การวิเคราะห์ข้อมูลการเข้าชั้นเรียนและการสำเร็จการศึกษาด้วยเทคนิคเหมืองข้อมูล: กรณีศึกษามหาวิทยาลัยฟาร์อีสเทอร์น
Class Attendance and Graduation Data Analytic with Data Mining Technique: Case Study of The Far Eastern University

ธีรวิชญ์ วงษา[1*]

[1]สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ มหาวิทยาลัยฟาร์อีสเทอร์น
เลขที่ 120 ถนนมหิดล อำเภอเมือง จังหวัดเชียงใหม่ 20100

บทคัดย่อ

บทความนี้นำเสนอการนำข้อมูลที่ถูกจัดเก็บไว้ในระบบจัดการสารสนเทศของมหาวิทยาลัยมาวิเคราะห์ โดยนำองค์ความรู้การวิเคราะห์ข้อมูลด้วยวิธีการทางเหมืองข้อมูลมาประยุกต์ใช้เพื่อหาความสัมพันธ์ของการเข้าชั้นเรียนกับการสำเร็จการศึกษาของนักศึกษามหาวิทยาลัยฟาร์อีสเทอร์น ผ่าน 2 กระบวนการหลัก ได้แก่ 1) การเตรียมข้อมูลก่อนการประมวลผลผ่านกระบวนการคัดกรองข้อมูล การรวมข้อมูล การดึงข้อมูล และการแปลงข้อมูลให้เหมาะสม 2) การหาความสัมพันธ์ระหว่างการเข้าชั้นเรียนกับการสำเร็จการศึกษาด้วยเทคนิคการหากฎความสัมพันธ์

คำสำคัญ

เหมืองข้อมูล ข้อมูลการสำเร็จการศึกษา ข้อมูลด้านการศึกษา

Abstract

This article represents how to transform data stored in the university information system into good use. We apply a data mining technique to class attendance and graduation of The Far Eastern University students' data. This is to discover association rules between them through 2 main processes: 1) Data pre-processing with data cleaning, data integration, data selection and data transformation, 2) Data association discovery between class attendance and graduation of students.

* ผู้เขียนหลัก
อีเมล์: teerawich@feu.edu

Keywords

Data Mining, Graduation Data, Education Data

บทนำ

การหากฎความสัมพันธ์เป็นหนึ่งในเทคนิคการวิเคราะห์ข้อมูลด้วยวิธีการทางเหมืองข้อมูล เพื่อค้นหาความสัมพันธ์ที่เกิดขึ้นจากข้อมูลที่ถูกจัดเก็บไว้ในระบบสารสนเทศ และนำผลลัพธ์ที่ได้มาใช้สนับสนุนการตัดสินใจ (Agrawal & Srikant, 1994 ; Agrawal, Imieliński & Swami, 1993 ; Usama Fayyad, 1996) ถูกนำมาใช้อย่างแพร่หลายในงานหลากหลายประเภท อาทิ ธุรกิจห้างสรรพสินค้านำข้อมูลการซื้อสินค้าของลูกค้ามาวิเคราะห์เพื่อการจัดวางสินค้าหรือการสร้างรายการส่งเสริมการขายกับสินค้า 2 ชนิดที่ลูกค้ามักซื้อร่วมกัน หรือการวิเคราะห์ข้อมูลการใช้งานสื่อสังคมออนไลน์ของผู้ใช้เพื่อการนำเสนอโฆษณาที่ตรงตามความสนใจของผู้ใช้ในขณะนั้น

สำหรับการนำเทคนิควิธีการทางเหมืองข้อมูลมาประยุกต์ใช้ในงานด้านการศึกษาในประเทศไทยยังพบว่ามีไม่มากนัก ทั้งที่ปัจจุบันสถานศึกษาส่วนใหญ่เปลี่ยนระบบการจัดเก็บข้อมูลมาใช้คอมพิวเตอร์มากขึ้นแต่ไม่มีการนำข้อมูลที่เก็บไว้เหล่านั้นมาใช้ให้เกิดประโยชน์

ดังนั้น ในบทความนี้จะนำเสนอตัวอย่างการวิเคราะห์ข้อมูลด้วยเทคนิคทางเหมืองข้อมูลเพื่อวิเคราะห์หาความสัมพันธ์ระหว่างการเข้าชั้นเรียนของนักศึกษาและการสำเร็จการศึกษาโดยนำข้อมูลย้อนหลัง 5 ปีของนักศึกษามหาวิทยาลัยฟาร์อีสเทอร์นมาใช้เป็นข้อมูลตัวอย่างประกอบ

เหมืองข้อมูล (Data Mining)

เหมืองข้อมูลเป็นหนึ่งในเทคนิคการวิเคราะห์ข้อมูลเพื่อค้นหาองค์ความรู้หรือประโยชน์ที่ซ่อนอยู่ในข้อมูลปริมาณมหาศาล เพื่อทำให้อยู่ในรูปแบบที่เข้าใจและสามารถนำไปใช้งานได้ (เอกสิทธิ์ พัชรวงศ์ศักดา, ม.ป.ป.)

