

การทำเหมืองข้อมูลเพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุด
มหาวิทยาลัยราชภัฏศรีสะเกษ

นายธีรพงศ์ สงัด



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2556

Data Mining Techniques to Support the Library Service of
Sisaket Rajabhat University

Mr. Teerapong Songputh



A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Information and Communication Technology
School of Science and Technology
Sukhothai Thammathirat Open University

2013

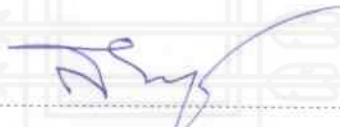
หัวข้อวิทยานิพนธ์ การทำเหมืองข้อมูลเพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุด
มหาวิทยาลัยราชภัฏศรีสะเกษ
ชื่อและนามสกุล นายธีรพงศ์ สงฆ์
แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร
สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์สำรวย กมลายุตต์
2. อาจารย์ ดร.ดวงดาว วิชาดากุล

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 28 มีนาคม 2557
คณะกรรมการสอบวิทยานิพนธ์



..... ประธานกรรมการ

(รองศาสตราจารย์ศิริภัทรา เหมือนมาลัย)



..... กรรมการ

(รองศาสตราจารย์สำรวย กมลายุตต์)



..... กรรมการ

(อาจารย์ ดร.ดวงดาว วิชาดากุล)



..... ประธานกรรมการบัณฑิตศึกษา

(ศาสตราจารย์ ดร.สิริวรรณ ศรีพล)

ชื่อวิทยานิพนธ์ การทำเหมืองข้อมูลเพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุด
มหาวิทยาลัยราชภัฏศรีสะเกษ

ผู้วิจัย นายธีรพงศ์ สงผัด **รหัสนักศึกษา** 2549600167

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)

อาจารย์ที่ปรึกษา (1) รองศาสตราจารย์สำรวย กมลายุตต์ (2) อาจารย์ ดร. ดวงดาว วิชาตากุล

ปีการศึกษา 2556

บทคัดย่อ

งานวิจัยนี้นำเสนอการใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) เพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ วัตถุประสงค์ของงานวิจัยคือ 1) การสร้างคลังข้อมูลผู้ใช้บริการห้องสมุด 2) การทำเหมืองข้อมูลเพื่อวิเคราะห์และจำแนกกลุ่มผู้ใช้บริการห้องสมุด 3) การทำเหมืองข้อมูลเพื่อวิเคราะห์และศึกษารูปแบบการยืมหนังสือของผู้ใช้บริการห้องสมุด 4) การจัดทำรายงานหลายมิติจากข้อมูลที่จัดเก็บในคลังข้อมูลตามความต้องการของผู้ใช้ เครื่องมือที่ใช้ในงานวิจัยได้แก่ โปรแกรม SQL Server 2008 และ SQL Service Analysis กระบวนการเริ่มจากการนำข้อมูลการใช้ห้องสมุดจากฐานข้อมูลห้องสมุดตั้งแต่ปี พ.ศ. 2553 - 2556 และข้อมูลทั่วไปของนักศึกษาจากฐานข้อมูลระบบทะเบียนนักศึกษามาสร้างคลังข้อมูลด้วยกระบวนการอีทีแอล (ETL) วัตถุประสงค์ของกระบวนการอีทีแอลคือ เพื่อให้ข้อมูลมีความถูกต้องสมบูรณ์ จากนั้นนำข้อมูลจากคลังข้อมูลมาสร้างรายงานหลายมิติตามความต้องการผ่านกระบวนการโอแอล สำหรับการทำให้เหมืองข้อมูลใช้แบบจำลองคริสป์-ดีเอ็ม (CRISP-DM Model) และเทคนิคการทำเหมืองข้อมูลที่ใช้ประกอบด้วย 3 เทคนิค ได้แก่ 1) การจัดกลุ่ม เพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ 2) กฎการหาความสัมพันธ์ วิเคราะห์ข้อมูลการยืมหนังสือเพื่อค้นหาความสัมพันธ์ที่น่าสนใจระหว่างหนังสือที่ถูกยืมว่า มีหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยกัน และ 3) อนุกรมเวลา เพื่อพยากรณ์ปริมาณผู้ใช้บริการสารสนเทศห้องสมุดจำแนกตามประเภทของผู้ใช้บริการ

ผลการวิจัยที่ได้คือ คลังข้อมูลผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษที่ใช้โครงสร้างแบบสโนว์เฟลก (Snowflake schema Model) ผู้ใช้บริการที่มีคุณลักษณะคล้ายกันถูกจัดออกเป็นทั้งหมด 6 กลุ่ม การหาความสัมพันธ์ของหนังสือที่มักถูกยืมไปด้วยกัน พบว่าที่ค่าความเชื่อมั่น ไม่น้อยกว่าร้อยละ 90 และค่าสนับสนุนไม่น้อยกว่า 9 ได้กฎความสัมพันธ์ทั้งหมด 20 กฎ และอนุกรมเวลาทำให้สามารถพยากรณ์ปริมาณผู้ใช้บริการสารสนเทศห้องสมุดในแต่ละกลุ่มหรือประเภทได้ การทำเหมืองข้อมูลทำให้บุคลากรทำงานในห้องสมุดได้รับความรู้ใหม่ ซึ่งสามารถนำมาประกอบในการแนะนำหนังสือหรือจัดกิจกรรมส่งเสริมการใช้บริการสารสนเทศของห้องสมุดตามกลุ่มของผู้ใช้ ยิ่งกว่านั้นยังช่วยเพิ่มความสะดวกในการค้นคืนสารสนเทศหรือรายชื่อหนังสือที่ต้องการ ลดเวลาในการสืบค้นช่วยสนับสนุนการวางแผนจัดกิจกรรมเพื่อส่งเสริมการใช้บริการและเพิ่มปริมาณผู้ใช้ นอกจากนี้แล้วยังทำให้ผู้บริหารได้รับรายงานหลายมิติที่สามารถมองภาพรวมและเจาะลึกในรายละเอียดตามมิติที่สนใจได้

คำสำคัญ การทำเหมืองข้อมูล การบริการสารสนเทศ ห้องสมุด แอสโซซิเอชันรูลส์ คลัสเตอร์ิง ไทม์ซีริน

Thesis title: Data Mining Techniques to Support the Library Service of Sisaket Rajabhat University

Researcher: Mr. Teerapong Songpath; **ID:** 2549600167;

Degree: Master of Science (Information and Communication Technology);

Thesis advisors: (1) Sumruay Komlayut, Associate Professor; (2) Dr. Duangdao Wichadakul;

Academic year: 2013

Abstract

This research presented the use of data mining techniques to support the information services of Sisaket Rajabhat University's library. The objectives of this research were: (1) to build a data warehouse of library users (2) to mine the data in order to analyze and cluster library users (3) to mine the data in order to analyze and study circulation patterns (4) to create a multi-dimensional report from data warehouse corresponding to user needs. The tools in this research were SQL Server 2008 and SQL Service Analysis. The processes started from extracting library users' data during 2010 - 2013 from the library database and students' personal data from the registration database system to build data warehouse using the ETL process in order to clean and to correct data. Then OLAP was used to produce multi - dimensional reports from the data warehouse. For data mining, the CRISP-DM model and data mining techniques were used to analyze data as follows: (1) clustering to group library users who had similar characteristics of circulation, (2) association rule to analyze circulation of data to discover interesting relations between books which were borrowed in bundled form and (3) time series to forecast the amount of library users classified by user type.

The results of the research were as follows: a library user data warehouse for Sisaket Rajabhat University based on a snowflake schema model was created and library users were clustered into 6 groups. According to association rules at a reliability of more than 90 percent and support values of at least 9, there were 20 relationship rules of borrowed books. Time series also forecasted the amount of library users by type. Data mining benefited the library staff by providing information which could be used to recommend books or to organize activities for promoting the use of library service, depending on user groups. Moreover, this also made it more convenient for users to retrieve information about desired books. It could thus decrease retrieval time. This knowledge supported activity planning to promote services and increase the amount of users. It also enabled administrators to receive multi-dimensional reports that could be drilled up and drilled down according to dimensions of interest.

Keywords: Data Mining techniques, Library Information Service, Association rules, Clustering, Time series

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์อย่างยิ่งจากท่านผู้ทรงคุณวุฒิหลายท่าน โดยเฉพาะอย่างยิ่งท่านอาจารย์ รองศาสตราจารย์ สำนวณ กมลายุตต์ อาจารย์ที่ปรึกษาหลัก และท่านอาจารย์ ดร.ดวงดาว วิชิตากุล อาจารย์ที่ปรึกษาร่วมวิทยานิพนธ์ฉบับนี้ ซึ่งได้กรุณาเสียสละเวลาในการให้คำปรึกษา ให้ข้อเสนอแนะต่างๆ ให้ความรู้ในการจัดทำวิทยานิพนธ์ตลอดจนแนะนำแนวทางในการศึกษาค้นคว้าเกี่ยวกับเนื้อหาและข้อมูลที่จะนำมาปรับปรุงแก้ไขข้อบกพร่องต่างๆ จนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบคุณมหาวิทยาลัยราชภัฏศรีสะเกษ ที่ได้กรุณาให้โอกาสในการศึกษาต่อ และขอขอบคุณท่านอาจารย์ ดร.อนันต์ศักดิ์ พวงอก หัวหน้าฝ่ายวิทยบริการและสารสนเทศ มหาวิทยาลัยราชภัฏศรีสะเกษ ที่เสียสละเวลาให้คำปรึกษา แนะนำ และให้ข้อมูลที่ครบถ้วนในการทำวิทยานิพนธ์ฉบับนี้ รวมทั้งขอขอบคุณเพื่อนร่วมงานทุกท่านที่เป็นกำลังใจ และให้ความช่วยเหลือ รวมทั้งผู้ที่มีส่วนร่วมทุกท่านที่มีได้เอื้อนามไว้ ณ ที่นี้

นอกจากนี้ผู้วิจัยขอขอบพระคุณคณาจารย์ ประจำสาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อนักศึกษา ตลอดจนผู้ที่มีส่วนเกี่ยวข้องในการทำวิทยานิพนธ์ครั้งนี้ที่ได้กรุณาให้การสนับสนุน ช่วยเหลือ และให้กำลังใจตลอดมาจนกระทั่งการทำวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ด้วยดี

ธีรพงศ์ สงฆ์

มีนาคม 2557

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ญ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของงานวิจัย.....	2
ขอบเขตของงานวิจัย.....	3
กรอบแนวคิดการวิจัย.....	3
นิยามคำศัพท์.....	4
ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง.....	5
การบริการห้องสมุดของมหาวิทยาลัยราชภัฏศรีสะเกษ.....	5
คลังข้อมูล (Data warehouse)	11
การทำเหมืองข้อมูล (Data mining)	17
ซอฟต์แวร์พัฒนาเหมืองข้อมูล.....	30
งานวิจัยที่เกี่ยวข้อง.....	32
บทที่ 3 วิธีดำเนินการวิจัย.....	36
ประชากรและกลุ่มตัวอย่าง.....	36
เครื่องมือที่ใช้ในการดำเนินงานวิจัย.....	36
ขั้นตอนการดำเนินงาน.....	37
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	57
การสร้างคลังข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด.....	57
การทำ Online Analytical Processing (OLAP).....	60
การทำเหมืองข้อมูล.....	62
บทที่ 5 สรุปผล อภิปราย และข้อเสนอแนะ.....	98
สรุปผลงานวิจัย.....	98
การอภิปรายผล.....	106
ข้อเสนอแนะ.....	109
บรรณานุกรม.....	110
ประวัติผู้วิจัย.....	113

สารบัญตาราง

	หน้า
ตารางที่ 3.1 แอ็ททริบิวส์ที่ใช้ในการวิเคราะห์การจัดกลุ่ม.....	47
ตารางที่ 3.2 แอ็ททริบิวส์ตารางข้อมูลการยืมของเล่มหนังสือ.....	48
ตารางที่ 3.3 แอ็ททริบิวส์ตารางข้อมูลการยืมของสมาชิก.....	49
ตารางที่ 3.4 แอ็ททริบิวส์ที่ใช้ในการวิเคราะห์การจัดกลุ่ม.....	50
ตารางที่ 4.1 ตาราง TB_Genier (Dimension Table) สำหรับเก็บข้อมูลเพศ.....	57
ตารางที่ 4.2 ตาราง TB_FBYBook (Dimension Table) สำหรับเก็บข้อมูลรหัสการยืม หนังสือ.....	58
ตารางที่ 4.3 ตาราง TB_Collection (Dimension Table) สำหรับเก็บข้อมูลหมวดหมู่ หนังสือ.....	58
ตารางที่ 4.4 ตาราง TB_Day (Dimension Table) สำหรับเก็บข้อมูลวัน.....	58
ตารางที่ 4.5 ตาราง TB_Item (Dimension Table) สำหรับเก็บข้อมูลหนังสือ.....	58
ตารางที่ 4.6 ตาราง TB_Month (Dimension Table) สำหรับเก็บข้อมูลเดือน.....	58
ตารางที่ 4.7 ตาราง TB_Program (Dimension Table) สำหรับเก็บข้อมูลโปรแกรม นักศึกษา.....	59
ตารางที่ 4.8 ตาราง TB_Student (Dimension Table) สำหรับเก็บข้อมูลนักศึกษา.....	59
ตารางที่ 4.9 ตาราง TB_Borrow (Fact Table) สำหรับเก็บข้อมูลการยืมหนังสือ.....	59
ตารางที่ 4.10 การจัดกลุ่ม (Clustering) อัลกอริธึมได้จำแนกออกเป็น 6 คลัสเตอร์.....	67
ตารางที่ 4.11 แสดงข้อมูลเกรดเฉลี่ยของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	74
ตารางที่ 4.12 แสดงข้อมูลหมวดหมู่ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	74
ตารางที่ 4.13 แสดงข้อมูลวันยืม (อาทิตย์ (1) ถึง เสาร์ (7)) ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	74
ตารางที่ 4.14 แสดงข้อมูลจำนวนหนังสือที่ยืมแต่ละครั้งของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	74
ตารางที่ 4.15 แสดงข้อมูลเพศของแต่ละคลัสเตอร์.....	75

สารบัญตาราง(ต่อ)

หน้า

ตารางที่ 4.16 แสดงข้อมูลเดือนที่ยืม (มกราคม (1) ถึง ธันวาคม (12)) ของแต่ละ คลัสเตอร์ 3 ลำดับสูงสุด.....	75
ตารางที่ 4.17 แสดงข้อมูลชั้นปีที่กำลังศึกษาอยู่ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	75
ตารางที่ 4.18 แสดงข้อมูลโปรแกรมวิชาที่สังกัดของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	75
ตารางที่ 4.19 แสดงข้อมูลความถี่ในการยืมหนังสือในแต่ละเดือนของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด.....	75
ตารางที่ 4.20 แสดงผลข้อมูลการหาความสัมพันธ์ (Association Rule)	79



สารบัญภาพ

หน้า

ภาพที่ 1.1 กรอบแนวคิดการวิจัย.....	3
ภาพที่ 2.1 วิวัฒนาการของสารสนเทศในคลังข้อมูล.....	12
ภาพที่ 2.2 กระบวนการในการพัฒนาคลังข้อมูล.....	13
ภาพที่ 2.3 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล.....	18
ภาพที่ 2.4 สถาปัตยกรรมระบบเหมืองข้อมูล.....	18
ภาพที่ 2.5 การวิเคราะห์พฤติกรรมกรรมการซื้อสินค้า.....	21
ภาพที่ 2.6 แสดงตัวอย่างการค้นหาทฤษฎีความสัมพันธ์.....	22
ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภท.....	24
ภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่ม.....	25
ภาพที่ 2.9 การแบ่งระดับงานของคริสป์-ดีเอ็ม.....	26
ภาพที่ 2.10 แบบจำลอง CRISP-DM.....	28
ภาพที่ 2.11 ตัวอย่างหน้าจอ Microsoft SQL Server 2008.....	30
ภาพที่ 3.1 แสดงขั้นตอนการจัดทำคลังข้อมูลการใช้บริการยืมคืนของห้องสมุด.....	38
ภาพที่ 3.2 แสดงโครงสร้างคลังข้อมูลแบบสโนว์เฟลค.....	38
ภาพที่ 3.3 การใช้เครื่องมือ Integration Service ในกระบวนการอีทีแอล.....	39
ภาพที่ 3.4 การนำข้อมูลจากไฟล์ Excel เข้าตาราง.....	40
ภาพที่ 3.5 แสดงรายการข้อมูลที่เป็นค่าว่างหรือไม่สามารถระบุข้อมูลได้ (เป็นค่า NULL) ...	41
ภาพที่ 3.6 คำสั่งลบข้อมูลที่เป็นค่าว่างหรือไม่สามารถระบุข้อมูลได้ (เป็นค่า NULL)	41
ภาพที่ 3.7 สร้าง Data Source ชื่อว่า DWLIBRARY.ds.....	43
ภาพที่ 3.8 สร้าง Data Source View ชื่อว่า DWLIBRARY.dsv.....	43
ภาพที่ 3.9 สร้าง Cube.....	44
ภาพที่ 3.10 Deploy โพรเจกต์.....	44
ภาพที่ 3.11 แสดงผลในลักษณะหลายมิติ.....	45
ภาพที่ 3.12 แสดงข้อมูลรหัสนักศึกษาเพื่อให้ทราบว่านักศึกษาและคนมายืมหนังสือกี่ครั้ง...	45

สารบัญภาพ (ต่อ)

	หน้า
ภาพที่ 3.13 แสดงการนำข้อมูลเข้าตาราง TB_DM_Custering โดยการแม็ปข้อมูล.....	48
ภาพที่ 3.14 แสดงการนำข้อมูลเข้าตาราง TB_DM_Borrow โดยการแม็ปข้อมูล.....	49
ภาพที่ 3.15 แสดงข้อมูลตาราง TB_DM_Timeseries.....	50
ภาพที่ 3.16 เลือกตารางหรือวิวที่ต้องการนำมาประมวลผลด้วยเทคนิคดาต้าไมนิ่ง.....	51
ภาพที่ 3.17 เลือกตารางที่ใช้ในการประมวลผลในอัลกอริธึมการจัดกลุ่ม (Clustering)	52
ภาพที่ 3.18 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึมการจัดกลุ่ม (Clustering)	52
ภาพที่ 3.19 การเลือกตาราง TB_DM_Borrow และตาราง TB_FBYBook.....	53
ภาพที่ 3.20 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึม Apriori.....	53
ภาพที่ 3.21 การเลือกตาราง TB_DM_Timeseries.....	54
ภาพที่ 3.22 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึมไทม์ซีรีส์ (Time Series)	54
ภาพที่ 3.23 แสดงการ deploy โพรเจกต์.....	55
ภาพที่ 4.1 แสดงผลที่ได้จากการสร้าง Data Source View ชื่อ DWLIBRARY.dsv.....	60
ภาพที่ 4.2 แสดงผลตัวอย่างการประมวลผลเชิงวิเคราะห์แบบออนไลน์.....	60
ภาพที่ 4.3 การแสดงผลข้อมูลหลายมิติ.....	61
ภาพที่ 4.4 แสดงผลข้อมูลการยืมหนังสือในแต่ละเดือนของนักศึกษาโปรแกรมวิชาต่างๆ.....	61
ภาพที่ 4.5 แสดงผลที่ได้จากการสร้าง Data Source View ชื่อ DM_LIBRARY.dsv.....	63
ภาพที่ 4.6 แสดงตัวอย่างข้อมูล ตาราง TB_DM_Custering.....	64
ภาพที่ 4.7 แสดงความสัมพันธ์ด้วยแอตทริบิวต์ชื่อ Borrow_id.....	64
ภาพที่ 4.8 แสดงตัวอย่างข้อมูล ตาราง TB_DM_Timeseries.....	65
ภาพที่ 4.9 กำหนดค่าพารามิเตอร์ การจัดกลุ่ม (Clustering)	66
ภาพที่ 4.10 การจัดกลุ่ม (Clustering) อัลกอริธึมได้จำแนกออกเป็น 6 คลัสเตอร์.....	67
ภาพที่ 4.11 รายละเอียดคลัสเตอร์ที่ 1.....	68
ภาพที่ 4.12 รายละเอียดคลัสเตอร์ที่ 2.....	69
ภาพที่ 4.13 รายละเอียดคลัสเตอร์ที่ 3.....	70

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

ห้องสมุด คือแหล่งสารสนเทศ บริการทรัพยากรสารสนเทศ ในรูปแบบต่างๆ เช่น หนังสือ วารสาร หนังสือพิมพ์ จุลสาร กฤตภาค วัสดุเทปและโทรทัศน์ CD-ROM DVD VCD รวมถึงไมโครฟิล์มด้วย โดยมีบรรณารักษ์เป็นผู้ดำเนินงานและบริหารงานต่างๆ ในห้องสมุด โดยจัดระบบเป็นหมวดหมู่ และระเบียบเรียบร้อย เพื่อให้ผู้ใช้ห้องสมุดมีความสะดวกสืบค้นได้ง่ายตรงกับความต้องการ ห้องสมุดในปัจจุบันทำหน้าที่เก็บรวบรวม จัดระบบ เพื่อให้บริการสื่อสารสนเทศต่างๆ ตลอดจนจนถึงเทคโนโลยีทางคอมพิวเตอร์และเทคโนโลยีทางการสื่อสารอีกทั้งยังมีเครื่องมือในการค้นหาและดำเนินการให้บริการสื่อต่างๆ เกิดประโยชน์สูงสุดแก่ผู้ใช้ห้องสมุด

ปัจจุบันการสืบค้นหนังสือในห้องสมุดมีความง่าย สะดวก และรวดเร็วมากกว่าที่ผ่านมา เนื่องจากมีการนำเทคโนโลยีสารสนเทศและการสื่อสารมาประยุกต์ โดยเฉพาะห้องสมุดหลายแห่งได้นำซอฟต์แวร์มาต่างๆ มาใช้งาน เพื่ออำนวยความสะดวกในการสืบค้นหนังสือและสื่อต่างๆ ให้แก่ผู้มารับบริการ เช่น ซอฟต์แวร์โอแพ็ก (OPAC : Online Public Access Catalog) เป็นระบบโปรแกรมที่ช่วยค้นหารายการหนังสือ และสื่อต่างๆ ในห้องสมุดแทนที่ระบบแบบเดิมที่เป็นบัตรรายการกระดาษ ซอฟต์แวร์ช่วยให้ผู้ใช้สามารถสืบค้นรายการหนังสือและสื่อต่างๆ ผ่านคอมพิวเตอร์ในห้องสมุด หรือสืบค้นจากคอมพิวเตอร์เครื่องอื่นๆ ผ่านทางเครือข่ายอินเทอร์เน็ตได้ ปัจจุบันเว็บไซต์ของห้องสมุดเกือบทุกที่จะมีบริการสืบค้นหนังสือออนไลน์ ซอฟต์แวร์โอแพ็กช่วยให้ผู้ใช้สามารถสืบค้นหนังสือและสื่อต่างๆ ได้ทุกที่ทุกเวลา อย่างไรก็ตามเมื่อสืบค้นได้รายชื่อหนังสือหรือสื่อที่ต้องการแล้ว เมื่อไปหาหนังสือหรือสื่อที่ตนเองค้นไว้ ผลปรากฏว่าหนังสือหรือสื่อที่สืบค้นได้นั้นมีเนื้อหาไม่ตรงกับความ ต้องการ ทำให้เสียเวลาและต้องสืบค้นใหม่ และอีกปัญหาหนึ่งก็คือ เมื่อสืบค้นหนังสือได้หนึ่งเล่มแล้วหากยังต้องการหนังสือที่เกี่ยวข้องเพิ่มอีกเป็นเล่มที่สองและเล่มที่สามตามมาผู้ใช้ต้องทำการสืบค้นรายการหนังสือหรือสื่อใหม่ทำให้เสียเวลาในการสืบค้นเป็นอย่างมาก

การทำเหมืองข้อมูลเป็นเทคนิคเพื่อค้นหา รูปแบบหรือแพทเทิร์น (Pattern) ความสัมพันธ์ของข้อมูลจากข้อมูลปริมาณมหาศาล เป็นการวิเคราะห์ชุดข้อมูลที่มีขนาดใหญ่ เพื่อค้นหาความสัมพันธ์ของข้อมูลที่ไม่เคยทราบมาก่อน การทำเหมืองข้อมูลเป็นการสำรวจหาความรู้ใหม่ๆ ที่มีประโยชน์ต่อการดำเนินงานของหน่วยงาน การทำเหมืองข้อมูลประกอบด้วยเทคนิคหลายเทคนิคที่พัฒนามาจากทฤษฎีด้านสถิติและด้านปัญญาประดิษฐ์ เช่น มาร์คอฟเชนส์ เครือข่ายประสาทเทียม เป็นต้น การทำเหมืองข้อมูลเป็นกระบวนการของการคัดแยกและจัดรูปแบบชุดข้อมูลจากแหล่งข้อมูลหลายๆ แหล่งให้เป็นรูปแบบเดียวกัน จากนั้นมีการนำชุดข้อมูลดังกล่าวเข้าสู่คลังข้อมูลที่ได้ออกแบบไว้ และสร้างแบบจำลองขึ้นมาวิเคราะห์ชุดข้อมูลนั้น ผลลัพธ์ที่ได้จากการวิเคราะห์คือความรู้ใหม่ที่ไม่เคยทราบมาก่อน และสามารถนำไปใช้ประโยชน์ได้ (Berry&Linoff, 2004: 101)

ปัจจุบันห้องสมุดโดยส่วนใหญ่ มีการนำระบบห้องสมุดอัตโนมัติมาให้บริการ ทั้งการบันทึกและจัดเก็บข้อมูลต่างๆ ที่เกี่ยวข้องไว้ เช่น รายละเอียดทรัพยากรสารสนเทศที่มีอยู่ ข้อมูลส่วนบุคคลของผู้ใช้บริการ รายการข้อมูลที่เกิดจากการที่ผู้ใช้แต่ละคนมาเยี่ยม คืบและจองทรัพยากรสารสนเทศ ซึ่งข้อมูลที่จัดเก็บดังกล่าวสามารถนำมาใช้ประโยชน์โดยการสร้างเป็นคลังข้อมูล และใช้เทคนิคการทำเหมืองข้อมูลในการวิเคราะห์และประมวลผลข้อมูลดังกล่าวเพื่อค้นหารูปแบบการเยี่ยมชมของผู้มาใช้บริการ เช่นการนำกระบวนการ Bibilomining (Nicholson, S. & Stanton, J. 2003: 250) มาประยุกต์กับข้อมูลที่เก็บไว้ เพื่อค้นหาความสัมพันธ์ของหนังสือที่ถูกยืมกับผู้ยืม จะช่วยทำให้สามารถแบ่งหรือจัดกลุ่มผู้ยืมได้และยังสามารถบอกได้ว่าแต่ละกลุ่มเป็นนักศึกษาหรือนักวิจัย หรืออาจารย์จากภาควิชาใด มีรายละเอียดการยืมหนังสือแตกต่างกันอย่างไรความรู้เหล่านี้จะช่วยทำให้ห้องสมุดทราบความต้องการของผู้ยืมหนังสือแต่ละกลุ่ม และใช้เป็นแนวทางในการจัดการ จัดซื้อและจัดหาหนังสือให้สอดคล้องกับความต้องการของผู้ยืมได้และสามารถนำมาใช้วางนโยบายต่างๆ ในการบริหารห้องสมุดได้อย่างมีประสิทธิภาพ (ไพโรจน์ เอี่ยมชัยมงคล, 2552: 3) ผลการวิเคราะห์ที่ได้ช่วยทำให้ผู้ให้บริการทราบถึงพฤติกรรมการยืมหนังสือของสมาชิกห้องสมุด ตลอดจนความต้องการใช้หนังสือของสมาชิกห้องสมุด โดยบรรณารักษ์สามารถนำข้อมูลเกี่ยวกับพฤติกรรมของสมาชิกห้องสมุดมาใช้ประโยชน์ในการให้บริการสืบค้นหนังสือเป็นการแนะนำให้แก่ผู้ใช้ว่ามีหนังสือเล่มใดบ้างที่มีเนื้อหาคล้ายกัน หรือนำประวัติการยืมหนังสือมาวิเคราะห์ว่าหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยกัน แล้วนำข้อมูลที่วิเคราะห์ได้จัดเก็บไว้ในระบบสืบค้นโอแพค (OPAC) เพื่อให้ระบบแสดงรายชื่อหนังสือที่เกี่ยวข้องกันให้ผู้ทราบด้วย ซึ่งจะเป็นเพิ่มทางเลือกให้กับผู้ใช้มากขึ้น หนังสือที่ได้จะมีเนื้อหาตรงตามความต้องการของผู้ใช้โดยไม่ต้องเสียเวลาในการสืบค้นหลายรอบ นอกจากนี้แล้ว อาจนำข้อมูลทั่วไปของสมาชิกและข้อมูลการใช้บริการมาวิเคราะห์ร่วมกันโดยใช้เทคนิคการทำเหมืองข้อมูลเพื่อจัดกลุ่มผู้มาใช้บริการที่มีคุณลักษณะคล้ายกัน ทำให้ผู้ให้บริการสามารถจัดบริการที่ตอบสนองต่อความต้องการของผู้ใช้บริการแต่ละกลุ่มได้ตรงกับความต้องการมากยิ่งขึ้นด้วย

จากที่กล่าวมาข้างต้น ผู้วิจัยจึงมีความสนใจที่จะพัฒนาค้นคลังข้อมูลและนำเทคนิคการทำเหมืองข้อมูลมาวิเคราะห์ข้อมูลผู้ใช้บริการและการค้นคืนสารสนเทศของผู้ใช้บริการ เพื่อค้นหาความรู้ใหม่โดยวิเคราะห์พฤติกรรมของผู้ใช้แต่ละกลุ่ม และจำแนกกลุ่มผู้ให้บริการภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ซึ่งความรู้ดังกล่าวสามารถนำไปใช้ประกอบในการพัฒนางานบริการที่มีประสิทธิภาพมากยิ่งขึ้น

2. วัตถุประสงค์การวิจัย

- 2.1 เพื่อสร้างคลังข้อมูลผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ
- 2.2 เพื่อวิเคราะห์และจำแนกกลุ่มผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ด้วยวิธีการทำเหมืองข้อมูลที่จัดเก็บในคลังข้อมูล
- 2.3 เพื่อวิเคราะห์และศึกษารูปแบบการยืมหนังสือของผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษด้วยวิธีการทำเหมืองข้อมูลที่จัดเก็บในคลังข้อมูล
- 2.4 เพื่อจัดทำรายงานหลายมิติตามความต้องการของผู้ใช้

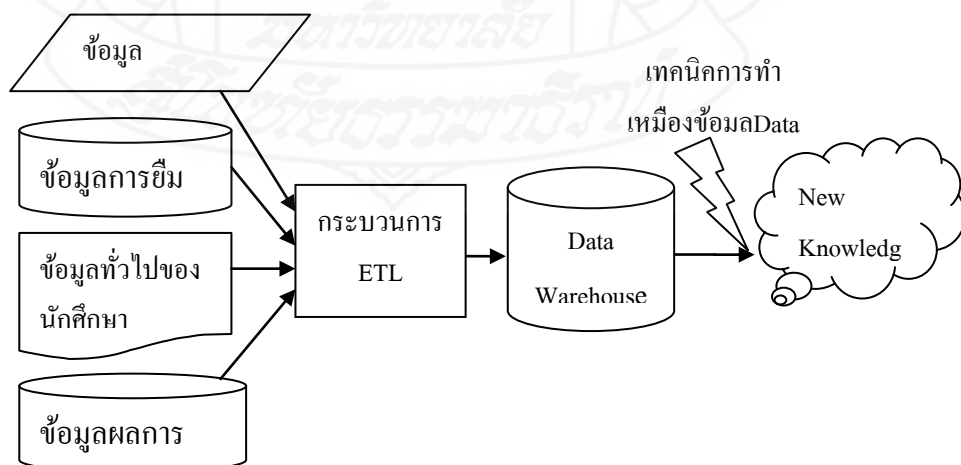
3. ขอบเขตของการวิจัย

คลังข้อมูลที่สร้างเป็นคลังข้อมูลที่ได้รวบรวมข้อมูลเกี่ยวกับผู้ใช้บริการห้องสมุดที่เป็นนักศึกษามหาวิทยาลัยราชภัฏศรีสะเกษ ได้แก่ ข้อมูลทั่วไปของนักศึกษา ข้อมูลการลงทะเบียนเรียน ข้อมูลผลการเรียนและข้อมูลการใช้บริการยืมคืนสารสนเทศของนักศึกษาที่เป็นสมาชิกห้องสมุดและมาใช้บริการยืมคืนสารสนเทศห้องสมุด โดยเก็บรวบรวมข้อมูลตั้งแต่วันที่ 1 มิถุนายน 2553 ถึง 31 ธันวาคม 2556

วิเคราะห์ข้อมูลของผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ โดยใช้แบบจำลองคริสป์-ดีเอ็ม (CRISP-DM) ซึ่งเป็นแบบจำลองการทำเหมืองข้อมูล เทคนิคการทำเหมืองข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วย การจัดกลุ่ม (Clustering) เพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ โดยใช้การวิเคราะห์ K-Mean Clustering กฎการหาความสัมพันธ์ (Association Rule) เพื่อวิเคราะห์ข้อมูลการยืมหนังสือ ว่าหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยด้วยกัน และไทม์ซีรีส์ (Time Series) เป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคต โดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาในอนาคต

4. กรอบแนวคิดการวิจัย

กรอบแนวคิดงานวิจัยประกอบด้วย ตัวแปรต้น คือข้อมูลนักศึกษา ได้แก่ ข้อมูลทั่วไปของนักศึกษา ข้อมูลการลงทะเบียนเรียน ข้อมูลผลการเรียนและข้อมูลการใช้บริการยืมคืนสารสนเทศของนักศึกษาที่ใช้บริการยืมคืนสารสนเทศห้องสมุดระหว่างปีการศึกษา 2553 จนถึงปีการศึกษา 2556 มาพัฒนาเป็นคลังข้อมูล (Data Warehouse) และตัวแปรตาม คือ ความรู้ที่ต้องการทราบเกี่ยวกับการยืมคืนหนังสือของนักศึกษา โดยใช้เทคนิคการทำเหมืองข้อมูลมาเป็นเครื่องมือค้นหาความรู้ดังกล่าว



ภาพที่ 1.1 กรอบแนวคิดการวิจัย

5. นิยามศัพท์เฉพาะ

5.1 **ทรัพยากรสารสนเทศ** หมายถึง หนังสือ วารสาร เอกสารวิชาการ สื่อต่างๆ ที่มีให้บริการภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

5.2 **การทำเหมืองข้อมูล (Data Mining)** หมายถึง กระบวนการวิเคราะห์ข้อมูลที่มีปริมาณมหาศาลเพื่อค้นหาความรู้ที่ไม่เคยทราบมาก่อน โดยใช้อัลกอริทึมที่พัฒนามาจากศาสตร์ด้านสถิติ และด้านปัญญาประดิษฐ์

5.3 **การบริการยืมคืน** หมายถึง การบริการยืมหรือคืนทรัพยากรสารสนเทศด้วยระบบคอมพิวเตอร์ให้แก่สมาชิกห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

5.4 **ผู้ใช้ห้องสมุด** หมายถึง นักศึกษามหาวิทยาลัยราชภัฏศรีสะเกษที่ใช้บริการยืมคืนสารสนเทศห้องสมุด

6. ประโยชน์ที่คาดว่าจะได้รับ

6.1 ได้คลังข้อมูลการบริการยืมสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

6.2 สามารถจำแนกคุณลักษณะของกลุ่มผู้ใช้บริการยืมสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

6.3 ทราบรูปแบบการยืมหนังสือของผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ซึ่งผลการวิเคราะห์สามารถนำไปประกอบการวางแผนการให้บริการต่อไป

6.4 สามารถเรียกดูรายงานได้ในรูปแบบหลายมิติทำให้ทราบผลการดำเนินงานของการบริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

บทที่ 2 วรรณกรรมที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาค้นคว้าข้อมูลเกี่ยวกับทฤษฎีและงานวิจัยที่เกี่ยวข้องจากเอกสาร บทความวิชาการและแหล่งสารสนเทศอินเทอร์เน็ต ในหัวข้อดังต่อไปนี้

1. การบริการห้องสมุดของมหาวิทยาลัยราชภัฏศรีสะเกษ
2. คลังข้อมูลและการพัฒนาคลังข้อมูล
3. การทำเหมืองข้อมูล
4. ซอฟต์แวร์พัฒนาเหมืองข้อมูล
5. งานวิจัยที่เกี่ยวข้อง

1. การบริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

1.1 ความเป็นมาของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

ห้องสมุดสถาบันราชภัฏศรีสะเกษ เดิมเรียกว่า “ห้องสมุดโครงการจัดตั้งสถาบันราชภัฏศรีสะเกษ” เปิดให้บริการครั้งแรกเมื่อวันที่ 10 มิถุนายน 2542 ณ อาคารโรงรถชั่วคราว และได้ย้ายอาคาร หลายครั้ง ดังนี้ เดือนตุลาคม พ.ศ. 2542 ย้ายมาที่อาคารเรียนชั่วคราว 1 (ปัจจุบันห้องประชุม 103) เดือนเมษายน พ.ศ. 2545 – พ.ศ. 2548 ย้ายมาประจำอยู่ที่อาคารหอสมุดซึ่งเป็นอาคารเอกเทศชั้นเดียว อยู่ติดกับอาคารสำนักงานอธิการบดี เดือนพฤษภาคม พ.ศ. 2548 ย้ายมาประจำอยู่ ณ อาคารบรรณราชนครินทร์ โดยได้เปิดให้บริการเพียงชั้นเดียว คือบริเวณชั้น 2 ขณะนั้นเรียกว่า “สำนักวิทยบริการและเทคโนโลยีสารสนเทศ” ซึ่งจัดเป็นสัดส่วนตามงานต่างๆ ดังนี้ ประตูทางเข้าออก เคาน์เตอร์ให้บริการยืมและคืน ทรัพยากรสารสนเทศ มุมวารสาร มุมสืบค้นข้อมูล ซึ่งมีเครื่องให้บริการจำนวน 3 เครื่อง มุมที่จัดเรียงหนังสือ ตั้งแต่หมวด 000 - 900 ไปจนถึง หนังสืออ้างอิง / วิจัย / วิทยานิพนธ์ / นวนิยาย / เรื่องสั้น และหนังสือสำหรับเด็ก และมีห้องให้บริการอินเทอร์เน็ต จำนวน 67 เครื่อง ห้องโสตทัศนศึกษา มีเครื่องให้บริการสื่อโสต จำนวน 20 เครื่อง

ต่อมาในปี 2551 มหาวิทยาลัยราชภัฏได้มีการเปลี่ยนโครงสร้างการบริการงานใหม่ และได้เปลี่ยนชื่อเดิมจาก “สำนักวิทยบริการและเทคโนโลยีสารสนเทศ” เป็น “ฝ่ายวิทยบริการและสารสนเทศ” (อยู่ภายใต้งานบริการการศึกษา จนถึงปัจจุบัน) วันที่ 16 พฤษภาคม 2553 ได้ขยายห้องสมุดขึ้นบริเวณชั้น 3 อาคารเดียวกันคืออาคารบรรณราช - นครินทร์ และได้มีการเปลี่ยนแปลงห้องต่างๆ ภายในชั้น 2 ดังนี้ ห้องโสตทัศนศึกษา ได้ตัดแปลงเป็นห้องสำหรับเด็ก ห้องอินเทอร์เน็ตตัดแปลงเป็นห้องสำหรับวารสาร และมุมวารสารเดิมได้ปรับเปลี่ยนให้เป็นชั้นวางเอกสารตำราสำหรับ Text book ส่วนบริเวณชั้น 3 ได้แบ่งสัดส่วน ให้มีห้องศูนย์ศรีสะเกษศึกษา

ห้องเอกสารตำราวิทยานิพนธ์ / วิจัย และหนังสืออ้างอิง ห้องค้นคว้าสำหรับกลุ่ม มุมให้บริการโซน Note book ห้องชมภาพยนตร์ (มินิเธียร์เตอร์) จัดมุมสำหรับให้บริการอินเทอร์เน็ต กว่า 160 เครื่อง จนถึงปัจจุบัน ฝ่ายวิทยบริการและสารสนเทศ มหาวิทยาลัยราชภัฏศรีสะเกษ (2555: 3)

1.2 ปรัชญา วิสัยทัศน์ พันธกิจ เป้าประสงค์ และประเด็นยุทธศาสตร์

ปรัชญาของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ คือ สารสนเทศเป็นเลิศ เปิดโลกการเรียนรู้ ควบคู่การบริการ วิสัยทัศน์คือ เป็นแหล่งเรียนรู้ที่สำคัญของสังคม มีทรัพยากรสารสนเทศที่หลากหลายและเทคโนโลยีที่ทันสมัย ก้าวไปสู่องค์กรแห่งการเรียนรู้ และสร้างคุณค่าด้วยการบริการที่มีคุณภาพ นอกจากนี้แล้วห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ยังได้กล่าวเกี่ยวกับพันธกิจ คือ ให้บริการผู้ใช้ทุกประเภท ด้วยทรัพยากรสารสนเทศที่หลากหลายและทันสมัย เพื่อตอบสนองภารกิจของมหาวิทยาลัยได้อย่างมีประสิทธิภาพ โดยมีเป้าประสงค์คือ (1) ทรัพยากรสารสนเทศมีเพียงพอต่อการเรียนการสอนทั้งในด้านปริมาณและคุณภาพ (2) นักศึกษาและบุคลากรมีทักษะการรู้สารสนเทศ (3) ให้บริการคอมพิวเตอร์ที่เพียงพอับความต้องการของผู้ใช้บริการ (4) ประชาชนในจังหวัดศรีสะเกษมีความรู้ด้านเทคโนโลยีสารสนเทศ เพื่อการประกอบอาชีพ (5) บุคลากรมีความรู้ด้านเทคโนโลยีสารสนเทศมากขึ้น (6) ผู้ใช้บริการได้รับความพึงพอใจสูงสุด (7) ห้องสมุดโรงเรียนในจังหวัดศรีสะเกษได้รับการพัฒนาให้เป็นระบบมากขึ้น (8) ห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษเป็นแหล่งการเรียนรู้ของคนทุกเพศทุกวัย (9) มีระบบบริหาร จัดการภายในองค์กรที่มีประสิทธิภาพและได้รับการพัฒนาอย่างต่อเนื่อง (10) มีการประกันคุณภาพประจำปี (11) ประสิทธิภาพในการดำเนินงานเป็นไปด้วยความคล่องตัว และมีประเด็นยุทธศาสตร์ คือ (1) มีจำนวนทรัพยากรสารสนเทศที่เพียงพอต่อการเรียนการสอน (2) ส่งเสริมทักษะการเรียนรู้สารสนเทศให้แก่ นักศึกษาและบุคลากร (3) การพัฒนาระบบเทคโนโลยีสารสนเทศ (4) การให้บริการเพื่อให้เกิดความพึงพอใจสูงสุดแก่ผู้ใช้บริการ (5) การให้บริการวิชาการแก่ชุมชนและท้องถิ่น (6) การพัฒนาห้องสมุดให้เป็นแหล่งเรียนรู้ตลอดชีวิต (7) การพัฒนาการบริหารและการจัดการองค์กร (8) การพัฒนาระบบประกันคุณภาพ (ฝ่ายวิทยบริการและสารสนเทศ มหาวิทยาลัยราชภัฏศรีสะเกษ, 2555: 2)

จากที่ได้กล่าวมาสรุปได้ว่า เพื่อให้การบริการของห้องสมุด มหาวิทยาลัยราชภัฏศรีสะเกษ เป็นไปตามปรัชญาว่าสารสนเทศเป็นเลิศ เปิดโลกการเรียนรู้ ควบคู่การบริการ ห้องสมุดจึงได้มีการปรับปรุงและพัฒนาการให้บริการเพื่อให้ตอบสนองต่อความต้องการของผู้ใช้บริการ ผู้วิจัยจึงได้ศึกษาถึงความพึงพอใจในใช้บริการห้องสมุด ของผู้ใช้บริการห้องสมุด ได้แก่ นักศึกษา อาจารย์ และบุคลากรของมหาวิทยาลัย ซึ่งเป็นกลุ่มเป้าหมายสำคัญในการใช้บริการของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ผลการศึกษาที่ได้จะเป็นแนวทางในการพัฒนา ในด้านสถานที่และสิ่งอำนวยความสะดวก ด้านการให้บริการ ด้านบุคลากร ด้านทรัพยากรสารสนเทศ ด้านระเบียบกฎเกณฑ์และการประชาสัมพันธ์ ของห้องสมุดให้สอดคล้องกับความต้องการของผู้ใช้บริการห้องสมุด ต่อไป

1.3 การให้บริการของห้องสมุด มหาวิทยาลัยราชภัฏศรีสะเกษ

ห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ มีให้บริการทั้งหมด 10 บริการ ดังนี้

1.3.1 บริการยืม - คืนทรัพยากรสารสนเทศ เป็นการให้บริการยืม - คืนทรัพยากรสารสนเทศประเภทต่างๆ แก่ผู้ใช้บริการให้เกิดประโยชน์สูงสุดได้แก่ หนังสือทั่วไปภาษาไทย ภาษาต่างประเทศ รายงานการวิจัย/วิทยานิพนธ์ และหนังสืออ้างอิง ซึ่งหนังสืออ้างอิงให้ยืมถ่ายเอกสารได้เท่านั้น โดยมีผู้ให้บริการ 2 คน คือ (1) นางสุนีย์ หุยวัน (2) นางสาวสุกัญญา ไชยพิมพ์ เป็นผู้ให้บริการ

1.3.2 บริการตอบคำถามและช่วยค้นคว้า เป็นบริการที่จัดขึ้นเพื่อให้ความช่วยเหลือแก่ผู้ใช้บริการในการสืบค้นสารสนเทศและการใช้บริการของห้องสมุด รวมถึงการให้ข้อมูลข่าวสารต่างๆ โดยมีบรรณารักษ์คอยตอบคำถามและให้คำแนะนำ โดยมีผู้ให้บริการ 3 คน คือ (1) นางสุนีย์ หุยวัน (2) นางสาวสุกัญญา ไชยพิมพ์ (3) นางสาววรรณิตา ดวงมณี เป็นผู้ให้บริการ

1.3.3 บริการวารสารและสิ่งพิมพ์ต่อเนื่อง เป็นการให้บริการวารสาร (Journals) หรือนิตยสาร (Magazine) สิ่งพิมพ์ที่มีเนื้อหาทันสมัยทันเหตุการณ์ จัดพิมพ์อย่างต่อเนื่อง มีกำหนดออกแน่นอนเป็นวาระต่างๆ เช่น รายสัปดาห์ รายปักษ์ รายเดือน ราย 2 เดือน ราย 3 เดือน เป็นต้น วารสารแต่ละฉบับประกอบด้วยบทความ ข้อเขียนทางวิชาการข่าว เกร็ดความรู้ จดหมาย ฯลฯ มีผู้เขียนหลายคน ประกอบด้วย วารสารภาษาไทย และวารสารภาษาต่างประเทศ แยกเป็นวารสารวิชาการ วารสารกึ่งวิชาการ และนิตยสาร หนังสือพิมพ์ (Newspaper) คือสิ่งพิมพ์ที่ออกรายวัน บางฉบับอาจออกเป็น รายสัปดาห์หรือรายอื่นๆ มุ่งเน้นเสนอข่าวที่ทันต่อเหตุการณ์ เสนอข่าวทุกชนิดที่เป็นข่าวสดที่น่าสนใจทั้งภายใน และภายนอกประเทศ ประกอบด้วย หนังสือพิมพ์ภาษาไทย และหนังสือพิมพ์ภาษาต่างประเทศ โดยมี นางสาวสุกัญญา ไชยพิมพ์ เป็นผู้ให้บริการ

1.3.4 บริการถ่ายเอกสาร เป็นบริการที่อำนวยความสะดวกให้แก่ผู้ใช้บริการเพื่อสะดวกในการค้นคว้าและการถ่ายสำเนาเอกสาร เพื่อลดปัญหาการฉีกทำลายสิ่งพิมพ์ของห้องสมุด ผู้ใช้บริการที่ต้องการถ่ายสำเนาเอกสาร จะต้องเสียค่าใช้จ่ายในการสำเนาเอกสารเองทั้งหมด โดยมีผู้ให้บริการ 2 คน คือ (1) นางสุนีย์ หุยวัน และ (2) นางสาวสุกัญญา ไชยพิมพ์ เป็นผู้ให้บริการ

1.3.5 บริการอินเทอร์เน็ต ห้องสมุดได้จัดสำหรับบริการอินเทอร์เน็ตที่มีคอมพิวเตอร์ – เเตอร์ สามารถเชื่อมต่อเครือข่ายอินเทอร์เน็ตความเร็วสูงกว่า 160 เครื่อง ไว้ให้บริการแก่ นักศึกษา อาจารย์ และบุคลากรของมหาวิทยาลัยในการสืบค้นหาข้อมูล โดยมี นายธีรพงศ์ สงผัด เป็นผู้ให้บริการ

1.3.6 บริการโสตทัศนศึกษาและสื่อการสอน ห้องสมุดได้จัดให้บริการโสตทัศนศึกษาประเภทต่างๆ ได้แก่ ซีดีรอมฝึกทักษะด้านภาษาอังกฤษ วีดิทัศน์ สื่อการเรียนการสอน แผ่นซีดี เพลง หรือภาพยนตร์ เพื่อให้ยืมภายในห้องสมุดเท่านั้น โดยมี นายอนุกุล เสียงหวาน เป็นผู้ให้บริการ

1.3.7 บริการสืบค้นทรัพยากรสารสนเทศ (OPAC) เป็นบริการสืบค้นฐานข้อมูลทรัพยากรสารสนเทศ ฐานข้อมูลทรัพยากรสารสนเทศ หรือเรียกว่า Web Opac ซึ่งเป็นระบบสืบค้นสารสนเทศออนไลน์ เพื่อความสะดวกในการสืบค้นทรัพยากรสารสนเทศทุกประเภทที่มีให้บริการ

ภายในห้องสมุด ก่อนจะทำการหยิบหนังสือจากชั้นหนังสือเพื่อความสะดวกในการหาหนังสือบนชั้น และสามารถเข้าสืบค้นได้ที่ <http://lib.sskru.ac.th>

1.3.8 บริการห้องประชุมกลุ่มย่อย เป็นห้องค้นคว้าส่วนบุคคลเรียนรู้รายบุคคล และรายกลุ่ม เป็นบริการห้องสำหรับประชุม ทำรายงานหรือค้นคว้าวิจัยเป็นรายบุคคล และกลุ่มตั้งแต่ 3 - 6 คน สำหรับนักศึกษา อาจารย์บุคลากรมหาวิทยาลัยมหาวิทยาลัย สถานที่ติดต่อขอใช้บริการ บริเวณชั้น 3 ของห้องสมุด ผู้ใช้บริการยื่นบัตรสมาชิกเพื่อทำการขอใช้บริการ เจ้าหน้าที่ที่ให้บริการ จะทำการเก็บบัตรสมาชิกไว้และมอบกุญแจเพื่อให้ผู้ใช้บริการใช้เปิดห้อง ผู้ใช้บริการสามารถเข้าใช้ห้อง ได้ตั้งแต่เวลา 08.00 น. ถึง 20.30 น. ในวันจันทร์ – ศุกร์ 08.00 น. – 17.00 น. ในวันเสาร์ – อาทิตย์ โดยมี นายธีรพงศ์ สงฆ์ เป็นผู้ให้บริการ

1.3.9 บริการห้องชมภาพยนตร์ การให้บริการห้องชมภาพยนตร์จะให้บริการอยู่ 2 แบบ คือ แบบที่ 1 รวมกลุ่ม 4 คน เลือกสื่อโสตฯ ที่ต้องการบนชั้นที่ให้บริการ ความสะดวก ประสงค์ใช้บริการที่เคาน์เตอร์ให้บริการ พร้อมบัตรประจำนักศึกษา สามารถเข้าชมได้เลย โดยมี เจ้าหน้าที่คอยอำนวยความสะดวกในการเปิด – ปิด โดยมี นายอนุกุล เสียงหวาน เป็นผู้ให้บริการ

1.3.10 บริการห้องสำหรับเด็ก ห้องสำหรับเด็กเหมาะสำหรับผู้ปกครองที่มาเรียน นำบุตรหลานมาด้วยจะนำบุตรหลานมาฝากไว้ เมื่อเรียนเสร็จก็จะมารับกลับสร้างความสะดวกสบายให้กับผู้ปกครอง ในห้องสำหรับเด็กมีให้บริการ 3 บริการดังนี้

- 1) มีบริการที่วีดีโอตลอดทั้งวัน
- 2) มีหนังสือสำหรับเด็ก เช่น นิทานและหนังสือที่เหมาะสมกับเยาวชนทั่วไป
- 3) มีโต๊ะอ่าน และลูกฟูกรองนั่งที่เป็นแผ่นยางรอง สำหรับนั่ง หรือนอนเล่น โดยมี นางสุนีย์ หุยวัน เป็นผู้ให้บริการ

1.4 ระเบียบการใช้ห้องสมุด

1.4.1 การเปิด - ปิด ห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ได้กำหนดระเบียบกฎเกณฑ์ ดังนี้

ระหว่างภาคเรียนที่มีการจัดการเรียน - สอน ตามปฏิทินการศึกษา

วันจันทร์ - ศุกร์ ให้บริการเวลา 08.00 – 20.30 น.

