บทที่ 3
วิธีการดำเนินงานโครงงาน

โครงงานการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลความคงอยู่ของลูกค้า กรณีศึกษาบริษัท Telco เป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมนิ่ง ซึ่งมีกระบวนการ วิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

3.2 แผนภาพกระแสข้อมูล Data Flow Diagram

3.3 การออกแบบเว็บไซต์

3.4 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP–DM

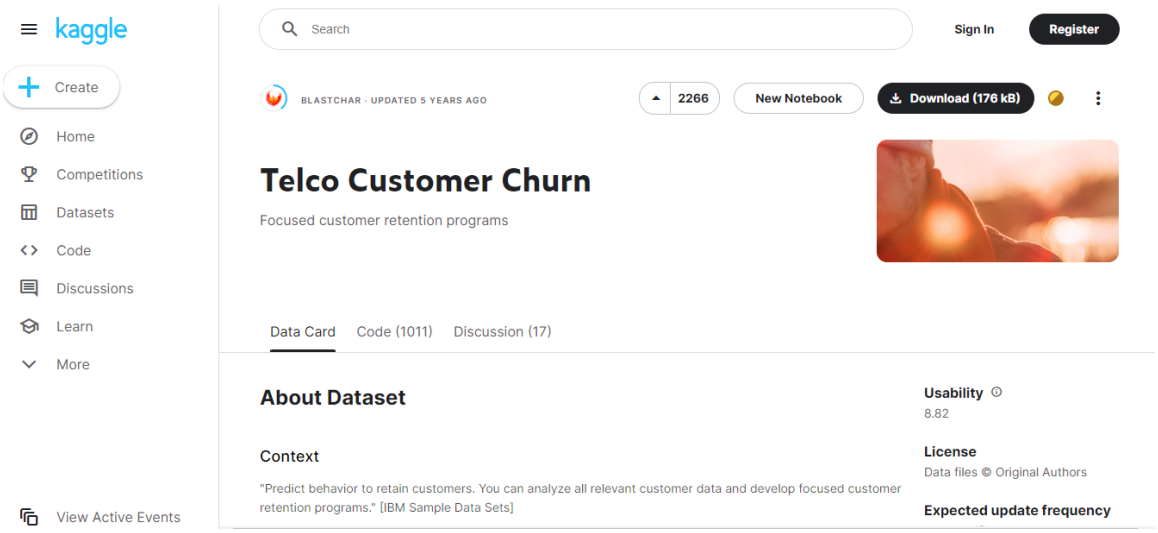
กระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Ch0rysler, SPSS และ NCR ที่มีการพัฒนาเป็น Workflow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดัง ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวมผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทาง Data Mining โดยการวิเคราะห์ข้อมูลในประเด็นนี้คือ การวิเคราะห์ข้อมูลความคงอยู่ของลูกค้าที่ยกเลิกหรือไม่ยกเลิกบริการ ซึ่งมีข้อมูลมาก จำนวนรายการทั้งหมด 7,043 รายการ ทำให้ไม่สามารถทำความเข้าใจกับข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็ว เช่น ต้องการทราบว่าลูกค้ายกเลิกบริการแบ่งเป็นสัญญาประเภทใดบ้าง

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบการพัฒนาเว็บไซต์ การพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลความคงอยู่ของลูกค้า กรณีศึกษาบริษัท Telco ทั้งเอกสารที่มีอยู่และศึกษาระบบงานเดิมเพื่อทำความเข้าใจถึงขั้นตอนการทำงานและทราบถึงจุดสำคัญของระบบ ทำให้การ

ออกแบบระบบใหม่มีความสอดคล้องกับความต้องการของผู้ใช้ โดยมีเนื้อหาครอบคลุมในเรื่องต่าง ๆ ดังนี้