บริษัทขนาดใหญ่เริ่มมีการนำเหมืองข้อมูลเข้ามาวิเคราะห์ข้อมูลเพื่อเจาะกลุ่มผู้บริโภค อาทิ ด้านการส่งเสริมการขาย สินเชื่อ และการประชาสัมพันธ์ ฯลฯ อีกทั้งเพื่อหาความสัมพันธ์ระหว่างปัจจัยภายใน (เช่น ราคาสินค้า การวางตำแหน่งผลิตภัณฑ์ ความสามารถของพนักงาน) กับปัจจัยภายนอก (เช่น ทิศทางเศรษฐกิจ สภาวะการแข่งขันในตลาด จำนวนลูกค้าเป้าหมาย) เพื่อให้บริษัทสามารถกำหนดปัจจัยที่มีผลต่อยอดขาย ความพึงพอใจของลูกค้า และคาดการณ์กำไรได้แม่นยำขึ้น

การทำเหมืองข้อมูลแบ่งออกได้เป็น 2 ประเภท ได้แก่ 1) เทคนิคการเรียนรู้แบบไม่มีผู้สอน เช่น Association Rule (ซึ่งเป็นวิธีการที่เลือกใช้ในบทความฉบับนี้) Clustering และ 2) เทคนิคการเรียนรู้แบบมีผู้สอน เช่น Classification, Regression (Usama Fayyad, 1996) ตามลำดับ

การหากฎความสัมพันธ์ (Association Rule)

การหากฎความสัมพันธ์เป็นหนึ่งในวิธีการทางเหมืองข้อมูล มีวัตถุประสงค์เพื่อใช้ค้นหาความสัมพันธ์ระหว่างชุดข้อมูลที่สนใจในข้อมูลขนาดใหญ่ (Agrawal, et al., 1993) มักถูกนำไปประยุกต์ใช้เพื่อค้นหาความสัมพันธ์ของการซื้อสินค้าหรือข้อมูลที่เป็นรายการเปลี่ยนแปลง (Transaction) เช่น "เมื่อลูกค้าซื้อขนมปังแล้วจะซื้อแยมด้วย" หรือ "เมื่อลูกค้าซื้อเบียร์แล้วจะซื้อผ้าอ้อมด้วย" (พัชรวงศ์ศักดา เอกสิทธิ์, ม.ป.ป.) เมื่อนำมาเขียนให้อยู่ในรูปแบบความสัมพันธ์จะได้ดังนี้

ลูกค้าซื้อขนมปัง => ลูกค้าซื้อแยม
ลูกค้าซื้อเบียร์ => ลูกค้าซื้อผ้าอ้อม

กฎความสัมพันธ์ประกอบด้วย 2 ส่วน ได้แก่ Left Hand Side: LHS คือ ส่วนที่อยู่ด้านซ้ายของเครื่องหมาย => เราจะเรียกว่า ข้ออ้าง (Premise) และ Right Hand Side: RHS คือ ส่วนที่อยู่ด้านขวาของเครื่องหมาย => เราจะเรียกว่า ข้อสรุป (Conclusion) ดังนั้น จากตัวอย่างกฎความสัมพันธ์ "ลูกค้าซื้อเบียร์ => ลูกค้าซื้อผ้าอ้อม" เราจะได้ว่า ข้ออ้าง คือ ลูกค้าซื้อเบียร์ และ ข้อสรุป คือ ลูกค้าซื้อผ้าอ้อม

เพื่อให้สามารถเลือกกฎความสัมพันธ์ที่สนใจจากกฎความสัมพันธ์ทั้งหมดที่เป็นไปได้ จะอาศัยตัวแปรที่เกี่ยวข้องดังนี้