วันเสาร์ - อาทิตย์ ให้บริการเวลา 08.00 – 17.00 น.

ระหว่างปิดภาคเรียน ตามปฏิทินการศึกษา

วันจันทร์ - อาทิตย์ ให้บริการเวลา 08.30 – 17.00 น.

วันนักขัตฤกษ์และวันหยุดตามประกาศมหาวิทยาลัย ปิดให้บริการ

1.4.2 การสมัครเป็นสมาชิก

1) **นักศึกษา และบุคลากรของมหาวิทยาลัยราชภัฏศรีสะเกษ** เป็นสมาชิกห้องสมุดโดยอัตโนมัติ

2) **บุคคลภายนอก** สมัครเป็นสมาชิกจะต้องชำระเงินค่าสมัคร 100 บาท/ปี นักเรียนหรือนักศึกษา ชำระเงินค่าสมัคร 50 บาท/ปี โดยแนบหลักฐาน รูปถ่ายขนาด 1 นิ้ว 2 รูป และสำเนาบัตรประจำตัวประชาชน 1 ใบ

3) อาจารย์พิเศษของมหาวิทยาลัยราชภัฏศรีสะเกษ สมัครเป็นสมาชิกได้ โดยไม่เสียค่าใช้จ่าย เพียงแต่จะต้องแนบเอกสารเป็นใบรับรองสถานภาพการเป็นอาจารย์พิเศษของมหาวิทยาลัยราชภัฏศรีสะเกษ จากคณบดี

1.4.3 สิทธิในการยืม

1) นักศึกษาระดับปริญญาตรี มีสิทธิยืมหนังสือได้ คนละไม่เกิน 5 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน และยืมวารสารล่วงหน้าได้คนละ 1 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน

2) นักศึกษาระดับปริญญาโท ป.บัณฑิต มีสิทธิยืมหนังสือได้ คนละไม่เกิน 7 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน และยืมวารสารล่วงหน้าได้คนละไม่เกิน 3 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน

3) เจ้าหน้าที่ของมหาวิทยาลัยราชภัฏศรีสะเกษ มีสิทธิยืมหนังสือได้ คนละไม่เกิน 5 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 14 วัน และยืมวารสารล่วงหน้าได้คนละไม่เกิน 3 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน

4) อาจารย์ และอาจารย์พิเศษ ของมหาวิทยาลัยราชภัฏศรีสะเกษ มีสิทธิยืมหนังสือได้ คนละไม่เกิน 15 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 1 ภาคการศึกษา และยืมวารสารล่วงหน้าได้คนละไม่เกิน 3 เล่ม/ครั้ง ระยะเวลาส่งคืนภายใน 7 วัน

5) บุคคลภายนอก ใช้สิทธิในการยืมถ้าเอกสารได้เท่านั้น ยืมได้ 3 ต่อหนึ่งบัตรที่หน่วยงานราชการออกให้ ระยะเวลาการส่งคืนภายใน 1 วัน

1.4.4 สิทธิในการใช้บริการอินเทอร์เน็ต

1) นักศึกษาทุกคน สามารถใช้อินเทอร์เน็ตได้โดยตรงคำว่า Username ให้ใส่ Student password ให้ใส่ รหัสนักศึกษาจำนวน 12 หลัก จึงจะสามารถเข้าใช้อินเทอร์เน็ตได้

2) บุคคลภายนอก ที่สมัครเป็นสมาชิกจะต้องไปขอ Password ได้ที่ศูนย์คอมพิวเตอร์ในการเข้าใช้บริการอินเทอร์เน็ตตลอดการเป็นสมาชิก

1.4.5 ระเบียบการยืม - คืน

1) สมาชิกต้องใช้บัตรสมาชิกของตนเองในการขอรับบริการเท่านั้น
2) สมาชิกทุกประเภทหากส่งทรัพย์สินยืมขาดหาย ชำรุดเสียหาย ค่าปรับเล่มละ 5 บาท/วัน/เล่ม

3) สมาชิกสามารถยืมหนังสือเรื่องเดียวกัน ติดต่อกันได้ไม่เกิน 3 ครั้ง

1.4.6 ระเบียบการใช้ห้องสมุดทั่วไป

1) สมาชิกทุกคนต้องแสดงบัตรห้องสมุดก่อนเข้าใช้บริการทุกครั้ง มิฉะนั้นจะต้องเสียเงินค่าเข้าใช้บริการ นักเรียน/นักศึกษา 5 บาท/วัน บุคคลภายนอกที่เป็นผู้ใหญ่ 10 บาท/วัน

2) ผู้ใช้บริการห้องสมุดต้องแต่งกายสุภาพ (ไม่ใส่กางเกงขาสั้น เสื้อกล้าม เสื้อแขนกุด และเสื้อสายเดี่ยวมาใช้บริการ)

3) ผู้ใช้บริการห้องสมุด ต้องไม่ส่งเสียงดังรบกวนคนอื่น

4) ห้ามผู้ใช้บริการห้องสมุด ฉีกหรือทำลายทรัพย์สินของห้องสมุด หรือทรัพย์สินของห้องสมุด

5) ห้ามผู้ใช้บริการขีดเขียนข้อความใดๆ หรือก่อให้เกิดความสกปรกแก่เฟอร์นิเจอร์ของห้องสมุดทุกชนิด

6) ห้ามผู้ใช้บริการห้องสมุด นำอาหารและเครื่องดื่ม เข้ามารับประทานภายในห้องสมุด

7) ห้ามผู้ใช้บริการห้องสมุดสูบบุหรี่ ภายในห้องสมุด

8) ห้ามผู้ใช้บริการห้องสมุด เปิดและปิดเครื่องปรับอากาศด้วยตนเอง

9) ผู้ใช้บริการห้องสมุดต้องให้เจ้าหน้าที่ตรวจกระเป๋าและสัมภาระทุกครั้งก่อนออกจากห้องสมุด

10) ผู้ใช้บริการห้องสมุดจะต้องรักษาทรัพย์สินอันมีค่าและตรวจสอบทรัพย์สินของตนเองทุกครั้งก่อนออกจากห้องสมุด โดยห้องสมุดจะไม่รับผิดชอบจากการสูญหายไม่ว่ากรณีใดๆ ทั้งสิ้น

1.4.7 บทลงโทษ

1) สมาชิกที่ทำทรัพยากรสารสนเทศจากการยืมสูญหาย ต้องจัดซื้อตามรายการที่ยืมนั้นและส่งคืนห้องสมุด หรือชำระค่าปรับเป็นเงิน 2 เท่า ของราคาทรัพยากรสารสนเทศรายการนั้น

2) นักศึกษาที่ค้างส่งทรัพยากรสารสนเทศหรือค้างชำระหนี้ของห้องสมุด จะถูกชะลอการสำเร็จการศึกษาจนกว่าจะดำเนินการให้เรียบร้อย

3) บุคคลที่ขโมยหรือทำลายทรัพย์สินของห้องสมุด ต้องได้รับโทษทางวินัยของมหาวิทยาลัยราชภัฏศรีสะเกษ หรือ ได้รับโทษตามกฎหมาย

การให้บริการของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษเน้นผู้ใช้เป็นหลัก ดังนั้นห้องสมุดจึงต้องมีการศึกษาความพึงพอใจของผู้ใช้บริการเพื่อนำผลที่ได้มาปรับการให้บริการเพื่อตอบสนองต่อความต้องการของผู้ใช้บริการ



2. คลังข้อมูล (Data warehouse)

คลังข้อมูล เป็นระบบข้อมูลขนาดใหญ่ที่เอื้ออำนวยให้ผู้ใช้งานสามารถใช้ข้อมูลได้อย่างมีประสิทธิภาพ เป็น ที่เก็บรวบรวมข้อมูลจากหลายแหล่งที่มา หลายช่วงเวลา มีคุณลักษณะของข้อมูลแตกต่างกัน นำมาไว้รวม ณ ที่เดียวกันใช้เพื่อการวิเคราะห์ (กิตติพงษ์ กลมกล่อม, 2552: 5)

คลังข้อมูล (Data warehouse) เป็นการเก็บข้อมูลในเชิงหัวข้อ (Subject - oriented) ซึ่งทำการรวบรวมข้อมูลมาจากแหล่งต่างๆ (Integrated) โดยจะเก็บข้อมูลเป็นระยะเวลานาน (Time - variant) และ ไม่มีการเปลี่ยนแปลงของชุดข้อมูลได้โดยง่าย (Non - volatile) ซึ่งเป็นข้อมูลสำหรับที่ใช้ช่วยการสนับสนุนในการดำเนินการตัดสินใจเพื่อการบริหาร (Roiger and Geatz, 2003: 148)

2.1 วิวัฒนาการของคลังข้อมูล

จากความเป็นมาของคลังข้อมูล แสดงให้เห็นว่าคลังข้อมูลมีวิวัฒนาการมาจากฐานข้อมูลปฏิบัติการ (สุวรรณณี อัครกุลชัย, 2555: 4) ในที่นี้จะกล่าวถึงการกำเนิดของฐานข้อมูล ดังนี้

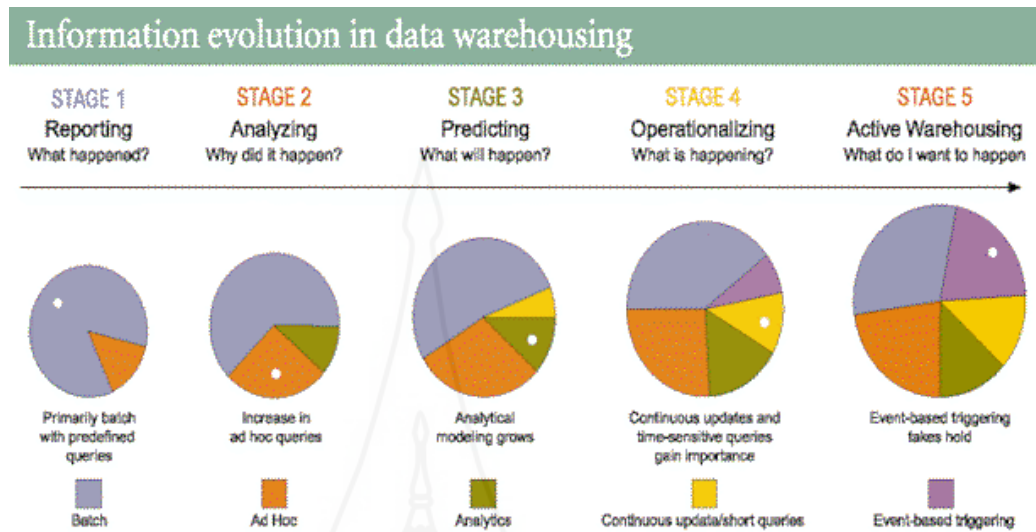
ปี พ.ศ. 2503 เริ่มมีการพัฒนาฐานข้อมูล มีการค้นคว้าและพัฒนาระบบฐานข้อมูลมาเรื่อยๆ เช่น ฐานข้อมูลแบบลำดับขั้น (Hierarchical database) ลักษณะคล้ายต้นไม้ที่คว่ำหัวลง ฐานข้อมูลแบบเครือข่าย (network database) และฐานข้อมูลเชิงสัมพันธ์ (relational database) ซึ่งมีเครื่องมือในการออกแบบฐานข้อมูลที่เรียกว่า อีอาร์ไดอะแกรม (Entity Relationship - ER - diagram) เป็นต้น มีเทคนิคการใช้อินเด็กซ์ (index) การบริหารข้อมูล และภาษาในการสืบค้นข้อมูล (query language)

ปี พ.ศ. 2513 คลังข้อมูลก่อกำเนิดที่สถาบันเทคโนโลยีแมสซาชูเซตส์หรือเอ็มไอที (Massachusetts Institute of Technology – MIT) โดยนักวิจัยพยายามแยกความแตกต่างระหว่างฐานข้อมูลปฏิบัติการกับกระบวนการวิเคราะห์ข้อมูล จึงได้มีการพัฒนาแนวทางของโครงสร้างโดยแบ่งออกเป็นชั้นๆ ให้จัดเก็บข้อมูลแยกต่างหากจากส่วนวิเคราะห์ ผลการศึกษาในเบื้องต้นมีอุปสรรคเนื่องจากความยุ่งยากในการประมวลผลที่จะต้องไปเรียกข้อมูลมาจากที่จัดเก็บหลายๆ ฐานข้อมูล จึงไม่ประสบความสำเร็จ

ปี พ.ศ. 2523 ฐานข้อมูลมีการพัฒนาให้มีศักยภาพมากขึ้น ได้มีการนำระบบจัดการฐานข้อมูลเชิงสัมพันธ์ (Relational Database Management System – RDBMS) เป็นแบบจำลองข้อมูลชนิดขั้นสูง (advanced data models) โดยพัฒนามาจากฐานข้อมูลเชิงสัมพันธ์ และได้นำหลักการที่นักวิจัยจากเอ็มไอทีคิดค้นในตอนต้นมาออกแบบส่วนของการเก็บข้อมูล (data capture) การเข้าถึงข้อมูล (data access) การกำหนดไดเรกทอรี (directory) และการให้บริการแก่ผู้ใช้ (user services) ซึ่งส่วนการนำเข้าข้อมูล นับเป็นส่วนที่เกี่ยวข้องกับฐานข้อมูลปฏิบัติการ ส่วนการเข้าถึงข้อมูล เป็นส่วนของกระบวนการวิเคราะห์ข้อมูล ซึ่งโครงสร้างนี้เริ่มเป็นการพัฒนาคลังข้อมูลขึ้นมาอีกระดับหนึ่ง

ปี พ.ศ. 2531 ถึงพ.ศ. 2534 ได้มีการนำเทคโนโลยีการทำคลังข้อมูล (Data warehousing) การทำเหมืองข้อมูล (Data mining) ฐานข้อมูลหลายมิติ (multidimensional databases) และฐานข้อมูลบนเว็บ (web databases) มาใช้ในธุรกิจ

ทั้งนี้ ถ้าพิจารณาเฉพาะส่วนของสารสนเทศในคลังข้อมูล พบว่ามีวิวัฒนาการในอีก รูปแบบหนึ่ง



ภาพที่ 2.1 วิวัฒนาการของสารสนเทศในคลังข้อมูล

ที่มา: <http://dssresources.com/papers/features/brobst&rarey01062003.html> Retrieved January 20, 2014

ยุคที่ 1 สารสนเทศในคลังข้อมูลที่รวบรวมไว้ ถูกนำไปใช้เพื่อแสดงรายงานเป็นระยะๆ ว่าเกิดอะไรขึ้น เช่น ในช่วงไตรมาสแรกของปี ได้รายได้เท่าไร เป็นต้น

ยุคที่ 2 เริ่มมีการวิเคราะห์สารสนเทศในคลังข้อมูลแบบทันทีทันใด สามารถตอบคำถามว่าทำไมถึงเกิดเหตุการณ์นั้นขึ้น โดยการใช้โอแลป (Online Analytical Processing – OLAP) มาช่วยวิเคราะห์

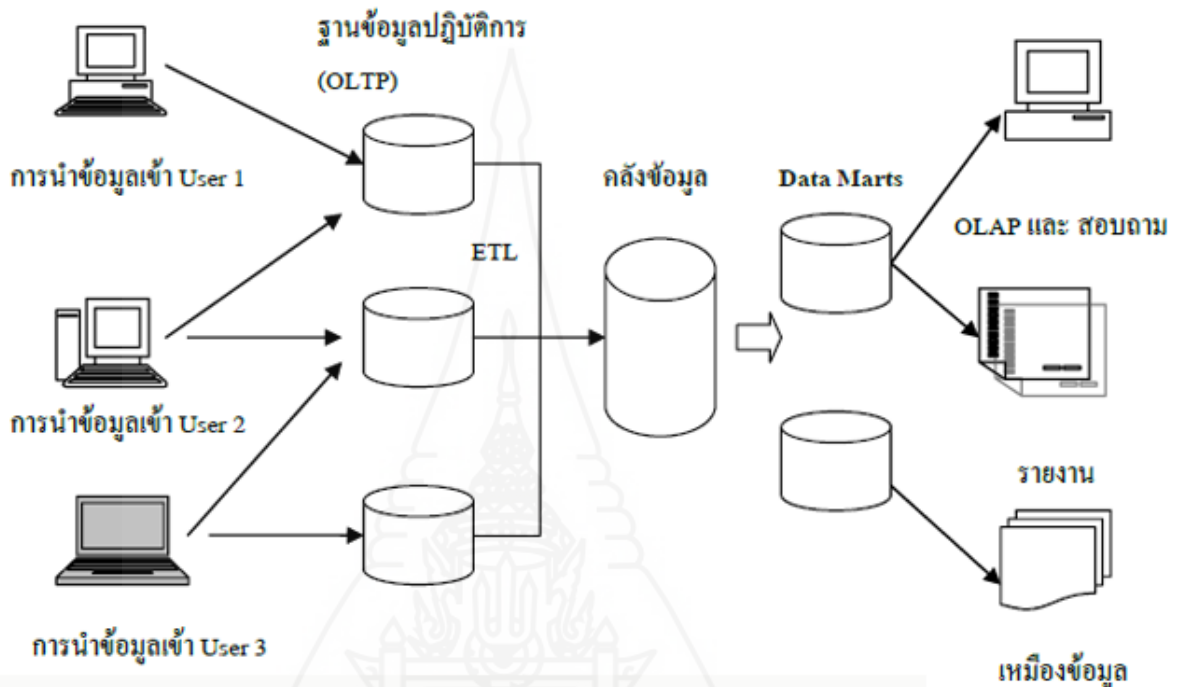
ยุคที่ 3 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์เพื่อการพยากรณ์โดยใช้หลักการของเหมืองข้อมูลมาช่วยวิเคราะห์ เช่น ต้องการสินค้าของลูกค้า เป็นต้น

ยุคที่ 4 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์มากขึ้นอย่างต่อเนื่อง เพื่อสามารถตอบได้ว่าจะเกิดอะไรขึ้นในอนาคตข้างหน้า

ยุคที่ 5 นำสารสนเทศในคลังข้อมูลมาใช้วิเคราะห์เพื่อกำหนดสิ่งที่จะเกิดขึ้นในอนาคต โดยผ่านระบบที่มีการติดต่อสื่อสารที่โต้ตอบกัน (Interactive) โดยตรง

2.2 กระบวนการในการพัฒนาคลังข้อมูล

เนื่องจากข้อมูลพื้นฐานของฐานข้อมูลในคลังข้อมูลประกอบด้วยข้อมูลจำนวนมาก จะต้องมีการออกแบบคลังข้อมูลเพื่อรวบรวมข้อมูลและวิเคราะห์หารูปแบบข้อมูลในคลังข้อมูล (สุวรรณณี อัครกุลชัย, 2555: 20)



ภาพที่ 2.2 กระบวนการในการพัฒนาคลังข้อมูล

ที่มา: สุวรรณณี อัครกุลชัย (2555: 20)

2.3 สถาปัตยกรรมคลังข้อมูล (Data Warehouse Architecture - DWA)

DWA เป็นโครงสร้างมาตรฐานที่ใช้อธิบาย เพื่อให้เข้าใจแนวคิด และกระบวนการของคลังข้อมูลนั้นๆ ซึ่งโดยทั่วไปแล้วคลังข้อมูลแต่ละระบบอาจจะมีรูปแบบที่ไม่เหมือนกันได้ เพื่อให้เหมาะสมกับองค์กรนั้นๆ (เลิศ เลิศศิริโสภณ, 2541: 95) ทั้งนี้ส่วนประกอบต่างๆ ภายใน DWA ที่สำคัญ ได้แก่

2.3.1 Operational database หรือ external database layer ทำหน้าที่จัดการกับข้อมูลในระบบงานปฏิบัติการหรือแหล่งข้อมูลภายนอกองค์กร

2.3.2 Information access layer เป็นส่วนที่ผู้ใช้ปลายทางติดต่อผ่านโดยตรง ประกอบด้วยฮาร์ดแวร์และซอฟต์แวร์ ที่ใช้ในการแสดงผลเพื่อการวิเคราะห์ โดยมีเครื่องมือช่วยเป็นตัวกลางที่ผู้ใช้ใช้ติดต่อกับคลังข้อมูล โดยในปัจจุบันเครื่องมือที่ได้รับความนิยมเพิ่มขึ้นอย่างรวดเร็ว นั่นคือ Online Analytical Processing Tool หรือ OLAP tool ซึ่งเป็นเครื่องมือที่มีความสามารถในการวิเคราะห์ที่ซับซ้อน และแสดงข้อมูลในรูปแบบหลายมิติ

2.3.3 Data access layer เป็นส่วนต่อประสานระหว่าง Information access layer กับ operational layer

2.3.4 Data director (metadata) layer เพื่อให้เข้าถึงข้อมูลได้ง่ายขึ้น และเป็นการเพิ่มความเร็วในการเรียกและดึงข้อมูลของคลังข้อมูล

2.3.5 Process management layer ทำหน้าที่จัดการกระบวนการทำงานทั้งหมด

2.3.6 Application messaging layer เป็นมิดเดิลแวร์ทำหน้าที่ในการส่งข้อมูลภายในองค์กรผ่านทางเครือข่าย

2.3.7 Data warehouse (physical) layer เป็นแหล่งเก็บข้อมูลของทั้ง Information data และ External data ในรูปแบบที่ง่ายแก่การเข้าถึงและยืดหยุ่นได้

2.3.8 Data staging layer เป็นกระบวนการแก้ไข และดึงข้อมูลจาก External database

2.4 เทคนิคในการสร้างคลังข้อมูล

การเคลื่อนที่ของข้อมูลในคลังข้อมูล ข้อมูลที่จะจัดเก็บภายในคลังข้อมูลมีการเคลื่อนที่ของข้อมูล (Information flow) 5 ประเภท (www.gotoknow.org/posts/452189) ดังนี้

2.4.1 in flow คือการนำข้อมูลจากฐานข้อมูลอื่นเข้าสู่คลังข้อมูลทั้งฐานข้อมูลภายในและภายนอกองค์กร โดยในขั้นนี้อาจมีการเปลี่ยนแปลงโครงสร้างข้อมูล การทำ demoralize การลบหรือเพิ่มฟิลด์เพื่อให้ข้อมูลทั้งหมดอยู่ในเนื้อหาที่สนใจเดียวกัน ในขั้นตอนนี้อาจใช้เครื่องมือที่เรียกว่า Data warehouse tool

2.4.2 up flow เมื่อข้อมูลที่เราต้องการอยู่ในคลังข้อมูลแล้วในบางครั้งอาจต้องมีการเพิ่มคุณค่าให้กับข้อมูลด้วยเพื่อให้ข้อมูลอยู่ในรูปแบบที่เป็นประโยชน์มากที่สุดต่อการนำเครื่องมือมาใช้ ซึ่งได้แก่การจัดกลุ่มข้อมูลหาค่าทางสถิติที่ซับซ้อน จัดข้อมูลให้อยู่ในรูปแบบหรือเทมเพลตมาตรฐาน

2.4.3 down flow เป็นขั้นตอนของการปรับปรุงเปลี่ยนแปลงข้อมูลที่เก่า และไม่อยู่ในเนื้อหาที่องค์กรสนใจออกไปจากคลังข้อมูลขององค์กร

2.4.4 out flow เป็นขั้นที่ผู้ใช้เรียกใช้ข้อมูลในคลังข้อมูลผ่านเครื่องมือต่างๆ โดยการเรียกใช้อาจมีเพียงขอเรียกเป็นครั้งคราวเป็นประจำทุกวัน/เดือน หรือแม้กระทั่งต้องการแบบทันที

2.4.5 Meta flow ข้อมูลที่จัดเก็บในคลังข้อมูลจะถูกทำข้อมูลไว้อีกชุดหนึ่ง เป็นแหล่งที่มาของข้อมูลนั้น หรือแม้กระทั่งที่อยู่ของข้อมูลนั้นในคลังข้อมูลและข้อมูลอื่นที่เกี่ยวข้อง

2.5 วิธีการออกแบบคลังข้อมูล

วิธีการนี้ถูกเสนอโดย Kimball ในปี 1996 เรียกว่าระเบียบวิธี 9 ขั้น หรือ Nine - Step Methodology (Connolly, 2002: 922) โดยวิธีการนี้เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่าดาต้ามาร์ท (Data mart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จแล้ว จึงนำมารวมกันเป็นคลังข้อมูลขององค์กรในขั้นสุดท้าย ซึ่งขั้นตอนทั้ง 9 ขั้นตอนนี้มีรายละเอียดดังนี้

2.5.1 กำหนดดาต้ามาร์ท คือการเลือกที่จะสร้างดาต้ามาร์ทของระบบงานใดบ้าง และระบบงานใดเป็นระบบงานแรกโดยองค์กรจะต้องสร้าง E - R model ที่รวมระบบงานทุกระบบขององค์กรไว้ แสดงการเชื่อมโยงของแต่ละระบบงานอย่างชัดเจน และสิ่งที่ต้องคำนึงถึงในการเลือกระบบงานที่จะเป็นดาต้ามาร์ทแรกนั้น มี 3 ปัจจัยที่เกี่ยวข้อง ได้แก่ จะต้องสามารถพัฒนาออกมาได้ทันตามเวลาที่ต้องการ โดยอยู่ในงบประมาณที่กำหนดไว้และต้องตอบปัญหาทางธุรกิจให้แก่องค์กรได้ ดังนั้นดาต้ามาร์ทแรกควรจะเป็นของระบบงานที่นำรายได้เข้ามาสู่องค์กร เช่น ระบบงานขาย เป็นต้น

2.5.2 กำหนด Fact Table ของดาต้ามาร์ท คือการกำหนดเนื้อหาหลักที่ควรจะเป็นของดาต้ามาร์ท โดยการเลือกเอนทิตีหลักและกระบวนการที่เกี่ยวกับเอนทิตีนั้นๆ ออกมาจาก E - R model ขององค์กร นั้นหมายถึงจะทำให้เราทราบถึง Dimension Table ที่ควรจะมีด้วย

2.5.3 กำหนดแอตทริบิวต์ที่จำเป็นของแต่ละ Dimension Table คือการกำหนดแอตทริบิวต์ที่บอกหรืออธิบายรายละเอียดของ dimension ได้ ทั้งนี้แอตทริบิวต์ที่เป็น primary key ควรเป็นค่าที่คำนวณได้ กรณีที่มีดาต้ามาร์ทมากกว่าหนึ่งดาต้ามาร์ทมี dimension เหมือนกัน นั้นหมายถึงว่าแอตทริบิวต์ใน dimension นั้นจะต้องเหมือนกันทุกประการ แต่นั่นก็ไม่อาจจะแก้ไขปัญหาการจัดเก็บข้อมูลซ้ำซ้อน อันจะนำมาสู่ความแตกต่างกันของข้อมูลชุดเดียวกัน ปัญหานี้จึงเป็นการดีที่จะมีการใช้ Dimension Table ร่วมกันในแต่ละ Fact Table ที่จำเป็นต้องมี dimension ดังกล่าว โดยเรียก Dimension Table ลักษณะแบบนี้ว่า conformed และเรียก Fact Table ว่า fact constellation เราสามารถกำหนดข้อดีของการใช้ Dimension Table ร่วมกันได้ดังนี้

- (1) แนใจได้ว่าในแต่ละรายงานจะออกมาสอดคล้องกัน
- (2) สามารถสร้างดาต้ามาร์ทในเวลาต่างๆ กันได้
- (3) สามารถเข้าถึงดาต้ามาร์ทโดยผู้พัฒนากลุ่มอื่นๆ
- (4) สามารถรวบรวมดาต้ามาร์ทหลายๆ อันเข้าด้วยกัน
- (5) สามารถออกแบบคลังข้อมูลร่วมกันได้

2.5.4 กำหนดแอตทริบิวต์ที่จำเป็นใน Fact Table โดยแอตทริบิวต์หลักใน Fact Table จะมาจาก primary key ในแต่ละ Dimension Table นอกจากนี้แล้ว ยังสามารถมีแอตทริบิวต์ที่จำเป็นอื่นๆ ประกอบอยู่ด้วย เช่นแอตทริบิวต์ที่ได้จากการคำนวณค่าเบื้องต้นที่จำเป็นสำหรับการคงอยู่ของแอตทริบิวต์อื่นใน Fact Table เรียกอีกอย่างหนึ่งว่า Measure การกำหนดแอตทริบิวต์นี้ไม่ควรจะเลือกแอตทริบิวต์ที่คำนวณค่าไม่ได้ เช่นเป็นตัวหนังสือหรือไม่ใช่ตัวเลข เป็นต้น และไม่ควรเลือกแอตทริบิวต์ที่ไม่เกี่ยวข้องกับเนื้อหาของ Fact Table ที่เราสนใจด้วย

2.5.5 จัดเก็บค่าการคำนวณเบื้องต้นใน Fact Table คือการจัดเก็บค่าที่ได้จากการคำนวณให้เป็นแอตทริบิวต์หนึ่งใน Fact Table ถึงแม้ว่าจะสามารถหาค่าได้จากแอตทริบิวต์อื่นๆ ก็ตาม ทั้งนี้เพื่อให้การสอบถามมีประสิทธิภาพมากขึ้น สามารถทำงานด้วยความเร็วที่เพิ่มขึ้น เนื่องจากไม่ต้องคำนวณค่าใหม่ทั้งหมด ถึงแม้ว่าจะเกิดความซ้ำซ้อนของข้อมูลในการจัดเก็บบ้างก็ตาม

2.5.6 เขียนคำอธิบายของ Dimension Table ทั้งนี้ก็เพื่อให้ผู้ใช้สามารถใช้งานดาต้ามาร์ทได้อย่างมีประสิทธิภาพเพราะเกิดความเข้าใจอย่างดีในส่วนต่างๆ

2.5.7 กำหนดระยะเวลาในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจจะเป็นการจัดเก็บเพียงช่วงระยะเวลา 1 - 2 ปี หรือนานกว่านั้น ขึ้นอยู่กับความต้องการขององค์กร เนื่องจากองค์กรแต่ละประเภทมีความต้องการในการจัดเก็บข้อมูลต่างช่วงเวลากัน ทั้งนี้ขึ้นอยู่กับความจำเป็นหรือข้อกำหนดในการดำเนินธุรกิจมีข้อสังเกตอยู่ 2 ประการที่น่าสนใจและสำคัญสำหรับการออกแบบแอตทริบิวต์ในเรื่องของการจัดเก็บข้อมูล ดังนี้

- 1) ข้อมูลที่ถูกจัดเก็บไว้นานเกินไปมักเกิดปัญหาการอ่าน หรือแปลข้อมูลนั้นๆ จากแฟ้มหรือเทปเก่า
- 2) เมื่อมีการนำรูปแบบเก่าของ Dimension Table มาใช้อาจเกิดปัญหาการเปลี่ยนแปลงของ dimension อย่างซ้ำๆ ได้

2.5.8 การติดตามปัญหาการเปลี่ยนแปลงของ dimension คือ การเปลี่ยนแอตทริบิวต์ของ Dimension Table เก่ามาใช้แล้วส่งผลกระทบต่อข้อมูลปัจจุบันของ Dimension Table โดยสามารถแบ่งประเภทของปัญหาที่เกิดขึ้นได้เป็น 3 ประเภท ดังนี้

- 1) เกิดการเขียนทับข้อมูลใหม่โดยข้อมูลเก่า
- 2) เกิดเรคอร์ดใหม่ๆ ขึ้นใน Dimension
- 3) เกิดเรคอร์ดที่มีทั้งค่าเก่าและใหม่ปนกันไป

2.5.9 กำหนดคิวรีเป็นการออกแบบด้านกายภาพ เพื่อให้ผู้ใช้เกิดความสะดวกในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพเมื่อดำเนินการทั้ง 9 ขั้นตอนสำหรับแต่ละดาต้ามาร์ทเสร็จแล้ว จึงจะนำทั้งหมดมารวมกันเป็นภาพของคลังข้อมูลขององค์กรต่อไป

2.6 การแปลงข้อมูลเข้าสู่ดาต้ามาร์ท

เมื่อเราออกแบบฐานข้อมูลสำหรับแต่ละดาต้ามาร์ทเสร็จแล้ว ขั้นตอนต่อไปที่สำคัญยิ่งก็คือการนำข้อมูลจากแหล่งข้อมูลไปแปลงให้อยู่ในแพลตฟอร์มของฐานข้อมูลที่ได้ออกแบบไว้ นั่นก็คือการแปลงข้อมูล หรือ Extraction Transformation and Loading (ETL) นั่นเอง โดยที่คุณภาพของการแปลงข้อมูลเป็นสิ่งที่สำคัญมากสำหรับการสร้างคลังข้อมูล ความซับซ้อนของการแปลงข้อมูลและโครงสร้างของข้อมูลจะแตกต่างกันไปตามคลังข้อมูลของแต่ละองค์กรต้องการ โดยที่การแปลงข้อมูล

หมายรวมถึงตั้งแต่การวิเคราะห์แหล่งข้อมูล กำหนดการส่งข้อมูลรวบรวมหรือสร้างข้อมูลภายนอก วางแผนและสร้างรูทีนของการแปลงข้อมูล และตรวจสอบความถูกต้องของข้อมูลที่ได้ (นงลักษณ์ พลอยปลื้ม, 2540: 126) สามารถสรุปเป็นขั้นตอนได้ดังนี้

2.6.1 วิเคราะห์แหล่งข้อมูล เช่น ปริมาณของข้อมูล จำนวนและชนิดของการเข้าถึงแหล่งข้อมูล แพลตฟอร์มและภาษาโปรแกรมที่ใช้ เป็นต้น

2.6.2 ย้ายข้อมูลที่ต้องการจากระบบเดิม มาไว้ในบริเวณที่ใช้ปรับแต่งข้อมูลหรือเรียกบริเวณนี้ว่า staging area เพื่อนำมาเลือกเฉพาะส่วนที่ต้องการแปลงข้อมูลและตรวจสอบความถูกต้อง หรือการทำความสะอาดข้อมูล

2.6.3 กำหนด primary key ของ Fact Table และ Dimension Table และกำหนด foreign key ระหว่าง Fact Table กับ Dimension Table

2.6.4 ย้ายข้อมูลที่ทำความสะอาดแล้ว จาก staging area ลงสู่เซิร์ฟเวอร์ของดาต้ามาร์ท

2.6.5 สร้าง metadata ของแต่ละดาต้ามาร์ท โดยเก็บรายละเอียดของข้อมูล การอัปเดตและส่งออกไว้ในดาต้ามาร์ท

2.6.6 ตรวจสอบความถูกต้องของข้อมูล ซึ่งจะต้องกระทำตลอดทั้งกระบวนการแปลงข้อมูลจะทำได้ดังนี้

- 1) ตรวจสอบผลรวมทั้งหมดของจำนวนข้อมูลที่ดึงมาจากแหล่งข้อมูลที่เพิ่มเข้าไป
- 2) ตรวจสอบแก้ไขข้อมูลในระบบเดิมของแหล่งข้อมูล ซึ่งควรจะเก็บข้อมูลในการตรวจแก้ไขไว้ใน metadata ของการแปลงข้อมูลด้วย
- 3) ตรวจสอบค่าของข้อมูลให้ถูกต้องในกระบวนการรวบรวมข้อมูล
- 4) ตรวจสอบผลรวมของข้อมูลหลังจากการย้ายข้อมูลลงสู่ดาต้ามาร์ทแล้ว

3. การทำเหมืองข้อมูล (Data mining)

3.1 ความหมาย และความสำคัญของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล คือ การค้นหาความรู้ในคลังข้อมูล (อังกฤษ: Knowledge Discovery in Databases - KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อมูลจำนวนมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และ การรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ (<http://th.wikipedia.org/wiki/การทำเหมืองข้อมูล>)

เหมืองข้อมูลเป็นเทคโนโลยีที่ช่วยให้การวิเคราะห์ข้อมูลทำได้โดยอัตโนมัติและมีประสิทธิภาพสูง ซึ่งส่งผลให้มีการนำเหมืองข้อมูลไปใช้อย่างแพร่หลายในทุกวงการ โดยเฉพาะในกรณีที่มีข้อมูลมีขนาดใหญ่มาก โดยระบบคอมพิวเตอร์จะทำหน้าที่ค้นหาแนวโน้มและลักษณะที่น่าสนใจต่างๆ ที่ปรากฏในข้อมูล ตลอดจนวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลนั้น ซึ่งจะช่วยให้สามารถพยากรณ์แนวโน้มของข้อมูลใหม่ที่จะเกิดขึ้น ในอนาคตได้ รวมถึงเข้าใจความสัมพันธ์ที่เชื่อมโยงข้อมูลแต่ละกลุ่มเข้าด้วยกัน (Nicholson, 2003: 247)

การทำเหมืองข้อมูล เป็นกระบวนการในการค้นหาความสัมพันธ์ใหม่ระหว่างข้อมูลด้วยกันรวมทั้งแพทเทิร์นและแนวโน้มโดยการกลั่นกรองข้อมูลปริมาณมหาศาลที่จัดเก็บอยู่ ด้วยการใช้เทคโนโลยีการรู้จำแพทเทิร์น ความรู้ทางสถิติและคณิตศาสตร์มาประยุกต์เข้าด้วยกัน (Gartner Group www.gartner.com อ้างใน Larose (2005) "Discovering Knowledge in Data: An Introduction to Data Mining")

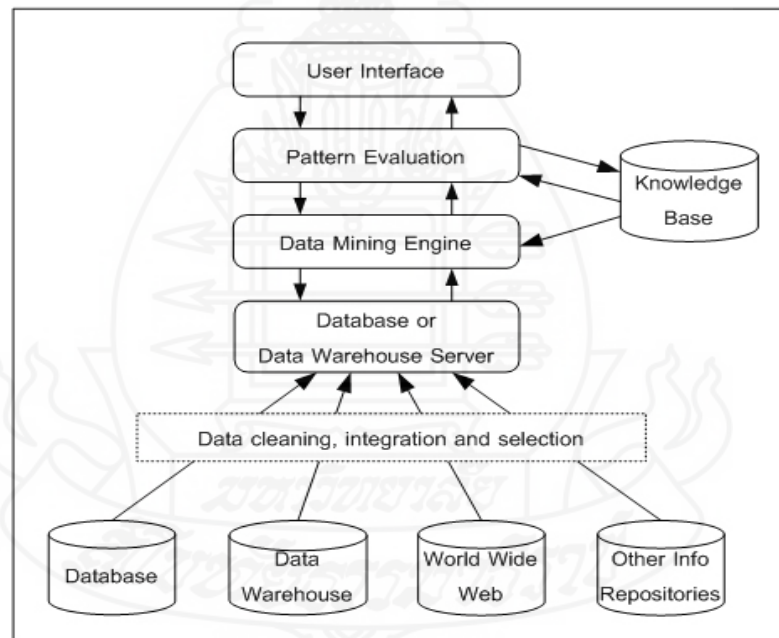
จากความหมายของการทำเหมืองข้อมูลที่กล่าวข้างต้นได้ว่า การทำเหมืองข้อมูลหมายถึงกระบวนการค้นหาความรู้ (knowledge) ซึ่งเป็นรูปแบบและความสัมพันธ์ที่น่าสนใจที่มีอยู่ในข้อมูลจากฐานข้อมูลขนาดใหญ่ที่ซับซ้อนจึงต้องอาศัยขั้นตอนวิธีทางคอมพิวเตอร์ผสมผสานกับวิธีการทางสถิติเพื่อสร้างตัวแบบ/แบบจำลอง (modeling) ที่อธิบายความเป็นไปหรือสภาพการณ์หนึ่งที่เกิดขึ้นแล้วหรือที่เราทราบคำตอบแล้วโดยต้องการนำแบบจำลองนี้มาใช้อธิบายสถานการณ์ที่ยังไม่

เกิดขึ้นหรือที่ยังไม่ทราบคำตอบซึ่งการสร้างแบบจำลองเหล่านี้มีกฎเกณฑ์รูปแบบในการดำเนินการ เพื่อให้ได้ความรู้แบบจำลองเหล่านี้อาจเป็นตัวแบบที่เรียบง่ายไปจนถึงที่ยุ่งยากซับซ้อนและมีการผสมผสานแนวคิดหรือเครื่องมือต่างๆ เข้าด้วยกันเพื่อที่จะสามารถสกัดข้อมูลความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ได้ โดยผู้วิเคราะห์ไม่จำเป็นต้องระบุรูปแบบจำลองด้วยตนเอง แต่เป็นการค้นหารูปแบบได้โดยอัตโนมัติ มีการพยากรณ์หรือคาดการณ์ผลลัพธ์จากข้อมูลให้ได้ ดังภาพที่ 2.3 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล



ภาพที่ 2.3 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล

3.2 สถาปัตยกรรมระบบเหมืองข้อมูล



ภาพที่ 2.4 สถาปัตยกรรมระบบเหมืองข้อมูล

ที่มา: <http://siripornk.blogspot.com/2010/08/data-mining.html> Retrieved January 20, 2014

3.2.1 แหล่งข้อมูลที่ใช้ทำเหมืองข้อมูล (data sources) เป็นแหล่งข้อมูลที่อาจจะจัดเก็บไว้อยู่ในฐานข้อมูล คลังข้อมูล เว็บไซต์ไวด์เว็บ (WWW) และคลังจัดเก็บข้อมูลอื่นๆ เนื้อหา

สาระภายในข้อมูลเป็นสิ่งที่ผู้ทำเหมืองข้อมูลมีความสนใจและต้องการค้นหาความรู้ที่ซ่อนเร้นอยู่ การดำเนินการเบื้องต้นเพื่อให้ข้อมูลเหล่านี้อยู่ในสภาพพร้อมใช้งาน นับตั้งแต่ การทำความสะอาดข้อมูล (data cleaning) เป็นการจัดการข้อมูลที่ไม่สมบูรณ์ รายละเอียดขาดหายไป เนื้อหาข้อมูลขัดแย้งกันเอง มาดำเนินการให้ข้อมูลใช้งานได้

1) *การบูรณาการข้อมูล (data integration)* การนำข้อมูลจากแหล่งต่างๆ ที่สัมพันธ์กันมาไว้ในที่เดียวกัน จำเป็นต้องปรับรูปแบบให้เป็นแบบเดียวกัน เพื่อให้ทำงานร่วมกันได้

2) *การคัดเลือกข้อมูล (data selection/extraction)* เป็นการนำข้อมูลที่ต้องการมาใช้งานมาจัดเก็บเพื่อเป็นแหล่งข้อมูล

3.2.2 คลังข้อมูล (data warehouse server) เป็นระบบแม่ข่ายที่ใช้จัดเก็บข้อมูล เพื่อรองรับข้อมูลทั้งหมดในการทำเหมืองข้อมูล

3.2.3 ฐานความรู้ (knowledge base) หมายถึง ความรู้ความเข้าใจในสิ่งหรือเรื่องที่ต้องการวิเคราะห์หรือค้นหา ความรู้ความเข้าใจเหล่านี้เกิดจากการเรียนรู้และประสบการณ์ในเรื่องดังกล่าว มีความเข้าใจในธรรมชาติของข้อมูลที่มีอยู่และความรู้ในการเลือกวิธีการทำเหมืองข้อมูลได้อย่างเหมาะสมกับลักษณะงานหรือข้อมูลที่ต้องการดำเนินการ

3.2.4 กลไกการทำเหมืองข้อมูล (data mining engine) เป็นวิธีการทำเหมืองข้อมูล ได้แก่ การกำหนดคุณสมบัติข้อมูล การวิเคราะห์ความสัมพันธ์ของข้อมูล การจำแนกหมวดหมู่ การจัดกลุ่ม การค้นหาความผิดปกติของข้อมูล และเครื่องมือที่ใช้ในการทำเหมืองข้อมูลแต่ละประเภท เช่น วิธีการทางสถิติ

3.2.5 ส่วนประเมินรูปแบบ (pattern evaluation module) ในการทำเหมืองข้อมูล สิ่งที่สำคัญคือ การสร้างแบบจำลองหรือตัวแบบ (model) เพื่อแสดงถึงสมมติฐานได้อย่างชัดเจนให้สามารถค้นพบความรู้จากแบบจำลองที่สร้างขึ้นมาได้ ทั้งนี้แบบจำลองที่สร้างขึ้นควรต้องมีความเหมาะสมในการพยากรณ์ได้ และเครื่องมือหรือกลไกเหล่านี้ ยังใช้ตรวจสอบและประเมินแบบจำลองที่จัดทำขึ้นมาด้วย เพื่อให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือ เหมาะสม และถูกต้องมากที่สุด

3.2.6 ส่วนที่ใช้ติดต่อกับผู้ใช้งาน (user interface) แบ่งเป็น

1) ส่วนรองรับการป้อนคำสั่งข่าของผู้ใช้งาน เช่น การค้นหา การใส่เงื่อนไขการทำเหมืองข้อมูล

2) ส่วนนำเสนอผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลในรูปแบบต่างๆ เช่น ตาราง แผนภูมิ และรูปแบบความสัมพันธ์ของข้อมูลที่ค้นหา

3.3 ประเภทการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นการค้นหารูปแบบที่เกิดขึ้นกับข้อมูลที่สนใจ ซึ่งแบ่งตามลักษณะงานที่ดำเนินการ (Data mining operation/functionality) เป็นประเภทต่างๆ ได้ดังนี้ (วิภา เจริญภัณฑารักษ์, 2555: 10)

3.3.1 การค้นหาคุณลักษณะ/รายละเอียดของข้อมูล เป็นการค้นหาคุณลักษณะเฉพาะของข้อมูล (class/concept description) ที่เราให้ความสนใจหรือต้องการดำเนินการ เพื่อเข้าใจในรายละเอียดที่เกี่ยวข้องกับข้อมูลเหล่านั้น การทำเหมืองข้อมูลประเภทนี้ แบ่งได้ 2 ลักษณะคือ

1) *การกำหนดคุณลักษณะเฉพาะข้อมูล (data characterization)* เป็นการค้นหาคุณลักษณะเฉพาะของข้อมูล เพื่ออธิบายรายละเอียดเจาะลึกของข้อมูลนั้น ทำให้เข้าใจภาพรวมของข้อมูลในเรื่องนั้น ตัวอย่างเช่น การค้นหาคุณลักษณะ/ประวัติของลูกค้า (customer profile) ที่ซื้อสินค้ากับบริษัท “STOU Electronics” มากกว่า \$1,000 ต่อปี โดยใช้คำสั่งเอสคิวแอล (SQL) เพื่อคัดกรองข้อมูลที่ต้องการในเบื้องต้น พบว่า ลูกค้ามีอายุระหว่าง 40 - 50 ปี มีงานทำ และมีประวัติ/เครดิต (credit rating) อยู่ในระดับดี เป็นต้น

2) *การกำหนดคุณลักษณะแตกต่าง (data discrimination)* เป็นการค้นหาและศึกษารายละเอียดของข้อมูลที่เราสนใจเปรียบเทียบกับข้อมูลเดิมที่มีคุณสมบัติที่แตกต่างหรือตรงกันข้าม ซึ่งการเปรียบเทียบข้อมูล มักแบ่งเป็น 2 กลุ่มคือ กลุ่มหลัก และกลุ่มที่แตกต่าง ทั้งนี้ผู้ทำเหมืองข้อมูลต้องเป็นผู้กำหนดเงื่อนไขเอง เช่น การศึกษาเปรียบเทียบรายละเอียดข้อมูลของกลุ่มลูกค้าที่ซื้อสินค้ากับบริษัท STOU Electronics มากกว่า 3 ครั้งต่อเดือน เทียบกับกลุ่มลูกค้าที่ซื้อสินค้าน้อยกว่า 3 ครั้งต่อปี เป็นต้น ผลที่ได้จากการค้นหารายละเอียดคือประวัติการซื้อสินค้าของลูกค้า เช่น กลุ่มที่ซื้อสินค้าประจำ จะเป็นผู้มีอายุระหว่าง 20 - 40 ปี จบการศึกษาระดับปริญญาตรี ขณะที่กลุ่มที่แตกต่าง จะมีประวัติของลูกค้าที่มีอายุต่ำกว่า 20 หรือ มากกว่า 60 ปี และจบการศึกษาในระดับที่ต่ำกว่าปริญญาตรี เป็นต้น ทั้งนี้อาจจะใช้ข้อมูลที่จัดเก็บในคลังข้อมูลมาดำเนินการโดยใช้คำสั่งเอสคิวแอล (SQL) ในการค้นหาตามเงื่อนไขที่ต้องการได้

เครื่องมือหรือวิธีการในการทำเหมืองข้อมูลแบบนี้ ใช้วิธีการทางสถิติ เช่น ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) ค่าฐานนิยม (mode) การหาค่าเบี่ยงเบนมาตรฐาน (standard deviation) การเขียนกราฟต่างๆ การหาความสัมพันธ์ระหว่างตัวแปรที่สนใจ และการเปรียบเทียบข้อมูลด้วยตารางหลายมิติ (multidimensional data cube) เป็นต้น การนำเสนอผลลัพธ์ที่ได้อาจจะเป็นแผนภูมิแบบต่างๆ เช่น แผนภูมิแท่ง เส้น และวงกลม เป็นต้น

3.3.2 การค้นหาความสัมพันธ์

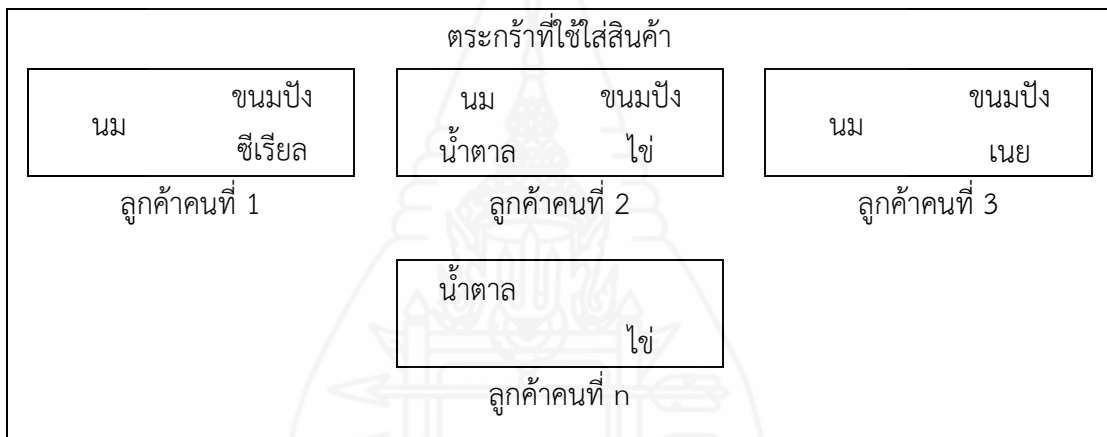
เป็นการทำเหมืองข้อมูลแบบการค้นหาความสัมพันธ์ (Frequent pattern and association) ประกอบด้วย การค้นหารูปแบบที่เกิดควบคู่กัน (frequent pattern) และสร้างความสัมพันธ์ที่เกิดขึ้น (association rule) มีรายละเอียดดังนี้

1) *การค้นหารูปแบบของข้อมูลที่เกิดขึ้นควบคู่กันในรายการเดียวกันเสมอ (frequent item set)* การทำเหมืองข้อมูลประเภทนี้นิยมใช้ในการวางแผนการตลาด การวางแผนการส่งซื้อสินค้าว่าควรส่งซื้อสินค้าใดร่วมกันบ้าง การจัดชั้นวางสินค้าอย่างเหมาะสม การจัดรายการส่งเสริมการขายสินค้า จึงมักนิยมเรียกการทำเหมืองข้อมูลประเภทนี้ว่า การวิเคราะห์ตะกร้าสินค้า (market basket analysis) เป็นการวิเคราะห์หรือค้นหาความสัมพันธ์ข้อมูลที่เกิดควบคู่กันในแต่ละรายการ (transaction) เช่น เมื่อซื้อนม มักจะซื้อ ขนมปัง ไข่ ร่วมกันด้วยเสมอ เป็นต้น

2) *การค้นหาความสัมพันธ์ (association)* หลังจากการค้นหารูปแบบที่เกิดขึ้นควบคู่กัน (Frequent Item Set) ได้แล้ว จะสร้างกฎความสัมพันธ์ของ โดยการกำหนดค่า “support” ซึ่งเป็นจำนวนรายการข้อมูลที่น่ามาดำเนินการ และค่า “confidence” เป็นค่าความน่าจะเป็นที่ใช้ในการวิเคราะห์ความสัมพันธ์ เช่น STOU Electronics ต้องการค้นหารายการข้อมูล (transaction) ในการซื้อสินค้าของลูกค้า ตามกฎความสัมพันธ์ต่อไปนี้ “Buys (X, ‘computer’) =>

buys (X, 'software') (support=1%, confidence=50%)” หมายความว่า 1% ของจำนวน ทรานแซกชันที่นำมาดำเนินการค้นหา (support=1%) พบว่าโอกาสหรือความน่าจะเป็นของลูกค้าที่ซื้อ “computer” แล้วจะซื้อ “software” มี 50% (confidence=50%) ผลที่ได้ทำให้ผู้วิเคราะห์มีความเข้าใจความเชื่อมโยงสัมพันธ์ของข้อมูลได้ดียิ่งขึ้น

