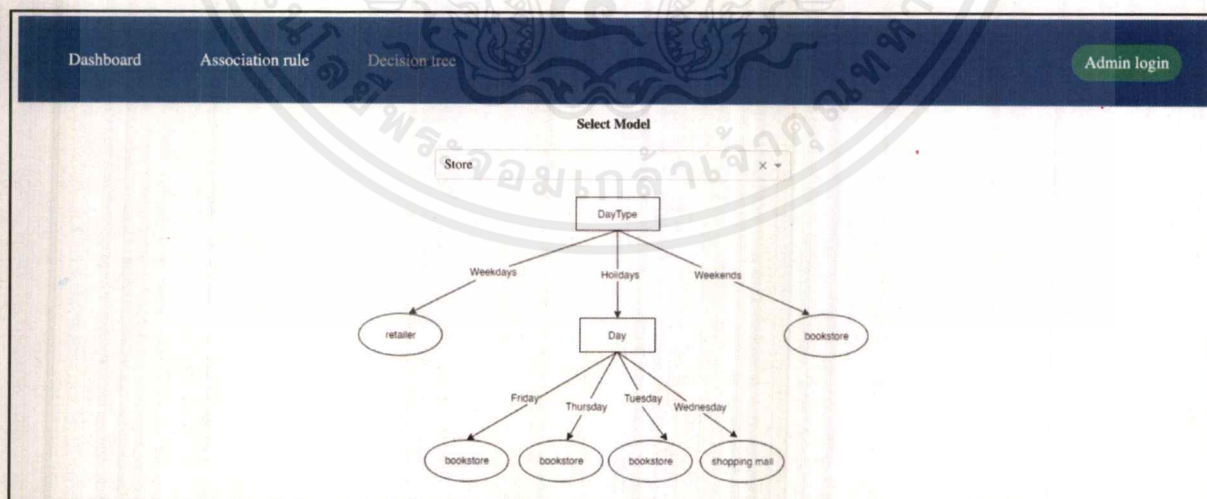
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Select Model					
Day Type					
Weekdays					
No.	Antecedent	Consequent	Support	Confidence	Lift
1	retailer, information	bookstore	0.233	0.901	1.834
2	shopping mall, photocopy, electrical appliance store, furniture	retailer	0.235	0.971	1.306
3	photocopy, electrical appliance store, furniture	retailer	0.285	0.960	1.291
4	photocopy, hardware	retailer	0.230	0.955	1.285
5	photocopy, furniture, bookstore	retailer	0.243	0.953	1.282
Weekends					
No.	Antecedent	Consequent	Support	Confidence	Lift
1	retailer, information	bookstore	0.233	0.901	1.834
2	shopping mall, photocopy	retailer	0.235	0.971	1.306

รูปที่ 4.6 หน้าแสดงผลกฎความสัมพันธ์

จากรูปคือการเลือกโมเดลเป็นกฎความสัมพันธ์ในชุดข้อมูลประเภทของวัน (Day Type) เพื่อแสดงกฎความสัมพันธ์ของวันธรรมดา (Weekdays) วันหยุดสุดสัปดาห์ (Weekends) และวันหยุดเทศกาล (Holidays)

การแสดงผลจะแสดงโมเดลต้นไม้ตัดสินใจ ผู้ใช้สามารถเลือกชุดข้อมูลของโมเดลต้นไม้ตัดสินใจที่จะแสดงได้ ซึ่งประกอบด้วย ศิลปะ การศึกษา แพชั่น บันเทิง การแพทย์ ร้านอาหาร บริการ ร้านค้า เทคโนโลยี การขนส่ง การท่องเที่ยว การบำรุงรักษารถยนต์ และชุดข้อมูลรวมทุกหมวดหมู่ ตัวอย่างแสดงดังรูปที่ 4.7



รูปที่ 4.7 หน้าแสดงโมเดลต้นไม้ตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปคือการเลือกโมเดลต้นไม้ตัดสินใจของชุดข้อมูลร้านค้า (Store) เพื่อแสดงโมเดลต้นไม้ตัดสินใจของชุดข้อมูลร้านค้า

