

บทที่ 3

ขั้นตอน และวิธีการดำเนินงาน

โครงการเรื่อง การเปรียบเทียบโมเดลในการวิเคราะห์ข้อมูลการยืมหนังสือในห้องสมุด ของ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เชียงใหม่ โดยการวิเคราะห์ข้อมูลแบบ CRISP-DM ซึ่งเป็นขั้นตอนมาตรฐานในการวิเคราะห์ข้อมูล เพื่อทำความเข้าใจรูปแบบและความสัมพันธ์ของข้อมูลในการยืมหนังสือ โมเดลที่ใช้ในการเปรียบเทียบ และวิเคราะห์ประกอบไปด้วย 2 โมเดลหลัก ได้แก่ 1.โมเดล Apriori และ 2. โมเดล FP-Growth เมื่อกระบวนการวิเคราะห์ข้อมูลเสร็จสมบูรณ์ ขั้นตอนต่อไปคือการออกแบบเว็บไซต์ รวมถึงการจัดรูปแบบการแสดงผล และสรุปผลจากกระบวนการดำเนินงาน โดยมีขั้นตอนดังต่อไปนี้

3.1 ขั้นตอน และวิธีการวิเคราะห์ข้อมูลด้วย CRISP-DM

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding)

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

3.1.3 การเตรียมข้อมูล (Data Preparation)

3.1.4 การสร้างแบบจำลอง (Modeling)

3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation)

3.1.6 การนำโมเดลไปใช้งานจริง (Deployment)

3.2 การวิเคราะห์และออกแบบระบบ

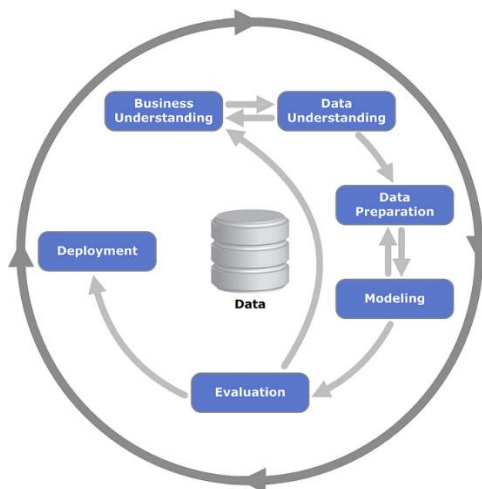
3.2.1 การวิเคราะห์และออกแบบฐานข้อมูล

3.2.2 การออกแบบเว็บไซต์

3.1 ขั้นตอน และวิธีการวิเคราะห์ข้อมูลด้วย CRISP-DM

CRISP-DM หรือ Cross Industry Standard Process for Data Mining ถูกสร้างขึ้นในปี 1996 จากความร่วมมือของหลายองค์กรที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล ได้แก่ SPSS, NCR, และ Daimler AG ซึ่งในขณะนั้นมีความต้องการมาตรฐานที่เป็นระบบในการทำ Data Mining เพื่อให้กระบวนการวิเคราะห์ข้อมูลมีความเป็นระบบมากขึ้นโดย CRISP-DM ได้รับการพัฒนาและเผยแพร่

สู่สาธารณะในปี 1999 ซึ่งได้กลายมาเป็นมาตรฐานที่ใช้กันอย่างแพร่หลายทั้งในวงการธุรกิจและวงการวิชาการ



ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

ที่มา : <https://n9.cl/3lly74>

ขั้นตอนและรายละเอียดของ CRISP-DM กระบวนการ CRISP-DM ประกอบไปด้วย 6 ขั้นตอนหลัก ซึ่งแต่ละขั้นตอนมีบทบาทสำคัญต่อความสำเร็จของโครงการวิเคราะห์ข้อมูล ดังรายละเอียดต่อไปนี้

3.1.1 กระบวนการศึกษาทำความเข้าใจธุรกิจ (Business Understanding) ขั้นตอนนี้เป็นขั้นตอนแรก เริ่มด้วยการทำความเข้าใจปัญหาทางธุรกิจที่ต้องการแก้ไข ซึ่งเป็นการตั้งคำถามที่ชัดเจนเกี่ยวกับสิ่งที่องค์กรต้องการรู้จากข้อมูล เช่น การวิเคราะห์ข้อมูลการยืมหนังสือในห้องสมุด วัตถุประสงค์หลักคือการใช้ข้อมูลเพื่อตรวจสอบพฤติกรรมกรรมการยืมหนังสือของผู้ใช้ ซึ่งจะช่วยให้ห้องสมุดสามารถจัดการกับการยืมหนังสือได้อย่างมีประสิทธิภาพยิ่งขึ้น อีกทั้งยังสามารถใช้ในการตัดสินใจเกี่ยวกับการเพิ่มหนังสือใหม่ ๆ หรือลดปริมาณหนังสือที่ไม่ค่อยมีคนยืมได้อย่างเหมาะสม

โดยคณะผู้จัดทำทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้ คือ ข้อมูลการยืมหนังสือในห้องสมุด ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ในช่วงประจำปี 2560-2567 ซึ่งมีจำนวนข้อมูลอยู่เป็นจำนวนมาก ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เกี่ยวข้องกับการรวบรวม และทำความเข้าใจกับข้อมูลที่มีอยู่ โดยเริ่มต้นจากการรวบรวมข้อมูลที่เกี่ยวข้องทั้งหมด จากนั้นทำการตรวจสอบคุณภาพของข้อมูล รวมถึงการวิเคราะห์เบื้องต้น เช่น การหาค่าทางสถิติพื้นฐาน การวิเคราะห์ข้อมูลที่หายไป หรือข้อมูลที่ผิดปกติ สิ่งนี้จะช่วยให้เราเข้าใจลักษณะ และโครงสร้างของข้อมูลได้ชัดเจนมากขึ้น และช่วยให้สามารถเลือกวิธีการวิเคราะห์ที่เหมาะสมในขั้นตอนถัดไป

โดยข้อมูลที่ทางคณะจัดทำทำการวิเคราะห์นั้น เป็นข้อมูลการยืมหนังสือในห้องสมุด ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ในช่วงประจำปี 2560-2567 มีจำนวนข้อมูลทั้งหมด 224,597 รายการ ประกอบด้วย 8 แอดทริบิวท์ ได้แก่ ประเภทบุคลากร, หลักสูตร, คณะ, เลขประจำหนังสือ, เลขเรียกหนังสือ, ชื่อหนังสือ, วันที่ยืม และวันที่ทำการคืน ข้อมูลจะได้รับการสำรวจและตรวจสอบว่าเป็นไปตามที่ต้องการหรือไม่ เช่น ข้อมูลขาดหายหรือข้อมูลที่ไม่ถูกต้อง เพื่อให้มั่นใจได้ว่าข้อมูลพร้อมสำหรับการวิเคราะห์ต่อไป

1	A	B	C	D	E	F	G	H
2	รายงานสถิติเอกสาร							
3	เล่มละ 1							
4	สืบไป 1 หน้า 2560 ถึง 30 ธันวาคม 2567							
5	ระบุคณะ : 224597 items							
6								
7	ปีรวม	PROGRAM	OFFICE	BARCODE	CALLNO	TITLE	CHECKOUTDATE	RETURNDATE
8	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5088844	HF415.1263 v337 2563	1-3 3 มุม Facebook / 3 มุม fan.	18/11/2563	24/11/2563
9	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5077595	v 4778v 2559	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	18/04/2567	23/04/2567
10	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5088952	v 4330v 2563	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	18/04/2567	23/04/2567
11	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	50005470	TS185 v 4821 2563	Package design 3 มุม fan. 3 มุม fan. 3 มุม fan.	08/05/2566	08/05/2566
12	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	8001 702992	654.4023 v328v 2562	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	18/11/2566	18/11/2566
13	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	505107390	GA78 v 03 v314v 2558	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	09/07/2560
14	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051083072	GA78 v 03 v872v 2552	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	09/07/2560
15	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051087793	GA78 v 03 v238 2556	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	09/07/2560
16	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5073493	GA78 v 03 v439v 2558	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	09/07/2560
17	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051087458	HF415.1263 v336v 2556	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	23/07/2561	31/07/2561
18	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051092371	HF415.1263 v341v 2556	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	23/07/2561	31/07/2561
19	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051093916	HF415.1263 v356v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	23/07/2561	31/07/2561
20	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051070332	HF415.1263 v367 2556	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	23/07/2561	31/07/2561
21	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5079239	HF415.1263 v342 2560	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	23/07/2561	31/07/2561
22	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051013622	TS182 v311	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	12/02/2567	12/02/2567
23	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051091568	TS182 v499v 2553	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	12/02/2567	12/02/2567
24	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051093972	LS259 v34	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	02/07/2560
25	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051093558	LS259 v445v 2552	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	02/07/2560
26	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051090893	LS259 v721v 2553	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	02/07/2560
27	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051071236	LS259 v899v 2559	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	28/06/2560	02/07/2560
28	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051093970	TS182 v311	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	12/02/2567	12/02/2567
29	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051090817	T222 v78 v336 2552	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/04/2562	17/04/2562
30	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5073472	8F408 v449v 2557	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/04/2562	17/04/2562
31	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5050000917805	743 v 4446v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	09/10/2563	16/10/2563
32	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5050000023432	709 8993 v311v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	08/08/2560	08/08/2560
33	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5050000000313	720 8993 v481v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	08/08/2560	08/08/2560
34	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	50500000706171	743 v 4282v 2559	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	09/10/2563	16/10/2563
35	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	50500000712344	743 v 4734v 2561	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	09/10/2563	16/10/2563
36	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051013603	GA78 v 03 v36v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	30/04/2561	30/04/2561
37	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051070477	GA78 v 03 v178v 2556	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	30/04/2561	30/04/2561
38	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051071309	GA78 v 03 v25v 2558	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	30/04/2561	30/04/2561
39	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051031723	TS182 v 2540	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/11/2560	17/11/2560
40	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051008691	HF5681 v31 3855v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/11/2560	17/11/2560
41	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051009861	HF5681 v31 3855v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/11/2560	17/11/2560
42	ปี 2560	เทคโนโลยีสารสนเทศ	เทคโนโลยีสารสนเทศ	5051009867	HF5681 v31 3855v	การจัดการงาน / 3 มุม fan. 3 มุม fan. 3 มุม fan.	17/11/2560	17/11/2560