ตัวอย่างของการค้นหากฎความสัมพันธ์ คือการวิเคราะห์พฤติกรรมของผู้บริโภค โดยการหาความสัมพันธ์ระหว่างสินค้าที่ต่างชนิดกัน แต่ลูกค้าจะเลือกซื้อไปพร้อมกัน ดังภาพที่ 2.5 โดยการค้นหาความสัมพันธ์นี้จะช่วยทำให้ทราบว่าสินค้าใดถูกลูกค้าเลือกซื้อไปคู่กันมากที่สุด และนำกฎความสัมพันธ์ที่ค้นพบนั้นมาใช้ในการปรับปรุงกลยุทธ์การขายสินค้าหรือใช้ประกอบการพิจารณาในการจัดวางสินค้า ทำให้เกิดความสะดวกในการเลือกซื้อสินค้า และเป็นการเพิ่มยอดขายให้กับร้านค้า



ภาพที่ 2.5 การวิเคราะห์พฤติกรรมการซื้อสินค้า

กฎความสัมพันธ์สามารถเขียนให้อยู่ในรูปไอเท็มเซตที่เป็นเหตุไปสู่ไอเท็มเซตที่เป็นผล เช่น ลูกค้าที่ซื้อนมส่วนใหญ่จะซื้อขนมปังด้วย ก็สามารถเขียนความสัมพันธ์ได้เป็น {นม} → {ขนมปัง} เป็นต้น

การอธิบายวิธีการค้นหาความสัมพันธ์จะต้องใช้นิยามดังต่อไปนี้

- 1) ไอเท็มเซต (Item sets: I) คือเซตที่มีไอเท็มทั้งหมดเป็นสมาชิก ซึ่งไอเท็มในที่นี้อาจเป็นชื่อของหน่วยใดๆ ที่นำมาใช้ในการเรียนรู้
- 2) ทรานแซกชัน (Transaction: T) เป็นเซตย่อยของไอเท็ม โดยที่ $T \subseteq I$
- 3) เซ็ตของข้อมูล (Data: D) คือเซตที่มีทรานแซกชันทุกตัวเป็นสมาชิก โดยที่ $D = \{t_1, t_2, \dots, t_3\}$
- 4) การกำหนดค่าสนับสนุน (Support) s_{min}' โดยที่ $0 \leq s_{min}' \leq 1$

- 5) การกำหนดค่าความเชื่อมั่น (Confidence) c_{min}' โดยที่ $0 \leq c_{min} \leq 1$
- 6) ทราจแซคชัน T บรรจุกเซตย่อยของไอเท็ม X ก็ต่อเมื่อ $X \subseteq t$
- 7) ค่าสนับสนุนของไอเท็มเซต X คือ $s(X) = \frac{|\{t \in D \mid X \subseteq t\}|}{|D|}$
- 8) กฎที่ได้อยู่ในรูปแบบ $L \rightarrow R$ เมื่อ $L \subset I, R \subset I$ และ $L \cap R = \emptyset$
- 9) ค่าความเชื่อมั่นของกฎ $L \rightarrow R$ คือ $c(L, R) = s(L \cup R) / s(L)$
- 10) ค่าสนับสนุนของกฎ $L \rightarrow R$ คือ $s(L, R) = s(L \cup R)$
- 11) ทุกกฎความสัมพันธ์ $L \rightarrow R$ เมื่อ $L \subset I, R \subset I$ และ $L \cap R = \emptyset$ แล้วจะต้องประกอบไปด้วยค่าสนับสนุนและค่าความเชื่อมั่น ซึ่งนิยามได้ดังนี้

(1) กฎความสัมพันธ์ $L \rightarrow R$ มีค่าสนับสนุนเท่ากับ s ก็ต่อเมื่อ $s\%$ ของทราจแซคชันใน D บรรจุก $L \cup R$

(2) กฎความสัมพันธ์ $L \rightarrow R$ มีค่าความเชื่อมั่นเท่ากับ c ก็ต่อเมื่อ $c\%$ ของทราจแซคชันใน D บรรจุก $L \cup R$

แต่ละกฎที่ถูกค้นพบเรียกว่ากฎความสัมพันธ์ โดยการค้นหาความสัมพันธ์ คือการค้นหาความสัมพันธ์ทั้งหมดในทราจแซคชันทุกตัว โดยกฎความสัมพันธ์ที่หาได้ทั้งหมดจะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่กำหนดไว้ ($s(L \cup R) \geq s_{min}$) และมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่กำหนดไว้ ($c(L, R) \geq c_{min}$) โดยกฎความสัมพันธ์ที่มีค่าทั้งสองสูงถึงเกณฑ์ที่กำหนดจะถูกเรียกว่า “แข็งแกร่ง” (Strong) ค่าสนับสนุนเรียกว่า “การปรากฏบ่อย”

Database D		จะได้กฎความสัมพันธ์ดังนี้	
tid	Item		
101	beer		$\{\text{beer}\} \rightarrow \{\text{milk, tomatoes}\}$
101	cheese	$s_{min} = 0.2$	ค่าสนับสนุน = $s(\{\text{beer, milk, tomatoes}\})$
102	beer	$c_{min} = 0.5$	= $1/3$
102	milk		= 33.33%
102	tomatoes		ค่าความเชื่อมั่น = $1/2$
103	cheese		= 50%

ภาพที่ 2.6 แสดงตัวอย่างการค้นหาความสัมพันธ์

เครื่องมือในการทำเหมืองข้อมูลประเภทนี้มักจะเป็นใช้เทคนิคหรืออัลกอริธึมมาช่วยรวบรวมและวิเคราะห์ความถี่ของการเกิดข้อมูล เช่น การหาจำนวนความถี่การเกิดรายการ (Frequent item set) ด้วยอัลกอริธึมอะโพรอริ (Apriori) อัลกอริธึมฟรีควอนแพตเทิร์นโกรท (Frequent Pattern Growth – FP - growth) กฎความสัมพันธ์ (association rule) เป็นต้น

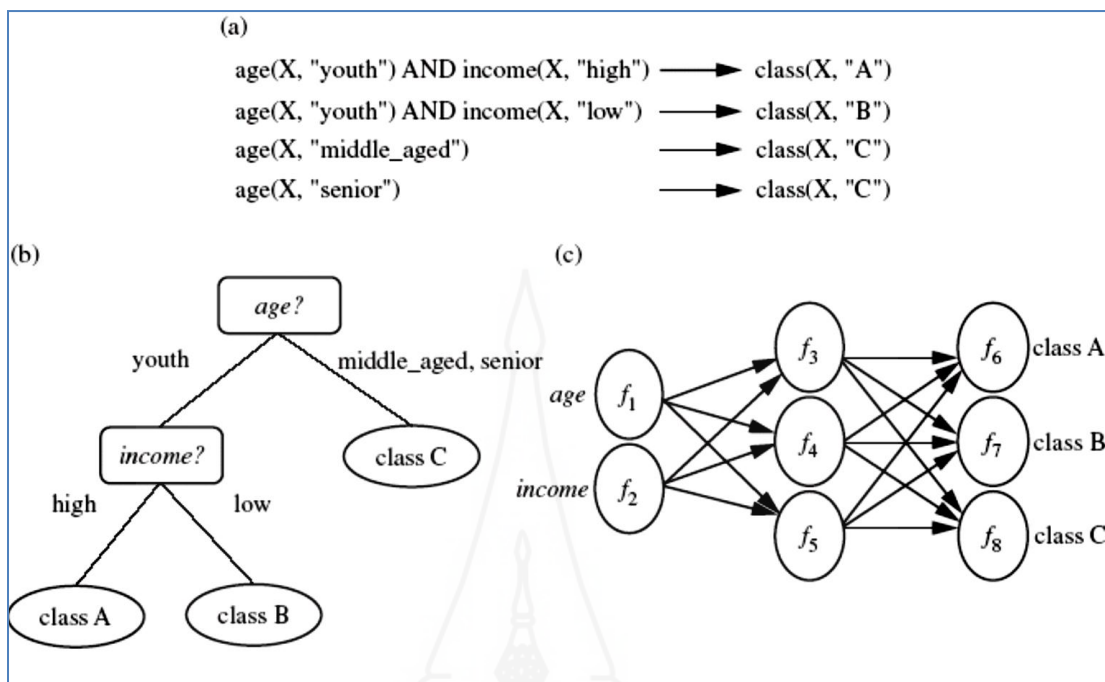
3.3.3 การจำแนกประเภทและการพยากรณ์ การจำแนกประเภทและการพยากรณ์ (Classification and Prediction) แบ่งเป็น

1) **การจำแนกประเภท (classification)** เป็นการทำให้เหมือนข้อมูลเพื่อจำแนกว่าข้อมูลที่ต้องการวิเคราะห์ว่าอยู่ในประเภท (class) ใด โดยอาศัยแบบจำลองหรือตัวแบบที่สร้างขึ้นมา การจำแนกประเภท (classification) จะใช้ข้อมูลที่มีอยู่จำนวนหนึ่ง หรือชุดข้อมูลเพื่อสร้างแบบจำลอง (training data set) หลังจากที่ได้แบบจำลองแล้วก็จะต้องทำการทดสอบแบบจำลองดังกล่าวด้วยข้อมูลอีกกลุ่ม (testing data set) เพื่อตรวจสอบความเที่ยงหรือความน่าเชื่อถือของแบบจำลองนั้น การทำให้เหมือนข้อมูลโดยวิธี จำแนกประเภท (classification) จะใช้สำหรับข้อมูลที่มีค่าไม่ต่อเนื่อง (discrete data) เช่น ข้อมูลที่ระบุเป็นผลการตัดสินใจว่า ใช่ หรือ ไม่ใช่ เสี่ยง หรือ ไม่เสี่ยง

ตัวอย่างเช่น ฝ่ายสินเชื่อของธนาคารต้องการวิเคราะห์ความเสี่ยงในการให้สินเชื่อลูกค้าของธนาคาร ซึ่งมีแบบจำลองที่สร้างขึ้นเพื่อกำหนดหรือใช้พยากรณ์ประเภท (class) โดยมีเงื่อนไข ดังนี้ หากเงินเดือนตั้งแต่ 20,000 ขึ้นไป และมีงานประจำ มีประเภท (class) ความเสี่ยงการให้สินเชื่อ คือ ไม่เสี่ยง แต่หากเงินเดือนต่ำกว่า 20,000 และงานที่ทำเป็นงานชั่วคราว มีประเภท (class) ความเสี่ยงการให้สินเชื่อ คือ เสี่ยง ดังนั้นหากต้องการพยากรณ์ ความเสี่ยงในการให้สินเชื่อแก่นาย A ซึ่งมีเงินเดือน 20,000 และมีอาชีพรับราชการ การพยากรณ์ประเภท (class) ความเสี่ยงของการให้สินเชื่อนาย A คือ ไม่เสี่ยง

เครื่องมือในการทำให้เหมือนข้อมูลแบบจำแนกประเภท (classification) ได้แก่ การใช้อัลกอริธึมสำหรับสร้างต้นไม้การตัดสินใจ (decision tree) ทฤษฎีเบย์ (Bayes Thorem) นาอิว์เบย์ (Naïve Bayes) โครงข่ายความเชื่อแบบเบย์ (Bayes Belief network) การวิเคราะห์แบบ rule base ในรูปแบบ IF_THEN rule โครงข่ายประสาทประดิษฐ์หรือนิวรัลเน็ตเวิร์ก (neural networks) เอสวีเอ็ม (SVM - Support Vector Machine) กฎความสัมพันธ์ (association rule) การหาคุณสมบัติที่มีความคล้ายคลึงหรือใกล้เคียงกันโดยวิธี k - nearest neighbor การหาเหตุและผลเพื่อตอบสนองมาตรฐานโดยใช้วิธี case base reasoning และอัลกอริธึมที่ใช้หลักการเชิงพันธุกรรม (genetic algorithm)

ภาพที่ 2.7 เป็นตัวอย่างเครื่องมือที่ใช้ในการจำแนกประเภท (classification) 2.7 (a) IF - THEN rule 2.7 (b) ต้นไม้การตัดสินใจ (decision tree) 2.7 (c) โครงข่ายประสาทประดิษฐ์ (neural networks)



ภาพที่ 2.7 ตัวอย่างเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภท

ที่มา: วิภา เจริญภักดิ์ (2555: 26)

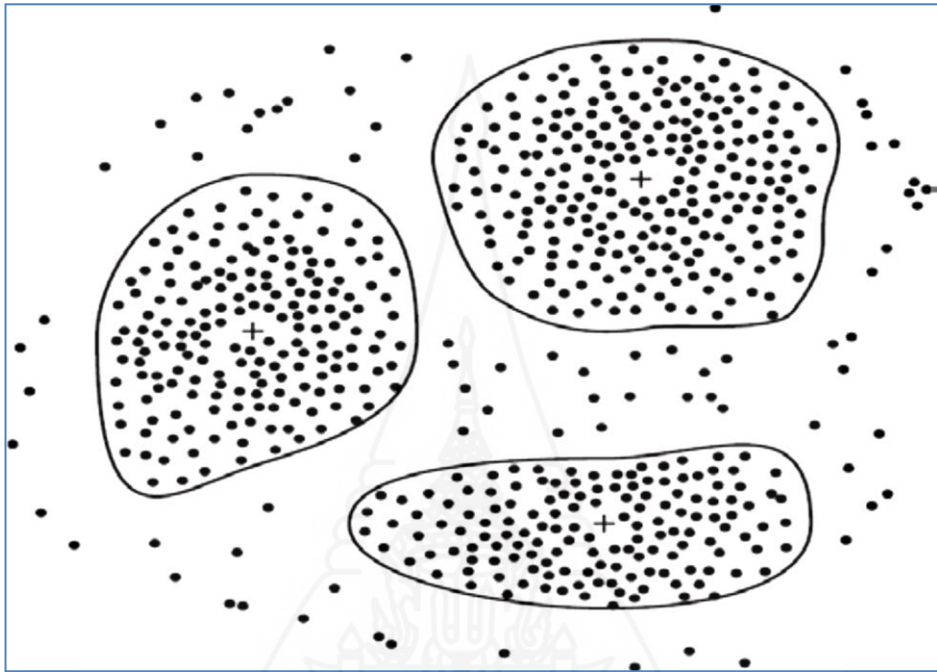
2) การพยากรณ์ (prediction) มีหลักการเช่นเดียวกับการจำแนกประเภท แต่จะใช้กับข้อมูลที่มีค่าต่อเนื่อง (continuous data) หรือค่าข้อมูลที่มีค่าเป็นลำดับการเกิดของข้อมูล (ordered data) นั้น จำแนกอยู่ในประเภท (class) ใด เช่น ข้อมูลอายุ เงินเดือน ความสูง เวลา รายได้จากการขายสินค้า เป็นต้น การทำเหมืองข้อมูลประเภทนี้เพื่อใช้พยากรณ์หรือประมาณการผลลัพธ์จากกลุ่มข้อมูลที่มีอยู่ในขณะนั้น เช่น การพยากรณ์ยอดขายสินค้าจากข้อมูลปัจจุบัน

เครื่องมือในการทำเหมืองข้อมูลแบบพยากรณ์ (Prediction) ได้แก่ การวิเคราะห์ความถดถอย (regression analysis) การวิเคราะห์และตรวจสอบด้วยการใช้ไคสแควร์ (Chi Square Automatic Interaction Detection – CHAID)

3.3.4 การจัดกลุ่ม

การทำเหมืองข้อมูลรูปแบบประเภทการจัดกลุ่ม (clustering) ไม่ได้กำหนดหรือแบ่งกลุ่มไว้ล่วงหน้า เป็นเทคนิคจัดกลุ่มสมาชิกที่มีความคล้ายคลึงกันสูงสุด (nearest neighbor) เพื่อจำแนกข้อมูลแต่ละหน่วยในชุดข้อมูล โดยวัดค่าความห่างจากจุดศูนย์กลาง (distance measurement) การวัดค่าเบี่ยงเบน (deviation measurement) เป็นการรวมหน่วยที่คล้ายกันมากที่สุดเข้าเป็นกลุ่มเดียวกัน ซึ่งบางครั้งเรียกเทคนิคนี้ว่า k - nearest neighbor ตัวอย่างเช่น การทำเหมืองข้อมูลเพื่อวิเคราะห์ตำแหน่งที่อยู่ของลูกค้าบริษัท STOU Electronics เพื่อจัดกลุ่มลูกค้าว่าอาศัยในในเขตพื้นที่ใดบ้าง ดังภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่มได้ จะ

เห็นว่าการจัดกลุ่มเหล่านี้ไม่ได้มีการกำหนดว่าจะต้องเป็นกลุ่มใดล่วงหน้า ในการเกิดกลุ่มเหล่านี้เป็นเพราะข้อมูลในแต่ละกลุ่มมีคุณลักษณะเป็นไปในทางเดียวกันจึงรวมตัวกันเป็นกลุ่มๆ ได้ ซึ่งต่างจากการจำแนกประเภท (classification) ที่มีการกำหนดกลุ่มหรือประเภท (class) ไว้ล่วงหน้า



ภาพที่ 2.8 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ซึ่งจัดได้ 3 กลุ่ม

ที่มา: วิชา เจริญภัณฑารักษ์ (2555: 27)

เครื่องมือในการทำเหมืองข้อมูลแบบการจัดกลุ่ม (Clustering) เช่น การใช้การวิเคราะห์ค่าเฉลี่ยแบบเค (k - mean analysis)

3.3.5 การค้นหาหน่วยข้อมูลที่ผิดปกติ

เป็นการทำเหมืองข้อมูลเพื่อค้นหาหน่วยข้อมูลที่แตกต่างหรือผิดปกติ (Outlier analysis) จากคุณสมบัติข้อมูลทั่วไปที่จัดเก็บในฐานข้อมูลอย่างมาก สาเหตุของการเกิดหน่วยเช่นนี้อาจจะมาจากความผิดพลาดในการบันทึกข้อมูล หรือความผิดพลาดอื่น (noise) หรือเป็นลักษณะผิดปกติจริง (exception) การค้นพบหน่วยผิดปกติและจัดการกับข้อมูลดังกล่าวอย่างเหมาะสมในการทำเหมืองข้อมูลจะทำให้ข้อความรู้ที่ได้มีคุณค่ามากขึ้น เช่น ติดตามการฉ้อโกงในการใช้บัตรเครดิต ที่มีรายการและยอดการใช้จ่ายที่พุ่งสูงขึ้นอย่างรวดเร็ว เมื่อเทียบกับพฤติกรรมการใช้เวลาปกติทั่วไป

เครื่องมือที่ใช้ในการทำเหมืองข้อมูลประเภทนี้ มักจะเป็นการใช้สถิติเพื่อวิเคราะห์ความน่าจะเป็น การวัดค่าความห่างจากจุดศูนย์กลาง (Distance measurement) และการวัดค่าเบี่ยงเบน (deviation measurement)

3.3.6 การวิเคราะห์การเปลี่ยนแปลงที่อิงตามกาลเวลา

เป็นการทำเหมืองข้อมูลเพื่อค้นหารูปแบบข้อมูลที่มีการเปลี่ยนแปลงตามเวลา (Evolution analysis) มักจะดำเนินการผสมผสานวิธีต่างๆ ที่กล่าวมาข้างต้นเพื่อค้นหารูปแบบเพื่อใช้ประโยชน์จากข้อมูลที่มีอยู่ เช่น การวิเคราะห์การเปลี่ยนแปลงของราคาหุ้นในแต่ละประเภทเพื่อการลงทุนที่เหมาะสม เป็นต้นเครื่องมือที่นิยมใช้ เช่น การวิเคราะห์อนุกรมเวลา (time series analysis)

3.4 กระบวนการทำเหมืองข้อมูล

การจัดทำเหมืองข้อมูลมีขั้นตอนหรือกระบวนการทำเหมืองข้อมูล ซึ่งมีประกอบด้วยรายละเอียดต่างๆ ในการดำเนินการแต่ละขั้นตอนอาจจะแตกต่างกันไป ดังนั้นเพื่อให้มีแนวทางการทำงานร่วมกันอย่างชัดเจน จึงได้มีการกำหนดกระบวนการที่เป็นมาตรฐานในการดำเนินการซึ่งมาตรฐานที่ว่่านี้ เรียกว่า คริสป์ - ดีเอ็ม (CRISP - DM – Cross Industry Standard Process for Data Mining) ก่อเกิดขึ้นด้วยเหตุผลเพื่อให้มีขั้นตอนมาตรฐาน ที่มีกรอบงานแต่ละส่วนช่วยให้ผู้ทำเหมืองข้อมูลใช้เป็นแนวทาง เพื่อช่วยให้ทำงานมีประสิทธิภาพขึ้น

3.4.1 หลักการของคริสป์ - ดีเอ็ม ในการกำหนดวิธีของการทำเหมืองข้อมูลจะใช้หลักการพื้นฐาน โดยการแบ่งงานเป็นลำดับชั้น (Hierarchical breakdown) เป็นกลุ่มงานย่อย (set of tasks) 4 ระดับ ดังภาพที่ 2.9

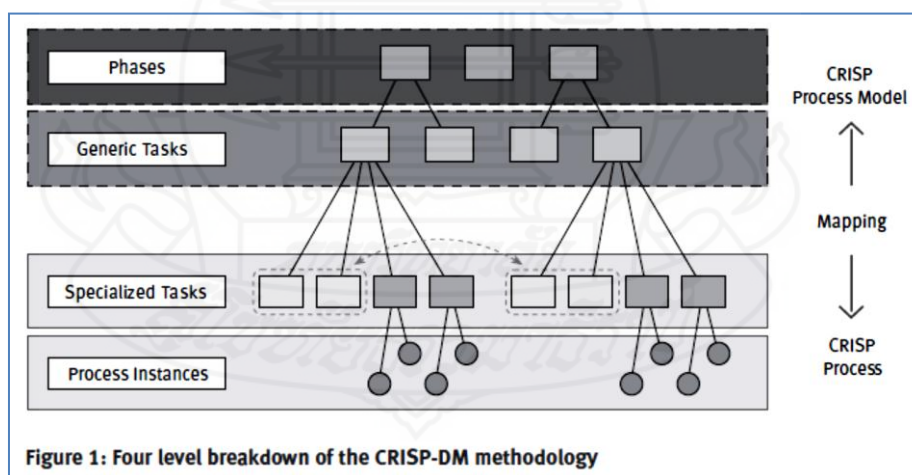


Figure 1: Four level breakdown of the CRISP-DM methodology

ภาพที่ 2.9 การแบ่งระดับงานของคริสป์ - ดีเอ็ม

ที่มา: [http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP - DM.pdf](http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf) Retrieved January 20, 2014

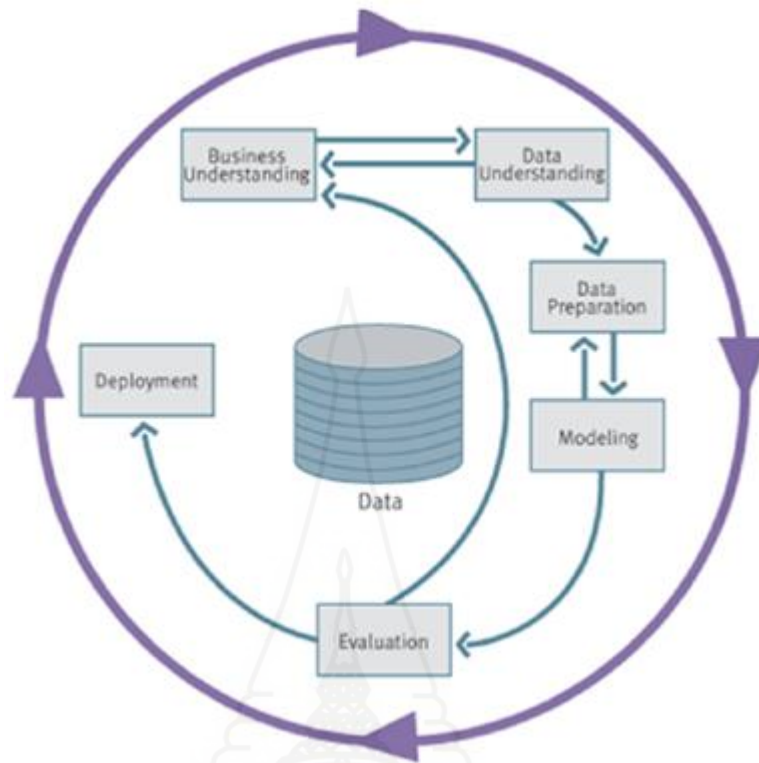
1) *ระดับที่ 1 งานหลัก (phase)* เป็นส่วนบนสุด ประกอบด้วยงานย่อยต่างๆ ที่ต้องดำเนินการในการทำเหมืองข้อมูล ซึ่งเป็นงานทั่วไป (generic task) ทั้งหมด เพื่อให้เห็นภาพรวมของงานที่มีอยู่

2) *ระดับที่ 2 งานทั่วไป (generic task)* ในส่วนนี้จะเป็นการกำหนดว่างานแต่ละงานจะทำเหมืองข้อมูลด้วยเทคนิคอะไรบ้าง ซึ่งบางครั้งอาจมีปัญหาคาดไม่ถึงเกิดขึ้นในระหว่างการทำงานก็ตาม แต่ก็ให้เห็นภาพของงานที่จะต้องทำได้ในระดับหนึ่ง และจะเชื่อมโยงกับระดับถัดไปได้ ตัวอย่างเช่น ระบุว่ามีการทำความสะอาดข้อมูล (clean data) จะต้องมีการกำหนดเทคนิคการทำเหมืองข้อมูลที่คาดว่าจะเหมาะสมสำหรับงานนั้น

3) *ระดับที่ 3 งานเฉพาะ (specific task)* ในส่วนนี้เป็นการขยายรายละเอียดของงานย่อยในระดับ 2 ว่าแต่ละงานย่อยมีวิธีการทำอะไรบ้างในแต่ละสถานการณ์ เช่น การทำความสะอาดข้อมูลที่เป็นตัวเลข ดังนั้นหากมีข้อมูลประเภทอื่นที่ไม่ใช่ตัวเลขแล้วจะมีวิธีการจัดการอย่างไร หรือตัวอย่างจากระดับที่สอง ระบุเทคนิคการทำเหมืองข้อมูลให้ชัดเจนว่าจะใช้เทคนิคใด เช่น การจัดกลุ่ม (clustering) การจำแนกประเภท (classification)

4) *ระดับที่ 4 รายละเอียดวิธีการ (process instance)* เป็นข้อมูลที่กำหนดรายละเอียดงานย่อยในระดับที่เหนือกว่าให้มีรายละเอียดมากขึ้น ซึ่งจะเป็นรายละเอียดเกี่ยวกับการดำเนินการ (action) ของแต่ละงานย่อย เงื่อนไข (decision) ผลลัพธ์ที่เกิดขึ้น (result of the actions) ของงานย่อยแต่ละอัน

3.4.2 ขั้นตอนของ คริสป์ - ดีเอ็ม แบบจำลอง CRISP - DM (Cross - Industry Standard Process for Data Mining) เป็นแบบจำลองแสดงวงจรชีวิตของโครงการทำเหมืองข้อมูลหรือดาต้าไมนิ่ง (วิชา เจริญภัณฑารักษ์, 2555: 35) ดังแสดงในภาพที่ 2.10 โดยแบบจำลอง CRISP - DM จะครอบคลุมระยะเวลาหรือเฟส (phase) การทำงานในโครงการตามลำดับ ของความสัมพันธ์ระหว่างงานทั้งหมดในโครงการ โดยวงจรชีวิตของโครงการทำเหมืองข้อมูล มีด้วยกัน 6 ระยะเวลาหรือเฟส โดยลำดับของงานในแต่ละเฟสไม่ตายตัว การเคลื่อนย้ายไปข้างหน้าหรือถอยหลังระหว่างเฟสสามารถกระทำได้ตลอดเวลา ผลที่ได้รับจากแต่ละเฟส ทำให้ทราบว่าต้องดำเนินการเฟสใดต่อไป ลูกศรที่แสดงในภาพชี้ให้เห็นความสำคัญและการขึ้นต่อกันระหว่างเฟสต่างๆ วงกลมที่อยู่รอบนอกเป็นสัญลักษณ์ของการทำเหมืองข้อมูลซึ่งต้องวนเป็นวงกลม โดยการทำเหมืองข้อมูลจะไม่สิ้นสุดเมื่อมีการนำผลไปใช้ แต่สิ่งที่ได้รับหรือเรียนรู้ในระหว่างกระบวนการและจากการนำผลไปใช้จะสามารถก่อให้เกิดสิ่งใหม่ๆ ซึ่งมักจะมุ่งเน้นเพื่อตอบคำถามให้แก่การดำเนินงานของหน่วยงานได้ โดยผลที่ได้รับจากกระบวนการทำเหมืองข้อมูลจะเป็นประโยชน์มากเมื่อนำประสบการณ์ที่ได้รับจากการวนซ้ำของกระบวนการนี้มาประกอบด้วยในกระบวนการทำเหมืองข้อมูลแบบคริสป์ - ดีเอ็ม ได้กำหนดไว้ 6 ขั้นตอน



ภาพที่ 2.10 แบบจำลอง CRISP - DM

ที่มา: <http://crisp-dm.org/Process/index.htm> Retrieved January 20, 2014

1) *ความเข้าใจธุรกิจ* เป็นการวิเคราะห์ความต้องการธุรกิจ จะช่วยให้เข้าใจสถานะของธุรกิจในปัจจุบัน และประเด็นความต้องการที่ผู้บริหารนำมาใช้ตัดสินใจเพื่อช่วยให้การดำเนินธุรกิจประสบความสำเร็จ และตัดสินใจได้ดียิ่งขึ้น เช่น ความต้องการเกี่ยวกับพฤติกรรมลูกค้า เพื่อพยากรณ์ว่าลูกค้ารายใดน่าจะซื้อสินค้าในรายการสินค้าใหม่ ทำให้สามารถวางแผนธุรกิจได้ตรงกับความต้องการ

2) *ความเข้าใจข้อมูล* เป็นการวิเคราะห์ความต้องการข้อมูลที่เกี่ยวข้อง เนื่องจากแหล่งข้อมูล ที่นำมาจัดทำเหมืองข้อมูลนั้นมีมากมาย และจัดเก็บในคลังข้อมูลหรือฐานข้อมูลขนาดใหญ่ มีข้อมูลที่หลากหลายซึ่งมีทั้งข้อมูลที่เป็นต้องใช้ในการทำเหมืองข้อมูล ปะปนกับข้อมูลอื่นซึ่งไม่ต้องการในการทำเหมืองข้อมูลในบางครั้ง จึงต้องมีขั้นตอนการกำหนดรายการและประเภทของข้อมูลที่จะนำมาใช้ โดยมีการตรวจสอบในด้านของคุณภาพของข้อมูล จำนวน ปริมาณเนื้อหาและการเข้าถึงข้อมูล เพื่อกำหนดเป็นข้อมูลที่ตรงกับความต้องการทำเหมืองข้อมูล ซึ่งอาจจะมีการคัดเลือก โดยการในคำสั่งสอบถามข้อมูลจากคลังข้อมูลที่มีข้อมูลจำนวนมาก โดยเลือกตัวอย่างข้อมูลมาทำเหมืองก่อน ซึ่งช่วยลดค่าใช้จ่าย และประหยัดเวลาในการดำเนินการได้

3) *การเตรียมข้อมูล* เป็นขั้นตอนเพื่อทำการแปลงหรือปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับวิธีการที่จะใช้ในการทำเหมืองข้อมูล เช่น ข้อมูล วัน/เดือน/ปี ที่มาจากแต่ละฐานข้อมูลมีรูปแบบแตกต่างกัน ได้แก่ 9 สิงหาคม 2554 9/08/54 08/9/2011 ในการจัดเตรียมข้อมูลจึงต้องแปลงให้มีรูปแบบเดียวกันเพื่อให้ใช้งานร่วมกันได้ โดยปกติแล้วการแปลงข้อมูลจะถูกกำหนดโดยเงื่อนไขของการปฏิบัติงาน และวิธีการทำเหมืองข้อมูลที่ต้องจัดเตรียมข้อมูลไว้ล่วงหน้าให้มีรูปแบบที่พร้อมใช้งาน

4) *การจัดทำแบบจำลอง* ในขั้นนี้เป็นการใช้เทคนิคการทำเหมืองข้อมูล แต่ละเทคนิคที่มีรูปแบบและวิธีการในการทำแตกต่างกัน เพื่อสร้างแบบจำลองขึ้นมา ซึ่งแบบจำลองที่ได้จะนำช่วยสกัดสาระสำคัญหรือความรู้ที่ซ่อนเร้นอยู่ออกจากฐานข้อมูล เช่น การค้นหาคำตอบว่าลูกค้าจะซื้อสินค้าหรือไม่ อาจต้องทำการวิเคราะห์โดยใช้เทคนิคการจัดกลุ่ม (clustering) ของลูกค้า หรือเทคนิคการจำแนกประเภทลูกค้าแต่ละคน (classification) ว่าจะซื้อหรือไม่ซื้อสินค้า ในการสร้างแบบจำลองด้วยเทคนิคต่างๆ ในการทำเหมืองข้อมูล จะแบ่งกลุ่มข้อมูลที่เกี่ยวข้องคือ ข้อมูลที่ใช้ในการสร้างแบบจำลอง (training data set) และกลุ่มข้อมูลที่ใช้ในการทดสอบแบบจำลอง (testing data set) เพื่อตรวจสอบความเที่ยงตรงหรือความน่าเชื่อถือของตัวแบบจำลองที่สร้างขึ้น การใช้เทคนิคต่างๆ ทำให้ได้แบบจำลองที่ช่วยตอบคำถามที่เกิดขึ้น ทั้งนี้ในขณะที่ทำเหมืองข้อมูล อาจมีความจำเป็นต้องเข้าถึงข้อมูลอื่นในคลังข้อมูล รวมทั้งการแปล หรือทดลองนำผลไปประยุกต์ใช้กับธุรกิจเพื่อทดสอบความน่าเชื่อถือ มีการนำเอาสารสนเทศที่ทำเหมืองข้อมูลที่ได้มาวิเคราะห์เพื่อตอบคำถามที่ผู้ตัดสินใจกำลังค้นหาคำตอบ ซึ่งการวิเคราะห์ในส่วนนี้จะครอบคลุมการกรองสารสนเทศที่เหมาะสมกับผู้ใช้งานและการแปลผล

5) *การประเมินผล* การประเมินผลลัพธ์จากการทำเหมืองข้อมูลที่เกี่ยวข้องกับ 2 ประเด็น คือ

(1) *การประเมินความรู้ที่ค้นพบจากการทำเหมืองข้อมูล* ว่าตรงกับที่ ต้อง การหรือไม่ หรือครอบคลุมประเด็นปัญหาที่มีหรือไม่

(2) *การประเมินวิธีการนำเสนอผลลัพธ์ในรูปแบบที่เหมาะสม* และง่าย ต่อการใช้งาน เช่น แผนภูมิต่างๆ ในเชิงรูปภาพ

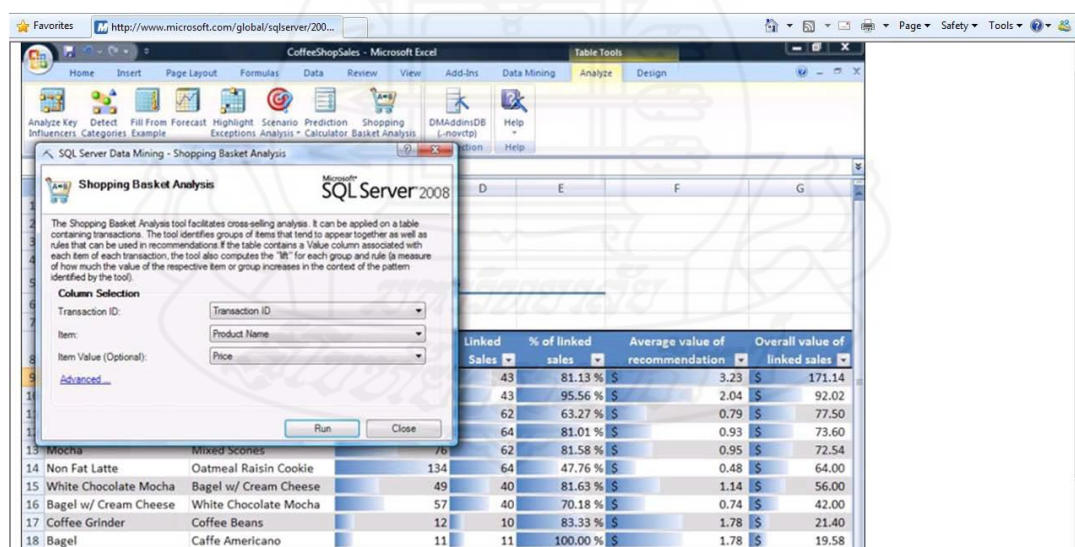
ถ้าผลลัพธ์ที่ได้ไม่เป็นที่พอใจ ก็จะต้องทำขั้นตอนนี้ซ้ำอีก รวมทั้ง ขั้นตอนก่อนหน้าด้วย หากเป็นคำตอบที่พอใจหรือน่าเชื่อถือ ก็จะสามารถดำเนินการขั้นต่อไป

6) *การนำแบบจำลองนำไปใช้งาน* เป็นขั้นตอนสุดท้ายของการทำเหมือง - ข้อมูล หลังจากการตรวจสอบจนแน่ใจว่าแบบจำลองที่สร้างขึ้นมาน่าเชื่อถือ ผู้ใช้งานซึ่งเป็นผู้ทำงาน เกี่ยวข้องกับการวิเคราะห์และตัดสินใจ สามารถนำไปใช้งานได้

4. ซอฟต์แวร์พัฒนาเหมืองข้อมูล

ซอฟต์แวร์พัฒนาค้นข้อมูลในงานวิจัยนี้ใช้ ซอฟต์แวร์ Microsoft SQL Server 2008 ซึ่งเป็นระบบจัดการฐานข้อมูลของบริษัทไมโครซอฟต์ ซึ่งมีฟังก์ชันสำหรับการทำเหมืองข้อมูล เช่น การหาความสัมพันธ์ (Association mining) การจำแนกประเภท (Classification) ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision Tree) การวิเคราะห์ความถดถอย (Regression) การจัดกลุ่ม (Clustering) การหารูปแบบตามลำดับเหตุการณ์ (Sequence pattern) เป็นต้น รวมทั้งยังสามารถทำงานร่วมกับซอฟต์แวร์ของบริษัทอื่นได้ ไมโครซอฟต์เอสคิวแอล เซิร์ฟเวอร์ 2008 (Microsoft SQL Server 2008) ประกอบด้วยเทคโนโลยีต่อไปนี้

- SQL Server Database Engine
- SQL Server Analysis Services
- SQL Server Integration Services
- SQL Server Replication
- SQL Server Reporting Services
- SQL Server Notification Services
- SQL Server Service Broker
- Full – Text Search
- SQL Server Tools and Utilities



ภาพที่ 2.11 ตัวอย่างหน้าจอ Microsoft SQL Server 2008

ที่มา: <http://www.microsoft.com/global/sqlserver/2008/en/us/PublishingImages/scrn - DataMining.jpg> Retrieved January 20, 2014

4.1 SQL Server Database Engine เป็นเทคโนโลยีที่ใช้ในการเก็บการประมวลผล และรักษาความปลอดภัยให้ข้อมูล อีกทั้งยังทำให้สามารถควบคุมการเข้าถึง และการประมวลผลที่รวดเร็วเพื่อสนองต่อความต้องการขององค์กร

4.2 SQL Server Analysis Services เป็นเทคโนโลยีที่ใช้ในการทำ Online Analytical Processing (OLAP) และ การทำเหมืองข้อมูลเพื่อนำไปใช้ด้านธุรกิจ Analysis Service สนับสนุนการทำ OLAP โดยให้ผู้ใช้สามารถออกแบบ สร้าง และจัดการข้อมูลที่ได้จากการรวบรวม ข้อมูลจากแหล่งอื่นๆ เข้าด้วยกัน สำหรับการทำให้เหมืองข้อมูลแล้ว Analysis Service ให้ผู้ใช้สามารถ ออกแบบ สร้าง และดูตัวแบบเหมืองข้อมูลที่สร้างได้ โดยสามารถนำข้อมูลจากแหล่งอื่นๆ มาทำ เหมืองข้อมูลด้วยขั้นตอนวิธี ที่เป็นมาตรฐานและใช้กันอย่างกว้างขวาง

4.3 SQL Server Integration Services เป็นเทคโนโลยีในการแปลงและรวบรวม ข้อมูลระดับองค์กร ซึ่งผู้ ใช้สามารถนำเครื่องมือไปทำการสกัด แปลง และรวบรวมข้อมูลจาก แหล่งข้อมูลหลายแหล่งที่อยู่แยกจากกันได้แล้วนำข้อมูลนี้ไปไว้ปลายทางที่เดียวกัน หรือแยกกันหลายๆ ที่ก็ได้

4.4 SQL Server Replication เป็นชุดของเทคโนโลยีที่มีไว้สำหรับการสำเนา และ กระจายข้อมูลหรือวัตถุในฐานข้อมูลจากฐานข้อมูลหนึ่งไปยังฐานข้อมูลอีกแห่งหนึ่ง และทำให้ข้อมูลใน ฐานข้อมูลหลายๆ แห่งมีความต้องการกัน (Consistency) นอกจากนี้ยังช่วยให้สามารถกระจายข้อมูลไป ยังที่ต่างๆ และสามารถใช้งานข้อมูลเหล่านั้น ผ่านระบบเครือข่าย เชื่อมโยงระยะไกล, เครือข่าย เชื่อมโยงระยะไกล, เครือข่ายไร้สาย หรือเครือข่ายอินเทอร์เน็ตได้

4.5 SQL Server Reporting Services เป็นเทคโนโลยีที่ทำให้สามารถสร้างรายงาน จากข้อมูลที่อยู่ในแหล่งข้อมูลเชิงสัมพันธ์หรือแหล่งข้อมูลหลายมิติได้ โดยสามารถสร้างและจัดการ รายงานที่อยู่ในรูปแบบ Tabular หรือ Matrix หรือ Graphic หรือ Free - Form ได้ ซึ่งรายงานที่ สร้างขึ้นมาสามารถดูและจัดการผ่านเว็บเพจได้

4.6 SQL Server Notification Services เป็นเทคโนโลยีสำหรับสร้างข้อความแจ้ง เตือนและนำส่งข้อความนั้นไปให้ผู้ใช้จำนวนมากในเวลาอันรวดเร็ว

4.7 SQL Server Service Broker เป็นเทคโนโลยีที่ช่วยให้งานที่กระทำบน ฐานข้อมูลนั้นมีความไว้วางใจได้ สามารถปรับขนาดงานได้ และทำให้งานมีความปลอดภัย ซึ่ง Service Broker นั้นเป็นเทคโนโลยีที่อยู่ในส่วนของ Database Engine อีกทั้ง Service Broker จะใช้ การสื่อสารด้วยข้อความเพื่องานต่างๆ ที่อยู่แยกกันให้ สามารถทำงานร่วมกันได้ นอกจากนี้ Service Broker ยังมีเครื่องมือต่างๆ ที่จำเป็นสำหรับสร้างงานบนฐานข้อมูลแบบกระจาย และเครื่องมือเหล่านี้ ช่วยลดระยะเวลาในการพัฒนาฐานข้อมูล ลงได้มาก นอกจากนี้ Service Broker ยังช่วยให้ผู้ใช้ สามารถปรับขนาดของงาน ให้เหมาะสมกับปริมาณงานที่ได้รับ

4.8 SQL Server Tools and Utilities SQL Server มีเครื่องมือต่างๆ ที่ช่วยให้ผู้ใช้ ออกแบบ พัฒนา นำไปประยุกต์ใช้ และบริหารฐานข้อมูลเชิงสัมพันธ์ , Analysis Services Cube, Data Transformation Packages, Replication Technologies, Reporting Servers และ Notification Servers ได้

5. งานวิจัยที่เกี่ยวข้อง

จรรยาพร จำปา(2553) ได้ทำการศึกษาเกี่ยวกับเรื่อง การพยากรณ์ความต้องการ ทรัพยากรห้องสมุดโดยวิธีภูควมสัมพันธ์ของห้องสมุดการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย โดยการ ทำเหมืองข้อมูล เพื่อวิเคราะห์พฤติกรรมการใช้ทรัพยากรห้องสมุดตามภูควมสัมพันธ์ของการยืม ทรัพยากรห้องสมุดที่มีผลกับหมวดหมู่ของทรัพยากรห้องสมุด กฟผ. และพยากรณ์ความต้องการ ทรัพยากรห้องสมุดในอนาคต โดยใช้ข้อมูลการยืมทรัพยากรห้องสมุดตั้งแต่ ปี พ.ศ. 2548 - 2553 ผล การศึกษาโดยเริ่มจากการหาความสัมพันธ์ด้วยอัลกอริทึม Apriori พบกฎทั้งสิ้น 4 กฎ คือ กฎที่ 1 ผู้ใช้บริการเมื่อยืมหมวดหมู่ 300 (สังคมศาสตร์) ก็จะยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือ เทคโนโลยี) กฎข้อที่ 2 เมื่อยืมหมวดหมู่ 000 (เบ็ดเตล็ดหรือความรู้ทั่วไป) ก็จะยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือเทคโนโลยี) กฎข้อที่ 3 เมื่อยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือ เทคโนโลยี) ก็จะยืมหมวดหมู่ 000 (เบ็ดเตล็ดหรือความรู้ทั่วไป) กฎข้อที่ 4 เมื่อยืมหมวดหมู่ 600 (วิทยาศาสตร์ประยุกต์หรือเทคโนโลยี) ก็จะยืมหมวดหมู่ 300 (สังคมศาสตร์) โดยมีค่าความเชื่อมั่น 18, 12, 7 และ 7 เปอร์เซนต์ตามลำดับ และพยากรณ์ความต้องการทรัพยากรห้องสมุดในอนาคต นำไปวิเคราะห์การถดถอยและวิเคราะห์อนุกรมเวลาเพื่อหาสมการพยากรณ์ความต้องการ โดยแต่ละ สมการมีการประเมินความแม่นยำโดยหาค่าความคลาดเคลื่อนสัมพัทธ์ (MRE) ได้เท่ากับ 36.36 16.47 16.46 และ 10.67 เปอร์เซนต์ตามลำดับ และนำไปสร้างโปรแกรมการพยากรณ์ความต้องการ ทรัพยากรห้องสมุดเพื่อใช้ในการบริหารจัดการ การจัดสรรงบประมาณต่างๆ ในการจัดหาหรือจัดซื้อ ทรัพยากรสารสนเทศให้ตรงตามความต้องการของผู้ใช้บริการ ดังนั้นการเลือกอัลกอริทึมในการสร้าง โมเดลการพยากรณ์นั้นควรทดสอบด้วยหลายๆ อัลกอริทึมเพื่อให้ได้โมเดลพยากรณ์ที่มีความเหมาะสม มากที่สุดในการวิเคราะห์พฤติกรรมการใช้ทรัพยากรห้องสมุดเพื่อนำมาปรับปรุงการบริหารจัดการ งบประมาณอย่างคุ้มค่าและตรงกับความต้องการของผู้ใช้บริการมากที่สุด และต้องทำการวิเคราะห์กฎ ใหม่ทุกปี การพยากรณ์จึงจะมีความแม่นยำมากขึ้น

คุณรัตน์ กรณทัณฑ์กันแสง (2553) ได้ทำการศึกษาเกี่ยวกับเรื่อง การประยุกต์ใช้ทฤษฎี เหมืองข้อมูลในบริการยืม - คืนของห้องสมุดกรณีศึกษา สำนักหอสมุด มหาวิทยาลัยธรรมศาสตร์ โดย วิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทั้งข้อมูลผู้ใช้บริการและข้อมูลการบริการยืม - คืนหนังสือ เพื่อให้ ทราบถึงการเข้าใช้บริการยืม - วัสดุสารสนเทศของคณะต่างๆ ใน มหาวิทยาลัยธรรมศาสตร์ การ พยากรณ์ความน่าจะเป็นในการยืม - คืนวัสดุสารสนเทศในแต่ละหมวด ซึ่งใช้ข้อมูลบริการยืม - คืน หนังสือตั้งแต่ไตรมาสที่ 1 ถึงไตรมาสที่ 2 ของปี พ.ศ. 2552 ประมาณ 60,000 รายการ โดยการนำ ข้อมูลมาทำเหมืองข้อมูลตามขั้นตอนของ CRISP-DM ด้วยอัลกอริทึมภูควมสัมพันธ์ และกฎการ แบ่งกลุ่ม ผลของการศึกษาด้วยอัลกอริทึมการแบ่งกลุ่มสามารถแบ่งผู้ยืมออกเป็น 4 กลุ่ม ซึ่งสามารถ ใช้เป็นข้อมูลประกอบการพิจารณาในการปรับปรุงบริการและการแนะนำหนังสือสำหรับผู้ใช้แต่ละ กลุ่ม และ สำหรับอัลกอริทึมการหาความสัมพันธ์พบกฎที่มีความเชื่อมั่นมากกว่า 80% โดย บรรณารักษ์สามารถใช้เป็นข้อมูลประกอบการพิจารณาในการปรับปรุงการบริการตามภูควมสัมพันธ์ที่ได้ เช่น กฎแสดงความสัมพันธ์ที่ว่า ถ้าร้อยละ 95 ของสมาชิกห้องสมุดที่ได้ยืมหนังสือ

หมวดรัฐศาสตร์ และหมวดกฎหมาย แล้วจะต้องยืมหนังสือหมวดการศึกษาด้วย ทำให้บรรณารักษ์สามารถนำข้อมูลรูปแบบการยืมข้างต้นมาใช้ประโยชน์โดยจัดเรียงหนังสือหมวดรัฐศาสตร์ หมวดกฎหมาย หมวดการศึกษาให้อยู่ใกล้กัน และกฎความสัมพันธ์ที่ว่า ถ้าร้อยละ 90 ของสมาชิกห้องสมุดที่ยืมหนังสือหมวดกฎหมายมักจะยืมหนังสือในหมวดวิศวกรรมหมวดสังคมศาสตร์ด้วย บรรณารักษ์สามารถนำข้อมูลรูปแบบการยืม หนังสือข้างต้นมาใช้ประโยชน์โดยการศึกษาถึงปริมาณหนังสือทั้ง 2 หมวดว่ามีสัดส่วนที่เท่ากันหรือไม่ ถ้าหนังสือในหมวดกฎหมายมีปริมาณน้อยไป บรรณารักษ์สามารถจะกันหนังสือในหมวดดังกล่าวไว้เป็นหนังสือสำรองหรือกำหนดวันคืนหนังสือให้เร็วขึ้นกว่าปกติเพื่อให้งานใช้งานเป็นไปอย่างทั่วถึง

Cunningham & Frank (1999) ได้ทำการศึกษาเกี่ยวกับเรื่อง Market basket analysis of library circulation data ซึ่งเป็นการวิเคราะห์ความสัมพันธ์ โดยใช้หลักการ Market basket analysis ที่ใช้ในการหาความสัมพันธ์ของสินค้าที่ถูกค้ำมักจะซื้อพร้อมกัน ซึ่งใช้ข้อมูลการยืม - คืนหนังสือช่วงปี 1998 จำนวน 52,518 เรคคอร์ด ของห้องสมุดมหาวิทยาลัยไวคาโต (University of Waikato) ประเทศนิวซีแลนด์มาทำการประมวลผล และใช้โปรแกรม WEKA เป็นเครื่องมือในการประมวลผลข้อมูล เพื่อค้นหาความสัมพันธ์ของหนังสือที่มักถูกยืมไปด้วยกัน ด้วยขั้นตอนวิธี Apriori โดยการหาความสัมพันธ์ของหมวดหมู่หนังสือ ซึ่งมีการจัดหมวดหมู่หนังสือแบบระบบการจัดหมวดหมู่หนังสือของหอสมุดรัฐสภาอเมริกัน (L.C.) ใช้ค่า Minimum support = 0.01 และค่า Minimum confidence = 0.01 การวิจัยเป็นการหาค่าความสัมพันธ์ของหมวดหมู่หนังสือ หัวเรื่องหนังสือ ที่มีข้อมูลเกิดขึ้นร่วมกันหรือมีความสัมพันธ์กันของหนังสือที่มีการยืมคืนจากห้องสมุด ผลของการวิจัย ได้นำไปใช้ประโยชน์การจัดวางตำแหน่งของหนังสือบนชั้นหนังสือให้เหมาะสมและตรงตามความต้องการของสารสนเทศของผู้ใช้บริการ เพื่อให้ผู้ใช้บริการเห็นหนังสือที่ต้องการโดยง่าย

Nicholson & Stanton (2003) ได้ทำการศึกษาเกี่ยวกับเรื่อง The Bibliomining Process: Data Warehousing and Data Mining for Library Decision - Making เพื่อศึกษาการทำเหมืองข้อมูลกับข้อมูลต่างๆ ที่มีอยู่ในฐานข้อมูลห้องสมุด ซึ่งการดำเนินการตามกระบวนการทำเหมืองข้อมูลในห้องสมุด (Bibliomining process) นั้นประกอบไปด้วย 6 ขั้นตอนคือ 1) การระบุเขตข้อมูล 2) การระบุแหล่งของข้อมูล 3) การสร้างคลังข้อมูล 4) การเลือกเครื่องมือในการวิเคราะห์ 5) การทำเหมืองข้อมูล 6) การวิเคราะห์และผลไปใช้งาน ซึ่งผลที่ได้จากการทำเหมืองข้อมูลสามารถนำไปเป็นข้อมูลช่วยในการตัดสินใจปรับปรุงห้องสมุดในด้านต่างๆ ดังนี้

