

บทที่ 3

ขั้นตอน และวิธีการดำเนินงาน

โครงการเรื่อง การวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้าด้วย Machine Learning และ RFM Analysis โดยเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) และการวิเคราะห์ข้อมูลแบบ CRISP-DM เพื่อทำการวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้า โมเดลที่ใช้ในการวิเคราะห์ประกอบไปด้วย 3 โมเดลหลัก ได้แก่ 1.โมเดล RFM Analysis 2.โมเดล K-mean Clustering 3.โมเดล Association Rule 4.โมเดล Decision Tree เมื่อกระบวนการวิเคราะห์ข้อมูลเสร็จสมบูรณ์ ขั้นตอนต่อไปคือการออกแบบเว็บไซต์ รวมถึงการจัดรูปแบบการแสดงผล และสรุปผลจากกระบวนการ ดำเนินงาน โดยมีขั้นตอนดังต่อไปนี้

3.1 ขั้นตอน และวิธีการวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 ER-Diagram

3.3 Context Diagram

3.4 Data Flow Diagram Level 0

3.5 การออกแบบเว็บไซต์

3.6 บทสรุป

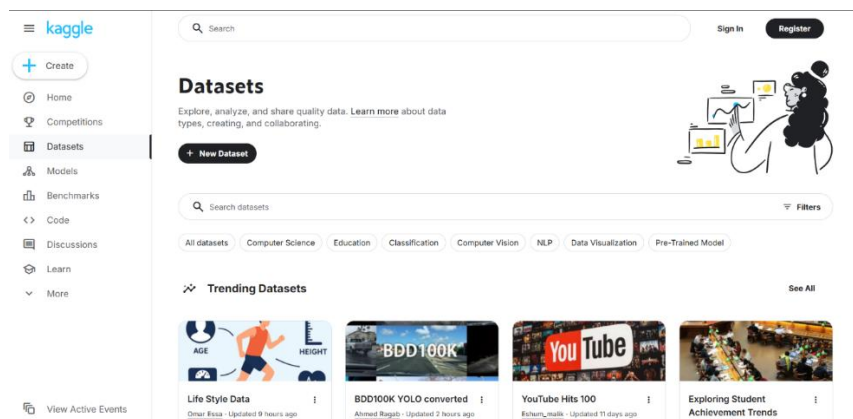
3.1 ขั้นตอน และวิธีการวิเคราะห์ข้อมูลด้วย CRISP-DM

การวิเคราะห์ข้อมูลด้วย CRISP-DM คือ กรอบการทำงานที่เป็นมาตรฐานสากล (Cross-Industry Standard Process for Data Mining) ซึ่งเป็นกระบวนการแบบเป็นขั้นตอน (มี 6 ขั้นตอนหลัก) ที่ใช้ในการวางแผน จัดการ และดำเนินโครงการชุดข้อมูลหรือโครงการวิทยาศาสตร์ข้อมูล เพื่อให้ได้ผลลัพธ์ที่มีคุณค่า โดยเน้นการทำความเข้าใจปัญหาทางธุรกิจ, ข้อมูล, การเตรียมข้อมูล, การสร้างโมเดล, การประเมินผลและการนำไปใช้งานจริง

3.1.1 การทำความเข้าใจธุรกิจ (Business Understanding) การทำความเข้าใจ ความเข้าใจ ธุรกิจปัญหาและวัตถุประสงค์ของโครงการจากมุมมองทางธุรกิจ จากนั้นแปลงปัญหาให้อยู่ในรูป ของโจทย์สำหรับการวิเคราะห์ข้อมูลและวางแผนการดำเนินงานเบื้องต้น

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูลตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกที่จะ ใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ ทำการรวบรวมข้อมูลพฤติกรรมอีคอมเมิร์ซการซื้อสินค้าอิเล็กทรอนิกส์จากห้างสรรพสินค้าแห่งหนึ่ง ในช่วงเดือนพฤศจิกายน ปี 2562 ที่ได้จากเว็บไซต์ชุดข้อมูล Kaggle <https://www.kaggle.com/datasets> ซึ่งเป็นเว็บไซต์รวบรวมชุดข้อมูลต่างๆ ที่เน้นเรื่อง Data Science, Machine Learning และ AI เพื่อที่ผู้วิเคราะห์จะสามารถทำความเข้าใจกับข้อมูลเหล่านั้น ผู้วิเคราะห์ได้ นำข้อมูลดังกล่าวมาทำการวิเคราะห์โดยผ่านเทคนิคการ Data mining



ภาพที่ 3.1 แสดงเว็บไซต์ Kaggle

ที่มา : <https://www.kaggle.com/datasets>

ซึ่งข้อมูลพฤติกรรมอีคอมเมิร์ซการซื้อสินค้าอิเล็กทรอนิกส์จากร้านค้าในช่วงเดือนพฤศจิกายน ปี 2562 มีจำนวนข้อมูลทั้งหมด 1,048,576 รายการ มีทั้งหมด 9 แอตทริบิวต์ ประกอบด้วย event_time (เวลาที่เหตุการณ์เกิดขึ้น), event_type (ประเภทกิจกรรม), product_id (รหัสของผลิตภัณฑ์), category_id (รหัสหมวดหมู่ของผลิตภัณฑ์), category_code (หมวดหมู่ของผลิตภัณฑ์), brand (ชื่อแบรนด์), price (ราคาของสินค้า), user_id (รหัสผู้ใช้), user_session (รหัสเซสชันของผู้ใช้ชั่วคราว)

event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session
2019-11-01 00:00:00 UTC	view	1003461	205301355561182655	electronics.smartphone	samsung	489.07	5.2E+08	4d3b303da-a5e4-439f-b7ab-ba5943f1d633
2019-11-01 00:00:00 UTC	view	5000088	205301356610086635	appliances.sewing_machine	janome	293.65	5.3E+08	8a58483-366c-4770-860e-ca7474742d83
2019-11-01 00:00:01 UTC	view	17320264	2053013553853497955	coated	ceded	28.31	5.62E+08	7542d2a7-9e40-477b-f9d2-46a8d8f97387
2019-11-01 00:00:01 UTC	view	3601930	2053013556381779293	appliances.kitchen.washer	lg	720.87	5.18E+08	3a2056a-7892-48cc-8020-2f5246a6e97f
2019-11-01 00:00:01 UTC	view	1004775	205301355561182655	electronics.smartphone	samsung	183.27	5.93E+08	313a2d81-6d8b-4608-84f6-cc7a879a72
2019-11-01 00:00:01 UTC	view	1306894	205301355820217191	computers.notebook	hp	360.09	5.27E+08	816a693-f5a6-4c0d-9e23-82a8c23d33c
2019-11-01 00:00:01 UTC	view	1304217	205301355820217191	computers.notebook	hp	514.56	5.14E+08	d918a2c-3804-4549-8d3c-6b5171873730
2019-11-01 00:00:02 UTC	view	1590063	2053013558190408249	roundell	roundell	30.86	5.19E+08	5d64f152-421c-4730-8c7f-85aa40c2588f
2019-11-01 00:00:02 UTC	view	12708937	205301355359896355	macshell	macshell	72.72	5.32E+08	0a899268-31ab-46de-896d-09a2d8d950624
2019-11-01 00:00:02 UTC	view	1004258	205301355561182655	electronics.smartphone	apple	732.07	5.33E+08	d2d3d2c6-631d-489e-9f05-09f340d8050d
2019-11-01 00:00:02 UTC	view	1120019	205301355978292471	appliances.kitchen.refrigerators	haier	437.33	5.19E+08	aab06025-b14e-46af-9f38-c04d19d8a73e
2019-11-01 00:00:03 UTC	view	2701517	2053013563911439225	appliances.kitchen.refrigerators	haier	155.11	5.18E+08	d8960296-247f-4044-9c91-bd5f38c6a98b
2019-11-01 00:00:03 UTC	view	16700260	2053013559901884381	furniture.kitchen.chair	haier	31.64	5.66E+08	173d872-1c87-4638-8207-ba8148b13c9f
2019-11-01 00:00:04 UTC	view	34600011	205008132018190480	electronics.notebook	acer	20.54	5.12E+08	4d82c2c7-e037-4632-a9d9-08d263d80091
2019-11-01 00:00:05 UTC	view	4600658	2053013563949936509	appliances.kitchen.dishwasher	samsung	411.83	5.27E+08	aab333da-29c3-4d50-8a41-8a2bc9256104
2019-11-01 00:00:06 UTC	view	24900193	2053013562183385881	electronics.notebook	acer	1.09	5.13E+08	6e03d15-f51a-4d6e-9404-cb586c35dafd
2019-11-01 00:00:07 UTC	view	27400066	2053013563391345499	electronics.notebook	acer	8.55	5.51E+08	3851271-5695-4a88-b0ba-499a36658f8f
2019-11-01 00:00:07 UTC	view	3700029	2053013553733466967	electronics.notebook	acer	22.68	5.21E+08	f5d0f0a-f32c-433a-993a-f62d44d4d936
2019-11-01 00:00:07 UTC	view	1004566	205301355561182655	electronics.smartphone	huawei	164.84	5.66E+08	52c27c6-b79e-4794-8f6f-bad76d3555a
2019-11-01 00:00:08 UTC	view	1307115	205301355820217191	computers.notebook	hp	411.59	5.14E+08	d8f18ac-3094-4549-8d3c-6b5171873730
2019-11-01 00:00:08 UTC	view	12719154	205301355359896355	electronics.notebook	yokohama	88.81	5.12E+08	f7e60aa-5103-4775-90ab-209f371700
2019-11-01 00:00:08 UTC	view	1004478	205301355561182655	electronics.smartphone	huawei	151.99	5.66E+08	aab10a8-5340-4533-ba74-b079d4a48433
2019-11-01 00:00:08 UTC	view	26022514	2053013562183385881	electronics.notebook	acer	1.07	5.64E+08	5d1a1d7d8-b414-ad4a-b17d-7d9380607d41

