



การเตรียมข้อมูลการซื้อขายอาหารผ่านบัตรสมาร์ทการ์ด
เพื่อการทำเหมืองกระบวนการ

Preparation of Smart Card Data for Food Purchase Analysis of Students
through Process Mining

นรภณ จ้อยฟอง
Norrapon Joyfong

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
บัณฑิตวิทยาลัย มหาวิทยาลัยสยาม
พุทธศักราช 2562

บทคัดย่อ

หัวข้อวิทยานิพนธ์ : การเตรียมข้อมูลการซื้อขายอาหารผ่านบัตรสมาชิกการ์ด
เพื่อการทำเหมืองกระบวนการ

ชื่อนักศึกษา : ว่าที่ร้อยตรี นรภณ จ้อยฟอง

รหัสประจำตัว : 6117600002

ชื่อปริญญา : วิทยาศาสตรมหาบัณฑิต

สาขาวิชา : เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษา : ศาสตราจารย์ ดร.วิเชียร เปรมชัยสวัสดิ์

งานวิจัยนี้นำเสนอเทคนิคในการเตรียมข้อมูลเพื่อทำเหมืองกระบวนการ จากการนำข้อมูลที่
ถูกจัดเก็บหลายรูปแบบและหลายแหล่งข้อมูลให้อยู่ในรูปแบบของไฟล์ CSV เพื่อประมวลผลด้วยโปรแกรม
DISCO ขั้นตอนการนำข้อมูลมาประกอบเข้าด้วยกันต้องใช้เทคนิคในการเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่
ถูกต้อง เนื่องจากข้อมูลมาจากหลายแหล่งข้อมูล และถูกจัดเก็บหลากหลายชนิดไฟล์ ข้อมูลมีรูปแบบการ
จัดเก็บที่แตกต่างกัน ไม่สามารถระบุคีย์หลักที่จะใช้ร่วมกันสำหรับอ้างอิงไฟล์ข้อมูลที่เกี่ยวข้องกัน ต้องใช้
เทคนิคการเตรียมข้อมูลเพื่อหาความสัมพันธ์ของข้อมูลที่สามารถใช้ร่วมกันได้เพื่อส่งออกไฟล์ให้อยู่ใน
รูปแบบ CSV นำข้อมูลที่ส่งออกประมวลผลต่อด้วยโปรแกรม Disco ซึ่งมีขั้นตอนการวิจัย ดังนี้ 1. รวบรวม
ข้อมูล 2. ปรับรูปแบบข้อมูล 3. จัดเก็บและเชื่อมความสัมพันธ์ของข้อมูลบนฐานข้อมูล 4. ส่งออกข้อมูล
และทดสอบนำเข้าข้อมูลด้วยโปรแกรม Disco วิธีการที่นำเสนอนี้ได้ทดลองตรวจสอบฐานข้อมูลด้วยการสุ่ม
ตรวจสอบข้อมูลย้อนกลับและตรวจสอบข้อมูลด้วยโปรแกรม Disco ได้ผลลัพธ์ที่ถูกต้อง และสามารถ
ส่งออกไฟล์ข้อมูลให้อยู่ในรูปแบบไฟล์ CSV ประมวลผลด้วยโปรแกรม Disco ได้ผลลัพธ์เป็นอย่างดี

ABSTRACT

Title : Preparation of Smart Card Data for Food Purchase Analysis of Students
through Process Mining

By : Mr. Norrapon Joyfong

Degree : Master of Science

Faculty : Graduate School of Information Technology

Advisor : *W. Premchaiswadi*
(Prof. Dr. Wichian Premchaiswadi)

This research presented data preparation techniques for the Data Mining Process of information that was stored in various formats and data sources in the form of CSV files for processing in DISCO Program and the process of combination of data need techniques find the correct information. Since the data came from various many sources and was stored in various file types, the data had different storage formats that could not specify a primary key to be used as a reference for related data files. It required data preparation techniques to find correlations of data that could be used together to export the file in CSV format. The exported data was processed with the Disco Program with the following research process as follows; 1) Collect data; 2) Adjust the data format; 3) Store and correlate the data on the database; 4) Export the data and test the importation of data by Disco. This process checked the database by random check back and reviewed the information via Disco Program, which had accurate results, could export the data file to CSV file format, and process with Disco Program with satisfactory results.

Keywords: Smart Card Data, Food Purchase Analysis, Process Mining, Disco tool



กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีเนื่องจากผู้วิจัยได้รับความอนุเคราะห์ช่วยเหลือเป็นอย่างดีเสมอมา ตลอดจนในคำปรึกษาต่างๆในการจัดทำสารนิพนธ์ โดยเฉพาะอาจารย์ที่ปรึกษา คือ ศ.ดร.วิเชียร เปรมชัยสวัสดิ์ และ อาจารย์ ภูริเดช อาภาสัจย์ ผู้ให้ความอนุเคราะห์ด้านข้อมูล รวมถึงคณาจารย์อีกหลายท่าน ที่คอยให้คำแนะนำ ข้อเสนอแนะและติดตามความก้าวหน้าในการดำเนินสารนิพนธ์ทำให้งานสารนิพนธ์มีความถูกต้องและความสมบูรณ์ยิ่งขึ้น ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาและการเสียสละเวลาของอาจารย์เป็นอย่างยิ่ง และขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณบิดา มารดา ที่คอยให้กำลังใจให้แก่ผู้วิจัยตลอดมาและขอบคุณพี่ๆเพื่อนๆ น้องๆที่คอยให้คำปรึกษา คำแนะนำที่มีประโยชน์และคอยให้กำลังใจผู้วิจัยมีความซาบซึ้งในความกรุณาและความปรารถนาดีของทุกๆท่านเป็นอย่างยิ่งจึงขอกราบขอบพระคุณไว้ในโอกาสนี้ด้วย

ว่าที่ร้อยตรี นรภณ จ้อยฟอง

มีนาคม 2563

สารบัญ

	หน้า
บทคัดย่อ	ก
Abstract	ข
กิตติกรรมประกาศ	ค
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของโครงการวิจัย	3
1.3 ขอบเขตของโครงการวิจัย	3
1.4 นิยามศัพท์เฉพาะ	3
1.4.1 Process Mining (เหมืองกระบวนการ)	3
1.4.2 Event Log (บันทึกเหตุการณ์)	4
1.4.3 Disco Fluxicon	4
1.4.4 Fuzzy Miner	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ	4
1.6 การวางแผนโครงการ	5
บทที่ 2 ทฤษฎีแนวคิดและงานวิจัยที่เกี่ยวข้อง	6
2.1 เหมืองกระบวนการ Process Mining	6
2.1.1 การค้นพบ (process discovery)	7
2.1.2 การตรวจสอบความถูกต้องของข้อมูล (conformance checking)	7
2.1.3 การปรับปรุงกระบวนการให้ดีขึ้น (enhancement)	7
2.2 บันทึกเหตุการณ์ (Event Log)	8
2.3 Disco	9
2.4 Fuzzy Miner	10
2.5 งานวิจัยที่เกี่ยวข้อง	14

สารบัญ(ต่อ)

	หน้า
บทที่ 3 วิธีการดำเนินการวิจัย	17
3.1 เครื่องมือที่ใช้ในงานวิจัย	17

3.1.1 Disco Fluxicon	17
3.1.2 Rapid Miner	19
3.2 ขั้นตอนการวิจัย	20
3.2.1 การเตรียมข้อมูล.....	20
3.2.2 การนำเข้าข้อมูล	20
3.2.3 ซึ่เฉพาะประเภทข้อมูล	21
3.3 การวิเคราะห์ข้อมูล	22
บทที่ 4 ผลการวิจัย	27
4.1 ผลการวิเคราะห์ข้อมูลจากโปรแกรม Disco Fluxicon.....	27
4.2 ผลการวิเคราะห์ข้อมูลจากโปรแกรม Rapid Miner.....	37
4.3 เปรียบเทียบความแตกต่างของผลการวิเคราะห์ข้อมูลระหว่าง โปรแกรม Disco fluxicon และโปรแกรม RapidMiner	40
บทที่ 5 สรุปผลและข้อเสนอแนะ	41
5.1 วัตถุประสงค์ของการวิจัย	41
5.2 สรุปผลการวิจัย.....	41
5.3 ข้อเสนอแนะ.....	42
บรรณานุกรม.....	43

สารบัญตาราง

หน้า

ตารางที่ 1.1 แสดงแผนและระยะเวลาในการดำเนินงานภาคนิพนธ์.....	3
ตารางที่ 2.1 ตารางการชี้เฉพาะประเภทของข้อมูล.....	15
ตารางที่ 4.1 ขั้นตอนการดำเนินกิจกรรม.....	30
ตารางที่ 4.2 ผลการดำเนินงานระหว่างกลุ่มที่มีประสิทธิภาพดีกับกลุ่มที่มีประสิทธิภาพต่ำจาก โปรแกรม Disco fluxicon.....	36

ตารางที่ 4.3 ผลการดำเนินงานระหว่างกลุ่มที่มีประสิทธิภาพดีกับกลุ่มที่มีประสิทธิภาพต่ำจาก โปรแกรม Rapid Miner	39
ตารางที่ 4.4 ตารางเปรียบเทียบความถี่ระหว่าง โปรแกรม Disco fluxicon และ โปรแกรม Rapid Miner	40



สารบัญรูปภาพ

	หน้า
รูปที่ 2.1 องค์ประกอบหลักของการทำเหมืองกระบวนการ.....	5
รูปที่ 2.2 ตัวอย่างการทำเหมืองข้อมูล: จัดหมวดหมู่เป้าหมายจะต้องกำหนดค่า.....	7
รูปที่ 2.3 ขั้นตอนการป้อนข้อมูลการทำเหมืองข้อมูล: Case ID กิจกรรมและเวลาจะต้องมีการระบุ..	8
รูปที่ 2.4 ขั้นตอนการเลือก Case ID.....	9
รูปที่ 2.5 ขั้นตอนการเลือก Activity.....	10

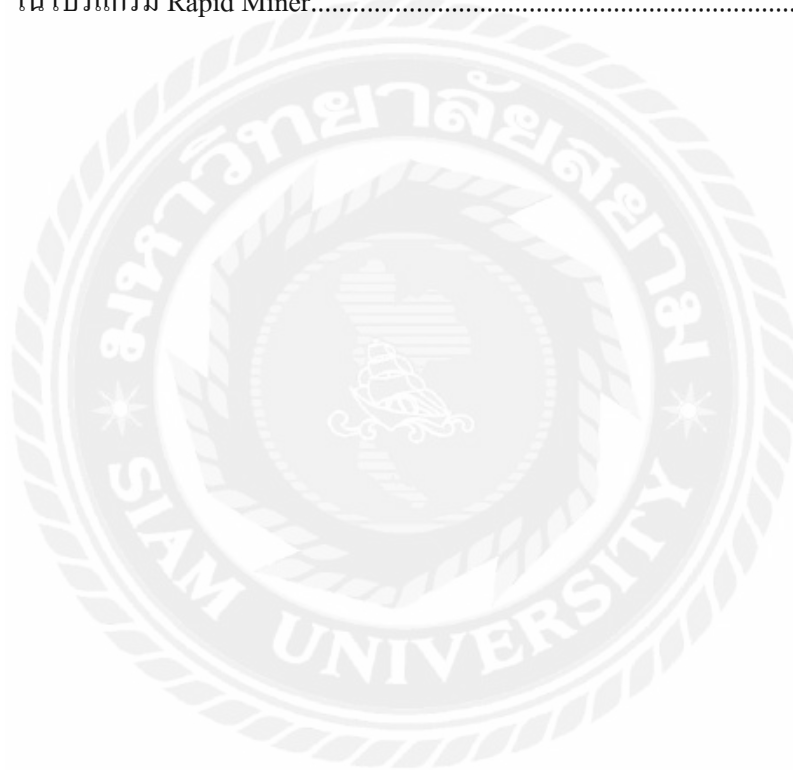
รูปที่ 2.6 ขั้นตอนการเลือก Timestamp.....	11
รูปที่ 2.7 ซ้าย: ได้ค้นพบรูปแบบกระบวนการสำหรับการสาธิต	12
รูปที่ 2.7 ขวา: เพิ่มเต็มรายละเอียดรูปแบบกระบวนการค้นพบขึ้น	12
รูปที่ 2.8 เมื่อได้สกัดข้อมูล จะสามารถนำเข้าไปใน Disco	13
รูปที่ 3.1 แสดงหน้าเว็บไซต์ดาวน์โหลดโปรแกรม Disco Fluxicon.....	18
รูปที่ 3.2 แสดงหน้าแรกของโปรแกรม Disco Fluxicon.....	18
รูปที่ 3.3 แสดงหน้าเว็บไซต์ดาวน์โหลดโปรแกรม Rapid Miner.....	19
รูปที่ 3.4 แสดงหน้าแรกของโปรแกรม Rapid Miner.....	19
รูปที่ 3.5 ขั้นตอนการนำข้อมูลออกจาก Microsoft SQL Server.....	20
รูปที่ 3.6 ภาพข้อมูลบันทึกเหตุการณ์ในโปรแกรม Disco Fluxicon.....	21
รูปที่ 3.7 ขั้นตอนการนำข้อมูลเข้าสู่โปรแกรม Disco Fluxicon.....	22
รูปที่ 3.8 แสดงรายละเอียดข้อมูลที่นำเข้า.....	22
รูปที่ 3.9 ภาพจำลองขั้นตอนกระบวนการขายรถยนต์ของโปรแกรม Disco Fluxicon รูปแบบความถี่ (Frequency).....	23
รูปที่ 3.10 ขั้นตอนการกำหนด Filter ที่จะทำการแสดงในแผนภาพ	24
รูปที่ 3.11 ขั้นตอนการ Export File จาก โปรแกรม Disco Fluxicon.....	24
รูปที่ 3.12 ขั้นตอนการเลือกโมดูลที่จะนำมาใช้ในโปรแกรม Rapid Miner.....	25
รูปที่ 3.13 ขั้นตอนการนำข้อมูลเข้าโปรแกรม Rapid Miner.....	26
รูปที่ 3.14 ภาพจำลองขั้นตอนกระบวนการขายรถยนต์ของโปรแกรม Rapid Miner.....	26
รูปที่ 4.1 แบบจำลองพฤติกรรมความถี่ของกระบวนการทำงาน	27
รูปที่ 4.2 ขั้นตอนการกำหนด Filter ที่จะทำการแสดงในแผนภาพ	28
รูปที่ 4.3 ภาพจำลองขั้นตอนกระบวนการขายรถยนต์ของโปรแกรม Disco Fluxicon รูปแบบความถี่ (Frequency).....	29
รูปที่ 4.4 การแสดงข้อมูลในแบบ Statistics Overview.....	30

สารบัญรูปภาพ (ต่อ)

หน้า

รูปที่ 4.5 การแสดงข้อมูลในแบบ Statistics Overview(ต่อ)	31
รูปที่ 4.6 ขั้นตอนกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายดี รูปแบบความถี่ (Frequency)	32
รูปที่ 4.7 ขั้นตอนกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายดี รูปแบบประสิทธิภาพ (Performance).....	37