ในส่วนของการนำเข้ากฎความสัมพันธ์และการนำเข้าโมเดลต้นไม้ตัดสินใจ นั้นจำเป็นต้องเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบก่อนที่จะทำการนำเข้ากฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ ดังรูปที่ 4.8 และกรณีการยืนยันในการเป็นผู้ดูแลระบบผิดพลาดจะแสดงข้อความแจ้งเตือน ดังรูปที่ 4.9 โดยหลังจากผู้ดูแลระบบเข้าสู่ระบบแล้ว ผู้ดูแลระบบจำเป็นต้องเลือกว่าจะนำเข้ากฎความสัมพันธ์หรือโมเดลต้นไม้ตัดสินใจ จากนั้นเลือกชุดข้อมูลที่จะนำเข้า และกดยืนยันการนำเข้าเพื่อเป็นอันเสร็จสมบูรณ์ ดังรูปที่ 4.10

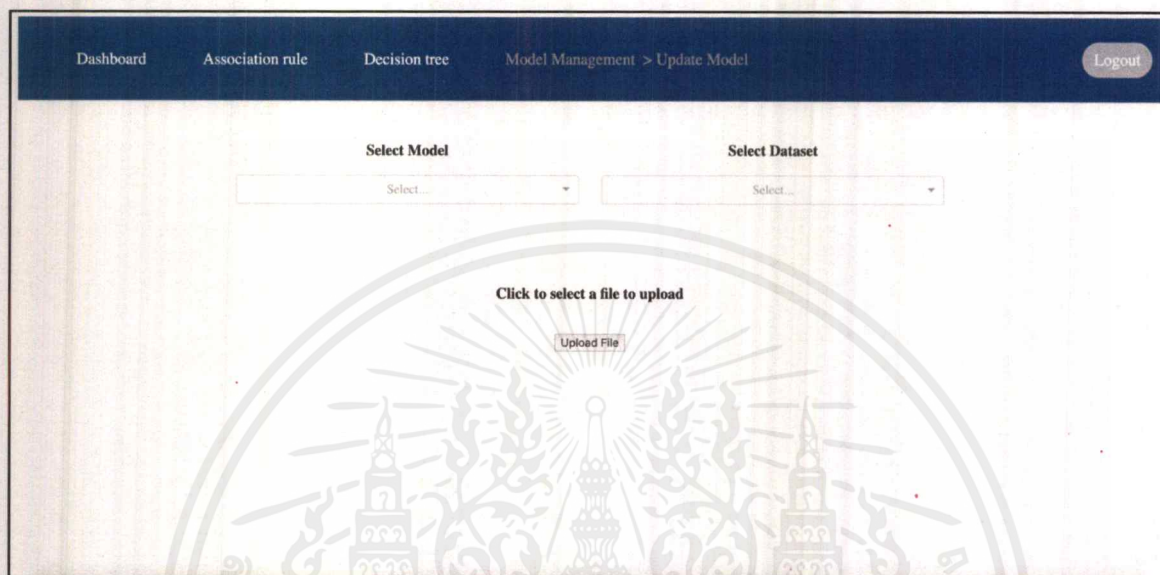
รูปที่ 4.8 หน้าเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบ

โดยในส่วนของการเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบนั้นจำเป็นต้องป้อนชื่อผู้ใช้ (username) และรหัสผ่าน (password) แล้วกดปุ่มเข้าสู่ระบบ (Login) เพื่อเข้าสู่ระบบ

รูปที่ 4.9 หน้าเข้าสู่ระบบเพื่อยืนยันในการเป็นผู้ดูแลระบบในกรณีการยืนยันผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปคือการการเป็นผู้ดูแลระบบนั้นถ้าป้อนชื่อผู้ใช้ (username) หรือป้อนรหัสผ่าน (password) ผิดพลาดจะมีข้อความแจ้งเตือนว่า Login Fail!! (Check username, password) เพื่อแจ้งให้ทราบว่ามีการป้อนชื่อผู้ใช้ หรือ รหัสผ่าน ผิดพลาด