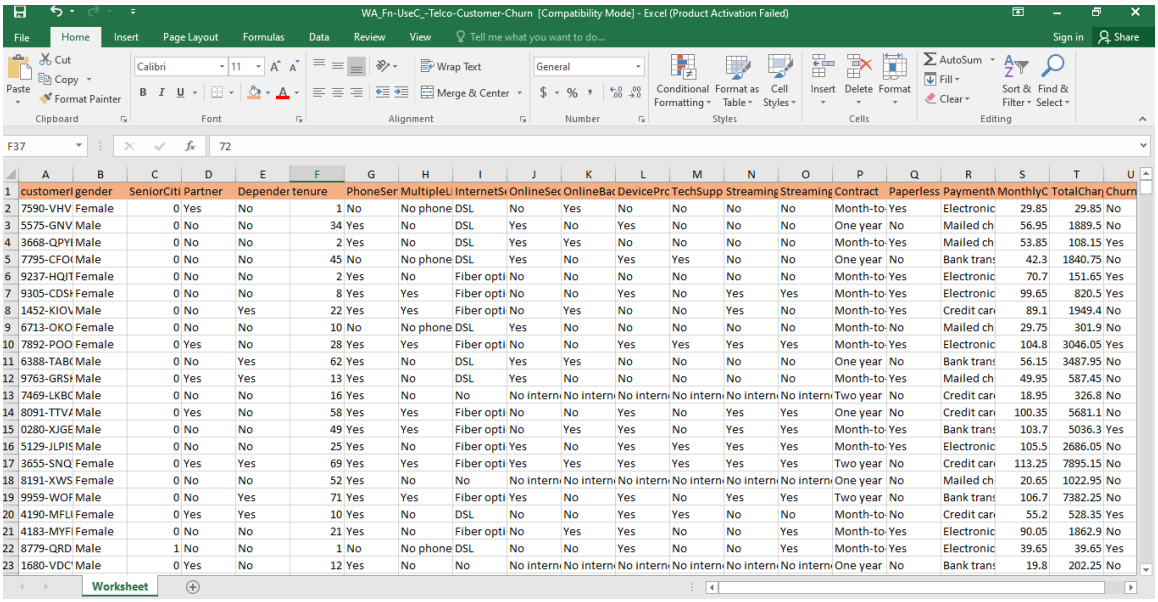
ผู้วิเคราะห์ข้อมูลทำการรวบรวมข้อมูลเพื่อตรวจสอบข้อมูลที่ได้จากเว็บไซต์ kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือ ชุดข้อมูล สำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภท ไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผย ได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไป



ภาพที่ 3.1 เว็บไซต์ kaggle.com

(ที่มา: https:www.kaggle.com)

ซึ่งข้อมูลการจำแนกลูกค้าที่ยกเลิกหรือคงอยู่มีจำนวนข้อมูล 7,043 รายการ ประกอบด้วย 21 แอตทริบิวท์ ประกอบด้วย customerID (รหัสลูกค้า) gender (เพศของลูกค้า) Partner (พันธะมิตร) Dependents(ผู้อยู่ในอุปการะ) tenure(ระยะเวลา) PhoneService (บริการโทรศัพท์) MultipleLines (การโทรหลายสาย) InternetService (บริการอินเทอร์เน็ต) OnlineSecurity (การรักษาความปลอดภัยออนไลน์) OnlineBackup (การสำรองข้อมูลออนไลน์) DeviceProtection (แผนคุ้มครองอุปกรณ์) TechSupport (การสนับสนุนด้านเทคนิค) StreamingTV (สตรีมมิ่งทีวี) StreamingMovies (สตรีมมิ่งภาพยนตร์) Contract (ประเภทสัญญา) PaperlessBilling (การเรียกเก็บเงินแบบไร้กระดาษ) PaymentMethod (วิธีการชำระเงิน) MonthlyCharges (ค่าบริการรายเดือน) TotalCharges (ค่าบริการทั้งหมด) Churn (ความคงอยู่ของลูกค้า)



ภาพที่ 3. 2 ข้อมูลความคงอยู่ของลูกค้า

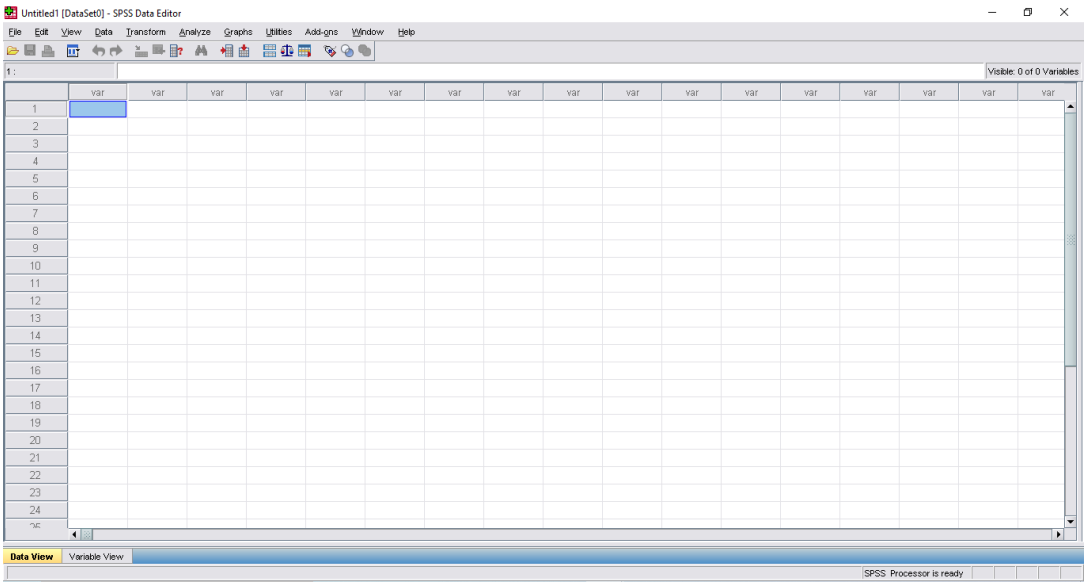
3.1.3 เตรียมข้อมูลให้พร้อมใช้งาน (Data Preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง โดยการทำ Data cleaning มีขั้นตอนดังนี้