- การทำนายความต้องการใช้หนังสือในอนาคต โดยดูจากความถี่ของหนังสือที่มีจำนวนครั้งการยืมสูง แล้วนำมาตัดสินใจว่าควรมีหนังสือในหมวดใด จำนวนใด เท่าใด จึงจะเพียงพอต่อการใช้งานของผู้ใช้บริการ

- การค้นหาว่าหนังสือเล่มใด ประเภทใดที่มักมีการส่งคืนช้ากว่ากำหนดเพื่อนำมาวางนโยบายต่างๆ ให้หนังสือหายน้อยลง หรือให้ผู้ใช้คืนหนังสือตรงเวลามากขึ้น

- การวิเคราะห์หาหนังสือที่ไม่เพียงพอต่อความต้องการ โดยการเก็บข้อมูลการสืบค้นหนังสือที่ผู้ใช้ต้องการแล้วหาไม่พบมาใช้ในการวิเคราะห์ข้อมูลเพื่อดำเนินการจัดหาหนังสือมาให้บริการ

- การวิเคราะห์ข้อมูลหนังสือที่มีถูกยืมไปด้วยกัน โดยการหาความสัมพันธ์หรือ Market Basket Analysis มาเป็นเครื่องมือในการวิเคราะห์ข้อมูลเพื่อให้สามารถแสดงข้อมูลหนังสือเล่มอื่นๆ ที่มีเนื้อหาเกี่ยวข้องกับหนังสือที่ผู้ใช้ต้องการยืม เพื่อเป็นทางเลือกให้ผู้ใช้บริการสามารถหาข้อมูลที่เกี่ยวข้องกันได้ โดยไม่ต้องเสียเวลาในการสืบค้น เพิ่มประสิทธิภาพ ประหยัดเวลา ประหยัดงบประมาณ และตอบสนองความต้องการของผู้ใช้ห้องสมุดได้

- การวิเคราะห์และค้นหารูปแบบการใช้งานเว็บไซต์ของห้องสมุด เพื่อนำข้อมูลที่ได้มาปรับปรุงเว็บไซต์ ให้เหมาะสม เช่นการวางข้อความที่สำคัญไว้ในตำแหน่งที่ผู้ชมมองเห็นง่าย หรือปรับปรุงให้เว็บไซต์น่าดึงดูด เพื่อไม่ให้ผู้ใช้ออกจากเว็บไซต์

- การแบ่งกลุ่มผู้ใช้บริการออกเป็นกลุ่มๆ แล้วนำข้อมูลที่ได้มาจัดสรรประเภทและจำนวนทรัพยากรให้สอดคล้องกับจำนวนผู้ใช้ในแต่ละกลุ่ม

จากผลการวิจัยที่ได้เกิดจากข้อมูลของแต่ละห้องสมุด ดังนั้นการดำเนินการตามกระบวนการทำเหมืองข้อมูลในห้องสมุด (Bibliomining process) บรรณารักษ์ต้องศึกษาข้อมูลที่น่ามาประมวลผลให้ตรงตามความต้องการของแต่ละห้องสมุด โดยเฉพาะอย่างยิ่งการสร้างคลังข้อมูล เพราะหากมีคลังข้อมูลที่มีประสิทธิภาพถือเก็บข้อมูลจำนวนมาก การดำเนินการตามกระบวนการทำเหมืองข้อมูลจะทำให้ข้อมูลมีความถูกต้องและการพยากรณ์จะมีความแม่นยำมากขึ้น

Chen & chen (2006) ได้ทำการศึกษาเกี่ยวกับเรื่อง A New efficient approach for data clustering in electronic library using ant colony clustering algorithm ซึ่งเป็นการทำเหมืองข้อมูล โดยใช้กระบวนการ KDD (Knowledge Discovery in Database) เพื่อหาวิธีที่มีประสิทธิภาพสำหรับการจัดกลุ่มข้อมูลในห้องสมุดอิเล็กทรอนิกส์ โดยการนำข้อมูลธุรกรรมการยืม – คืนสื่ออิเล็กทรอนิกส์ของผู้ใช้บริการห้องสมุด มาวิเคราะห์ข้อมูลการสร้างคลังข้อมูล โดยใช้ข้อมูลรหัสนักศึกษา (ID) ในการจำแนกข้อมูลนักศึกษา สังกัดหรือหน่วยงานภายในมหาวิทยาลัย และข้อมูลหมวดหนังสือ (CNO) ในการจัดเก็บข้อมูลหมวดหมู่หรือประเภทหนังสือ และข้อมูลการยืมหนังสือที่เก็บในฐานข้อมูลมหาวิทยาลัย มาใช้ในการทำคลังข้อมูลเพื่อหาความคล้ายคลึงกันของข้อมูลในการแบ่งกลุ่มผู้ใช้โดยการเทคนิคการทำเหมืองข้อมูลและใช้ขั้นตอนวิธี Ant Colony Clustering และ K - Means Clustering ในการแบ่งกลุ่มผู้ใช้บริการ ผลการวิเคราะห์พบว่าพฤติกรรมกรการยืมหนังสือของนักศึกษาที่มีลักษณะคล้ายคลึงกันสามารถแบ่งกลุ่มนักศึกษาออกเป็น 3 กลุ่ม และกลุ่มที่เกิดจากการแบ่งกลุ่มด้วยขั้นตอนวิธี Ant Colony Clustering นั้นสมาชิกในกลุ่มจะมีลักษณะคล้ายกันมากกว่า และช่วยให้ผู้ใช้ที่ค้นหาสิ่งที่พวกเขา กำลังมองหาได้อย่างมีประสิทธิภาพมากขึ้นเมื่อเทียบกับกลุ่มสมาชิกที่มีการแบ่งด้วยขั้นตอนวิธี K - Means Clustering แต่ทั้งนี้การเลือกใช้เทคโนโลยีหรือเทคนิคต่างๆ ในการจัดการต้องเลือกเทคนิคที่เหมาะสมในการใช้งาน เพื่อตอบสนองความต้องการของผู้ใช้ได้อย่างถูกต้อง มีคุณค่าและเหมาะสม

Chen & chen (2007) ได้ทำการศึกษาเกี่ยวกับ Using data mining technology to provide a recommendation service in the digital library โดยนำธุรกรรมการยืม –คืนหนังสือ และวัสดุสารสนเทศของสมาชิกห้องสมุดซึ่งเป็นการทำเหมืองข้อมูล ด้วยขั้นตอนวิธี Ant Colony Clustering เป็นการแยกกลุ่มผู้ใช้ออกเป็นกลุ่มๆ ตามความสัมพันธ์หรือความสอดคล้อง และคล้ายคลึงกันของพฤติกรรมกรยืมหนังสือ ซึ่งข้อมูลที่เหมือนกันจะถูกจัดอยู่ในกลุ่มเดียวกัน ซึ่งสามารถจัดกลุ่มข้อมูลทั้งหมดได้ 3 กลุ่ม แล้วนำข้อมูลที่ได้อมาในแต่ละกลุ่มหากกฎความสัมพันธ์ (Association Rule) ของหนังสือและวัสดุสารสนเทศ โดยใช้ขั้นตอนวิธี Apriori เพื่อหาความสัมพันธ์ของหนังสือที่มีการยืมไปด้วยกันของแต่ละกลุ่ม ซึ่งเซตค่าความเชื่อมั่น (Confidence) ร้อยละ 80 และมีค่าสนับสนุน (Support) ร้อยละ 50 ผลของการศึกษานำไปสร้างระบบให้คำแนะนำสำหรับการค้นคืนหนังสือผ่านเว็บ เป็นแนวทางสำหรับห้องสมุดในการวางแผนการซื้อหนังสือได้ตรงตามความต้องการด้วยงบประมาณที่จำกัด และตอบสนองความต้องการของผู้อ่านไปพร้อมกับการส่งเสริมการบริการห้องสมุดดิจิทัล เพื่อสนับสนุนการให้บริการคำแนะนำในการค้นหาหนังสือที่เหมาะสมที่สุดกับสมาชิกห้องสมุด

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังที่กล่าวมาแล้วนั้นมีการทำเหมืองข้อมูลในห้องสมุดจำนวนมาก ในการทำเหมืองข้อมูลส่วนใหญ่ใช้โปรแกรม WEKA เป็นเครื่องมือในการวิเคราะห์ข้อมูลซึ่งหากต้องทดสอบข้อมูลแต่ละครั้งต้องเสียเวลาในการสร้างไฟล์เพื่อทดสอบตามรูปแบบของอัลกอริทึมต่างๆ ซึ่งมีการใช้ทั้งวิธีการแบ่งกลุ่มและกฎการหาความสัมพันธ์ของข้อมูล ซึ่งงานวิจัยที่ศึกษาต่างมีวัตถุประสงค์เพื่อตอบสนองความต้องการของผู้ใช้บริการทั้งสิ้น และเพื่อให้งานบริการห้องสมุดมีประสิทธิภาพและมีบริการที่ตรงตามความต้องการของผู้ใช้หรือสมาชิก แต่งานวิจัยที่กล่าวมานั้นส่วนใหญ่ใช้ข้อมูลพฤติกรรมบริการยืมคืนหนังสือของสมาชิก ผู้วิจัยจึงมีความสนใจศึกษาเกี่ยวกับการใช้เทคโนโลยีเหมืองข้อมูลในการสนับสนุนการค้นคืนสารสนเทศของห้องสมุด โดยนำข้อมูลการยืมคืนหนังสือและข้อมูลส่วนตัวของสมาชิกห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ มาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล โดยใช้เครื่องมือ SQL Server 2008 ในการวิเคราะห์ข้อมูลตามขั้นตอนของ CRISP-DM ด้วยอัลกอริทึมกฎความสัมพันธ์ กฎการแบ่งกลุ่ม และวิเคราะห์หอนุกรมเวลา ซึ่งเป็นขั้นตอนที่มีการใช้งานอย่างแพร่หลาย จึงทำให้มีความเชื่อมั่นว่าผลของการศึกษาวิจัยจะมีความถูกต้อง ช่วยให้นำข้อมูลไปประยุกต์ใช้ในการค้นคืนสารสนเทศภายในห้องสมุด และช่วยให้ผู้ให้บริการสามารถเข้าถึงสารสนเทศที่ตรงตามความต้องการของผู้ใช้ได้อย่างถูกต้อง รวดเร็ว ตรงตามความต้องการของผู้ใช้

บทที่ 3

วิธีดำเนินงานวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงพัฒนา (Development research) โดยเป็นการพัฒนาคลังข้อมูลนักศึกษาที่มาใช้บริการยืมหนังสือของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษเพื่อนำมาวิเคราะห์หาความรู้ใหม่โดยการทำให้เหมือนข้อมูลตามขั้นตอนของแบบจำลองของคริสป์ - ดีเอ็ม (CRISP - DM Model) ซึ่งรายละเอียดของวิธีดำเนินการวิจัย ประกอบด้วย 3 ส่วน ได้แก่

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการดำเนินงานวิจัย
3. ขั้นตอนการดำเนินงาน

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร

ข้อมูลการใช้บริการยืมหนังสือของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และข้อมูลผู้ใช้บริการจากฐานข้อมูลการเข้าใช้บริการยืมหนังสือห้องสมุด และฐานข้อมูลงานทะเบียนมหาวิทยาลัยราชภัฏศรีสะเกษ

1.2 กลุ่มตัวอย่าง

สุ่มข้อมูลแบบเจาะจง (Purposive sampling) โดยใช้ข้อมูลการบริการยืมคืนสารสนเทศของนักศึกษาที่เป็นสมาชิกห้องสมุดจากโปรแกรมห้องสมุดอัตโนมัติ WALAI AUTOLIB ช่วงเวลาตั้งวันที่ 1 มิถุนายน 2553 ถึง 31 ธันวาคม 2556

2. เครื่องมือที่ใช้ในการดำเนินงานวิจัย

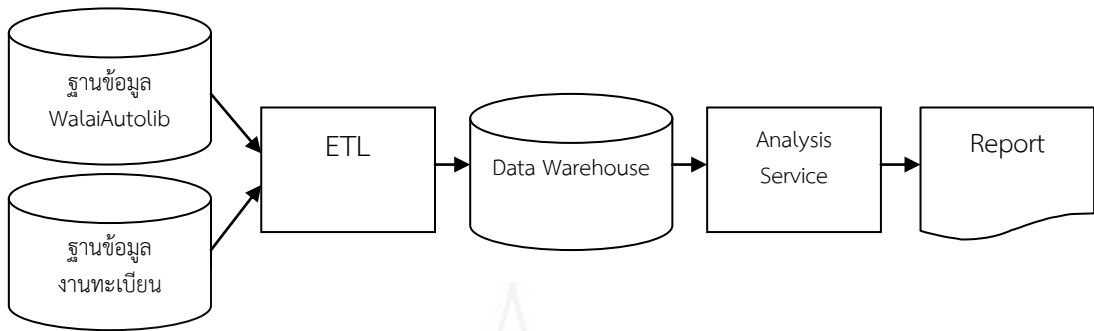
ซอฟต์แวร์ Microsoft SQL Server 2008 เพื่อพัฒนาคลังข้อมูลผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และซอฟต์แวร์ Microsoft Visual Studio 2008 เพื่อทำเหมืองข้อมูล (Data Mining Tool) โดยการวิเคราะห์และจำแนกข้อมูลกลุ่มผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และศึกษารูปแบบความสัมพันธ์ของข้อมูลให้ทราบถึงพฤติกรรมการยืมหนังสือของผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และใช้ซอฟต์แวร์ SQL Server Analysis Services สำหรับการทำให้ Online Analytical Processing (OLAP) ให้ผู้ใช้สามารถออกแบบ เรียกดูรายงานได้หลายมิติทำให้ทราบผลการดำเนินงานของการบริการยืมหนังสือของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ

3. ขั้นตอนการดำเนินงาน

งานวิจัยนี้ประกอบด้วยขั้นตอนสำคัญ 3 ขั้นตอนคือ 1) การสร้างคลังข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด 2) การทำ Online Analytical Processing (OLAP) 3) และการทำเหมืองข้อมูล

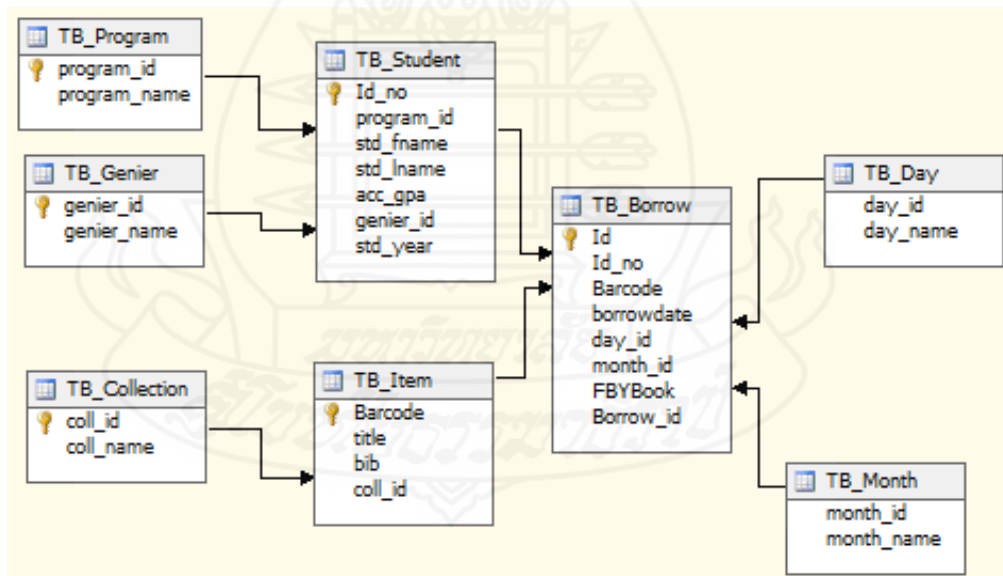
3.1 การสร้างคลังข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด

คลังข้อมูลที่มีคุณภาพนั้น ขึ้นอยู่กับคุณภาพของข้อมูลที่จัดเก็บในคลัง กระบวนการสำคัญในการนำเข้าสู่ข้อมูลสู่คลังข้อมูลคือกระบวนการอีทีแอลซึ่งประกอบด้วย 3 ขั้นตอนคือ 1) การสกัดข้อมูล (extraction) เป็นขั้นตอนการสกัดหรือเลือกเฉพาะข้อมูลที่ต้องการหรือสนใจจากแหล่งข้อมูลต่างๆ ที่มีอยู่มาจัดเก็บไว้ในคลังข้อมูล 2) การจัดฟอร์แมตข้อมูล (transformation) หรือบางครั้งเรียกว่าการทำความสะอาดข้อมูล (cleansing) เป็นการขจัดความไม่สอดคล้องกันของข้อมูลเพื่อให้ข้อมูลที่มาจากแหล่งจัดเก็บที่แตกต่างกัน แต่เป็นข้อมูลในเรื่องเดียวกัน เช่น ข้อมูลวันเดือนปี ข้อมูลเพศที่อาจใช้รหัสการแทนเพศที่แตกต่างกัน เป็นต้น โดยข้อมูลอาจมีฟอร์แมต (format) การจัดเก็บที่ต่างกันไป ดังนั้นจึงต้องมีการจัดฟอร์แมตใหม่ให้มีความสอดคล้องกัน เพื่อป้องกันการผลลัพธ์ที่อาจผิดพลาดเนื่องจากการประมวลผลข้อมูลที่มีฟอร์แมตแตกต่างกัน 3) การนำเข้าข้อมูล (loading) หลังจากที่มีการเลือกและจัดฟอร์แมตข้อมูลเสร็จเรียบร้อยแล้วขั้นตอนต่อไปคือการนำข้อมูลดังกล่าวเข้าสู่คลังข้อมูลที่ได้ออกแบบไว้แล้วก่อนที่จะนำไปจัดทำรายงานหลายมิติหรือวิเคราะห์ด้วยเครื่องมือต่างๆ ต่อไป ดังนั้น กระบวนการอีทีแอลจึงเป็นการเตรียมข้อมูลที่จะนำเข้าสู่คลังข้อมูลที่อยู่บนพื้นฐานความต้องการของผู้ใช้ว่าจะวิเคราะห์ข้อมูลใดบ้าง ภายใต้วัตถุประสงค์ในการทำเหมืองข้อมูลและอ้างอิงจากโครงสร้างเดิมการแก้ไขข้อมูลจะไม่มีเปลี่ยนแปลงความหมายของข้อมูล อาจเป็นการสร้างตารางใหม่จากการรวมกันของตารางหลายๆ ตารางหรือหลายๆ ระเบียบ ซึ่งในงานวิจัยนี้ผู้วิจัยนำข้อมูลจากฐานข้อมูลของโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib ที่จัดเก็บข้อมูลการบริการยืมคืนสารสนเทศภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และข้อมูลนักศึกษาจากฐานข้อมูลทะเบียนมหาวิทยาลัยราชภัฏศรีสะเกษ นำมาสร้างระบบคลังข้อมูลใหม่ ซึ่งในการออกแบบสถาปัตยกรรมของระบบคลังข้อมูลการยืมคืนสารสนเทศของนักศึกษาภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ มีการนำข้อมูลมาจากฐานข้อมูลโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib และฐานข้อมูลงานทะเบียน มาผ่านกระบวนการอีทีแอล (Extract Transform and Load : ETL) เพื่อเลือกหรือสกัดข้อมูลที่ต้องการและจัดการกับฟอร์แมตข้อมูล หรือทำความสะอาดข้อมูลให้มีความถูกต้องสอดคล้องกัน จากนั้นจึงนำข้อมูลที่ได้เข้าสู่คลังข้อมูลเพื่อบันทึกจัดเก็บ แล้วจึงดำเนินการนำข้อมูลในคลังข้อมูลไปประมวลผลในด้วยเครื่องมือต่างๆ สำหรับเครื่องมือที่ใช้ในงานวิจัยนี้คือ Analysis Service Data Mining และการทำรายงานหลายมิติ ดังภาพที่ 3.1



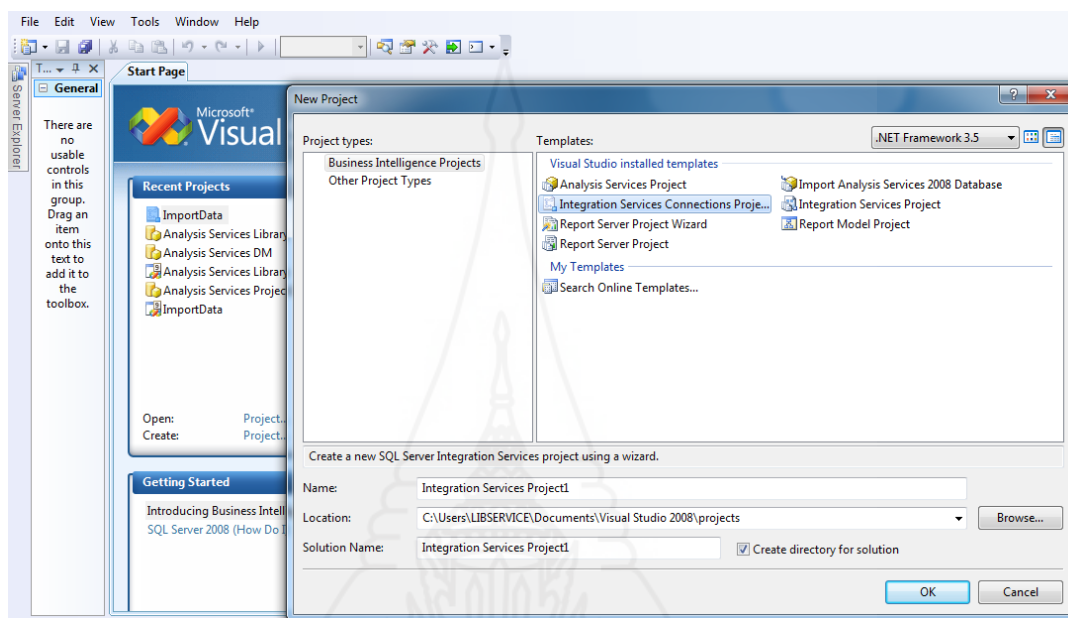
ภาพที่ 3.1 แสดงขั้นตอนการจัดทำคลังข้อมูลการให้บริการยืมคืนของห้องสมุด

ในการสร้างคลังข้อมูลดังกล่าวผู้วิจัยได้ออกแบบโครงสร้างคลังข้อมูลโดยใช้แผนภาพอีอาร์ (ER Diagram) และจากนั้นจึงนำแผนภาพอีอาร์มาแมปปีง (Mapping) เป็นสคีมา (Schema) โดยใช้โครงสร้างคลังข้อมูลแบบสโนว์เฟลค (Snowflake Schema) ที่มีตารางข้อเท็จจริง (Fact Table) หนึ่งตารางคือ การยืมหนังสือ และมีตารางมิติ (Dimension) อื่นๆ อยู่โดยรอบ ซึ่งตารางมิติบางตารางจะมีตารางมิติอื่นเชื่อมโยงต่อไปอีก เช่น ตารางมิตินักศึกษา มีตารางมิติโปรแกรมการศึกษา (TB_Student) และตารางมิติเพศ (TB_Genier) เชื่อมโยงอีก ดังแสดงในภาพที่ 3.2



ภาพที่ 3.2 แสดงโครงสร้างคลังข้อมูลแบบสโนว์เฟลค

ผู้วิจัยได้ดำเนินการนำข้อมูลเข้าสู่คลังข้อมูลโดยใช้กระบวนการอีทีแอล (Extract Transform and Load: ETL) คัดเลือก ทำความสะอาดและนำเข้าข้อมูล กระบวนการอีทีแอลในงานวิจัยนี้ใช้เครื่องมือ Integration Service ของ Microsoft SQL Server ดังแสดงในภาพที่ 3.3

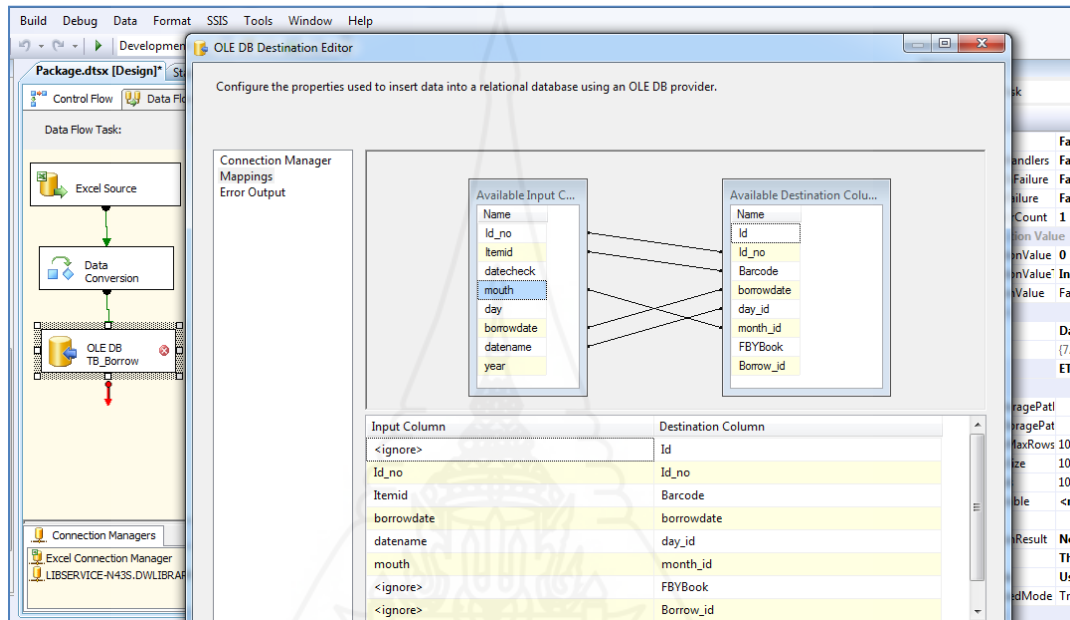


ภาพที่ 3.3 การใช้เครื่องมือ Integration Service ในกระบวนการอีทีแอล

3.1.1 ขั้นตอนการสกัดข้อมูล (Data Extraction) คือการนำข้อมูลออกมาจากแหล่งข้อมูล ซึ่งข้อมูลที่ใช้ในการดำเนินงานวิจัยในครั้งนี้เป็นข้อมูลการบริการยืมคืนสารสนเทศของนักศึกษาที่เป็นสมาชิกห้องสมุดจากโปรแกรมห้องสมุดอัตโนมัติ WALAI AUTOLIB และข้อมูลเกี่ยวกับนักศึกษา ได้แก่ข้อมูลทั่วไปของนักศึกษา ข้อมูลการลงทะเบียนเรียน ข้อมูลผลการเรียนของนักศึกษาที่เป็นสมาชิกห้องสมุดจากฐานข้อมูลงานทะเบียนนักศึกษา ช่วงเวลาตั้งแต่วันที่ 1 มิถุนายน 2553 ถึง 31 ธันวาคม 2556 โดยมีขั้นตอนการดำเนินการดังนี้

1) จัดเตรียมข้อมูล ตามความต้องการที่ได้จากการวิเคราะห์ความต้องการและขอบเขตงานของการวิจัย ซึ่งข้อมูลเดิมอยู่ที่โปรแกรมห้องสมุดอัตโนมัติ WALAI AUTOLIB โดยจัดเก็บในฐานข้อมูล ซึ่งประกอบไปด้วย 1) ข้อมูลสมาชิก ที่เป็นข้อมูลทั่วไปของนักศึกษาที่เข้ามาใช้บริการห้องสมุดประกอบด้วย เพศ รหัสนักศึกษา ชื่อสมาชิก นามสกุลสมาชิก โปรแกรมวิชา 2) ข้อมูลหนังสือหรือสารสนเทศ ที่เป็นข้อมูลทั่วไปของหนังสือหรือสารสนเทศที่มีให้บริการของห้องสมุด ประกอบด้วย ชื่อหนังสือ รหัสหนังสือ ประเภทหนังสือ 3) ข้อมูลการใช้บริการยืมสารสนเทศ ที่เป็นข้อมูลการใช้บริการยืมสารสนเทศภายในห้องสมุด ประกอบด้วย รหัสสมาชิก รหัสหนังสือ วันที่ยืม วันที่ยืม เดือนที่ยืม จำนวนเล่มหนังสือที่ยืม รหัสการยืม 4) ข้อมูลนักศึกษาจากฐานข้อมูลงานทะเบียนที่ประกอบด้วย รหัสนักศึกษา ชื่อนักศึกษา เพศ เกรดเฉลี่ย ซึ่งได้ Export ข้อมูลออกมาเป็นไฟล์ Excel

2) นำข้อมูลจากไฟล์ Excel เข้าสู่คลังข้อมูลที่สร้างไว้ สำหรับใช้เป็นระบบแหล่งข้อมูลของงานวิจัย (source system) โดยการนำเข้าข้อมูล (import) ตามการจัดเตรียมโครงสร้างข้อมูลที่ออกแบบไว้ และดำเนินการใช้กระบวนการจัดการกับ ข้อมูล ซึ่งใช้เครื่องมือที่เป็นซอฟต์แวร์ SQL Server Integration Services (SSIS) เพื่อนำข้อมูลต่างๆ เข้าสู่ที่พักข้อมูล (staging area) โดยใช้ชื่อว่า “ETL Excel”



ภาพที่ 3.4 การนำข้อมูลจากไฟล์ Excel เข้าตาราง

3.1.2 ขั้นตอนการแปลงข้อมูล (Data transformation) และการทำความสะอาดข้อมูล (data cleansing) คือการนำข้อมูลที่ได้มาจากการสกัด (extract) มาจัดฟอร์แมตให้ถูกต้องสอดคล้องกันด้วยวิธีการต่างๆ เช่น การแมปข้อมูล (Data Mapping) เป็นการนำข้อมูลที่มีความหมายเดียวกันแต่อยู่ในรูปแบบที่แตกต่างกันให้อยู่ในรูปแบบเดียวกัน โดยมีกระบวนการ Data Cleansing เพื่อการตรวจสอบและแก้ไขข้อมูลให้ถูกต้องซึ่งในส่วนนี้มีการดำเนินการดังนี้

1) **คัดเลือกข้อมูลที่ได้จากที่พักข้อมูล** โดยจะมีการคัดเลือกข้อมูลที่มีความสอดคล้องกับงานวิจัยว่าจะต้องดำเนินการรวบรวมข้อมูลจากที่พักข้อมูลในส่วนใดบ้างมาเตรียมข้อมูลไว้ เพื่อดำเนินการต่อในส่วนของการทำ ETL เป็นการเตรียมข้อมูลก่อนการนำข้อมูลเข้าสู่คลังข้อมูล โดยใช้ความสามารถของ Microsoft SQL 2008 Integration Service มาดำเนินการในส่วนนี้

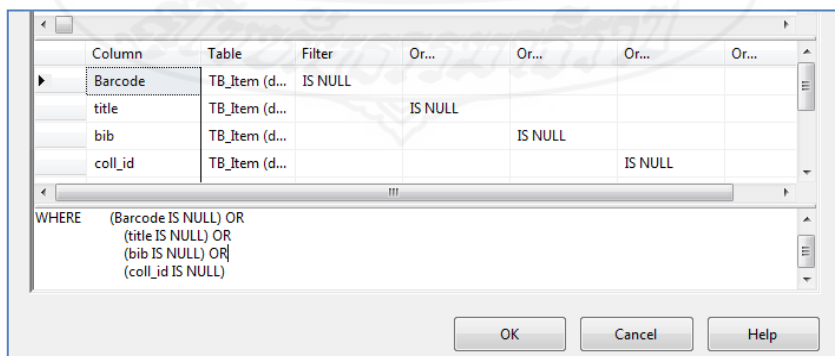
2) **เมื่อคัดเลือกข้อมูลได้ตามความต้องการแล้ว** จะพักข้อมูลไว้ในที่พักของข้อมูลจากนั้นจะนำข้อมูล มาดำเนินการตรวจสอบเพื่อหาค่าที่ไม่ถูกต้องหรือที่เรียกว่าการทำความสะอาดข้อมูล (data cleansing) โดยมีการดำเนินการคือ คัดเลือกข้อมูลที่ถูกต้องไปไว้ในฐานข้อมูล และแยกข้อมูลที่ไม่ถูกต้องออกมาวิเคราะห์เพื่อกำจัดความไม่ถูกต้อง การทำความสะอาดข้อมูลก่อนที่จะนำข้อมูลเข้าสู่คลังข้อมูลผู้วิจัยได้ทำการวิเคราะห์ข้อมูลเพื่อกำจัดความไม่ถูกต้องของข้อมูลโดยมี

กระบวนการแก้ไขข้อมูลให้ถูกต้องสมบูรณ์ ได้แก่ การแก้ไขค่าว่างของข้อมูลให้สมบูรณ์ และจัดการแก้ไขข้อมูลที่ไม่ถูกต้องตรงตามข้อมูลอ้างอิง โดยเปรียบเทียบค่าที่ไม่สมบูรณ์ให้ตรงกับข้อมูลที่กำหนดไว้หรือลบข้อมูลที่มีปัญหานั้นออกไป เพื่อให้ข้อมูลไม่เกิดความผิดพลาดเวลานำข้อมูลเข้าสู่ระบบ ในงานวิจัยนี้พบปัญหาในส่วนของความไม่สมบูรณ์หรือความไม่ถูกต้องของข้อมูล มีเพียงตารางเดียวคือ ตาราง TB_Item ที่ใช้เก็บข้อมูลหนังสือหรือสารสนเทศภายในห้องสมุด ซึ่งมีข้อมูลที่บันทึกรายละเอียดไม่สมบูรณ์ เป็นข้อมูลจำนวนทั้งสิ้น 31 รายการที่มีข้อมูลเป็นค่าว่างหรือไม่สามารถระบุข้อมูลได้ (เป็นค่า NULL) ดังที่แสดงในภาพที่ 3.5

Barcode	title	bib	coll_id
i00002111	"ระบบปฏิบัติการยูนิกซ์ / อภิศาดิ อิศวาศิศยงกร และ Ad...	NULL	NULL
Engineering Group ."	b00001201	0	NULL
i00002112	"ระบบปฏิบัติการยูนิกซ์ / อภิศาดิ อิศวาศิศยงกร และ Ad...	NULL	NULL
Engineering Group ."	b00001201	0	NULL
i00002114	"ระบบปฏิบัติการยูนิกซ์ / อภิศาดิ อิศวาศิศยงกร และ Ad...	NULL	NULL
Engineering Group ."	b00001201	0	NULL
i00005509	"อิริตหลังความตาย : \$	NULL	NULL
i00005511	"อิริตหลังความตาย : \$	NULL	NULL
i00005513	"อิริตหลังความตาย : \$	NULL	NULL
i00005514	"อิริตหลังความตาย : \$	NULL	NULL
i00005650	"อิริตหลังความตาย : \$	NULL	NULL
i00005651	"อิริตหลังความตาย : \$	NULL	NULL

ภาพที่ 3.5 แสดงรายการข้อมูลที่เป็นค่าว่างหรือไม่สามารถระบุข้อมูลได้ (เป็นค่า NULL)

เมื่อตรวจสอบพบข้อมูลที่มีปัญหา จึงได้นำข้อมูลไปตรวจสอบพบว่าเป็นข้อมูลที่มีการจัดเก็บผิดพลาด ไม่สามารถดำเนินการแก้ไขได้ ดังนั้นเพื่อให้ระบบสามารถทำงานได้อย่างถูกต้องและมีประสิทธิภาพจึงได้ดำเนินการลบข้อมูลที่มีปัญหาดังกล่าวออกจากระบบ โดยใช้คำสั่ง “DELETE FROM dbo.TB_Item WHERE (Barcode IS NULL) OR (title IS NULL) OR (bib IS NULL) OR (coll_id IS NULL) ” ดังที่แสดงในภาพที่ 3.6



ภาพที่ 3.6 คำสั่งลบข้อมูลที่เป็นค่าว่างหรือไม่สามารถระบุข้อมูลได้ (เป็นค่า NULL)

3.1.3 ขั้นตอนการนำเข้าข้อมูล (Data loading) คือการนำข้อมูลที่ผ่านการแปลงและตรวจสอบความถูกต้องแล้วเข้าสู่คลังข้อมูล เป็นการดำเนินการนำข้อมูลจากที่พักข้อมูลที่มีการแปลงข้อมูลให้ถูกต้องตามโครงสร้างของคลังข้อมูล และ มีการทำความสะอาดข้อมูลที่ไม่ถูกต้องหรือไม่ถูกโครงสร้างเรียบร้อยแล้วเพื่อดำเนินการนำข้อมูลขึ้นระบบไว้ในคลังข้อมูลขั้นตอนในการดำเนินการในส่วนนี้ประกอบไปด้วย

1) การพิจารณานำข้อมูลจากที่พักข้อมูล มาเข้าในคลังข้อมูลโดยดำเนินการส่วนที่เรียกว่าการทำ Source System Mapping คือกระบวนการในการพิจารณาว่าข้อมูลที่จะนำลงในคลังข้อมูล นั้นต้องนำมาจากตารางใดบ้าง

2) การดำเนินการนำข้อมูลเข้าสู่คลังข้อมูลที่ชื่อว่า (DWLIBRARY) โดยการพิจารณาข้อมูลจากการวิเคราะห์ในส่วนของการทำ Source System Mapping โดยดำเนินการสร้างฟังก์ชันในการนำเข้าข้อมูล

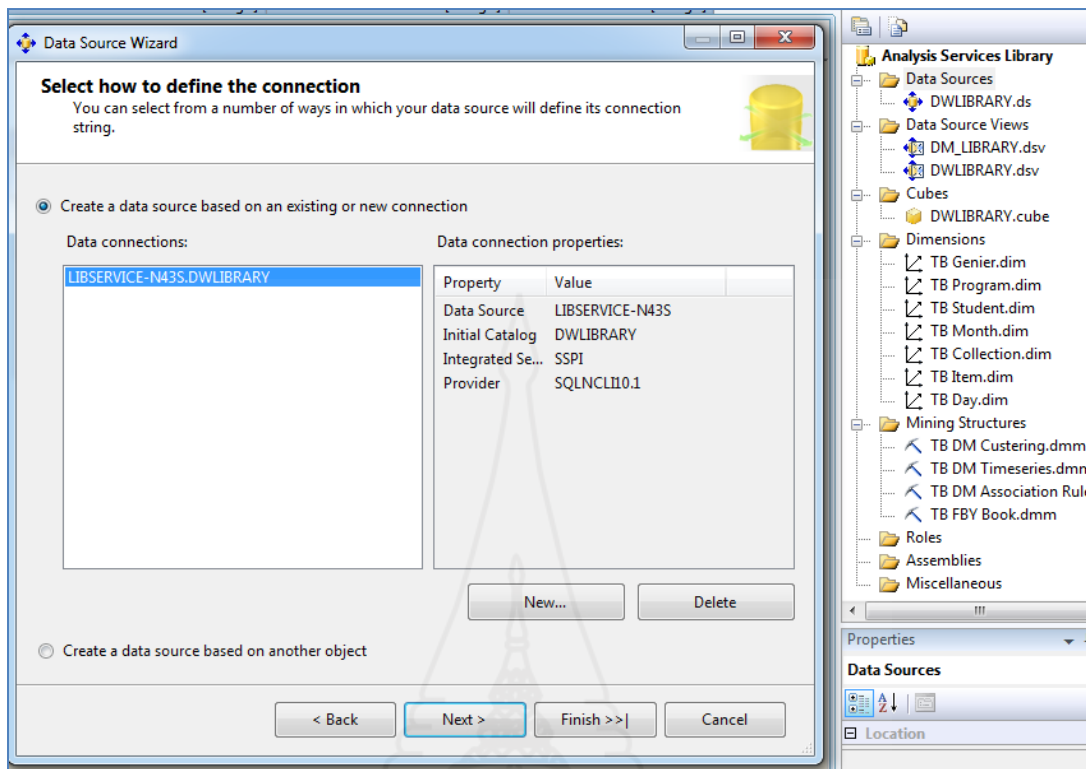
3.2 การทำ Online Analytical Processing (OLAP)

หลังจากสร้างคลังข้อมูล (Data warehouse) ชื่อ DWLIBRARY ด้วย Microsoft SQL Server Management Studio แล้ว จึงใช้ Microsoft SQL Server 2008 Analysis Services (SSAS) ซึ่งเป็นเครื่องมือสำหรับการประมวลผลเชิงวิเคราะห์แบบออนไลน์หรือโอแลป (Online Analytical Processing – OLAP) โดยดำเนินการดังนี้

3.2.1 สร้างโปรเจกต์ (Project) โดยการเลือกแบบเทมเพลต (Templates) เป็น Analysis Services Project กำหนดชื่อโปรเจกต์ กำหนดโฟลเดอร์ที่ใช้เก็บข้อมูลโปรเจกต์ และกำหนดไดเรกทอรีสำหรับเก็บโซลูชัน (solution) ของโปรเจกต์

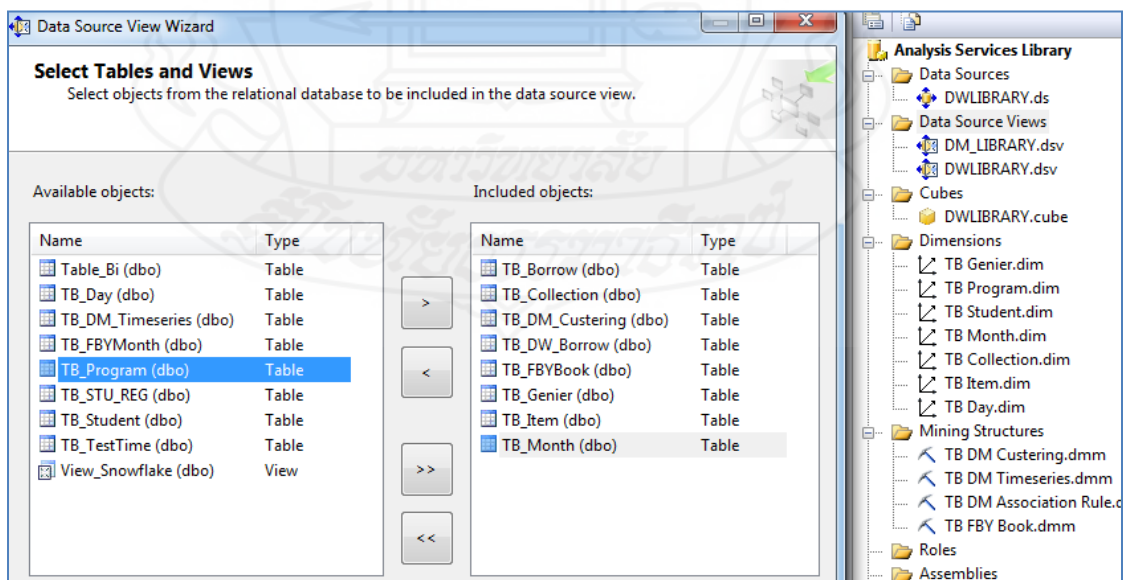
3.2.2 สร้าง Data Source ชื่อว่า DWLIBRARY.ds โดยการระบุ Server Name และฐานข้อมูลที่ต้องการ ดังภาพที่ 3.7





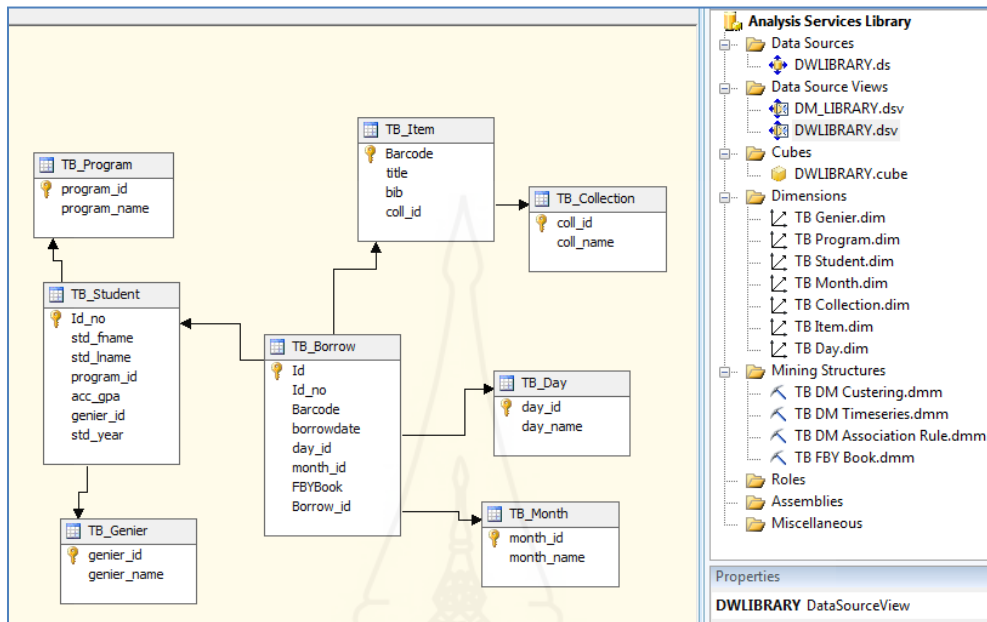
ภาพที่ 3.7 สร้าง Data Source ชื่อว่า DWLIBRARY.ds

3.2.3 สร้าง Data Source View ชื่อว่า DWLIBRARY.dsv โดยระบุตารางต่างๆ จากฐานข้อมูลที่เลือกไว้ ดังภาพที่ 3.8



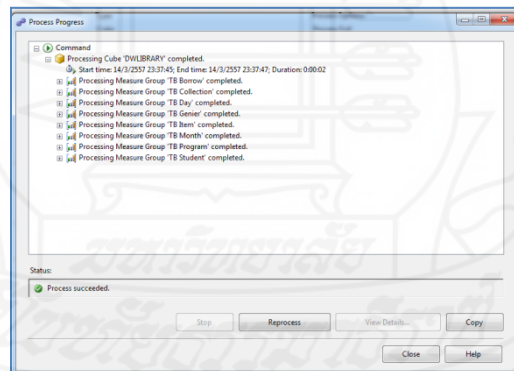
ภาพที่ 3.8 สร้าง Data Source View ชื่อว่า DWLIBRARY.dsv

3.2.4 สร้าง Cube เพื่อกำหนด Measures และ Dimensions ดังภาพที่ 3.9



ภาพที่ 3.9 สร้าง Cube

3.2.5 Deploy โปรเจกต์ ดังภาพที่ 3.10



ภาพที่ 3.10 Deploy โปรเจกต์

3.2.6 เรียกดูข้อมูลจากการประมวลผลเชิงวิเคราะห์แบบออนไลน์ โดยผ่านแท็บ เบราเซอร์โดยเลือก Measures และ Dimensions มาแสดงผลในลักษณะหลายมิติได้ตามต้องการ ดังภาพที่ 3.11 ทั้งนี้หากมีการเปลี่ยนแปลงโครงสร้างใดๆ ใน Data Source, Data Source View หรือ Cube ควร deploy อีกครั้งเพื่ออัปเดต Cube

Program Id	Genier Id	Id No	TB Borrow Count
4101			5374
4111			9818
4121			13436
4122			5390
4126			7749
4130			8
4131			16858
4132			9985
4135			16941
5161			54
5166			1494
6161			2352
6162			242
6171			1191
6352			363
6353			5
6414			898
6601			72
Grand Total			120372

ภาพที่ 3.11 แสดงผลในลักษณะหลายมิติ

จากภาพที่ 3.11 การแสดงผลในลักษณะหลายมิติ ผู้วิจัยแสดงผลข้อมูลจำนวนการยืมหนังสือของแต่ละโปรแกรมได้อย่างง่ายดาย โดยการเลือกข้อมูลจากตาราง TB_Borrow และข้อมูลโปรแกรมต่างๆ จากตาราง TB_Program มาแสดงผล ทั้งนี้การแสดงผลในลักษณะหลายมิติยังสามารถแสดงผลข้อมูลต่างๆ จากตารางที่มีมาแสดงผลข้อมูลได้ เช่น การดึงข้อมูลเพศจากตาราง TB_Genier และตาราง TB_Student ในการแสดงผลข้อมูลการยืมหนังสือของนักศึกษาแต่ละโปรแกรมมีเพศหญิงกี่คนและเพศชายกี่คน พร้อมทั้งแสดงข้อมูลรหัสนักศึกษาเพื่อให้ทราบว่านักศึกษาแต่ละคนมายืมหนังสือกี่ครั้ง ดังภาพที่ 3.12

Program Id	Genier Id	Id No	TB Borrow Count
2001	1		37
	2	5612001254	1
		5612001349	2
		5612001408	3
		5612001412	1
		5612001431	1
		5622001121	1
		Total	9
Total			46
2122			90
2211			11
2240	1		20
	2		8
	Total		28
2252	1	5012252104	1
		5012252106	4

ภาพที่ 3.12 แสดงข้อมูลรหัสนักศึกษาเพื่อให้ทราบว่านักศึกษาและคนมายืมหนังสือกี่ครั้ง

3.3 การทำเหมืองข้อมูล

การทำเหมืองข้อมูลคือการค้นหาความรู้ที่ไม่เคยทราบมาก่อน สำหรับงานวิจัยนี้ ขั้นตอนการทำเหมืองข้อมูลดำเนินการตามแบบจำลองของคริสป์ - ดีเอ็ม (CRISP - DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.3.1 ทำความเข้าใจปัญหา ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องและข้อมูลเกี่ยวกับการค้นคืนสารสนเทศภายในห้องสมุด และวิเคราะห์ปัญหาที่เกิดขึ้นจริงกับการปฏิบัติงานของห้องสมุด ดังนี้

1) สารสนเทศไม่เพียงพอต่อความต้องการของผู้ใช้ ในบางช่วงเหตุการณ์ เช่น ช่วงเวลาสอบ หนังสือไม่เพียงพอต่อการให้บริการแก่ผู้ใช้บริการ เพราะไม่สามารถพยากรณ์จำนวนนักศึกษาที่มาใช้บริการห้องสมุดในแต่ละช่วงเวลาได้ ทำให้ขาดข้อมูลประกอบในการปรับปรุงและเตรียมการบริการ

2) นักศึกษาค้นหาหนังสือไม่ตรงตามความต้องการ เพราะต้องระบุหัวเรื่องในการค้นหา หากมีการแบ่งกลุ่มนักศึกษาออกเป็นกลุ่มๆ จะทำให้สามารถค้นหาหนังสือที่น่าสนใจในแต่ละกลุ่มได้และนักศึกษาได้หนังสือที่ตรงตามความต้องการอีกด้วย

3) หากนักศึกษาต้องการหนังสือหลายเล่มต้องเสียเวลาในการสืบค้นข้อมูลหลายรอบ จึงทำให้เสียเวลามากในการค้นหาข้อมูลหนังสือ

3.3.2 ทำความเข้าใจข้อมูล ผู้วิจัยดำเนินการรวบรวมข้อมูลที่เกี่ยวข้องเพื่อนำไปวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งมีด้วยกันหลายอัลกอริทึม และแต่ละอัลกอริทึมก็มีวัตถุประสงค์ที่ต่างกัน จึงต้องเริ่มจากการทำความเข้าใจกับข้อมูลที่เกี่ยวข้องในคลังข้อมูล เพื่อวิเคราะห์ว่าควรมีแอตทริบิวต์ใดที่ควรนำมาวิเคราะห์ด้วยอัลกอริทึมนั้นๆ ซึ่งจะทำให้ได้มาซึ่งความรู้ตามวัตถุประสงค์ที่กำหนดไว้ จากการศึกษาพบว่า ข้อมูลที่จัดเก็บอยู่ในฐานข้อมูลโปรแกรมห้องสมุดนั้นเป็นการจัดเก็บข้อมูลการยืมคืน ข้อมูลหนังสือหรือสารสนเทศ และข้อมูลสมาชิกที่เก็บไว้เพียงรหัสนักศึกษาหรือรหัสสมาชิกแค่นั้นเอง ผู้วิจัยจึงดำเนินการขอข้อมูลส่วนตัวของนักศึกษาหรือสมาชิกจากฐานข้อมูลงานทะเบียนของมหาวิทยาลัยราชภัฏศรีสะเกษ ซึ่งมีข้อมูลทั้งรหัสนักศึกษา ชื่อนามสกุล เพศ โปรแกรมวิชา เกรดเฉลี่ย