ค่าสนับสนุน (Support) คือ ค่าที่แสดงถึงความถี่ของกลุ่มข้อมูลว่ามีการเกิดขึ้นมากน้อยแค่ไหนจากชุดข้อมูลทั้งหมด สามารถคำนวณได้จากสมาการที่ 1

$$support(X) = \frac{|t \in T : X \subseteq t|}{|T|} \quad (1)$$

ค่าความมั่นใจ (Confidence) แสดงร้อยละของความเชื่อมั่นของกฎความสัมพันธ์เมื่อรูปแบบทางด้านซ้ายของกฎความสัมพันธ์ (LHS) เกิดขึ้น แล้วรูปแบบทางด้านขวาของกฎความสัมพันธ์ (RHS) จะเกิดขึ้น สามารถคำนวณได้จากสมการที่ 2

$$confidence(LHS => RHS) = \frac{support(LHS,RHS)}{support(LHS)} \quad (2)$$

ค่าลิฟท์ (Lift) คือ ค่าที่แสดงให้ว่าการเกิดรูปแบบ LHS และ RSH มีความสัมพันธ์ต่อกันมากเพียงใด โดยถ้าค่าลิฟท์เข้าใกล้ 1 แสดงว่ารูปแบบของ LHS และ RHS ไม่ขึ้นต่อกัน (Independent) แต่ถ้าค่าลิฟท์มีค่าเกิน 1 มากๆ แสดงว่ารูปแบบทั้งสองมีความสัมพันธ์กัน และสามารถนำไปใช้ประโยชน์เพื่อการทำนายผลข้อมูลในอนาคตได้ สามารถคำนวณได้จากสมการที่ 3

$$lift(LHS => RHS) = \frac{support(LHS,RHS)}{support(LHS) \times support(RHS)} \quad (3)$$

เทคนิคการหากฎความสัมพันธ์ด้วยวิธี Apriori Algorithm (Agrawal & Srikant, 1994) เป็นเทคนิคที่ถูกพัฒนาขึ้นโดย Agrawal และ Srikant ตั้งแต่ปี ค.ศ. 1994 และยังมีผู้นำวิธีการนี้มาพัฒนามาอย่างต่อเนื่อง โดยวิธีการทำงานด้วยวิธีนี้มีหลักการทำงาน 2 ขั้นตอนใหญ่ ได้แก่ 1) การหารูปแบบที่เกิดขึ้นร่วมกันบ่อย 2) การสร้างกฎความสัมพันธ์

1. **การหารูปแบบที่เกิดขึ้นร่วมกันบ่อย**

เป็นการหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อยจากชุดข้อมูลขนาดใหญ่โดยจะถูกเลือกขึ้นมาเมื่อมีค่าสนับสนุนมากกว่าค่าสนับสนุนต่ำสุด (Minimum Support) ที่ผู้ใช้กำหนด โดยจะแสดงอยู่ในรูปของเซตข้อมูล เช่น {ขนมปัง} {แยม} {ขนมปัง, แยม}

2. **การสร้างกฎความสัมพันธ์**

นำรูปแบบที่หาได้จากข้อ 1 มาสร้างเป็นกฎความสัมพันธ์ เช่น ขนมปัง => แยม หมายความว่า เมื่อลูกค้าซื้อขนมปังแล้ว ลูกค้าจะซื้อแยมร่วมด้วย

การเตรียมข้อมูลก่อนการประมวลผล (Data Pre-Processing)

การจัดเตรียมข้อมูลก่อนนำไปประมวลผลเป็นหนึ่งในขั้นตอนที่มีความสำคัญต่อกระบวนการทำเหมืองข้อมูล (Pyle, 1999) ประกอบไปด้วย 1) Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป 2) Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน 3) Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้ และ 4) Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน

สาเหตุที่ต้องมีการทำ Data Pre-Processing เนื่องจากในกระบวนการเก็บรวบรวมข้อมูลจากฐานข้อมูลนั้น อาจมีการดึงข้อมูลที่ไม่เกี่ยวข้องติดมาด้วย เช่น ค่าที่เกิดจากขอบเขต (อายุ: -1) ค่าที่เป็นไปไม่ได้ (เพศ: ชาย ตั้งครรภ์: จริง) เป็นต้น ดังนั้นกระบวนการวิเคราะห์ข้อมูลที่ไม่ใส่ใจต่อกระบวนการ คัดกรองข้อมูลอาจส่งผลกระทบต่อผลลัพธ์ที่คลาดเคลื่อนและใช้การไม่ได้

ตัวอย่างการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อการวิเคราะห์ข้อมูลทางการศึกษา