ภาพที่ 3.2 แสดงข้อมูลการยืม-คืนหนังสือจากห้องสมุด ประจำปี 2560 ถึง 2567

3.1.3 การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่สำคัญสำหรับการวิเคราะห์ข้อมูล โดยข้อมูลที่เก็บมาอาจมีปัญหาในหลายด้าน เช่น ข้อมูลขาดหาย ข้อมูลซ้ำซ้อน หรือข้อมูลที่ไม่ตรงกับความต้องการ ขั้นตอนนี้จะมุ่งเน้นไปที่การทำความสะอาดข้อมูล การแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสม และการเลือกฟีเจอร์หรือคุณลักษณะที่สำคัญในการสร้างโมเดล ตัวอย่างเช่น การเข้ารหัสค่าที่เป็นข้อความให้เป็นตัวเลข หรือการลบข้อมูลที่ไม่สมบูรณ์

3.1.3.1 การคัดเลือกข้อมูล (Data Selection) คือเลือกข้อมูลที่สำคัญและมีความเกี่ยวข้องกับปัญหาหรือคำถามที่ต้องการวิเคราะห์ คณะผู้จัดทำทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูล ข้อมูลการยืม-คืนหนังสือจากห้องสมุด ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ประจำปี 2560-2567 โดยตัดส่วนที่ไม่จำเป็นออก ให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ โดยผู้จัดทำจะเพิ่มแอตทริบิวต์ ID เข้าไป ในภาพรวมจำนวน 7 แอตทริบิวต์ ได้แก่ ID, ประเภทบุคลากร, หลักสูตร, คณะ, ชื่อหนังสือ, วันที่ยืม และวันที่ทำการคืน โดยใช้ข้อมูลในปี 2560-2567 จำนวน 224,597 รายการ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

ID	ประเภท	PROGRAM	OFFICE	TITLE	CHECKOUTDATE	RETURNDATE
1	นักศึกษา	วศ.บ.วิศวกรรมอิเล็กทรอนิกส์	คณะวิศวกรรมศาสตร์	1-2-3 คำถาม Facebook	18/11/2020	24/11/2020
2	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	เจาะลึกร่างกายมนุษย์ /	19/4/2024	23/4/2024
3	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	หนังสือในพิภพอวกาศ ลว	19/4/2024	23/4/2024
4	นักศึกษา	ออกแบบบรรจุภัณฑ์(4ปี)	คณะศิลปกรรมและสถาปัตยกรรมศาสตร์	Package design เพือก'	8/5/2023	8/5/2023
5	นักศึกษา	บวส.การบัญชี	คณะบริหารธุรกิจและศิลปศาสตร์	วิศวกรรมคุณภาพและกา	19/12/2023	19/12/2023
6	นักศึกษา	วศ.บ.วิศวกรรมอาหาร	คณะวิศวกรรมศาสตร์	คู่มือใช้งาน Access 201	26/6/2017	3/7/2017
7	นักศึกษา	วศ.บ.วิศวกรรมอาหาร	คณะวิศวกรรมศาสตร์	ระบบฐานข้อมูล = Data	26/6/2017	3/7/2017
8	นักศึกษา	วศ.บ.วิศวกรรมอาหาร	คณะวิศวกรรมศาสตร์	บริหารและจัดการฐานขั	26/6/2017	3/7/2017
9	นักศึกษา	วศ.บ.วิศวกรรมอาหาร	คณะวิศวกรรมศาสตร์	คู่มือใช้งาน Access 201	26/6/2017	3/7/2017
10	นักศึกษา	วศ.บ.วิศวกรรมอาหาร	คณะวิศวกรรมศาสตร์	เริ่มต้นทำเงินกับ amazon	23/7/2018	31/7/2018
11	นักศึกษา	ศส.บ.การท่องเที่ยวและการโรงแรม	คณะบริหารธุรกิจและศิลปศาสตร์	เริ่มต้นทำเงินกับ amazon	23/7/2018	31/7/2018
12	นักศึกษา	ศส.บ.การท่องเที่ยวและการโรงแรม	คณะบริหารธุรกิจและศิลปศาสตร์	เช็คร้านออนไลน์บน Fac	23/7/2018	31/7/2018
13	นักศึกษา	ศส.บ.การท่องเที่ยวและการโรงแรม	คณะบริหารธุรกิจและศิลปศาสตร์	การตลาด 2.1 : ประยุ	23/7/2018	31/7/2018
14	นักศึกษา	ศส.บ.การท่องเที่ยวและการโรงแรม	คณะบริหารธุรกิจและศิลปศาสตร์	E-commerce and Onli	23/7/2018	31/7/2018
15	นักศึกษา	ศส.บ.การท่องเที่ยวและการโรงแรม	คณะบริหารธุรกิจและศิลปศาสตร์	Digital marketing : con	23/7/2018	31/7/2018
16	นักศึกษา	วิศวกรรมเครื่องจักรกลเกษตร	คณะวิศวกรรมศาสตร์	การดำเนินงาน TPM	12/2/2024	12/2/2024
17	นักศึกษา	วิศวกรรมเครื่องจักรกลเกษตร	คณะวิศวกรรมศาสตร์	TPM สำหรับโรงงานแบบ	12/2/2024	12/2/2024
18	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	เทคนิคการเขียนรายงาน	25/9/2019	2/10/2019
19	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	ฉลาดคิด พิชิตแกล้ง A /	25/9/2019	2/10/2019
20	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	เคล็ดลับการเขียนรายงาน	25/9/2019	2/10/2019
21	นักศึกษา	บธ.บ.การจัดการ	คณะบริหารธุรกิจและศิลปศาสตร์	หลักการและเทคนิคการ	25/9/2019	2/10/2019
22	นักศึกษา	วศ.บ.วิศวกรรมไฟฟ้า	คณะวิศวกรรมศาสตร์	วิศวกรรมไฟฟ้าภาคปฏิบิ	17/4/2019	17/4/2019
23	นักศึกษา	วศ.บ.วิศวกรรมไฟฟ้า	คณะวิศวกรรมศาสตร์	เรียนรู้และใช้งาน PLC ล	17/4/2019	17/4/2019
24	นักศึกษา	วศ.บ.วิศวกรรมไฟฟ้า	คณะวิศวกรรมศาสตร์	รักษาความถี่ตัวนำแนว	17/4/2019	17/4/2019
25	นักศึกษา	ไม่มีข้อมูล		ออกแบบผลิตภัณฑ์จุด	9/10/2020	16/10/2020
26	นักศึกษา	ไม่มีข้อมูล		ศิลปกรรมผลงานจากร	8/9/2017	8/9/2017
27	นักศึกษา	ไม่มีข้อมูล		สถาปัตยกรรมไทย / แ	8/9/2017	8/9/2017
28	นักศึกษา	ไม่มีข้อมูล		ประวัติศาสตร์การออกแบบ	9/10/2020	16/10/2020
29	นักศึกษา	ไม่มีข้อมูล		การเขียนแบบและออกน	9/10/2020	16/10/2020
30	นักศึกษา	ไม่มีข้อมูล		คำาเขียนด้วยตนเอง รัช	30/4/2018	30/4/2018
31	นักศึกษา	ไม่มีข้อมูล		คู่มือเขียนโครงการข้อมู	30/4/2018	30/4/2018
32	นักศึกษา	ไม่มีข้อมูล		หลักการพื้นฐานของวิ	30/4/2018	30/4/2018