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูล และทำการ Data Cleaning โดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 14 แอตทริบิวท์ ได้แก่ gender(เพศของลูกค้า) PhoneService(บริการโทรศัพท์) MultipleLines (การโทรหลายสาย) InternetService (บริการอินเทอร์เน็ต) OnlineSecurity (การรักษาความปลอดภัยออนไลน์) OnlineBackup (การสำรองข้อมูลออนไลน์) DeviceProtection (แผนคุ้มครองอุปกรณ์) TechSupport (การสนับสนุนด้านเทคนิค) StreamingTV (สตรีมมิ่งทีวี) StreamingMovies (สตรีมมิ่งภาพยนตร์) Contract (ประเภทสัญญา) PaperlessBilling (การเรียกเก็บเงินแบบไร้กระดาษ) PaymentMethod (วิธีการชำระเงิน) Churn (ความคงอยู่ของลูกค้า) ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	gender	PhoneSer	MultipleLines	InternetService	OnlineSec	OnlineBac	DevicePro	TechSupp	Streaming	StreamingMovies	Contract	Paperless	PaymentMethod	Churn
2	Female	No	No phone ser	DSL	No	Yes	No	No	No	No	Month-to-month	Yes	Electronic check	No
3	Male	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check	No
4	Male	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed check	Yes
5	Male	No	No phone ser	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (au	No
6	Female	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	Yes
7	Female	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check	Yes
8	Male	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-month	Yes	Credit card (autor	No
9	Female	No	No phone ser	DSL	Yes	No	No	No	No	No	Month-to-month	No	Mailed check	No
10	Female	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electronic check	Yes
11	Male	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank transfer (au	No
12	Male	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-month	Yes	Mailed check	No
13	Male	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No internet service	Two year	No	Credit card (autor	No
14	Male	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	One year	No	Credit card (autor	No
15	Male	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-month	Yes	Bank transfer (au	Yes
16	Male	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electronic check	No
17	Female	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit card (autor	No
18	Female	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No internet service	One year	No	Mailed check	No
19	Male	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank transfer (au	No
20	Female	Yes	No	DSL	No	No	Yes	Yes	No	No	Month-to-month	No	Credit card (autor	Yes

ภาพที่ 3.3 การคัดเลือกข้อมูล (Data Selection)

จากนั้นใช้โปรแกรม SPSS มาเป็นตัวช่วยในการเลือกแอตทริบิวต์ที่จำเป็นต่อการตัดสินใจความคงอยู่ของลูกค้า



ภาพที่ 3.4 โปรแกรม SPSS

ตารางที่ 3.1 แสดงค่าสัมประสิทธ์การถดถอย

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.396	.041		34.425	.000
	gender	.007	.009	.007	.689	.491
	PhoneService	-.057	.018	-.038	-3.168	.002
	MultipleLines	.000	.000	.014	1.341	.180
	InternetService	-.100	.011	-.107	-8.849	.000
	OnlineBackup	.002	.002	.011	1.019	.308
	DeviceProtection	.000	.000	.007	.700	.484
	TechSupport	.001	.001	.015	1.360	.174
	StreamingMovies	.000	.000	.016	1.475	.140
	Contract	.172	.006	.325	28.059	.000
	PaperlessBilling	.102	.010	.114	10.390	.000
	PaymentMethod	.048	.004	.126	10.926	.000
	StreamingTV	.000	.000	.017	1.561	.118

จากตารางที่ 3.1 แสดงค่าสัมประสิทธ์การถดถอย เป็นตารางที่แสดงค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวในสมการมีดังนี้

Model สมการผลการวิเคราะห์

B ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวที่จะนำมาสร้างสมการพยากรณ์ ซึ่งเป็นค่า Unstandardized Coefficient จะเป็นการเขียนสมการในรูปคะแนนดิบหรือค่าจริง

Beta ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระแต่ละตัวที่จะนำมาสร้างสมการพยากรณ์ ซึ่งค่าStandardized Coefficient จะเป็นการเขียนสมการในรูปคะแนนมาตรฐาน

t ค่าสถิติ t เป็นสถิติที่ใช้ทดสอบว่าตัวแปรอิสระใดบ้างที่สามารถใช้พยากรณ์ตัวแปรตามได้บ้าง เป็นการทดสอบค่าสัมประสิทธิ์การถดถอย (β) โดยมีสมมติฐานการทดสอบดังนี้

$H_0 : \beta_i = 0$ หรือ ตัวแปรอิสระตัวที่ i ไม่มีผลต่อตัวแปรตาม (ไม่สามารถใช้พยากรณ์ตัวแปรตามได้)