รูปที่ 4.10 หน้าการนำเข้ากฎความสัมพันธ์และการนำเข้าโมเดลต้นไม้ตัดสินใจ

โดยขั้นตอนการนำเข้ากฎความสัมพันธ์และการนำเข้าโมเดลต้นไม้ตัดสินใจ มีขั้นตอนดังนี้ เลือกประเภทข้อมูลที่จะนำเข้า (Select Import type) เลือกคุณลักษณะ (Select Attribute) และทำการกดอัปโหลดข้อมูล (Upload File) เพื่อเลือกข้อมูลที่ต้องการนำเข้าและอัปโหลดข้อมูล หลังจากกดอัปโหลดข้อมูล จะเพิ่มไฟล์ข้อมูล que เลือกไว้ในที่เก็บข้อมูลและมีชื่อตามประเภทข้อมูลที่จะนำเข้าและคุณลักษณะ เช่น ถ้าประเภทข้อมูลที่จะนำเข้า เป็น กฎความสัมพันธ์ (Association rule) และเลือกคุณลักษณะฤดูที่เป็นฤดูใบไม้ผลิ (Spring) จะเพิ่มไฟล์ข้อมูลชื่อ AssociationRule_spring ไว้ในที่เก็บข้อมูล

4.4 อภิปรายผลการดำเนินงาน

ในส่วนนี้เป็นการอภิปรายผลการดำเนินงาน แบ่งออกเป็นการอภิปรายผลการดำเนินงานของการหา
 กฎความสัมพันธ์ การสร้างโมเดลต้นไม้ตัดสินใจ

4.4.1 การอภิปรายผลการดำเนินงานของการหาความสัมพันธ์

จากตารางแสดงกฎความสัมพันธ์ทั้งหมด ได้กฎความสัมพันธ์ที่มีความน่าสนใจที่เกิดขึ้นในกฎ
 ความสัมพันธ์ของชุดข้อมูลวันหยุดสุดสัปดาห์ 2 กฎความสัมพันธ์ ได้แก่ กฎความสัมพันธ์ที่ 1 ถ้ามี
 รายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่โรงแรม (hotel) และร้านอาหาร (restaurant) แล้วจะ
 มีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านอาหารจานด่วน (fast food restaurant) กฎ
 ความสัมพันธ์ที่ 2 ถ้ามีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่โรงแรม (hotel) และร้านอาหาร
 จานด่วน (fast food restaurant) แล้วจะมีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านอาหาร
 (restaurant) จากกฎความสัมพันธ์ทั้ง 2 กฎมีค่า Lift เท่ากับ 13.066 และ 12.501 ตามลำดับ โดย
 กฎความสัมพันธ์ทั้ง 2 กฎนั้นถือว่าเป็นกฎที่มีความน่าสนใจ

และในส่วนชุดข้อมูลที่เกิดกฎความสัมพันธ์ที่มีการเกิดขึ้นร่วมกันนั้น คือชุดข้อมูลต้นเดือน
 กลางเดือน และปลายเดือน โดยกฎความสัมพันธ์ที่เกิดขึ้นร่วมกัน ได้แก่ กฎความสัมพันธ์ ถ้ามีรายการ
 การใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านค้าปลีก (retailer) และบริการกู้ข้อมูล (repair data) แล้ว
 จะมีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านหนังสือ (bookstore) ดังนั้นในการหา
 กฎความสัมพันธ์ของข้อมูลการแบ่งชุดข้อมูลออกตามช่วงของเดือนจะได้กฎความสัมพันธ์ที่ไม่แตกต่างกัน
 ในแต่ละชุดข้อมูล ในการหาความสัมพันธ์ของข้อมูลจึงไม่จำเป็นต้องแบ่งชุดข้อมูลตามช่วงของ
 เดือน

จากกฎความสัมพันธ์ทั้งหมดจะเห็นได้ว่าในหมวดหมู่ที่เกิดขึ้นส่วนมากเป็นหมวดหมู่ที่
 เกี่ยวข้องกับการศึกษา เพราะชุดข้อมูลที่นำมาใช้ในการหาความสัมพันธ์เป็นชุดข้อมูลที่เก็บมาจาก
 การซื้อสินค้าและบริการผ่านบัตรเครดิตของนักศึกษา และข้อมูลที่มีจะเป็นข้อมูลในช่วงวัยที่เป็นช่วง
 วัยนักศึกษาเท่านั้น ทำให้ข้อมูลที่เกิดขึ้นมีหมวดหมู่ที่เกี่ยวข้องกับการศึกษามากกว่าหมวดหมู่อื่น