ภาพที่ 3.3 ข้อมูลการยืม-คืนหนังสือจากห้องสมุด
ประจำปี 2560-2567 ที่ถูกคัดเลือกมา

3.1.3.2 การกลั่นกรองข้อมูล (Data Cleaning) คือกระบวนการในการปรับปรุงคุณภาพของข้อมูลดิบที่ได้มา โดยทำการตรวจสอบ แก้ไข หรือลบข้อมูลที่ไม่ถูกต้อง ขาดหาย หรือมีความไม่สมบูรณ์ เพื่อให้ข้อมูลมีความถูกต้องและมีความพร้อมสำหรับการนำไปวิเคราะห์ต่อไป กระบวนการนี้มีความสำคัญมาก เพราะข้อมูลดิบที่มีข้อผิดพลาดสามารถทำให้ผลลัพธ์ที่ได้จากการวิเคราะห์หรือสร้างโมเดลมีความคลาดเคลื่อนได้ ทางคณะผู้จัดทำได้ดำเนินการ ดังนี้

1) คณะผู้จัดทำได้ดำเนินการตรวจสอบค่าที่ขาดหายไปภายในชุดข้อมูล และพบว่าจำนวน 8,193 รายการ ซึ่งอาจเกิดจากข้อผิดพลาดในการบันทึกหรือกระบวนการเก็บรวบรวมข้อมูลที่ไม่สมบูรณ์ การมีค่าที่ขาดหายไปอาจส่งผลกระทบต่อความถูกต้องและคุณภาพของข้อมูลโดยรวม ดังนั้น เพื่อให้ชุดข้อมูลมีความสมบูรณ์และพร้อมสำหรับการใช้งาน ได้

ดำเนินการจัดการค่าที่ขาดหายไปตามแนวทางที่เหมาะสม หลังจากการทำความสะอาดข้อมูล ในช่วงปี 2560-2567 ทำให้เหลือข้อมูลที่สามารถนำไปใช้ต่อได้จำนวน 216,404 รายการ

Row No.	ประเภท	PROGRAM	OFFICE	TITLE	CHECKOUT...	RETURNDA...
1	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	1-2-3 สำนัก Fa...	Nov 18, 2563	Nov 24, 2563
2	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	เจาะลึกงานบ...	Apr 19, 2567	Apr 23, 2567
3	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	กรณีศึกษา	Apr 19, 2567	Apr 23, 2567
4	นักศึกษา	เอกแบบบรรจุภัณฑ์	คณะศิลปกรรม	Package desi...	May 8, 2566	May 8, 2566
5	นักศึกษา	ปวช.ช่างอิเล็กทรอนิกส์	คณะบริหารงาน	โครงการสหภาพ	Dec 19, 2566	Dec 19, 2566
6	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	คู่มือใช้งาน Acces...	Jun 26, 2560	Jul 3, 2560
7	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	รายงานเรื่อง = C	Jun 26, 2560	Jul 3, 2560
8	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	บัญชีและจัดการ	Jun 26, 2560	Jul 3, 2560
9	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	คู่มือใช้งาน Acces...	Jun 26, 2560	Jul 3, 2560
10	นักศึกษา	ศ.บ.การต่อเครื่อง	คณะบริหารงาน	เริ่มทำเงินกับ si	Jul 23, 2561	Jul 31, 2561
11	นักศึกษา	ศ.บ.การต่อเครื่อง	คณะบริหารงาน	เปิดร้านออนไลน์	Jul 23, 2561	Jul 31, 2561
12	นักศึกษา	ศ.บ.การต่อเครื่อง	คณะบริหารงาน	การตลาด 2.1 : ประ	Jul 23, 2561	Jul 31, 2561
13	นักศึกษา	ศ.บ.การต่อเครื่อง	คณะบริหารงาน	E-commerce ...	Jul 23, 2561	Jul 31, 2561
14	นักศึกษา	ศ.บ.การต่อเครื่อง	คณะบริหารงาน	Digital marke...	Jul 23, 2561	Jul 31, 2561
15	นักศึกษา	บริหารงานเอกสาร	คณะบริหารงาน	การส่งเงินบริการ	Feb 12, 2567	Feb 12, 2567
16	นักศึกษา	บริหารงานเอกสาร	คณะบริหารงาน	TPM สำหรับโรง	Feb 12, 2567	Feb 12, 2567
17	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	เทคนิคการเขียน	Sep 25, 2562	Oct 2, 2562
18	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	ฉลาดคิด หรือเก่ง	Sep 25, 2562	Oct 2, 2562
19	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	เคล็ดลับการเขียน	Sep 25, 2562	Oct 2, 2562
20	นักศึกษา	น.บ.การจัดการ	คณะบริหารงาน	หลักการและเทคนิค	Sep 25, 2562	Oct 2, 2562
21	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	บริหารงานฝ่าย	Apr 17, 2562	Apr 17, 2562
22	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	เริ่มเขียนคู่มือ	Apr 17, 2562	Apr 17, 2562
23	นักศึกษา	ท.บ.บริหารงาน	คณะบริหารงาน	รายงานผลสัมฤทธิ์	Apr 17, 2562	Apr 17, 2562
24	นักศึกษา	ไม่มีข้อมูล	ไม่มีข้อมูล	เอกแบบเลิกทำ	Oct 9, 2563	Oct 16, 2563
25	นักศึกษา	ไม่มีข้อมูล	ไม่มีข้อมูล	คัมภีร์งาน	Sep 8, 2560	Sep 8, 2560

ExampleSet (216,404 examples, 0 special attributes, 6 regular attributes)

ภาพที่ 3.4 แสดงการจัดการค่าที่ขาดหายไป

2) นอกเหนือจากค่าที่ขาดหายไป คณะผู้จัดทำยังได้ทำการตรวจสอบข้อมูลที่ไม่มีการระบุค่า แต่ไม่ใช่ค่าว่าง (เช่น ข้อมูลที่มีเครื่องหมายแทนที่หรือมีการบันทึกที่ไม่สมบูรณ์) ซึ่งพบว่ามีจำนวน 655 รายการ ที่ขาดข้อมูลสำคัญ เพื่อรักษาคุณภาพของข้อมูล จึงดำเนินการลบข้อมูลดังกล่าวออกจากชุดข้อมูล ส่งผลให้หลังจากการทำความสะอาดข้อมูล จำนวนข้อมูลที่เหลืออยู่ลดลงเหลือ 215,749 รายการ

Open in Turbo Prep Auto Model Interactive Analysis Filter (215,749 / 215,749 examples):

Row No.	ประเภท	PROGRAM	OFFICE	TITLE	CHECKOUT...	RETURNDATE
215735	ลูกจ้างชั่วคราว (อ	ภาษาอังกฤษเหนือ	คณะบริหารธุรกิจแ	Introducing L...	Sep 2, 2567	Oct 23, 2567
215736	ลูกจ้างชั่วคราว (อ	ภาษาอังกฤษเหนือ	คณะบริหารธุรกิจแ	Hospitality in ...	Sep 9, 2567	Sep 13, 2567
215737	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	คู่มือการบริหารการ	Jun 26, 2567	Oct 27, 2567
215738	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	หลักการผลิตและก	Jun 26, 2567	Oct 27, 2567
215739	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
215740	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
215741	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การวางแผนและคว	Jun 26, 2567	Oct 27, 2567
215742	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
215743	พนักงานมหาวิทยาลัย	วิศวกรรมไฟฟ้า	คณะวิศวกรรมศาส	วิศวกรรมไฟฟ้าแรง	Aug 21, 2567	Oct 25, 2567
215744	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	Control syste...	Sep 17, 2567	Sep 30, 2567
215745	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	Control syste...	Sep 30, 2567	Oct 23, 2567
215746	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบอัตโนมัติ	Sep 17, 2567	Sep 30, 2567
215747	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบอัตโนมัติ	Sep 30, 2567	Oct 23, 2567
215748	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบงานคว	Sep 17, 2567	Sep 30, 2567
215749	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบงานคว	Sep 30, 2567	Oct 23, 2567

ExampleSet (215,749 examples,0 special attributes,6 regular attributes)