ภาพที่ 3.2 แสดงข้อมูลข้อมูลพฤติกรรมอีคอมเมิร์ซการซื้อสินค้าอิเล็กทรอนิกส์จากร้านค้าในช่วงเดือนพฤศจิกายน ปี 2562

3.1.3 การเตรียมข้อมูล (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมา และ เลือกไว้ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) โดยมีขั้นตอนดังนี้

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่ เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลพฤติกรรมอีคอมเมิร์ซการซื้อสินค้าอิเล็กทรอนิกส์จากร้านค้า ในช่วงเดือนพฤศจิกายน ปี 2562 โดยใช้ข้อมูลทั้งหมดในการวิเคราะห์ในภาพรวม มีทั้งหมด 9 แอตทริบิวต์

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้อง ออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางผู้วิเคราะห์ ข้อมูลได้ดำเนินการ ดังนี้

คณะผู้จัดทำได้ดำเนินการตรวจสอบค่าที่ขาดหายไปภายในชุดข้อมูล และพบว่า มีจำนวน 395,293 รายการ ซึ่งอาจเกิดจากข้อผิดพลาดในการบันทึกหรือ กระบวนการเก็บ รวบรวมข้อมูลที่ไม่สมบูรณ์ การมีค่าที่ขาดหายไปอาจส่งผลกระทบต่อความ ถูกต้องและคุณภาพ ของข้อมูลโดยรวม ดังนั้น เพื่อให้ชุดข้อมูลมีความสมบูรณ์และพร้อม สำหรับการใช้งาน ได้ดำเนินการจัดการค่าที่ขาดหายไปตามแนวทางที่เหมาะสม หลังจากการ ทำความสะอาดข้อมูล ทำให้เหลือข้อมูลที่สามารถนำไปใช้ต่อได้จำนวน 653,283 รายการ หลังจากนั้นทำการแยกคอลัมน์ event_time ซึ่งมีทั้งวันที่และเวลาที่รวมกันให้แยกออกมาเป็น วันที่และเวลาอย่างละคอลัมน์โดยใช้ชื่อคอลัมน์ใหม่คือ date คือวันที่ time คือเวลาและแยก คอลัมน์ category_code ที่บอกในส่วนของประเภทของสินค้าซึ่งแต่เดิมประเภทสินค้านั้นรวมหมวด กันในคอลัมน์เดียวจึงทำการแยกออกได้อีกเป็น 3 คอลัมน์ย่อยคือ category_code_other1, category_code_other2, category_code_other3 เพื่อให้สามารถนำข้อมูลไปวิเคราะห์ได้ละเอียด ยิ่งขึ้น รวมมีทั้งหมด 13 แอตทริบิวต์ ประกอบด้วย date (วันที่), time(เวลาที่เกิดเหตุการณ์), event_type (ประเภทกิจกรรม), product_id (รหัสของผลิตภัณฑ์), category_id (รหัสหมวดหมู่ ของผลิตภัณฑ์), category_code (หมวดหมู่ของผลิตภัณฑ์), category_code_other1 (หมวดหมู่ ของผลิตภัณฑ์ย่อย 1), category_code_other2 (หมวดหมู่ของผลิตภัณฑ์ย่อย 2), category_code_other3 (หมวดหมู่ของผลิตภัณฑ์ย่อย 3), brand (ชื่อแบรนด์), price (ราคาของ สินค้า), user_id (รหัสผู้ใช้), user_session (รหัสเซสชันของผู้ใช้ชั่วคราว) ดังภาพที่ 3.4

row_no	event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session
1	Nov 1, 2019 1.	view	1003461	2003013055	electronics.s	xiaomi	489.070	520088904	4d3030a0-af...
2	Nov 1, 2019 1.	view	5000088	2003013056	appliances.s	janome	293.650	530496790	8d2983-306...
3	Nov 1, 2019 1.	view	3601530	2003013063	appliances.s	lg	712.870	519085991	30b5dc4-78...
4	Nov 1, 2019 1.	view	1004775	2003013055	electronics.s	xiaomi	163.270	558856683	313a20f1-68...
5	Nov 1, 2019 1.	view	1306884	2003013056	computers.s	hp	360.990	520772885	816a09f1-6a...
6	Nov 1, 2019 1.	view	1004241	2003013056	computers.s	hp	514.900	514629257	08f16ac-306...
7	Nov 1, 2019 1.	view	1004568	2003013056	electronics.s	apple	712.070	532647354	4d3030a0-af...
8	Nov 1, 2019 1.	view	4600603	2003013063	appliances.s	samsung	411.530	526395647	aa3330a-29...
9	Nov 1, 2019 1.	view	1004566	2003013056	electronics.s	huawei	154.640	56250988	52d276a-07...
10	Nov 1, 2019 1.	view	1307115	2003013056	computers.s	hp	411.590	514629257	08f16ac-306...
11	Nov 1, 2019 1.	view	1004708	2003013055	electronics.s	huawei	151.990	566143627	aa3330a-29...
12	Nov 1, 2019 1.	view	1005115	2003013055	electronics.s	apple	468.470	562078884	0664462-63...
13	Nov 1, 2019 1.	view	1307350	2003013056	computers.s	acer	514.790	513741043	6279f70-346...
14	Nov 1, 2019 1.	view	1307340	2003013056	computers.s	acer	360.340	520772885	816a09f1-6a...
15	Nov 1, 2019 1.	view	16700826	2003013059	furniture.kit	niko	47.770	56625282	173d7f2-1d...
16	Nov 1, 2019 1.	view	4804194	2003013054	electronics.s	acer	69.240	565731881	56b70a7-ca...
17	Nov 1, 2019 1.	view	3701016	2003013065	appliances.s	tefal	108.090	566278257	5cafe7a-b04...
18	Nov 1, 2019 1.	view	1002544	2003013055	electronics.s	apple	458.280	548256216	6d30d54-cd...
19	Nov 1, 2019 1.	cat	1005014	2003013055	electronics.s	samsung	503.090	533226659	69d20a2-29...
20	Nov 1, 2019 1.	view	3801174	2003013062	appliances.s	casio	32.860	513200477	742ba02-72...
21	Nov 1, 2019 1.	view	4000903	2003013063	appliances.s	bosch	684.440	526995647	aa3330a-29...