รูปที่ 4.8 ขั้นตอนกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายต่ำ รูปแบบความถี่ (Performance).....	34
รูปที่ 4.9 ขั้นตอนกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายต่ำ รูปแบบประสิทธิภาพ (Performance).....	35
รูปที่ 4.10 แสดงแบบจำลองความถี่พฤติกรรมใน โปรแกรม Rapid Miner	37
รูปที่ 4.11 ภาพจำลองกระบวนการขายรถยนต์ใน โปรแกรม Rapid Miner.....	38
รูปที่ 4.12 แสดงแบบจำลองกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายดี ในโปรแกรม Rapid Miner.....	38
รูปที่ 4.13 แสดงแบบจำลองกระบวนการขายของกลุ่มที่ประสิทธิภาพการขายต่ำ ในโปรแกรม Rapid Miner.....	39



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ข้อมูล คือ ข้อเท็จจริง รายละเอียดหรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้น ซึ่งข้อเท็จจริงเหล่านี้อาจอยู่ในรูปแบบต่าง ๆ เช่น ตัวเลข ข้อความ ภาพ เสียง และสื่อบันทึกต่าง ๆ ดังนั้นการเก็บข้อมูลจึงเป็นการเก็บรวบรวมเกี่ยวกับข้อเท็จจริงของสิ่งที่เราสนใจนั่นเอง ข้อมูลจึงหมายถึงตัวแทนของข้อเท็จจริง หรือความเป็นไปของสิ่งที่เราสนใจมีการบันทึกไว้ในรูปแบบต่าง ๆ ตั้งแต่ข้อข้อความที่บันทึกไว้ในกระดาษ จนถึง การบันทึกข้อมูลในรูปแบบสื่ออิเล็กทรอนิกส์ด้วยโปรแกรมต่าง ๆ ตามการใช้งานของแต่ละฝ่าย โรงเรียนเป็นสถานที่แห่งหนึ่งที่มีการเก็บรวบรวมข้อมูลไว้ เช่น เมื่อนักเรียนสมัครเรียน โรงเรียนจะมีการบันทึกประวัตินักเรียนผ่านแบบฟอร์มรับสมัคร ข้อมูลของนักเรียนจะประกอบไปด้วยรายละเอียดเกี่ยวกับชื่อ ที่อยู่ บ้านเลขที่ ชื่อผู้ปกครอง บิดา มารดา จัดทำเป็นฐานข้อมูลเพื่อสืบค้นและนำข้อมูลไปใช้ในระบบอื่น ๆ เพื่อลดความซ้ำซ้อนกันของข้อมูล นอกจากนี้ยังมีการบันทึกข้อมูล การมาเรียนของนักเรียน บันทึกผลการเรียน ข้อมูลเหล่านี้ล้วนเป็นข้อเท็จจริงที่เกิดขึ้น สามารถนำข้อมูลมาใช้ประโยชน์ได้ในภายหลัง

ข้อมูลหลักที่นำมาวิเคราะห์ในงานวิจัยในครั้งนี้คือข้อมูลการซื้ออาหารผ่านบัตรเครดิตดิจิทัลของนักเรียนแต่ละคน โดยได้รับการสนับสนุนข้อมูลจากผู้ให้บริการบัตรเครดิตดิจิทัล ข้อจำกัดของข้อมูลที่ได้รับมาคือ ไม่สามารถระบุได้ว่ารายการซื้อสินค้าแต่ละรายการเป็นของนักเรียนคนใด เนื่องจากข้อมูลที่ได้รับมามีเพียงรหัสประจำบัตร รหัสร้านค้า เวลาที่ซื้อ ร้านค้าที่ซื้อ และยอดเงินเท่านั้น ไม่สามารถทราบได้เลยว่าเป็นรายการซื้ออาหารจากบัตรเครดิตดิจิทัลของนักเรียนคนใด ข้อมูลที่สำคัญที่สามารถเชื่อมโยงรหัสประจำตัวประชาชนของผู้ถือบัตรได้นั้น บริษัทผู้ให้บริการไม่สามารถส่งให้ได้เนื่องจากเป็นข้อมูลที่สำคัญและถือเป็นข้อมูลปกปิด ไม่สามารถส่งออกนอกองค์กรได้ ผู้ทำวิจัยจึงได้ทำการรวบรวมข้อมูลจากหลายแหล่งข้อมูล และหลากหลายรูปแบบไฟล์เพื่อให้สามารถเชื่อมโยงข้อมูลของผู้ถือบัตรกับข้อมูลนักเรียนเพื่อให้สามารถระบุตัวตนของผู้ถือบัตรเครดิตดิจิทัลได้อย่างถูกต้องและสอดคล้องกับรายการอาหารที่ซื้อนอกจากนี้ยังสามารถเชื่อมโยงผลตรวจสุขภาพประจำปีรายบุคคลได้เพื่อให้ง่ายต่อการวิเคราะห์หาสาเหตุหรือปัจจัยที่อาจส่งผลให้เกิดภาวะน้ำหนักเกินเกณฑ์ในเด็กวัยเรียน

การเตรียมข้อมูล (Data Preparation) เพื่อทำเหมืองกระบวนการ (Process Mining) จึงมีความสำคัญ เพื่อให้ได้สารสนเทศที่ตรงตามความต้องการ การเตรียมข้อมูลต้องสนับสนุนและจัดรูปแบบข้อมูลให้อยู่ในรูปแบบที่ง่ายต่อการประมวลผลด้วยโปรแกรมที่สนับสนุนเทคนิคในการทำเหมืองกระบวนการ

1.2 วัตถุประสงค์

1. เพื่อรวบรวมไฟล์ข้อมูลซื้ออาหารที่บันทึกผ่านการใช้บัตรเครดิตดิจิทัลของนักเรียน
2. เพื่อปรับรูปแบบข้อมูลที่รวบรวมจากหลายแหล่งข้อมูล หลายรูปแบบข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับฐานข้อมูล
3. เพื่อนำข้อมูลจัดเก็บในรูปแบบฐานข้อมูลและเชื่อมโยงความสัมพันธ์ของข้อมูล
4. เพื่อส่งออกข้อมูลจากโปรแกรมฐานข้อมูลและทดสอบนำเข้าข้อมูลด้วยโปรแกรม Disco

1.3 ขอบเขตการวิจัย

1. รวบรวมไฟล์ข้อมูลซื้ออาหารที่บันทึกผ่านการใช้บัตรเครดิตดิจิทัลของนักเรียน
2. สร้างไฟล์ข้อมูลใหม่ที่สามารถเชื่อมโยงข้อมูลการซื้ออาหารผ่านการใช้บัตรเครดิตดิจิทัล จาก การสำรวจข้อมูลที่ไม่เคยมีการบันทึกมาก่อน
3. ปรับรูปแบบไฟล์ข้อมูลที่หลากหลายชนิดให้อยู่ในรูปแบบฐานข้อมูลเพื่อให้ง่ายต่อการประมวลผล
4. เชื่อมความสัมพันธ์ของข้อมูลและส่งออกข้อมูลให้อยู่ในรูปแบบ CSV ไฟล์
5. ใช้เทคนิคการเตรียมข้อมูลเพื่อให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมเพื่อนำไปประมวลผลโดยใช้เทคนิคเหมืองกระบวนการ

1.4 คำนิยามศัพท์เฉพาะ

1.4.1 เหมือนกระบวนการ (Process Mining)

เหมือนกระบวนการ (Process Mining) เป็นเทคนิคการวิเคราะห์กระบวนการในการค้นหาคุณค่า (value) ที่จะสามารถหาโมเดลของกระบวนการใหม่ จากข้อมูลที่เกิดขึ้นจริงในบันทึกเหตุการณ์ (event log) ทำให้สามารถมองเห็นกระบวนการทางธุรกิจได้อย่างชัดเจนมากขึ้น โดยการนำข้อมูลมาวิเคราะห์นั้น เพื่อเป็นการเปรียบเทียบรูปแบบการทำงานแบบเดิมที่เป็นอยู่และนำมาปรับปรุงการทำงานเพื่อเพิ่มประสิทธิภาพมากขึ้น

1.4.2 บันทึกเหตุการณ์ (Event Log)

บันทึกเหตุการณ์ (Event Log) คือ ข้อมูลที่บันทึกลำดับเหตุการณ์การทำงานของกิจกรรมไว้ในระบบซึ่งระบบสารสนเทศอาจอยู่ในรูปแบบที่ต่างออกไปขึ้นอยู่กับระบบสารสนเทศที่ใช้งาน เมื่อนำข้อมูลเหล่านี้มาคัดกรองด้วยเทคนิค Process Mining จะทำให้เห็นวิธีการพัฒนาประสิทธิภาพในการทำงานมากยิ่งขึ้น

1.4.3 Disco Fluxicon

Disco Fluxicon เป็นเครื่องมือที่ใช้ในการทำเหมือนกระบวนการที่มีประสิทธิภาพสูงและง่ายต่อการใช้งาน เป็นเครื่องมือที่จะช่วยจัดการกับข้อมูลบันทึกเหตุการณ์ที่มีขนาดใหญ่และมีความซับซ้อน โดยโปรแกรม จะมีส่วนเสริมตัวกรองข้อมูล เพื่อใช้ในการจัดการบันทึกเหตุการณ์ที่เกิดขึ้น เช่น Time Frame, Variation, Performance ซึ่งภายในตัวโปรแกรมมีอัลกอริทึมหลักทั้งหมด 2 ตัว คือ Fuzzy Miner, Time Performance ใช้หาข้อเท็จจริงของบันทึกเหตุการณ์

1.4.4 Fuzzy Miner

Fuzzy Miner เป็นอัลกอริทึมหนึ่งในโปรแกรม Disco Fluxicon ที่ทำหน้าที่ในการแสดงการไหลของ กระบวนการแสดงความถี่และการทำซ้ำของบันทึกเหตุการณ์อย่างละเอียดสามารถนำไปวิเคราะห์เบื้องต้น

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 สามารถรวบรวมไฟล์สำหรับการประมวลด้วยเทคนิคเหมือนกระบวนการได้

1.5.2 สามารถเปลี่ยนรูปแบบการจดบันทึกข้อมูลเป็นไฟล์ข้อมูลจัดเก็บในระบบได้

1.5.3 สามารถเพิ่มข้อมูลรายการอาหารที่สำคัญในระบบเพื่อประโยชน์ในการวิเคราะห์ข้อมูลในอนาคต

1.5.4 สามารถรวบรวมข้อมูลด้วยเทคนิคการเตรียมข้อมูลและนำข้อมูลประมวลผลข้อมูลด้วยเทคนิคเหมือนกระบวนการได้เป็นอย่างดี

1.6 การวางแผนโครงการ

ระยะเวลาขั้นตอนการดำเนินงานวิจัยการเตรียมข้อมูลการซื้อขายอาหารผ่านบัตรสมาร์ตการ์ดเพื่อการทำเหมืองกระบวนการ เพื่อรวบรวมข้อมูลจากหลายแหล่งข้อมูล จากหลายรูปแบบไฟล์ ใช้เทคนิคการเตรียมข้อมูล ปรับข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการทำเหมืองกระบวนการ โดยมีขั้นตอนการดำเนินงานดังตารางที่ 1.1

ตารางที่ 1.1 ระยะเวลาขั้นตอนการดำเนินงาน

ปี 2562-2563	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.
ศึกษาและรวบรวมข้อมูล	←————→										
จัดเตรียมบันทึกเหตุการณ์	←————→										
ปรับรูปแบบไฟล์ให้อยู่ในรูปแบบฐานข้อมูล	←————→										
ตรวจสอบข้อมูลย้อนกลับจากบันทึกเหตุการณ์				←————→							
จัดทำเล่มสารนิพนธ์ฉบับสมบูรณ์					←————→						

บทที่ 2

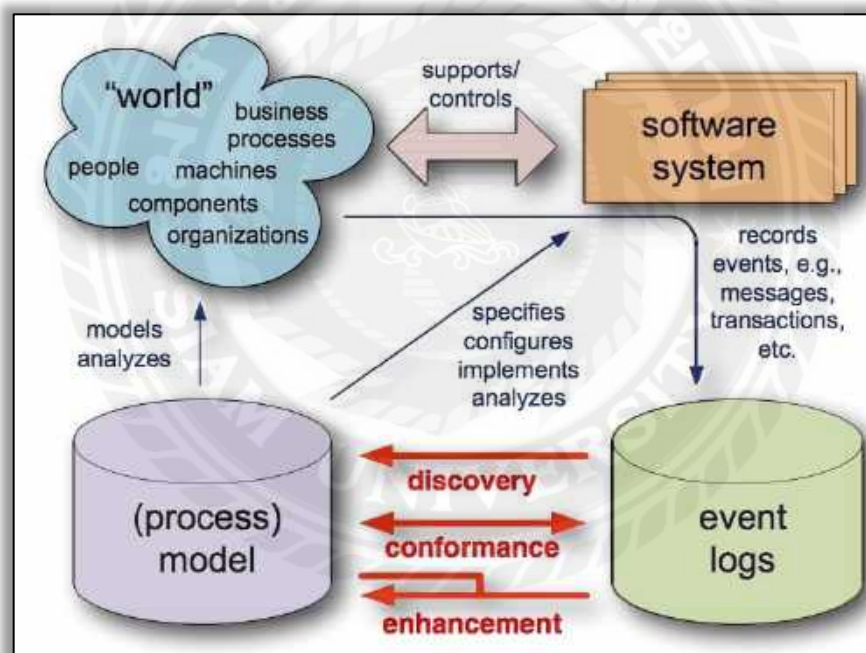
ทฤษฎีแนวคิดและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เน้นการนำเสนอเทคนิคการเตรียมข้อมูลเพื่อนำข้อมูลเข้าสู่ขั้นตอนการทำเหมืองกระบวนการด้วยโปรแกรม Disco ผู้วิจัยได้ศึกษาเอกสารและวรรณกรรมที่เกี่ยวข้องดังต่อไปนี้

2.1 เหมืองกระบวนการ (Process Mining) [1]

เหมืองกระบวนการ [1] เป็นเทคโนโลยีที่อุบัติขึ้นมาใหม่ เมื่อไม่นานมานี้ เป็นเทคนิคที่ใช้ในการค้นหาคุณค่า (value) จากข้อมูลที่เกิดขึ้นจริงในบันทึกเหตุการณ์ โดยมีวัตถุประสงค์หลัก 3 ประการ คือ การค้นพบกระบวนการ (process discovery), การตรวจสอบความสอดคล้อง (conformance checking), และการปรับปรุงให้ดีขึ้น (enhancement) ภาพรวมการทำงานของการทำงานเหมืองกระบวนการสามารถแสดงได้ในรูปที่

2.1



รูปที่ 2.1 ภาพรวมของการทำงานเหมืองกระบวนการ[1]

เป็นเทคนิคการวิเคราะห์กระบวนการในการค้นหาคุณค่า (value) ที่จะสามารถหาโมเดลของกระบวนการใหม่ จากข้อมูลที่เกิดขึ้นจริงในบันทึกเหตุการณ์ (event log) ทำให้สามารถมองเห็นกระบวนการการซื้ออาหารของนักเรียนได้อย่างชัดเจนมากขึ้น โดยการนำข้อมูลมาวิเคราะห์นั้น เพื่อเป็นการเปรียบเทียบรูปแบบการทำงานแบบเดิมที่เป็นอยู่และนำมาปรับปรุงการกำหนดนโยบายการซื้ออาหารภายในโรงเรียน เพื่อนำไปสู่การสร้างความสำเร็จหรือขอความร่วมมือแก่ผู้ปกครองในการบริโภคอาหารของนักเรียนที่มีน้ำหนักเกินเกณฑ์ให้เกิดพฤติกรรมบริโภคที่มีคุณภาพของนักเรียนอย่างยั่งยืน