4.4.2 การอภิปรายผลการดำเนินงานของการสร้างโมเดลต้นไม้ตัดสินใจ

จากผลการดำเนินงาน พบว่าโมเดลต้นไม้ตัดสินใจบางโมเดลมีค่าความถูกต้อง (Accuracy) สูง แต่โมเดลมีค่า F-Measure ที่ต่ำ เนื่องจากข้อมูลในชุดข้อมูลมีจำนวนของคลาสคำตอบในแต่ละหมวดหมู่ไม่สมดุลกัน และการแบ่งหมวดหมู่ไม่ตรงตามหมวดหมู่ที่เหมาะสม

ข้อมูลในชุดข้อมูลมีจำนวนของคลาสคำตอบในแต่ละหมวดหมู่ไม่สมดุลกัน มักจะเกิดปัญหาในการทำโมเดลต้นไม้ตัดสินใจ แสดงตามตารางที่ 4.12

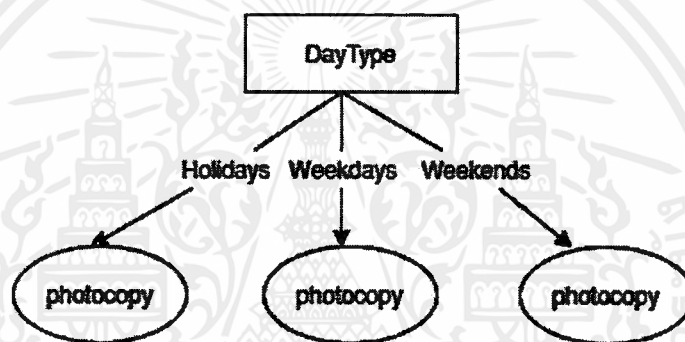
ตารางที่ 4.12 อัตราส่วนข้อมูลที่ไม่สมดุลกัน

Dataset	#Class	Class proportion	F-Measure
all	12	10:40:4:3:24:10:1:114:32:4:60:13	0.070
art	4	86:1:5:6	0.250
education	4	345:27:22:2	0.233
entertainment	3	2:04:29	0.301
fashion	7	4:3:10:4:10:2:1	0.155
medical	5	2000:140:6:143:40	0.185
restaurant	6	2:2:7:43:2:43	0.138
service	2	5:07	0.393
store	8	15:23:1:23:7:20:21:4	0.081
technology	4	7:10:3:12	0.196
transportation	5	17:5:10:4	0.390
travel	5	1580:2800:1200:3:260	0.216
car maintenance	6	81:3:2:25:9:9	0.159

จากการสังเกตจะเห็นว่าหมวดหมู่การศึกษามีอัตราส่วนของข้อมูลเป็น 345 : 27 : 22 : 2 ซึ่งแสดงให้เห็นว่าข้อมูลมีความเอนเอียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่สรุปได้ว่าโมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่ศิลปะ โมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่การศึกษา โมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่บันเทิง โมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่การแพทย์ โมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่การท่องเที่ยว และโมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่การบำรุงรักษารถยนต์ เป็นชุดข้อมูลที่ไม่มีความสมดุลกัน ดังนั้นโมเดลต้นไม้ตัดสินใจที่แบ่งตามหมวดหมู่ที่กล่าวข้างต้นเป็นโมเดลที่มีประสิทธิภาพน้อยกว่า โมเดลต้นไม้ตัดสินใจที่แบ่งตามหมวดหมู่อื่น ๆ เพราะมีการทนายคลาสค่าตอบที่เอนเอียงมากเป็นค่าตอบเดียว แสดงดังรูปที่ 4.11 แสดงการทนายคลาสค่าตอบที่เอนเอียงมากเป็นค่าตอบเดียวโดยยกตัวอย่างโมเดลต้นไม้ตัดสินใจของหมวดหมู่การศึกษา