ภาพที่ 3.3 แสดงการจัดการค่าที่ขาดหายไป

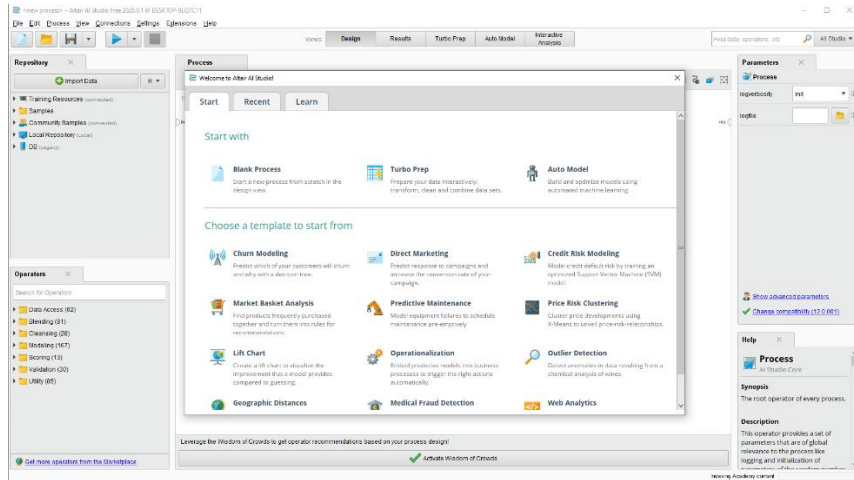
date	time	event_type	product_id	category_id	category_code	category_code_other1	category_code_other2	category_code_other3	brand	price	user_id	user_session
1/11/2019	0:00:00	view	1003461	2,00301E+18	electronics	smartphone	-	-	xiaomi	489.07	520088904	4d3030a0-af161633
1/11/2019	0:00:00	view	5000088	2,00301E+18	appliances	sewing_machine	-	-	janome	293.65	530496790	8d2983-306-ca71741283
1/11/2019	0:00:01	view	3601530	2,00301E+18	appliances	kitchen	-	-	lg	712.87	519085991	30b5dc4-78bc-8020-1746d6e07f
1/11/2019	0:00:01	view	1004775	2,00301E+18	electronics	smartphone	-	-	xiaomi	163.27	558856683	313a20f1-68ba-4604-9416-ca72a879e0f2
1/11/2019	0:00:01	view	1306421	2,00301E+18	computers	notebook	-	-	hp	360.99	520772885	816a09f1-6a0c-9b23-836a0c23d33c
1/11/2019	0:00:01	view	1306421	2,00301E+18	computers	notebook	-	-	hp	514.56	514629257	08f16ac-3064-4549-8a0c-6d5171877376
1/11/2019	0:00:02	view	1004258	2,00301E+18	electronics	smartphone	-	-	apple	712.07	532647354	d2c2d26-c31d-480e-9f0c-8d308050a0
1/11/2019	0:00:05	view	4600603	2,00301E+18	appliances	kitchen	-	-	huawei	411.83	526395647	aa3330a-29bc-29c1-4659-8a0c-18a2b292501a
1/11/2019	0:00:07	view	1004566	2,00301E+18	electronics	smartphone	-	-	huawei	164.84	566250908	52c2276c-b79e-4794-86ff-ba3c7d37525a
1/11/2019	0:00:08	view	1307115	2,00301E+18	computers	notebook	-	-	hp	411.59	514629257	08f16ac-306a-4509-8a0c-6d511877376
1/11/2019	0:00:08	view	1004708	2,00301E+18	electronics	smartphone	-	-	huawei	151.99	566143627	aa3330a-29bc-29c1-4659-8a0c-18a2b292501a
1/11/2019	0:00:10	view	1005115	2,00301E+18	electronics	smartphone	-	-	apple	468.47	562078884	0664462-63bc-4131-8a0c-b794d07995c3
1/11/2019	0:00:10	view	1307350	2,00301E+18	computers	notebook	-	-	acer	514.79	513741043	6279f70-346a-41f1-9e38-3a7f46b7276
1/11/2019	0:00:11	view	1307340	2,00301E+18	computers	notebook	-	-	acer	360.34	520772885	816a09f1-6a0c-9b23-836a0c23d33c
1/11/2019	0:00:12	view	16700826	2,00301E+18	furniture	kitchen	-	-	niko	47.77	56625282	173d7f2-1d87-4e38-8207-8a8148b13c9f
1/11/2019	0:00:13	view	4804194	2,00301E+18	electronics	audio	headphone	-	acer	69.24	565731881	56b70a7-ca04-4705-8998-33c07607e16
1/11/2019	0:00:13	view	3701016	2,00301E+18	appliances	environment	vacuum	-	tefal	108.09	566278257	5cafe7a-b04e-7a-3c4e-4b47-af0a-ab300c43506c
1/11/2019	0:00:14	view	1002544	2,00301E+18	electronics	smartphone	-	-	apple	458.28	548256216	6d30d54-cd09-464c-89c5-48f2c318f1
1/11/2019	0:00:14	cat	1005014	2,00301E+18	electronics	smartphone	-	-	samsung	503.09	533226659	69d20a2-29bc-4e4b-8296-0af0726a22a
1/11/2019	0:00:15	view	3901174	2,00301E+18	appliances	environment	water_heater	-	casio	32.86	513200477	742ba02-727b-4d15-9223-948ba8a0173
1/11/2019	0:00:15	view	4600603	2,00301E+18	appliances	kitchen	dishwasher	-	bosch	684.44	526995647	aa3330a-29bc-29c1-4659-8a0c-18a2b292501a
1/11/2019	0:00:17	view	1005181	2,00301E+18	electronics	smartphone	-	-	xiaomi	211.92	513251129	6a67a79b-1930-4e20-976c-8b4163aa11d
1/11/2019	0:00:17	view	21460629	2,00301E+18	electronics	chocs	-	-	tesco	957.9	51318532	4c14672e-2620-450a-9204-046356a206a
1/11/2019	0:00:17	view	1800729	2,00301E+18	electronics	video	tv	-	samsung	289.33	515782889	f2081e0-1a6c-48f1-9b40-4c5e38f9a6c3
1/11/2019	0:00:17	view	4200021	2,146666E+18	apparel	shirt	-	-	goodoot	8.73	566175330	6080144-699d-4911-836e-16a4d864e2d5
1/11/2019	0:00:18	view	1003317	2,00301E+18	electronics	smartphone	-	-	apple	928.18	532647354	d2c2d26-c31d-480e-9f0c-8d308050a0
1/11/2019	0:00:18	view	16700826	2,00301E+18	furniture	kitchen	chair	-	niko	47.77	56625282	173d7f2-1d87-4e38-8207-8a8148b13c9f
1/11/2019	0:00:20	view	1300795	2,00301E+18	computers	notebook	-	-	lenovo	385.85	520772885	816a09f1-6a0c-9b23-836a0c23d33c
1/11/2019	0:00:23	view	1004856	2,00301E+18	electronics	smartphone	-	-	samsung	128.42	436701163	50da79b-1b0c-420b-b0d0-8b5d2a321e
1/11/2019	0:00:23	view	12300928	2,00301E+18	construction	tools	drill	-	de Walt	98.51	517081124	6d1a99b3-3c06-4e1c-9f8e-8c797f5274dc

ภาพที่ 3.4 แสดงการแยกคอดัชนีที่มีออกมาเป็นคอดัชนีย่อยๆ

3.1.4 การสร้างแบบจำลอง (Modeling) การเปรียบเทียบโมเดลเป็นขั้นตอนสำคัญใน กระบวนการทำ Data Mining และการวิเคราะห์ข้อมูล เพื่อประเมินประสิทธิภาพของแต่ละ อัลกอริทึม และเลือกวิธีการที่เหมาะสมกับลักษณะของข้อมูล ใน การศึกษานี้ คณะผู้จัดทำได้ เลือกใช้อัลกอริทึม FP-Growth ในการค้นหากฎความสัมพันธ์ (Association Rules) และใช้อัลกอริทึม RFM Analysis และ K-mean Clustering ในการค้นหาการ จับกลุ่ม (Clustering) ประเภทการซื้อซ้ำของลูกค้าจาก รวมไปถึงอัลกอริทึม Decision Tree ใช้ สำหรับการจำแนกประเภท (Classification) ข้อมูลที่มีการเตรียมไว้แล้ว โดยใช้โปรแกรม RapidMiner เป็นเครื่องมือในการตรวจสอบ โดยใช้วิธีการดังต่อไปนี้

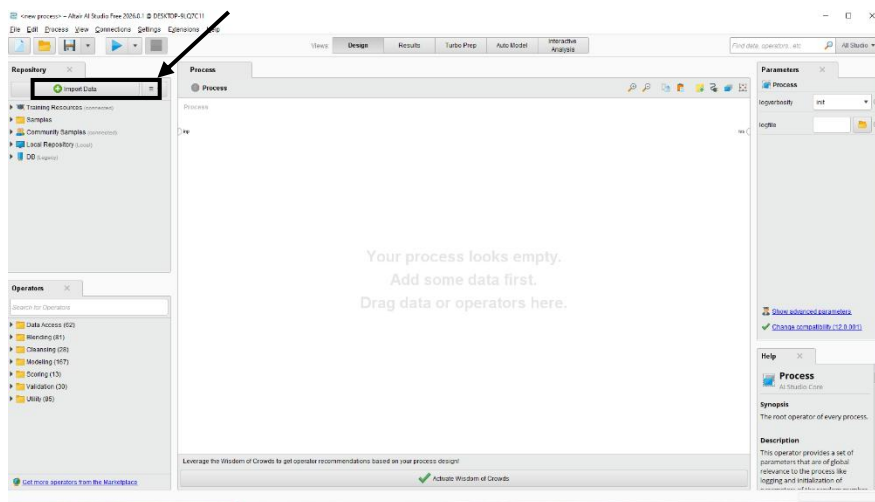
3.1.4.1 การนำข้อมูลเข้าโปรแกรม RapidMiner มีขั้นตอนดังต่อไปนี้

- 1) เริ่มต้นด้วยการเปิดโปรแกรม RapidMiner กดเลือก Blank Process เริ่มต้นกระบวนการทำงานใหม่ตั้งแต่เริ่มต้น แสดงดังภาพที่ 3.5



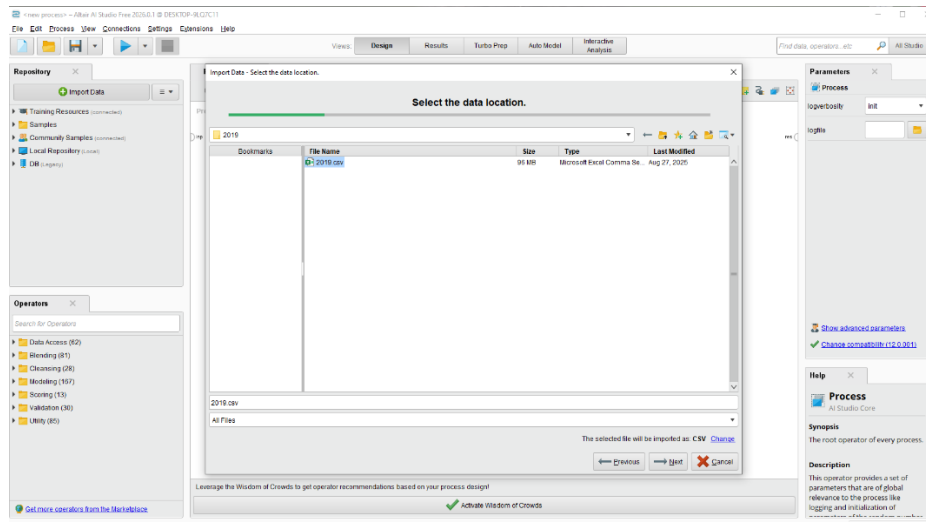
ภาพที่ 3.5 แสดงหน้าหลักของโปรแกรม RapidMiner