โครงสร้างการจัดเตรียมข้อมูลสำหรับการจัดทำ Process mining ประกอบด้วยข้อมูลที่เป็นบันทึกในระบบสารสนเทศที่มีการใช้งานในชีวิตประจำวันของโรงเรียน เครื่องมือที่ใช้งาน โครงสร้างโรงเรียน ซอร์ฟแวร์ที่ใช้งานในโรงเรียนรูปแบบข้อมูล หรือระบบจัดการฐานข้อมูลที่ใช้งาน ซึ่งข้อมูลทั้งหมดนี้เป็นแหล่งเริ่มต้นของข้อมูลเพื่อจัดทำ Process mining ในกระบวนการจัดเตรียม โดยประเภทของ Process Mining แบ่งออกเป็น 3 ประเภทของการทำงานดังนี้

2.1.1 การค้นพบ (process discovery)

เทคนิคการค้นพบจะนำบันทึกเหตุการณ์ (Event Log) มาสร้างแบบจำลองการทำงาน ของกระบวนการได้ การค้นพบกระบวนการเป็นประเภทของเหมืองกระบวนการ เพื่อช่วยในการค้นพบ ปัญหาคอขวดหรือข้อผิดพลาดที่มีอยู่ในระบบ หลายองค์กรให้ความสนใจเมื่อได้เห็นเทคนิคการค้นพบกระบวนการนี้เพียงแค่อ่านบันทึกเหตุการณ์ในการทำงานจริง

2.1.2 การตรวจสอบความถูกต้องของข้อมูล (conformance checking)

ทำการตรวจสอบและเปรียบเทียบแบบจำลองของกระบวนการกับบันทึกเหตุการณ์ที่เกิดขึ้นจริง โดยการตรวจสอบความสอดคล้องสามารถใช้ตรวจสอบว่ากระบวนการทำงานที่บันทึกไว้ในบันทึกเหตุการณ์มีความสอดคล้องกับแบบจำลองหรือไม่ และสามารถตรวจสอบแบบจำลองได้เช่นเดียวกันเนื่องจากอาจมีแบบจำลองหลายประเภทที่ผิดพลาด การพิจารณาตรวจสอบความสอดคล้องสามารถนำมาประยุกต์ใช้กับ แบบจำลององค์กร แบบจำลองขั้นตอน หรือ นโยบายทางธุรกิจ และอื่น ๆ

2.1.3 การปรับปรุงกระบวนการให้ดีขึ้น (enhancement)

การปรับปรุงหรือขยายแบบจำลองกระบวนการที่มีอยู่ด้วยการใช้บันทึกเหตุการณ์ของกระบวนการจริง เป็นการขยายขีดความสามารถของแบบจำลองกระบวนการที่มีอยู่ที่เน้นไปในเรื่องของการไหลของการควบคุมการ (control flow) ไปในมุมมองอื่น ๆ เช่น มุมมองด้านองค์กร เวลาเครือข่ายสังคม เป็นต้น หรือการออกแบบ แบบจำลองใหม่เพื่อแก้ปัญหาต่าง ๆ ในแบบจำลองเดิม เช่น ปัญหาคอขวด เป็นต้น

2.2 บันทึกเหตุการณ์ (Even Log)[2]

บันทึกเหตุการณ์ คือชุดข้อมูลที่เกิดจากการใช้งานของผู้ใช้บนระบบซอฟต์แวร์ และถูกบันทึกลงบนฐานข้อมูล เพื่อนำมาแสดงรายงานหรือนำมาวิเคราะห์ผลด้วยเทคนิคต่าง ๆ โดยส่วนสำคัญประกอบด้วย กรณี (Case), กิจกรรม (Activity) ทรัพยากร (Resource) และประทับเวลา (Timestamp) โดยมักเก็บในรูปแบบไฟล์ สามารถส่งออกบันทึกเหตุการณ์ได้ในหลายรูปแบบ เช่น HTML, CSV ดังรูปที่ 2.3

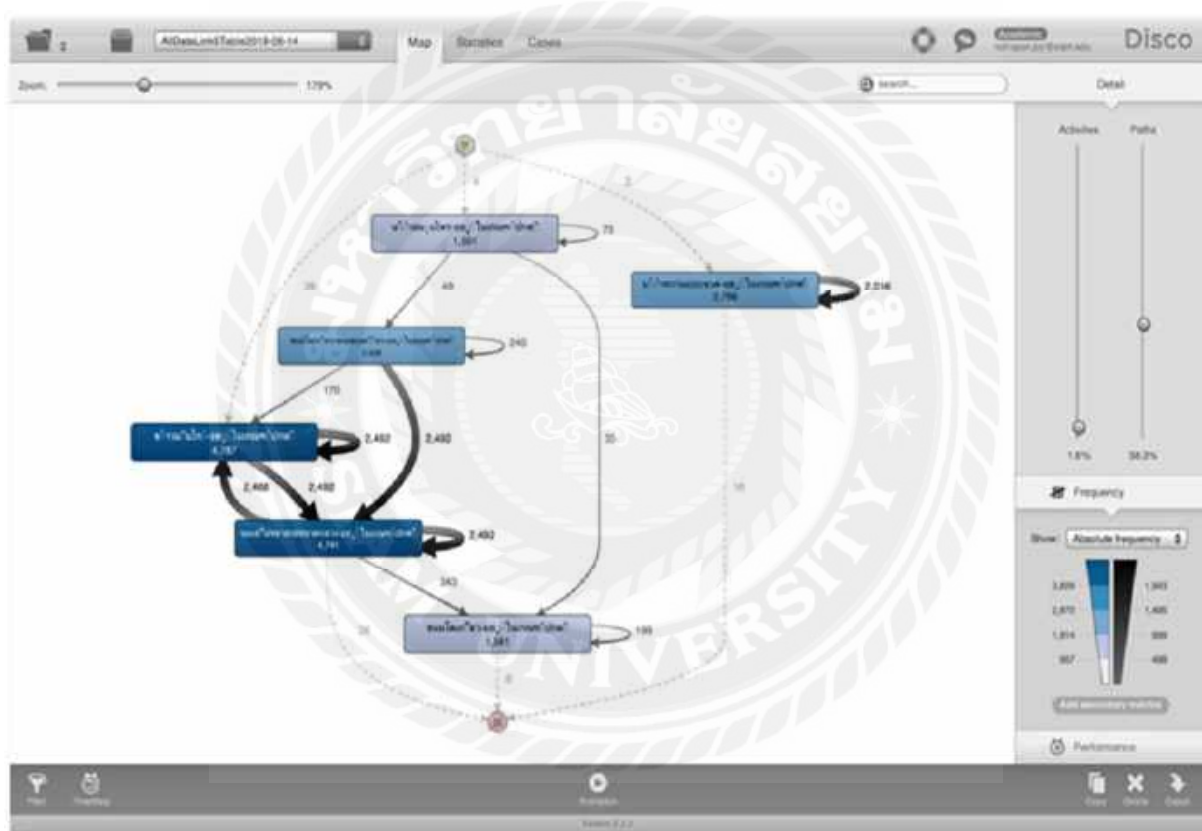


SVC NO#	MERCHANT	TRANS_DATE_TIME	CASH	BREAK
4629510117173070	BAKERY-SHOP	07/13/2018 6:45:48	20	MORNING
4629510117173770	BEVERAGE-SHOP	08/07/2018 6:48:32	10	MORNING
4629510089888850	BEVERAGE-SHOP	09/17/2018 6:52:43	70	MORNING
4629510089888850	BEVERAGE-SHOP	09/17/2018 6:53:10	10	MORNING
4629510089895530	BAKERY-SHOP	11/02/2018 6:57:08	40	MORNING
4629510089889910	NOODLE	09/13/2018 6:57:50	30	MORNING
4629510117172560	BEVERAGE-SHOP	06/07/2018 6:57:59	5	MORNING
4629510117174170	BEVERAGE-SHOP	08/08/2018 6:59:57	5	MORNING
4629510117172990	BEVERAGE-SHOP	07/10/2018 7:00:41	5	MORNING
4629510089896190	BEVERAGE-SHOP	07/19/2018 7:00:53	20	MORNING
4629510117172990	BEVERAGE-SHOP	11/12/2018 7:01:10	32	MORNING
4629510117172990	BAKERY-SHOP	11/05/2018 7:01:27	45	MORNING
4629510089889910	NOODLE	12/03/2018 7:01:50	30	MORNING
4629510117172990	BEVERAGE-SHOP	11/13/2018 7:02:40	38	MORNING
4629510089893960	BAKERY-SB	09/21/2018 7:03:25	20	MORNING
4629510089895530	BEVERAGE-SHOP	09/17/2018 7:03:42	45	MORNING
4629510089895530	BAKERY-SHOP	12/06/2018 7:04:08	20	MORNING
4629510117172990	BAKERY-SHOP	12/06/2018 7:04:19	10	MORNING
4629510089895530	BAKERY-SHOP	12/20/2018 7:04:25	30	MORNING
4629510117172990	BEVERAGE-SHOP	08/07/2018 7:05:52	30	MORNING
4629510089890110	BAKERY-SHOP	06/28/2018 7:06:08	10	MORNING
4629510089888850	BAKERY-SHOP	02/11/2019 7:06:08	10	MORNING
4629510117172990	BEVERAGE-SHOP	08/23/2018 7:06:10	15	MORNING
4629510117173650	BAKERY-SHOP	06/14/2018 7:06:57	30	MORNING
4629510117173070	BAKERY-SHOP	06/07/2018 7:07:06	30	MORNING

รูปที่ 2.3 ตัวอย่างบันทึกเหตุการณ์

2.3 DISCO[3]

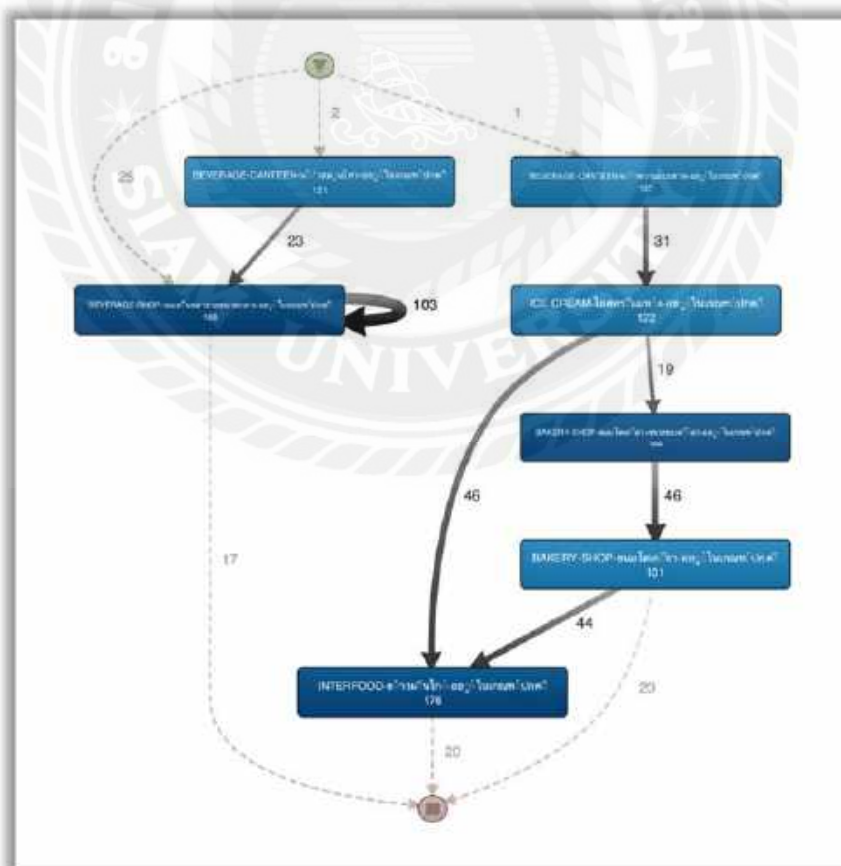
Disco เป็นโปรแกรมที่สนับสนุนเทคนิคในการทำเหมืองกระบวนกรที่มีกรออกแบบส่วนเชื่อมต่อกับผู้ใช้งานและสามารถเรียนรู้ได้โดยไม่ต้องมีความรู้และประสบการณ์ในการทำเหมืองกระบวนกรมากนัก แต่มีข้อจำกัดอยู่ที่เป็น โปรแกรมที่สนับสนุนการทำงานหลัก ๆ เฉพาะอัลกอริทึม Fuzzy miner เท่านั้น ในส่วนของข้อมูลเข้า สามารถรับได้ในรูปแบบของ MXML, SA-MXML หรือ XES แต่ที่สำคัญและเป็นจุดเด่นของ Disco อีกประการหนึ่งคือ Disco สามารถรับข้อมูลที่อยู่ในรูปแบบของ CSV ได้ ทำให้มีความสะดวกในการใช้งาน เนื่องจากสามารถใช้ข้อมูลจาก spreadsheet ได้ นอกจากนี้โปรแกรม Disco ยังสามารถ export ข้อมูลออกไปในรูปแบบต่าง ๆ เพื่อนำไปใช้ในการวิเคราะห์ด้วยเครื่องมือทำเหมืองกระบวนกรที่รองรับอัลกอริทึมอื่น ๆ ได้อีกด้วย แสดงตัวอย่างโปรแกรม Disco ดังรูปที่ 2.4



รูปที่ 2.4 โปรแกรม Disco

2.4 Fuzzy miner[4]

Fuzzy miner เป็นหนึ่งใน อัลกอริทึมในการค้นพบกระบวนการระยะแรก เป็นอัลกอริทึมโดยตรงที่ทำหน้าที่แก้ไขปัญหาค้นหาของตัวเลขขนาดใหญ่ กิจกรรมและเหตุการณ์ที่ไม่มีโครงสร้างระดับสูง (รูปที่ 2.3) เพื่อช่วยผู้ใช้งานสามารถวิเคราะห์กระบวนการจากบันทึกเหตุการณ์(Event log) ที่มีปริมาณข้อมูลจำนวนมากได้ง่าย ถือเป็นเครื่องมือที่ใช้ในการทำเหมืองกระบวนการที่มีประสิทธิภาพสูงและง่ายต่อการใช้งาน ช่วยจัดการกับข้อมูลบันทึกเหตุการณ์(Event log) ที่มีขนาดใหญ่และมีความซับซ้อน โดยโปรแกรม จะมีส่วนเสริมตัวกรองข้อมูล เพื่อใช้ในการจัดการบันทึกเหตุการณ์ที่เกิดขึ้น เช่น Time Frame, Variation, Performance ซึ่งภายในตัวโปรแกรมมีอัลกอริทึมหลักทั้งหมด 2 ตัว คือ Fuzzy Miner, Time Performance ใช้หาข้อเท็จจริงของบันทึกเหตุการณ์ และยังมีส่วนเสริมในการช่วยสรุปสถิติภายในตัวโปรแกรม ที่สำคัญอีกประการหนึ่ง โปรแกรมที่ใช้อัลกอริทึม Fuzzy miner ในการประมวลผลด้วยเทคนิคเหมืองกระบวนการ คือ Disco ซึ่งโปรแกรมสามารถรับข้อมูลที่อยู่ในรูปแบบของ csv ได้ ทำให้มีความสะดวกในการใช้งาน พร้อมทั้งสามารถ export ข้อมูลออกไปเพื่อใช้ในการวิเคราะห์ด้วยอัลกอริทึมอื่น ๆ ได้อีกด้วย เช่น โปรแกรมสำหรับประมวลผลด้วยเทคนิคเหมืองกระบวนการ ProM เป็นต้น ตัวอย่างการประมวลผลโดยใช้โปรแกรม Disco ซึ่งมีโครงสร้างอัลกอริทึม Fuzzy miner แสดงได้ดังรูปที่ 2.5