ภาพที่ 3.5 แสดงการจัดการข้อมูลที่ไม่มีการระบุค่าแต่ไม่ใช่ค่าว่าง

3) นอกเหนือจากการจัดการค่าที่ขาดหายไปและข้อมูลที่ไม่มีการระบุค่า คณะผู้จัดทำยังได้ตรวจสอบ และคัดกรองข้อมูลที่ไม่เกี่ยวข้องกับหนังสือ เช่น บริการ และบอร์ดเกมในห้องสมุด ซึ่งพบว่ามีจำนวน 9,518 รายการ ที่ไม่ตรงกับเกณฑ์การคัดเลือกข้อมูล ข้อมูลเหล่านี้ถูกลบออกเพื่อให้ชุดข้อมูลมีความถูกต้อง และสอดคล้องกับวัตถุประสงค์ในการจัดเก็บข้อมูลที่เกี่ยวข้องกับหนังสือโดยเฉพาะ หลังจากดำเนินการทำความสะอาดข้อมูลระหว่างปี 2560-2567 จากเดิมที่มีข้อมูลพร้อมใช้งาน 215,749 รายการ ขณะนี้เหลือข้อมูลที่ผ่านการคัดกรองและสามารถนำไปใช้ได้จำนวน 206,231 รายการ

Open in Turbo Prep Auto Model Interactive Analysis Filter (206,231 / 206,231 examples):

Row No.	att1	ประเภท	PROGRAM	OFFICE	TITLE	CHECKOUT...	RETURNDATE
206217	215734	ลูกจ้างชั่วคราว (อ	ภาษาอังกฤษเหนือ	คณะบริหารธุรกิจแ	Introducing L...	Sep 2, 2567	Oct 23, 2567
206218	215735	ลูกจ้างชั่วคราว (อ	ภาษาอังกฤษเหนือ	คณะบริหารธุรกิจแ	Hospitality in ...	Sep 9, 2567	Sep 13, 2567
206219	215736	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	คู่มือการบริหารการ	Jun 26, 2567	Oct 27, 2567
206220	215737	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	หลักการผลิตและก	Jun 26, 2567	Oct 27, 2567
206221	215738	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
206222	215739	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
206223	215740	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การวางแผนและคว	Jun 26, 2567	Oct 27, 2567
206224	215741	อาจารย์พนักงาน	การจัดการธุรกิจ	คณะบริหารธุรกิจแ	การบริหารการผลิต	Jun 26, 2567	Oct 27, 2567
206225	215742	พนักงานมหาวิทยาลัย	วิศวกรรมไฟฟ้า	คณะวิศวกรรมศาส	วิศวกรรมไฟฟ้าแรง	Aug 21, 2567	Oct 25, 2567
206226	215743	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	Control syste...	Sep 17, 2567	Sep 30, 2567
206227	215744	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	Control syste...	Sep 30, 2567	Oct 23, 2567
206228	215745	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบอัตโนมัติ	Sep 17, 2567	Sep 30, 2567
206229	215746	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบอัตโนมัติ	Sep 30, 2567	Oct 23, 2567
206230	215747	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบงานคว	Sep 17, 2567	Sep 30, 2567
206231	215748	อาจารย์พนักงาน	วิศวกรรมเครื่องกล	คณะวิศวกรรมศาส	พัฒนาระบบงานคว	Sep 30, 2567	Oct 23, 2567

ExampleSet (206,231 examples,0 special attributes,7 regular attributes)

ภาพที่ 3.6 แสดงการลบข้อมูลที่ไม่เกี่ยวข้องกับหนังสือ

4) ทำการจัดกลุ่มของ 3 คอลัมน์ ได้แก่ ประเภท, PROGRAM, และ OFFICE โดยกำหนดให้คอลัมน์ใดคอลัมน์หนึ่งมีข้อมูลที่ไม่เหมือนกัน จะสามารถจัดกลุ่มได้ 437 กลุ่ม

	A	B	C	D
358	357	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
359	358	อาจารย์อำนวยการฯ	ผ.บ.ภาคกลาง	ศูนย์วิทยุชุมชน
360	359	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
361	360	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
362	361	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
363	362	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
364	363	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
365	364	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
366	365	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
367	366	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
368	367	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
369	368	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
370	369	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
371	370	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
372	371	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
373	372	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
374	373	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
375	374	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
376	375	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
377	376	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
378	377	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
379	378	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
380	379	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
381	380	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
382	381	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
383	382	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
384	383	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
385	384	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
386	385	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
387	386	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
388	387	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
389	388	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
390	389	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
391	390	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
392	391	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
393	392	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
394	393	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
395	394	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
396	395	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
397	396	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
398	397	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
399	398	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
400	399	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
401	400	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
402	401	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน
403	402	อาจารย์อำนวยการฯ	วิทยุชุมชน	ศูนย์วิทยุชุมชน

ภาพที่ 3.7 แสดงการจัดกลุ่มของ 3 คอลัมน์

5) ทำการรวบรวมรายการหนังสือที่แต่ละกลุ่มได้ทำการยื่น โดยกำหนดให้หนังสือแต่ละเล่มที่นำมารวมกันภายในแต่ละกลุ่มต้องมีชื่อที่ไม่ซ้ำกัน จากนั้นทำการแปลงรูปแบบการแสดงผลของรายการหนังสือแต่ละเล่มที่ได้จากตารางข้อมูลในขั้นตอนที่ 4 ให้อยู่ในรูปแบบแนวนอน

The image shows a spreadsheet with columns labeled GUB through GUX. The data is organized into rows, with the first row containing a header and subsequent rows containing numerical values. The spreadsheet is viewed from a top-down perspective, showing the layout of the data as it appears on the screen.

ภาพที่ 3.8 แสดงการแปลงรายการหนังสือแต่ละเล่มจากตารางข้อมูล

6) นำรายการหนังสือที่ถูกแปลงจากขั้นตอนที่ 5 ไปจัดเรียงในตารางข้อมูลรายการยื่นหนังสือ เพื่อค้นหากลุ่มที่ยื่นหนังสือในแต่ละเล่ม

Transacton_ID	PROGRAM	OFFICE	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource
1	PROGRAM	OFFICE				
2	PROGRAM	OFFICE				
3	PROGRAM	OFFICE				
4	PROGRAM	OFFICE				
5	PROGRAM	OFFICE				
6	PROGRAM	OFFICE				
7	PROGRAM	OFFICE				
8	PROGRAM	OFFICE				
9	PROGRAM	OFFICE				
10	PROGRAM	OFFICE				
11	PROGRAM	OFFICE				
12	PROGRAM	OFFICE				
13	PROGRAM	OFFICE				
14	PROGRAM	OFFICE				
15	PROGRAM	OFFICE				
16	PROGRAM	OFFICE				
17	PROGRAM	OFFICE				
18	PROGRAM	OFFICE				
19	PROGRAM	OFFICE				
20	PROGRAM	OFFICE				
21	PROGRAM	OFFICE				
22	PROGRAM	OFFICE				
23	PROGRAM	OFFICE				
24	PROGRAM	OFFICE				
25	PROGRAM	OFFICE				
26	PROGRAM	OFFICE				
27	PROGRAM	OFFICE				
28	PROGRAM	OFFICE				
29	PROGRAM	OFFICE				
30	PROGRAM	OFFICE				
31	PROGRAM	OFFICE				
32	PROGRAM	OFFICE				
33	PROGRAM	OFFICE				
34	PROGRAM	OFFICE				
35	PROGRAM	OFFICE				
36	PROGRAM	OFFICE				
37	PROGRAM	OFFICE				
38	PROGRAM	OFFICE				
39	PROGRAM	OFFICE				
40	PROGRAM	OFFICE				
41	PROGRAM	OFFICE				
42	PROGRAM	OFFICE				
43	PROGRAM	OFFICE				
44	PROGRAM	OFFICE				
45	PROGRAM	OFFICE				

ภาพที่ 3.9 แสดงการจัดเรียงในตารางข้อมูลรายการยืมหนังสือ

7) ตรวจสอบรายการยืมหนังสือของแต่ละกลุ่มโดยใช้ข้อมูลจากขั้นตอนที่ 6 มาตรวจสอบ หากพบว่ากลุ่มนั้นได้ยืมหนังสือเล่มดังกล่าว ให้แทนค่าเป็น 1 หากกลุ่มดังกล่าวไม่ได้ยืมให้แทนค่าเป็น 0

Transacton_ID	PROGRAM	OFFICE	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource
1	PROGRAM	OFFICE				
2	PROGRAM	OFFICE				
3	PROGRAM	OFFICE				
4	PROGRAM	OFFICE				
5	PROGRAM	OFFICE				
6	PROGRAM	OFFICE				
7	PROGRAM	OFFICE				
8	PROGRAM	OFFICE				
9	PROGRAM	OFFICE				
10	PROGRAM	OFFICE				
11	PROGRAM	OFFICE				
12	PROGRAM	OFFICE				
13	PROGRAM	OFFICE				
14	PROGRAM	OFFICE				
15	PROGRAM	OFFICE				
16	PROGRAM	OFFICE				
17	PROGRAM	OFFICE				
18	PROGRAM	OFFICE				
19	PROGRAM	OFFICE				
20	PROGRAM	OFFICE				
21	PROGRAM	OFFICE				
22	PROGRAM	OFFICE				
23	PROGRAM	OFFICE				
24	PROGRAM	OFFICE				
25	PROGRAM	OFFICE				
26	PROGRAM	OFFICE				
27	PROGRAM	OFFICE				
28	PROGRAM	OFFICE				
29	PROGRAM	OFFICE				
30	PROGRAM	OFFICE				
31	PROGRAM	OFFICE				
32	PROGRAM	OFFICE				
33	PROGRAM	OFFICE				
34	PROGRAM	OFFICE				
35	PROGRAM	OFFICE				
36	PROGRAM	OFFICE				
37	PROGRAM	OFFICE				
38	PROGRAM	OFFICE				
39	PROGRAM	OFFICE				
40	PROGRAM	OFFICE				
41	PROGRAM	OFFICE				
42	PROGRAM	OFFICE				
43	PROGRAM	OFFICE				
44	PROGRAM	OFFICE				
45	PROGRAM	OFFICE				