$H_i : \beta_i \neq 0$ ตัวแปรอิสระตัวที่ i มีผลต่อตัวแปรตาม (สามารถใช้พยากรณ์ตัวแปรตามได้)

การสรุปและตัดสินใจ จะเปรียบเทียบค่า Sig. ในผลการวิเคราะห์กับค่าระดับนัยสำคัญทางสถิติ(α นิยมกำหนดเกณฑ์ที่ 0.05) ถ้าผลการวิเคราะห์พบว่าค่า Sig. มีค่าน้อยกว่าค่า α จะปฏิเสธ H_0 ยอมรับ H_Aแสดงว่าตัวแปรอิสระตัวที่ i มีผลต่อตัวแปรตาม (สามารถใช้พยากรณ์) ตัวแปรตามได้อย่างมีนัยสำคัญทางสถิต

Sig. ค่าความน่าจะเป็นในการปฏิเสธสมมติฐาน H_0 ที่ได้จากการคำนวณจากข้อมูลตัวอย่างที่ น ามาวิเคราะห์สถิติ t ที่ใช้ทดสอบว่าตัวแปรอิสระใดบ้างที่สามารถใช้พยากรณ์ตัวแปรตาม ได้บ้าง ซึ่งเป็นการทดสอบค่าสัมประสิทธิ์การถดถอย (βi) ค่าน้อยกว่า α สรุปว่า ปฏิเสธ H_0 ยอมรับ H_A : βi≠ 0 แสดงว่าตัวแปรอิสระตัวที่ i มีผลต่อตัวแปรตาม (สามารถใช้ พยากรณ์) ตัวแปรตามได้อย่างมีนัยสำคัญทางสถิติ หรือ βi≠ 0

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูล เป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และ สถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การจำแนก (Classification) การแบ่งกลุ่ม (Clustering) และการสร้างความสัมพันธ์ (Association rule)

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัยที่สามารถทำให้รู้ได้ว่าจำนวนลูกค้าที่ยกเลิกบริการ จากเทคนิคการจำแนกประเภทข้อมูล ด้วยการสร้างโมเดล Decision Tree เพื่อจัดกลุ่มข้อมูลที่มีความสัมพันธ์กัน โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูลด้วยชุดข้อมูลที่คัดเลือก

PhoneSe	InternetServic	Contract	Paperless	PaymentMethod	Churn
No	DSL	Month-to-month	Yes	Electronic check	No
Yes	DSL	One year	No	Mailed check	No
Yes	DSL	Month-to-month	Yes	Mailed check	Yes
No	DSL	One year	No	Bank transfer (automatic)	No
Yes	Fiber optic	Month-to-month	Yes	Electronic check	Yes
Yes	Fiber optic	Month-to-month	Yes	Electronic check	Yes
Yes	Fiber optic	Month-to-month	Yes	Credit card (automatic)	No
No	DSL	Month-to-month	No	Mailed check	No
Yes	Fiber optic	Month-to-month	Yes	Electronic check	Yes
Yes	DSL	One year	No	Bank transfer (automatic)	No
Yes	DSL	Month-to-month	Yes	Mailed check	No
Yes	No	Two year	No	Credit card (automatic)	No
Yes	Fiber optic	One year	No	Credit card (automatic)	No
Yes	Fiber optic	Month-to-month	Yes	Bank transfer (automatic)	Yes
Yes	Fiber optic	Month-to-month	Yes	Electronic check	No
Yes	Fiber optic	Two year	No	Credit card (automatic)	No
Yes	No	One year	No	Mailed check	No
Yes	Fiber optic	Two year	No	Bank transfer (automatic)	No

ภาพที่ 3.5 ชุดข้อมูลที่คัดเลือกมาทำการวิเคราะห์ข้อมูล

จากรูปภาพที่ 3.5 ประกอบด้วย 6 แอตทริบิวต์ คือ

– PhoneService แสดงสมัครการใช้บริการโทรศัพท์ ประกอบด้วย 2 ค่า คือ yes “ใช่” No “ไม่ใช่”

– InternetService แสดงสมัครใช้บริการอินเทอร์เน็ต ประกอบด้วย 3 ค่าคือ No” ไม่ใช่” DSL”สายโทรศัพท์” Fiber optic”สายไฟเบอร์ออฟติก”

– Contract แสดงประเภทสัญญา ประกอบด้วย 3 ค่า คือ Month–to–Month”เดือนต่อเดือน” One Year”1ปี” Two Year”2ปี”

– PaperlessBilling แสดงลูกค้าเลือกการเรียกเก็บเงินแบบไร้กระดาษ ประกอบด้วย 2 ค่า คือ yes “ใช่” No “ไม่ใช่”

–PaymentMethod แสดงวิธีที่ลูกค้าชำระบิล ประกอบไปด้วย 4 ค่า คือ Bank Withdrawal’เงินสด’ Credit Card ”บัตรเครดิต” Electronic check”โอนเงิน” Mailed Check” เช็ค”