รูปที่ 2.5 แสดง Fuzzy miner จากโปรแกรม Disco

2.5 CSV[5]

CSV ย่อมาจาก Comma Separated Value เป็นไฟล์ข้อความประเภทหนึ่งที่ใช้สำหรับเก็บข้อมูลในรูปแบบตาราง ใช้เครื่องหมายจุลภาค หรือคอมม่า (,) ในการแบ่งแต่ละคอลัมน์ สามารถบันทึกไฟล์จาก Microsoft Excel เป็น CSV ไฟล์ได้โดยตรง หรือ อาจได้ไฟล์ CSV จากการ export ไฟล์จากระบบฐานข้อมูลอื่น ๆ การสร้างไฟล์ CSV จาก Microsoft Excel สามารถทำได้โดยเลือกรูปแบบการบันทึกไฟล์ชนิด CSV UTF-8(Comma delimited)(.csv) ไฟล์ที่ได้จากการบันทึกในรูปแบบ CSV นี้ สามารถนำไปประมวลผลต่อด้วยโปรแกรม Disco ได้ ตัวอย่างไฟล์ CSV ดังรูปที่ 2.6

	SVC NO	ID NO	PersonCode	First Name	Last Name	GenderName	Class	MERCHANT
2	4629510089894800	1300100012090	TT0074			หญิง	ครูไทย	BEVERAGE-CANTEEN
3	4629510089888080	1119600190063	802			ชาย	ประถมศึกษาปีที่ 6/2	BEVERAGE-CANTEEN
4	4629510089894800	1300100012090	TT0074			หญิง	ครูไทย	BAKERY-SB
5	4629510089894290	1103500105073	1477			หญิง	ประถมศึกษาปีที่ 1/3	BEVERAGE-CANTEEN
6	4629510089895400	1740101124751	1268			ชาย	ประถมศึกษาปีที่ 2/1	BEVERAGE-CANTEEN
7	4629510089886080	1119600190063	802			ชาย	ประถมศึกษาปีที่ 6/2	INTERFOOD
8	4629510089895730	1101700504808	1282			หญิง	ประถมศึกษาปีที่ 2/1	BEVERAGE-CANTEEN
9	4629510089892950	1749901014753	891			ชาย	มัธยมศึกษาปีที่ 2/1	BEVERAGE-CANTEEN
10	4629510117174800	1100201800672	632			หญิง	มัธยมศึกษาปีที่ 2/1	KHAOKAENG
11	4629510089896580	1104000118185	647			หญิง	มัธยมศึกษาปีที่ 2/1	NOODLE
12	4629510089896450	981015198277	ET0104			หญิง	English Teacher	KHAOKAENG
13	4629510117174790	1102170043109	1348			ชาย	มัธยมศึกษาปีที่ 2/2	KHAOKAENG
14	4629510089896450	981015198277	ET0104			หญิง	English Teacher	BEVERAGE-CANTEEN
15	4629510089896590	1579901474471	1388			หญิง	ประถมศึกษาปีที่ 2/2	KHAOKAENG
16	4629510089896590	1579901474471	1388			หญิง	ประถมศึกษาปีที่ 2/2	BEVERAGE-CANTEEN
17	4629510089896540	1100401511429	1150			ชาย	ประถมศึกษาปีที่ 3/1	KHAOKAENG
18	4629510089896600	1102200248862	827			ชาย	ประถมศึกษาปีที่ 6/1	KHAOKAENG
19	4629510089895510	1102170144070	1274			หญิง	ประถมศึกษาปีที่ 2/1	NOODLE
20	4629510089896940	1102004290614	1301			หญิง	ประถมศึกษาปีที่ 2/2	KHAOKAENG
21	4629510089897710	1102004290614	TT0074			หญิง	ครูไทย	BEVERAGE-SHOP

รูปที่ 2.6 ไฟล์ CSV

2.6 Data cleansing

data cleaning หมายถึง การทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูลตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล เพราะหมายถึงความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่น ๆ เป็นต้น จึงต้องมีการแทนที่การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องเหล่านี้ออกไป เพื่อให้ข้อมูลมีคุณภาพ การทำความสะอาดข้อมูลเกิดขึ้นเพื่อแก้ปัญหาจากความไม่สอดคล้องกันของข้อมูล ซึ่งอาจเกิดจากข้อผิดพลาดจากการบันทึกข้อมูล การส่งข้อมูล หรือการให้ความหมายของข้อมูลที่ถูกจัดเก็บแตกต่างกันออกไป หากมีการใช้ข้อมูลจากฐานข้อมูลมากกว่าหนึ่งแหล่งหรือข้อมูลที่มาจากแหล่งข้อมูลภายนอก จึงมีโอกาสสูงที่จะเกิดข้อมูลที่ไม่สะอาดเกิดขึ้นกับข้อมูล ความถูกต้องของข้อมูลจึงเป็นสิ่งสำคัญหากจะนำข้อมูลเพื่อไปประมวลผลต่อ เพื่อหลีกเลี่ยงการสรุปข้อมูลที่ผิดพลาด ตัวอย่างเช่น ข้อมูลที่เกิดการซ้ำซ้อนหรือขาดหายไปจะส่งผลให้มีสถิติที่ไม่ถูกต้องหรือทำให้เข้าใจผิดได้ เนื่องจากข้อมูลในคลังข้อมูลมีจำนวนมาก จำนวนข้อมูลที่เกิดความไม่สอดคล้องกันจึงมากตามไปด้วย การทำความสะอาดจึงเป็นปัญหาใหญ่ที่สุดของการเตรียมข้อมูล ซึ่งหมายถึงการรวบรวมข้อมูลจากแหล่งต่าง ๆ การแปลงข้อมูลและการนำข้อมูลเข้า จึงมีโอกาสสูงที่จะพบความผิดพลาดของข้อมูลมากขึ้น การทำความสะอาดข้อมูลจึงใช้เวลานานมากให้ขั้นตอนการเตรียมข้อมูลเพื่อให้ได้สารสนเทศที่เหมาะสมต่อการนำไปประมวลผลต่อไป

บทที่ 3

วิธีการดำเนินการวิจัย

วิธีการดำเนินงานเพื่อจัดเตรียมข้อมูลจากหลายแหล่งข้อมูลให้อยู่ในรูปแบบไฟล์ CSV เพื่อให้สามารถนำไฟล์ไปวิเคราะห์ได้ด้วยโปรแกรม Disco ได้นั้น มีขั้นตอนการดำเนินงานทั้งหมด 4 ขั้นตอน ดังนี้

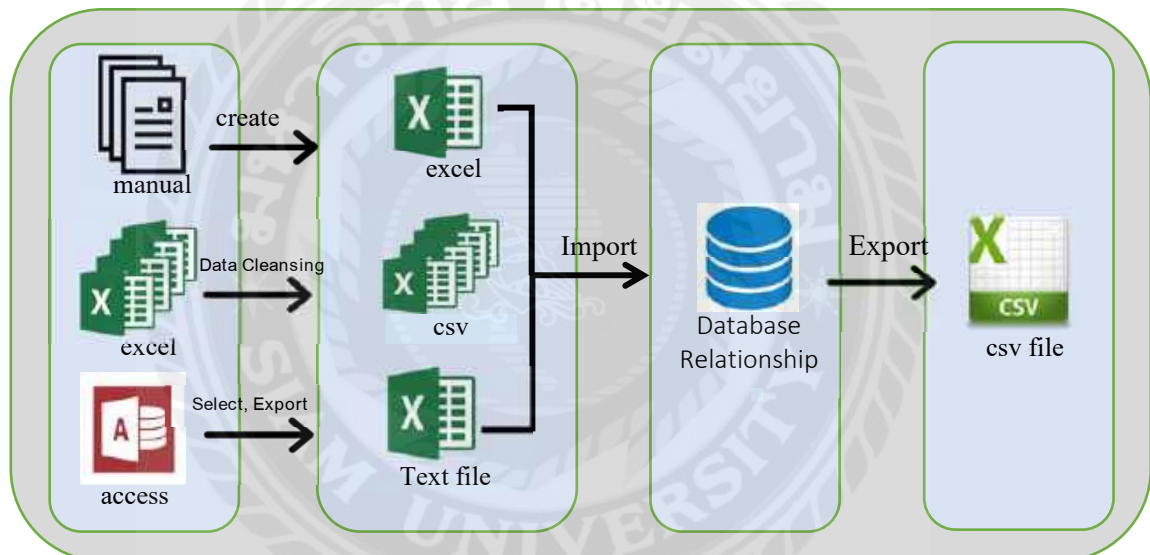
3.1 รวบรวมไฟล์ข้อมูล

3.2 ปรับรูปแบบข้อมูล

3.3 จัดเก็บและเชื่อมความสัมพันธ์ของข้อมูลบนฐานข้อมูล

3.4 ส่งออกข้อมูลเป็นไฟล์ CSV และทดสอบนำเข้าข้อมูลด้วยโปรแกรม Disco

แสดงได้ดังรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนการดำเนินการวิจัย

3.1 รวบรวมไฟล์ข้อมูล

การรวบรวมไฟล์ข้อมูล คือ ขั้นตอนสำคัญในการเตรียมข้อมูลเพื่อทำเหมืองกระบวนการ เนื่องจากข้อมูลถูกจัดเก็บจากหลายหน่วยงานและหลายรูปแบบไฟล์ ข้อมูลเกิดความซ้ำซ้อนและไม่เป็นปัจจุบัน ในบางหน่วยงานไม่เคยสร้างเป็นไฟล์ข้อมูลมาก่อน ผู้วิจัยได้รวบรวมไฟล์ข้อมูลที่จำเป็นเพื่อจัดเตรียมข้อมูลสำหรับทำเหมืองกระบวนการมีการรวบรวมข้อมูลและสร้างข้อมูล ดังนี้

3.1.1 รวบรวมไฟล์ข้อมูลการซื้ออาหารผ่านบัตรเงินสดดิจิทัล

ไฟล์ข้อมูลการซื้ออาหารผ่านบัตรเงินสดดิจิทัล คือไฟล์บันทึกเหตุการณ์การซื้ออาหารโดยใช้บัตรเงินสดดิจิทัล รวบรวมบันทึกเหตุการณ์ที่เกิดขึ้นเป็นระยะเวลา 1 ปีการศึกษาระหว่างวันที่ 14 May 2018 – 22 March 2019 ข้อมูลดังกล่าวได้รับจากผู้ดูแลระบบบัตรเงินสดดิจิทัล โดยข้อมูลในคอลัมน์ SVCNO เป็นข้อมูลรหัสที่แสดงอยู่บนหน้าบัตรถือเป็นข้อมูลหลัก (Primary Key) ในการแยกรายการซื้ออาหารในแต่ละรายการ ไฟล์ดังกล่าว เป็นชนิด CSV แสดงข้อมูลดังรูปที่ 3.2

Date	SVC Number	จำนวนครั้ง	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)	จำนวนครั้ง	จำนวนเงิน (บาท)	จำนวนครั้ง
20170104	4629510074707127						
20170104	4629510074707135						
20170104	4629510074707143						
20170104	4629510074707150						
20170104	4629510074707200						
20170104	4629510074707291						
20170104	4629510074707309			1		20	
20170104	4629510074707317						
20170104	4629510074707325						
20170104	4629510074707507						
20170104	4629510074707531						
20170104	4629510074707580					1	55
20170104	4629510074707689						
20170104	4629510074707713						
20170104	4629510074707739						
20170104	4629510074707762						
20170104	4629510074707879					1	40
20170104	4629510074707945					2	50
20170104	4629510074707960						
20170104	4629510074708034						

รูปที่ 3.2 ข้อมูลการซื้ออาหารผ่านบัตรเงินสดดิจิทัลชนิด CSV

3.1.2 รวบรวมไฟล์ข้อมูลผู้ถือบัตรเงินสดดิจิทัล

ไฟล์ข้อมูลผู้ถือบัตรเงินสดดิจิทัล คือ ข้อมูลประจำบัตรที่ถูกบันทึกไว้ในฐานข้อมูล Microsoft Access ใช้สำหรับยืนยันตัวตนของผู้ถือบัตรคือนักเรียนหรือนุคลากรคนใด ข้อมูลบัตรถูกเก็บไว้ในชิปของบัตรแต่ละใบ เมื่อทาบบัตรเข้ากับเครื่องบันทึกเวลา สามารถอ่านค่าข้อมูลของบัตรเชื่อมโยงและบันทึกข้อมูลเวลาเข้าโรงเรียนสู่โปรแกรมฐานข้อมูลเก็บข้อมูลเข้าและออกจากโรงเรียนได้ แสดงดังรูปที่ 3.3



รูปที่ 3.3 บันทึกเวลาเข้า-ออกโรงเรียนด้วยบัตรเงินสดดิจิทัล

ผู้วิจัยได้สืบค้นและส่งออกข้อมูลบัตรเงินสดดิจิทัลที่อยู่ในรูปแบบตารางฐานข้อมูลออกมาเป็นรูปแบบไฟล์ Excel ซึ่งแสดงดังรูปที่ 3.4



USERID	Badgenumbe	SSN	Name	Gender	TITLE	P
841	1001243	01416		M	อ.บ.	097-
842	1001244	01417		M	อ.บ.	083-
843	1001245	01424		F	อ.ญ.	081-
844	1001246	01423		F	อ.ญ.	081-
845	1001247	01431		F	อ.ญ.	088-
846	1001248	01429		F	อ.ญ.	062-
847	1001249	01438		F	อ.ญ.	
848	1001250	01420		M	อ.บ.	083-
849	1001251	01421		M	อ.บ.	
850	1001252	01428		M	อ.ญ.	097-
851	1001253	01436		F	อ.ญ.	098-
852	1001254	01427		F	อ.ญ.	
853	1001255	01415		M	อ.บ.	
854	1001256	01418		M	อ.บ.	
855	1001257	01434		M	อ.ญ.	
856	1001258	01426		F	อ.ญ.	
857	1001259	01425		F	อ.ญ.	
858	1001260	01422		M	อ.บ.	
859	1001261	01437		F	อ.ญ.	
860	1001262	01433		F	อ.ญ.	