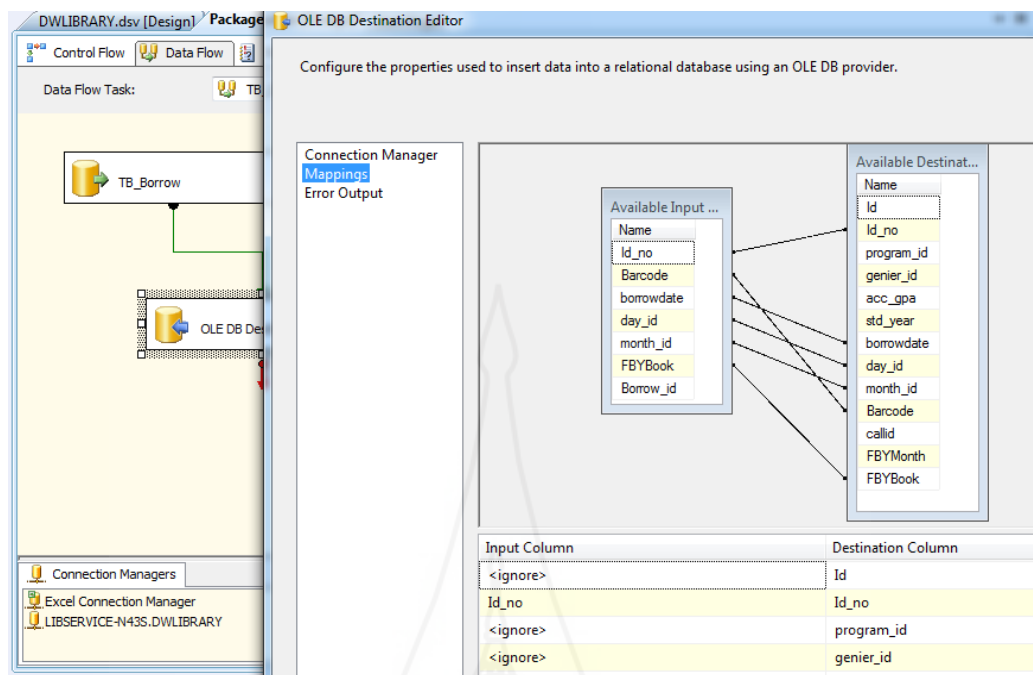
3.3.3 เตรียมข้อมูล เมื่อสร้างคลังข้อมูล (Data Warehouse) ผู้วิจัยได้นำข้อมูลที่มีการจัดเก็บไว้มาปรับโครงสร้างที่เหมาะสมเพื่อใช้กับแบบจำลองการทำเหมืองข้อมูลตามอัลกอริทึม เพราะข้อมูลที่ได้จากที่ใช้ในงานวิจัย ซึ่งประกอบด้วย 3 อัลกอริทึม คือ 1) คลัสเตอริง (Clustering) 2) แอสโซซิเอชันรูลส์ (Association Rules) และ 3) ไทม์ซีรีส์ (Time Series) ดังนั้น ในแต่ละอัลกอริทึมจึงต้องมีการเตรียมข้อมูลหรือดาต้ามาร์ตให้เหมาะสมกับวัตถุประสงค์การใช้งานของแต่ละอัลกอริทึม โดยเริ่มจากการคัดเลือกข้อมูลที่ต้องการใช้งานมาไว้ในดาต้ามาร์ต โดยการรวมกันของหลายๆ ตารางหรือหลายๆ ระเบียบ โดยในขั้นตอนนี้ผู้วิจัยได้ดำเนินการใช้กระบวนการอีทีแอล (Extract Transform and Load : ETL) ที่สร้างขึ้นด้วยเครื่องมือ Integration Service ของ Microsoft SQL Server มาประยุกต์ เพื่อปรับโครงสร้างให้เหมาะสมกับแบบจำลองการทำเหมืองข้อมูลของแต่ละอัลกอริทึม ซึ่งมีรายละเอียดดังต่อไปนี้

1) การจัดกลุ่ม (Clustering) เป็นการจัดการข้อมูลเพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ ซึ่งแอตทริบิวต์ที่ผู้วิจัยนำมาวิเคราะห์ด้วยอัลกอริธึมนี้ ประกอบด้วยแอตทริบิวต์ต่างๆ ดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 แอตทริบิวต์ที่ใช้ในการวิเคราะห์การจัดกลุ่ม

Col.	Name	Description
1	Id	ลำดับการยืม
2	Id_no	รหัสนักศึกษา
3	program_id	โปรแกรมวิชา
4	genier_id	เพศ
5	acc_gpa	เกรดเฉลี่ย
6	std_year	นักศึกษาชั้นปี
7	Barcode	รหัสหนังสือ
8	callid	หมวดหมู่
9	borrowdate	วันที่ยืม
10	day_id	วันยืม
11	month_id	เดือนยืม
12	FBYMonth	ความถี่ในการยืมแต่ละเดือน
13	FBYBook	จำนวนหนังสือในการยืมแต่ละครั้ง

จากตาราง 3.1 แอตทริบิวต์ที่ใช้ในการวิเคราะห์การจัดกลุ่มเพื่อให้ได้ข้อมูลตรงตามความต้องการ ผู้วิจัยได้สร้างตาราง TB_DM_Custering ขึ้นมาและนำข้อมูลเข้าตารางโดยการแมปข้อมูล (Data Mapping) ตารางการยืมหนังสือ (TB_Borrow) ตารางหนังสือ (TB_Item) และตารางข้อมูลนักศึกษาหรือสมาชิก (TB_Student) ซึ่งใช้เครื่องมือที่เป็นซอฟต์แวร์ SQL Server Integration Services (SSIS) ดังภาพที่ 3.13



ภาพที่ 3.13 แสดงการนำข้อมูลเข้าตาราง TB_DM_Custering โดยการแม็ปข้อมูล

2) กฎการหาความสัมพันธ์ (Association Rule) เป็นการจัดการข้อมูลเพื่อวิเคราะห์ข้อมูลการยืมหนังสือว่า มีหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยด้วยกัน ซึ่งแอตทริบิวต์ที่ผู้วิจัยนำมาวิเคราะห์ด้วยอัลกอริทึมนี้ประกอบด้วยแอตทริบิวต์ต่างๆ ดังแสดงในตารางที่ 3.2 และ ตารางที่ 3.3

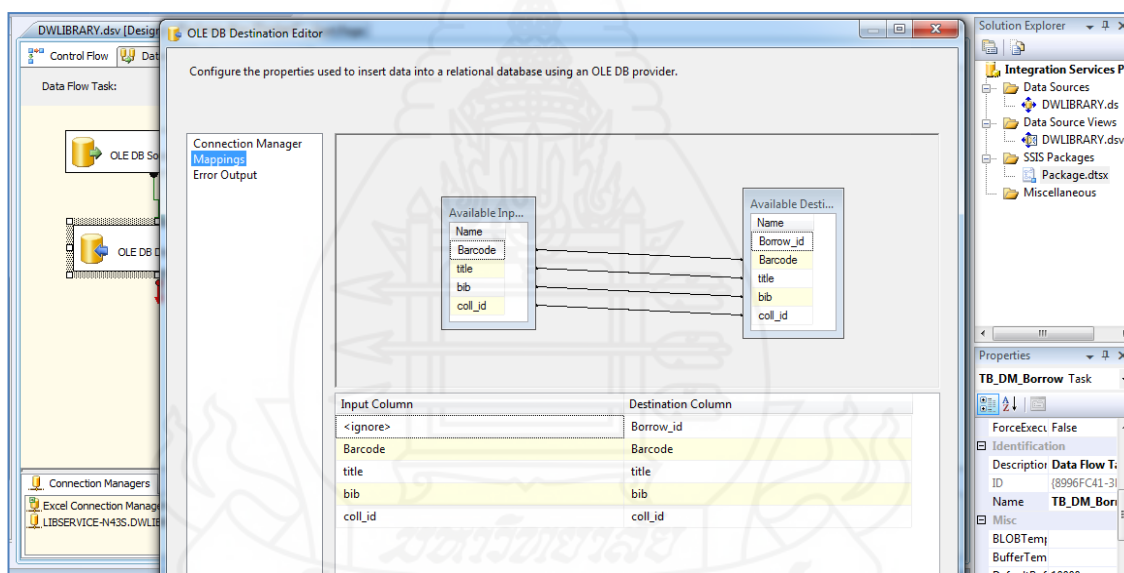
ตารางที่ 3.2 แอตทริบิวต์ตารางข้อมูลการยืมของเล่มหนังสือ

Col.	Name	Description
1	Id	ลำดับการยืม
2	Borrow_id	รหัสลำดับการยืม
3	Barcode	รหัสเล่มหนังสือ
4	Title	ชื่อหนังสือหนังสือ
5	Bib	รหัสชื่อหนังสือ
6	Coll_id	หมวดหนังสือ

ตารางที่ 3.3 แอ็ททริบิวต์ตารางข้อมูลการยืมของสมาชิก

Col.	Name	Description
1	Id_no	รหัสสมาชิก
2	Borrow_id	รหัสลำดับการยืม
3	Program_id	รหัสโปรแกรมที่สังกัด
4	Genier_id	รหัสเพศ

จากตารางที่ 3.3 แอ็ททริบิวต์ที่ใช้ในการวิเคราะห์การจัดกลุ่มเพื่อให้ได้ข้อมูลตรงตามความต้องการ ผู้วิจัยได้สร้างตาราง TB_DM_Borrow และตาราง TB_FBYBook ขึ้นมาและนำข้อมูลเข้าตารางโดยการแม็ปข้อมูล (Data Mapping) ตารางการยืมหนังสือ (TB_Borrow) ซึ่งใช้เครื่องมือที่เป็นซอฟต์แวร์ SQL Server Integration Services (SSIS) ดังภาพที่ 3.14



ภาพที่ 3.14 แสดงการนำข้อมูลเข้าตาราง TB_DM_Borrow โดยการแม็ปข้อมูล

3) *ไทม์ซีรีส์ (Time Series)* เป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคตโดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาในอนาคต ซึ่งแอ็ททริบิวส์ที่ผู้วิจัยนำมาวิเคราะห์ด้วยอัลกอริธึมนี้ประกอบด้วยแอ็ททริบิวส์ต่างๆ ดังแสดงในตารางที่ 3.4

ตารางที่ 3.4 แอ็ททริบิวส์ที่ใช้ในการวิเคราะห์การจัดกลุ่ม

Col.	Name	Description
1	Timeindex	เดือน (มี.ย.53 – ธ.ค.56)
2	Year_id	ปี
3	Month_id	เดือน
4	Quntity	จำนวนการยืมทั้งหมด
5	Qun_PromgramA	จำนวนการยืมของโปรแกรม
6	Qun_PromgramB	จำนวนการยืมของโปรแกรม
7	Qun_PromgramC	จำนวนการยืมของโปรแกรม

จากตารางที่ 3.4 แอ็ททริบิวส์ที่ใช้ในการวิเคราะห์การจัดกลุ่มเพื่อให้ได้ข้อมูลตรงตามความต้องการผู้วิจัยได้สร้างตาราง TB_DM_Timeseries ขึ้นมาและใช้คำสั่ง SQL ในการดึงข้อมูลแล้วนำข้อมูลที่ได้ออกสู่ตาราง TB_DM_Timeseries ดังภาพที่ 3.15

	Timeindex	Year_id	Month_id	Quntity	Qun_PromgramA	Qun_PromgramB	Qun_PromgramC
1	255306	2010	6	1654	168	335	380
2	255307	2010	7	1650	106	194	303
3	255308	2010	8	2478	170	267	517
4	255309	2010	9	1154	80	126	227
5	255310	2010	10	2057	159	272	568
6	255311	2010	11	2084	137	220	508
7	255312	2010	12	2183	205	354	579
8	255401	2011	1	3808	782	605	1228
9	255402	2011	2	556	14	28	44
10	255403	2011	3	607	12	46	17
11	255404	2011	4	541	34	18	18
12	255405	2011	5	468	47	41	53
13	255406	2011	6	5916	651	1361	1983
14	255407	2011	7	5680	550	1563	1231

ภาพที่ 3.15 แสดงข้อมูลตาราง TB_DM_Timeseries

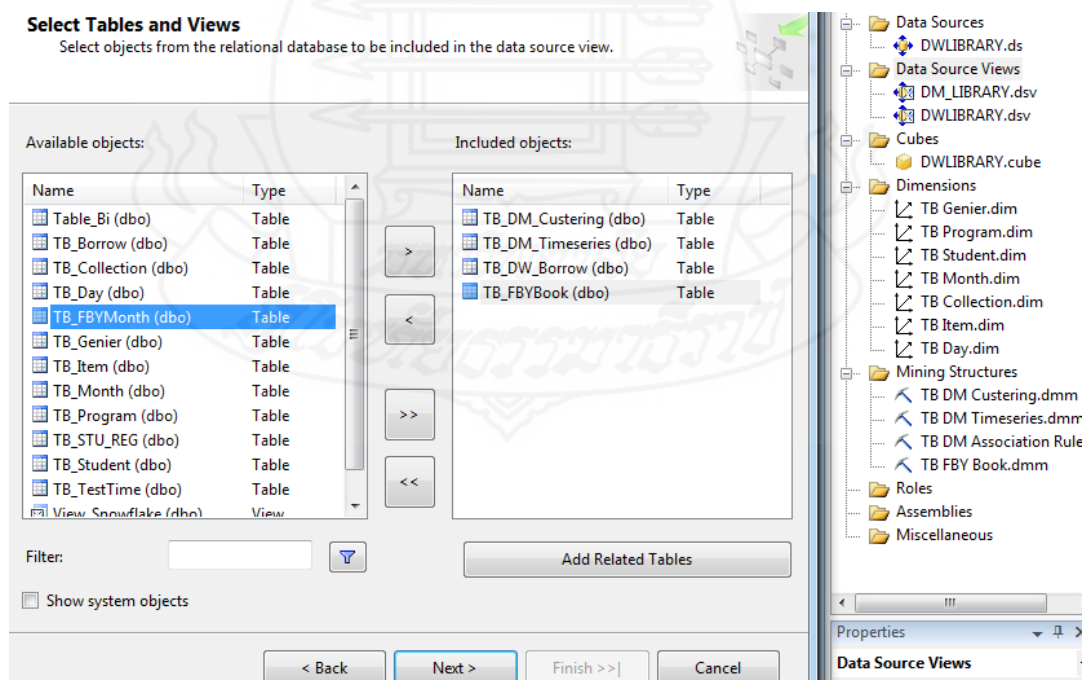
3.3.4 สร้างแบบจำลอง เป็นขั้นตอนของการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ได้แก่การสร้างแบบจำลองที่สอดคล้องกับอัลกอริทึมแต่ละชนิด ซึ่งจากการศึกษาและทำความเข้าใจกับข้อมูลแล้ว ผู้วิจัยได้เลือกอัลกอริทึมที่จะนำมาประมวลผล 3 อัลกอริทึม คือการจัดกลุ่ม (Clustering) , กฎการหาความสัมพันธ์ (Association Rule) , อนุกรมเวลา (Time Series) ขั้นตอนในการประมวลผลโดยใช้เทคนิคดาต้าไมนิง ด้วยซอฟต์แวร์ Microsoft SQL Server 2008 มีดังต่อไปนี้

1) สร้างโปรเจกต์ (Project) โดยเลือกแบบเทมเพลต (Templates) เป็น Analysis Project กำหนดชื่อโปรเจกต์ กำหนดโฟลเดอร์ที่ใช้เก็บข้อมูลโปรเจกต์ และกำหนดไดเรกทอรีสำหรับเก็บโซลูชัน (solution) ของโปรเจกต์

2) กำหนดชื่อแหล่งข้อมูลใหม่ โดยการคลิกขวาที่โฟลเดอร์ Data Sources แล้วเลือก New Data Sources โดยตั้งชื่อว่า DWLIBRARY.ds เป็นแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล

3) เปิดดูรายละเอียดข้อมูลที่อยู่ในแหล่งข้อมูลที่กำหนดมาก่อนในขั้นตอนที่แล้ว เพื่อทำการเลือกตารางหรือวิวของข้อมูลที่ต้องการนำมาใช้ในการทำเหมืองข้อมูล โดยคลิกขวาที่โฟลเดอร์ Data Sources View แล้วเลือก New Data Source View จะมีรายชื่อตารางและวิวทั้งหมดที่อยู่ในแหล่งข้อมูลที่กำหนดไว้ปรากฏขึ้นมา โดยตั้งชื่อว่า DM_LIBRARY.dsv

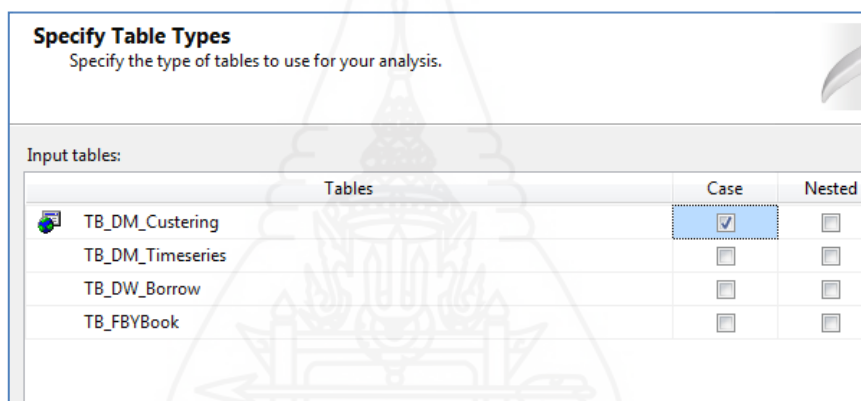
4) เลือกตารางหรือวิวที่ต้องการนำมาประมวลผล ด้วยเทคนิคดาต้าไมนิง คือตาราง TB_DM_Custering ตาราง TB_DM_Timeseries ตาราง TB_DM_Borrow และตาราง TB_FBYBook ดังภาพที่ 3.16



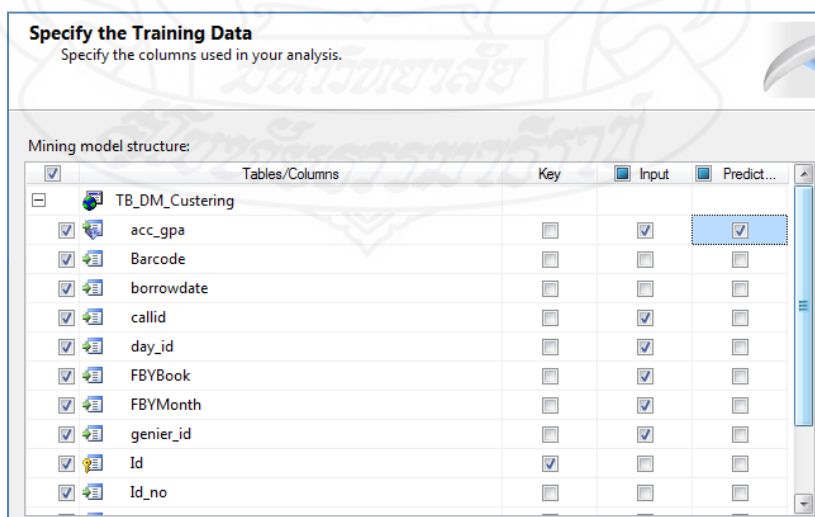
ภาพที่ 3.16 เลือกตารางหรือวิวที่ต้องการนำมาประมวลผลด้วยเทคนิคดาต้าไมนิง

5) เลือกอัลกอริธึมที่จะใช้ประมวลผล โดยเลือกจากรายชื่ออัลกอริธึมที่ปรากฏโดยการคลิกขวาที่โพลเดอร์ Mining Structure คลิกเลือก New Mining Structure และในแต่ละอัลกอริธึมจะต้องกำหนดว่า แอตทริบิวต์ใดเป็นส่วนของ key ส่วนของ input และส่วนของ predict นอกจากนี้แล้วยังสามารถกำหนดค่าพารามิเตอร์ (parameter) ที่เกี่ยวข้องได้ด้วย ซึ่งต้องดำเนินการที่ละอัลกอริธึม ให้ครบทั้งหมด ดังนี้

6) การจัดกลุ่ม (Clustering) เพื่อแบ่งกลุ่มของสมาชิกที่มีลักษณะของการยืมหนังสือคล้ายกัน หรือมีข้อมูลทั่วไปที่เหมือนกันออกเป็นกลุ่มๆ โดยการเลือกจากราย TB_DM_Custering ที่ใช้ในการประมวลผลในอัลกอริธึมการจัดกลุ่ม (Clustering) ดังภาพที่ 3.17 และให้กำหนดแอตทริบิวต์ ID เป็นคีย์ (key) ซึ่งเป็นค่าที่เก็บลำดับการยืมหนังสือ ให้กำหนดแอตทริบิวต์ acc_gpa เป็นอินพุต (input) และพยากรณ์ (prediction) ซึ่งเป็นผลการเรียนนักศึกษาที่มีการยืมหนังสือ ดังภาพที่ 3.18

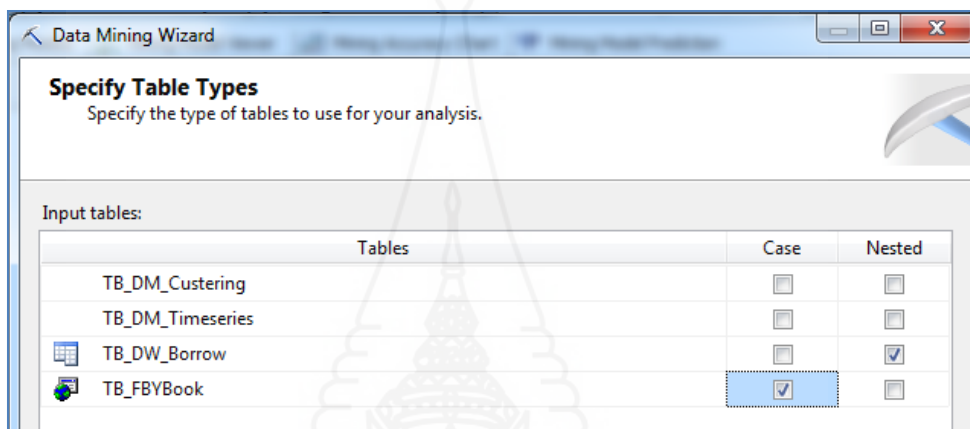


ภาพที่ 3.17 เลือกตารางที่ใช้ในการประมวลผลในอัลกอริธึมการจัดกลุ่ม (Clustering)

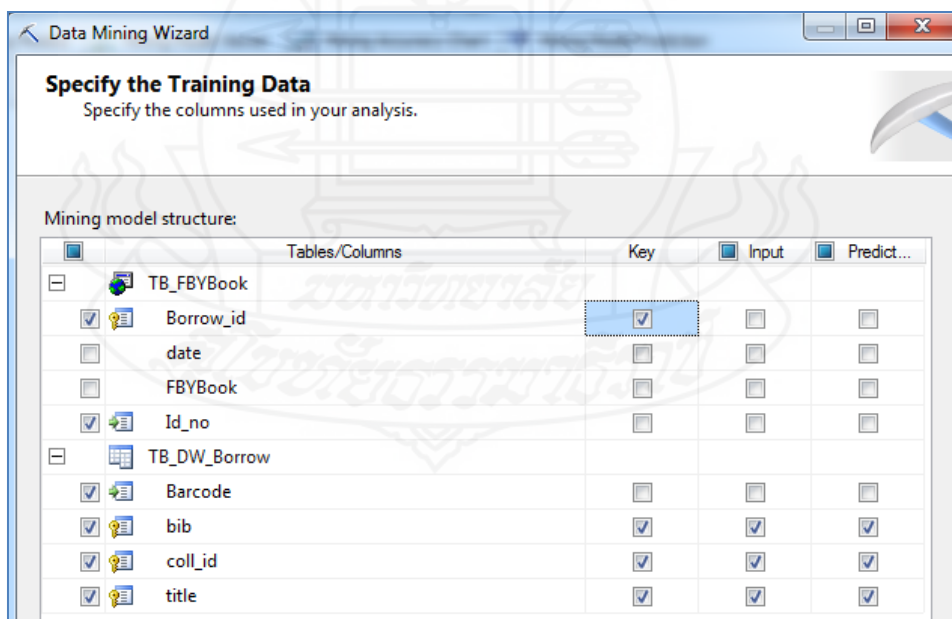


ภาพที่ 3.18 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึมการจัดกลุ่ม (Clustering)

7) กฎการหาความสัมพันธ์ (Association Rule) เพื่อวิเคราะห์ข้อมูลการยืมหนังสือ หรือค้นหารูปแบบความสัมพันธ์ของพฤติกรรมกรยืม - คืนหนังสือของสมาชิก เพื่อหาความน่าจะเป็นในการยืมหนังสือว่าเล่มใดบ้างที่มีถูกยืมไปด้วยด้วยกัน โดยการเลือกตาราง TB_DM_Borrow และตาราง TB_FBYBook ที่ใช้ในการประมวลผลในอัลกอริธึม Apriori ดังภาพที่ 3.19 และให้กำหนดแอตทริบิวต์ Borrow_id เป็นคีย์ (key) ซึ่งเป็นค่าที่เก็บรหัสการยืมหนังสือ ให้กำหนดแอตทริบิวต์ bib เป็นอินพุต (input) และพยากรณ์ (prediction) ซึ่งเป็นค่ารหัสหนังสือที่ใช้แสดงผล ดังภาพที่ 3.20

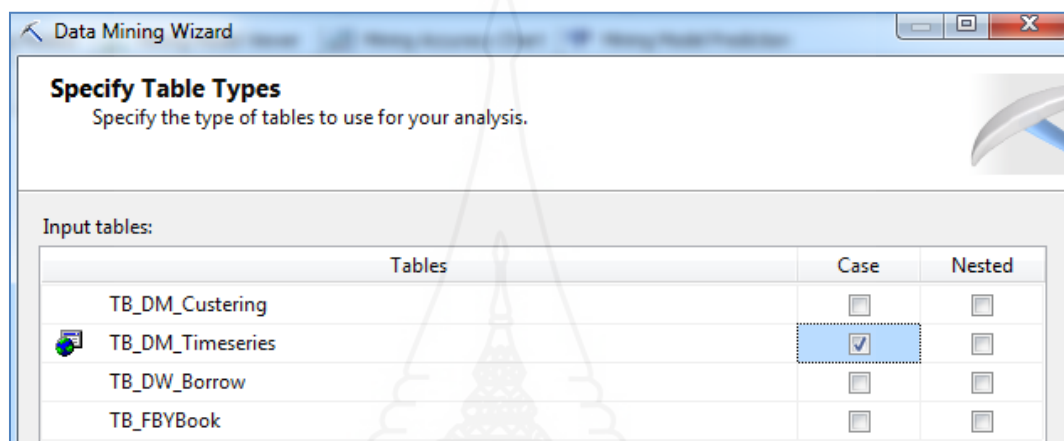


ภาพที่ 3.19 การเลือกตาราง TB_DM_Borrow และตาราง TB_FBYBook

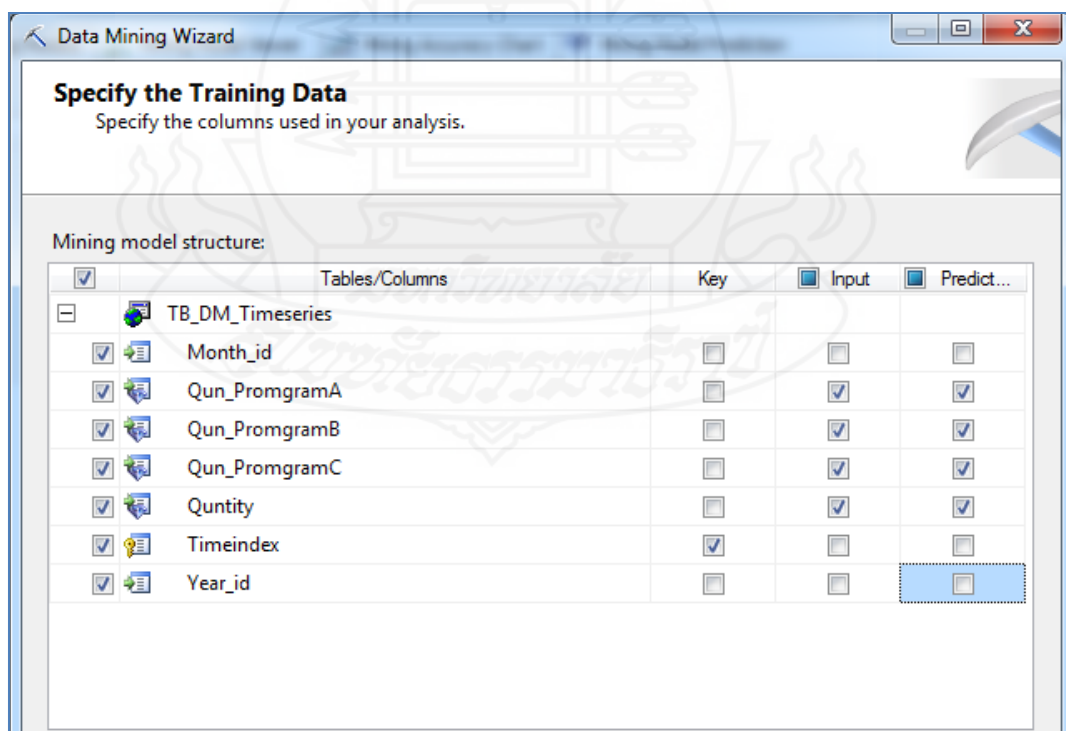


ภาพที่ 3.20 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึม Apriori

8) *ไทม์ซีรีส์ (Time Series)* หรืออนุกรมเวลาเป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคต โดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษา โดยการเลือกตาราง TB_DM_Timeseries ที่ใช้ในการประมวลผลในอัลกอริธึมไทม์ซีรีส์ (Time Series) ดังภาพที่ 3.21 และให้กำหนดแอตทริบิวต์ ID เป็นคีย์ (key) ซึ่งเป็นค่าที่เก็บลำดับการยืมหนังสือ ให้กำหนดแอตทริบิวต์ acc_gpa เป็นอินพุต (input) และพยากรณ์ (prediction) ซึ่งเป็นผลการเรียนนักศึกษาที่มีการยืมหนังสือ ดังภาพที่ 3.22

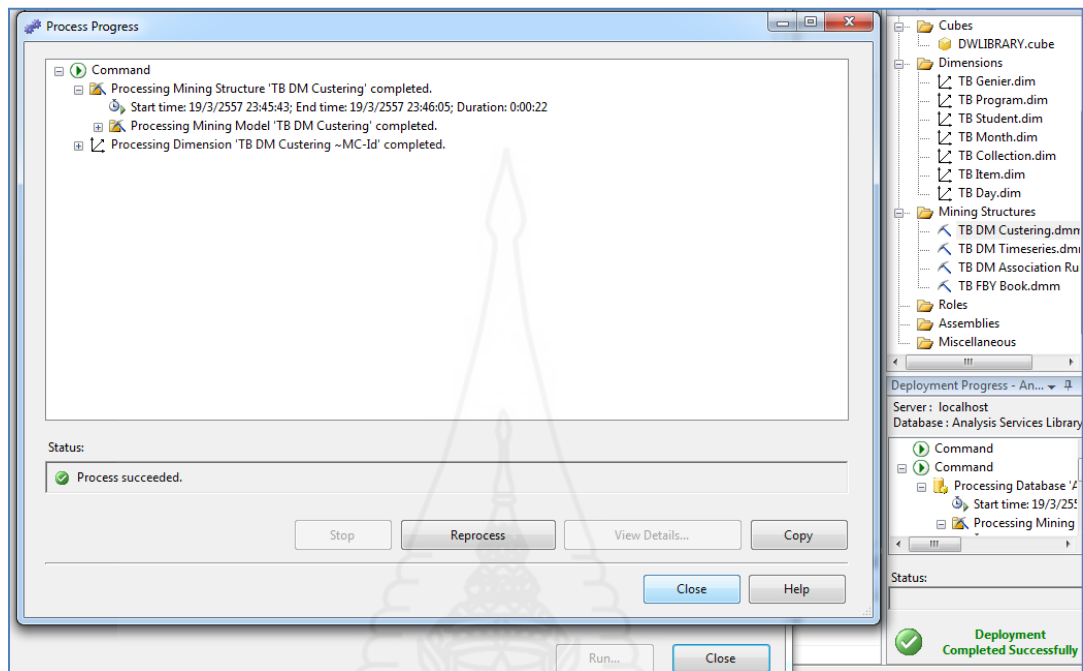


ภาพที่ 3.21 การเลือกตาราง TB_DM_Timeseries



ภาพที่ 3.22 กำหนดคีย์, อินพุต และพยากรณ์ อัลกอริธึมไทม์ซีรีส์ (Time Series)

9) เมื่อกำหนดเรียบร้อยแล้วให้ Deploy โพรเจกต์ และเมื่อ Deploy เสร็จ โดยไม่มีข้อผิดพลาด ก็สามารถเปิดดูผลลัพธ์ได้จากแท็บต่างๆ ดังภาพที่ 3.23



ภาพที่ 3.23 แสดงการ deploy โพรเจกต์

3.3.5 ประเมินผลลัพธ์ที่ได้ เป็นการประเมินผลลัพธ์จากแบบจำลองและอัลกอริธึมที่ใช้วิเคราะห์ข้อมูลว่า ผลลัพธ์ที่ได้นั้นครอบคลุมและสามารถตอบวัตถุประสงค์ที่กำหนดไว้หรือไม่ จากนั้นจึงนำเสนอประเมินแบบจำลองที่ได้จากการทำเหมืองข้อมูลเพื่อพิจารณาถึงความเหมาะสม ในการนำแบบจำลองไปประยุกต์ใช้ว่าแบบจำลองที่ได้นั้น มีความแม่นยำในการทำนายมากน้อยเพียงใด โดยในการประเมินแบบจำลองนั้น จะแบ่งตามลักษณะของการทำเหมืองข้อมูล ซึ่งในงานวิจัยนี้การแบ่งกลุ่มสามารถประเมินผลที่ได้จากระยะห่างจากจุดศูนย์กลางของกลุ่ม กฎความสัมพันธ์สามารถประเมินผลที่ได้จากการพิจารณาค่าความเชื่อมั่น และค่าสนับสนุน ส่วนไหมซีรีส์สามารถประเมินผลโดยการเปรียบเทียบผลการคาดการณ์ที่ได้กับข้อมูลจริงที่เกิดขึ้น

3.3.6 นำไปใช้งาน เป็นการนำผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ไปใช้งานจริง ในการจำแนกกลุ่มผลลัพธ์ที่ได้ ทำให้สามารถจำแนกกลุ่มผู้ให้บริการและทราบข้อมูลการใช้บริการของผู้ให้บริการแต่ละกลุ่มในด้านต่างๆ สำหรับการวิเคราะห์ด้วยกฎความสัมพันธ์โดยเฉพาะด้านการค้ำคืนสินสารสนเทศ การวิเคราะห์ประวัติการยืมหนังสือว่าหนังสือเล่มใดที่ถูกยืมด้วยกันเสมอ ซึ่งผู้ให้บริการสามารถนำข้อมูลผลลัพธ์ที่ได้ไปให้คำแนะนำสำหรับผู้ให้บริการต่อไปได้ ทำให้ผู้ให้บริการไม่เสียเวลาค้นคืนหนังสือหรือสื่อที่ต้องการและได้หนังสือมากพอต่อความต้องการ ผลการวิเคราะห์ที่ได้สามารถนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้ำคืนสินสารสนเทศ ของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ด้วยการประยุกต์และพัฒนารูปแบบการค้ำคืนสินสารสนเทศ ที่ตรงตามความต้องการของผู้ใช้มากขึ้น ส่วนผลลัพธ์ที่ได้จากการพยากรณ์ยอดการใช้บริการยืมหนังสือของนักศึกษาจากการวิเคราะห์อนุกรมเวลา ให้ทราบจำนวนผู้มาใช้บริการในแต่ละช่วงเวลา ซึ่งผู้ให้บริการสามารถนำไปเตรียมความพร้อมในด้านต่างๆ หรือบริการที่เกี่ยวข้องให้เพียงพอต่อความต้องการ



บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยเรื่อง การทำเหมืองข้อมูลเพื่อสนับสนุนการบริการคณิศนสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ผู้วิจัยได้พัฒนาคั้งข้อมูลนักศึกษาที่มาใช้บริการยืมหนังสือของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ เพื่อนำมาวิเคราะห์หาความรู้ใหม่โดยการทำเหมืองข้อมูลตามรูปแบบและขั้นตอนของแบบจำลองของคริสป์ - ดีเอ็ม (CRISP - DM Model) ซึ่งผลการวิจัยและการวิเคราะห์ข้อมูล ประกอบด้วย 3 ส่วน ได้แก่

1. การสร้างคั้งข้อมูล
2. การทำ Online Analytical Processing (OLAP)
3. การทำเหมืองข้อมูล

1. การสร้างคั้งข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด

งานวิจัยนี้ผู้วิจัยนำข้อมูลจากฐานข้อมูลของโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib ที่จัดเก็บข้อมูลการบริการยืมคณิศนสารสนเทศภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และข้อมูลนักศึกษาจากฐานข้อมูลทะเบียนมหาวิทยาลัยราชภัฏศรีสะเกษ นำมาสร้างระบบคั้งข้อมูล มีการนำข้อมูลมาจากฐานข้อมูลโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib และฐานข้อมูลงานทะเบียน มาผ่านระบบอีทีแอล (Extract Transform and Load: ETL) เพื่อจัดการกับโครงสร้างของข้อมูลจัดการกับฟอร์แมตข้อมูล รวมทั้งการทำความสะอาดข้อมูลให้มีความถูกต้องสมบูรณ์ แล้วจึงนำข้อมูลที่ได้ออกแบบโครงสร้างคั้งข้อมูลแบบสโนว์เฟลก (Snowflake Schema) ที่มีตารางข้อเท็จจริง (Fact Table) 1 ตารางคือ ตาราง TB_Borrow และมีตารางมิติ (Dimension) 7 ตาราง คือ ตาราง TB_FBYBook, ตาราง TB_Collection ตาราง TB_Day ตาราง TB_Genier ตาราง TB_Item ตาราง TB_Month ตาราง TB_Program และตาราง TB_Student

ตารางที่ 4.1 ตาราง TB_Genier (Dimension Table) สำหรับเก็บข้อมูลเพศ

Col.	Name	Type	Description
1	Genier_id (PK)	nvarchar (50)	รหัสเพศ
2	Genier_name	nvarchar (50)	เพศ

ตารางที่ 4.2 ตาราง TB_FBYBook (Dimension Table) สำหรับเก็บข้อมูลรหัสการยืมหนังสือ

Col.	Name	Type	Description
1	Borrow_id (PK)	Int	รหัสการยืม
2	Id_no	nvarchar (50)	รหัสนักศึกษา
3	Borrowdate	Date	วันที่ยืม
4	FBYBook	Int	จำนวนเล่มที่ยืม

ตารางที่ 4.3 ตาราง TB_Collection (Dimension Table) สำหรับเก็บข้อมูลหมวดหมู่หนังสือ

Col.	Name	Type	Description
1	Coll_id (PK)	nvarchar (50)	รหัสหมวดหมู่หนังสือ
2	Coll_name	nvarchar (50)	ชื่อหมวดหมู่หนังสือ

ตารางที่ 4.4 ตาราง TB_Day (Dimension Table) สำหรับเก็บข้อมูลวัน

Col.	Name	Type	Description
1	Day_id (PK)	nvarchar (50)	รหัสวัน
2	Day_name	nvarchar (50)	ชื่อวัน

ตารางที่ 4.5 ตาราง TB_Item (Dimension Table) สำหรับเก็บข้อมูลหนังสือ

Col.	Name	Type	Description
1	Bacode (PK)	nvarchar (50)	รหัสลำดับการยืม
2	Title	nvarchar (max)	ชื่อหนังสือหนังสือ
3	Bib	nvarchar (max)	รหัสชื่อหนังสือ
4	Coll_id	nvarchar (50)	หมวดหนังสือ

ตารางที่ 4.6 ตาราง TB_Month (Dimension Table) สำหรับเก็บข้อมูลเดือน

Col.	Name	Type	Description
1	Month_id (PK)	nvarchar (50)	รหัสเดือน
2	Month_name	nvarchar (50)	ชื่อเดือน

ตารางที่ 4.7 ตาราง TB_Program (Dimension Table) สำหรับเก็บข้อมูลโปรแกรมนักศึกษา

Col.	Name	Type	Description
1	Program_id (PK)	nvarchar (50)	รหัสโปรแกรมนักศึกษา
2	Program_name	nvarchar (max)	ชื่อโปรแกรมนักศึกษา

ตารางที่ 4.8 ตาราง TB_Student (Dimension Table) สำหรับเก็บข้อมูลนักศึกษา

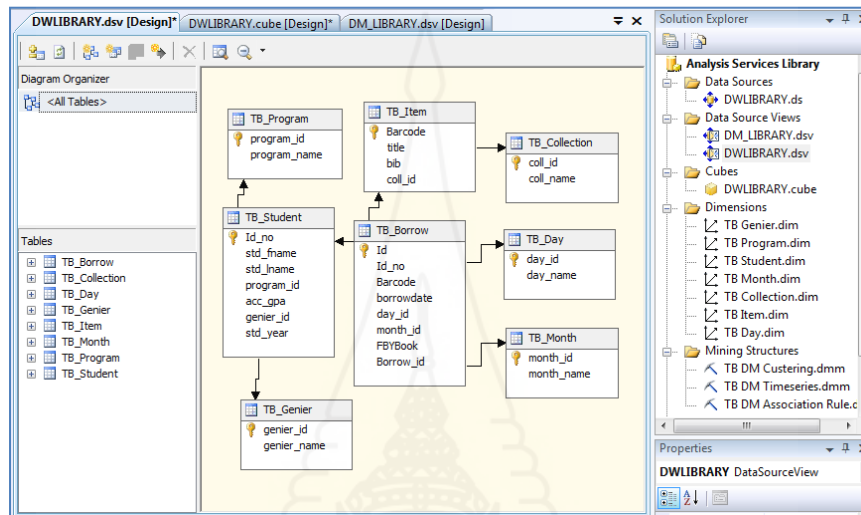
Col.	Name	Type	Description
1	Id_no (PK)	nvarchar (50)	รหัสนักศึกษา
2	std_fname	nvarchar (50)	ชื่อนักศึกษา
3	std_lname	nvarchar (50)	นามสกุลนักศึกษา
4	program_id	nvarchar (50)	รหัสโปรแกรมนักศึกษา
5	acc_gpa	Float	เกรดเฉลี่ย
6	genier_id	nvarchar (50)	รหัสเพศ
7	std_year	nvarchar (50)	ชั้นปี

ตารางที่ 4.9 ตาราง TB_Borrow (Fact Table) สำหรับเก็บข้อมูลการยืมหนังสือ

Col.	Name	Type	Description
1	Id (PK)	Int	ลำดับการยืม
2	Id_no	nvarchar (50)	รหัสนักศึกษา
3	Barcode	nvarchar (50)	รหัสเล่มหนังสือ
4	Borrowdate	Date	วันที่ยืม
5	Day_id	nvarchar (50)	วันยืม
6	Month_id	nvarchar (50)	เดือนที่ยืม
7	FBYBook	Int	จำนวนเล่มที่ยืม
8	Borrow_id	Int	รหัสการยืม

2. การทำ Online Analytical Processing (OLAP)

ผลจากการทำวิจัยได้สร้าง Data Source ชื่อ DWLIBRARY.ds ใน SSAS เชื่อมต่อกับคลังข้อมูล (data warehouse) จากนั้นสร้าง Data Source View ชื่อ DWLIBRARY.dsv ดังแสดงในภาพที่ 4.1 เพื่อเลือกตารางจากคลังข้อมูลทั้งที่เป็น Fact Tables และ Dimension Tables



ภาพที่ 4.1 แสดงผลที่ได้จากการสร้าง Data Source View ชื่อ DWLIBRARY.dsv

จากนั้นสร้าง Cube แล้วจะสามารถดูข้อมูลที่เป็นของการประมวลผลเชิงวิเคราะห์แบบออนไลน์ได้ ในลักษณะหลายมิติผ่านแท็บ Browser ดังแสดงในภาพที่ 4.2 ทั้งนี้ผู้ใช้สามารถเปลี่ยน Measures และ Dimension ที่ต้องการวิเคราะห์โดยลากจากหน้าต่างซ้ายมือไปวางในหน้าต่างขวามือ เพื่อดูข้อมูลต่างๆ เกี่ยวกับนักศึกษาที่มีการยืมคืนหนังสือได้ในหลายมิติ

	Month Id ▾	Day Id ▾				
	1	10	11	12	2	3
Coll Id ▾	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow
0	519	536	801	593	89	116
1	505	695	1103	629	77	39
2	154	111	192	117	24	40
3	2903	3306	4912	3254	567	319
4	892	1199	2037	1410	110	91
5	650	609	1240	812	178	30
6	660	791	1153	668	120	255
7	180	126	290	269	19	27
8	413	382	519	407	49	39
9	699	388	1102	746	84	32
Fic	1420	1491	2634	2310	343	323
J	57	86	137	115	16	7
R	11	14	19	6	4	2
Sc	1015	1832	2160	1372	197	197

ภาพที่ 4.2 แสดงผลตัวอย่างการประมวลผลเชิงวิเคราะห์แบบออนไลน์

จากภาพที่ 4.2 แสดงให้เห็นข้อมูลในมิติที่ด้านบนหรือคอลัมน์เป็นข้อมูลของเดือนและวันที่มีการยืมหนังสือและด้านซ้ายหรือแถวเป็นข้อมูลของหมวดหนังสือที่มีการยืม แสดงให้เห็นถึงจำนวนการยืมหนังสือในแต่ละหมวดในแต่ละเดือน เช่น หนังสือในหมวดความรู้ทั่วไป (000) เดือนมกราคม (1) มีจำนวนการยืมหนังสือเท่ากับ 519 ครั้ง หนังสือในหมวดสังคมศาสตร์ (300) เดือนตุลาคม (10) มีจำนวนการยืมหนังสือเท่ากับ 3,306 ครั้ง ซึ่งในการแสดงผลข้อมูลหลายมิติสามารถเปลี่ยนข้อมูลหรือมิติในการดูข้อมูลได้ หากต้องการดูข้อมูลว่าแต่ละโปรแกรมมีการยืมหนังสือจำนวนเท่าไรในแต่ละก็สามารถปรับเปลี่ยนรูปแบบการแสดงผลข้อมูลได้ ดังภาพที่ 4.3 และหากต้องการดูข้อมูลถึงลงไปอีกก็สามารถปรับเปลี่ยนรูปแบบการแสดงผลได้เช่นกันโดยการเลือกข้อมูลที่ต้องการแสดงผลใน Measure Group คลีกลากเขาใส่ก็จะทำให้ได้ผลของข้อมูลต่างๆ ที่ต้องการมาแสดงอย่างง่ายดาย ดังภาพที่ 4.4 แสดงผลข้อมูลการยืมหนังสือในแต่ละเดือนของนักศึกษาโปรแกรมวิชาต่างๆ และบอกรายละเอียดด้วยว่าเพศชายและเพศหญิงยืมจำนวนเท่าไร

		Month Id ▾		Day Id ▾			
		1	10	11	12	2	
Program Id ▾	Genier Id ▾	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	
2001				6	7		
2122		2	20	19	5	3	
2211		2	2	2	2		
2240			2	2	2		
2252		102	164	192	142	13	
2253		35	49	48	47	18	
2254		51	152	237	142	2	
2255			3	6			
2311		2	9	9	31		
	[TB Day].[Day Id]	4					
		12	57	94	51	2	
	Attribute Hierarchy: Day Id	87	59	262	289	9	
		93	65	166	88	9	
	2351						

ภาพที่ 4.3 การแสดงผลข้อมูลหลายมิติ

		Month Id ▾		Day Id ▾			
		1	10				
Program Id ▾	Genier Id ▾	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Count	TB Borrow Co
2001							
2122	1	2		1	10		4
	2						2
	Total	2		1	10		4
2211			2				
2240							
2252		102	25	28	38	20	23
2253		35	4	5	8	15	4
2254		51	6	25	18	51	12
2255						3	
2311		2	3	2	2	1	
2334		12		8	14	13	
2336		87		9	25	9	
2351		93		3	3	57	
2352		32	5	10	3	9	
2355		10			4		

ภาพที่ 4.4 แสดงผลข้อมูลการยืมหนังสือในแต่ละเดือนของนักศึกษาโปรแกรมวิชาต่างๆ

3. การทำเหมืองข้อมูล

งานวิจัยนี้ดำเนินการตามแบบจำลองของคริสพ์ - ดีเอ็ม (CRISP - DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.1 ทำความเข้าใจธุรกิจ

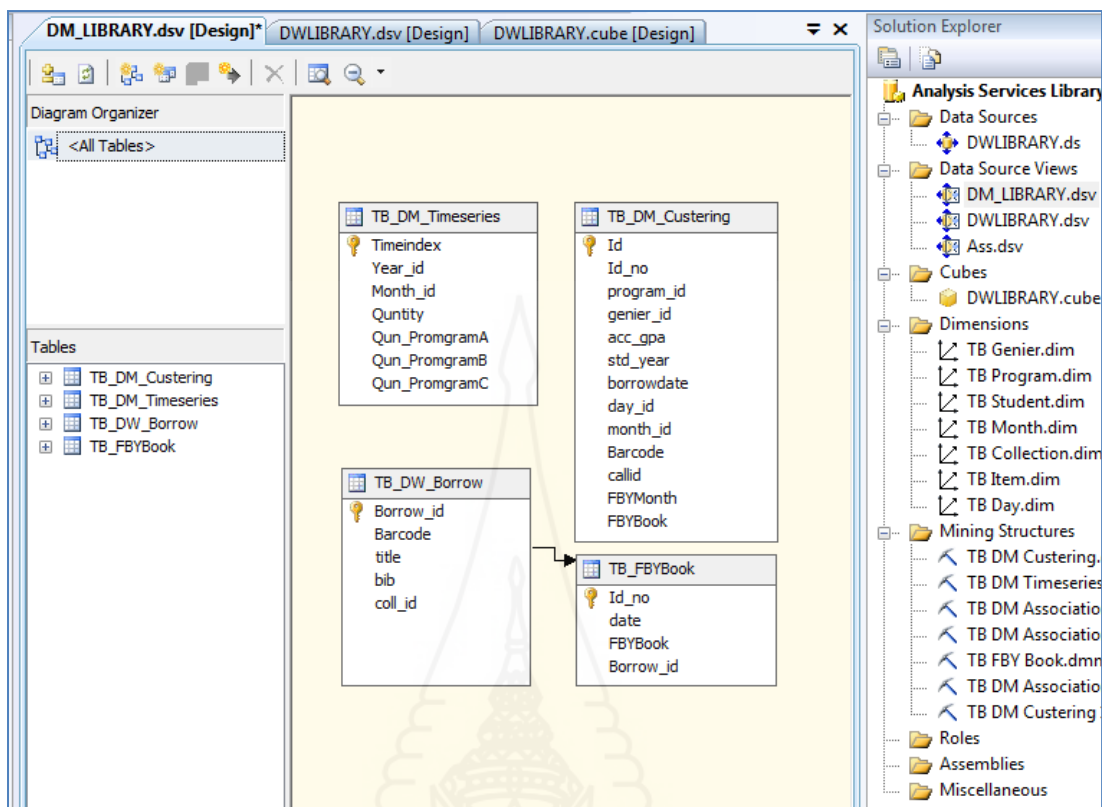
ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องและศึกษาข้อมูลเกี่ยวกับการค้นคืนสารสนเทศภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และวิเคราะห์ปัญหาที่เกิดขึ้นจริงกับการปฏิบัติงานของห้องสมุด พบว่าสารสนเทศไม่เพียงพอต่อความต้องการของผู้ใช้ในบางช่วงเหตุการณ์ เช่น ช่วงเวลาสอบ หนังสือไม่เพียงพอต่อการให้บริการแก่ผู้ใช้บริการ เพราะไม่สามารถพยากรณ์จำนวนนักศึกษาที่มาใช้บริการห้องสมุดในแต่ละช่วงเวลาได้ ทำให้ขาดข้อมูลประกอบในการปรับปรุงและเตรียมการบริการ นักศึกษาค้นหาหนังสือไม่ตรงตามความต้องการ หากมีการแบ่งกลุ่มนักศึกษาออกเป็นกลุ่มๆ จะทำให้สามารถค้นหาหนังสือที่น่าสนใจในแต่ละกลุ่มได้และนักศึกษาได้หนังสือที่ตรงตามความต้องการอีกด้วย หากนักศึกษาต้องการหนังสือหลายเล่มต้องเสียเวลาในการสืบค้นข้อมูลหลายรอบ จึงทำให้เสียเวลามากในการค้นหาข้อมูลหนังสือ

3.2 ทำความเข้าใจข้อมูล

ผู้วิจัยดำเนินการรวบรวมข้อมูลที่เกี่ยวข้องเพื่อนำไปวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งมีด้วยกันหลายอัลกอริทึม และแต่ละอัลกอริทึมก็มีวัตถุประสงค์ที่แตกต่างกัน ผู้วิจัยจึงได้สร้างคลังข้อมูลขึ้นมาดังที่กล่าวมาแล้วนั้น ซึ่งเป็นข้อมูลหนังสือ ข้อมูลสมาชิกและข้อมูลการใช้บริการยืมหนังสือของสมาชิกภายในห้องสมุด