ในบทความนี้ได้นำเทคนิคการสร้างกฎความสัมพันธ์ซึ่งเป็นหนึ่งในวิธีการทางเหมืองข้อมูลเพื่อนำมาวิเคราะห์และหาความสัมพันธ์ระหว่างการเข้าชั้นเรียนและการจบการศึกษา

สำหรับข้อมูลที่นำมาใช้เพื่อสร้างกฎความสัมพันธ์ได้แก่ ข้อมูลการเข้าชั้นเรียนและข้อมูลการสำเร็จการศึกษาของนักศึกษามหาวิทยาลัยฟาร์อีสเทอร์น โดยเป็นข้อมูลย้อนหลัง 5 ปี (พ.ศ. 2551 – พ.ศ. 2555) ดังแสดงในตารางที่ 1 โดยในชุดข้อมูลที่ได้รับประกอบด้วย Attribute ลำดับที่ รหัส ชื่อ-นามสกุล สถานะ อาจารย์ที่ปรึกษา รหัสวิชา ชื่อวิชา มา ขาด สาย ลา รวม อาจารย์ผู้สอน เทอม ตามลำดับ

ตารางที่ 1

ตัวอย่างข้อมูลการเข้าชั้นเรียนของนักศึกษาแยกตามรายวิชา

ที่	รหัส	ชื่อ-นามสกุล	...	มา	ขาด	สาย	ลา	รวม	...	เทอม
...	...	...	...	...	...	...	...	...	...	...
5431	5101031047	นายธีราธร อุ่นเมือง	...	14	4	0	1	19	...	253
5432	5101031047	นายธีราธร อุ่นเมือง	...	9	1	0	0	10	...	254
5433	5101031047	นายธีราธร อุ่นเมือง	...	23	1	0	2	26	...	253
5434	5101031047	นายธีราธร อุ่นเมือง	...	20	5	0	0	25	...	252
5435	5101031047	นายธีราธร อุ่นเมือง	...	19	6	0	0	25	...	252
5436	5101031047	นายธีราธร อุ่นเมือง	...	8	12	0	0	20	...	252
5437	5101031047	นายธีราธร อุ่นเมือง	...	11	2	0	0	13	...	352
5438	5101031701	นายธนกร วุทธานนท์	...	6	12	0	0	18	...	151
...	...	...	...	...	...	...	...	...	...	...

ข้อมูลที่ได้รับมานั้นถือเป็นข้อมูลดิบ ยังไม่สามารถนำมาสร้างกฎความสัมพันธ์ได้ โดยต้องนำข้อมูลดิบดังกล่าวมาทำ Data Pre-Processing เพื่อรวมข้อมูลที่ซ้ำกัน ลบข้อมูลที่ผิดพลาดและลบ Attribute ที่ไม่เกี่ยวข้องออก จากนั้นจะทำการรวมชุดข้อมูลของนักศึกษาที่ซ้ำกันให้เป็น 1 ชุดข้อมูล เช่น ข้อมูลของนักศึกษารหัส 5101031047 ประกอบด้วย 7 ชุดข้อมูล โดยในแต่ละชุดข้อมูล คือ รายละเอียดการเช็คชื่อเข้าชั้นเรียนในแต่ละรายวิชา ผลลัพธ์จากการรวมชุดข้อมูลของนักศึกษารหัส 5101031047 ที่ประกอบด้วย Attribute การสำเร็จการศึกษา มา ขาด สาย ลา เท่ากับ 213, 140, 109, และ 112 ตามลำดับ ดังแสดงในตารางที่ 2

ตารางที่ 2

ตัวอย่างข้อมูลการเข้าชั้นเรียนที่จัดรูปแบบแล้ว

รหัส	Graduate	Attend	Absent	Late	Leave
5101031047	TRUE	213	140	109	112
5101031701	FALSE	31	57	0	0
5101033001	TRUE	373	68	0	7
5101033002	FALSE	9	75	0	0
5101033003	TRUE	381	65	1	1
5101033014	FALSE	7	77	0	0
5101033015	TRUE	318	123	4	3
...	...	...	...	...	...