ภาพที่ 3.10 แสดงการตรวจสอบรายการยืมหนังสือของแต่ละกลุ่ม

8) ทำการลบคอลัมน์ที่ไม่จำเป็น เช่น คอลัมน์ประเภท, PROGRAM, OFFICE โดยจะเหลือคอลัมน์ที่จะนำไปใช้ทำโมเดลคือ หมายเลขกลุ่มและหนังสือแต่ละเล่มที่ถูกยืมตามภาพที่ปรากฏ

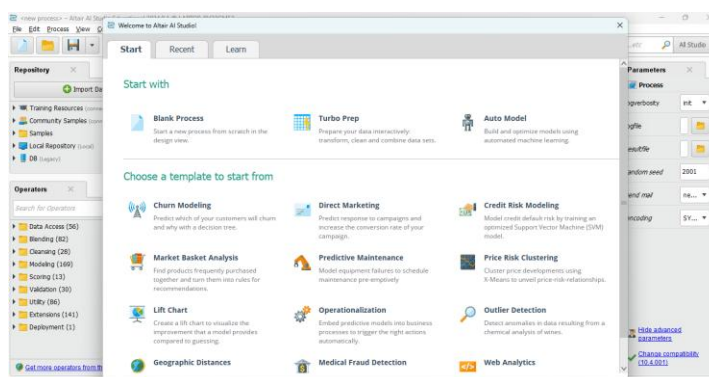
Transacton_ID	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource	Material of motor / unit Resource
200	200	0	0	0
161	300	0	0	0
162	200	0	0	0
163	300	0	0	0
164	300	0	0	0
165	300	0	0	0
166	200	0	0	0
167	300	0	0	0
168	200	0	0	0
169	200	0	0	0
170	200	0	0	0
171	200	0	0	0
172	271	0	0	0
173	321	0	0	0
174	214	0	0	0
175	373	0	0	0
176	373	0	0	0
177	373	0	0	0
178	373	0	0	0
179	274	0	0	0
180	274	0	0	0
181	290	0	0	0
182	301	0	0	0
183	382	0	0	0
184	382	0	0	0
185	382	0	0	0
186	382	0	0	0
187	382	0	0	0
188	382	0	0	0
189	382	0	0	0
190	382	0	0	0
191	382	0	0	0
192	382	0	0	0
193	382	0	0	0
194	382	0	0	0
195	382	0	0	0
196	382	0	0	0
197	382	0	0	0
198	382	0	0	0
199	382	0	0	0
200	382	0	0	0
201	382	0	0	0
202	382	0	0	0
203	382	0	0	0
204	382	0	0	0
205	382	0	0	0
206	382	0	0	0
207	382	0	0	0
208	382	0	0	0
209	382	0	0	0
210	382	0	0	0
211	382	0	0	0
212	382	0	0	0
213	382	0	0	0
214	382	0	0	0
215	382	0	0	0
216	382	0	0	0
217	382	0	0	0
218	382	0	0	0
219	382	0	0	0
220	382	0	0	0
221	382	0	0	0
222	382	0	0	0
223	382	0	0	0
224	382	0	0	0
225	382	0	0	0
226	382	0	0	0
227	382	0	0	0
228	382	0	0	0
229	382	0	0	0
230	382	0	0	0
231	382	0	0	0
232	382	0	0	0
233	382	0	0	0
234	382	0	0	0
235	382	0	0	0
236	382	0	0	0
237	382	0	0	0
238	382	0	0	0
239	382	0	0	0
240	382	0	0	0

ภาพที่ 3.11 แสดงรูปแบบข้อมูลที่นำไปใช้ทำโมเดล

3.1.4 การสร้างแบบจำลอง (Modeling) การเปรียบเทียบโมเดลเป็นขั้นตอนสำคัญในกระบวนการทำ Data Mining และการวิเคราะห์ข้อมูล เพื่อประเมินประสิทธิภาพของแต่ละอัลกอริทึม และเลือกวิธีการที่เหมาะสมกับลักษณะของข้อมูล ในการศึกษาครั้งนี้ คณะผู้จัดทำได้เลือกใช้อัลกอริทึม Apriori และ FP-Growth ในการค้นหากฎความสัมพันธ์ (Association Rules) จากข้อมูลที่มีการเตรียมไว้แล้ว โดยใช้โปรแกรม RapidMiner เป็นเครื่องมือในการตรวจสอบ และเปรียบเทียบอัลกอริทึมทั้งสอง โดยใช้วิธีการดังต่อไปนี้

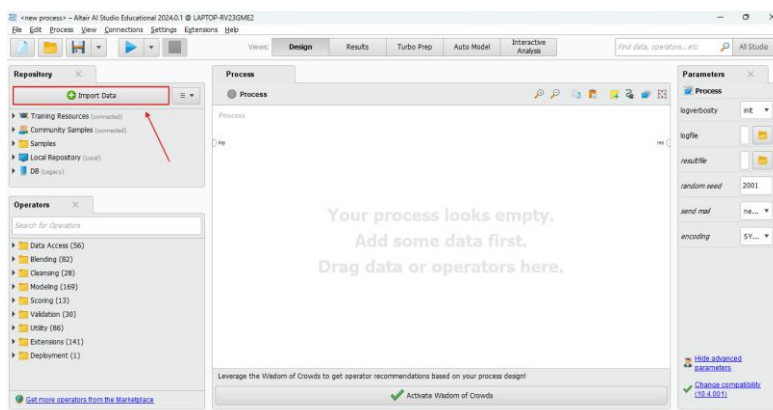
3.1.4.1 การนำข้อมูลเข้าโปรแกรม RapidMiner มีขั้นตอนดังต่อไปนี้

1) เริ่มต้นด้วยการเปิดโปรแกรม RapidMiner กดเลือก Blank Process เริ่มต้นกระบวนการทำงานใหม่ตั้งแต่เริ่มต้น แสดงดังภาพที่ 3.12



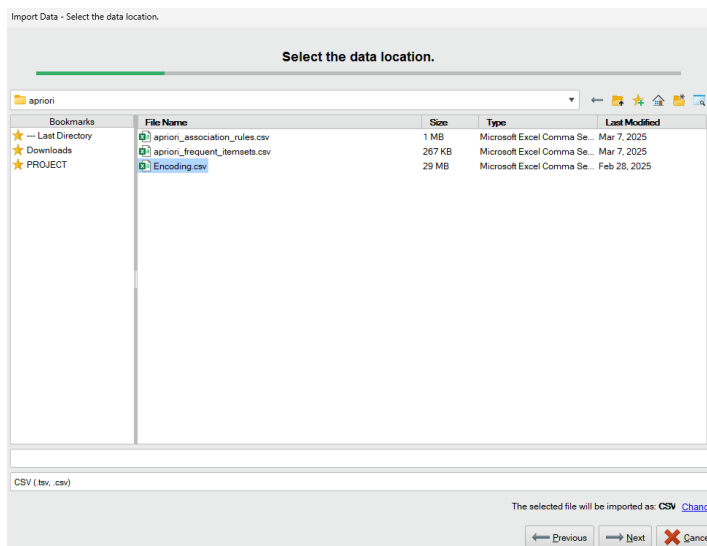
ภาพที่ 3.12 แสดงหน้าหลักของโปรแกรม RapidMiner

2) Import Dataset ที่เราทำการ Clean ข้อมูลแล้วเรียบร้อยแล้วในโปรแกรม โดยเลือก Import Data แสดงดังภาพที่ 3.13