รูปที่ 3.4 ข้อมูลผู้ถือบัตรเงินสดดิจิทัลที่อยู่ในรูปแบบฐานข้อมูล

3.1.3 รวบรวมไฟล์ข้อมูลประวัตินักเรียน

ไฟล์ข้อมูลประวัตินักเรียน คือ ข้อมูลประวัตินักเรียนของโรงเรียน ข้อมูลถูกเก็บบันทึกอยู่ในรูปแบบฐานข้อมูล Microsoft Access ไม่สามารถเข้าถึงโปรแกรมฐานข้อมูลโดยตรงได้ ผู้วิจัยได้ขอความอนุเคราะห์ในการขอสืบค้นข้อมูลและส่งออกข้อมูลให้อยู่ในรูปแบบ Text File เพื่อนำข้อมูลทำการเชื่อมโยงเข้ากับข้อมูลในส่วนอื่น ๆ ประกอบด้วย รหัสนักเรียน (Person Code) เป็น Primary key แสดงดังรูปที่ 3.5

PersonCode	FirstName	LastName	GenderName	PID	BirthDate
01266	เด็กชาย	[REDACTED]	,ชาย,	1 1008 01687 58 3,	11/8/2553
01273	เด็กหญิง	[REDACTED]	,หญิง,	1 1020 04209 05 1,	10/22/2553
01277	เด็กหญิง	[REDACTED]	,หญิง,	1 1021 70141 10 1,	10/4/2553
01280	เด็กหญิง	[REDACTED]	,หญิง,	1 1022 00305 86 6,	8/13/2553
01281	เด็กหญิง	[REDACTED]	,หญิง,	1 1035 00101 81 7,	3/9/2554
01295	เด็กหญิง	[REDACTED]	,หญิง,	1 1024 00260 53 0,	8/6/2554
01296	เด็กหญิง	[REDACTED]	หญิง,	2 1040 00008 91 4,	4/4/2554
01297	เด็กหญิง	[REDACTED]	หญิง,	1 1021 70155 86 1,	7/30/2554
01300	เด็กหญิง	[REDACTED]	,หญิง,	1 1004 01537 83 5,	1/3/2554
01307	เด็กหญิง	[REDACTED]	,หญิง,	1 1020 04227 65 3,	1/2/2554
01324	เด็กหญิง	[REDACTED]	หญิง,	1 1008 01682 24 7,	9/6/2553
01337	เด็กชาย	[REDACTED]	จ,ชาย,	1 1020 03968 83 8,	5/29/2551
01338	เด็กชาย	[REDACTED]	,ชาย,	1 1004 01413 40 1,	
01339	เด็กชาย	[REDACTED]	ชาย,	1 1015 01383 64 4,	4/24/2551
01491	เด็กหญิง	[REDACTED]	,หญิง,	1 1022 00246 55 0,	

รูปที่ 3.5 ข้อมูลประวัตินักเรียนอยู่ในรูปแบบ Text File

3.1.4 ไฟล์ข้อมูลผลตรวจสุขภาพประจำปี

ไฟล์ข้อมูลผลตรวจสุขภาพประจำปี คือ รายงานผลตรวจสุขภาพจากโรงพยาบาลนครชน จากนโยบายการตรวจผลตรวจสุขภาพประจำปีของโรงเรียน ผลการตรวจสุขภาพเป็นไฟล์ชนิด excel เพื่อให้ทราบถึงการเจริญเติบโตของนักเรียนว่าอยู่ในเกณฑ์ใด โรงเรียนและผู้ปกครองสามารถร่วมกันส่งเสริมโภชนาการเพื่อให้นักเรียนมีภาวะโภชนาการอยู่ในเกณฑ์ที่เหมาะสม โดยมีรหัสนักเรียน (HID) เป็นคีย์หลัก (Primary Key) ดังรูปที่ 3.6

ช่วงเวลาเบรค	ชื่อเบรค
05:00 - 07:45	MORNING
07:46 - 09:59	ในเวลาเรียน
10:00 - 10:15	เบรคเช้า
10:16 - 10:59	ในเวลาเรียน
11:00 - 11:49	Lunch ประถม
11:50 - 12:40	Lunch มัธยม
12:41 - 13:59	ในเวลาเรียน
14:00 - 14:15	เบรคบ่าย
14:16 - 15:29	ในเวลาเรียน
15:30 - 18:00	หลังเลิกเรียน

รูปที่ 3.7 ข้อมูลเวลาขายอาหาร

3.2 ปรับรูปแบบไฟล์ข้อมูลให้อยู่ในรูปแบบข้อมูล

ในการรวบรวมข้อมูลจากหลายแหล่งข้อมูล เพื่อให้สามารถนำเข้าฐานข้อมูลได้ต้องมีการปรับรูปแบบไฟล์ให้ตรงตามรูปแบบฐานข้อมูล เช่น ชื่อคอลัมน์ ควรเป็นภาษาอังกฤษที่สื่อความหมาย การปรับรูปแบบข้อมูลวันที่และเวลา ไม่ควรมีข้อมูลที่เป็นค่าว่าง ดังรูปที่ 3.8

PersonCode	PersonName	PersonSex	PersonAge	PG	BirthDate	PersonName	PersonSex	PersonAge	PersonID	PersonSex	PersonAge	PersonID	PersonSex	PersonAge	GuardID	Cur							
00002	00004	00014	00017	00020	00023	00026	00032	00037	00043	00044	00048	00061	00068	00070	00082	00084	00104	00108	00118	00113	00117	00128	00137

รูปที่ 3.8 ปรับรูปแบบข้อมูลให้อยู่ในรูปแบบฐานข้อมูล

3.2.1 การทำความสะอาดข้อมูล (Data Cleaning)

เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือข้อมูลในฐานข้อมูล ซึ่งถือเป็นหลักสำคัญของการจัดเก็บข้อมูลบนฐานข้อมูล หากนำข้อมูลดิบที่ไม่ผ่านการตรวจสอบหรือแก้ไขให้ถูกต้อง จะส่งผลให้เกิดความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่น ๆ ด้วย ในขั้นตอนการทำ Data Cleaning สามารถแยกกระบวนการย่อยได้ ดังนี้

3.2.1.1 ข้อมูลที่หายไป (Missing Data)

เมื่อเลือกชุดข้อมูลโดยผู้เชี่ยวชาญแล้ว ขั้นตอนต่อไปคือ การตรวจสอบข้อมูลและการทำความสะอาดข้อมูล ปัญหาที่พบได้บ่อยคือ ความไม่สมบูรณ์ของข้อมูล การขาดหายไปของข้อมูล อาจเกิดได้จาก ผู้บันทึก (human error), ความไม่พร้อมในการใช้งานของข้อมูลในขณะบันทึก, อุปกรณ์ไม่พร้อมหรือการขาดข้อกำหนดในการบันทึกข้อมูลที่ตรงกัน เป็นต้น ดังตัวอย่างที่ผู้วิจัยพบได้จากไฟล์ส่วนใหญ่ที่รวบรวมมา ดังรูปที่ 3.9

PersonCode	Gender	Age	Birthdate	Nationality	NationalName	Religion	Father's ID	Mother's ID
50002			13/07/2539	ไทย	พุทธ	นิกาย	5102100197568	5109900249372
50004			14/06/2540	ไทย	พุทธ	นิกาย	5104600107533	5102100197570
50014			3/12/2538	ไทย	พุทธ	นิกาย	5104600107533	5104000107573
50021			31/12/2539	ไทย	พุทธ	นิกาย	5101200812090	5101200812091
50026			13/11/2542	ไทย	พุทธ	นิกาย	5104000207076	5104000207076
50025			16/07/2540	ไทย	พุทธ	นิกาย	5101610066888	5109901064475
50034			5/05/2540	ไทย	พุทธ	นิกาย	5105000539865	5101016640715
50032				ไทย	พุทธ	นิกาย	5102199018862	5107600011564
50037			20/02/2537	ไทย	พุทธ	นิกาย	5109900126204	5109900126204
50043			5/10/2537	ไทย	พุทธ	นิกาย	5109900209114	5109900211424
50044			27/06/2541	ไทย	พุทธ	นิกาย	5101203182124	5101203182124
50049			28/05/2541	ไทย	พุทธ	นิกาย	5110200460439	5110100382871
50061			20/07/2541	ไทย	พุทธ	นิกาย	5102002499619	5101302053632
50068				ไทย	พุทธ	นิกาย	5102101603055	5101300254578
50070			0/05/2542	ไทย	พุทธ	นิกาย	5109904327641	5101800358776
50082			0/01/2539	ไทย				
50084			13/05/2542	ไทย	พุทธ	นิกาย	5101200016988	5109900391982
50108			10/10/2540	ไทย	พุทธ	นิกาย	5101202536909	5104000255036
50108			29/09/2541	ไทย	พุทธ	นิกาย	5101501982666	51020400746366
50110			25/07/2540	ไทย	พุทธ	นิกาย	5102200003337	5120700497143
50113			0/06/2542	ไทย	พุทธ	นิกาย	51009100175112	5102201613684
50115				ไทย	พุทธ	นิกาย	5100501185155	5104120060365
50117			3/06/2540	ไทย	พุทธ	นิกาย	5101900640871	5104000254969
50129			8/06/2543	ไทย	พุทธ	นิกาย	5100400059847	5102201171738
50137			17/07/2543	ไทย	พุทธ	นิกาย	5109900734307	51013011054005

รูปที่ 3.9 Missing Data ของข้อมูลที่ขาดหายไป

การแก้ไขข้อมูลเมื่อพบว่ามีข้อผิดพลาดของข้อมูล ผู้วิจัยได้ทำการลบข้อมูลนั้นออกไปเฉพาะในส่วนที่ไม่ส่งผลกระทบต่อผลการประมวลผล จำนวนข้อมูลตัวอย่างมีจำนวนไม่มากผู้วิจัยสามารถแก้ไขได้ด้วยตนเองโดยไม่ต้องใช้โปรแกรมอื่นช่วย

3.2.1.2 ค่าผิดปกติ (Noisy Data)

นอกเหนือจากข้อมูลที่ขาดหายไป อีกหนึ่งข้อมูลที่ควรวิเคราะห์คือค่าผิดปกติของข้อมูล เช่น ค่าน้อยเกินไป หรือมากเกินไปจากค่าปกติหรือค่ากลาง ผู้วิจัยพบความผิดปกติจากไฟล์บันทึกเหตุการณ์การซื้ออาหารเช่น ราคาอาหารถูกบันทึกเป็นราคา 1 บาท ซึ่งน้อยเกินไปจากการตรวจสอบ ไม่มีรายการอาหารใดที่มีราคา 1 บาท ดังรูปที่ 3.10 หรือค่าผิดปกติของข้อมูลการบันทึกราคาที่สูงเกินไป เป็นจำนวนหลายพันบาท ดังตัวอย่างรูปที่ 3.11

	A	B	C	D	E	F	G	H	I
1	MERCHANT	SVC_NBR	TRANS_DATE_TIME	TRANS_CODE	CAL	BREAK			
2	BAKERY-SB	4629510089892020	05/14/2018 11:43:00	Purchase	1	Lunch ปรธรม			
3	BAKERY-SB	4629510089897280	05/14/2018 15:59:33	Purchase	1	หลังเลิกเรียน			
4	BEVERAGE-SHOP	4629510089896180	05/16/2018 16:31:19	Purchase	1	หลังเลิกเรียน			
5	BEVERAGE-SHOP	4629510089890760	05/17/2018 11:23:28	Purchase	1	Lunch ปรธรม			
6	BEVERAGE-SHOP	4629510089889750	05/17/2018 16:36:44	Purchase	1	หลังเลิกเรียน			
7	BEVERAGE-SHOP	4629510089895500	05/17/2018 16:37:38	Purchase	1	หลังเลิกเรียน			
8	BEVERAGE-SHOP	4629510089896240	05/17/2018 16:52:23	Purchase	1	หลังเลิกเรียน			
9	BAKERY-SB	4629510089893250	05/18/2018 12:02:56	Purchase	1	Lunch มัธยม			
10	BEVERAGE-SHOP	4629510089895520	05/21/2018 11:26:58	Purchase	1	Lunch ปรธรม			
11	BEVERAGE-SHOP	4629510089889130	05/21/2018 12:10:06	Purchase	1	Lunch มัธยม			
12	BEVERAGE-SHOP	4629510089889240	05/21/2018 16:07:21	Purchase	1	หลังเลิกเรียน			
13	BEVERAGE-SHOP	4629510089897560	05/23/2018 12:12:54	Purchase	1	Lunch มัธยม			
14	BEVERAGE-SHOP	4629510104880520	05/23/2018 12:14:34	Purchase	1	Lunch มัธยม			
15	BAKERY-SHOP	4629510117173010	05/23/2018 16:42:12	Purchase	1	หลังเลิกเรียน			
16	BEVERAGE-SHOP	4629510089896790	05/23/2018 16:53:56	Purchase	1	หลังเลิกเรียน			
17	BEVERAGE-SHOP	4629510089890350	05/24/2018 16:59:47	Purchase	1	หลังเลิกเรียน			
18	BAKERY-SB	4629510089892040	05/25/2018 11:30:09	Purchase	1	Lunch ปรธรม			

รูปที่ 3.10 Noisy Data ค่าผิดปกติที่มีค่าน้อยเกินไป

	A	B	C	D	E	F	G	H	I
1	MERCHANT	SVC_NBR	TRANS_DATE_TIME	TRANS_CODE	CAL	BREAK			
2	ICE CREAM	4629510089891040	03/12/2019 12:38:48	Purchase	3500	Lunch มัธยม			
3	INTERFOOD	4629510089891060	12/04/2018 16:21:32	Purchase	3200	หลังเลิกเรียน			
4	NOODLE	4629510117174680	03/12/2019 10:20:08	Purchase	2200	ในเวลาเรียน			
5	NOODLE	4629510089891050	03/12/2019 11:38:02	Purchase	2100	Lunch ปรธรม			
6	INTERFOOD	4629510089891060	08/14/2018 10:57:24	Purchase	2000	ในเวลาเรียน			
7	NOODLE	4629510089888030	03/15/2019 11:46:06	Purchase	1600	Lunch ปรธรม			
8	NOODLE	4629510089888030	03/20/2019 11:37:57	Purchase	1600	Lunch ปรธรม			
9	NOODLE	4629510089888030	03/21/2019 11:49:32	Purchase	1600	Lunch ปรธรม			
10	NOODLE	4629510089888030	03/22/2019 11:45:16	Purchase	1600	Lunch ปรธรม			
11	INTERFOOD	4629510089891060	06/06/2018 11:32:40	Purchase	1500	Lunch ปรธรม			
12	INTERFOOD	4629510089891060	09/21/2018 11:20:31	Purchase	1500	Lunch ปรธรม			
13	INTERFOOD	4629510089888160	02/07/2019 11:00:36	Purchase	1500	Lunch ปรธรม			
14	NOODLE	4629510089888030	03/13/2019 11:42:11	Purchase	1500	Lunch ปรธรม			
15	NOODLE	4629510089888030	03/15/2019 11:32:32	Purchase	1500	Lunch ปรธรม			
16	INTERFOOD	4629510089891060	05/24/2018 12:07:53	Purchase	1480	Lunch มัธยม			
17	INTERFOOD	4629510089891060	05/31/2018 11:40:37	Purchase	1410	Lunch มัธยม			
18	INTERFOOD	4629510089888030	02/13/2019 10:50:16	Purchase	1400	ในเวลาเรียน			