รูปที่ 4.11 โมเดลต้นไม้ตัดสินใจหมวดหมู่การศึกษา

สาเหตุที่โมเดลที่มีประสิทธิภาพต่ำ เนื่องมาจากการแบ่งหมวดหมู่ไม่ตรงตามหมวดหมู่ที่เหมาะสม เพราะหมวดหมู่มีความคล้ายคลึงกันจึงเกิดปัญหาการแบ่งหมวดหมู่ย่อยไม่เหมาะสมตามหมวดหมู่ ยกตัวอย่างในหมวดหมู่การขนส่ง จะประกอบด้วยหมวดหมู่ย่อย 4 หมวด คือ สถานีบริการรถแท็กซี่ ค่าทางด่วน และบริการขนส่ง และในหมวดหมู่ท่องเที่ยว ประกอบด้วยหมวดหมู่ย่อย 5 หมวดหมู่ คือ สายการบิน โรงแรม ค่าเช่ารถ รถไฟ และตัวแทนการท่องเที่ยว และในหมวดหมู่การบำรุงรักษารถยนต์ ประกอบด้วยหมวดหมู่ย่อย 6 หมวด คือ อุปกรณ์ตกแต่งรถยนต์ ตัวแทนจำหน่ายรถยนต์ ล้างรถ น้ำมัน ค่าจอดรถ และซ่อมรถ โดยที่หมวดหมู่ย่อยรถไฟเหมาะสมที่จะจัดอยู่หมวดหมู่การขนส่งกว่าอยู่ในหมวดหมู่ท่องเที่ยว และหมวดหมู่ย่อยน้ำมันเหมาะสมที่จะจัดอยู่หมวดหมู่การขนส่งกว่าอยู่ในหมวดหมู่การบำรุงรักษารถยนต์

สรุปผลการดำเนินงานและข้อเสนอแนะ

5.1 สรุปผลการดำเนินงาน

ปัญหาพิเศษนี้นำเสนอวิธีการวิเคราะห์ข้อมูลสำหรับการวิเคราะห์พฤติกรรมผู้ใช้งานบัตรเครดิต โดยใช้เทคนิคการหาความสัมพันธ์ของข้อมูล การสร้างโมเดลต้นไม้ตัดสินใจ และการพัฒนาเว็บแอปพลิเคชัน

จากการดำเนินการหาความสัมพันธ์ ได้กฎความสัมพันธ์ที่มีความน่าสนใจที่เกิดขึ้นในกฎความสัมพันธ์ของชุดข้อมูลวันหยุดสุดสัปดาห์ 2 กฎความสัมพันธ์ ได้แก่ กฎความสัมพันธ์ที่ 1 ถ้ามีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่โรงแรม (hotel) และร้านอาหาร (restaurant) แล้วจะมีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านอาหารจานด่วน (fast food restaurant) กฎความสัมพันธ์ที่ 2 ถ้ามีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่โรงแรม (hotel) และร้านอาหารจานด่วน (fast food restaurant) แล้วจะมีรายการการใช้จ่ายผ่านบัตรเครดิตในหมวดหมู่ร้านอาหาร (restaurant) จากกฎความสัมพันธ์ทั้ง 2 กฎมีค่า Lift เท่ากับ 13.066 และ 12.501 ตามลำดับ ซึ่งค่า Lift เป็นค่าที่บ่งบอกว่าการเกิดรูปแบบเหตุการณ์ส่วนช้ายและเหตุการณ์ส่วนขวามีความสัมพันธ์มากน้อยแค่ไหนโดยถ้าค่ามากกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนช้ายเป็นการส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าน้อยกว่า 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนช้ายไม่ได้การส่งเสริมการเกิดรูปแบบเหตุการณ์ส่วนขวา ถ้าค่าเท่ากับ 1 แสดงว่าการเกิดรูปแบบเหตุการณ์ส่วนช้ายเหตุการณ์ส่วนขวาไม่มีความสัมพันธ์กัน