ข้อมูลที่รวมแล้วจะถูกนำมาคำนวณให้เป็นร้อยละและจัดลำดับความถี่ของการมาเรียน การขาดเรียน การมาช้า และการลา โดยแบ่งความถี่ออกเป็น 5 ระดับ ดังตารางที่ 3 และเปลี่ยนสถานะของนักศึกษาให้เหลือ 2 สถานะจากทั้งหมด 13 สถานะ ได้แก่ สำเร็จการศึกษาและไม่สำเร็จการศึกษา ผลลัพธ์ของข้อมูลที่ได้นี้จะนำไปใช้เพื่อหากฎความสัมพันธ์มีรูปแบบข้อมูลดังแสดงในตารางที่ 4

ตารางที่ 3
ระดับความถี่ต่อร้อยละ

ความถี่	**ร้อยละ**
very high	80-100
high	60-79
medium	40-59
low	20-39
very low	0-19

ตารางที่ 4
ตัวอย่างข้อมูลที่จะนำไปวิเคราะห์ด้วยวิธีการเหมืองข้อมูล

Graduated	Attend	Absent	Late	Leave
FALSE	very low	very high	very low	very low
FALSE	very low	very high	very low	very low
FALSE	medium	medium	very low	very low
FALSE	very low	very high	very low	very low
FALSE	high	low	very low	very low
TRUE	high	low	very low	very low
TRUE	high	low	very low	very low
...	...	...	...	...

การสร้างกฎความสัมพันธ์

ข้อมูลที่นำมาใช้ในการสร้างกฎความสัมพันธ์ประกอบด้วยข้อมูลการสำเร็จการศึกษาและอัตราการเข้าชั้นเรียนของนักศึกษาย้อนหลัง 5 ปี (พ.ศ. 2551- พ.ศ. 2555) จำนวน 4,769 ชุดข้อมูล ผลการสร้างกฎความสัมพันธ์แสดงในตารางที่ 5

ตารางที่ 5

กฎความสัมพันธ์ที่ได้จากการวิเคราะห์ข้อมูล

No.	LHS	Symbol	RHS	Confidence	Lift
1	Graduated=TRUE Attend=very high	==>	Absent=very low	0.99	2.60
2	Attend=very high	==>	Absent=very low	0.99	2.59
3	Absent=very high	==>	Attend=very low	0.98	5.57
4	Graduated=FALSE Absent=very high	==>	Attend=very low	0.98	5.57
5	Absent=very high	==>	Graduated=FALSE	0.98	1.83
6	Attend=very low Absent=very high	==>	Graduated=FALSE	0.98	1.83
7	Attend=very low	==>	Graduated=FALSE	0.98	1.83
8	Graduated=FALSE Attend=very low	==>	Absent=very high	0.97	5.58
9	Absent=very high	==>	Graduated=FALSE Attend=very low	0.96	5.58
10	Attend=very low	==>	Absent=very high	0.96	5.57
11	Attend=very low	==>	Graduated=FALSE Absent=very high	0.94	5.57
12	Absent=medium	==>	Attend=medium	0.90	6.05
13	Graduated=TRUE Absent=low	==>	Attend=high	0.90	3.46

จากตารางที่ 5 พบว่าเราได้กฎความสัมพันธ์มาทั้งหมด 13 ข้อ โดยถ้าดูเฉพาะกฎที่เกี่ยวข้องกับการสำเร็จการศึกษาจะมีกฎทั้งหมด 9 ข้อ ได้แก่ กฎข้อที่ 1, 4, 5, 6, 7, 8, 9, 11, 13 ตามลำดับ เช่น กฎข้อที่ 5 คือ Absent=very high ==> Graduated=FALSE หมายความว่า ถ้าขาดเรียนมากที่สุดแล้วจะไม่สำเร็จการศึกษา ที่ค่าความเชื่อมั่นร้อยละ 98 หรือกฎข้อที่ 7 คือ Attend=very low ==> Graduated=FALSE หมายความว่า ถ้าเข้าเรียนน้อยที่สุดแล้วจะไม่สำเร็จการศึกษาที่ค่าความเชื่อมั่นร้อยละ 98 เป็นต้น