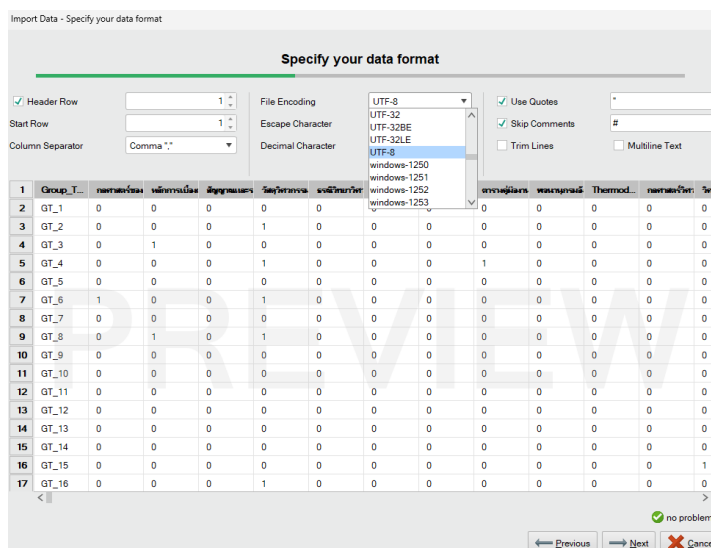
ภาพที่ 3.13 Import Data

3) เลือก Encoding.csv เพื่อนำข้อมูลที่ทำความสะอาดแล้วมา ใช้กับโมเดลที่ต้องการแสดงดังภาพที่ 3.14



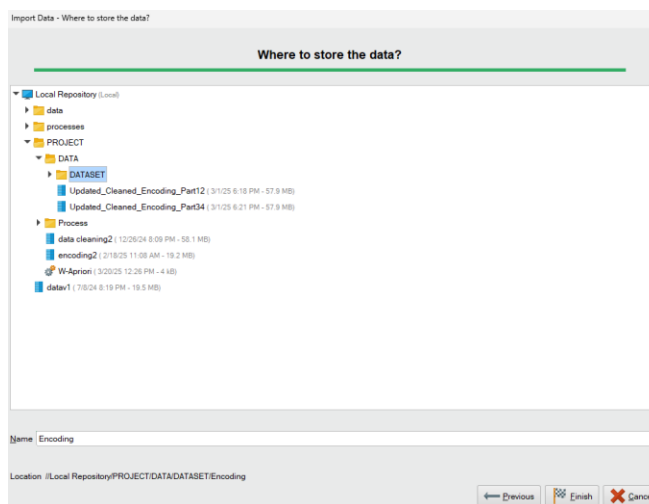
ภาพที่ 3.14 แสดงข้อมูล Encoding ที่จัดเก็บไว้ใน Folder Data

4) ตั้งค่าให้รองรับภาษาไทยโดยใช้ UTF-8 เพื่อแสดงผลตัวอักษรภาษาไทยได้อย่างถูกต้องครบถ้วน แสดงดังภาพที่ 3.15



ภาพที่ 3.15 แสดงการเลือก UTF-8

5) เลือก Folder Data ที่สร้างไว้ เพื่อนำข้อมูล Encoding.csv จัดเก็บไว้เพื่อความสะดวกในการใช้งาน แสดงดังภาพที่ 3.16



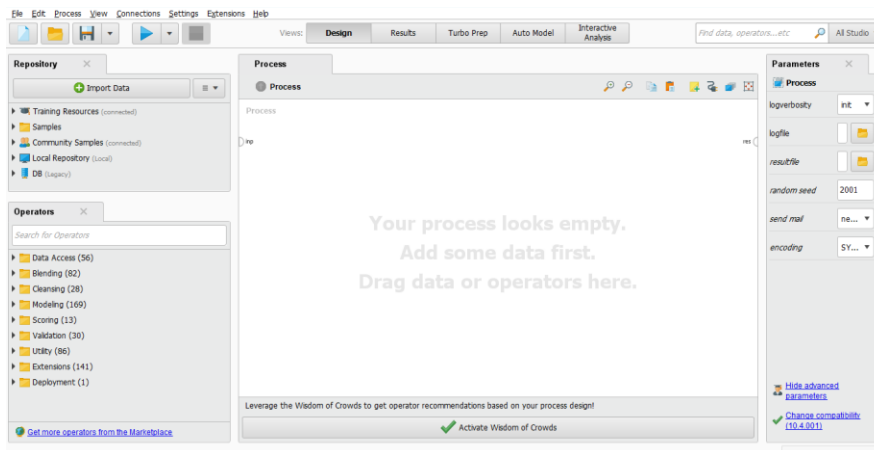
ภาพที่ 3.16 แสดงข้อมูล Encoding ที่จัดเก็บไว้ใน Folder Data

6) ข้อมูลที่ได้จากการนำเข้าข้อมูลจะมีทั้งหมด จำนวน 437 แถว 29,753 แอททริบิวต์ แสดงดังภาพที่ 3.17

File No.	Class_Type	nameofuser	ชื่อรายการสินค้า	ประเภทของสินค้า	Size/Volume = E	ชนิดสี/สีของสินค้า	ชนิด/ชนิดของสินค้า	ชนิด/ชนิดของสินค้า	ขนาด/ปริมาณสินค้า	ขนาดของสินค้า	Thermoflex	ขนาด/ปริมาณสินค้า	ขนาด/ปริมาณสินค้า	ขนาด/ปริมาณสินค้า
1	GT_1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	GT_2	0	0	0	1	0	0	0	0	0	0	0	0	0
3	GT_3	0	1	0	0	0	0	0	0	0	0	0	0	0
4	GT_4	0	0	0	1	0	0	0	1	0	0	0	0	0
5	GT_5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	GT_6	1	0	0	1	0	0	0	0	0	0	0	0	0
7	GT_7	0	0	0	0	0	0	0	0	0	0	0	0	1
8	GT_8	0	1	0	1	0	0	0	0	0	0	0	0	0
9	GT_9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	GT_10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	GT_11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	GT_12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	GT_13	0	0	0	0	0	0	0	0	0	0	0	0	0
14	GT_14	0	0	0	0	0	0	0	0	0	0	0	0	0
15	GT_15	0	0	0	0	0	0	0	0	0	0	0	1	0
16	GT_16	0	0	0	1	0	0	0	0	0	0	0	0	0
17	GT_17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	GT_18	0	0	0	0	0	0	0	0	0	0	0	0	0
19	GT_19	0	0	0	0	0	0	0	0	0	0	0	0	0
20	GT_20	0	0	0	0	0	0	0	0	0	0	0	0	0
21	GT_21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	GT_22	0	0	0	0	0	0	0	0	0	0	0	0	0
23	GT_23	0	0	0	0	0	0	0	0	0	0	0	0	0
24	GT_24	0	0	0	0	0	0	0	0	0	0	0	0	0
25	GT_25	0	0	0	0	0	0	0	0	0	0	0	0	0

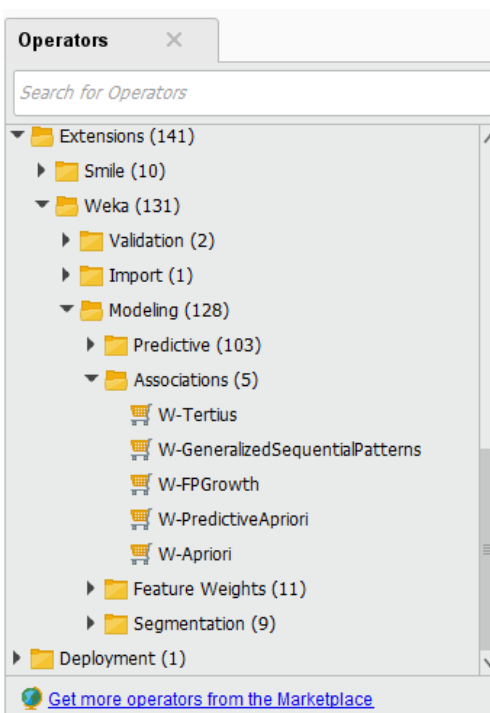
ภาพที่ 3.17 แสดงข้อมูลทั้งหมดที่นำเข้า

7) เลือก Blank Process เป็นพื้นที่ทำงานที่ว่างเปล่า แสดงดังภาพที่ 3.18



ภาพที่ 3.18 แสดงหน้า Blank Process

8) ลากโอเปอเรเตอร์จาก Operators Panel ลงมาและเชื่อมต่อ เพื่อสร้าง Process แสดงดังภาพที่ 3.19



ภาพที่ 3.19 แสดง Operator ต่าง ๆ

3.1.4.2 ขั้นตอนการดำเนินงานของโมเดลทั้ง 2 โมเดล

1) Apriori แสดงกระบวนการสร้างและทดสอบใน RapidMiner เป็นกระบวนการวิเคราะห์ข้อมูลที่ใช้สำหรับค้นหาความสัมพันธ์ระหว่างรายการข้อมูลที่เกิดขึ้นร่วมกัน โดยอาศัยหลักการของ Association Rule Mining ซึ่งสามารถนำไปใช้ในการวิเคราะห์และค้นหาความสัมพันธ์ของข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ

1.1 การดึงข้อมูล (Retrieve Dataset) กระบวนการนำข้อมูลจากแหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