จากการดำเนินการสร้างโมเดลต้นไม้ตัดสินใจ ในชุดข้อมูลที่มีความเอนเอียง (imbalanced dataset) จะทำให้โมเดลต้นไม้ตัดสินใจมีการทำนายแค่คลาสคำตอบที่มีความเอนเอียงมากเท่านั้น ประกอบไปด้วยโมเดลต้นไม้ตัดสินใจที่แบ่งข้อมูลตามหมวดหมู่ศิลปะ หมวดหมู่การศึกษา หมวดหมู่บันเทิง หมวดหมู่การแพทย์ หมวดหมู่การท่องเที่ยว และหมวดหมู่ยานพาหนะ เป็นชุดข้อมูลที่ข้อมูลที่ไม่มีความสมดุลกัน ดังนั้นโมเดลต้นไม้ตัดสินใจที่แบ่งตามหมวดหมู่ที่กล่าวข้างต้นเป็นโมเดลที่มีการทำนายแค่คลาสคำตอบที่มีความเอนเอียงเท่านั้น

จากการดำเนินการพัฒนาเว็บแอปพลิเคชัน สามารถแบ่งส่วนของเว็บแอปพลิเคชันตามผู้ใช้งานได้ ดังนี้ ส่วนของผู้ใช้งานทั่วไป (user) แบ่งออกเป็น หน้าแสดงผลภาพรวมของข้อมูล หน้าแสดงผลกฎความสัมพันธ์ และหน้าแสดงผลโมเดลต้นไม้ตัดสินใจ และส่วนของผู้ดูแลระบบ (admin) แบ่งออกเป็น ส่วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อัปโหลดข้อมูลกฎความสัมพันธ์และโมเดลต้นไม้ตัดสินใจ และส่วนของการลบข้อมูลกฎความสัมพันธ์และโมเดลต้นไม้ตัดสินใจ

5.2 ข้อเสนอแนะ

จากผลการดำเนินงานจะพบว่าในการแปลงข้อมูลในส่วนของคุณลักษณะหมวดหมู่ เป็นกระบวนการที่ต้องทำอย่างรอบคอบ ถ้าแปลงคุณลักษณะหมวดหมู่แล้วยังมีความกำกวมอยู่ จะทำให้ผลการดำเนินงานออกมาไม่มีประสิทธิภาพ เช่น หมวดหมู่การขนส่ง (Transportation) และการท่องเที่ยว (Travel) จะต้องแปลงอย่างชัดเจนเช่น รถแท็กซี่ (taxi) ควรอยู่ในหมวดหมู่การขนส่งมากกว่าหมวดหมู่การท่องเที่ยวหรือไม่



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

บริษัท กรุงเทพธุรกิจ มีเดีย จำกัด. 2562. กระบวนการสากลเพื่อการวิเคราะห์ข้อมูล. [Online]. Available: <http://www.bangkokbiznews.com/blog/detail/646353>. เข้าถึงเมื่อวันที่ 5 ธ.ค. 62.

บริษัท ฐานเศรษฐกิจ มัลติมีเดีย จำกัด. 2560. ปรับกลยุทธ์สู่รูปรีดอิมตัว. [Online]. Available: <http://www.thansettakij.com/content/236650>. เข้าถึงเมื่อวันที่ 2 ธ.ค. 62.

บริษัท อู๊ปส์ เน็ตเวิร์ค จำกัด. 2558. ข้อมูลมหาศาลที่ทำให้การทำการตลาดนั้นอัจฉริยะขึ้น. [Online]. Available: <https://www.marketingoops.com/news/biz-news/big-data-and-better-marketing>. เข้าถึงเมื่อวันที่ 1 ธ.ค. 62.

เอกสิทธิ์ พิชรวงศ์ศักดิ์. 2557. ขั้นตอนการหากฎความสัมพันธ์. [Online]. Available: <http://dataminingtrend.com/2014/association-rules>. เข้าถึงเมื่อวันที่ 7 ธ.ค. 62.

เอกสิทธิ์ พิชรวงศ์ศักดิ์. 2557. ขั้นตอนการสร้างโมเดลต้นไม้ตัดสินใจ. [Online]. Available: <http://dataminingtrend.com/2014/decision-tree-model>. เข้าถึงเมื่อวันที่ 10 ธ.ค. 62.