2) Import Dataset ที่เราทำการ Clean ข้อมูลแล้วเรียบร้อยแล้ว
มาใส่ ใน โปรแกรม โดยเลือก Import Data แสดงดังภาพที่ 3.6



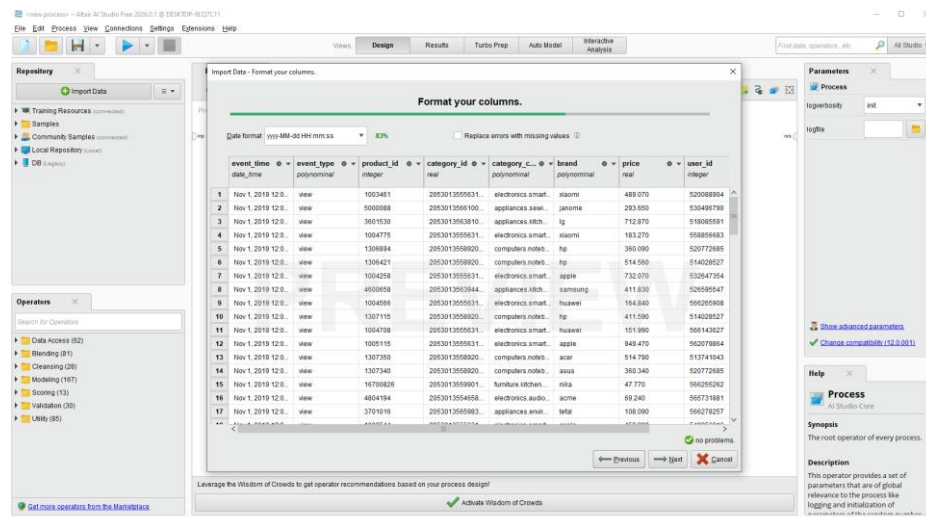
ภาพที่ 3.6 แสดงการนำเข้าข้อมูล

3) เลือก 2019.csv เพื่อนำข้อมูลที่ทำความสะอาดแล้วมา
ใช้กับโมเดล ที่ต้องการแสดงดังภาพที่ 3.7



ภาพที่ 3.7 แสดงข้อมูลที่จัดเก็บไว้ใน Folder Data

4) เลือก Folder Data ที่สร้างไว้ เพื่อนำข้อมูล Encoding.csv จัดเก็บไว้เพื่อ ความสะดวกในการใช้งาน แสดงดังภาพที่ 3.8



ภาพที่ 3.8 แสดงรายละเอียดข้อมูลที่นำเข้ามา

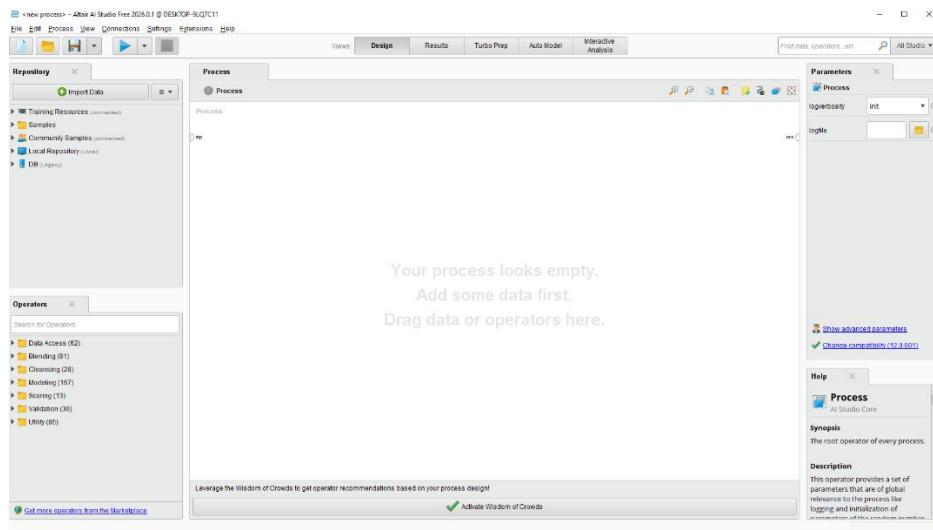
5) ข้อมูลที่ได้จากการนำเข้ามาข้อมูลจะมีทั้งหมด จำนวน 653,283 แถว 9 แอททริบิวต์ แสดงดังภาพที่ 3.9

Row No.	event_time	event_type	product_id	category_id	category_cd	brand	price	user_id	user_session
1	Nov 1, 2019 1..	view	1003461	2053013555..	electronics.s.	xiaomi	489.070	520688904	4c2b304a-ef..
2	Nov 1, 2019 1..	view	5000988	2053013556..	appliances.s.	panoma	293.959	530489790	6d938b3-366..
3	Nov 1, 2019 1..	view	3019530	2053013553..	appliances.h.	lg	712.870	518085591	30b58c4-78..
4	Nov 1, 2019 1..	view	1004770	2053013555..	electronics.s.	xiaomi	183.270	558855683	3136291f-68..
5	Nov 1, 2019 1..	view	1306894	2053013558..	computers.n.	hp	360.090	520772985	811a493-45a..
6	Nov 1, 2019 1..	view	1306421	2053013558..	computers.n.	hp	514.560	514228527	d8f18ac-369..
7	Nov 1, 2019 1..	view	1004258	2053013555..	electronics.s.	apple	732.070	532847354	d2d3d26-63..
8	Nov 1, 2019 1..	view	4600958	2053013553..	appliances.h.	samsung	413.830	529095547	aab33a9-29..
9	Nov 1, 2019 1..	view	1004566	2053013555..	electronics.s.	huawei	164.840	566209098	52c227c-87..
10	Nov 1, 2019 1..	view	1307110	2053013558..	computers.n.	hp	413.590	514228527	d8f18ac-369..
11	Nov 1, 2019 1..	view	1004768	2053013555..	electronics.s.	huawei	151.960	568142627	aab10d3-5c..
12	Nov 1, 2019 1..	view	1005110	2053013555..	electronics.s.	apple	640.070	602070984	06a4a4d-83..
13	Nov 1, 2019 1..	view	1307260	2053013556..	computers.n.	acer	514.730	613141543	8207072-34e..
14	Nov 1, 2019 1..	view	1307340	2053013558..	computers.n.	asus	360.340	520772985	811a493-45a..
15	Nov 1, 2019 1..	view	1670626	2053013558..	furniture.s.	ikea	47.770	646245262	173a752-14..
16	Nov 1, 2019 1..	view	4804184	2053013554..	electronics.s.	sony	69.240	665731881	516d47a-ca..
17	Nov 1, 2019 1..	view	3701016	2053013556..	appliances.s.	mtl	108.090	646278257	5cab674-30a..
18	Nov 1, 2019 1..	view	1002544	2053013555..	electronics.s.	apple	458.280	548258216	0c0d044-ca..
19	Nov 1, 2019 1..	cat	1005014	2053013555..	electronics.s.	samsung	503.990	533216659	69d3043-2b..
20	Nov 1, 2019 1..	view	3001174	2053013552..	appliances.s.	casio	32.660	513203477	742ab02-72..
21	Nov 1, 2019 1..	view	4600953	2053013553..	appliances.h.	toshiba	664.440	529095547	aab33a9-29..

ภาพที่ 3.9 แสดงข้อมูลทั้งหมดที่นำเข้า

7) เลือก Blank Process เป็นพื้นที่ทำงานที่ว่างเปล่า แสดง

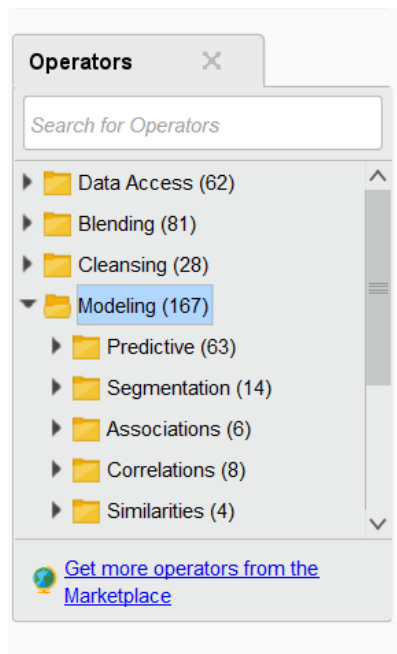
ดังภาพที่ 3.10



ภาพที่ 3.10 แสดงหน้า Blank Process

8) ลากไอคอนเรเตอร์จาก Operators Panel ลงมาและ

เชื่อมต่อ เพื่อสร้าง Process แสดงดังภาพที่ 3.11



ภาพที่ 3.11 แสดง Operator ต่าง ๆ

3.1.4.2 ขั้นตอนการดำเนินงานของโมเดล

1) RFM (Recency, Frequency, Monetary) เป็นเทคนิคที่ใช้วิเคราะห์พฤติกรรมลูกค้าโดยอ้างอิงจากข้อมูลธุรกรรม ซึ่งแบ่งเป็น 3 มิติ ได้แก่ **Recency** คือ ระยะเวลาซื้อครั้งล่าสุด, **Frequency** คือ ความถี่ในการซื้อ และ **Monetary** คือ มูลค่าการใช้จ่ายรวม โดยการนำทั้ง 3 เกณฑ์มาคำนวณร่วมกันจะช่วยจำแนกลูกค้าออกเป็นกลุ่มต่าง ๆ

1.1 การดึงข้อมูล (Retrieve Dataset)