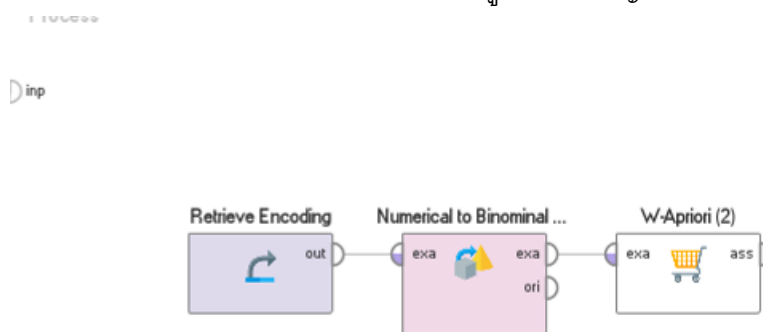
1.2 ใช้เครื่องมือ Numerical to Binominal เพื่อแปลงค่าข้อมูลจาก Numerical เป็น Binominal การแปลงข้อมูลในขั้นตอนนี้เป็นสิ่งจำเป็น เนื่องจากอัลกอริทึม Associate รองรับเฉพาะข้อมูลที่อยู่ในรูปแบบ Binominal (ค่า True/False หรือ 0/1) เท่านั้น

1.3 เชื่อมต่ออัลกอริทึม Apriori เข้ากับชุดข้อมูลที่ได้รับการแปลงให้อยู่ในรูปแบบที่เหมาะสม

1.4 กำหนดค่าความเชื่อมั่น (Confidence) เท่ากับ 0.01 ซึ่งใช้ในการคำนวณระดับความน่าเชื่อถือของกฎความสัมพันธ์ และตั้งค่าการสนับสนุนขั้นต่ำ (Minimum Support) เท่ากับ 0.02 เพื่อระบุเกณฑ์ขั้นต่ำของข้อมูลที่รูปแบบข้อมูลต้องเกิดขึ้นในชุดข้อมูลทั้งหมด

1.5 เริ่มกระบวนการวิเคราะห์โดยกด Start เพื่อให้ระบบทำการประมวลผลและสร้างกฎความสัมพันธ์จากข้อมูล

1.6 ประเมินผลลัพธ์ของโมเดล โดยพิจารณาจากระยะเวลาในการประมวลผล และประสิทธิภาพของโมเดลในการจัดการข้อมูลขนาดใหญ่



ภาพที่ 3.20 แสดงการวัดประสิทธิภาพของโมเดล Apriori

2) FP-Growth เป็นอัลกอริทึมที่ใช้สำหรับค้นหารูปแบบที่เกิดขึ้นบ่อย (Frequent Patterns) ภายในชุดข้อมูลขนาดใหญ่ โดยเป็นหนึ่งในวิธีการของ Association Rule Mining ซึ่งช่วยให้สามารถค้นหาความสัมพันธ์ระหว่างรายการข้อมูลที่เกิดขึ้นร่วมกันได้อย่างรวดเร็วและมีประสิทธิภาพ อัลกอริทึมนี้ใช้โครงสร้าง Frequent Pattern Tree (FP-Tree) เพื่อลดความจำเป็นในการสแกนข้อมูลซ้ำหลายรอบ ทำให้สามารถค้นหารูปแบบที่เกิดขึ้นบ่อยได้โดยไม่ต้องสร้างชุดข้อมูลย่อยจำนวนมาก

2.1 การดึงข้อมูล (Retrieve Dataset) กระบวนการนำข้อมูลจากแหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

2.2 ใช้เครื่องมือ Numerical to Binominal เพื่อแปลงค่าข้อมูลให้อยู่ในรูปแบบ Binominal (ค่า True/False หรือ 0/1) ซึ่งเหมาะสมกับการวิเคราะห์ความสัมพันธ์ของข้อมูล

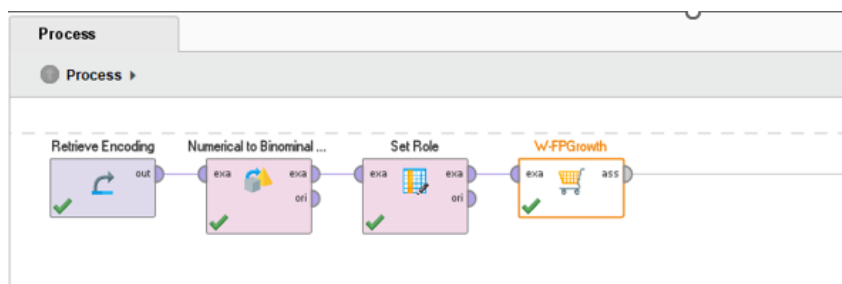
2.2 กำหนดค่าคอลัมน์ Group_Type ให้เป็น ID Attribute เพื่อให้ระบบสามารถใช้เป็นตัวระบุรายการที่เกิดขึ้นร่วมกัน

2.3 ใช้อัลกอริทึม FP-Growth ใน RapidMiner เพื่อทำการวิเคราะห์ข้อมูลที่ผ่านการเตรียมไว้ได้อย่างเหมาะสม

2.4 กำหนดค่าความเชื่อมั่น (Confidence) เพื่อใช้คำนวณระดับความน่าเชื่อถือของกฎความสัมพันธ์ และตั้งค่าการสนับสนุนขั้นต่ำ (Minimum Support) เพื่อใช้เป็นเกณฑ์เบื้องต้นในการระบุความถี่ของรูปแบบข้อมูลที่ปรากฏในชุดข้อมูลทั้งหมด

2.5 เริ่มกระบวนการวิเคราะห์โดยกด Start เพื่อเริ่มกระบวนการ และให้ระบบทำการสร้างโครงสร้าง FP-Tree เพื่อระบุความสัมพันธ์ของข้อมูล

2.6 ประเมินผลลัพธ์ของโมเดล โดยพิจารณาจาก ระยะเวลาในการประมวลผล และประสิทธิภาพของโมเดลในการจัดการข้อมูลขนาดใหญ่



ภาพที่ 3.21 แสดงการวัดประสิทธิภาพของโมเดล FP-Growth

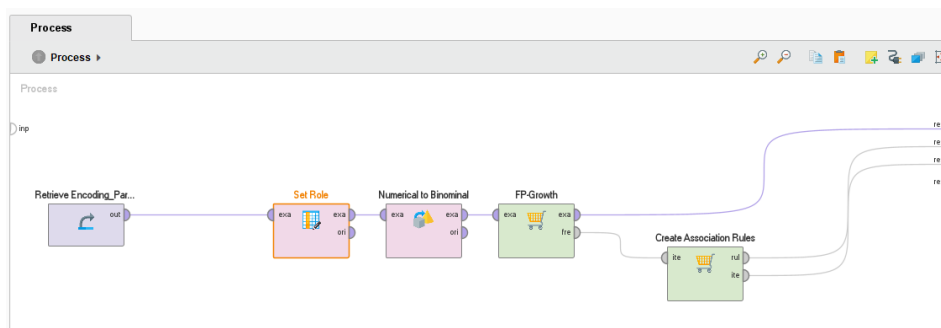
3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation) เป็นกระบวนการสำคัญในการประเมินคุณภาพของกฎความสัมพันธ์ที่สร้างขึ้นจากทั้งสองโมเดล เพื่อเลือกโมเดลที่ดีที่สุดสำหรับการวิเคราะห์และเปรียบเทียบผลลัพธ์ ผู้จัดทำได้ใช้โปรแกรม RapidMiner ในการวัดประสิทธิภาพของโมเดล โดยพิจารณาตัวชี้วัดหลัก ได้แก่ เวลาในการประมวลผล ซึ่งวัดระยะเวลาที่โปรแกรมใช้ในการประมวลผลข้อมูล และประสิทธิภาพในการจัดการข้อมูลขนาดใหญ่ ซึ่งประเมินว่าโมเดลสามารถวิเคราะห์ และประมวลผลข้อมูลจำนวนมากได้อย่างมีประสิทธิภาพโดยไม่ลดทอนความเร็วและความถูกต้อง ปัจจัยเหล่านี้ช่วยให้สามารถเลือกโมเดลที่เหมาะสมที่สุดสำหรับการวิเคราะห์ข้อมูลในครั้งนี้

ตารางที่ 3.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล

	Apriori	FP-Growth
เวลาในการประมวลผล		
ประสิทธิภาพในการจัดการข้อมูลขนาดใหญ่		

3.1.6 การนำโมเดลไปใช้งานจริง (Deployment) เป็นขั้นตอนสุดท้ายในกระบวนการวิเคราะห์ข้อมูล และการทำ Data Mining หลังจากที่โมเดลได้ถูกสร้างและประเมินประสิทธิภาพแล้ว โดยในโครงการนี้ คณะผู้จัดทำได้นำโมเดล FP-Growth ไปใช้เพื่อค้นหากฎความสัมพันธ์ในชุดข้อมูลที่เกี่ยวข้องกับหนังสือที่ถูกยืมจากห้องสมุด โดยได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลทั้งสอง และพบว่า FP-Growth ให้ผลลัพธ์ที่ดีที่สุดในการวิเคราะห์ความสัมพันธ์ของข้อมูล เนื่องจากอัลกอริธึมนี้ใช้โครงสร้างข้อมูลแบบ FP-Tree ซึ่งช่วยให้การค้นหากฎความสัมพันธ์มีประสิทธิภาพมากขึ้นเมื่อทำงานกับชุดข้อมูลขนาดใหญ่ โดยการดำเนินงานของโมเดล FP-Growth ใน RapidMiner มีขั้นตอนดังนี้