รูปที่ 3.11 Noisy Data ค่าผิดปกติที่มีค่าสูงเกินไป

การแก้ปัญหา Noisy Data หรือค่าผิดปกติที่พบ ผู้วิจัยได้แก้ปัญหาด้วยวิธีการลบข้อมูลนั้นออกไปเฉพาะในส่วนที่ไม่ส่งผลกระทบต่อการประมวลผล จำนวนข้อมูลตัวอย่างมีจำนวนไม่มาก ผู้วิจัยสามารถแก้ไขได้ด้วยตนเองโดยไม่ต้องใช้โปรแกรมอื่นช่วย

4.2.1.3 ข้อมูลที่ไม่สอดคล้องกัน (Inconsistent Data)

เนื่องจากในขั้นตอนของการเตรียมข้อมูล เพื่อนำไปประมวลผล มีการรวบรวมข้อมูลจากหลายแหล่ง รูปแบบการจัดเก็บของแต่ละฝ่ายมีความเป็นอิสระต่อกัน ข้อมูลอาจมีความซ้ำซ้อนและเรียกชื่ข้อมูลไม่เหมือนกัน เช่น รหัสนักเรียน ฝ่ายทะเบียนใช้คำว่า ID ฝ่ายวิชาการใช้เรียก ST-ID เป็นต้น เพื่อไม่ให้เกิดความซ้ำซ้อน การเรียกใช้ข้อมูลอาจใช้ข้อมูลของฝ่ายใดฝ่ายหนึ่ง และกำหนดเป็นชื่อเดียวกัน หรือการกำหนดรูปแบบการป้อนข้อมูลบัตรประจำตัวประชาชนแสดงดังรูปที่ 3.12

	A	B	C	D	E	F	G
	PersonCode	FirstName	LastName	GenderName	PID	BirthDate	
1	01266			ชาย	1 1008 01687 58 3	11/8/2553	
2	01273			หญิง	1 1020 04209 05 1	10/22/2553	
3	01277			หญิง	1 1021 70141 10 1	10/4/2553	
4	01280			หญิง	1 1022 00305 86 6	8/13/2553	
5	01281			หญิง	1 1035 00101 81 7	3/9/2554	
6	01295			หญิง	1 1026 00260 53 0	8/6/2554	
7	01296			หญิง	2 1040 00008 91 4	4/4/2554	
8	01297			หญิง	1 1021 70155 86 1	7/30/2554	
9	01516			ชาย	E44397828		
10	01517			ชาย	E44397829		
11	01518			หญิง	EF1785631		
12	01113			ชาย	0101951681033	12/09/2550	
13	00043			ชาย	1100200991792	4/10/2537	
14	01156			หญิง	1100201082705	24/02/2548	
15	00113			หญิง	1100201458117	6/04/2542	
16	01371			ชาย	1100201624220	22/10/2544	
17	00632			หญิง	1100201800672	2/02/2548	
18	00000			หญิง	1100000762287	9/05/2548	

รูปที่ 3.12 Inconsistent Data ข้อมูลที่ไม่สอดคล้องกัน

ผู้วิจัยได้ปรับรูปแบบการบันทึกให้เหมือนกัน ที่เป็นรูปแบบเดียวกัน เพื่อแก้ปัญหาให้ข้อมูลมีความสอดคล้องกัน

3.2.1.4 การรวมข้อมูล (Data Integration)

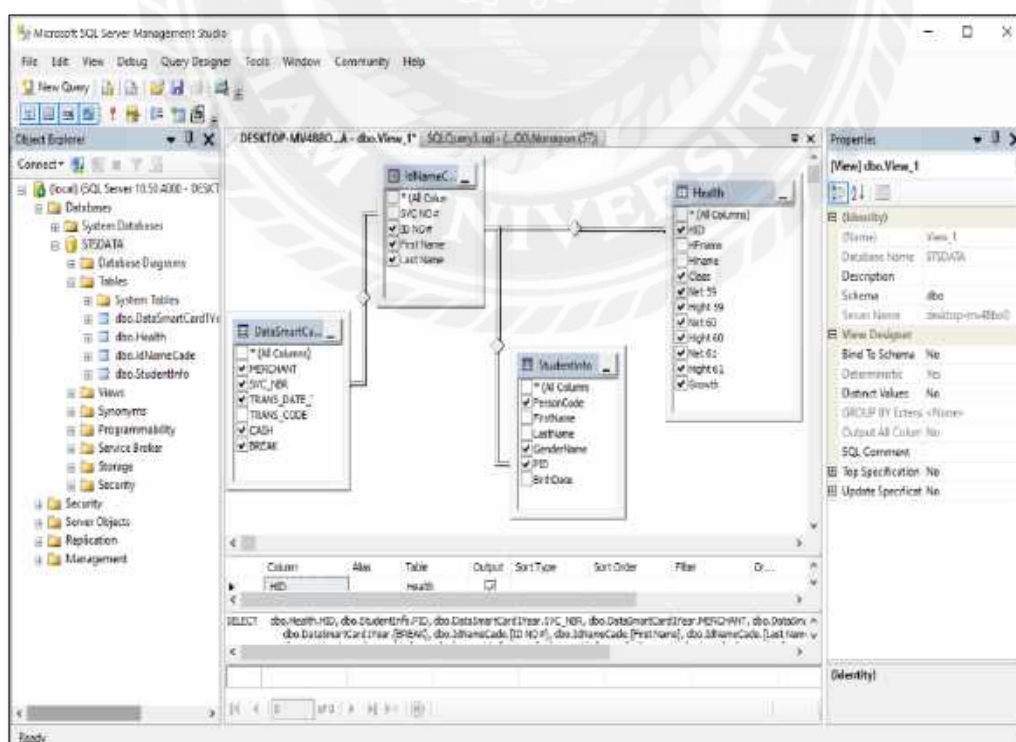
เนื่องจากข้อมูลที่นำมาประกอบการวิเคราะห์ ต้องผ่านขั้นตอนการจัดเตรียมข้อมูล ข้อมูลชนิดเดียวกันอาจมีชื่อเรียกเหมือนกัน หรือต่างกันได้ ในการเรียกใช้ข้อมูล ต้องมีการปรับให้มีรูปแบบที่เหมือนกันก่อนจากหลาย ๆ ตารางนำมารวมกัน

3.2.1.5 การสุมข้อมูล (Data Reduction)

ข้อมูลที่รวบรวมมาจากหลายแหล่งข้อมูลมีความเป็นไปได้ที่จะมีการเอนเอียงไปในข้อมูลชนิดใดชนิดหนึ่งสูง หรือเพื่อให้กลุ่มตัวอย่างแคบลงและไม่ส่งผลกระทบต่อผลการวิจัย การเลือกข้อมูลมาใช้ประกอบการวิเคราะห์จึงต้องสุมข้อมูลเพื่อให้เหมาะสมในการนำไปใช้งาน ดังตัวอย่างที่ผู้วิจัยเลือกกลุ่มข้อมูลเฉพาะกลุ่ม นั่นคือกลุ่มนักเรียนทุกคน โดยไม่นำข้อมูลของครูและบุคลากรในโรงเรียนมาร่วมวิเคราะห์ด้วย เนื่องจากในการประมวลผลด้วยเทคนิคเหมืองกระบวนกรหากนำข้อมูลการรับประทานอาหารซ้ำที่มีผลต่อภาวะน้ำหนักเกินเกณฑ์ บุคลากรกับนักเรียนจะมีพฤติกรรมการบริโภคที่ต่างกันมากเกิน ไป ผู้วิจัยจึงสุมหรือกำหนดเฉพาะกลุ่มของข้อมูลที่เป็นนักเรียนเท่านั้น

3.3 จัดเก็บและเชื่อมความสัมพันธ์ของข้อมูลบนฐานข้อมูล

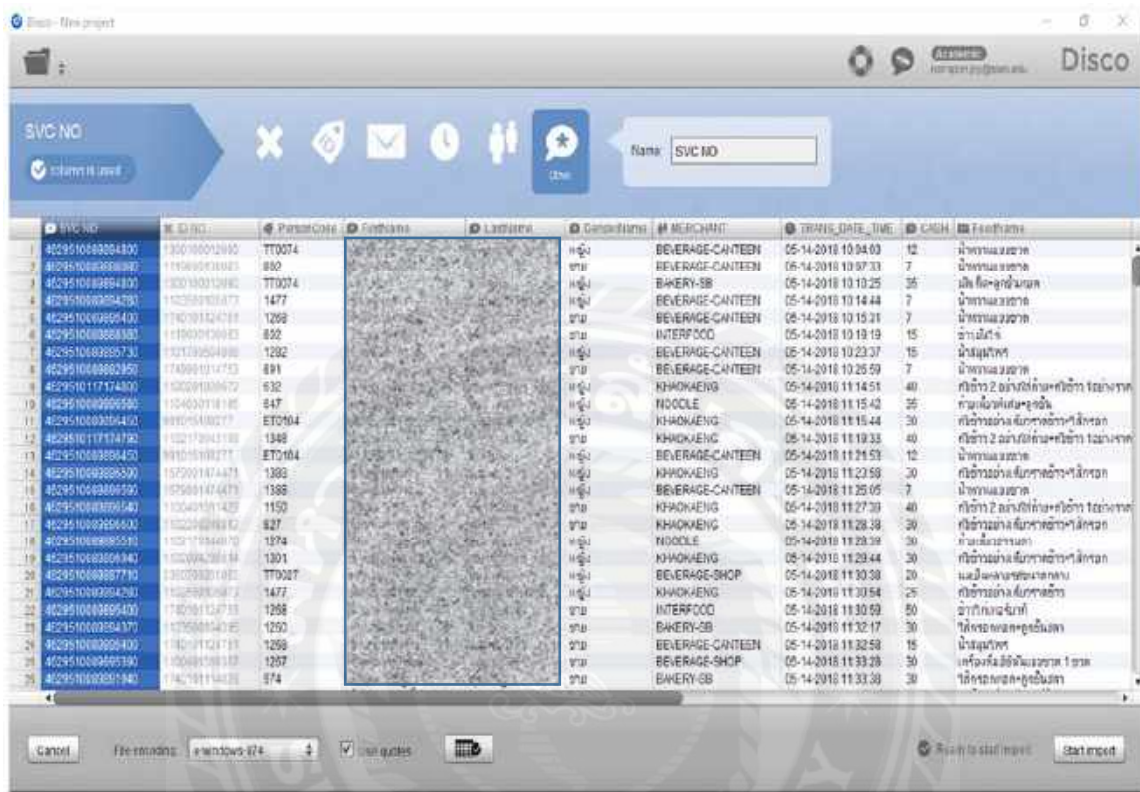
เพื่อให้ข้อมูลที่รวบรวมมาจากหลายแหล่งข้อมูลสามารถเชื่อมโยงถึงกันได้ ขั้นตอนนี้คือส่วนการนำเข้า (Import) ข้อมูลจากไฟล์ excel จัดเก็บข้อมูลเข้าสู่โปรแกรมฐานข้อมูล เชื่อมข้อมูลจากหลายตารางให้เป็นชุดข้อมูลเดียวกันและส่งออกข้อมูลเป็นไฟล์ CSV ดังรูปที่ 3.13



รูปที่ 3.13 จัดเก็บและเชื่อมความสัมพันธ์ของข้อมูลบนฐานข้อมูล

3.4 ส่งออกข้อมูลเป็นไฟล์ CSV และทดสอบนำเข้าข้อมูลด้วยโปรแกรม Disco

หลังจากเลือกข้อมูลจากหลายตารางที่เชื่อมความสัมพันธ์กัน ขั้นตอนสำคัญคือส่งออกข้อมูลจากโปรแกรมฐานข้อมูลออกมาในรูปแบบไฟล์ CSV และทดสอบนำเข้าข้อมูลสู่โปรแกรมวิเคราะห์ข้อมูลด้วยเทคนิคเหมือนกระบวนการคือโปรแกรม Disco ดังรูปที่ 3.14



รูปที่ 3.14 นำเข้าไฟล์ .csv เพื่อวิเคราะห์ข้อมูลด้วยโปรแกรม Disco

บทที่ 4 ผลการวิจัย

ผลการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อรวบรวมไฟล์ที่เกี่ยวข้องกับการซื้ออาหารของนักเรียนผ่านบัตรเงินสดดิจิทัลหรือบัตรสมาร์ตการ์ด ซึ่งเดิมข้อมูลนั้นถูกจัดเก็บแยกออกจากกัน ถูกบันทึกเป็นไฟล์ต่างชนิดกัน ไม่สะดวกต่อการนำมาวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการรวบรวมข้อมูลจากหลายแหล่งดังกล่าว ใช้เทคนิคการจัดเตรียมข้อมูล สามารถสรุปผลการวิจัยได้ดังนี้

4.1 ผลการเก็บรวบรวมข้อมูล

4.2 ผลการเตรียมข้อมูล

4.1 ผลการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อนำมาประกอบการวิจัย คือขั้นตอนที่ต้องใช้ความละเอียดและรอบคอบในการเลือกว่าข้อมูลชนิดใดควรเก็บ ข้อมูลใดเกี่ยวข้อง และไม่เกี่ยวข้องต่อการนำไปวิเคราะห์ การเก็บรวบรวมข้อมูล เป็นกระบวนการย่อยของ ขั้นตอนการจัดเตรียมข้อมูล เพื่อให้ได้ข้อมูลที่ถูกต้อง จำเป็นอย่างยิ่งที่ต้องมีผู้เชี่ยวชาญของระบบหรือเจ้าของข้อมูลทำหน้าที่เลือกข้อมูล เพื่อนำไปสู่ขั้นตอนการจัดเตรียมข้อมูลต่อไป จากผลการเก็บรวบรวมข้อมูลสามารถรวบรวมไฟล์ข้อมูล ได้ดังนี้

4.1.1 STS-Txns_ByDate_2017-2018.csv คือ ข้อมูลบันทึกเหตุการณ์การซื้ออาหารของนักเรียนรวมระยะเวลา 1 ปีการศึกษา ได้รับจากบริษัทผู้ให้บริการบัตรสมาร์ตการ์ด ตัวอย่างดังรูปที่ 4.1

Date	SVC Number	จำนวนครั้ง	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)	จำนวนเงิน (บาท)
20170104	4629510074707127							
20170104	4629510074707139							
20170104	4629510074707143							
20170104	4629510074707150							
20170104	4629510074707200							
20170104	4629510074707291							
20170104	4629510074707309				1	20		
20170104	4629510074707317							
20170104	4629510074707325							
20170104	4629510074707307							
20170104	4629510074707531							
20170104	4629510074707580						1	55
20170104	4629510074707689							
20170104	4629510074707713							
20170104	4629510074707739							
20170104	4629510074707762							
20170104	4629510074707879						1	40
20170104	4629510074707949						2	90
20170104	4629510074707960							
20170104	4629510074708034							

รูปที่ 4.1 ข้อมูลบันทึกเหตุการณ์การซื้ออาหาร

4.1.2 SVCNo-ID-Name.xlsx คือ ข้อมูลการทำบัตรนักเรียนจากบัตรสมาร์ทการ์ด ได้รับข้อมูลจากฝ่ายธุรการ โดยไฟล์ดังกล่าวบันทึกข้อมูลการทำบัตรนักเรียนและพนักงานทุกคน อ้างอิงจากระหัสหน้าบัตรกับรหัสประจำตัวประชาชน โดยบัตรนี้ใช้เป็นบัตรประจำตัวนักเรียนและสามารถเติมเงิน และใช้ซื้ออาหารแทนเงินสดได้ ดังตัวอย่างข้อมูลในรูปที่ 4.2