3.3 เตรียมข้อมูล

ผู้วิจัยดำเนินการเมื่อสร้างคลังข้อมูล (Data Warehouse) ผู้วิจัยได้นำข้อมูลที่มีการจัดเก็บไว้มาปรับโครงสร้างที่เหมาะสมเพื่อใช้กับแบบจำลองการทำเหมืองข้อมูลตามอัลกอริทึม เพราะข้อมูลที่ได้จากที่ใช้ในงานวิจัย ซึ่งประกอบด้วย 3 อัลกอริทึม คือ 1) คลัสเตอร์ริง (Clustering) 2) แอสโซซิเอชันรูลส์ (Association Rules) และ 3) ไทม์ซีรีส์ (Time Series) ผู้วิจัยจึงสร้าง Data Source ชื่อ DWLIBRARY.ds ใน SSAS เชื่อมต่อกับคลังข้อมูล (data warehouse) จากนั้นสร้าง Data Source View ชื่อ DM_LIBRARY.dsv เพื่อเลือกตารางจากคลังข้อมูลในการวิเคราะห์ข้อมูลแต่ละอัลกอริทึมดังนี้



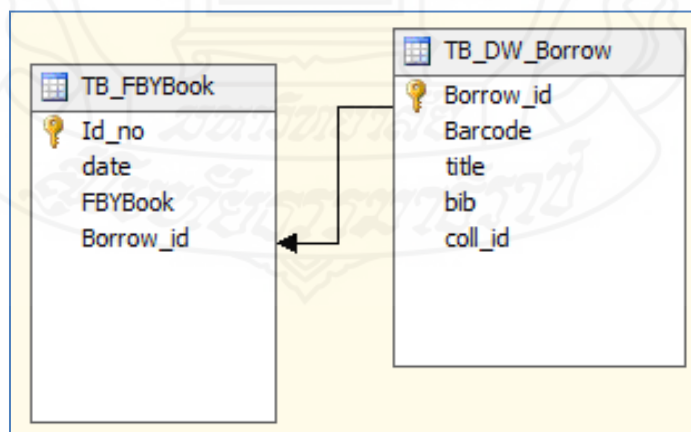
ภาพที่ 4.5 แสดงผลที่ได้จากการสร้าง Data Source View ชื่อ DM_LIBRARY.dsv

3.3.1 การจัดกลุ่ม (Clustering) เป็นการจัดการข้อมูลเพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ สำหรับข้อมูลที่ใช้เป็นข้อมูลเกี่ยวกับสมาชิกและข้อมูลเกี่ยวกับการใช้บริการยืมคืนของสมาชิกโดยดึงข้อมูลจากรายการ TB_DM_Custering ซึ่งประกอบด้วยข้อมูลที่จัดเก็บรหัสนักศึกษา โปรแกรมวิชา เพศ ช่วงเกรดเฉลี่ย ชั้นปี วันที่ยืม วันยืมเดือนที่ยืม รหัสหนังสือ หมวดหมู่หนังสือ จำนวนเล่มที่ยืมต่อเดือนและจำนวนเล่มที่ยืมในแต่ละครั้ง ซึ่งมีข้อมูลการให้บริการยืมคืนจำนวน 122,763 รายการ

Id	Id_no	program_id	genier_id	acc_gpa	std_year	borrowdate	day_id	month_id	Barcode	callid	FBYMonth	FBYBook
1	4914101109	4101	1	<2.5	0	2010-09-01	4	9	i00020695	3	1	1
2	4914101128	4101	1	<2.5	0	2010-07-03	7	7	i00014354	9	1	1
3	4914101128	4101	1	<2.5	0	2010-08-27	6	8	i00031990	Sc	3	3
4	4914101128	4101	1	<2.5	0	2010-08-27	6	8	i00034271	Sc	3	3
5	4914101128	4101	1	<2.5	0	2010-08-27	6	8	i00041904	8	3	3
6	4914101133	4101	1	<2.5	0	2010-09-04	7	9	i00016026	3	3	3
7	4914101133	4101	1	<2.5	0	2010-09-04	7	9	i00019433	3	3	3
8	4914101133	4101	1	<2.5	0	2010-09-04	7	9	i00020002	3	3	3
9	4914101134	4101	2	<2.5	0	2010-09-04	7	9	i00019433	3	1	1
10	4914121116	4121	1	<2.5	0	2010-11-26	6	11	i00014777	7	3	3
11	4914121116	4121	1	<2.5	0	2010-11-26	6	11	i00029828	7	3	3
12	4914121116	4121	1	<2.5	0	2010-11-26	6	11	i00034758	6	3	3
13	5012240109	2240	1	<2.5	0	2010-07-05	2	7	i00024380	6	2	2
14	5012240109	2240	1	<2.5	0	2010-07-05	2	7	i00034583	Sc	2	2
15	5012240111	2240	1	<2.5	0	2010-07-05	2	7	i00000039	6	3	3
16	5012240111	2240	1	<2.5	0	2010-07-05	2	7	i00012797	Fic	3	3
17	5012240111	2240	1	<2.5	0	2010-07-05	2	7	i00026884	6	3	3
18	5012240111	2240	1	<2.5	0	2010-08-05	5	8	i00038841	Sc	1	1

ภาพที่ 4.6 แสดงตัวอย่างข้อมูล ตาราง TB_DM_Custering

3.3.2 กฎการหาความสัมพันธ์ (Association Rule) เป็นการจัดการข้อมูลเพื่อวิเคราะห์ข้อมูลการยืมหนังสือ ว่าหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยด้วยกัน สำหรับข้อมูลที่ใช้เป็นข้อมูลจาก 2 ตาราง คือ ตาราง TB_FBYBook เป็นข้อมูลเกี่ยวกับการยืมหนังสือ ซึ่งประกอบไปด้วยข้อมูลรหัสนักศึกษา วันที่ยืม จำนวนเล่มที่ยืมและรหัสการยืม กับตาราง TB_DW_Borrow เป็นข้อมูลเกี่ยวกับหนังสือที่มีการยืมประกอบด้วยข้อมูลรหัสการยืม รหัสหนังสือ ชื่อหนังสือ รหัสเล่มหนังสือและหมวดหมู่หนังสือ ซึ่งทั้ง 2 ตารางต้องสร้างความสัมพันธ์ด้วยแอตทริบิวต์ชื่อ Borrow_id โดยความสัมพันธ์ที่สร้างขึ้นเป็นแบบหลายต่อหนึ่ง (many - to - one) ดังภาพที่ 4.7



ภาพที่ 4.7 แสดงความสัมพันธ์ด้วยแอตทริบิวต์ชื่อ Borrow_id

3.3.3 ไทม์ซีรีส์ (Time Series) เป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคตโดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาในอนาคต สำหรับข้อมูลที่ใช้เป็นข้อมูลเกี่ยวกับสถิติหรือข้อมูลช่วงเวลาเป็นข้อมูลจากตาราง TB_DM_Timeseries เป็นข้อมูลเกี่ยวกับการเก็บข้อมูลการใช้บริการยืมคืนของนักศึกษาในแต่ละเดือนตั้งแต่ช่วงเดือน มิถุนายน 2553 ถึง เดือน ธันวาคม 2556 ซึ่งเป็นข้อมูลการยืมคืนทั้งหมดและข้อมูลการยืมคืนของนักศึกษา 3 โปรแกรมวิชาที่มีการยืมสูงสุด

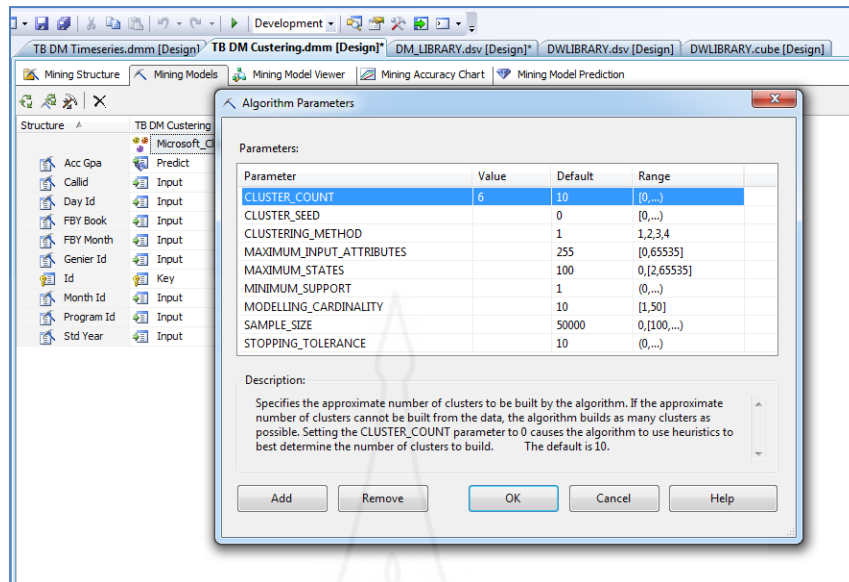
	Timeindex	Year_id	Month_id	Quntity	Qun_PromgramA	Qun_PromgramB	Qun_PromgramC
1	255306	2010	6	1654	168	335	380
2	255307	2010	7	1650	106	194	303
3	255308	2010	8	2478	170	267	517
4	255309	2010	9	1154	80	126	227
5	255310	2010	10	2057	159	272	568
6	255311	2010	11	2084	137	220	508
7	255312	2010	12	2183	205	354	579
8	255401	2011	1	3808	782	605	1228
9	255402	2011	2	556	14	28	44
10	255403	2011	3	607	12	46	17
11	255404	2011	4	541	34	18	18
12	255405	2011	5	468	47	41	53
13	255406	2011	6	5916	651	1361	1983
14	255407	2011	7	5680	550	1563	1231
15	255408	2011	8	4793	454	1077	1242
16	255409	2011	9	1331	102	268	407
17	255410	2011	10	4636	422	786	1657
18	255411	2011	11	4716	311	1096	1823
19	255412	2011	12	1915	98	386	794
20	255501	2012	1	1931	161	546	669
21	255502	2012	2	460	16	83	114

ภาพที่ 4.8 แสดงตัวอย่างข้อมูล ตาราง TB_DM_Timeseries

3.4 สร้างแบบจำลอง

การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ได้แก่การสร้างแบบจำลองที่สอดคล้องกับอัลกอริธึมแต่ละชนิด ซึ่งจากการศึกษาและทำความเข้าใจกับข้อมูลแล้ว คณะผู้วิจัยได้เลือกอัลกอริธึมที่จะนำมาประมวลผล 3 อัลกอริธึม

3.4.1 ผลการจับกลุ่ม (Clustering) เพื่อแบ่งกลุ่มของสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกัน หรือมีข้อมูลทั่วไปเหมือนกันออกเป็นกลุ่มๆ จากตาราง TB_DM_Custering ซึ่งมีข้อมูลการใช้บริการยืมคืนจำนวน 122,763 รายการ โดยกำหนดค่าพารามิเตอร์ดังที่แสดงในภาพที่ 4.9



ภาพที่ 4.9 กำหนดค่าพารามิเตอร์ การจัดกลุ่ม (Clustering)

CLUSTER_COUNT ซึ่งเป็นค่าของจำนวนคลัสเตอร์ที่ต้องการให้อัลกอริทึมจำแนกให้ ค่าดีฟอลท์คือ 10 แต่ถ้ากำหนดไว้เท่ากับ 0 หมายความว่าให้อัลกอริทึมกำหนดจำนวนคลัสเตอร์ที่ต้องการจำแนกเอง ด้วยการใช้หลักการของฮิวริสติกส์ในการจำแนก ซึ่งในงานวิจัยนี้ผู้วิจัยได้ทดสอบการตั้งค่าต่างๆ ของคลัสเตอร์ตั้ง 4 - 10 คลัสเตอร์ พบว่าการตั้งค่าที่ 6 คลัสเตอร์เป็นจำนวนคลัสเตอร์ที่เหมาะสมที่สุดเนื่องจากมีความแตกต่างของแต่ละคลัสเตอร์ชัดเจนที่สุด

CLUSTER_SEED เป็นค่าตัวเลขสุ่มเริ่มต้นที่ใช้ในการจำแนกคลัสเตอร์ ค่าดีฟอลท์คือ 0

CLUSTERING_METHOD เป็นการกำหนดวิธีการจำแนกคลัสเตอร์ที่จะใช้ในอัลกอริทึม ค่าดีฟอลท์คือ ค่า 1 หมายถึงใช้วิธี Scalable EM (Expected Mean) ค่า 2 หมายถึง Non - scalable EM ค่า 3 หมายถึง Scalable K - means และค่า 4 หมายถึง Non - scalable K - means

MAXIMUM_INPUT_ATTRIBUTES คือค่าของจำนวนแอตทริบิวต์สูงสุดที่สามารถนำมาใช้ในการจำแนกคลัสเตอร์ ค่าดีฟอลท์คือ 255

MAXIMUM_STATES คือค่าที่ใช้ควบคุมจำนวนสถานะทั้งหมดที่แอตทริบิวต์แต่ละแอตทริ - บิวต์ สามารถเปลี่ยนสถานะได้ ค่านี้นี้ไม่ควรกำหนดไว้สูงเพราะจะมีผลต่อหน่วยความจำที่ต้องใช้ในการทำงานของอัลกอริทึมอาจไม่พอ ค่าดีฟอลท์คือ 100

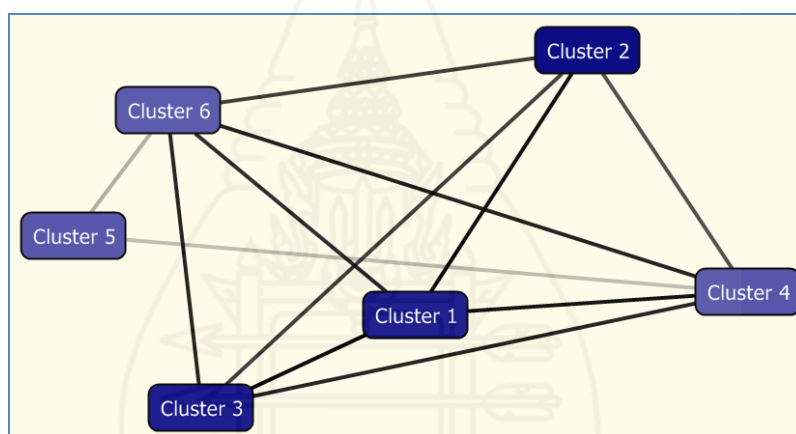
MINIMUM_SUPPORT คือ ค่าที่กำหนดว่าในแต่ละคลัสเตอร์เมื่อจำแนกแล้วควรมีจำนวนข้อมูลไม่น้อยกว่าค่าที่กำหนดในพารามิเตอร์ตัวนี้ ดังนั้นค่านี้นี้จึงใช้ในการพิจารณาว่าคลัสเตอร์ใดควรตัดทิ้งหรือจำแนกใหม่ การกำหนดค่านี้นี้สูงเกินไปอาจทำให้ได้ผลลัพธ์ที่ไม่ดีนัก ค่าดีฟอลท์คือ 1

MODELLING_CARDINALITY เป็นค่าที่ควบคุมจำนวนแบบจำลองที่อัลกอริทึมจะสร้างขึ้นในระหว่างที่มีการจำแนกคลัสเตอร์ หากกำหนดค่านี้นี้ลดลงก็จะเพิ่มสมรรถนะในการทำงานของอัลกอ - ริทึม ค่าดีฟอลท์คือ 10

SAMPLE_SIZE เป็นค่าแสดงจำนวนของเรคอร์ดข้อมูลที่จะใช้ในแต่ละขั้นตอนของกรอบงานซึ่งปรับเปลี่ยนได้ ถ้ากำหนดค่านี้น้อยลงจะทำให้อัลกอริธึมลดการทำงานลงเพราะว่าการทำงานของอัลกอริธึมจะไม่ประมวลผลเรคอร์ดข้อมูลทั้งหมด โดยเฉพาะอย่างยิ่งหากกำหนดร่วมกับค่า STOPPING_TOLERANCE ที่กำหนดให้มีค่ามาก ค่าดีฟอลท์คือ 50,000 เรคอร์ด

STOPPING_TOLERANCE เป็นค่าที่ใช้กำหนดว่าเมื่อใดอัลกอริธึมควรจะหยุดทำงาน โดยค่านี้จะแทนจำนวนเรคอร์ดสูงสุดที่สามารถเปลี่ยนไปเป็นสมาชิกในอีกคลัสเตอร์ได้ก่อนที่จะพิจารณาว่าแบบจำลองนั้นต้องหยุดทำงานแล้ว ค่าดีฟอลท์คือ 10

สำหรับรายละเอียดหรือผลที่ได้จากการใช้เทคนิค Clustering มีดังต่อไปนี้ เมื่อคลิกที่แท็บ Model Mining Viewer พบว่า อัลกอริธึมได้จำแนกออกเป็น 6 คลัสเตอร์ และเมื่อเลื่อนสไลด์บาร์ด้านซ้ายมือซึ่งเป็นสไลด์บาร์ที่แสดงความสัมพันธ์ระหว่างคลัสเตอร์ โดยเลื่อนลงจะพบว่าความสัมพันธ์ดังแสดงในภาพที่ 4.10 โดยแต่ละกลุ่มมีจำนวนครั้งของการยืมหนังสือแสดงไว้ในตารางที่



ภาพที่ 4.10 การจัดกลุ่ม (Clustering) อัลกอริธึมได้จำแนกออกเป็น 6 คลัสเตอร์

ตารางที่ 4.10 การจัดกลุ่ม (Clustering) อัลกอริธึมได้จำแนกออกเป็น 6 คลัสเตอร์

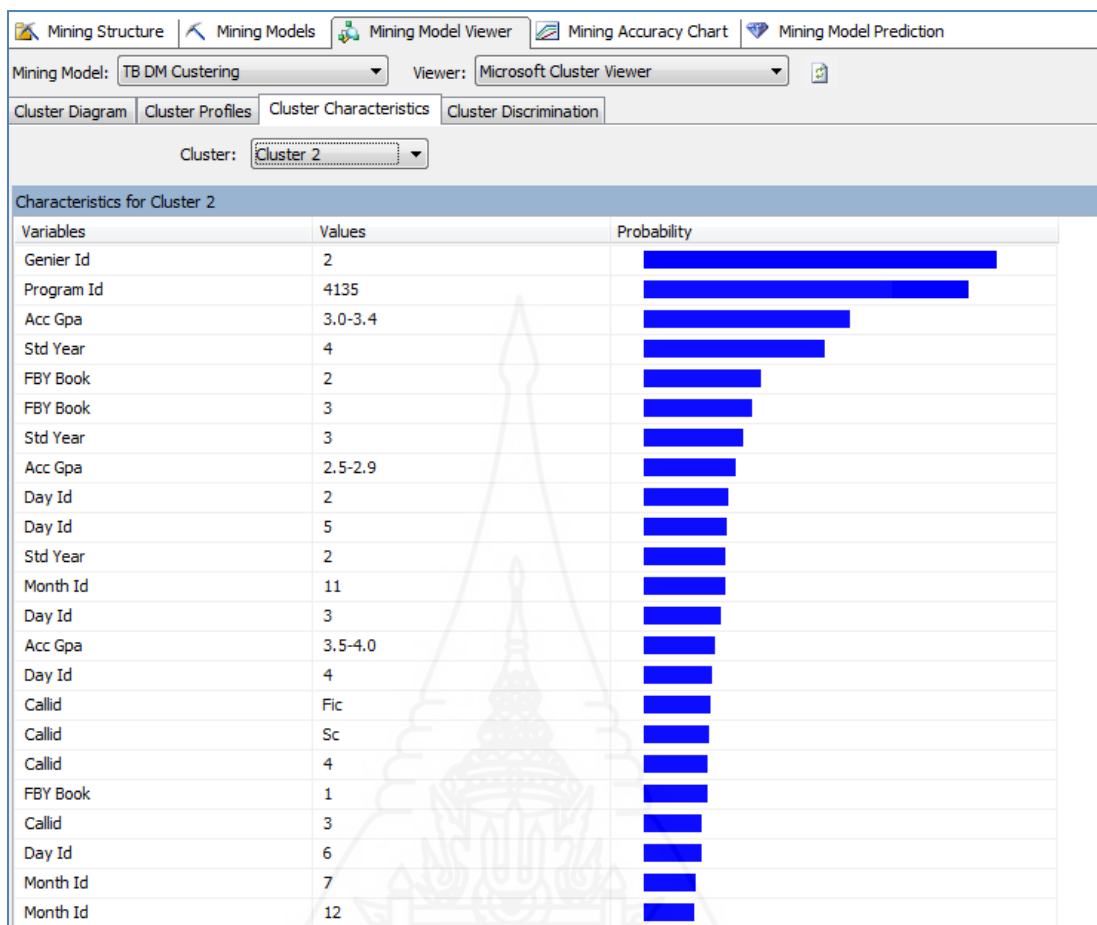
Cluster ที่	จำนวนครั้งของการยืมหนังสือ
1	22,999
2	25,335
3	23,397
4	16,716
5	17,395
6	16,921
ทั้งหมด	122,763

จากตารางที่ 4.10 พบว่าประชากรที่นำมาประมวลผลโดยใช้เทคนิค Clustering มีข้อมูลจำนวน 122,763 เรคอร์ด โดยภาพรวมนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 – 3.4 และมีเพศหญิงเป็นส่วนใหญ่ โดยเป็นนักศึกษาชั้นปีที่ 3 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ชอบยืมหนังสือในวันพฤหัสบดี และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 - 3 เล่ม ผู้วิจัยได้ศึกษารายละเอียดแต่ละคลัสเตอร์ โดยคลิกที่แท็บ Cluster Characteristics จะมีหน้าต่างใหม่เปิดขึ้นมา สามารถเลือกคลัสเตอร์ที่ต้องการศึกษาได้ ดังนี้

Variables	Values	Probability
Genier Id	2	[Bar]
Std Year	3	[Bar]
Acc Gpa	2.5-2.9	[Bar]
Program Id	4135	[Bar]
Callid	3	[Bar]
Std Year	2	[Bar]
FBY Book	2	[Bar]
Acc Gpa	3.0-3.4	[Bar]
FBY Book	3	[Bar]
Day Id	5	[Bar]
Month Id	7	[Bar]
Day Id	3	[Bar]
FBY Book	1	[Bar]
Day Id	4	[Bar]
Month Id	11	[Bar]
Day Id	6	[Bar]
Acc Gpa	<2.5	[Bar]

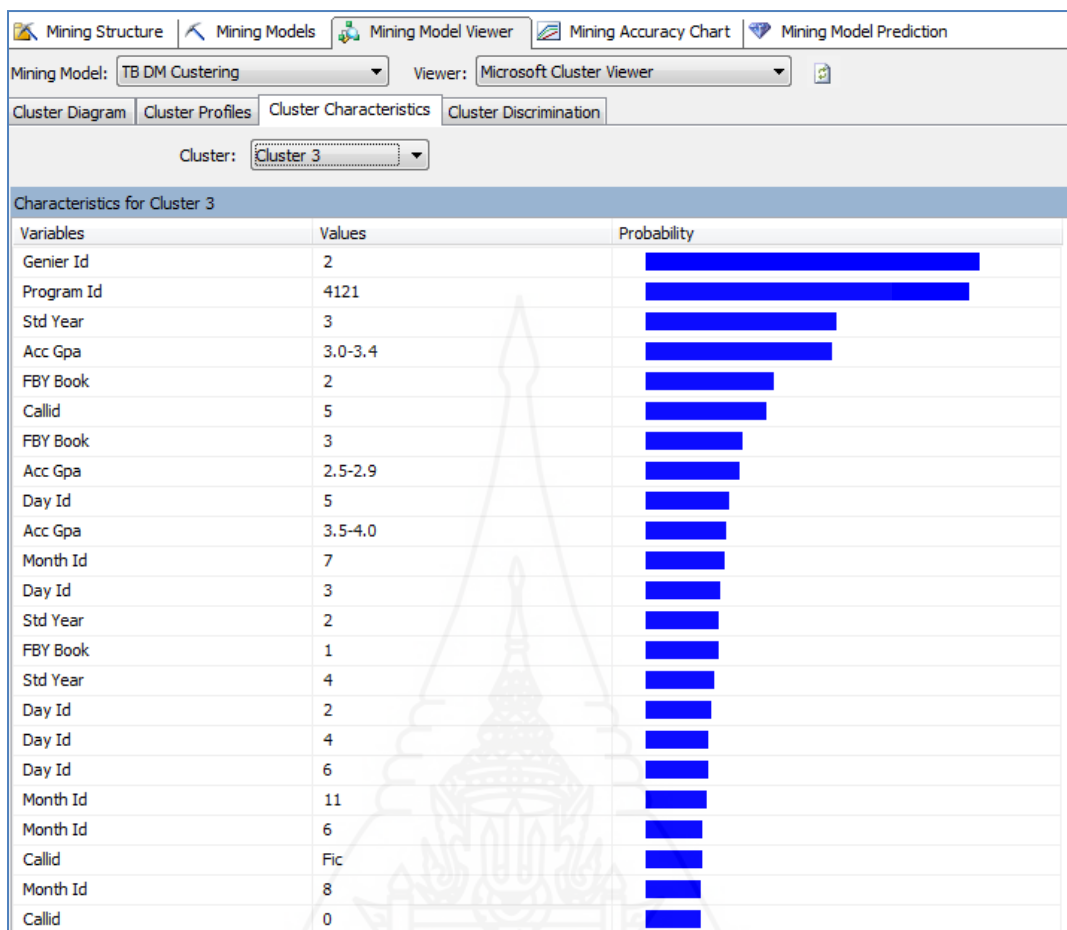
ภาพที่ 4.11 รายละเอียดคลัสเตอร์ที่ 1

คลัสเตอร์ที่ 1 สรุปผลว่านักศึกษาที่มีการยืมหนังสือจำนวนทั้งหมด 22,999 เรคอร์ด โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 – 2.9 ร้อยละ 52.9 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 – 3.4 ร้อยละ 30.2 และมีเพศหญิงเป็นส่วนใหญ่ ร้อยละ 86.7 โดยเป็นนักศึกษาชั้นปีที่ 3 ร้อยละ 54.7 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ร้อยละ 50.3 นักศึกษาชอบยืมหนังสือในวันพฤหัสบดี ร้อยละ 21.9 รองลงมาชอบยืมหนังสือในวันอังคาร ร้อยละ 20.5 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 32.9 รองลงมาจะยืมหนังสือครั้งละ 3 เล่ม ร้อยละ 25.2 นักศึกษาในคลัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนกรกฎาคม ร้อยละ 21.6 รองลงมาเป็นเดือนพฤศจิกายน ร้อยละ 16.7 และนักศึกษาที่มีการยืมหนังสือของคลัสเตอร์ที่ 1 ส่วนใหญ่ยืมหนังสือหมวดสังคมศาสตร์ (300) ร้อยละ 37



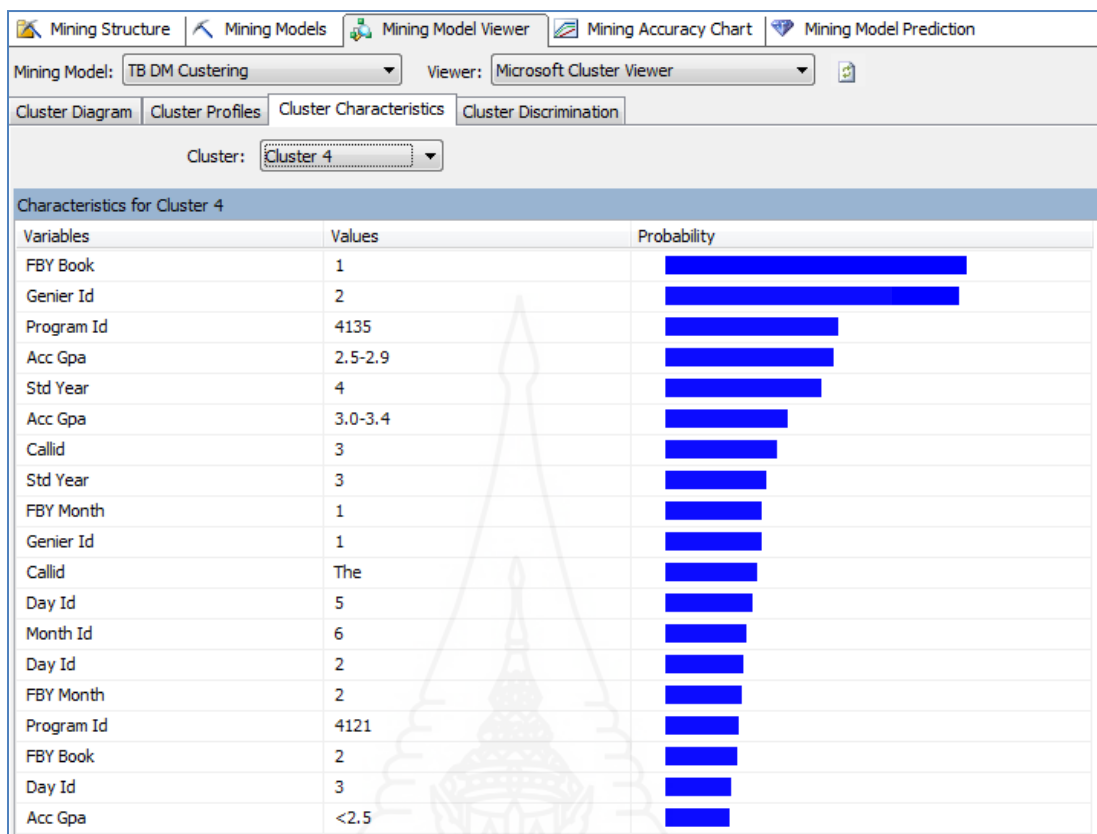
ภาพที่ 4.12 รายละเอียดคัสเตอร์ที่ 2

คัสเตอร์ที่ 2 สรุปผลว่านักศึกษาที่มีการยืมหนังสือจำนวนทั้งหมด 25,335 เรคอร์ด โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 – 3.4 ร้อยละ 54.6 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 – 2.9 ร้อยละ 24.4 และมีเพศหญิงเป็นส่วนใหญ่ ร้อยละ 93 โดยเป็นนักศึกษาชั้นปีที่ 4 ร้อยละ 47.8 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ร้อยละ 85.8 นักศึกษาชอบยืมหนังสือในวันจันทร์ ร้อยละ 22.3 รองลงมาชอบยืมหนังสือในวันพฤหัสบดี ร้อยละ 21.8 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 30.9 รองลงมาจะยืมหนังสือครั้งละ 3 เล่ม ร้อยละ 28.6 นักศึกษาในคัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนพฤศจิกายน ร้อยละ 21.5 รองลงมาเป็นเดือนกรกฎาคม ร้อยละ 14 และนักศึกษาที่มีการยืมหนังสือของคัสเตอร์ที่ 2 ส่วนใหญ่ยืมหนังสือหมวดนวนิยาย (Fic) ร้อยละ 17.6



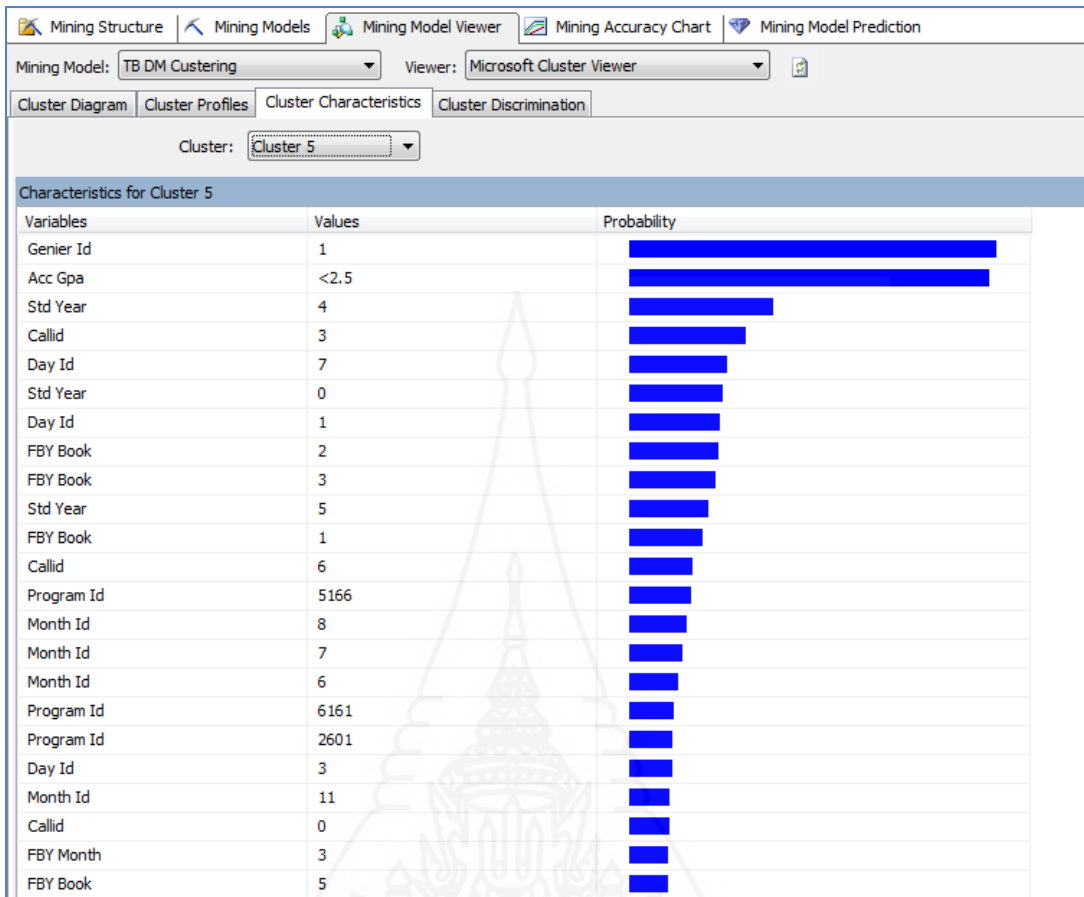
ภาพที่ 4.13 รายละเอียดคัสเตอร์ที่ 3

คัสเตอร์ที่ 3 สรุปผลว่านักศึกษาที่มีการยืมหนังสือจำนวนทั้งหมด 23,397 เรคอร์ด โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 – 3.4 ร้อยละ 48.9 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 – 2.9 ร้อยละ 24.9 และมีเพศหญิงเป็นส่วนใหญ่ ร้อยละ 87.8 โดยเป็นนักศึกษาชั้นปีที่ 3 ร้อยละ 50.3 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาคณิตศาสตร์ ร้อยละ 85.1 นักศึกษาชอบยืมหนังสือในวันพฤหัสบดี ร้อยละ 22 รองลงมาชอบยืมหนังสือในวันอังคาร ร้อยละ 19.6 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 33.8 รองลงมาจะยืมหนังสือครั้งละ 3 เล่ม ร้อยละ 25.7 นักศึกษาในคัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนกรกฎาคม ร้อยละ 20.9 รองลงมาเป็นเดือนพฤศจิกายน ร้อยละ 16.1 และนักศึกษาที่มีการยืมหนังสือของคัสเตอร์ที่ 3 ส่วนใหญ่ยืมหนังสือหมวดวิทยาศาสตร์ (500) ร้อยละ 32



ภาพที่ 4.14 รายละเอียดคลัสเตอร์ที่ 4

คลัสเตอร์ที่ 4 สรุปผลว่านักศึกษาที่มีการยืมหนังสือจำนวนทั้งหมด 16,716 เรคอร์ด โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 - 2.9 ร้อยละ 43.1 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 - 3.4 ร้อยละ 31.5 และมีเพศหญิงเป็นส่วนใหญ่ ร้อยละ 75.3 โดยเป็นนักศึกษาชั้นปีที่ 4 ร้อยละ 39.9 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ร้อยละ 44.1 นักศึกษาชอบยืมหนังสือในวันพฤหัสบดี ร้อยละ 22.6 รองลงมาชอบยืมหนังสือในวันจันทร์ ร้อยละ 20.1 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 1 เล่ม ร้อยละ 77 รองลงมาจะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 18.6 นักศึกษาในคลัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนมิถุนายน ร้อยละ 20.8 รองลงมาเป็นเดือนพฤศจิกายน ร้อยละ 14.5 และนักศึกษาที่มีการยืมหนังสือของคลัสเตอร์ที่ 4 ส่วนใหญ่ยืมหนังสือหมวดสังคมศาสตร์ (300) ร้อยละ 28.8



ภาพที่ 4.15 รายละเอียดคลัสเตอร์ที่ 5

คลัสเตอร์ที่ 5 สรุปผลว่านักศึกษาที่มีการยืมหนังสือจำนวนทั้งหมด 17,395 เรคอร์ด โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยน้อยกว่า 2.5 ร้อยละ 97.6 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 – 2.9 ร้อยละ 2 และมีเพศชายเป็นส่วนใหญ่ ร้อยละ 99.7 โดยเป็นนักศึกษาชั้นปีที่ 4 ร้อยละ 39.2 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาประกาศนียบัตรบัณฑิตวิชาชีพครู ร้อยละ 16.9 นักศึกษาชอบยืมหนังสือในวันเสาร์ ร้อยละ 26.8 รองลงมาชอบยืมหนังสือในวันอาทิตย์ ร้อยละ 24.7 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 24.5 รองลงมาจะยืมหนังสือครั้งละ 3 เล่ม ร้อยละ 23.6 นักศึกษาในคลัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนสิงหาคม ร้อยละ 15.8 รองลงจะเป็นเดือนกรกฎาคม ร้อยละ 14.6 และนักศึกษาที่มีการยืมหนังสือของคลัสเตอร์ที่ 5 ส่วนใหญ่ยืมหนังสือหมวดสังคมศาสตร์ (300) ร้อยละ 31.9

ตารางที่ 4.11 แสดงข้อมูลเกรดเฉลี่ยของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
เกรดเฉลี่ย	ร้อยละ	เกรดเฉลี่ย	ร้อยละ	เกรดเฉลี่ย	ร้อยละ	เกรดเฉลี่ย	ร้อยละ	เกรดเฉลี่ย	ร้อยละ	เกรดเฉลี่ย	ร้อยละ
2.5 - 2.9	52.90	3.0 - 3.4	54.60	3.0 - 3.4	48.91	2.5 - 2.9	43.18	<2.5	97.68	3.0 - 3.4	45.51
3.0 - 3.4	30.20	2.5 - 2.9	24.40	2.5 - 2.9	24.90	3.0 - 3.4	31.51	2.5 - 2.9	2.07	3.5 - 4.0	26.21
<2.5	16.08	3.5 - 4.0	18.99	3.5 - 4.0	21.12	<2.5	16.64			2.5 - 2.9	17.54

ตารางที่ 4.12 แสดงข้อมูลหมวดหมู่ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
หมวดหมู่	ร้อยละ	หมวดหมู่	ร้อยละ	หมวดหมู่	ร้อยละ	หมวดหมู่	ร้อยละ	หมวดหมู่	ร้อยละ	หมวดหมู่	ร้อยละ
3	37.07	Fic	17.61	5	32.01	3	28.88	3	31.90	3	27.00
Fic	14.56	Sc	17.26	Fic	14.91	The	23.79	6	17.24	Fic	18.50
Sc	10.31	4	17.16	0	14.77	0	6.94	0	11.18	Sc	13.58

ตารางที่ 4.13 แสดงข้อมูลวันเยี่ยม (อาทิตย์ (1) ถึง เสาร์ (7)) ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
วันเยี่ยม	ร้อยละ	วันเยี่ยม	ร้อยละ	วันเยี่ยม	ร้อยละ	วันเยี่ยม	ร้อยละ	วันเยี่ยม	ร้อยละ	วันเยี่ยม	ร้อยละ
5	21.93	2	22.34	5	22.04	5	22.64	7	26.87	7	21.71
3	20.50	5	21.88	3	19.60	2	20.12	1	24.75	1	20.81
4	17.98	3	20.41	2	17.52	3	17.11	3	11.92	3	13.67

ตารางที่ 4.14 แสดงข้อมูลจำนวนหนังสือที่เยี่ยมแต่ละครั้งของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
จำนวนหนังสือ	ร้อยละ	จำนวนหนังสือ	ร้อยละ	จำนวนหนังสือ	ร้อยละ	จำนวนหนังสือ	ร้อยละ	จำนวนหนังสือ	ร้อยละ	จำนวนหนังสือ	ร้อยละ
2	33.00	2	30.91	2	33.82	1	77.04	2	24.55	3	31.00
3	25.29	3	28.65	3	25.72	2	18.60	3	23.62	2	21.01
1	19.44	1	16.87	1	19.17	3	4.18	1	19.94	5	20.75

ตารางที่ 4.15 แสดงข้อมูลเพศของแต่ละคลัสเตอร์

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
เพศ	ร้อยละ	เพศ	ร้อยละ	เพศ	ร้อยละ	เพศ	ร้อยละ	เพศ	ร้อยละ	เพศ	ร้อยละ
2	86.78	2	93.00	2	87.85	2	75.35	1	99.72	2	95.56
1	13.22	1	7.00	1	12.15	1	24.65			1	4.44

ตารางที่ 4.16 แสดงข้อมูลเดือนที่ยืม (มกราคม (1) ถึง ธันวาคม (12)) ของแต่ละคลัสเตอร์

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
เดือน ยืม	ร้อยละ	เดือน ยืม	ร้อยละ	เดือน ยืม	ร้อยละ	เดือน ยืม	ร้อยละ	เดือน ยืม	ร้อยละ	เดือน ยืม	ร้อยละ
7	21.68	11	21.58	7	20.97	6	20.84	8	15.81	1	13.82
11	16.73	7	14.05	11	16.16	11	14.54	7	14.63	11	13.48
8	13.57	12	13.45	6	15.21	8	13.03	6	13.30	8	12.13

ตารางที่ 4.17 แสดงข้อมูลชั้นปีที่กำลังศึกษาอยู่ของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
ชั้นปี	ร้อยละ	ชั้นปี	ร้อยละ	ชั้นปี	ร้อยละ	ชั้นปี	ร้อยละ	ชั้นปี	ร้อยละ	ชั้นปี	ร้อยละ
3	54.76	4	47.88	3	50.33	4	39.91	4	39.29	4	48.69
2	36.04	3	26.22	2	19.40	3	25.78	0	25.50	2	17.44
1	7.22	2	21.79	4	17.99	2	14.38	5	21.67	5	13.18

ตารางที่ 4.18 แสดงข้อมูลโปรแกรมวิชาที่สังกัดของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
สังกัด	ร้อยละ	สังกัด	ร้อยละ	สังกัด	ร้อยละ	สังกัด	ร้อยละ	สังกัด	ร้อยละ	สังกัด	ร้อยละ
4135	50.29	4135	85.85	4121	85.15	4135	44.16	5166	16.98	4111	23.08
2501	9.82	4111	8.23	4111	4.25	4121	18.77	6161	12.47	4101	16.92
4121	8.99	2361	2.31	2252	3.72	4101	9.20	2601	12.05	2601	16.84

ตารางที่ 4.19 แสดงข้อมูลความถี่ในการยืมหนังสือในแต่ละเดือนของแต่ละคลัสเตอร์ 3 ลำดับสูงสุด

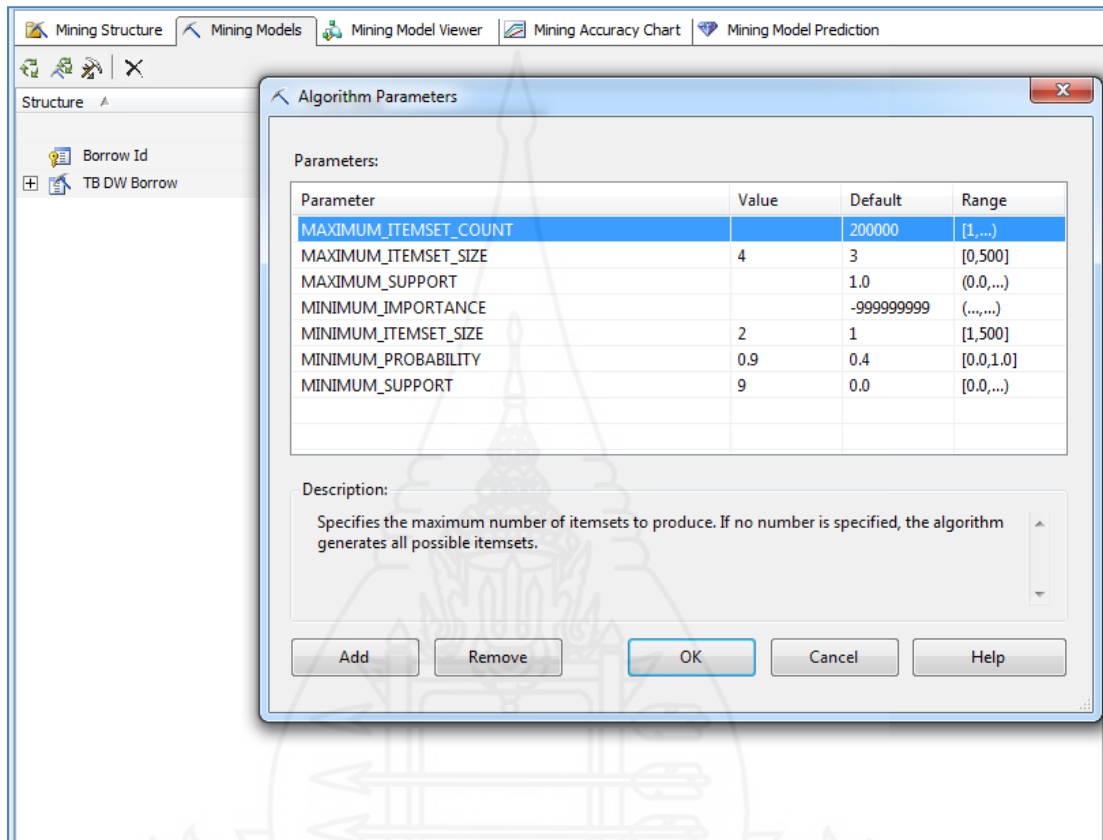
Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
ความถี่	ร้อยละ	ความถี่	ร้อยละ	ความถี่	ร้อยละ	ความถี่	ร้อยละ	ความถี่	ร้อยละ	ความถี่	ร้อยละ
5	10.01	12	6.29	3	7.45	1	24.91	3	10.82	5	8.05
3	9.79	11	5.84	4	6.89	2	19.59	2	10.01	4	7.45
4	9.63	14	5.21	7	6.01	3	12.48	4	8.17	3	7.17

ผลการวิเคราะห์ข้อมูลจากการใช้เทคนิค Clustering เราสามารถแสดงข้อมูลออกมา รายบุคคลได้ ทำให้ทราบว่าสมาชิกที่มีพฤติกรรมคล้ายกันจะอยู่กลุ่มเดียวกัน ตัวอย่าง นักศึกษา โปรแกรมวิชา (4101) เป็นเพศชาย ชอบยืมหนังสือวันพุธ เดือนกันยายน มีเกรดเฉลี่ยน้อยกว่า 2.5 จะถูกจัดอยู่ใน Cluster ที่ 5 ซึ่งการจัดกลุ่มจะแบ่งตามพฤติกรรมในการใช้บริการยืมหนังสือและ ข้อมูลส่วนตัวของนักศึกษา เช่น นักศึกษาที่สังกัดโปรแกรมวิชา (4101) แต่เป็นเพศหญิง มีความชอบ ยืมหนังสือในวันเสาร์ ชอบยืมหนังสือในเดือนกันยายนเหมือนกัน และมีเกรดเฉลี่ย 2.5 - 2.9 จะถูกจัด อยู่ใน Cluster ที่ 4 ดังภาพที่ 4.17

program_id	genier_id	day_id	month_id	Acc Gpa	\$CLUSTER
4101	1	4	9	<2.5	Cluster 5
4101	1	7	7	<2.5	Cluster 5
4101	1	6	8	<2.5	Cluster 5
4101	1	6	8	<2.5	Cluster 5
4101	1	6	8	<2.5	Cluster 5
4101	1	7	9	<2.5	Cluster 5
4101	1	7	9	<2.5	Cluster 5
4101	1	7	9	<2.5	Cluster 5
4101	2	7	9	2.5-2.9	Cluster 4
4121	1	6	11	<2.5	Cluster 5
4121	1	6	11	<2.5	Cluster 5
4121	1	6	11	<2.5	Cluster 5
2240	1	2	7	<2.5	Cluster 5
2240	1	2	7	<2.5	Cluster 5

ภาพที่ 4.17 แสดงข้อมูลสมาชิกที่มีพฤติกรรมคล้ายกันจะอยู่กลุ่มเดียวกัน

3.4.2 ผลการหาความสัมพันธ์ (Association Rule) เพื่อวิเคราะห์ข้อมูลการยืมหนังสือ หรือคั่นหารูปแบบความสัมพันธ์ของพฤติกรรมกรยืม - คั่นหนังสือของสมาชิก เพื่อหาความน่าจะเป็นในการยืมหนังสือว่าเล่มใดบ้างที่มีถูกยืมไปด้วยด้วยกัน ซึ่งมีข้อมูลการยืมคั่นจำนวน 64,026 ครั้ง โดยกำหนดค่าพารามิเตอร์ดังที่แสดงในรูปภาพที่ 4.18



ภาพที่ 4.18 กำหนดค่าพารามิเตอร์การหาความสัมพันธ์ (Association Rule)

MAXIMUM_ITEMSET_COUNT เป็นการกำหนดค่าสูงสุดของชุด Item ที่จะให้ อัลกอริธึมนี้ ถ้าไม่กำหนดค่านี้ไว้ อัลกอริธึมจะสร้างชุดของ Item ทั้งหมดโดยพิจารณาจากค่า MINIMUM_SUPPORT การกำหนดค่านี้จะช่วยป้องกันไม่ให้อัลกอริธึมสร้างจำนวนของชุด item ที่มีจำนวนมากเกินไป ในกรณีที่มีชุดของ Item มากเกินไป อัลกอริธึมจะเก็บเฉพาะชุดของ item เพียง n ชุดเท่านั้นโดยดูจากคะแนนของค่า importance เป็นพื้นฐาน

MAXIMUM_ITEMSET_SIZE เป็นการกำหนดจำนวนสูงสุดของ item ที่จะให้อัลกอริธึม สร้างขึ้นมาในชุดของ Item ในที่นี้กำหนดขนาดเท่ากับ 5 หมายความว่า อัลกอริธึมจะสร้างชุดของ item ขึ้นมาตั้งแต่ 1 item 2 items 3 items 4 items และสูงสุดไม่เกิน 5 items ในการกำหนดถ้า ปรับเปลี่ยนค่านี้ลดลงจะช่วยลดเวลาในการประมวลผลลง ค่าดีฟอลท์คือ 0 ซึ่งหมายความว่าไม่มีการ จำกัดขนาดของชุด Item

จากภาพที่ 4.19 แสดงผลข้อมูลการหาความสัมพันธ์ (Association Rule) โดยกำหนดความน่าจะเป็น (Probability) ไว้ที่ 0.9 หรือมีค่าความเชื่อมั่น (Confidence) ไม่น้อยกว่าร้อยละ 90 และกำหนดค่าความสำคัญ (Importance) ไว้ที่ 3.90 หรือมีค่าความน่าสนใจ (Lift) ไม่น้อยกว่า 3.90 พบว่ามีกฎความสัมพันธ์ที่เกิดขึ้นทั้งหมด 20 กฎ แสดงไว้ในตารางที่ 4.20

ตารางที่ 4.20 แสดงผลข้อมูลการหาความสัมพันธ์ (Association Rule)

ค่าความเชื่อมั่น	ค่าความน่าสนใจ	กฎความสัมพันธ์
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล, - สู่ทิศทางใหม่ของการบริหารการศึกษา - > - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education - and state of human knowledge: an analysis , - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis - > - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล, - สู่ทิศทางใหม่ของการบริหารการศึกษา - > - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา

ตารางที่ 4.20 (ต่อ)

ค่าความ เชื่อมั่น	ค่าความ น่าสนใจ	กฎความสัมพันธ์
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล, - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.075	- สู่ทิศทางใหม่ของการบริหารการศึกษา, - ความเป็นมืออาชีพในการจัดและบริหารการศึกษา ยุคปฏิรูป การศึกษา -> - การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล -> - ป่วย อึ้งภากรณ์: ทศนะว่าด้วยการศึกษา

ตารางที่ 4.20 (ต่อ)

ค่าความ เชื่อมั่น	ค่าความ น่าสนใจ	กฎความสัมพันธ์
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ, - การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง - > - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง, - บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis - > - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล -> - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
100 %	4.464	- บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้มนุษย์ = Education and state of human knowledge: an analysis , - สู่ทิศทางใหม่ของการบริหารการศึกษา -> - ป่วย อิงภรณ์: ทศนะว่าด้วยการศึกษา
91 %	3.961	- เมืองพระนคร นครวัด นครธม = Angkor: an introduction , - นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท -> - คู่มือนำเที่ยวนครวัด - นครธม
91 %	4.438	- คู่มือนำเที่ยวนครวัด - นครธม, - นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท -> - เมืองพระนคร นครวัด นครธม = Angkor: an introduction

จากตารางแสดงผลข้อมูลการหาความสัมพันธ์ (Association Rule) คือ ตารางที่ 4.20 สามารถวิเคราะห์ผลลัพธ์จากกฎความสัมพันธ์การยืมหนังสือของนักศึกษา ได้ดังนี้

บริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล แล้วจะยืม หนังสือเรื่อง ป่วย อึ้งภากรณ์: ทักษะว่าด้วยการศึกษาด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

กฎข้อที่ 18 ถ้าผู้ใช้ยืมหนังสือเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้ มนุษย์ = Education and state of human knowledge: an analysis และ หนังสือเรื่อง สูติศทางใหม่ของการบริหารการศึกษา แล้วจะยืม หนังสือเรื่อง ป่วย อึ้งภากรณ์: ทักษะว่าด้วยการศึกษาด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

กฎข้อที่ 19 ถ้าผู้ใช้ยืมหนังสือเรื่อง เมืองพระนคร นครวัด นครธม = Angkor : an introduction และหนังสือเรื่อง นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท แล้วจะยืม หนังสือเรื่อง คู่มือนำเที่ยวนครวัด - นครธมด้วยค่าความเชื่อมั่นร้อยละ 91 และค่าความน่าสนใจ 3.961

กฎข้อที่ 20 ถ้าผู้ใช้ยืมหนังสือเรื่อง คู่มือนำเที่ยวนครวัด - นครธม และ หนังสือเรื่อง นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท แล้วจะยืม หนังสือเรื่อง เมืองพระนคร นครวัด นครธม = Angkor: an introduction ด้วยค่าความเชื่อมั่นร้อยละ 91 และค่าความน่าสนใจ 4.438

จากผลการใช้กฎการหาความสัมพันธ์ (Association Rule) เพื่อวิเคราะห์ข้อมูลการยืมหนังสือ หรือค้นหารูปแบบความสัมพันธ์ของพฤติกรรมการยืม - คินหนังสือของสมาชิกพบว่าข้อมูลมีค่าความสัมพันธ์ (Support) ที่น้อยมาก ซึ่งมีค่าความสัมพันธ์สูงสุดเพียง 41 ครั้งเท่านั้น จึงอาจสรุปได้ว่าข้อมูลที่นำมาทดสอบปริมาณที่ไม่มากและการยืมหนังสือเล่มเดิมซ้ำๆ กันมีปริมาณที่น้อยเพราะมีการผลิตหนังสือเล่มใหม่ๆ ออกมาเยอะหรือมีการปรับปรุงข้อมูลหนังสือเล่มใหม่อยู่เสมอทำให้สมาชิกไม่ยืมหนังสือที่เป็นเล่มเก่าซ้ำกันหลายรอบ และผลการวิจัยพบว่าหนังสือที่ยืมมีความสัมพันธ์กันไม่เกิน 4 เล่มเท่านั้น อาจเป็นกับกฎระเบียบของห้องสมุดที่มีการจำกัดจำนวนการยืมหนังสือของสมาชิก ผู้วิจัยได้แสดงรายละเอียดการหาความสัมพันธ์ (Association Rule) ที่มีค่า item=2 ค่า item=3 และค่า item=4 โดยเลือกข้อมูล Item ที่มีค่า Support อันดับสูงสุด 10 อันดับแรกมาแสดง

Support	Size	Itemset
41	2	รักปมฯ ฉบับกวนเด็กทอ 2 /may112., รักปมฯ ฉบับกวนเด็กทอ 1 /may112 [นามแฝง].
37	2	ลิซิดแห่งฟ้าชะตาแห่งรัก /Belly Mommy., ลิซิดแห่งฟ้าชะตาแห่งรัก /เขียน: Belly Mommy ; แปล: LoUKpla .
31	2	การวัดผลการศึกษา /สมนึก ถัทยธณิน., หลักการวัดและประเมินผลการศึกษา /พิเชิด ฤทธิ์จรูญ.
31	2	ความเป็นครู ตามหลักสูตรสถาบันราชภัฏ ฉบับปรับปรุง พุทธศักราช 2536 รหัสวิชา 1011101 (ฉบับปรับปรุงครั้งที่ 2) /ยนต์ ชุมจิต., ความเป็นครู
30	2	คู่มือศ1 /คิดภา., คู่มือศ2 /คิดภา.
25	2	จิตวิทยาพัฒนาการชีวิตทุกช่วงวัย เล่ม 1 /ศรีเรือน แก้วกังวาล., ทฤษฎีจิตวิทยาพัฒนาการ /พรรณทิพย์ ศิริวรรณยศ.
20	2	การเรียนรู้ร่วมชั้นของเด็กพิเศษ/ประกอติ พูลพัฒน์., การศึกษาสำหรับเด็กพิเศษ /วาริ สิริจิตตร.
18	2	จิตวิทยาสร้างสรรคการเรียนการสอน /อาริ พันธมณ., จิตวิทยาการศึกษา /สุรางค์ ไคว์ตระกูล.
16	2	สสดีเพื่อการวิจัย /จาเรญ ฐานแก้ว., การวิจัยเบื้องต้น /บุญชม ศรีสะอาด.
16	2	การเรียนรู้สำหรับเด็กปฐมวัย /ศรียา นิยมธรรม., การศึกษาสำหรับเด็กพิเศษ /วาริ สิริจิตตร.