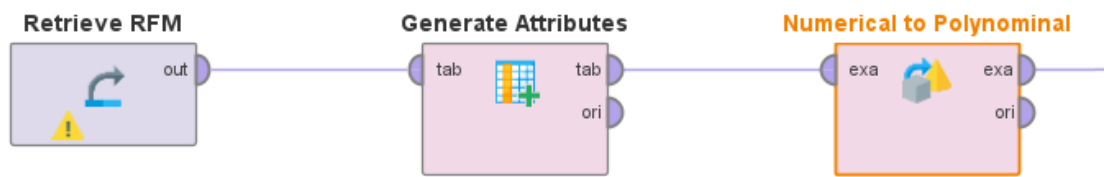
กระบวนการนำข้อมูลจาก แหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

1.2 ใช้เครื่องมือ Generate Attributes เพื่อสร้างตัวแปรใหม่ (Attributes) ขึ้นจากการคำนวณหรือการประมวลผลด้วยสูตรที่กำหนด

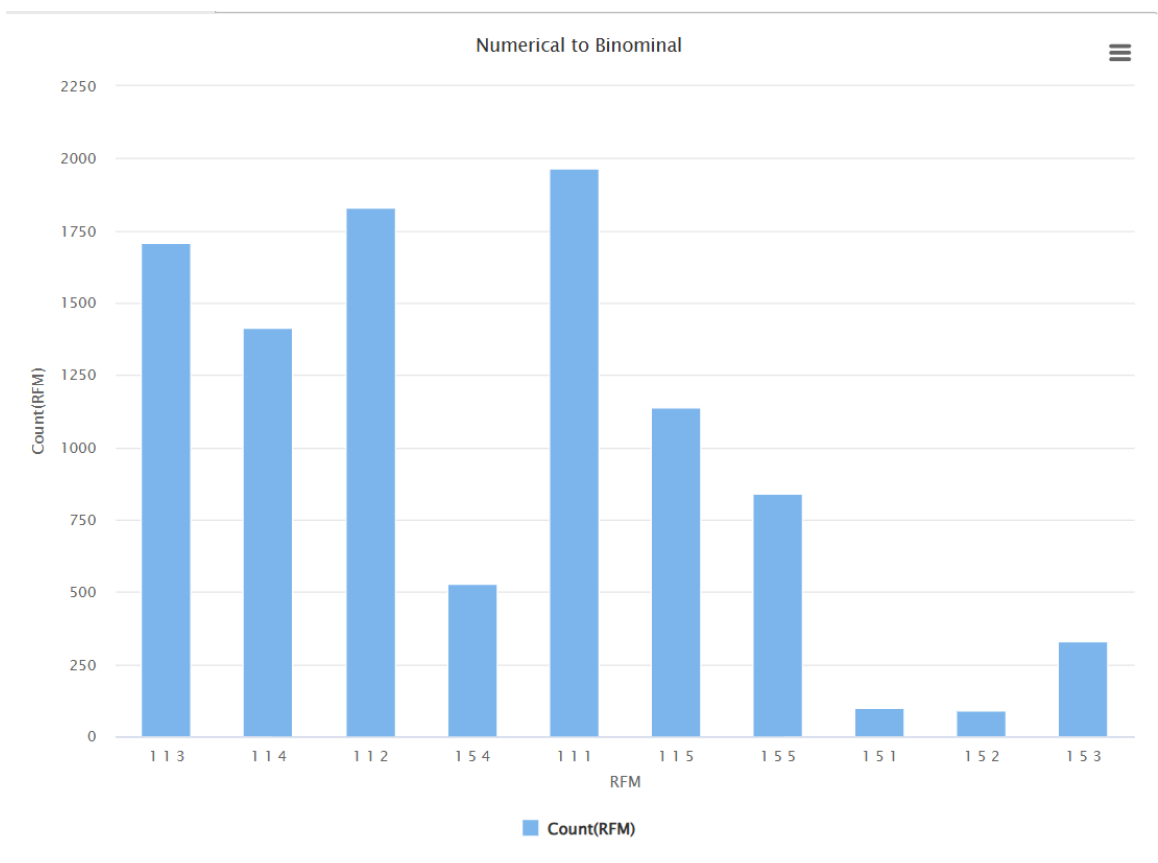
1.3 ใช้เครื่องมือ Nominal to Numerical เพื่อแปลงค่าข้อมูลจากชนิด Nominal (ข้อมูลที่เป็นข้อความหรือเชิงหมวดหมู่ เช่น ชื่อหมวดสินค้าหรือรหัสกลุ่ม) ให้อยู่ในรูปแบบ Numerical (ตัวเลข)

1.4 เริ่มกระบวนการวิเคราะห์โดยกด Start เพื่อให้ระบบทำการ ประมวลผลและสร้างผลลัพธ์ออกมา

1.5 ได้ผลลัพธ์ของโมเดล ที่วิเคราะห์ออกมาเป็นค่า RFM ที่ระบุประเภทลูกค้าดังรูปที่ 3.13



ภาพที่ 3.12 แสดงเครื่องมือของโมเดล RFM



ภาพที่ 3.13 แสดงผลลัพธ์ค่า RFM ที่ใช้ระบุประเภทลูกค้าที่ซื้อสินค้า

2) K-Means Clustering เป็นเทคนิคการทำเหมืองข้อมูลแบบไม่ต้องสอน (Unsupervised Learning) ที่ใช้สำหรับแบ่งข้อมูลออกเป็นกลุ่มย่อย (Clusters) และจัดข้อมูลที่มีความคล้ายกันให้อยู่ด้วยกัน

1.1 การดึงข้อมูล (Retrieve Dataset)

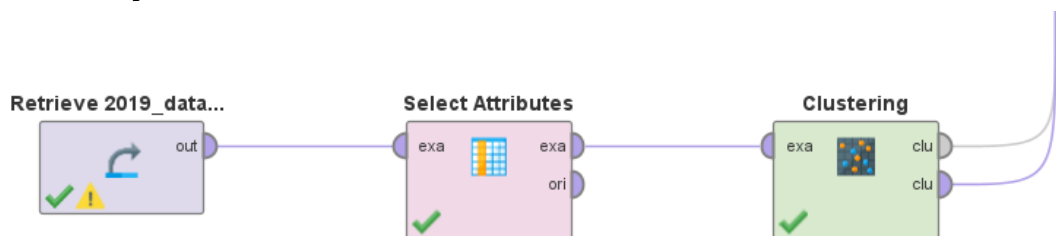
กระบวนการนำข้อมูลจาก แหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

1.2 ใช้เครื่องมือ Select Attributes เลือกเฉพาะคอลัมน์ที่จำเป็นสำหรับการวิเคราะห์ เพื่อให้ข้อมูลสะอาดและลดความซับซ้อนของชุดข้อมูล

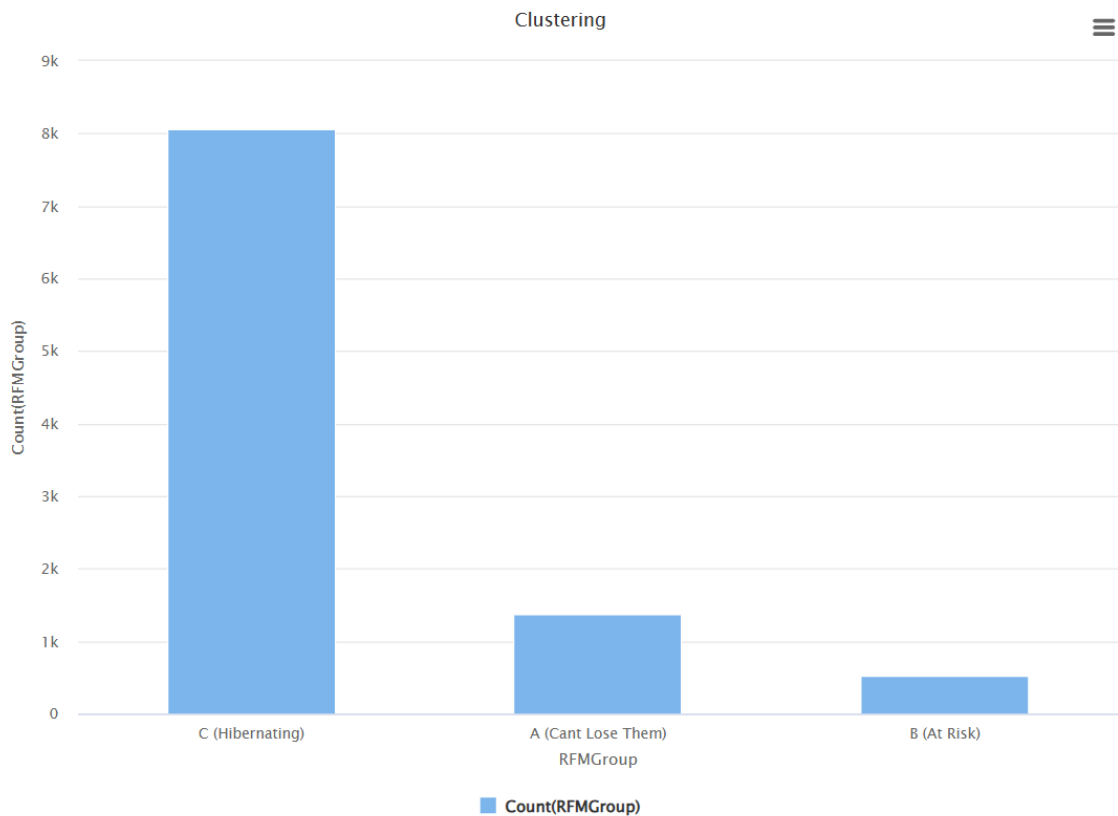
1.3 ใช้เครื่องมือ Clustering เพื่อแบ่งกลุ่มประเภทการซื้อของลูกค้า