- 1) โหลดข้อมูล ที่ผ่านการทำความสะอาดแล้วจากไฟล์ CSV, Excel หรือฐานข้อมูล
- 2) ใช้ Set Role กำหนด ID ให้กับคอลัมน์ Transaction_ID เพื่อบ่งบอกหมายเลขของแต่ละกลุ่มรายการที่ถูกเชื่อมโยง
- 3) ใช้ Numerical To Binominal เพื่อแปลงค่าข้อมูลให้เป็นประเภท Binominal ตามข้อกำหนดของ FP-Growth
- 4) ใช้ FP-Growth เพื่อแสดง Frequent Itemsets ที่ได้จากการสร้าง FP-tree
- 5) ใช้ Create Association Rules เพื่อสร้างกฎความสัมพันธ์ของหนังสือแต่ละเล่ม
- 6) ประเมินผลลัพธ์โดยใช้ตัวชี้วัดประสิทธิภาพ ได้แก่ Support ความถี่ของรายการที่ปรากฏร่วมกัน และ Confidence ความน่าเชื่อถือของกฎความสัมพันธ์ที่ได้



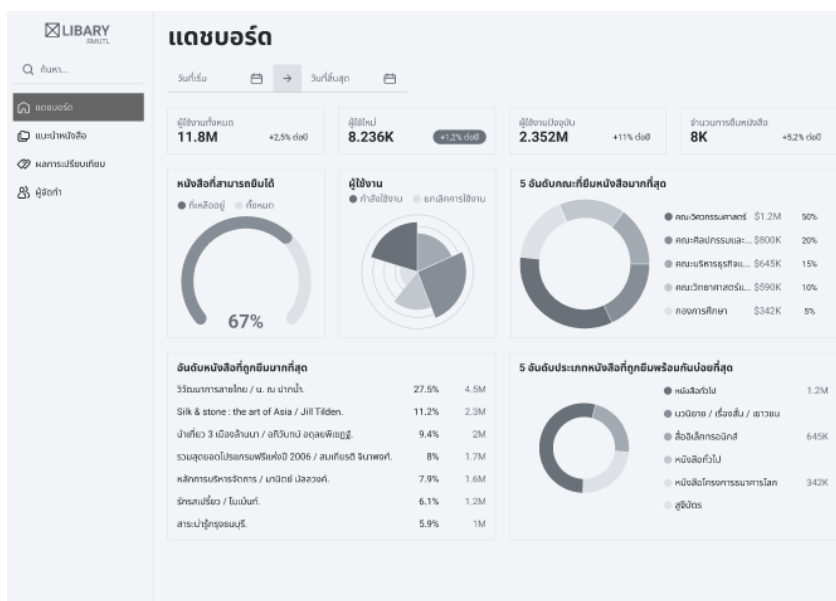
ภาพที่ 3.22 แสดงโมเดลของ FP-Growth

3.2 การวิเคราะห์และออกแบบระบบ

3.2.1 การวิเคราะห์และออกแบบฐานข้อมูล ระบบฐานข้อมูลสำหรับการยืม-คืนหนังสือ ได้รับการออกแบบให้สามารถจัดเก็บและบริหารข้อมูลได้อย่างมีประสิทธิภาพ รองรับปริมาณข้อมูลจำนวนมาก และสามารถนำไปวิเคราะห์เชิงลึกผ่านเครื่องมือภายนอกได้ โครงสร้างฐานข้อมูลประกอบด้วยข้อมูลสำคัญ เช่น รหัสรายการยืม ประเภทบุคลากร สาขา คณะ ชื่อหนังสือ วันที่ยืม และวันที่คืน ซึ่งช่วยให้สามารถติดตามและจัดการธุรกรรมการยืม-คืนได้อย่างเป็นระบบ สิทธิการใช้งานแบ่งออกเป็นสองระดับหลัก ได้แก่ ผู้ดูแลระบบ (Admin) ที่สามารถเพิ่ม แก้ไข และ

ลบข้อมูลหนังสือ ข้อมูลเชิงวิเคราะห์ อัปโหลดทฎความสัมพันธ์ และส่งออกข้อมูลเป็น CSV ในขณะที่ ผู้ใช้งานทั่วไป (User) สามารถล็อกอินเพื่อดูข้อมูลการยืมของตนเองผ่านแดชบอร์ด แต่ไม่มีสิทธิ์แก้ไขหรือลบข้อมูล ระบบยังถูกออกแบบให้สามารถรองรับการจัดการข้อมูลขนาดใหญ่ และบูรณาการเข้ากับเครื่องมือวิเคราะห์ เพื่อช่วยให้ห้องสมุดสามารถบริหารจัดการข้อมูลได้อย่างสะดวก รวดเร็ว และตอบสนองต่อความต้องการของผู้ใช้ได้อย่างมีประสิทธิภาพ

3.2.2 การออกแบบเว็บไซต์ เว็บไซต์สำหรับการเผยแพร่ข้อมูลการยืม-คืนหนังสือถูกออกแบบมาเพื่อให้ผู้ใช้สามารถเข้าถึงและตรวจสอบข้อมูลของตนเองได้อย่างสะดวก โดยใช้เทคโนโลยี HTML, CSS และ PHP ในการพัฒนาโครงสร้างและระบบการแสดงผล โครงสร้างของเว็บไซต์ถูกออกแบบโดยการใช้โปรแกรม Figma ซึ่งช่วยในการสร้าง Wireframe เพื่อวางรูปแบบหน้าจอและเมนูหลักให้ใช้งานง่าย หน้าหลักของเว็บไซต์จะแสดงข้อมูลสำคัญ เช่น เมนูนำทางไปยังหน้าแดชบอร์ด, รายการหนังสือที่ผู้ใช้ได้ยืม, สถานะการคืนหนังสือ, และประวัติการยืม-คืน ข้อมูลทั้งหมดจะถูกนำเสนอในรูปแบบที่เข้าใจง่าย เช่น ตาราง แผนภูมิ หรือสถิติสรุป ที่เกี่ยวข้องกับการยืมหนังสือในห้องสมุดของ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่ ดังนี้



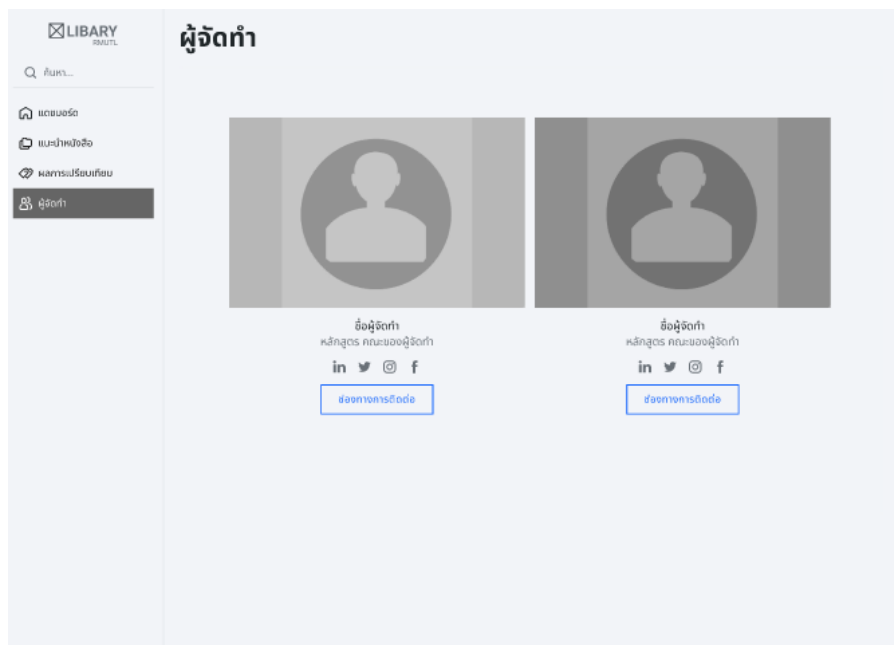
ภาพที่ 3.23 Wireframe หน้า Index



ภาพที่ 3.24 Wireframe หน้า Recommended books



ภาพที่ 3.25 Wireframe หน้า Comparison Results



ภาพที่ 3.26 Wireframe หน้า Contact Us