SVC NO.	ID NO.	First Name	Last Name
4629510089896477	0991014935671		
4629510117173261			
4629510117172040	P0506188A		
4629510089895099	0991005964396		
4629510089896469	0991004983591		
4629510117171992	0991011883980		
4629510117174111	HG032024		
4629510117174095	0991015856383		
4629510117173576			
4629510089896089	0991014713201		
4629510117174053			
4629510089890280			
4629510117172156	1100201059166		
4629510089889274	1100201800672		
4629510089890330	1100401165865		

รูปที่ 4.2 ไฟล์ข้อมูลการทำบัตรนักเรียนจากบัตรสมาร์ทการ์ด

4.1.3 FullData-Student1.csv คือข้อมูลนักเรียนที่ได้จากระบบทะเบียนและวัดผล ดังรูปที่ 4.3 โดยได้รับความอนุเคราะห์จากฝ่ายทะเบียน ส่งออกข้อมูลจากระบบทะเบียนและวัดผลเป็นไฟล์ชนิด CSV ดังตัวอย่างข้อมูลตามรูปที่ 4.4



รูปที่ 4.3 ระบบทะเบียนและวัดผลของนักเรียน

4.2 ผลการเตรียมข้อมูล(Data preprocessing)

เมื่อเลือกแหล่งข้อมูลและรวบรวมข้อมูลได้แล้ว ขั้นตอนถัดมาคือการเตรียมข้อมูล เพื่อเตรียมพร้อมสำหรับการนำข้อมูลเพื่อไปประมวลผลต่อ ข้อมูลที่ผ่านขั้นตอนการเตรียมข้อมูลจะต้องมีความถูกต้องครบถ้วนและไม่มีข้อมูลที่มีค่าผิดปกติ ดังนั้นก่อนนำข้อมูลไปวิเคราะห์หรือประมวลผลต่อด้วยวิธีการใดก็ตาม ต้องมีการจัดเตรียมข้อมูลและทำความสะอาดข้อมูลอย่างถูกต้องและเป็นขั้นตอน จากตัวอย่างข้อมูลที่รวบรวมมา เมื่อผ่านการทำความสะอาดข้อมูลแล้วผู้วิจัยได้ปรับรูปแบบการบันทึกข้อมูลให้เป็นชนิดเดียวกันทั้งหมดโดยบันทึกให้อยู่ในรูปแบบไฟล์ excel ได้ข้อมูลดังนี้

4.2.1 STS-Txns_ByDate_2017-2018.xlsx คือ ข้อมูลบันทึกเหตุการณ์การซื้ออาหารของนักเรียนรวมระยะเวลา 1 ปีการศึกษา ผู้วิจัยได้ใช้เทคนิคในการเตรียมข้อมูลและทำความสะอาดข้อมูลได้ผลลัพธ์ดังรูปที่ 4.6

MERCHANT	SVC_NBR	TRANS_DATE_TIME	TRANS_CODE	CASH	BREAK
INTERFOOD	4629510117714480	05/08/2018 11:07:12	Purchase	44	Lunch ประถม
BEVERAGE-CANTEEN	4629510117714480	05/08/2018 11:13:28	Purchase	44	Lunch ประถม
BEVERAGE-SB	4629510117714480	05/08/2018 11:18:59	Purchase	44	Lunch ประถม
SPARE	4629510117714480	05/08/2018 11:26:38	Purchase	44	Lunch ประถม
GIFT SHOP	4629510117714480	05/08/2018 11:36:38	Purchase	44	Lunch ประถม
ICE CREAM	4629510117714480	05/08/2018 11:42:44	Purchase	44	Lunch ประถม
BEVERAGE-SHOP	4629510117714480	05/08/2018 11:48:48	Purchase	44	Lunch ประถม
BEVERAGE-SHOP	4629510117714480	05/08/2018 11:49:27	Purchase	44	Lunch ประถม
BEVERAGE-SHOP	4629510117714480	05/08/2018 11:50:26	Purchase	44	Lunch มัธยม
KHAKAENG	4629510117714480	05/08/2018 12:02:57	Purchase	44	Lunch มัธยม
BAKERY-SHOP	4629510082717720	05/08/2018 12:32:38	Purchase	4	Lunch มัธยม
ORDER-CANTEEN	4629510117714480	05/08/2018 12:56:45	Purchase	44	ในเวลาเรียน
SPARE	4629510082717720	05/08/2018 12:59:46	Purchase	4	ในเวลาเรียน
NOODLE	4629510082717720	05/08/2018 13:33:06	Purchase	4	ในเวลาเรียน
NOODLE	4629510117714480	05/08/2018 13:33:51	Purchase	44	ในเวลาเรียน
BEVERAGE-SHOP	4629510116884220	05/14/2018 7:14:04	Purchase	30	MORNING

รูปที่ 4.6 ข้อมูลบันทึกเหตุการณ์การซื้ออาหารของนักเรียนหลังใช้เทคนิคการเตรียมข้อมูล

4.2.2 SVCNo-ID-Name.xlsx คือ ข้อมูลการทำบัตรนักเรียนจากบัตรสมาร์ทการ์ด เมื่อผ่านขั้นตอนการเตรียมข้อมูล ผู้วิจัยได้เลือกชุดข้อมูลเฉพาะที่อยู่ในช่วงที่สนใจเนื่องจาก ข้อมูลที่ได้รับจากฝ่ายธุรการเป็นข้อมูลรวมของนักเรียนทั้งหมดจำนวนหลายปีที่มีการทำบัตรสมาร์ทการ์ดทำให้มีข้อมูลที่ไม่จำเป็นมากเกินไป ซึ่งได้ผลลัพธ์ดังรูปที่ 4.7

SVC NO.	ID NO.	First Name	Last Name
4629510089896469	0991004983591		
4629510089895099	0991005964396		
4629510089891023	0991011883858		
4629510117171992	0991011883980		
4629510089896014	0991013776455		
4629510089896089	0991014713201		
4629510089896584	0991014713295		
4629510089890645	0991014715379		
4629510089897483	0991014901580		
4629510089896477	0991014935671		
4629510089896450	0991015199277		
4629510089896352	0991015199463		
4629510117174095	0991015856383		
4629510117172156	1100201059166		
4629510089889274	1100201800672		
4629510089890330	1100401165865		
4629510089897673	1100401170150		
4629510089889878	1100401203686		
4629510089897632	1100401257964		

รูปที่ 4.7 ข้อมูลการทำบัตรนักเรียนจากบัตรสมาร์ทการ์ดหลังใช้เทคนิคการเตรียมข้อมูล

4.2.3 FullData-Student1.xlsx คือข้อมูลนักเรียนที่ได้จากระบบทะเบียนและวัดผล ฝ่ายทะเบียนได้ทำการเก็บบันทึกข้อมูลผ่านโปรแกรมสำเร็จรูป ผู้วิจัยได้รับข้อมูลส่งออกจากระบบเป็นรูปแบบ csv ไฟล์ หลังใช้เทคนิคการเตรียมข้อมูลได้ผลลัพธ์ดังรูปที่ 4.8

PersonCode	FirstName	LastName	GenderName	PID	BirthDate	NationName
00253			ชาย	1102400126141	19/11/2543	ไทย
00256			หญิง	1100801364351	9/09/2543	ไทย
00259			หญิง	1100703111326	31/10/2543	ไทย
00263			หญิง	1100401081629	7/05/2544	ไทย
00264			ชาย	1579900915724	12/03/2544	ไทย
00265			ชาย	1100801305788	23/05/2542	ไทย
00270			ชาย	1101500745232	12/06/2537	ไทย
00272			ชาย	1103900063408	7/08/2543	ไทย
00274			ชาย	1100401092345	30/07/2544	ไทย
00275			ชาย	1103900067110	25/09/2543	ไทย
00278			หญิง	1100703074960	11/08/2543	ไทย
00279			หญิง	1103900085061	23/07/2544	ไทย
00280			หญิง	1104000036537	3/08/2543	ไทย
00281			หญิง	1102170019241	23/11/2543	ไทย
00283			หญิง	1103500033676	26/04/2544	ไทย
00284			ชาย	135137327	1/03/2544	ไทย
00287			หญิง	1103500025975	20/05/2540	ไทย
00289			หญิง	1102200148345	18/05/2539	ไทย
00292			หญิง	1100400752287	9/05/2538	ไทย

รูปที่ 4.8 ข้อมูลนักเรียนจากระบบทะเบียนและวัดผลหลังใช้เทคนิคการเตรียมข้อมูล

4.2.4 HealthAll61.xlsx คือ ข้อมูลผลการตรวจสุขภาพของประจำปีของนักเรียน หลังใช้เทคนิคการเตรียมข้อมูลได้ผลลัพธ์ดังรูปที่ 4.9

HID	Hfname	Hlname	Class	Net 59	Hight 59	Net 60	Hight 60	Net 61	Hight 61	Growth
01165			ประถมศึกษาปีที่ 3/2	19.2	114.3	21.2	120.9	25.8	128.8	อยู่ในเกณฑ์ปกติ
01166			ประถมศึกษาปีที่ 3/2	25.3	118.1	30.8	125.5	40.2	133	น้ำหนักเกินเกณฑ์
01168			ประถมศึกษาปีที่ 3/2	27.5	120.5	33.6	126.4	41.3	133	น้ำหนักเกินเกณฑ์
01169			ประถมศึกษาปีที่ 3/2	33.1	123.8	35.4	129	38	134.8	น้ำหนักเกินเกณฑ์
01170			ประถมศึกษาปีที่ 3/2	22.7	115.7	25.6	121	26.7	130	อยู่ในเกณฑ์ปกติ
01171			ประถมศึกษาปีที่ 3/2	25.5	117	29.9	123.1	37.7	130.7	น้ำหนักเกินเกณฑ์
01172			ประถมศึกษาปีที่ 3/2	30.2	120.8	30.7	126.8	34.8	134.7	อยู่ในเกณฑ์ปกติ
01173			ประถมศึกษาปีที่ 3/2	25.1	112.1	31.4	119.8	37.9	130	น้ำหนักเกินเกณฑ์
01188			ประถมศึกษาปีที่ 3/2	35.4	125.8	42	133.4	49.4	141	น้ำหนักเกินเกณฑ์
01209			ประถมศึกษาปีที่ 3/2	21.3	115.1	24.3	122	29	131.6	อยู่ในเกณฑ์ปกติ
01175			ประถมศึกษาปีที่ 3/2	22.9	116.3	26	123.6	33.8	131	อยู่ในเกณฑ์ปกติ
01177			ประถมศึกษาปีที่ 3/2	17.6	110.9	19.5	118.5	22.5	126.1	อยู่ในเกณฑ์ปกติ
01178			ประถมศึกษาปีที่ 3/2	32	120	36.5	128.1	43.5	136	น้ำหนักเกินเกณฑ์
01179			ประถมศึกษาปีที่ 3/2	29.3	120.5	27	127.6	39.6	134.2	น้ำหนักเกินเกณฑ์
01180			ประถมศึกษาปีที่ 3/2	17.3	106	20.1	113.7	24.3	122	อยู่ในเกณฑ์ปกติ
01181			ประถมศึกษาปีที่ 3/2	18.1	112.1	21.7	120.2	30	128.5	อยู่ในเกณฑ์ปกติ
01182			ประถมศึกษาปีที่ 3/2	22.3	117.6	22.7	127.3	31.6	135	อยู่ในเกณฑ์ปกติ

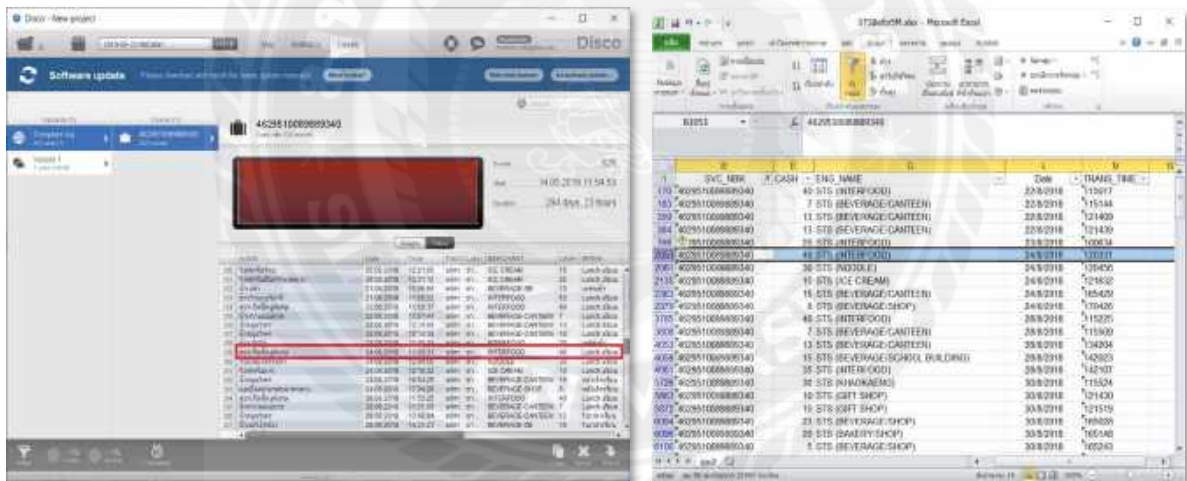
รูปที่ 4.9 ข้อมูลผลตรวจสุขภาพประจำปีของนักเรียนหลังใช้เทคนิคการเตรียมข้อมูล

สรุปอภิปรายผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้นำเสนอเทคนิคการจัดเตรียมข้อมูลเพื่อนำข้อมูลไปประมวลผลด้วยโปรแกรมเทคนิคเหมืองกระบวนกร เมื่อผ่านขั้นตอนการจัดเตรียมข้อมูลแล้ว ข้อมูลจากหลายส่วน จะเข้าสู่กระบวนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกันหรือไฟล์เดียวกัน และนำเข้าข้อมูลเพื่อวิเคราะห์ด้วยโปรแกรม Disco ต่อไป