1.4 เริ่มกระบวนการวิเคราะห์โดยกด Start เพื่อให้ระบบทำการ ประมวลผลและสร้างผลลัพธ์ออกมา

1.5 ได้ผลลัพธ์ของโมเดล ที่วิเคราะห์แบ่งกลุ่มลูกค้าจาก event_type ที่แบ่งออกมาได้ 3 ประเภทคือ A (Can't Lose Them), B (At Risk), C (Hibernating) ดังรูปที่ 3.15



ภาพที่ 3.14 แสดงเครื่องมือของโมเดล K-Means Clustering



ภาพที่ 3.15 แสดงการแบ่งกลุ่ม Cluster

3) FP-Growth เป็นอัลกอริทึมที่ใช้ค้นหารูปแบบที่เกิดขึ้นบ่อย (Frequent Patterns) ในชุดข้อมูลขนาดใหญ่ โดยเป็นส่วนหนึ่งของ Association Rule Mining ที่ช่วยวิเคราะห์ความสัมพันธ์ระหว่างรายการสินค้าที่มักถูกซื้อพร้อมกัน จุดเด่นคือการสร้างโครงสร้าง FP-Tree เพื่อบีบอัดข้อมูล ทำให้ค้นหารูปแบบได้รวดเร็วและมีประสิทธิภาพโดยไม่ต้องสร้างชุดข้อมูลย่อยจำนวนมาก

1.1 การดึงข้อมูล (Retrieve Dataset)

กระบวนการนำข้อมูลจาก แหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

1.2 ใช้เครื่องมือ Filter Examples สำหรับ

คัดเลือกข้อมูลตัวอย่างที่ต้องการ เช่น การเลือกเฉพาะธุรกรรมที่มี event_type = purchase เพื่อตัดข้อมูลที่ไม่เกี่ยวข้องออกไปและโฟกัสเฉพาะธุรกรรมที่ต้องการวิเคราะห์

1.3 ใช้เครื่องมือ Select Attributes เลือกเฉพาะ

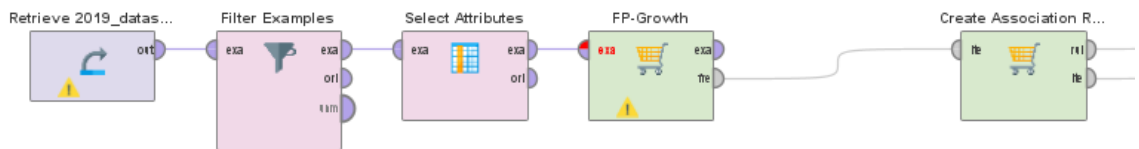
คอลัมน์ที่จำเป็นสำหรับการวิเคราะห์ เพื่อให้ข้อมูลสะอาดและลดความซับซ้อนของชุดข้อมูล

1.4 ใช้เครื่องมือ FP-Growth สำหรับค้นหาชุดข้อมูลที่เกิดบ่อย (Frequent Itemsets) โดยสร้าง FP-Tree เพื่อหาความสัมพันธ์ของสินค้าที่มักถูกซื้อพร้อมกันในธุรกรรม

1.5 ใช้เครื่องมือ Create Association Rules สร้างกฎความสัมพันธ์ (Association Rules) จาก Frequent Itemsets ที่ได้ โดยกำหนดค่าต่าง ๆ เช่น Support และ Confidence เพื่อคัดเลือกเฉพาะกฎที่มีความน่าเชื่อถือ

1.6 เริ่มกระบวนการวิเคราะห์ที่โดยกด Start เพื่อให้ระบบทำการประมวลผลและสร้างผลลัพธ์ออกมา

1.7 ได้ผลลัพธ์ของโมเดล ที่วิเคราะห์ออกมาผลการวิเคราะห์ด้วย FP-Growth ดังภาพที่ 3.17



ภาพที่ 3.16 แสดงเครื่องมือของโมเดล FP-Growth

No.	Premises	Conclusion	Support ↓	Conf. ↓	LaPla...	Gain	p-s	Lift	Convict...
1	RFM = 1 1 1	RFMGroup = C	0.196	1	1	-0.199	0.038	1.234	∞
2	RFM = 1 1 2	RFMGroup = C	0.184	1	1	-0.184	0.035	1.234	∞
3	RFM = 1 1 3	RFMGroup = C	0.172	1	1	-0.172	0.033	1.234	∞
4	RFM = 1 1 4	RFMGroup = C	0.142	1	1	-0.142	0.027	1.234	∞
5	RFM = 1 1 5	RFMGroup = C	0.115	1	1	-0.115	0.022	1.234	∞
6	RFM = 1 5 5	RFMGroup = A	0.084	1	1	-0.084	0.073	7.275	∞
7	RFM = 1 5 4	RFMGroup = A	0.053	1	1	-0.053	0.046	7.275	∞
8	RFM = 1 5 3	RFMGroup = B	0.033	1	1	-0.033	0.031	19.062	∞
9	RFM = 1 5 1	RFMGroup = B	0.010	1	1	-0.010	0.010	19.062	∞

ภาพที่ 3.17 แสดงการวัดประสิทธิภาพของโมเดล FP-Growth

4) Decision Tree (ต้นไม้ตัดสินใจ) เป็นอัลกอริทึมประเภทหนึ่งของ Machine Learning ที่อยู่ในกลุ่ม Supervised Learning (การเรียนรู้แบบมีผู้สอน) โดยเป็นแบบจำลองที่

ใช้ได้ทั้งสำหรับการ จำแนกประเภท (Classification) และการ พยากรณ์ค่า (Regression) ของ ข้อมูล

1.1 การดึงข้อมูล (Retrieve Dataset)

กระบวนการนำข้อมูลจาก แหล่งข้อมูลหรือฐานข้อมูลมาใช้งานเป็นการรวบรวมข้อมูลจากไฟล์ ต่างๆ เช่น CSV, Excel หรือนำ Dataset ที่ทำการ Import เข้ามาเพื่อใช้งาน

1.2 ใช้เครื่องมือ Select Attributes เลือกเฉพาะ

คอลัมน์ที่จำเป็นสำหรับการวิเคราะห์ เพื่อให้ข้อมูลสะอาดและลดความซับซ้อนของชุดข้อมูล

1.3 ใช้เครื่องมือ Set Role เป็นการกำหนด

บทบาทของตัวแปรในชุดข้อมูล โดยกำหนดตัวแปรเป้าหมาย (Label)

1.4 ใช้เครื่องมือ Decision Tree สร้างแบบจำลอง

ต้นไม้ตัดสินใจ (Decision Tree) เพื่อเรียนรู้รูปแบบและกฎการตัดสินใจจากข้อมูลและตัวแปรตัว อื่น ๆ

1.5 RFMDuplicate เป็นชุดข้อมูล RFM ที่ถูกทำ

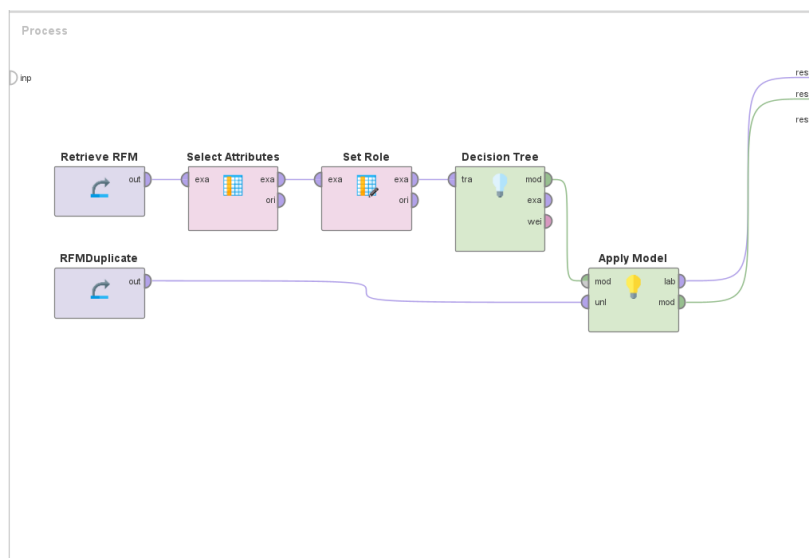
สำเนาไว้โดยยังไม่กำหนดป้ายกำกับ (Label) เพื่อใช้เป็นข้อมูลสำหรับการนำโมเดลที่สร้างแล้ว ไปทำนายผล โดยข้อมูลชุดนี้จะถูกส่งเข้าไปในขั้นตอน Apply Model