กฎความสัมพันธ์ที่ได้นี้แสดงให้เห็นว่านักศึกษาควรเข้าชั้นเรียนให้มาก เนื่องจากในการเรียนการสอนอาจารย์ผู้สอนอาจมีการบรรยายหรืออธิบายเพิ่มเติมจากในเนื้อหา หรือยกตัวอย่างกรณีศึกษาที่น่าสนใจเพื่อใช้ประกอบการเรียนการสอน เพื่อเปิดมุมมองในรายวิชาให้แก่นักศึกษาในหลายๆ มิติ และยังเปิดโอกาสให้สามารถซักถามได้เมื่อนักศึกษาเกิดข้อสงสัยใดๆ ในรายวิชา หรือในอีกกรณีหนึ่งอาจตีความได้ว่า นักศึกษาอาจไม่เหมาะกับการศึกษาด้วยตนเอง โดยอาจมีสาเหตุมาจากปัจจัยหลายๆ อย่างเช่น พื้นฐานความรู้ของนักศึกษาที่มีไม่เท่ากัน ระยะเวลาที่ใช้ในการศึกษา สิ่งยั่วยุต่างๆ ที่อาจทำให้ละเลยหรือใช้เวลากับการศึกษาด้วยตนเองน้อยลง เป็นต้น

สรุป

ในบทความนี้ได้นำเสนอตัวอย่างการนำข้อมูลที่มหาวิทยาลัยเก็บไว้มาวิเคราะห์ด้วยเทคนิคทางเหมืองข้อมูลเพื่อหาความสัมพันธ์ระหว่างการเข้าเรียนและการสำเร็จการศึกษาของนักศึกษามหาวิทยาลัยฟาร์อีสเทอร์น โดยใช้ข้อมูลในอดีต ตั้งแต่พี พ.ศ. 2551 ถึง พ.ศ. 2555 แบ่งออกเป็น 2 กระบวนการหลัก ได้แก่ 1) การเตรียมข้อมูลก่อนการประมวลผล (Data Pre-Processing) ผ่านกระบวนการคัดกรองข้อมูล การรวมข้อมูล การดึงข้อมูล และการแปลงข้อมูลให้เหมาะสม 2) การหาความสัมพันธ์ระหว่างการเข้าชั้นเรียนกับการสำเร็จการศึกษา โดยการนำเทคนิคการหากฎความสัมพันธ์ (Association Rule) มาประยุกต์ใช้ทำให้ได้กฎความสัมพันธ์มา 13 ข้อ

ปัญหาที่พบจากการนำข้อมูลมาวิเคราะห์พบว่า ข้อมูลของนักศึกษาบางคนมีน้อยกว่าที่ควรจะเป็น เช่น ในหนึ่งหลักสูตร นักศึกษาที่จะสำเร็จการศึกษาได้จะต้องมีหน่วยกิตสะสมไม่ต่ำกว่า 131 หน่วยกิต หรือคิดเป็นวิชาได้ 44 วิชา แต่ข้อมูลที่ได้รับมานั้น นักศึกษาบางคนมีข้อมูลการเข้าชั้นเรียนไม่ครบ เช่น มีข้อมูลการเข้าชั้นเรียน 7 วิชาบ้าง หรือมี 10 วิชาบ้าง เป็นต้น และบางครั้งข้อมูลที่ได้มาก็เป็นข้อมูลที่ไม่น่าเป็นไปได้ด้วย เช่น ข้อมูลนักศึกษารหัส 5101033025 ในรายวิชาหนึ่ง มีข้อมูลการ มา ขาด สาย ลา เป็น 0, 0, 0, 0 ตามลำดับ ดังนั้นเพื่อให้การวิเคราะห์ข้อมูลมีความถูกต้องแม่นยำมากขึ้นในอนาคต หน่วยงานที่เกี่ยวข้องจึงจำเป็นที่จะต้องใส่ใจและให้ความสำคัญกับการบันทึกข้อมูลที่ถูกต้องลงในระบบด้วย

เอกสารอ้างอิง

Agrawal, R. & Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. **Proceedings of the 20th International Conference on Very Large Data Bases, VLDB. September.** (pp. 487-499). Chile: Santiago.

Agrawal, R. ; Imielińnski, T. & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. **Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93.** (pp. 207-216). Washington, D.C.: ACM.

Pyle, D. (1999). **Data Preparation for Data Mining.** Los Altos, California: Morgan Kaufmann.

Usama Fayyad, G. P.-S. (1996). From Data Mining to Knowledge Discovery in Databases (PDF). **American Association for Artificial Intelligence.** (pp. 37-54). Retrieved Febuary 7, 2017, from http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf.

เอกสิทธ์ พัชรวงศ์ศักดา. (ม.ป.ป.). **การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิงเบื้องต้น**. ม.ป.ท.: ม.ป.พ.

เอกสิทธิ์ พัชรวงศ์ศักดา. (2556). **คู่มือการใช้งาน Weka Explorer เบื้องต้น**. กรุงเทพฯ: เอเซีย ดิจิตอล การพิมพ์.