ภาพที่ 4.20 ผลการหาความสัมพันธ์ (Association Rule) ค่า item=2

การหาความสัมพันธ์ (Association Rule) สามารถวิเคราะห์ผลลัพธ์จากกฎความสัมพันธ์การยืมหนังสือที่มีจำนวน 2 Item โดยเลือกข้อมูล Item ที่มีค่า Support อันดับสูงสุด 10 อันดับแรกมีรายละเอียดดังนี้

ลำดับที่ 1. ถ้ายืมหนังสือเรื่อง รักป่วนๆ ฉบับกวนเด็กหอ 2 แล้วจะยืมหนังสือเรื่อง รักป่วนๆ ฉบับกวนเด็กหอ 1 มีค่าความสัมพันธ์ (Support) เท่ากับ 41

ลำดับที่ 2. ถ้ายืมหนังสือเรื่อง ลิขิตแห่งฟ้าชะตาแห่งรัก แล้วจะยืมหนังสือเรื่อง ลิขิตแห่งฟ้าชะตาแห่งรัก มีค่าความสัมพันธ์ (Support) เท่ากับ 37

ลำดับที่ 3. ถ้ายืมหนังสือเรื่อง ความเป็นครู ตามหลักสูตรสถาบันราชภัฏ ฉบับปรับปรุง พุทธศักราช 2536 รหัสวิชา 1011101 (ฉบับปรับปรุงครั้งที่ 2) แล้วจะยืมหนังสือเรื่อง ความเป็นครูไทย มีค่าความสัมพันธ์ (Support) เท่ากับ 31

ลำดับที่ 4. ถ้ายืมหนังสือเรื่อง การวัดผลการศึกษา แล้วจะยืมหนังสือเรื่อง หลักการวัดและประเมินผลการศึกษา มีค่าความสัมพันธ์ (Support) เท่ากับ 31

ลำดับที่ 5. ถ้ายืมหนังสือเรื่อง คู่พยศ1 แล้วจะยืมหนังสือเรื่อง คู่พยศ2 มีค่าความสัมพันธ์ (Support) เท่ากับ 30

ลำดับที่ 6. ถ้ายืมหนังสือเรื่อง จิตวิทยาพัฒนาการชีวิตทุกช่วงวัย เล่ม 1 แล้วจะยืมหนังสือเรื่อง ทฤษฎีจิตวิทยาพัฒนาการมีค่าความสัมพันธ์ (Support) เท่ากับ 25

ลำดับที่ 7. ถ้ายืมหนังสือเรื่อง การเรียนร่วมชั้นของเด็กพิเศษ แล้วจะยืมหนังสือเรื่อง การศึกษาสำหรับเด็กพิเศษ มีค่าความสัมพันธ์ (Support) เท่ากับ 20

ลำดับที่ 8. ถ้ายืมหนังสือเรื่อง จิตวิทยาสร้างสรรค์การเรียนการสอน แล้วจะยืมหนังสือเรื่อง จิตวิทยาการศึกษา มีค่าความสัมพันธ์ (Support) เท่ากับ 18

ลำดับที่ 9. ถ้ายืมหนังสือเรื่อง แคลคูลัส 3 แล้วจะยืมหนังสือเรื่อง เรขาคณิตวิเคราะห์และแคลคูลัส III มีค่าความสัมพันธ์ (Support) เท่ากับ 16

ลำดับที่ 10. ถ้ายืมหนังสือเรื่อง การศึกษาและความเป็นครูไทย แล้วจะยืมหนังสือเรื่อง ความเป็นครูไทย มีค่าความสัมพันธ์ (Support) เท่ากับ 16

Rules		Itemsets		Dependency Network	
Minimum support:	1	Filter Itemset:			
Minimum itemset size:	3	Show:	Show attribute name only		
Maximum rows:	2000	<input type="checkbox"/> Show long name			
A	Support	A	S	Itemset	
12	3			การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนันต์ชัย พงศ์สุวรรณ , สูทิศทางใหม่ของการบริหารการศึกษา /ชาญชัย อาจิมมาจาร, ความเป็นมืออาชีพ...	
11	3			เมืองพระนคร นครธม = Angkor : an introduction /อริชัย เชคเดส, แต่ง ; ปรานี วงษ์เทศ, แปล , คู่มือปฎิบัติของนครวัด-นครธม /ศรัณย์ บุญประเสริฐ...	
11	3			คำอธิบายกฎหมายลักษณะประกันภัย : ศึกษาแบบเรียงมาตรา /สรพล สุขทรศนันท์, คำอธิบายกฎหมายลักษณะหนี้ /อรพรม พันธ์พัฒนา, คำอธิบายกฎหมาย...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้น...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้น...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจ...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจ...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนันต์ชัย พงศ์สุวรรณ , สูทิศทาง...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนัน...	
9	3			การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนันต์ชัย พงศ์สุวรรณ , บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ = Education and state...	
9	3			บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ = Education and state of human knowledge : an analysis /เสนท์ จามริก , การบริหารการศึ...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ = Education and state...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนัน...	
9	3			การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนันต์ชัย พงศ์สุวรรณ , การบริหารการศึกษา : นโยบายและยุทธศาสตร์เพื่อการบรรลุ /วิโรจน์ สารรัตนะ, ...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจ...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนัน...	
9	3			การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ /มานะ ทองรักษุ์ ; ธีรชัย มาตจวิสัย บรรณาธิการ , การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนัน...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง /อนันต์ชัย พงศ์สุวรรณ , การบริหาร...	
9	3			ป่วย อึ้งภากรณ์ : ทัศนะว่าด้วยการศึกษา /ปกปiong จันวิทย์, บรรณาธิการ, บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ = Education and state...	

ภาพที่ 4.21 ผลการหาความสัมพันธ์ (Association Rule) ค่า item=3

การหาความสัมพันธ์ (Association Rule) สามารถวิเคราะห์ผลลัพธ์จากกฎความสัมพันธ์การยืมหนังสือที่มีจำนวน 3 Item โดยเลือกข้อมูล Item ที่มีค่า Support อันดับสูงสุด 10 อันดับแรกมีรายละเอียดดังนี้

ลำดับที่ 1. ถ้ายืมหนังสือเรื่อง การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง และเรื่อง สูทิศทางใหม่ของการบริหารการศึกษา แล้วจะยืมหนังสือเรื่อง ความเป็นมืออาชีพในการจัดและบริหารการศึกษา ยุคปฏิรูปการศึกษา มีค่าความสัมพันธ์ (Support) เท่ากับ 12

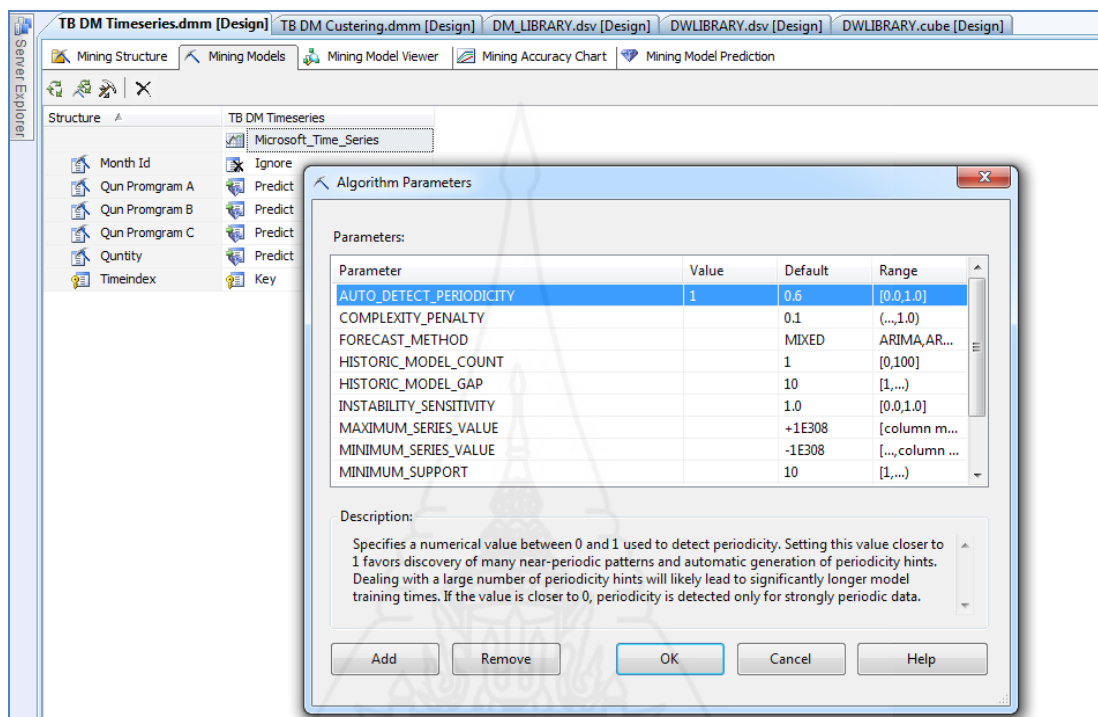
ลำดับที่ 2. ถ้ายืมหนังสือเรื่อง คำอธิบายกฎหมายลักษณะประกันภัย: ศึกษาแบบเรียงมาตรา และเรื่อง คำอธิบายกฎหมายลักษณะหนี้ แล้วจะยืมหนังสือเรื่อง คำอธิบายกฎหมาย จ้างแรงงาน จ้างทำของ รับขน มีค่าความสัมพันธ์ (Support) เท่ากับ 11

ลำดับที่ 3. ถ้ายืมหนังสือเรื่อง เมืองพระนคร นครวัด นครธม และเรื่อง คู่มือปฎิบัติของนครวัด - นครธม แล้วจะยืมหนังสือเรื่อง นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท มีค่าความสัมพันธ์ (Support) เท่ากับ 11

ลำดับที่ 4. ถ้ายืมหนังสือเรื่อง ป่วย อึ้งภากรณ์: ทัศนะว่าด้วยการศึกษา และเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ แล้วจะยืมหนังสือเรื่อง สูทิศทางใหม่ของการบริหารการศึกษา มีค่าความสัมพันธ์ (Support) เท่ากับ 9

ลำดับที่ 5. ถ้ายืมหนังสือเรื่อง การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง และเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสถานะความรู้นุษย์ แล้วจะยืมหนังสือเรื่อง สูทิศทางใหม่ของการบริหารการศึกษา มีค่าความสัมพันธ์ (Support) เท่ากับ 9

3.4.3 ผลการพยากรณ์ไทม์ซีรีส์ (Time Series) ค่าตัวเลขการยืมหนังสือในอนาคต โดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษา โดยกำหนดค่าพารามิเตอร์ดังที่แสดงในภาพที่ 4.23



ภาพที่ 4.23 กำหนดค่าพารามิเตอร์ไทม์ซีรีส์ (Time Series)

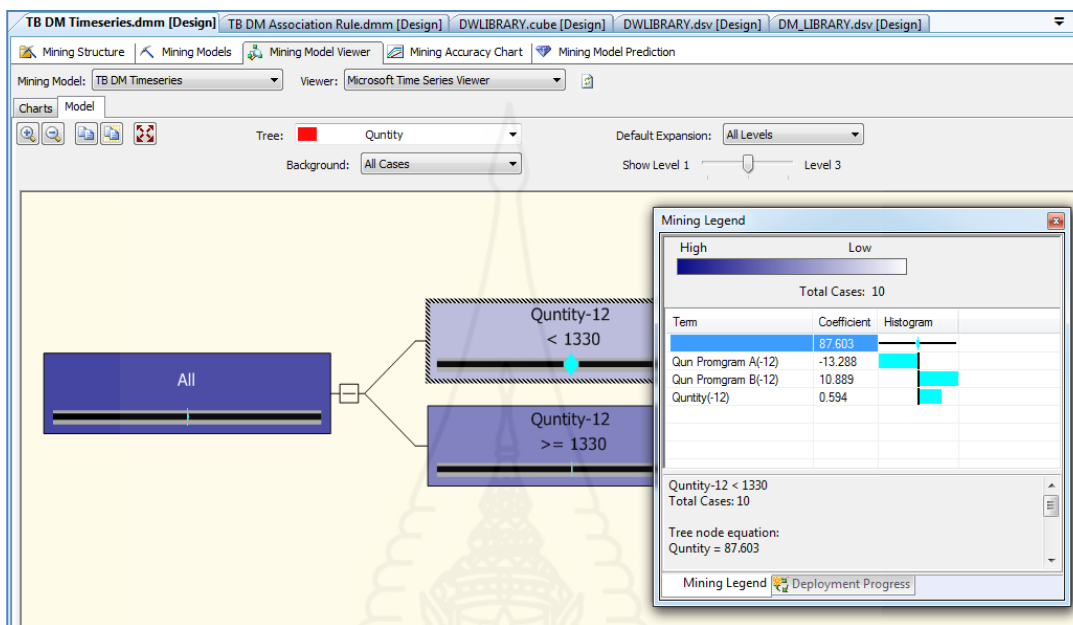
AUTO_DETECT_PERIODICITY เป็นค่าที่ใช้ในตรวจหาช่วงระยะเวลา ถ้ากำหนดค่าเข้าใกล้ 1 จะค้นหาแพทเทิร์นของข้อมูลทั้งหมดที่มีช่วงระยะเวลาเข้ามาเกี่ยวข้อง แต่ถ้ากำหนดค่าเข้าใกล้ 0 การตรวจหาแพทเทิร์นระยะเวลาที่เกี่ยวข้องกับข้อมูลจะแสดงเฉพาะข้อมูลที่มีระยะเวลาเข้ามาเกี่ยวข้องมากๆ เท่านั้น

COMPLEXITY_PENALTY เป็นค่าที่ใช้ในการควบคุมการขยายขนาดของ Auto regression trees ค่าที่กำหนดจะอยู่ในช่วง [0,1]

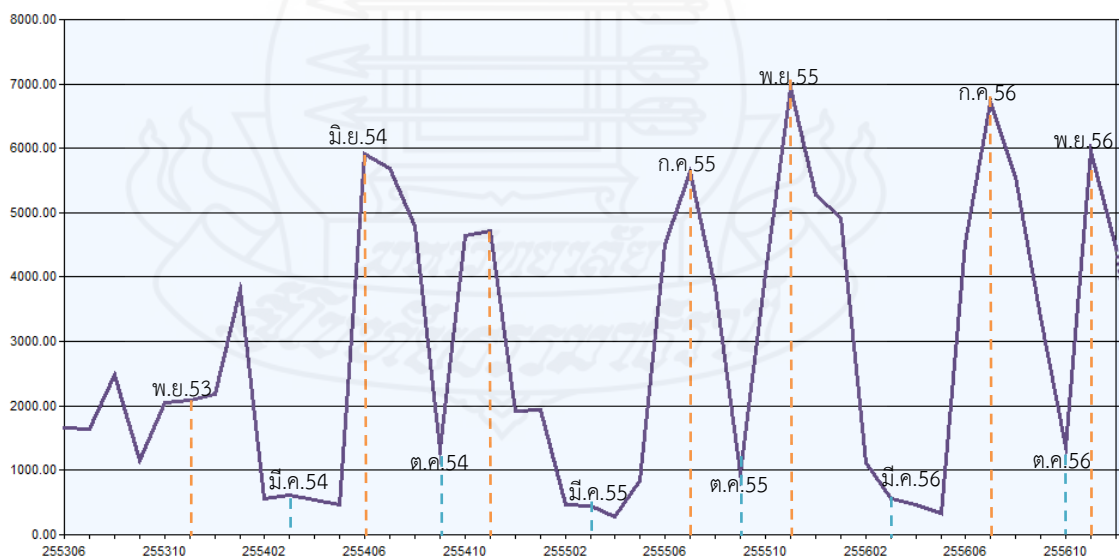
MISSING_VALUE_SUBSTITUTION กำหนดวิธีการที่จะใช้ในการเติมค่าที่สูญหายไป วิธีการที่มีให้เลือกได้แก่ การใช้ค่าของช่วงเวลาที่แล้ว (previous value) การใช้ค่าเฉลี่ย (mean value) การกำหนดค่าคงที่ (constant value)

PERIODICITY_HINT เป็นค่าที่บอกอัลกอริทึมเกี่ยวกับลักษณะของข้อมูลว่ามีเรื่องของฤดูกาลเข้ามาเกี่ยวข้องด้วย โดยรูปแบบที่กำหนดคือ {n [, n]} ซึ่งค่าใน [] เป็นค่าที่ใส่หรือไม่ใส่ก็ได้ และค่า n คือ ค่าตัวเลขจำนวนเต็มหรือค่าตัวเลขทศนิยมใดๆ เช่น เงินรายได้ของบริษัทที่นำมาประมวลผลมีแพทเทิร์นเป็นรายไตรมาสและรายปี ดังนั้นควรกำหนดค่าเป็น {3, [12]}

ผลที่ได้จากการใช้เทคนิคในการพยากรณ์การยืมหนังสือของนักศึกษา สามารถดูในรูปแบบ Decision Trees ดังภาพที่ 4.24 หรือแสดงผลในรูปแบบกราฟแสดงแนวโน้มของจำนวนนักศึกษาที่ใช้บริการยืมหนังสือ ดังภาพที่ 4.25



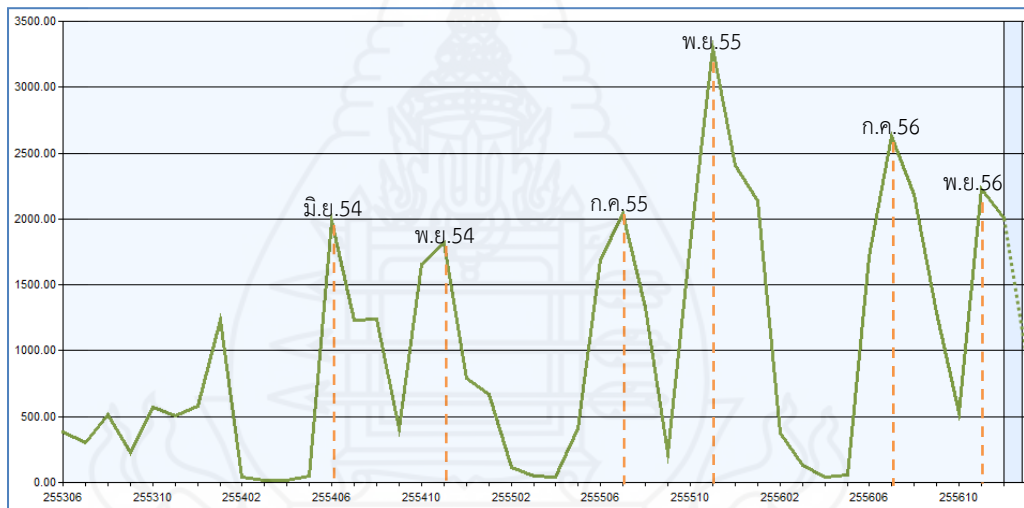
ภาพที่ 4.24 รูปแบบ Decision Trees ของจำนวนนักศึกษาที่ใช้บริการยืมหนังสือ



ภาพที่ 4.25 รูปแบบกราฟแสดงแนวโน้มของจำนวนนักศึกษาที่ใช้บริการยืมหนังสือ

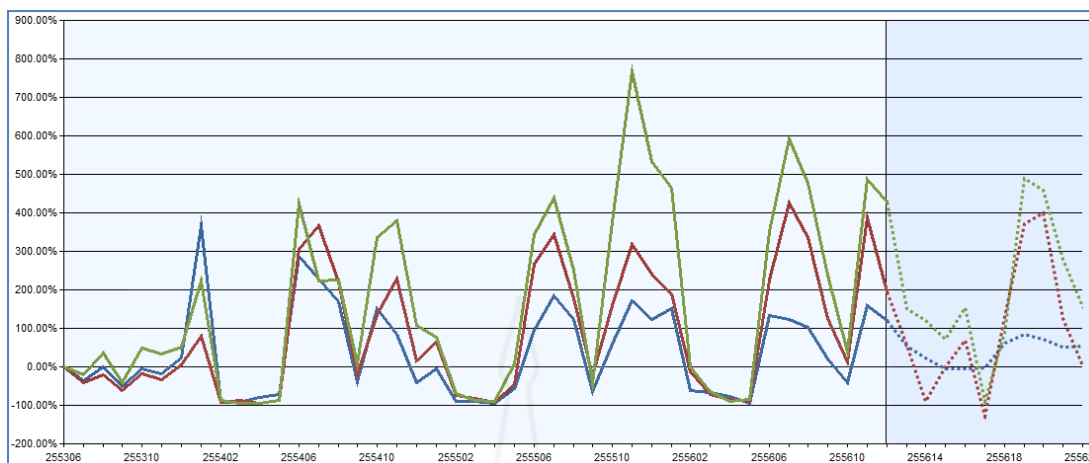
จากรูปภาพที่ 4.25 สามารถสรุปผลการวิจัยตามเส้นกราฟที่แสดงผลเปรียบเทียบแนวโน้มการยืมหนังสือของนักศึกษาในแต่ละเดือน ซึ่งเป็นข้อมูลระหว่างเดือนมิถุนายน ปี 2553 ถึงเดือนธันวาคม ปี 2556 เส้นกราฟแสดงให้เห็นชัดเจนว่าในแต่ละเดือนมีการยืมคืนหนังสือในรูปแบบเดียวกัน คือ นักศึกษามีการยืมหนังสือในช่วงเดือนพฤษภาคมของทุกปีและเพิ่มปริมาณขึ้นเรื่อยๆจนถึงเดือนมิถุนายนและเดือนกรกฎาคมจากนั้นนักศึกษาจะมีการยืมหนังสือน้อยลงเรื่อยๆจนถึงเดือนกันยายนแล้วจึงเริ่มมีการยืมหนังสืออีกครั้งและเพิ่มขึ้นเรื่อยๆ จนถึงเดือนพฤศจิกายนจากนั้นการยืมหนังสือก็จะเริ่มลดลงเรื่อยๆ จนถึงเดือนกุมภาพันธ์จากนั้นนักศึกษาจะยืมหนังสือสม่ำเสมอแต่เป็นปริมาณที่ไม่มากนักจนถึงเดือนเมษายน ซึ่งสามารถสรุปได้ว่านักศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียน

ทั้งนี้เพื่อศึกษาข้อมูลช่วงเวลาการยืมคืนที่ชัดเจนขึ้นผู้วิจัยได้ทำการศึกษาข้อมูลการยืมคืนหนังสือของนักศึกษาแต่ละโปรแกรมวิชา โดยผู้วิจัยเลือกข้อมูลของโปรแกรมวิชาที่มีการยืมคืนหนังสือสูงสุด ได้แก่ โปรแกรมวิชาสังคมศึกษา (4135) โปรแกรมวิชาคณิตศาสตร์ (4121) และโปรแกรมวิชาภาษาอังกฤษ (4111)



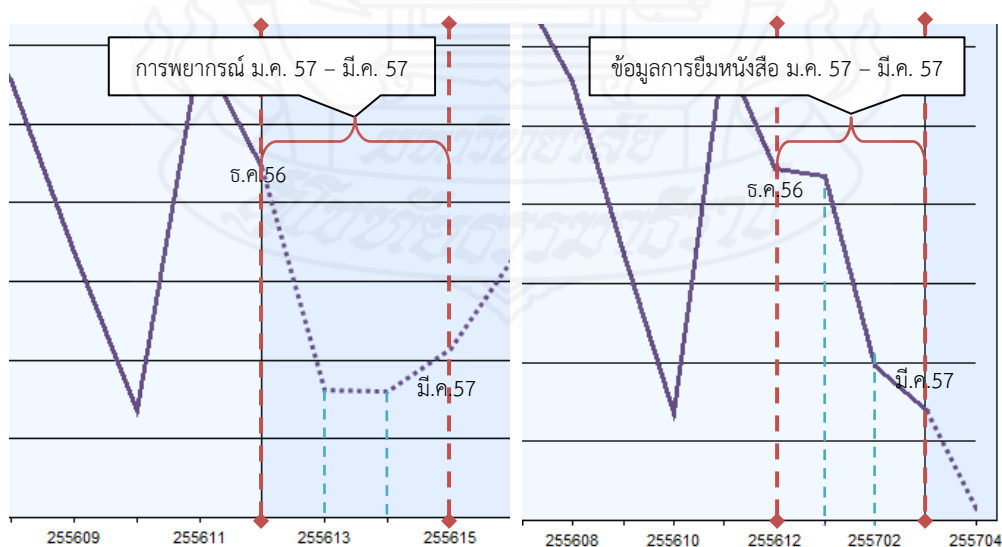
ภาพที่ 4.26 รูปแบบกราฟแสดงแนวโน้มมีการใช้บริการยืมหนังสือโปรแกรมวิชาสังคมศึกษา (4135)

จากรูปภาพที่ 4.26 สามารถสรุปผลการวิจัยตามเส้นกราฟแสดงให้เห็นชัดเจนว่าในแต่ละเดือนมีการยืมคืนหนังสือในรูปแบบเดียวกัน คือ นักศึกษาโปรแกรมวิชาสังคมศึกษามีการยืมหนังสือในช่วงเดือนพฤษภาคมของทุกปีและเพิ่มปริมาณขึ้นเรื่อยๆจนถึงเดือนมิถุนายนและเดือนกรกฎาคมจากนั้นนักศึกษาโปรแกรมวิชาสังคมศึกษาจะมีการยืมหนังสือน้อยลงเรื่อยๆจนถึงเดือนกันยายนแล้วจึงเริ่มมีการยืมหนังสืออีกครั้งและเพิ่มขึ้นเรื่อยๆ จนถึงเดือนพฤศจิกายนจากนั้นการยืมหนังสือก็จะเริ่มลดลงเรื่อยๆ จนถึงเดือนกุมภาพันธ์จากนั้นนักศึกษาโปรแกรมวิชาสังคมศึกษาจะยืมหนังสือสม่ำเสมอแต่เป็นปริมาณที่ไม่มากนักจนถึงเดือนเมษายน ซึ่งสามารถสรุปได้ว่านักศึกษาโปรแกรมวิชาสังคมศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียนเหมือนกับข้อมูลการยืมคืนหนังสือโดยรวมของนักศึกษามหาวิทยาลัยราชภัฏศรีสะเกษ



ภาพที่ 4.27 รูปแบบกราฟแสดงแนวโน้มการใช้บริการยืมหนังสือทั้ง 3 โปรแกรมวิชา

จากภาพที่ 4.27 แสดงให้เห็นว่าแนวโน้มการยืมหนังสือของแต่ละโปรแกรมวิชาทั้ง 3 โปรแกรมวิชา คือ โปรแกรมวิชาสังคมศึกษา (4135) โปรแกรมวิชาคณิตศาสตร์ (4121) และโปรแกรมวิชาภาษาอังกฤษ (4111) โดยเปรียบเทียบข้อมูลการยืมหนังสือแต่ละโปรแกรมวิชาจะพบว่าตามเส้นกราฟแสดงให้เห็นชัดเจนว่าในแต่ละเดือนมีการยืมคืนหนังสือในรูปแบบเดียวกันหรือเป็นการเคลื่อนไหวเป็นไปค่าการผันแปรตามวัฏจักร (Cyclical Variation) เพราะนักศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียน ซึ่งจากกราฟแสดงข้อมูลการพยากรณ์ข้อมูลในอนาคตจะเห็นได้ว่ามีรูปแบบการยืมหนังสือที่ไม่แตกต่างกันเลยซึ่งผู้วิจัยได้นำข้อมูลช่วงเดือนมกราคม 2557 ถึงเดือนมีนาคม 2557 มาเปรียบเทียบกับข้อมูลการพยากรณ์ ดังภาพที่ 4.28

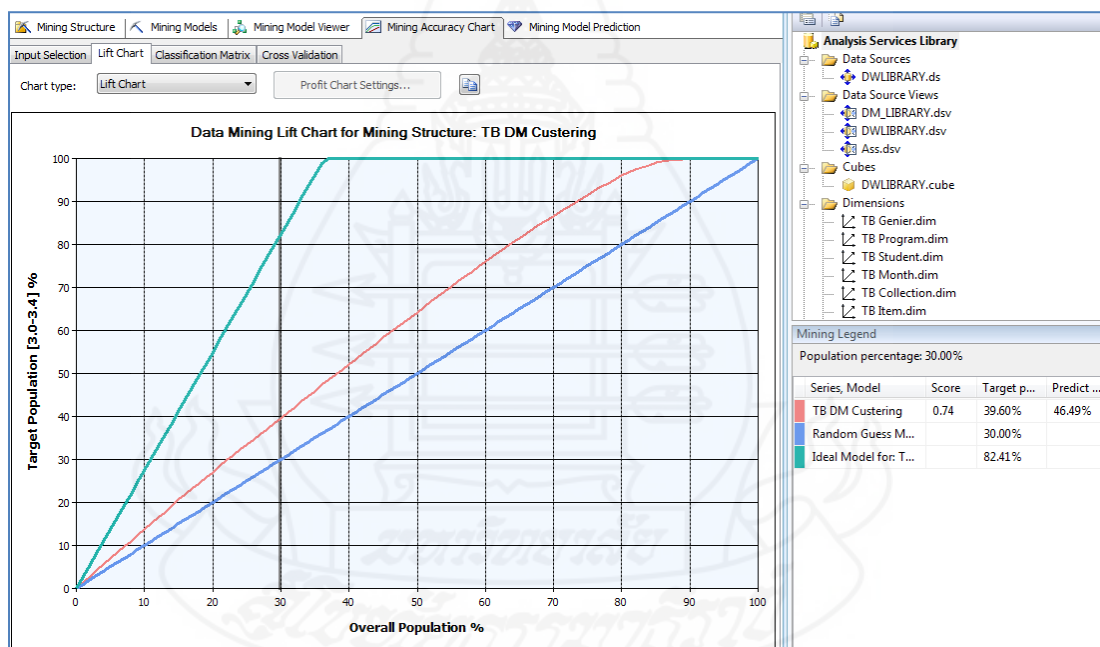


ภาพที่ 4.28 กราฟแสดงข้อมูลการพยากรณ์ข้อมูลในอนาคต

3.5 ประเมิน

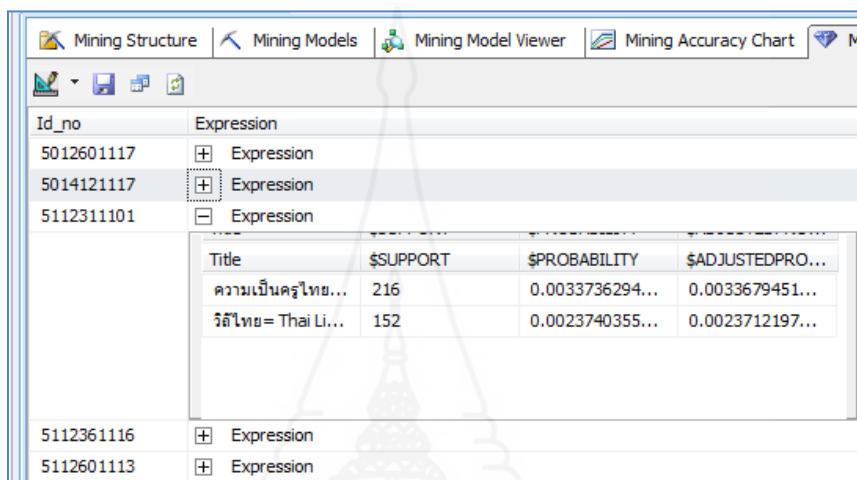
เป็นการประเมินผลลัพธ์จากแบบจำลองและอัลกอริธึมที่ใช้วิเคราะห์ข้อมูลว่าครอบคลุมและสามารถตอบวัตถุประสงค์ที่กำหนดไว้หรือไม่ จากนั้นจึงนำเสนอประเมินแบบจำลองที่ได้จากการทำเหมืองข้อมูลเพื่อพิจารณาถึงความเหมาะสม ในการนำแบบจำลองไปประยุกต์ใช้ว่าแบบจำลองที่ได้นั้น มีความแม่นยำในการทำนายมากน้อยเพียงใด โดยในการประเมินแบบจำลอง นั้น จะแบ่งตามลักษณะของการทำเหมืองข้อมูล ดังนี้

3.5.1 การแบ่งกลุ่ม (Clustering) สามารถประเมินผล โดยการนำโมเดลการทำเหมืองข้อมูลนั้นไปทำการทดสอบด้วยชุดข้อมูลสำหรับการทดสอบ (Testing data set) ซึ่งได้กำหนดเอาไว้ 30% เพื่อดูว่าแพตเทิร์นได้จากการสร้างโมเดลการทำเหมืองข้อมูลนั้นมีความถูกต้องเพียงใด การทดสอบความถูกต้องของการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม (Clustering) ใน SQL Server นั้น สามารถดูจากลิฟต์ชาร์ต (Lift Chart) ผลการทดสอบพบว่าโมเดลการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม มีความน่าจะเป็นของการพยากรณ์ 46.49% และมีค่า Score ที่ใช้วัดประสิทธิภาพของโมเดลเท่ากับ 0.74 ดังภาพที่ 4.29



ภาพที่ 4.29 แสดงลิฟต์ชาร์ต (Lift Chart) การทดสอบโมเดลการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม

3.5.2 กฎความสัมพันธ์ สามารถประเมินโมเดลการทำเหมืองข้อมูลนี้ โดยการนำโมเดลนี้ไปพยากรณ์ด้วยการสร้างแบบสอบถามหรือคิวรี (Query) ซึ่งใช้ฟังก์ชันในการพยากรณ์ชื่อว่าการพยากรณ์ความสัมพันธ์ (Predict Association) และกำหนดเงื่อนไขที่ต้องการการแสดงผล ตัวอย่างนี้ผู้วิจัยได้กำหนดเงื่อนไขให้นำข้อมูลมาแสดงผลคือ ข้อมูลรหัสนักศึกษา และรายการหนังสือเพียง 2 ชื่อเรื่องที่ชอบยืมไปด้วยกันพร้อมทั้งแสดงค่าสถิติด้วย ดังภาพที่ 4.30

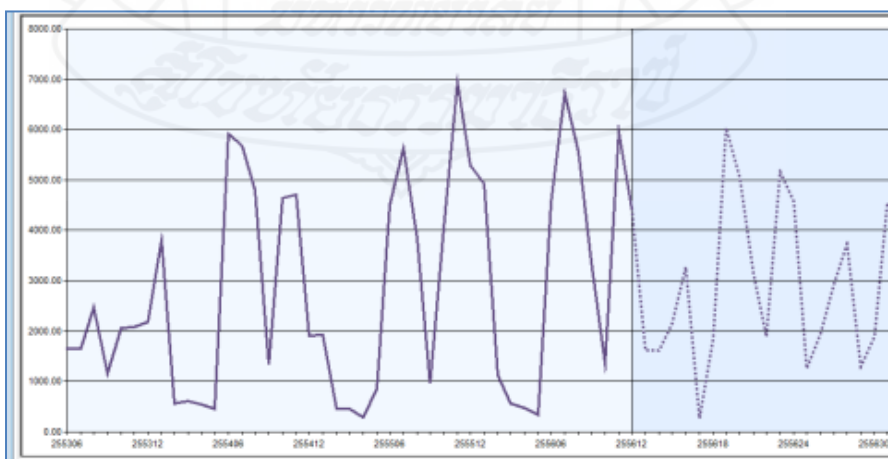


Id_no	Expression
5012601117	+
5014121117	+
5112311101	-
5112361116	+
5112601113	+

Title	\$SUPPORT	\$PROBABILITY	\$ADJUSTEDPRO...
ความเป็นครูไทย...	216	0.0033736294...	0.0033679451...
วิถีไทย= Thai Li...	152	0.0023740355...	0.0023712197...

ภาพที่ 4.30 แสดงการสร้างแบบสอบถามหรือคิวรี (Query)

3.5.3 ไทม์ซีรีส์ (Time Series) สามารถประเมินโมเดลการทำเหมืองข้อมูลนี้ โดยการนำโมเดลนี้ไปพยากรณ์แนวโน้มการยืมหนังสือในอนาคต จะเห็นว่าข้อมูลที่ได้จากกราฟ ดังภาพที่ 4.31 มีรูปแบบกราฟที่คล้ายกับข้อมูลเดิมที่มีอยู่ ซึ่งทำให้มั่นใจได้ว่าการใช้ ไทม์ซีรีส์ (Time Series) ในการพยากรณ์นั้นให้ความแม่นยำและสามารถนำไปใช้ในการพยากรณ์แนวโน้มการยืมหนังสือในอนาคตได้



ภาพที่ 4.31 แสดงการพยากรณ์แนวโน้มการยืมหนังสือในอนาคต

3.6 นำไปใช้งาน

เป็นการนำผลลัพธ์หรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ไปใช้งานจริง ทำให้สามารถจำแนกกลุ่มผู้ใช้บริการและทราบข้อมูลการใช้บริการของผู้ใช้บริการแต่ละกลุ่มในด้านต่างๆ โดยเฉพาะด้านการค้นคืนสารสนเทศ จากการศึกษาพฤติกรรมกรายหนังสือของนักศึกษาโดยการสร้างคลังข้อมูล และทำเหมืองข้อมูล ซึ่งสามารถแบ่งกลุ่มผู้ใช้ออกเป็น 6 กลุ่ม และการหาความสัมพันธ์ของรายการหนังสือที่มักถูกยืมไปด้วยกันโดยเลือกข้อมูลที่มีค่าความเชื่อมั่นไม่ต่ำกว่าร้อยละ 90 รวมถึงการสร้างรูปแบบการรายงานหลายมิติเพื่อตอบสนองความต้องการของผู้ใช้ได้ ผู้วิจัยจึงนำเสนอผลของการทำเหมืองข้อมูลมาประยุกต์ใช้ เพื่อสนับสนุนการค้นคืนสารสนเทศภายในห้องสมุดดังต่อไปนี้

3.6.1 การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้นคืนสารสนเทศ

โดยการแนะนำหนังสือที่เหมาะสมสำหรับนักศึกษาตามกลุ่มต่างๆ 6 กลุ่มที่มีการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม (Clustering) โดยแนวทางการประยุกต์ใช้คือดำเนินการปรับปรุงเว็บโอแพ็ค (WebOPAC) เมื่อสมาชิกดำเนินการเข้าสู่ระบบโดยการ Login ระบบจะดำเนินการเชื่อมโยงข้อมูลสมาชิกห้องสมุดกับกฎการแบ่งกลุ่มที่สร้างขึ้น จะทำให้ทราบได้ว่าสมาชิกอยู่กลุ่ม (Cluster) ไດและระบบจะแนะนำหนังสือที่เหมาะสมสำหรับนักศึกษาแต่ละคน ดังภาพที่ 4.32

The screenshot shows the library's web interface. At the top, it says 'Sisaket Rajabhat University Library' and 'welcome Mr. Teerapong'. Below the search bar, it shows 'Search Results' with 'Found : 19 Titles'. On the left, there is a user profile for 'Mr. Teerapong', Administrator, Office: Library sskru. The main content area is titled 'Books for you' and lists several book recommendations with their covers and details:

- Book 1:** Title: 200 เมนูเด็ดที่ขีดเบาหวาน =200 diabetic recipes /ชื่อทิพย์วรรณ พันธุ์แก้ว. Author: ชื่อทิพย์วรรณ พันธุ์แก้ว,บรรณาธิการ. Published: กรุงเทพมหานคร.2008 [2551]. Call Number: 641.5631 ส473. Location: มหาวิทยาลัยราชภัฏศรีสะเกษ.
- Book 2:** Title: Principles of accounting /Belverd E. Needles, Marian Powers, Susan V. Crosson. Author: Powers, Marian. Published: [S.l.] :South-Western College Pub.2008. Call Number: 657 P957 2008. Location: มหาวิทยาลัยราชภัฏศรีสะเกษ.
- Book 3:** Title: AutoCAD 2008 essentials /by Munir M. Hamad. Author: Hamad, Munir M. Published: Hingham, MA :Infinity Science Press, c2008. Call Number: 620.0042 H352A 2008. Location: มหาวิทยาลัยราชภัฏศรีสะเกษ.
- Book 4:** Title: Colors /Suluck Visavapattamawon, editor. Author: Suluck Visavapattamawon.

ภาพที่ 4.32 ระบบแนะนำหนังสือที่เหมาะสมสำหรับนักศึกษาแต่ละคน

3.6.2 การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้นคืนสารสนเทศ

โดยการแนะนำหนังสือที่มีความสัมพันธ์กันจากการทำเหมืองข้อมูลด้วยการหาความสัมพันธ์ (Association rule) ของหนังสือแต่ละเล่มที่มักถูกยืมไปด้วยกัน โดยแนวทางการประยุกต์ใช้คือ ดำเนินการปรับปรุงเว็บโอแพ็ค (WebOPAC) เมื่อนักศึกษาสืบค้นข้อมูลหนังสือที่ต้องการได้หนึ่งเล่ม เมื่อนักศึกษาเลือกหนังสือเล่มใดเล่มหนึ่งแล้ว ระบบจะทำการแนะนำหนังสือที่มักถูกยืมคู่กับหนังสือที่ถูกเลือกนั้นเสมอ เป็นการเพิ่มความสะดวกในการค้นคืนสารสนเทศและได้สารสนเทศหรือหนังสือที่ต้องการอย่างรวดเร็วไม่ต้องเสียเวลาในการสืบค้นใหม่หลายๆ รอบ ตัวอย่างจากกฎข้อที่ 1 ถ้าผู้ใช้ยืมหนังสือเรื่อง การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ และ หนังสือเรื่อง การบริหาร การศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล และ หนังสือเรื่อง สู่ทิศทางใหม่ของการบริหาร การศึกษา แล้วจะยืมหนังสือเรื่อง ป่วย อังภาภรณ์: ทศนะว่าด้วยการศึกษา ด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

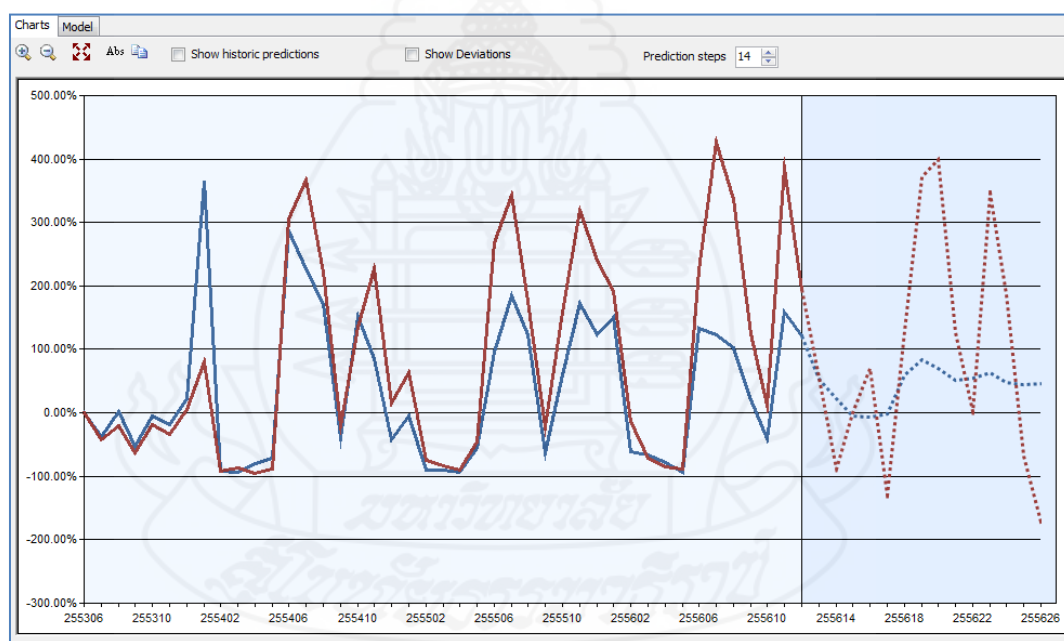
เมื่อนักศึกษาสืบค้นและเลือกหนังสือชื่อเรื่อง “การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ” ระบบจะแนะนำหนังสือที่มีความสัมพันธ์กันให้กับนักศึกษา ได้แก่ หนังสือเรื่อง “การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล” หนังสือเรื่อง “สู่ทิศทางใหม่ของการบริหารการศึกษา” และหนังสือเรื่อง “ป่วย อังภาภรณ์: ทศนะว่าด้วยการศึกษา” โดยอัตโนมัติ ดังภาพที่ 4.33

The screenshot shows the library's interface for a book titled "การบริหารจัดการโรงเรียนเพื่อพัฒนาสู่ความเป็นเลิศ" (Management of Schools for Development to Excellence) by M. Thongkiet. The book details include: Publication: กรุงเทพฯ: อารักษ์, 2549; Edition: พิมพ์ครั้งที่ 1; Detail: 205 หน้า, 26 ซม.; Subject: โรงเรียนการบริหาร(+); Source Types: Book. The interface also displays statistics (8 ค.ศ. 2555, 11 copies, 1 issue), tags, and a "Related Items" section with three recommended book covers: "การบริหารการศึกษา" (Management of Education), "สู่ทิศทางใหม่ของการบริหารการศึกษา" (New Direction of Educational Management), and "ป่วย อังภาภรณ์" (Ploy Angpakorn).