1.6 ใช้เครื่องมือ Apply Model นำแบบจำลอง

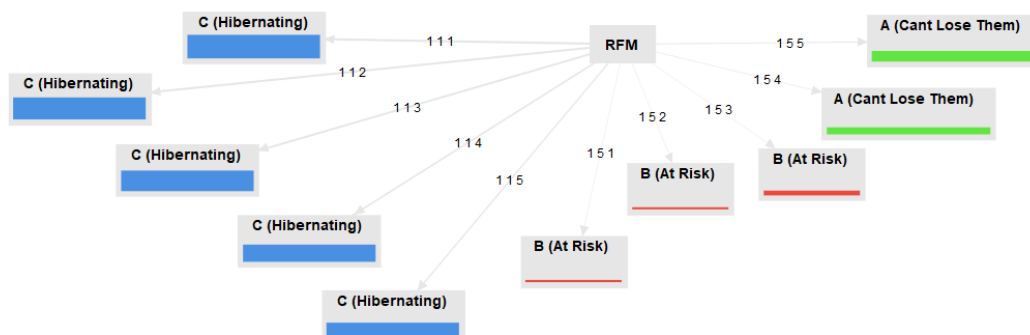
Decision Tree ที่สร้างเสร็จแล้ว มาประยุกต์ใช้กับข้อมูล RFMDuplicate เพื่อจำแนกประเภท กลุ่มลูกค้า

1.7 ได้ผลลัพธ์ของโมเดล ที่วิเคราะห์ออกมาผล

การวิเคราะห์ด้วย Decision Tree ดังภาพที่ 3.19



ภาพที่ 3.18 แสดงเครื่องมือของโมเดล Decision Tree

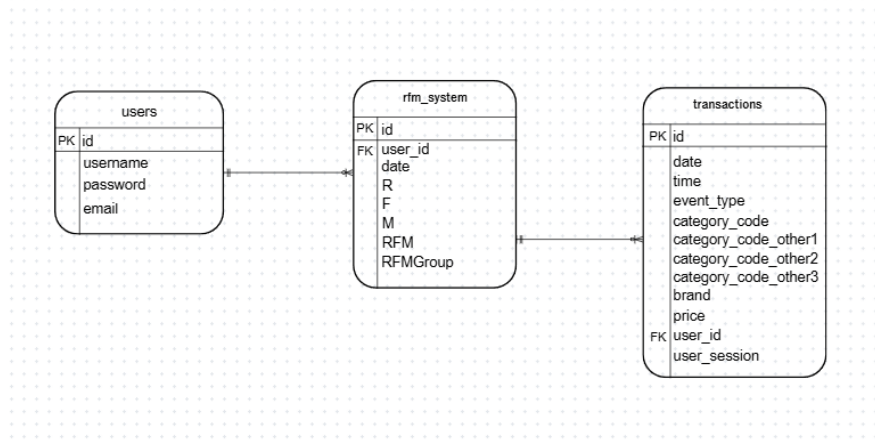


ภาพที่ 3.19 แสดงผลลัพธ์ของโมเดล Decision Tree

3.1.5 การวัดประสิทธิภาพของโมเดล (Evaluation) การวัดประสิทธิภาพของโมเดลเป็นขั้นตอนสำคัญในการประเมินคุณภาพและความเหมาะสมของโมเดลที่ใช้วิเคราะห์ โดยในโครงการนี้ได้เลือกใช้ 3 โมเดลหลัก ได้แก่ RFM Analysis, K-Means Clustering, FP-Growth และ Decision Tree ซึ่งแต่ละโมเดลจะมีการวัดผลที่แตกต่างกันไป สำหรับ RFM Analysis ประสิทธิภาพของโมเดลจะพิจารณาจากความสามารถในการจัดกลุ่มลูกค้าตามเกณฑ์ Recency, Frequency และ Monetary ที่สะท้อนพฤติกรรมจริงของลูกค้าได้อย่างเป็นรูปธรรม ขณะที่ K-Means Clustering ใช้การประเมินจากค่าเฉลี่ย RFM ของแต่ละกลุ่มและความแตกต่างระหว่าง Cluster เพื่อดูว่าการแบ่งกลุ่มสามารถแยกประเภทลูกค้าได้ชัดเจนและสะท้อนความเป็นจริงมากน้อยเพียงใด สำหรับ FP-Growth ใช้การวัดผลด้วยค่าตัวชี้วัดทางสถิติ ได้แก่ Support, Confidence และ Lift เพื่อบ่งชี้ความแข็งแกร่งและความน่าเชื่อถือของกฎความสัมพันธ์ที่ค้นพบ และ Decision Tree ใช้เป็นการจำแนกประเภทกลุ่มลูกค้า โดยการพิจารณาผลลัพธ์เหล่านี้ช่วยให้สามารถเลือกใช้โมเดลได้อย่างเหมาะสมที่สุดต่อการวิเคราะห์พฤติกรรมการซื้อของลูกค้า

3.1.6 การนำโมเดลไปใช้งานจริง (Deployment) การนำแบบจำลองไปใช้งานเป็นขั้นตอนสุดท้ายของกระบวนการ CRISP-DM ซึ่งมุ่งเน้นการนำผลการวิเคราะห์หรือโมเดลที่สร้างขึ้นมาใช้ประโยชน์จริงในเชิงปฏิบัติ โดยวัตถุประสงค์หลักคือการเปลี่ยนผลลัพธ์เชิงข้อมูลให้เป็นเครื่องมือที่สนับสนุนการตัดสินใจทางธุรกิจได้อย่างมีประสิทธิภาพ ในโครงการนี้ การนำแบบจำลองไปใช้งานอยู่ในรูปของการนำผลการวิเคราะห์ RFM Analysis, K-Means Clustering และ FP-Growth ไปใช้ในการแบ่งกลุ่มลูกค้าและค้นหาความสัมพันธ์ของสินค้า

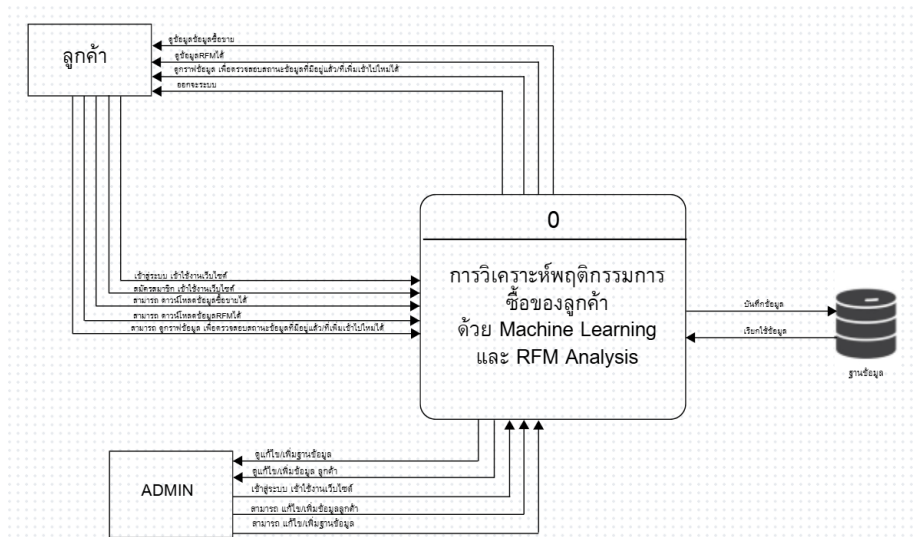
3.2 ER-Diagram



ภาพที่ 3.20 ภาพของ ER-Diagram

จากแผนภาพ โครงสร้างฐานข้อมูลของระบบประกอบด้วย 3 ตารางหลัก ได้แก่ ตาราง users สำหรับจัดเก็บข้อมูลบัญชีผู้ดูแลระบบ, ตาราง transactions สำหรับเก็บรายละเอียดประวัติการทำธุรกรรมการซื้อขายของลูกค้าทั้งหมด และตาราง rfm_system สำหรับเก็บผลลัพธ์คะแนน R, F, M และการจัดกลุ่มเป้าหมาย โดยข้อมูลการซื้อและผลการวิเคราะห์จะมีความสัมพันธ์เชื่อมโยงกันผ่านรหัสลูกค้า (user_id) เพื่อให้ระบบสามารถนำประวัติการทำธุรกรรมมาประมวลผลและระบุกลุ่มลูกค้าได้อย่างถูกต้อง

3.3 Context Diagram



ภาพที่ 3.21 ภาพของ Context Diagram

ตารางที่ 3.1 แสดงคำอธิบายการประมวลผลของโปรเซส 1.0 : ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	ระบบการวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้าด้วย Machine Learning และ RFM Analysis
DFD number	1.0
Process name	ตรวจสอบการเข้าสู่ระบบ
Input data flows	ข้อมูลการเข้าสู่ระบบ ข้อมูลสมัครสมาชิก ข้อมูลการแก้ไข/เพิ่มข้อมูล
Output data flows	สถานะการเข้าสู่ระบบ ข้อมูลการบันทึก แก้ไข/เพิ่มข้อมูล
Data stores user	ฐานข้อมูล
Description	โปรเซสในการตรวจสอบสิทธิ์การใช้งาน ของผู้ดูแลระบบ รวมถึงการจัดการข้อมูล

ตารางที่ 3.2 แสดงคำอธิบายการประมวลผลของโปรเซส 2.0 : เข้าชมเว็บไซต์

Process Description	
System	ระบบการวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้าด้วย Machine Learning และ RFM Analysis
DFD number	2.0
Process name	เข้าชมเว็บไซต์
Input data flows	คำสั่งการเข้าชมเว็บไซต์
Output data flows	ข้อมูลสำหรับการแสดงผลหน้าเว็บไซต์
Data stores user	ฐานข้อมูล
Description	โปรเซสในการรับคำสั่งร้องขอเข้าชมหน้า เว็บไซต์เพื่อดึงข้อมูลจากฐานข้อมูลมา แสดงผล