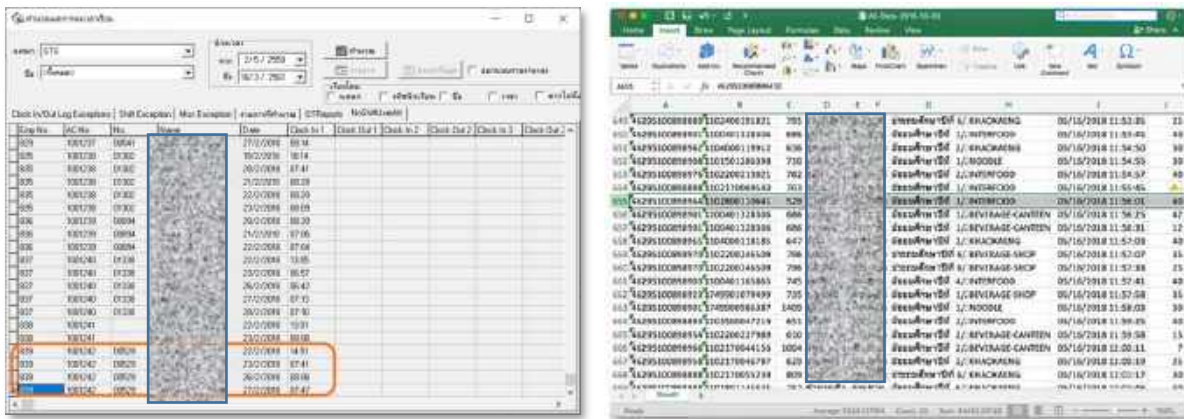
5.1 สรุปผลการวิจัย

การใช้เทคนิคการจัดเตรียมข้อมูลสำหรับจัดเตรียมไฟล์ต่าง ๆ เพื่อนำไปวิเคราะห์ด้วยเทคนิคเหมืองกระบวนกร ไฟล์ข้อมูลที่พร้อมสำหรับการนำไปวิเคราะห์ต่อจะเหลือเพียง 1 ตารางอยู่ในรูปแบบไฟล์ csv ผู้วิจัยได้ทำการทดสอบนำข้อมูลส่งเข้าไปประมวลผลด้วยโปรแกรม Disco เพื่อตรวจสอบย้อนกลับโดยการเปรียบเทียบจากชุดข้อมูลบันทึกเหตุการณ์ (Even log) ว่ามีข้อมูลถูกต้องตรงกันกับข้อมูลที่นำไปประมวลผลด้วยโปรแกรม Disco หรือไม่ แสดงดังรูปที่ 5.1



รูปที่ 5.1 เปรียบเทียบข้อมูลบันทึกเหตุการณ์กับชุดข้อมูลที่ประมวลผลด้วยโปรแกรม Disco

ผลการตรวจสอบพบว่า ข้อมูลที่ปรากฏอยู่ใน โปรแกรม Disco เป็นข้อมูลที่ถูกต้องตรงกันกับข้อมูลที่อยู่ในบันทึกเหตุการณ์ ผู้วิจัยได้ตรวจสอบกับข้อมูลในไฟล์ข้อมูลอื่น ๆ ด้วย เพื่อให้การตรวจสอบข้อมูลย้อนกลับถูกต้องและครบถ้วน โดยตรวจสอบความถูกต้องของผู้ถือบัตรจากบันทึกเหตุการณ์ว่าตรงตามชื่อนักเรียนในระบบหรือไม่ แสดงได้ดังรูปที่ 5.2



รูปที่ 5.2 ตรวจสอบข้อมูลบันทึกเหตุการณ์กับรายชื่อนักเรียนในระบบ

5.2 ข้อเสนอแนะ

จากการรวบรวมข้อมูลที่มาจากหลายแหล่งข้อมูล เพื่อใช้เทคนิคการเตรียมข้อมูลสำหรับนำข้อมูลไปประมวลผลต่อด้วยเทคนิคเหมืองกระบวนการ ขั้นตอนการรวบรวมข้อมูลและทำความสะอาดข้อมูล เป็นขั้นตอนที่ใช้เวลาในการทำงานมากถึง 80% ของการวิเคราะห์ข้อมูล ผู้วิจัยได้พบกับปัญหาที่เกิดจากการจัดเตรียมข้อมูลสามารถแบ่งปัญหาและข้อเสนอแนะได้ดังนี้

5.2.1 ข้อมูลบันทึกเหตุการณ์ซื้อขายอาหาร

ข้อมูลหลักที่ใช้ในการประมวลผลของงานวิจัยในครั้งนี้คือข้อมูลการซื้อขายอาหารผ่านบัตรเงินสดดิจิทัล ซึ่งเป็นข้อมูลที่ได้จากบันทึกเหตุการณ์ของระบบซื้อขายอาหารผ่านระบบการตัดและเติมเงินของผู้ให้บริการภายนอก เมื่อโรงเรียนต้องการทราบข้อมูลการซื้อขายรายวัน บริษัทผู้ให้บริการจะจัดส่งไฟล์ชนิด pdf ในรูปแบบตารางมาให้ฝ่ายบริหาร เมื่อต้องนำข้อมูลมาใช้งานเพื่อประมวลผลต่อ ผู้วิจัยต้องแปลงข้อมูลให้อยู่ในรูปแบบของ work sheet ทำให้เพิ่มขั้นตอนการทำงานมากขึ้น และผลลัพธ์ของการแปลงไฟล์ไม่มีความสมบูรณ์ นอกจากนี้ยังคงพบปัญหาการปกปิดข้อมูลบางส่วนที่ไม่สามารถทราบได้ว่ารายการซื้อขายอาหารนั้น ๆ เป็นของนักเรียนคนใดซึ่งไม่สามารถร้องขอเพิ่มเติมจากผู้ให้บริการได้

รายการข้อมูลในบันทึกเหตุการณ์มีความผิดพลาดหลายอย่าง สิ่งที่สำคัญคือการบ่อนราคาลินค้าก่อนชำระเงินของร้านค้าไม่ถูกต้อง มีการบ่อนราคาน้อยเกินไป มากเกินไปจำนวนมาก แสดงให้เห็นถึงความผิดพลาดของพนักงาน ที่ก่อให้เกิดความสูญเสียรายได้จำนวนมาก

ข้อเสนอแนะ ควรมีการฝึกอบรมการใช้งานเครื่องคิดเงินอย่างถูกต้องให้กับพนักงานที่เกี่ยวข้อง มีระบบฝึกอบรมพนักงานใหม่ทุกครั้งเพื่อป้องกันความสูญเสียรายได้

5.2.2 ข้อมูลผู้ถือบัตรเงินสดดิจิทัล

ข้อมูลผู้ถือบัตรเงินสดดิจิทัล คือข้อมูลที่มีความสำคัญมากในการเชื่อมโยงกับบันทึกเหตุการณ์การซื้อขายอาหาร เนื่องจากสามารถระบุได้ว่ารายการชื้อขายนั้น ๆ ก็ือรายการของนักเรียนคนใด ทำให้ง่ายต่อการนำไปประมวลผลต่อ จากการรวบรวมข้อมูล แฝ้มข้อมูลดังกล่าวไม่มีการสร้างขึ้นทำให้การอ้างอิงการห้สมัครบัตรห้สำนักเรียนที่เป็นผู้ถือบัตรไม่มีความเชื่อมโยงกัน บัตรเงินสดดิจิทัลเป็นการ์ดที่

ออกโดยฝ่ายกิจการนักเรียน สิ่งพิมพ์และป้อนข้อมูลนักเรียนลงบนบัตร หน้าบัตรมีหมายเลขประจำบัตรจะปรากฏในบันทึกเหตุการณ์การซื้ออาหาร แต่เมื่อนำบัตรไปทาบเพื่อใช้ในการบันทึกเวลาเรียนจะได้ค่าตัวเลขอีกค่าหนึ่งซึ่งไม่ใช่เลขเดียวกันกับที่ปรากฏหน้าบัตร ส่งผลให้เกิดความยุ่งยากในการเชื่อมโยงข้อมูล

ข้อเสนอแนะ ในขั้นตอนการทำบัตรนักเรียน ควรมีการสร้างไฟล์และลงรายละเอียดข้อมูลให้ครอบคลุมทั้งหมด เพื่อให้ง่ายต่อการนำข้อมูลไปประมวลผลต่อไป

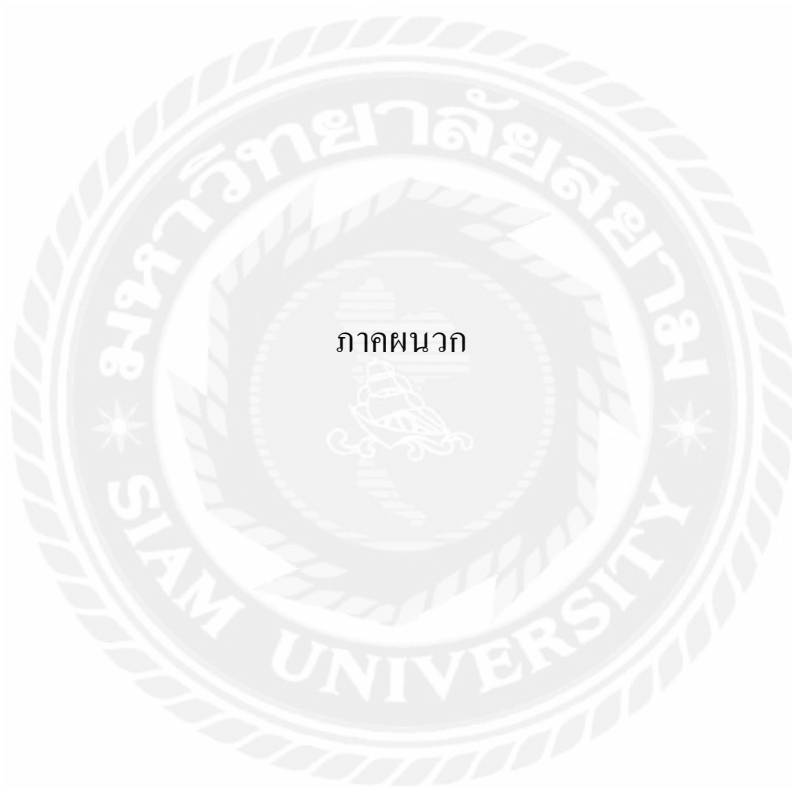
สรุป การใช้เทคนิคการจัดเตรียมข้อมูลสำหรับจัดเตรียมไฟล์ต่าง ๆ เพื่อนำไปวิเคราะห์ด้วยเทคนิคเหมืองกระบวนการ ในขั้นตอนการเลือกไฟล์ที่จะนำมาใช้ในขั้นตอนการเตรียมข้อมูลจำเป็นอย่างยิ่งที่ต้องให้ผู้เชี่ยวชาญของระบบนั้น ๆ เป็นผู้เลือกและระบุไฟล์ที่สามารถนำไปใช้ได้ เมื่อผ่านขั้นตอนการทำความสะอาดข้อมูล ประโยชน์ที่ได้รับโดยตรงของขั้นตอนนี้คือ การมองเห็นความผิดปกติที่พบจากไฟล์ข้อมูล ผู้เชี่ยวชาญของแต่ละระบบสามารถบอกได้ว่าเกิดอะไรขึ้น เกิดจากปัญหาอะไร และมีวิธีแก้ไขเพื่อให้ข้อมูลมีความถูกต้องสมบูรณ์ในอนาคต ข้อมูลที่ผ่านเทคนิคการจัดเตรียมข้อมูลที่ต้องการสามารถนำไปวิเคราะห์ประมวลผลต่อไปได้มีประสิทธิภาพ



บรรณานุกรม

- กระบวนการขาย. (ม.ป.ป.). เข้าถึงได้จาก <https://sites.google.com/site/rabobpatibatkan/kar-khay-beuxng-tn-2/krabwn>
- ปูลวิษข์ ทองแดง และจันทร์จิรา สีสว่าง. (2012). ภาวะน้ำหนักเกินในเด็กไทย. *รามาศิษย์พยาบาลสาร*, 18(3), 287-297.
- ภูริเดช อากาศัตย์, และนุชรี เปรมชัยสวัสดิ์. (2562). การวิเคราะห์เนื้อหาของเอกสารประกอบการสอนด้วยเทคนิคเหมืองกระบวนการ. *วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยสยาม*, 20(1/38), 54-59.
- ภูริเดช อากาศัตย์, นุชรี เปรมชัยสวัสดิ์ และวิเชียร เปรมชัยสวัสดิ์. (2556). การใช้เหมืองกระบวนการเพื่อค้นพบการทำงานร่วมกันของนักศึกษาในชั้นเรียน. *วารสารวิทยาศาสตร์และเทคโนโลยี มทร. ชัยบุรี*, 9(1). 91-100.
- วิเชียร เปรมชัยสวัสดิ์. (2558). เหมืองกระบวนการ. *วารสารวิศวกรรมศาสตร์มหาวิทยาลัยสยาม*, 16(1/30), 1-10.
- เอกราช เจริญผล. (2559, 31 มกราคม). Process mining คืออะไร [เว็บไซต์]. เข้าถึงได้จาก <http://siam-processmining.blogspot.com/p/process-mining.html>
- เอนก นามขันธ์ และทมนี สุโกส. (2559). วิเคราะห์แนวทางแก้ปัญหากระบวนการทำงานของระบบด้วย Fuzzy Miner ที่มีต่อการตัดสินใจรับวัดกรรมการจัดการจัดการเรียนการสอนผ่านเว็บ. Retrieved from http://www.elcls.ssru.ac.th/tommanee_so/pluginfile.php/17/block_html/content/Anake_TRU_BENJAMITRA2016_NO106-1.pdf
- เอนก นามขันธ์ และเสาวภา เมืองแก่น. (2560). กระบวนการจัดการข้อมูลเพื่อใช้วิเคราะห์ผลด้วยวิธีกระบวนการเหมืองข้อมูล. *วารสารวิชาการมหาวิทยาลัยธนบุรี (วิทยาศาสตร์และเทคโนโลยี)*, 1(1), 33-41.
- Bartolini, G. (2001). Web usage mining and discovery of association rules from http servers logs. Melbourne: Monash University.
- Cooley, R., Tan, P., & Srivastava, J. (1999). Websift: The web site information filter system internet. Retrieved from http://dmr.cs.umn.edu/Papers/P1999_16.pdf
- Günther, C. W., & Wil MP Van Der Aalst. Fuzzy mining—adaptive process simplification based on multi-perspective metrics. *Business Process Management. Springer Berlin Heidelberg*, 328-343. Retrieved from https://link.springer.com/chapter/10.1007%2F978-3-540-75183-0_24

- Han, J., & Kamber, M. (2006). *Data mining concepts and techniques* (2nd ed.). San Francisco: Morgan Kaufman.
- Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM SIGKDD*, 2(1), 1-15.
- Madtthanawongsakorn, C., Saguansakdiyotin, N., Porouhan, P., Poohridate, A., & Premchaiswad, W. (2019). Applying process mining to investigate the relation between food purchase behavior and children's weight based on the food digital cards. In *Conference: 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)* (pp. 194-198). doi 10.1109/ICTKE47035.2019.8966860
- Process Mining Group. (2016). Fuzzy Miner. Retrieved from <http://www.processmining.org/online/Fuzzyminer>
- Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining: Discovery and application of usage patterns from web data. *SIGKDD Exploration*, 1(2), 12-23.
- van der Aalst, W. (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Springer-Verlag, Berlin. Retrieved from <https://link.springer.com/book/10.1007/978-3-642-19345-3>
- Wikipedia. (n.d.). Process mining. Retrieved June 15, 2019 from https://en.wikipedia.org/wiki/Process_mining
- Wikipedia. (n.d.). Rapid miner. Retrieved June 15, 2019 from <https://en.wikipedia.org/wiki/RapidMiner>
- Wvdaalst. (2016, August 24). Processmining: Event logs [Blog post]. Retrieved from <http://www.processmining.org/logs/start>



ภาคผนวก

ประวัติผู้วิจัย

ชื่อ – สกุล

ว่าที่ ร.ต.นรภณ จ้อยฟอง

วันเดือนปีเกิด

27 กรกฎาคม พ.ศ. 2524

วุฒิการศึกษา

ปริญญาตรี

มหาวิทยาลัยธนบุรี(วิทยาลัยเทคโนโลยีธนบุรี, 2547)

คณะวิทยาศาสตร์และเทคโนโลยี

สาขาเทคโนโลยีสารสนเทศ