ภาพที่ 4.33 ระบบแนะนำหนังสือที่มักถูกยืมคู่กับหนังสือที่ถูกเลือก

3.6.3 การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการคั่นคืนสารสนเทศ

โดยการแนะนำหนังสือหรือจัดกิจกรรมส่งเสริมการยืมหนังสือตามช่วงเดือนต่างๆ เพื่อให้ผู้ใช้บริการหรือนักศึกษามีความสนใจหรือได้หนังสือที่ต้องการ โดยแนวทางการประยุกต์ใช้คือการจัดกิจกรรมการส่งเสริมกิจกรรมการยืมหนังสือให้แต่ละโปรแกรมวิชาที่มีการยืมหนังสือน้อยในแต่ละเดือน เพราะแต่ละโปรแกรมมีการยืมหนังสือไม่เท่ากัน ตัวอย่าง เมื่อเปรียบเทียบข้อมูลการยืมหนังสือของโปรแกรมวิชาสังคมศึกษา (4135) กับ โปรแกรมวิชาคณิตศาสตร์ (4121) ดังภาพที่ 4.34 พบว่าการยืมหนังสือของทั้งสองโปรแกรมมีลักษณะคล้ายกันน่าจะเกิดจากนโยบายการยืมคืนหนังสือของห้องสมุดที่กำหนดช่วงเวลาให้ยืมหรือเกิดจากการเปิดปิดภาคเรียนของนักศึกษา แต่เมื่อดูจากเส้นกราฟโปรแกรมวิชาคณิตศาสตร์จะมีการยืมหนังสือที่อยู่ในปริมาณที่สูงและมีการยืมเพิ่มขึ้นเรื่อยๆ ซึ่งมากกว่าโปรแกรมวิชาสังคมศึกษาที่มีการยืมหนังสือลดลงเรื่อยๆ หากดูกราฟที่เป็นจุดไขว่ปลาที่เป็นกราฟการพยากรณ์แนวโน้มในการยืมหนังสือ จะทำให้เห็นชัดเจนเลยว่าโปรแกรมวิชาสังคมศึกษาจะมีการยืมหนังสือที่น้อยลง จากผลที่ได้บรรณารักษ์อาจนำข้อมูลที่ได้มาจัดกิจกรรมส่งเสริมการให้บริการกับนักศึกษาโปรแกรมวิชาสังคมศึกษาเพื่อให้เข้าใช้บริการยืมหนังสือมากขึ้นได้



ภาพที่ 4.34 กราฟเปรียบเทียบข้อมูลการยืมหนังสือของโปรแกรมวิชาสังคมศึกษากับ โปรแกรมคณิตศาสตร์

บทที่ 5

สรุปผลการวิจัย

การศึกษาวิจัยเรื่อง การทำเหมืองข้อมูลเพื่อสนับสนุนการบริการคั่นคืนสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ ผู้วิจัยได้พัฒนาคลังข้อมูลนักศึกษาที่มาใช้บริการยืมหนังสือของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษเพื่อนำมาวิเคราะห์หาความรู้ใหม่โดยการทำเหมืองข้อมูลได้ข้อสรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะดังนี้

1. สรุปผลงานวิจัย

1.1 การสร้างคลังข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด

งานวิจัยนี้ผู้วิจัยนำข้อมูลจากฐานข้อมูลของโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib ที่จัดเก็บข้อมูลการบริการยืมคืนสารสนเทศภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และข้อมูลส่วนตัวนักศึกษาจากฐานข้อมูลทะเบียนมหาวิทยาลัยราชภัฏศรีสะเกษ นำข้อมูลที่ได้มาสร้างระบบคลังข้อมูล โดยมีการนำข้อมูลมาจากรฐานข้อมูลโปรแกรมห้องสมุดอัตโนมัติ WalaiAutolib และฐานข้อมูลงานทะเบียน มาผ่านระบบอีทีแอล (Extract Transform and Load: ETL) ซึ่งผู้วิจัยได้ออกแบบส่วนของการดึงข้อมูลให้มีการคัดเลือกข้อมูลแต่ละตารางเพียงส่วนใดส่วนหนึ่งของตารางเฉพาะข้อมูลที่จำเป็นเท่านั้น โดยการทำความเข้าใจข้อมูลที่นำเข้าสู่คลังข้อมูล เพื่อจัดการกับโครงสร้างของข้อมูลและจัดการกับ format ข้อมูล เพื่อกำหนดเงื่อนไขต่างๆในการตรวจสอบข้อมูล ความถูกต้องของข้อมูลแต่ละตาราง หากข้อมูลมีความไม่ถูกต้องจะทำการเปลี่ยนแปลงหรือตัดข้อมูลออกไปถือว่าเป็นการทำความสะอาดข้อมูลให้มีความถูกต้องสมบูรณ์เพื่อให้แน่ใจว่าข้อมูลมีความถูกต้องแล้วจึงนำข้อมูลที่ได้เข้าสู่คลังข้อมูล ซึ่งในการทำระบบอีทีแอล (ETL) ผู้วิจัยใช้เครื่องมือ SQL Server Integration Service การนำข้อมูลเข้าสู่คลังข้อมูล

1.2 การทำ Online Analytical Processing (OLAP)

ผู้วิจัยได้ใช้เครื่องมือ SQL Server Analysis Service (SSAS) เพื่อสร้าง Analysis Service Project สำหรับวิเคราะห์สร้าง OLAP Cube ที่เป็นฐานข้อมูลหลายมิติ และกำหนดข้อมูลจากตารางข้อมูลที่อยู่ในคลังข้อมูลที่ได้ออกแบบโครงสร้างคลังข้อมูล แบบสโนว์เฟลก (Snowflake Schema) ที่มีตารางข้อเท็จจริง (Fact Table) 1 ตารางคือ ตาราง TB_Borrow และมีตารางมิติ (Dimension) 7 ตาราง คือ ตาราง TB_FBYBook ตาราง TB_Collection ตาราง TB_Day ตาราง TB_Genier ตาราง TB_Item ตาราง TB_Month ตาราง TB_Program และตาราง TB_Student และผู้ใช้สามารถนำข้อมูลมาแสดงผลแบบหลายมิติในลักษณะของ Pivot Table ได้ และเรียกดูข้อมูลการยืมหนังสือของนักศึกษาตามคุณลักษณะของนักศึกษาหรือเรียกดูข้อมูลตามพฤติกรรมการยืมหนังสือของนักศึกษาได้แบบหลายมิติ เพราะเมื่อสร้าง Cube แล้วจะสามารถดูข้อมูลที่เป็นของการประมวลผลเชิงวิเคราะห์แบบออนไลน์ได้ ในลักษณะหลายมิติผ่านแท็บ Browser ทั้งนี้ผู้ใช้สามารถ

เปลี่ยน Measures และ Dimension ที่ต้องการวิเคราะห์โดยลากจากหน้าต่างทางซ้ายมือไปวางในหน้าต่างทางขวามือ เพื่อดูข้อมูลต่างๆ เกี่ยวกับนักศึกษาที่มีการยืมคืนหนังสือได้ในหลายมิติ

1.3 การทำเหมืองข้อมูล

ผู้วิจัยได้ใช้เครื่องมือ SQL Server Analysis Service (SSAS) เพื่อสร้าง Analysis Service Project เพื่อค้นหาข้อมูลที่ซ่อนอยู่ในคลังข้อมูลงานวิจัยนี้ดำเนินการตามแบบจำลองของคริสป์ - ดีเอ็ม (CRISP - DM Model)

1.3.1 ทำความเข้าใจธุรกิจ ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องและศึกษาข้อมูลเกี่ยวกับการค้นคืนสารสนเทศภายในห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ และวิเคราะห์ปัญหาที่เกิดขึ้นจริงกับการปฏิบัติงานของห้องสมุด พบว่าสารสนเทศไม่เพียงพอต่อความต้องการของผู้ใช้ในช่วงเหตุการณ์ เช่น ช่วงเวลาสอบ หนังสือไม่เพียงพอต่อการให้บริการแก่ผู้ใช้บริการ เพราะไม่สามารถพยากรณ์จำนวนนักศึกษาที่มาใช้บริการห้องสมุดในแต่ละช่วงเวลาได้ ทำให้ขาดข้อมูลประกอบในการปรับปรุงและเตรียมการบริการ นักศึกษาค้นหาหนังสือไม่ตรงตามความต้องการ หากมีการแบ่งกลุ่มนักศึกษาออกเป็นกลุ่มๆ จะทำให้สามารถค้นหาหนังสือที่น่าสนใจในแต่ละกลุ่มได้และนักศึกษาได้หนังสือที่ตรงตามความต้องการอีกด้วย หากนักศึกษาต้องการหนังสือหลายเล่มต้องเสียเวลาในการสืบค้นข้อมูลหลายรอบ จึงทำให้เสียเวลามากในการค้นหาข้อมูลหนังสือ

1.3.2 ทำความเข้าใจข้อมูล ผู้วิจัยดำเนินการรวบรวมข้อมูลที่เกี่ยวข้องเพื่อนำไปวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งมีด้วยกันหลายอัลกอริทึม และแต่ละอัลกอริทึมก็มีวัตถุประสงค์ที่แตกต่างกัน ผู้วิจัยจึงได้สร้างคลังข้อมูลขึ้นมาดังที่กล่าวมาแล้วนั้น ซึ่งเป็นข้อมูลหนังสือข้อมูลสมาชิกและข้อมูลการใช้บริการยืมหนังสือของสมาชิกภายในห้องสมุด

1.3.3 เตรียมข้อมูล ผู้วิจัยดำเนินการเมื่อสร้างคลังข้อมูล (Data Warehouse) ผู้วิจัยได้นำข้อมูลที่มีการจัดเก็บไว้มาปรับโครงสร้างที่เหมาะสมเพื่อใช้กับแบบจำลองการทำเหมืองข้อมูลตามอัลกอริทึมต่างๆ เพราะข้อมูลที่จำเป็นต้องใช้ในงานวิจัย ซึ่งประกอบด้วย 3 อัลกอริทึม คือ 1) คลัสเตอร์ริง (Clustering) 2) แอสโซซิเอชันรูลส์ (Association Rules) และ 3) ไทม์ซีรีส์ (Time Series) ใช้เครื่องมือ SQL Server Analysis Service (SSAS) เพื่อสร้าง Analysis Service Project เพื่อค้นหาข้อมูลที่ซ่อนอยู่ในคลังข้อมูล

1) *การจัดกลุ่ม (Clustering)* เป็นการจัดการข้อมูลเพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ สำหรับข้อมูลที่ใช้เป็นข้อมูลเกี่ยวกับสมาชิกและข้อมูลเกี่ยวกับการการใช้บริการยืมคืนของสมาชิกโดยดึงข้อมูลจากตาราง TB_DM_Custering ซึ่งประกอบด้วยข้อมูลที่จัดเก็บรหัสนักศึกษา โปรแกรมวิชา เพศ ช่วงเกรดเฉลี่ย ชั้นปี วันที่ยืม วันยืมเดือนที่ยืม รหัสหนังสือ หมวดหมู่หนังสือ จำนวนเล่มที่ยืมต่อเดือนและจำนวนเล่มที่ยืมในแต่ละครั้ง ซึ่งมีข้อมูลการใช้บริการยืมคืนจำนวน 122,763 รายการ

2) *กฎการหาความสัมพันธ์ (Association Rule)* เป็นการจัดการข้อมูลเพื่อวิเคราะห์ข้อมูลการใช้บริการยืมหนังสือ ว่าหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยด้วยกัน สำหรับข้อมูลที่ใช้เป็นข้อมูลจาก 2 ตาราง คือ ตาราง TB_FBYBook เป็นข้อมูลเกี่ยวกับการยืมหนังสือ ซึ่งประกอบไปด้วยข้อมูลรหัสนักศึกษา วันที่ยืม จำนวนเล่มที่ยืมและรหัสการยืม กับตาราง TB_DW_Borrow เป็นข้อมูล

เกี่ยวกับหนังสือที่มีการยืมประกอบด้วยข้อมูลรหัสการยืม รหัสหนังสือ ชื่อหนังสือ รหัสเล่มหนังสือ และหมวดหมู่หนังสือ

3) *ไทม์ซีรีส์ (Time Series)* เป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคตโดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาในอนาคต สำหรับข้อมูลที่ใช้เป็นข้อมูลเกี่ยวกับสถิติหรือข้อมูลช่วงเวลาเป็นข้อมูลจากตาราง TB_DM_Timeseries เป็นข้อมูลเกี่ยวกับการเก็บข้อมูลการใช้บริการยืมคืนของนักศึกษาในแต่ละเดือนตั้งแต่ช่วงเดือน มิถุนายน 2553 ถึง เดือน ธันวาคม 2556 ซึ่งเป็นข้อมูลการยืมคืนทั้งหมดและข้อมูลการยืมคืนของนักศึกษา 3 โปรแกรมวิชาที่มีการยืมสูงสุด

1.3.4 สร้างแบบจำลอง

การวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining ได้แก่การสร้างแบบจำลองที่สอดคล้องกับอัลกอริธึมแต่ละชนิด ซึ่งจากการศึกษาและทำความเข้าใจกับข้อมูลแล้ว คณะผู้วิจัยได้เลือกอัลกอริธึมที่จะนำมาประมวลผล 3 อัลกอริธึม

1) *การจัดกลุ่ม (Clustering)* เพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกัน หรือมีข้อมูลทั่วไปเหมือนกันออกเป็น 6 คลัสเตอร์ จากประชากรที่นำมาประมวลผล โดยใช้เทคนิค Clustering มีข้อมูลจำนวน 122,763 เกรอर्ड โดยภาพรวมนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 - 3.4 และมีเพศหญิงเป็นส่วนใหญ่ โดยเป็นนักศึกษาชั้นปีที่ 3 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ชอบยืมหนังสือในวันพฤหัสบดี และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 - 3 เล่ม ผู้วิจัยได้ศึกษารายละเอียดคลัสเตอร์ที่ 2 เพราะมีความน่าสนใจ คือมีจำนวนนักศึกษาใช้บริการการยืมหนังสือมากที่สุดจำนวน 25,335 เกรอर्ड โดยนักศึกษาที่มีการยืมหนังสือจะมีเกรดเฉลี่ยอยู่ระหว่าง 3.0 - 3.4 ร้อยละ 54.6 รองลงมาจะมีเกรดเฉลี่ยอยู่ระหว่าง 2.5 - 2.9 ร้อยละ 24.4 และมีเพศหญิงเป็นส่วนใหญ่ ร้อยละ 93 โดยเป็นนักศึกษาชั้นปีที่ 4 ร้อยละ 47.8 และส่วนใหญ่เป็นนักศึกษาโปรแกรมวิชาสังคมศึกษา ร้อยละ 85.8 นักศึกษาชอบยืมหนังสือในวันจันทร์ ร้อยละ 22.3 รองลงมาชอบยืมหนังสือในวันพฤหัสบดี ร้อยละ 21.8 และในการยืมหนังสือแต่ละครั้งส่วนใหญ่จะยืมหนังสือครั้งละ 2 เล่ม ร้อยละ 30.9 รองลงมาจะยืมหนังสือครั้งละ 3 เล่ม ร้อยละ 28.6 นักศึกษาในคลัสเตอร์นี้ชอบยืมหนังสือในช่วงเดือนพฤศจิกายน ร้อยละ 21.5 รองลงจะเป็นเดือนกรกฎาคม ร้อยละ 14 และนักศึกษาที่มีการยืมหนังสือ คลัสเตอร์ที่ 2 ส่วนใหญ่ยืมหนังสือหมวดนวนิยาย (Fic) ร้อยละ 17.6 แสดงให้เห็นว่านักศึกษาในคลัสเตอร์นี้มีจำนวนมากและเกรดเฉลี่ยอยู่ในระดับที่ดีด้วย แต่กลับมีการยืมหนังสือหมวดนวนิยายมากด้วยและผลการวิเคราะห์ข้อมูลจากการใช้เทคนิค Clustering ยังแสดงข้อมูลออกมารายบุคคลได้ ทำให้ทราบว่าสมาชิกที่มีพฤติกรรมคล้ายกันจะอยู่กลุ่มเดียวกัน ตัวอย่าง นักศึกษาโปรแกรมวิชา (4101) เป็นเพศชาย ชอบยืมหนังสือวันพุธ เดือนกันยายน มีเกรดเฉลี่ยน้อยกว่า 2.5 จะถูกจัดอยู่ใน Cluster ที่ 5 ซึ่งการจัดกลุ่มจะแบ่งตามพฤติกรรมในการใช้บริการยืมหนังสือและข้อมูลส่วนตัวของนักศึกษา เช่น นักศึกษาที่สังกัดโปรแกรมวิชา (4101) แต่เป็นเพศหญิง มีความชอบยืมหนังสือในวันเสาร์ ชอบยืมหนังสือในเดือนกันยายนเหมือนกัน และมีเกรดเฉลี่ย 2.5 - 2.9 จะถูกจัดอยู่ใน Cluster ที่ 4

2) *กฎการหาความสัมพันธ์ (Association Rule)* เพื่อวิเคราะห์ข้อมูลการยืมหนังสือ หรือค้นหารูปแบบความสัมพันธ์ของพฤติกรรมยืม - คืนหนังสือของสมาชิก เพื่อหา

กฎข้อที่ 16 ถ้าผู้ใช้ยืมหนังสือเรื่อง การบริหารการศึกษาในคลื่นความเปลี่ยนแปลง และ หนังสือเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับปัญหาสภาวะความรู้มนุษย์ = Education and state of human knowledge: an analysis แล้วจะยืม หนังสือเรื่อง ป่วย อึ้ง ภากรณ์: ทศนะว่าด้วยการศึกษาด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

กฎข้อที่ 17 ถ้าผู้ใช้ยืมหนังสือเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับ ปัญหาสภาวะความรู้มนุษย์ = Education and state of human knowledge: an analysis และ หนังสือเรื่อง การบริหารการศึกษา: นโยบายและยุทธศาสตร์เพื่อการบรรลุผล แล้วจะยืม หนังสือเรื่อง ป่วย อึ้ง ภากรณ์: ทศนะว่าด้วยการศึกษาด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

กฎข้อที่ 18 ถ้าผู้ใช้ยืมหนังสือเรื่อง บทวิเคราะห์ว่าด้วยการศึกษากับ ปัญหาสภาวะความรู้มนุษย์ = Education and state of human knowledge: an analysis และ หนังสือเรื่อง สู่ทิศทางใหม่ของการบริหารการศึกษา แล้วจะยืม หนังสือเรื่อง ป่วย อึ้ง ภากรณ์: ทศนะว่าด้วยการศึกษาด้วยค่าความเชื่อมั่นร้อยละ 100 และค่าความน่าสนใจ 4.464

กฎข้อที่ 19 ถ้าผู้ใช้ยืมหนังสือเรื่อง เมืองพระนคร นครวัด นครธม = Angkor: an introduction และหนังสือเรื่อง นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท แล้ว จะยืม หนังสือเรื่อง คู่มือนำเที่ยวนครวัด - นครธมด้วยค่าความเชื่อมั่นร้อยละ 91 และค่าความ น่าสนใจ 3.961

กฎข้อที่ 20 ถ้าผู้ใช้ยืมหนังสือเรื่อง คู่มือนำเที่ยวนครวัด - นครธม และ หนังสือเรื่อง นครวัด นครธม: กัมพูชา ดินแดนแห่งปราสาท แล้วจะยืม หนังสือเรื่อง เมืองพระนคร นครวัด นครธม = Angkor: an introduction ด้วยค่าความเชื่อมั่นร้อยละ 91 และค่าความน่าสนใจ 4.438

จากผลการใช้กฎการหาความสัมพันธ์ (Association Rule) เพื่อ วิเคราะห์ข้อมูลการยืมหนังสือ หรือค้นหารูปแบบความสัมพันธ์ของพฤติกรรมการยืม - คืนหนังสือของ สมาชิกพบว่าข้อมูลมีค่าความสัมพันธ์ (Support) ที่น้อยมาก ซึ่งมีค่าความสัมพันธ์สูงสุดเพียง 41 ครั้ง เท่านั้น จึงอาจสรุปได้ว่าข้อมูลที่นำมาทดสอบปริมาณที่ไม่มากและการยืมหนังสือเล่มเดิมซ้ำๆ กันมี ปริมาณที่น้อยเพราะมีการผลิตหนังสือเล่มใหม่ๆ ออกมาเยอะหรือมีการปรับปรุงข้อมูลหนังสือเล่มใหม่ อยู่เสมอทำให้สมาชิกไม่ยืมหนังสือที่เป็นเล่มเก่าซ้ำกันหลายรอบ และผลการวิจัยพบว่าหนังสือที่ยืมมี ความสัมพันธ์กันไม่เกิน 4 เล่มเท่านั้น อาจเป็นกับกฎระเบียบของห้องสมุดที่มีการจำกัดจำนวนการยืม หนังสือของสมาชิก

3) *ไทม์ซีรีส์ (Time Series)* การพยากรณ์ค่าตัวเลขการยืมหนังสือใน อนาคต โดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาใน อนาคต ซึ่งเปรียบเทียบแนวโน้มการยืมหนังสือของนักศึกษาในแต่ละเดือน ซึ่งเป็นข้อมูลระหว่างเดือน มิถุนายน ปี 2553 ถึงเดือนธันวาคม ปี 2556 เส้นกราฟแสดงให้เห็นชัดเจนว่าในแต่ละเดือนมีการยืม คืนหนังสือในรูปแบบเดียวกัน คือ นักศึกษามีการยืมหนังสือในช่วงเดือนพฤษภาคมของทุกปีและเพิ่ม ปริมาณขึ้นเรื่อยๆจนถึงเดือนมิถุนายนและเดือนกรกฎาคมจากนั้นนักศึกษาจะมีการยืมหนังสือน้อยลงเรื่อยๆจนถึงเดือนกันยายนแล้วจึงเริ่มมีการยืมหนังสืออีกครั้งและเพิ่มขึ้นเรื่อยๆจนถึงเดือนพฤศจิกายน จากนั้นการยืมหนังสือก็จะเริ่มลดลงเรื่อยๆ จนถึงเดือนกุมภาพันธ์จากนั้นนักศึกษาจะยืมหนังสือ

สม่ำเสมอแต่เป็นปริมาณที่ไม่มากนักจนถึงเดือนเมษายน ซึ่งสามารถสรุปได้ว่านักศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียน

ทั้งนี้เพื่อศึกษาข้อมูลช่วงเวลาการยืมคืนที่ชัดเจนขึ้นผู้วิจัยได้ทำการศึกษาข้อมูลการยืมคืนหนังสือของนักศึกษาแต่ละโปรแกรมวิชา โดยผู้วิจัยเลือกข้อมูลของโปรแกรมวิชาที่มีการยืมคืนหนังสือสูงสุด ได้แก่ โปรแกรมวิชาสังคมศึกษา(4135) โปรแกรมวิชาคณิตศาสตร์ (4121) และโปรแกรมวิชาภาษาอังกฤษ (4111) และผลการวิเคราะห์ข้อมูลในแต่ละเดือนมีการยืมคืนหนังสือในรูปแบบเดียวกันกับนักศึกษาส่วนใหญ่ คือ นักศึกษาโปรแกรมวิชาสังคมศึกษามีการยืมหนังสือในช่วงเดือนพฤษภาคมของทุกปีและเพิ่มปริมาณขึ้นเรื่อยๆจนถึงเดือนมิถุนายนและเดือนกรกฎาคม จากนั้นนักศึกษาโปรแกรมวิชาสังคมศึกษาจะมีการยืมหนังสือน้อยลงเรื่อยๆจนถึงเดือนกันยายนแล้วจึงเริ่มมีการยืมหนังสืออีกครั้งและเพิ่มขึ้นเรื่อยๆจนถึงเดือนพฤศจิกายนจากนั้นการยืมหนังสือก็จะเริ่มลดลงเรื่อยๆ จนถึงเดือนกุมภาพันธ์จากนั้นนักศึกษาโปรแกรมวิชาสังคมศึกษาจะยืมหนังสือสม่ำเสมอแต่เป็นปริมาณที่ไม่มากนักจนถึงเดือนเมษายน ซึ่งสามารถสรุปได้ว่านักศึกษาโปรแกรมวิชาสังคมศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียนเหมือนกับข้อมูลการยืมคืนหนังสือโดยรวมของนักศึกษามหาวิทยาลัยราชภัฏศรีสะเกษ ซึ่งสรุปได้ว่าเป็นการเคลื่อนไหวเป็นไปค่าการผันแปรตามวัฏจักร (Cyclical Variation) เพราะนักศึกษามีการยืมคืนหนังสือในช่วงที่มีการเรียนการสอนและไม่ค่อยใช้บริการยืมคืนหนังสือในช่วงที่ปิดภาคเรียน ซึ่งจากกราฟแสดงข้อมูลการพยากรณ์ข้อมูลในอนาคตจะเห็นได้ว่ามีรูปแบบการยืมหนังสือที่ไม่แตกต่างกัน

1.3.5 ประเมิน

ผู้วิจัยดำเนินการประเมินโดยแบ่งตามลักษณะของการทำเหมืองข้อมูล ดังนี้

1) การแบ่งกลุ่ม (Clustering) ประเมินผล โดยการนำโมเดลการทำเหมืองข้อมูลนั้นไปทำการทดสอบด้วยชุดข้อมูลสำหรับการทดสอบ (Testing data set) ซึ่งได้กำหนดเอาไว้ 30% เพื่อดูว่าแพตเทิร์นได้จากการสร้างโมเดลการทำเหมืองข้อมูลนั้นมีความถูกต้องเพียงใด การทดสอบความถูกต้องของการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม (Clustering) ใน SQL Server นั้นสามารถดูจากลิฟต์ชาร์ต (Lift Chart) ผลการทดสอบพบว่าโมเดลการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม มีความน่าจะเป็นของการพยากรณ์ 46.49% และมีค่า Score ที่ใช้วัดประสิทธิภาพของโมเดลเท่ากับ 0.74

2) กฎความสัมพันธ์ ประเมินโมเดลการทำเหมืองข้อมูลนี้ โดยการนำโมเดลนี้ไปพยากรณ์ด้วยการสร้างแบบสอบถามหรือคิวรี (Query) ซึ่งใช้ฟังก์ชันในการพยากรณ์ชื่อว่าการพยากรณ์ความสัมพันธ์ (Predict Association) และกำหนดเงื่อนไขที่ต้องการการแสดงผล ตัวอย่างนี้ผู้วิจัยได้กำหนดเงื่อนไขให้หาข้อมูลมาแสดงผลคือ ข้อมูลรหัสนักศึกษา และรายการหนังสือเพียง 2 ชื่อเรื่องที่ขอบยืมไปด้วยกันพร้อมทั้งแสดงค่าสถิติด้วย ซึ่งแสดงผลได้ถูกต้อง

3) ไลน์ซีรีส์ (Time Series) สามารถประเมินโมเดลการทำเหมืองข้อมูลนี้ โดยการนำโมเดลนี้ไปพยากรณ์แนวโน้มการยืมหนังสือในอนาคต ผลที่แสดงมีรูปแบบกราฟที่คล้ายกับข้อมูลเดิมที่มีอยู่ ซึ่งทำให้มั่นใจได้ว่าการใช้ ไลน์ซีรีส์ (Time Series) ในการพยากรณ์นั้นให้ความแม่นยำและสามารถนำไปใช้ในการพยากรณ์แนวโน้มการยืมหนังสือในอนาคตได้

1.3.6 นำไปใช้งาน

ผู้วิจัยจึงนำเสนอผลของการทำเหมืองข้อมูลมาประยุกต์ใช้เพื่อสนับสนุนการค้นคืนสารสนเทศภายในห้องสมุด โดยใช้ข้อมูลจากการวิเคราะห์ข้อมูลด้วยเทคนิค Data mining ไปใช้งานจริง ทำให้สามารถจำแนกกลุ่มผู้ใช้บริการและทราบข้อมูลการใช้บริการของผู้ใช้บริการแต่ละกลุ่มในด้านต่างๆ โดยเฉพาะด้านการค้นคืนสารสนเทศ จากการศึกษาพฤติกรรมการยืมหนังสือของนักศึกษาโดยการสร้างคลังข้อมูล และทำเหมืองข้อมูล ซึ่งสามารถแบ่งกลุ่มผู้ใช้ออกเป็น 6 กลุ่ม และการหาความสัมพันธ์ของรายการหนังสือที่มีถูกยืมไปด้วยกันโดยเลือกข้อมูลที่มีความเชื่อมั่นไม่ต่ำกว่าร้อยละ 90 รวมถึงการสร้างรูปแบบการรายงานหลายมิติเพื่อตอบสนองความต้องการของผู้ใช้ได้ดังต่อไปนี้

1) การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้นคืนสารสนเทศ โดยการแนะนำหนังสือที่เหมาะสมสำหรับนักศึกษาตามกลุ่มต่างๆ 6 กลุ่มที่มีการทำเหมืองข้อมูลด้วยการแบ่งกลุ่ม (Clustering) โดยแนวทางการประยุกต์ใช้คือดำเนินการปรับปรุงเว็บโอแพ็ค (Web OPAC) เมื่อสมาชิกดำเนินการเข้าสู่ระบบโดยการ Login ระบบจะดำเนินการเชื่อมโยงข้อมูลสมาชิกห้องสมุดกับกฎการแบ่งกลุ่มที่สร้างขึ้น จะทำให้ทราบได้ว่าสมาชิกอยู่กลุ่ม (Cluster) ไດและระบบจะแนะนำหนังสือที่เหมาะสมสำหรับนักศึกษาแต่ละคน

2) การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้นคืนสารสนเทศ โดยการแนะนำหนังสือที่มีความสัมพันธ์กันจากการทำเหมืองข้อมูลด้วยการหาความสัมพันธ์ (Association rule) ของหนังสือแต่ละเล่มที่มีถูกยืมไปด้วยกัน โดยแนวทางการประยุกต์ใช้คือดำเนินการปรับปรุงเว็บโอแพ็ค (Web OPAC) เมื่อนักศึกษาสืบค้นข้อมูลหนังสือที่ต้องการได้หนึ่งเล่ม เมื่อนักศึกษาเลือกหนังสือเล่มใดเล่มหนึ่งแล้ว ระบบจะทำการแนะนำหนังสือที่มีถูกยืมคู่กับหนังสือที่ถูกเลือกนั้นเสมอ เป็นการเพิ่มความสะดวกในการค้นคืนสารสนเทศและได้สารสนเทศหรือหนังสือที่ต้องการอย่างรวดเร็วไม่ต้องเสียเวลาในการสืบค้นใหม่หลายๆรอบ

3) การนำไปใช้เป็นแนวทางในการปรับปรุงการให้บริการค้นคืนสารสนเทศ โดยการแนะนำหนังสือหรือจัดกิจกรรมส่งเสริมการยืมหนังสือตามช่วงเดือนต่างๆ เพื่อให้ผู้ใช้บริการหรือนักศึกษามีความสนใจหรือได้หนังสือที่ต้องการ โดยแนวทางการประยุกต์ใช้คือการจัดกิจกรรมการส่งเสริมกิจกรรมการยืมหนังสือให้แต่ละโปรแกรมวิชาที่มีการยืมหนังสือน้อยในแต่ละเดือน เพราะแต่ละโปรแกรมมีการยืมหนังสือไม่เท่ากัน ผลที่ได้บรรณารักษ์อาจนำข้อมูลที่ได้มาจัดกิจกรรมส่งเสริมการใช้บริการกับนักศึกษาเพื่อให้เข้าใช้บริการยืมหนังสือมากขึ้นได้

2. การอภิปรายผล

2.1 การสร้างคลังข้อมูลนักศึกษาที่มาใช้บริการห้องสมุด

การพัฒนาคลังข้อมูลโดยใช้ซอฟต์แวร์ Microsoft SQL Server นั้นสามารถพัฒนาคลังข้อมูลได้ครบถ้วนทุกขั้นตอน ตั้งแต่ การคัดแยก (extracting) ข้อมูลจากแหล่งที่มาของข้อมูล (source systems) ตามความต้องการของผู้ใช้เพื่อเก็บไว้ในที่พักข้อมูล (staging area) เพื่อดำเนินการทำความสะอาดข้อมูล (cleansing) แล้วโหลด (loading) เข้าสู่คลังข้อมูล โดยใช้เครื่องมือ Database Engine และ SQL Server Integration Services และสามารถพัฒนาแอปพลิเคชันสำหรับผู้ใช้แต่ละกลุ่มเพื่อใช้ค้นคืนข้อมูล/ความรู้จากคลังข้อมูล ทั้งที่เป็นการประมวลผลเชิงวิเคราะห์แบบออนไลน์หรือโอแอลพี (OLAP) และการสร้างรายงานต่างๆ โดยใช้ SQL Server Analysis Services การพัฒนาคลังข้อมูลโดยใช้เทคนิคเหมือนข้อมูลนั้น มีประโยชน์ในการวิเคราะห์แพทเทิร์นหรือความรู้ที่ซ่อนอยู่ในคลังข้อมูลด้วยวิธีการทำเหมืองข้อมูลด้วย ดังนั้นหากหน่วยงานใดมีการจัดเก็บข้อมูลในรูปแบบของคลังข้อมูลจะทำให้สามารถนำข้อมูลมาแสดงผลในรูปแบบหลายมิติในลักษณะของ Pivot Table ได้ เพราะเมื่อสร้าง Cube แล้วจะสามารถดูข้อมูลที่เป็นของการประมวลผลเชิงวิเคราะห์แบบออนไลน์ได้

2.2 การทำ Online Analytical Processing (OLAP)

ผู้วิจัยใช้เครื่องมือ SQL Server Analysis Service (SSAS) เพื่อสร้าง Analysis Service Project สำหรับวิเคราะห์สร้าง OLAP Cube ที่เป็นฐานข้อมูลหลายมิติ และกำหนดข้อมูลจากตารางข้อมูลที่อยู่ในคลังข้อมูลที่ได้ออกแบบโครงสร้างคลังข้อมูล และผู้ใช้สามารถนำข้อมูลมาแสดงผลแบบหลายมิติในลักษณะของ Pivot Table ได้ และเรียกดูข้อมูลการยืมหนังสือของนักศึกษาตามคุณลักษณะของนักศึกษาหรือเรียกดูข้อมูลตามพฤติกรรมการยืมหนังสือของนักศึกษาได้แบบหลายมิติ ดังนั้น Online Analytical Processing (OLAP) จึงเป็นเทคโนโลยีที่ประกอบด้วยเครื่องมือที่ช่วยดึงและนำเสนอข้อมูลในหลายมิติ (Multidimensional) จากหลายๆ มุมมอง โดยที่ OLAP ได้รับการออกแบบมาสำหรับผู้ใช้ในระดับของผู้บริหารหรือหน่วยงานในองค์กร ที่ต้องวิเคราะห์ข้อมูลเพื่อใช้ประกอบการตัดสินใจในระดับสูง เพราะสนองตอบกับการแสดง และนำเสนอข้อมูลได้เป็นอย่างดีต่อการทำรายงาน ทำการวิเคราะห์ข้อมูลได้ในทุกมุมมองขององค์กรประกอบที่มีความเกี่ยวข้องทั้งหมดมาช่วยคาดคะเน หรือทำนายผลที่จะตามมาในภาคหน้าได้

2.3 การทำเหมืองข้อมูล

เทคนิคดาต้าไมนิ่ง เป็นเทคนิคที่สามารถนำมาวิเคราะห์ข้อมูลปริมาณมหาศาลเพื่อค้นหาความรู้ใหม่ ๆ หรือแพทเทิร์นของข้อมูลที่หน่วยงานไม่เคยทราบมาก่อน เหมือน Berry&Linoff ที่กล่าวว่า การทำเหมืองข้อมูลเป็นกระบวนการของการคัดแยกและจัดรูปแบบชุดข้อมูลจากแหล่งข้อมูลหลาย ๆ แหล่งให้เป็นรูปแบบเดียวกัน จากนั้นมีการนำชุดข้อมูลดังกล่าวเข้าสู่คลังข้อมูลที่ได้ออกแบบไว้ และสร้างแบบจำลองขึ้นมาวิเคราะห์ชุดข้อมูลนั้น ผลลัพธ์ที่ได้จากการวิเคราะห์คือความรู้ใหม่ที่ไม่เคยทราบมาก่อน และสามารถนำไปใช้ประโยชน์ได้ ซึ่งจากงานวิจัยนี้ได้แนะนำเทคนิคต่างๆ ของดาต้าไมนิ่งมาวิเคราะห์ข้อมูลการยืมหนังสือของนักศึกษา ทำให้ได้ทราบคุณลักษณะของนักศึกษาและพฤติกรรมการยืมหนังสือของนักศึกษา ทราบแพทเทิร์นการยืมหนังสือ ทราบแนวโน้ม

การยืมหนังสือ รวมถึงความสัมพันธ์ของหนังสือแต่ละเล่มด้วย สามารถนำไปใช้ในการแนะนำหนังสือ สนับสนุนการค้นคว้าสารสนเทศหรือแม้แต่การดำเนินงานของห้องสมุดได้ ซึ่งสอดคล้องกับที่ ไพโรจน์ เอี่ยมชัยมงคล ที่กล่าวว่า การนำกระบวนการ Bibliomining มาประยุกต์กับข้อมูลที่เก็บไว้ เพื่อค้นหา ความสัมพันธ์ของหนังสือที่ถูกยืมกับผู้ยืม จะช่วยให้สามารถแบ่งหรือจัดกลุ่มผู้ยืมได้และยังสามารถ บอกได้ว่าแต่ละกลุ่มเป็นนักศึกษาหรือนักวิจัย หรืออาจารย์จากภาควิชาใด มีรายละเอียดการยืม หนังสือแตกต่างกันอย่างไรความรู้เหล่านี้จะช่วยทำให้ห้องสมุดทราบความต้องการของผู้ยืมหนังสือแต่ ละกลุ่ม และใช้เป็นแนวทางในการจัดการ จัดซื้อและจัดหาหนังสือให้สอดคล้องกับความต้องการของผู้ ยืมได้และสามารถนำมาใช้วางนโยบายต่างๆ ในการบริหารห้องสมุดได้อย่างมีประสิทธิภาพ งานวิจัยนี้ ได้ใช้เทคนิคการวิเคราะห์ประกอบด้วย 3 อัลกอริธึม คือ 1) คลัสเตอร์ริง (Clustering) 2) แอสโซซิเอชัน รูลส์ (Association Rules) และ 3) ไทม์ซีรีส์ (Time Series)

2.3.1 การจัดกลุ่ม (Clustering Analysis) ซึ่งเป็นอัลกอริธึมหนึ่งของดาต้าไม นิ่ง เพื่อจัดกลุ่มนักศึกษาตามพฤติกรรมการยืมหนังสือและข้อมูลส่วนตัวของนักศึกษา ทำให้ทราบ ข้อมูลและความแตกต่างของนักศึกษาในแต่ละกลุ่มและเมื่อนำข้อมูลมาวิเคราะห์จะทำให้สามารถ ทราบถึงความต้องการของนักศึกษาในการยืมหนังสือแต่ละเล่ม หรือแต่ละหมวดได้ ซึ่ง Chen & chen ได้ทำเหมืองข้อมูล โดยใช้กระบวนการ KDD (Knowledge Discovery in Database) เพื่อหา วิธีที่มีประสิทธิภาพสำหรับการจัดกลุ่มข้อมูลผู้ใช้บริการห้องสมุด มาวิเคราะห์ข้อมูลผลการวิเคราะห์ พบว่ากลุ่มที่เกิดจากการแบ่งกลุ่มด้วยขั้นตอนวิธี Ant Colony Clustering นั้นสมาชิกในกลุ่มจะมี ลักษณะคล้ายกันมากกว่าเมื่อเทียบกับกลุ่มสมาชิกที่มีการแบ่งด้วยขั้นตอนวิธี K - Means Clustering และสอดคล้องกับ ดุลรัตน์ กรณท์กันแสง ได้ทำการศึกษาเกี่ยวกับเรื่อง การประยุกต์ใช้ทฤษฎีเหมือง ข้อมูลในบริการยืม - คืนของห้องสมุดกรณีศึกษา สำนักหอสมุด มหาวิทยาลัยธรรมศาสตร์ โดย วิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทั้งข้อมูลผู้ใช้บริการและข้อมูลการบริการยืม - คืนหนังสือ เพื่อให้ ทราบถึงการเข้าใช้บริการยืม - วัสดุสารสนเทศของคณะต่างๆ ใน มหาวิทยาลัยธรรมศาสตร์โดยการนำ ข้อมูลมาทำเหมืองข้อมูลตามขั้นตอนของ CRISP-DM ด้วยอัลกอริธึมกฎความสัมพันธ์ และกฎการ แบ่งกลุ่ม ผลของการศึกษาด้วยอัลกอริธึมการแบ่งกลุ่มสามารถแบ่งผู้ยืมออกเป็น 4 กลุ่ม ซึ่งสามารถ ใช้เป็นข้อมูลประกอบการพิจารณาในการปรับปรุงบริการสำหรับผู้ใช้แต่ละกลุ่ม เช่นเดียวกับงานวิจัยนี้ ที่ใช้อัลกอริธึมในการจัดกลุ่มสามารถแบ่งกลุ่มนักศึกษาที่มาใช้บริการห้องสมุดที่มีคุณลักษณะ คล้ายๆกันไว้ด้วยกัน ทำให้ทราบว่ากลุ่มผู้ใช้บริการห้องสมุดของมหาวิทยาลัยราชภัฏศรีสะเกษเป็น นักศึกษากลุ่มใด ซึ่งความรู้ที่ได้สามารถนำมาใช้ส่งเสริมการใช้บริการกลุ่มที่มาใช้บริการห้องสมุดและ จัดกิจกรรมเพื่อส่งเสริมการใช้บริการห้องสมุดในกลุ่มที่ยังไม่ค่อยมาใช้บริการมากนัก

2.3.2 แอสโซซิเอชัน รูลส์ (Association Rules) เป็นการจัดการข้อมูล เพื่อ วิเคราะห์ข้อมูลการยืมหนังสือ หรือค้นหารูปแบบความสัมพันธ์ของพฤติกรรมการยืม - คืนหนังสือของ สมาชิก เพื่อหาความน่าจะเป็นในการยืมหนังสือว่าเล่มใดบ้างที่มีถูกยืมไปด้วยด้วยกัน ซึ่งมีงานวิจัยที่ มีการศึกษาค้นคว้า กันคืองานวิจัยของ Cunningham & Frank ได้ทำการศึกษาเกี่ยวกับเรื่อง Market basket analysis of library circulation data ซึ่งเป็นการวิเคราะห์ความสัมพันธ์ โดยใช้ หลักการ Market basket analysis ที่ใช้ในการหาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อพร้อม กัน ซึ่งใช้ข้อมูลการยืม - คืนหนังสือโดยการหาความสัมพันธ์ของหมวดหมู่หนังสือ เพื่อนำไปใช้

ประโยชน์ในการจัดวางตำแหน่งของหนังสือบนชั้นหนังสือให้เหมาะสมและตรงตามความต้องการสารสนเทศของผู้ใช้บริการ เพื่อให้ผู้ใช้บริการเห็นหนังสือที่ต้องการได้ง่าย และงานวิจัยของ Chen & chen ได้ทำเหมืองข้อมูลโดยทำการศึกษาเกี่ยวกับ Using data mining technology to provide a recommendation service in the digital library โดยนำธุรกรรมการยืม –คืนหนังสือ และวัสดุสารสนเทศของสมาชิกห้องสมุดซึ่งเป็นการทำเหมืองข้อมูล ด้วยขั้นตอนวิธี Clustering แล้วนำข้อมูลในแต่ละกลุ่มมาหากฎความสัมพันธ์ (Association Rule) ของหนังสือและวัสดุสารสนเทศ โดยใช้ขั้นตอนวิธี Apriori เพื่อเป็นแนวทางสำหรับห้องสมุดในการวางแผนการซื้อหนังสือได้ตรงตามความต้องการด้วยงบประมาณที่จำกัด และตอบสนองความต้องการของผู้อ่านไปพร้อมกับการส่งเสริมบริการห้องสมุดดิจิทัลและงานวิจัยของจริยาพร จำปา ได้ทำการศึกษาเกี่ยวกับเรื่อง การพยากรณ์ความต้องการทรัพยากรห้องสมุด โดยนำวิธีกฎความสัมพันธ์ของห้องสมุดการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย เพื่อวิเคราะห์พฤติกรรมการใช้ทรัพยากรห้องสมุดตามกฎความสัมพันธ์ของการยืมทรัพยากรห้องสมุดที่มีผลกับหมวดหมู่ของทรัพยากรห้องสมุด กฟผ. ซึ่งสามารถจัดหาหนังสือในแต่ละหมวดให้ตรงตามความต้องการของผู้ใช้ สอดคล้องกับผลการวิจัยนี้ที่หาความสัมพันธ์ของหนังสือแต่ละเล่มที่มักถูกยืมไปด้วยกันเสมอมาใช้สำหรับแนะนำให้กับสมาชิก เพื่อให้ได้หนังสือที่ตรงตามความต้องการของสมาชิก

2.3.3 ไทม์ซีรีส์ (Time Series) หรืออนุกรมเวลาเป็นการพยากรณ์ค่าตัวเลขการยืมหนังสือในอนาคต โดยพยากรณ์การยืมหนังสือในแต่ละเดือนทำให้เห็นแนวโน้มการยืมหนังสือของนักศึกษาในอนาคต ซึ่งเปรียบเทียบกับแนวโน้มการยืมหนังสือของนักศึกษาในแต่ละเดือน ทำให้สามารถวางแผนการดำเนินงานหรือจัดกิจกรรมต่างๆ ส่งเสริมการยืมหนังสือของนักศึกษาได้ ซึ่งสอดคล้องกับจริยาพร จำปา ได้ทำการศึกษาเกี่ยวกับเรื่อง การพยากรณ์ความต้องการทรัพยากร ห้องสมุดโดยวิธีกฎความสัมพันธ์ของห้องสมุดการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย เพื่อวิเคราะห์พฤติกรรมการใช้ทรัพยากรห้องสมุดตามกฎความสัมพันธ์ของการยืมทรัพยากรห้องสมุดที่มีผลกับหมวดหมู่ของทรัพยากรห้องสมุด กฟผ. และพยากรณ์ความต้องการทรัพยากรห้องสมุดในอนาคตและนำไปสร้างโปรแกรมการพยากรณ์ความต้องการทรัพยากรห้องสมุดเพื่อใช้ในการบริหารจัดการ การจัดสรรงบประมาณต่างๆ ในการจัดหาหรือจัดซื้อทรัพยากรสารสนเทศให้ตรงตามความต้องการของผู้ใช้บริการ

ผลการวิเคราะห์ด้วยเทคนิคดาต้าไมนิ่งในงานวิจัยนี้ เป็นเพียงกระบวนการที่ทำให้ผู้ใช้เห็นแพทเทิร์นความสัมพันธ์ของข้อมูลเท่านั้น เช่น แพทเทิร์นการแบ่งกลุ่มผู้ใช้ แพทเทิร์นการพยากรณ์จำนวนนักศึกษาที่มีการยืมหนังสือในอนาคต แต่ไม่ใช่เครื่องมือที่สามารถใช้ในการค้นหาคำตอบจากข้อมูลที่องค์การจัดเก็บอยู่เพื่อแก้ปัญหาต่างๆ ได้อย่างอัตโนมัติ การทำเหมืองข้อมูลเป็นเพียงเครื่องมือที่ช่วยวิเคราะห์ข้อมูลเท่านั้น ผลที่ได้หากจะนำไปประยุกต์ใช้งานจริงจึงควรศึกษาปัจจัยต่างๆ ของธุรกิจหรือทำความเข้าใจธุรกิจก่อนนำข้อมูลไปใช้จริง

3. ข้อเสนอแนะ

3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

3.1.1 การศึกษาครั้งนี้มีการนำเสนอแนวทางการประยุกต์ใช้แบบจำลองที่ได้จากการทำเหมืองข้อมูล แต่ยังไม่ได้นำแบบจำลองดังกล่าวไปใช้จริง ดังนั้นหากนำไปใช้จริงแล้วควรประเมินแบบจำลองอีกครั้งเพื่อนำผลการประเมินมาพัฒนาแนวทางการทำเหมืองข้อมูล และการประยุกต์ใช้ให้แบบจำลองมีความแม่นยำยิ่งขึ้น

3.1.2 ผู้บริหารระดับสูงและผู้เกี่ยวข้องควรให้การสนับสนุน ในการจัดเก็บและนำข้อมูลต่างๆ ที่มีในมหาวิทยาลัยหรือในองค์กรต่างๆ ซึ่งมีปริมาณมากๆ และมีการจัดเก็บอยู่ในระบบฐานข้อมูลอยู่แล้วไปใช้ประโยชน์ให้มากที่สุด โดยการสร้างคลังข้อมูลขององค์กร เพื่อใช้สำหรับประกอบการวิเคราะห์ที่จะเป็นประโยชน์ต่อการบริหารงานอย่างมีประสิทธิภาพมากยิ่งขึ้น

3.1.3 การจัดทำรายงานในรูปแบบหลายมิติ ซึ่งเป็นรายงานที่ผู้บริหารสามารถนำไปประกอบการพิจารณาตัดสินใจได้มากขึ้น ผู้บริหารสามารถที่จะเจาะลึกรายละเอียดในแต่ละมิติที่สนใจได้ และสามารถที่จะดูภาพรวมได้ด้วย ดังนั้นควรนำไปใช้กับการบริหารจัดการข้อมูลอื่นๆ ภายในมหาวิทยาลัย เช่น รายงานข้อมูลทางการเงิน รายงานข้อมูลด้านงานทะเบียน เป็นต้น เพื่อให้ผู้บริการสามารถดูข้อมูลในรูปแบบต่างๆ ในแต่ละมิติที่สนใจนำไปประกอบการพิจารณาตัดสินใจในการบริหารจัดการองค์กรได้อย่างมีประสิทธิภาพ

3.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

3.2.1 การศึกษาครั้งนี้ศึกษาเฉพาะสมาชิกห้องสมุดที่นักศึกษาเท่านั้น เพื่อให้เกิดประสิทธิภาพสูงสุดกับห้องสมุดควรขยายคลังข้อมูลโดยพัฒนาโครงสร้างให้ครอบคลุมต่อการจัดเก็บข้อมูลสมาชิกทุกกลุ่ม ซึ่งจะทำให้ผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลหรือการจัดทำรายงานหลายมิติมีความสมบูรณ์ ครบถ้วนและเป็นประโยชน์ต่อการปรับปรุงบริการของห้องสมุด และนำผลลัพธ์ที่ได้จากแบบจำลองไปเป็นแนวทางในการพัฒนาบริการการค้นคืนทรัพยากรสารสนเทศของห้องสมุด มหาวิทยาลัยราชภัฏศรีสะเกษ ทั้งนี้เพื่อประโยชน์ต่อผู้ใช้บริการ เช่น พัฒนาเครื่องมือสำหรับสืบค้นข้อมูลหนังสือ (Web OPAC) ตามความต้องการ/พฤติกรรมของกลุ่มผู้ใช้แต่ละกลุ่ม

3.2.2 การศึกษาในงานวิจัยควรนำอัลกอริธึมอื่นๆ มาประยุกต์ เพื่อค้นหาความรู้อื่นๆ ที่เกี่ยวกับผู้ใช้บริการ เพื่อให้ได้องค์ความรู้ที่หลากหลายมากขึ้น เช่น การทำคลาสสิฟิเคชันในการทำนายรูปแบบการใช้บริการของนักศึกษา เป็นต้น



บรรณานุกรม

บรรณานุกรม

- กิตติพงษ์ กลมกล่อม (2552) *การออกแบบและพัฒนาค้างข้อมูล = Data warehouse*
กรุงเทพมหานคร เคทีพีคอมพิวเตอร์ คอนซัลท์
- จริยาพร จำปา (2553) “การพยากรณ์ความต้องการทรัพยากรห้องสมุดโดยวิธีกฎความสัมพันธ์ของ
ห้องสมุดการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย” *ปริญญาวิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต*
ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัย
เทคโนโลยีพระจอมเกล้าพระนครเหนือ
- ดุลยรัตน์ กรณธ์แสง (2553) “การประยุกต์ใช้ทฤษฎีเหมืองข้อมูลในบริการยืม-คืนของห้องสมุด
กรณีศึกษา สำนักหอสมุด มหาวิทยาลัยธรรมศาสตร์” *การค้นคว้าอิสระวิทยาศาสตร์*
มหาบัณฑิต สาขาบริหารเทคโนโลยี วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์
- นงลักษณ์ พลอยปลื้ม (2540) “การแปลงข้อมูลเข้าสู่ Data Warehouse” *BCM Magazine* 8,
103 (กันยายน): 126
- ไพโรจน์ เอี่ยมชัยมงคล (2552) “การประยุกต์กระบวนการ Bibliomining กรณีศึกษา หอสมุด
มหาวิทยาลัยราชภัฏลำปาง” *วารสารวิจัยสมาคมห้องสมุดแห่งประเทศไทย* 2, 1
(มกราคม-มิถุนายน): 1-12
- มหาวิทยาลัยราชภัฏศรีสะเกษ (2555) รายงานประจำปี 2555 (1 ตุลาคม 2554 – 31 ตุลาคม
2555) ศรีสะเกษ ฝ่ายวิทยบริการและสารสนเทศ มหาวิทยาลัยราชภัฏศรีสะเกษ
- เลิศ เลิศศิริโสภณ (2541) “ถึงเวลาของดาต้าแวร์เฮาส์แล้วหรือยัง” *BCM Magazine* 9, 115
(กันยายน): 95
- วิภา เจริญภัณฑารักษ์ (2555) “หลักการพื้นฐานของการทำเหมืองข้อมูล” ใน *ประมวลสาระชุด
วิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ=Data warehouse data mining and
business intelligence* หน่วยที่ 8 หน้า 1-33 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช
บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี
- สุวรรณิ อัครกุลชัย (2555) “หลักการพื้นฐานของคลังข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล
เหมืองข้อมูล และธุรกิจอัจฉริยะ=Data warehouse data mining and business
intelligence* หน่วยที่ 1 หน้า 1-30 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช
บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี
- สุวรรณิ อัครกุลชัย (2555) *การพัฒนาแบบจำลองแบบพลวัตในการพยากรณ์ใช้อุปาทานของน้ำเพื่อ
อุตสาหกรรมในประเทศไทย* กรุงเทพมหานคร คณะวิศวกรรมศาสตร์ มหาวิทยาลัย
หอการค้าไทย

- Berry, M. and Linoff, G. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Indianapolis: Wiley.
- Connolly, Thomas M. and Begg, Carolyn E. (2002). *Database Systems: A Practical Approach to Design, Implementation, and Management 3rd*. MA : Addison-Wesley.
- Chen, A.P. and Chen, C.C. (2006). A New efficient approach for data clustering in electronic library using ant colony clustering algorithm. *The Electronic Library*
- Chen, C.C. and Chen, A.P. (2007). Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*.
- Cunningham, S.J. and Frank, E. (1999). Market basket analysis of library circulation data. *Proceedings of the sixth International Conference on Neural Information Processing 1999, Volume II, Perth, Western Australia*.
- <http://crisp-dm.org/Process/index.htm> Retrieved January 20, 2014
- <http://dssresources.com/papers/features/brobst&rarey01062003.html> Retrieved January 20, 2014
- <http://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf> Retrieved January 20, 2014
- <http://gotoknow.org/posts/452189> Retrieved January 20, 2014
- <http://microsoft.com/global/sqlserver/2008/en/us/PublishingImages/scrn-DataMining.jpg> Retrieved January 20, 2014
- <http://th.wikipedia.org/wiki/การทำเหมืองข้อมูล> Retrieved January 20, 2014
- <http://siripornk.blogspot.com/2010/08/data-mining.html> Retrieved January 20, 2014
- Larose, D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Indianapolis: Wiley.
- Nicholson, S. (2003). "The Bibliomining process: Data warehousing and data mining for library decision-making" *Information Technology and Libraries*. 22, 4: 4-9.
- Nicholson, S. and Stanton, J. (2003). *Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries*. Hershey PA: Idea Group Publishing.
- Roiger and Geatz. (2003). *Data mining – A tutorial-based primer*. Pearson International Edition, USA: Addison-Wesley

ประวัติผู้วิจัย

ชื่อ	นายธีรพงศ์ สงฆ์
วัน เดือน ปีเกิด	6 กรกฎาคม 2523
สถานที่เกิด	อำเภอโพธิ์ศรีสุวรรณ จังหวัดศรีสะเกษ
ประวัติการศึกษา	บธ.บ. สถาบันราชภัฏศรีสะเกษ พ.ศ. 2548
สถานที่ทำงาน	มหาวิทยาลัยราชภัฏศรีสะเกษ ตำบลโพธิ์ อำเภอเมือง จังหวัดศรีสะเกษ
ตำแหน่ง	เจ้าหน้าที่คอมพิวเตอร์ กลุ่มงานวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏศรีสะเกษ