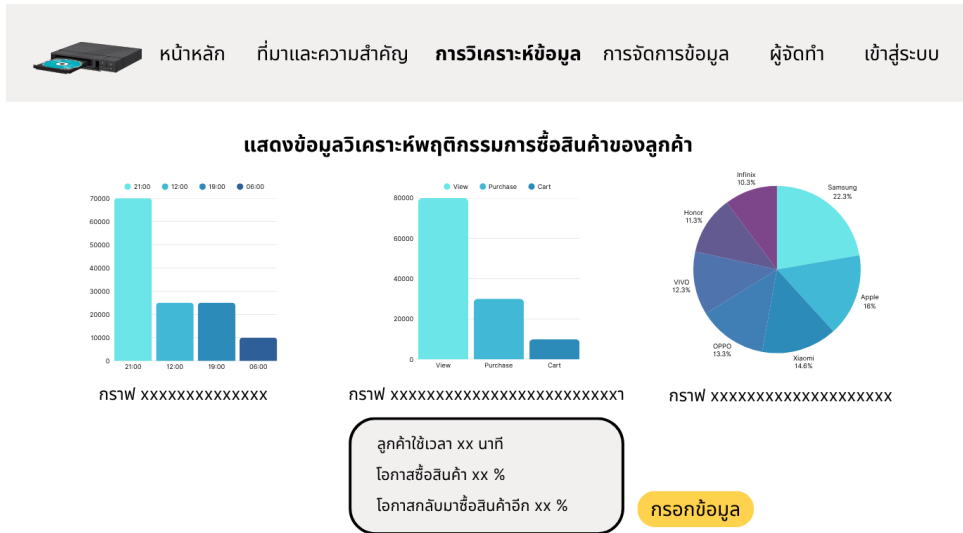
ตารางที่ 3.3 แสดงคำอธิบายการประมวลผลของโปรเซส 3.0 : ข้อมูลกราฟ

Process Description	
System	ระบบการวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้าด้วย Machine Learning และ RFM Analysis
DFD number	3.0
Process name	ข้อมูลกราฟ
Input data flows	ข้อมูลกราฟ ข้อมูลซื้อขาย
Output data flows	ข้อมูลการแสดงผลกราฟ ข้อมูลประวัติการณ์ซื้อขาย
Data stores user	ข้อมูลกราฟ ข้อมูลซื้อขาย
Description	โปรเซสในการดึงข้อมูลประวัติการณ์ซื้อขาย และข้อมูลสรุปผลในรูปแบบกราฟมาแสดงผลให้ลูกค้า

ตารางที่ 3.4 แสดงคำอธิบายการประมวลผลของโปรเซส 4.0 : ข้อมูล RFM

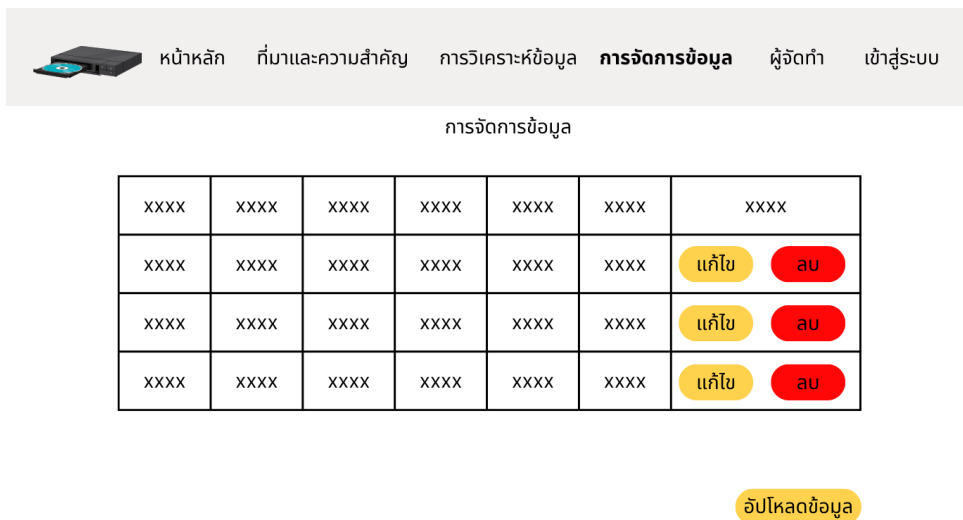
Process Description	
System	ระบบการวิเคราะห์พฤติกรรมการณ์ซื้อของลูกค้าด้วย Machine Learning และ RFM Analysis
DFD number	4.0
Process name	ข้อมูล RFM
Input data flows	ข้อมูล RFM
Output data flows	ข้อมูลการแสดงผล RFM ข้อมูลไฟล์ดาว์นโหลด RFM
Data stores user	ข้อมูล RFM
Description	โปรเซสในการดึงผลลัพธ์การวิเคราะห์พฤติกรรมการณ์ลูกค้าตามเกณฑ์ RFM มาแสดงผลและอนุญาตให้ดาว์นโหลด

3) หน้าการวิเคราะห์ข้อมูล แสดงผลลัพธ์ในรูปแบบกราฟต่าง ๆ เช่น กราฟแท่ง แผนภูมิวงกลม เพื่อให้ผู้ใช้งานเห็นแนวโน้มและพฤติกรรมที่ชัดเจน



ภาพที่ 3.25 หน้าการวิเคราะห์ข้อมูล

4) หน้าการจัดการข้อมูล สำหรับผู้ดูแลระบบ (Administrator) ใช้ในการตรวจสอบ แก้ไข หรือลบข้อมูลในฐานข้อมูล รวมถึงการอัปเดตข้อมูลใหม่ โดยออกแบบให้ง่ายต่อการจัดการ พร้อมปุ่ม “แก้ไข” และ “ลบ” เพื่อความสะดวกในการดูแลระบบ



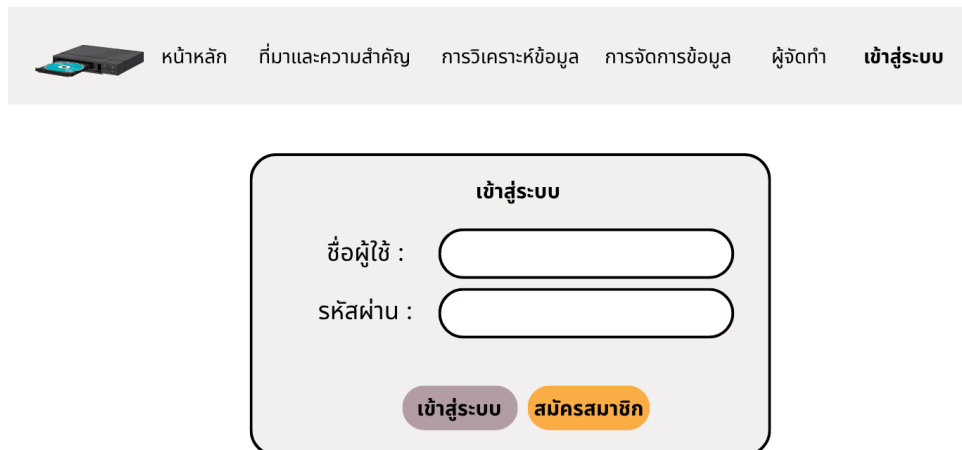
ภาพที่ 3.26 หน้าการจัดการข้อมูล

5) หน้าผู้จัดทำ แสดงข้อมูลและรูปภาพของผู้จัดทำโครงการงาน เพื่อแสดงความรับผิดชอบและให้เครดิตกับทีมผู้พัฒนา



ภาพที่ 3.27 หน้าผู้จัดทำ

6) หน้าเข้าสู่ระบบ เป็นหน้าที่ใช้สำหรับผู้ใช้งานที่มีบัญชีอยู่แล้วในการเข้าสู่ระบบ โดยผู้ใช้ต้องกรอกชื่อผู้ใช้ (Username) และรหัสผ่าน (Password) เพื่อยืนยันตัวตนก่อนเข้าใช้งานระบบ หน้านี้มีปุ่ม “เข้าสู่ระบบ” ใช้สำหรับควบคุมสิทธิ์การเข้าถึงข้อมูล และเพิ่มความปลอดภัยของระบบ



ภาพที่ 3.28 หน้าเข้าสู่ระบบ

3.5 บทสรุป

จากกระบวนการที่ดำเนินงาน โครงการนี้เริ่มจากการศึกษาและทำความเข้าใจข้อมูล พฤติกรรมการซื้อสินค้าของลูกค้า จากนั้นนำมาวิเคราะห์ด้วยกระบวนการ CRISP-DM เพื่อสร้างแบบจำลองข้อมูลด้วยเทคนิค RFM Analysis, K-Means Clustering, FP-Growth และ Decision Tree ที่เหมาะสม ผลลัพธ์ที่ได้ถูกนำไปประยุกต์ใช้งานในรูปแบบเว็บไซต์เพื่อแสดงผล การวิเคราะห์และสนับสนุนการตัดสินใจทางธุรกิจ ช่วยให้ผู้ประกอบการหรือผู้สนใจสามารถ นำไปปรับใช้ได้จริง ทั้งยังเพิ่มประสิทธิภาพในการวางกลยุทธ์การตลาด การรักษาลูกค้า และเป็นประโยชน์ต่อการเรียนรู้ด้านการวิเคราะห์ข้อมูลของนักศึกษาในอนาคต