– Churn แสดงการตัดสินใจความคงอยู่ของลูกค้าที่ใช้บริการต่อ ประกอบด้วย 2 ค่า คือ yes “ใช้ต่อ” No “ไม่ใช้ต่อ”

การสร้างโมเดล Decision tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์ กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการดังนี้

ตารางที่ 3.2 สมการการหาความสัมพันธ์ของแอตทริบิวต์

การคำนวณ	สมการ
Entropy	entropy(c1) = −p(c1) log p(c1) และ p(c2) คือ ค่าความน่าจะเป็นของ c1
Information Gain	IG (parent, child) = entropy(parent) − [p(c1) x entropy(c1) + p(c2) x entropy(c2) + ...]

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มาก ที่สุดมาเป็น root ของ Decision กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นใช้ต่อและไม่ใช้ต่อ ดังนี้

1. คำนวณค่า IG ของแอตทริบิวต์ PhoneService จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(churn) = −p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)
= −[0.27 * $\log_2$ (0.27) + 0.74 * $\log_2$ (0.74)]
= −[0.27 * − 1.89 + 0.74 * − 0.43]
= 0.83

entropy(phone = yes) = −p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)
= −[0.27 * $\log_2$ (0.27) + 0.73 * $\log_2$ (0.73)]
= −[0.27 * − 1.89 + 0.73 * − 0.45]
= 0.84

entropy(phone = no) = −p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)
= −[0.25 * $\log_2$ (0.25) + 0.75 * $\log_2$ (0.75)]
= −[0.25 * − 2 + 0.75 * − 0.42]
= 0.82

IG(churn , phone) = entropy(churn) − [p(phone = yes) * entropy(phone = yes) + p(phone = no) * entropy(phone = no)]

= 0.83 – [0.89 * 0.84 + 0.1 *0.82]

= 0.83 – [0.75 + 0.08]

= 0.83 – 0.83

= 0

2. คำนวณค่า IG ของแอตทริบิวต์ InternetService จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(churn) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.27 * $\log_2$ (0.27) + 0.74 * $\log_2$ (0.74)]

= –[0.27 * – 1.89 + 0.74 * – 0.43]

= 0.83

entropy(internet = DSL) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.19 * $\log_2$ (0.19) + 0.81 * $\log_2$ (0.81)]

= –[0.19 * – 2.39 + 0.81 * – 0.3]

= 0.7

entropy(internet = fiber) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.42 * $\log_2$ (0.42) + 0.58 * $\log_2$ (0.58)]

= –[0.42 * – 1.25 + 0.58 * – 0.79]

= 0.98

entropy(internet = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.07 * $\log_2$ (0.07) + 0.93 * $\log_2$ (0.93)]

= –[0.07 * – 3.84 + 0.93 * – 0.12]

= 0.38

IG(churn , internet) = entropy(churn) – [p(internet = DSL) * entropy (internet =DSL) + p(internet = fiber) * entropy(internet = fiber) + p(internet = no) * entropy(internet =no)]

= 0.83 – [0.34 * 0.7 + 0.44 *0.98+0.22*0.38]

= 0.83 – 0.75

= 0.08

3. คำนวณค่า IG ของแอตทริบิวต์ Contract จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(churn) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.27 * $\log_2$ (0.27) + 0.74 * $\log_2$ (0.74)]

= –[0.27 * – 1.89 + 0.74 * – 0.43]

= 0.83

entropy(contract = month–to–month) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.43 * $\log_2$ (0.43) + 0.57 * $\log_2$ (0.57)]

= –[0.43 * – 1.222 + 0.57 * – 0.81]

= 0.98

entropy(contract = one year) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.11 * $\log_2$ (0.11) + 1 * $\log_2$ (1)]

= –[0.11 * – 3.18 + 1 * – 0]

= 0

entropy(contract = two year) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.03 * $\log_2$ (0.03) + 1 * $\log_2$ (1)]

= –[0.03 * – 5.1 + 1* – 0]

= 0

IG(churn , contract) = entropy(contract) – [p(contract = month–to–month) * entropy (contract = month–to–month) + p(contract = one year) * entropy(contract = one year) + p(contract = two year) * entropy(contract = two year)]

= 0.83 – [0.55 * 0.98 + 0.21 *0 + 0.24 * 0]

= 0.83 – 0.55

= 0.28

4. คำนวณค่า IG ของแอตทริบิวต์ PaperlessBillingจากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(churn) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.27 * $\log_2$ (0.27) + 0.74 * $\log_2$ (0.74)]

= –[0.27 * – 1.89 + 0.74 * – 0.43]

= 0.83

entropy(paperbill = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.34 * $\log_2$ (0.34) + 0.66 * $\log_2$ (0.66)]

= –[0.34 * – 1.56 + 0.66 * – 0.6]

= 0.93

entropy(paperbill = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.16 * $\log_2$ (0.16) + 0.84 * $\log_2$ (0.84)]

= –[0.16 * – 2.64 + 0.84 * – 0.25]

= 0.63

IG(churn , paperbill) = entropy(churn) – [p(paperbill =yes) * entropy(paperbill =yes) + p(paperbill = no) * entropy(paperbill = no)]

= 0.83 – [0.59 * 0.93 + 0.41 *0.63]

= 0.83 – 0.81

= 0.02

5. คำนวณค่า IG ของแอตทริบิวต์ PaymentMethodจากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(churn) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.27 * $\log_2$ (0.27) + 0.74 * $\log_2$ (0.74)]

= –[0.27 * – 1.89 + 0.74 * – 0.43]

= 0.83

entropy(payment =bank) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.17 * $\log_2$ (0.17) + 0.83 * $\log_2$ (0.83)]

= –[0.17 * – 2.56 + 0.83 * – 0.27]

= 0.66

entropy(payment =credit card) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.15 * $\log_2$ (0.15) + 0.85 * $\log_2$ (0.85)]

= –[0.15 * – 2.74 + 0.85 * – 0.23]

= 0.61

entropy(payment =electronic check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.45 * $\log_2$ (0.45) + 0.55 * $\log_2$ (0.55)]

= –[0.45 * – 1.15 + 0.55 * – 0.86]

= 1

entropy(payment =mailed check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.19 * $\log_2$ (0.19) + 0.81 * $\log_2$ (0.81)]

= –[0.19 * – 2.39 + 0.81 * – 0.3]

= 0.7

IG(churn & payment) = entropy(payment) – [p(payment = bank) * entropy (payment = bank) + p(payment = credit card) * entropy(payment = credit card) + p(payment = electronic check) * entropy(payment = electronic check) + p(payment = mailed check) * entropy(payment = mailed check)]

= 0.83 – [0.23 * 0.66 + 0.22 *0.61 + 0.34 * 1 + 0.23 * 0.7]

= 0.83 – 0.79

= 0.04

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ Contract(ประเภทสัญญา) มีค่ามากที่สุด (0.28) ดังนั้นจึงเลือกแอตทริบิวต์ Contract ขึ้นมาเป็นโหนด Root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนด มีคลาสคำตอบเดียวกัน จึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ Contract

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 กับจำนวนข้อมูลทั้งหมดโดยใช้ ผลลัพธ์เป็น Churn (การตัดสินใจความคงอยู่ของลูกค้าที่ใช้บริการต่อ)

1. คำนวณค่า IG ของแอตทริบิวต์ Contract และ PhoneService จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

   entropy(contract = month–to–month) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

   = –[0.43 * $\log_2$ (0.43) + 0.57 * $\log_2$ (0.57)]

   = –[0.43 * – 1.22 + 0.57 * – 0.81]

= 1

entropy(phone = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.43 * $\log_2$ (0.43) + 0.57* $\log_2$ (0.57)]

= –[0.43 * – 1.23 + 0.57 * – 0.81]

= 1

entropy(phone = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.41 * $\log_2$ (0.41) + 0.6 * $\log_2$ (0.6)]

= –[0.41 * – 1.29 + 0.6 * – 0.74]

= 1

IG(contract & phone) = entropy(contract & phone)–[p(phone= yes) *entropy (phone =yes) + p(phone = no) * entropy(phone = no)]

= 1 – [1 * 1 + 0.1 *1]

= 1 – 1

= 0

2. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract = month–to–month) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.43 * $\log_2$ (0.43) + 0.57 * $\log_2$ (0.57)]

= –[0.43 * – 1.22 + 0.57 * – 0.81]

= 1

entropy(internet = DSL) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.32 * $\log_2$ (0.32) + 0.68 * $\log_2$ (0.68)]

= –[0.32 * – 1.64 + 0.68 * – 0.56]

= 1

entropy(internet = fiber) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.55 * $\log_2$ (0.55) + 0.45 * $\log_2$ (0.45)]

= –[0.55 * – 0.86 + 0.45 * – 1.15]

= 1

entropy(internet = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.19 * $\log_2$ (0.19) + 0.81 * $\log_2$ (0.81)]

= –[0.19 * – 2.39 + 0.81 * – 0.3]

= 0

IG(contract & internet) = entropy(contract & internet) – [p(internet = DSL) * entropy (internet =DSL) + p(internet = fiber) * entropy(internet = fiber) + p(internet = no) * entropy(internet =no)]

= 1 – [0.32 * 1 + 0.55 *1 + 0.14 * 0.7]

= 1 – 0.96

= 0.03

3. คำนวณค่า IG ของแอตทริบิวต์ Contract และ PaperlessBilling จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract = month–to–month) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.43 * $\log_2$ (0.43) + 0.57 * $\log_2$ (0.57)]

= –[0.43 * – 1.22 + 0.57 * – 0.81]

= 1

entropy(paperbill = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.48 * $\log_2$ (0.48) + 0.52 * $\log_2$ (0.52)]

= –[0.48 * – 1 + 0.52 * – 1]

= 1

entropy(paperbill = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.32 * $\log_2$ (0.32) + 0.69 * $\log_2$ (0.69)]

= –[0.32 * – 1.64 + 0.69 * – 0.54]

= 1

IG(contract & paperbill) = entropy(contract & paperbill) – [p(paperbill = yes) * entropy(paperbill =yes) + p(paperbill = no) * entropy(paperbill = no)]

= 1 – [0.67 * 1 + 0.33 * 1]

= 1 – 1

= 0

4. คำนวณค่า IG ของแอตทริบิวต์ Contract และ PaymentMethod จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract = month–to–month) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.43 * $\log_2$ (0.43) + 0.57 * $\log_2$ (0.57)]

= –[0.43 * – 1.22 + 0.57 * – 0.81]

= 1

entropy(payment =bank) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.34 * $\log_2$ (0.34) + 0.66 * $\log_2$ (0.66)]

= –[0.34 * – 1.56 + 0.66 * – 0.6]

= 1

entropy(payment =credit card) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.33 * $\log_2$ (0.33) + 0.67 * $\log_2$ (0.67)]

= –[0.33 * – 1.6 + 0.67 * – 0.58]

= 1

entropy(payment =electronic check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.54 * $\log_2$ (0.54) + 0.46 * $\log_2$ (0.46)]

= –[0.54 * – 0.8 + 0.46 * – 1.12]

= 1

entropy(payment =mailed check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.32 * $\log_2$ (0.32) + 0.68 * $\log_2$ (0.68)]

= –[0.32 * – 1.46 + 0.68 * – 0.56]

= 1

IG(contract & payment) = entropy(contract & payment) – [p(payment = bank) * entropy (payment = bank) + p(payment = credit card) * entropy(payment = credit card) + p(payment = electronic check) * entropy(payment = electronic check) + p(payment = mailed check) * entropy(payment = mailed check)]

= – [0.15 * 1 + 0.14 *1 + 0.48 * 1 + 0.23 * 1]

= 1 – 1

= 0

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ InternetService (สมัครบริการอินเตอร์เน็ต) มีค่ามากที่สุด (0.03) ดังนั้นจึงเลือกแอตทริบิวต์ InternetService ขึ้นมาเป็นโหนด Root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนด มีคลาสคำตอบเดียวกัน จึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ InternetService

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ ผลลัพธ์เป็น Churn (การตัดสินใจความคงอยู่ของลูกค้าที่ใช้บริการต่อ)

1. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService และ PhoneService จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract & internet = DSL) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.32 * $\log_2$ (0.32) + 0.68 * $\log_2$ (0.68)]

= –[0.32 * – 1.64 + 0.68 * – 0.56]

= 1

entropy(phone = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.29 * $\log_2$ (0.29) + 0.72* $\log_2$ (0.72)]

= –[0.29 * – 1.79 + 0.72 * – 0.47]

= 0.9

entropy(phone = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.41 * $\log_2$ (0.41) + 0.59 * $\log_2$ (0.59)]

= –[0.41 * – 1.29 + 0.59 * – 0.76]

= 1

IG(contract & internet & phone) = entropy(contract & internet & phone)– [p(phone=yes)*entropy (phone =yes) + p(phone = no) * entropy(phone = no)]

= 1 – [0.7 * 0.9 + 0.3 *1]

= 1 – 0.9

= 0.08

2. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService และ PaperlessBilling จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract & internet = DSL) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.32 * $\log_2$ (0.32) + 0.68 * $\log_2$ (0.68)]

= –[0.32 * – 1.64 + 0.68 * – 0.56]

= 1

entropy(paperbill = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.34 * $\log_2$ (0.34) + 0.65 * $\log_2$ (0.65)]

= –[0.34 * – 1.56 + 0.65 * – 0.62]

= 1

entropy(paperbill = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.29 * $\log_2$ (0.29) + 0.71 * $\log_2$ (0.71)]

= –[0.29 * – 1.79 + 0.71 * – 0.49]

= 1

IG(contract & internet & paperbill) = entropy(contract & internet & paperbill) – [p(paperbill = yes) * entropy(paperbill = yes) + p(paperbill = no) * entropy(paperbill = no)]

= 1 – [0.6 * 1 + 0.4 * 1]

= 1 – 1

= 0

3. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService และ PaymentMethod จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract & internet = DSL) = $-p(\text{ใช้ต่อ})*\log_2 p(\text{ใช้ต่อ}) + p(\text{ไม่ใช้ต่อ})*\log_2 p(\text{ไม่ใช้ต่อ})$

$= -[0.32 * \log_2 (0.32) + 0.68 * \log_2 (0.68)]$

$= -[0.32 * -1.64 + 0.68 * -0.56]$

$= 1$

entropy(payment =bank) = $-p(\text{ใช้ต่อ})*\log_2 p(\text{ใช้ต่อ}) + p(\text{ไม่ใช้ต่อ})*\log_2 p(\text{ไม่ใช้ต่อ})$

$= -[0.2 * \log_2 (0.2) + 0.8 * \log_2 (0.8)]$

$= -[0.2 * -2.32 + 0.8 * -0.32]$

$= 1$

entropy(payment =credit card) = $-p(\text{ใช้ต่อ})*\log_2 p(\text{ใช้ต่อ}) + p(\text{ไม่ใช้ต่อ})*\log_2 p(\text{ไม่ใช้ต่อ})$

$= -[0.27 * \log_2 (0.27) + 0.73 * \log_2 (0.73)]$

$= -[0.27 * -1.89 + 0.73 * -0.45]$

$= 1$

entropy(payment =electronic check) = $-p(\text{ใช้ต่อ})*\log_2 p(\text{ใช้ต่อ}) + p(\text{ไม่ใช้ต่อ})*\log_2 p(\text{ไม่ใช้ต่อ})$

$= -[0.4 * \log_2 (0.4) + 0.59 * \log_2 (0.59)]$

$= -[0.4 * -1.32 + 0.59 * -0.76]$

$= 1$

entropy(payment =mailed check) = $-p(\text{ใช้ต่อ})*\log_2 p(\text{ใช้ต่อ}) + p(\text{ไม่ใช้ต่อ})*\log_2 p(\text{ไม่ใช้ต่อ})$

$= -[0.31 * \log_2 (0.31) + 0.69 * \log_2 (0.69)]$

= –[0.31 * – 1.69 + 0.69 * – 0.54]

= 1

IG(contract & internet & payment) = entropy(contract & internet & payment) – [p(payment = bank) * entropy (payment = bank) + p(payment = credit card) * entropy(payment = credit card) + p(payment = electronic check) * entropy(payment = electronic check) + p(payment = mailed check) * entropy(payment = mailed check)]

= – [0.16 * 1 + 0.15 *1 + 0.45 * 1 + 0.3 * 1]

= 1 – 1

= 0

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ PhoneService (สมัครบริการโทรศัพท์) มีค่ามากที่สุด (0.08) ดังนั้นจึงเลือกแอตทริบิวต์ PhoneService ขึ้นมาเป็นโหนด Root และจะต้องทำการแตกกิ่งจาก โหนด root ออกไปจนข้อมูลในแต่ละโหนด มีคลาสคำตอบเดียวกัน จึงทำการสร้างโหนดในระดับถัดไปของแอตทริบิวต์ PhoneService

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 4 กับจำนวนข้อมูลทั้งหมดโดยใช้ ผลลัพธ์เป็น Churn (การตัดสินใจความคงอยู่ของลูกค้าที่ใช้บริการต่อ)

1. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService และ PhoneService และ PaymentMethod จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy(contract & Internet &phone = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)* $\log_2$ p(ไม่ใช้ต่อ)

= –[0.41 * $\log_2$ (0.41) + 0.59 * $\log_2$ (0.59)]

= –[0.41 * – 1.29 + 0.59 * – 0.76]

= 1

entropy(payment =bank) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.33 * $\log_2$ (0.33) + 0.67 * $\log_2$ (0.67)]

= –[0.33 * – 1.6 + 0.67 * – 0.57]

= 1

entropy(payment =credit card) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.31 * $\log_2$ (0.31) + 0.68 * $\log_2$ (0.68)]

= –[0.31 * – 1.69 + 0.68 * – 0.56]

= 1

entropy(payment =electronic check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.5 * $\log_2$ (0.5) + 0.5 * $\log_2$ (0.5)]

= –[0.5 * – 1 + 0.5 * – 1]

= 1

Entropy (payment =mailed check) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.34 * $\log_2$ (0.34) + 0.66 * $\log_2$ (0.66)]

= –[0.34 * – 1.56 + 0.66 * – 0.6]

= 1

IG (contract & internet &phone & payment) = entropy (contract & internet &phone & payment) – [p (payment = bank) * entropy (payment = bank) + p(payment = credit card) * entropy(payment = credit card) + p(payment = electronic check) * entropy(payment = electronic check) + p(payment = mailed check) * entropy(payment = mailed check)]

= –[0.14 * 1 + 0.15 *1 + 0.47 * 1 + 0.24 * 1]

= 1 – 0.9

= 0.01

2. คำนวณค่า IG ของแอตทริบิวต์ Contract และ InternetService และ PhoneService และ PaperlessBilling จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

   Entropy (contract & Internet &phone = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)* $\log_2$ p(ไม่ใช้ต่อ)

   = –[0.41 * $\log_2$ (0.41) + 0.59 * $\log_2$ (0.59)]

= –[0.41 * – 1.29 + 0.59 * – 0.76]

= 1

entropy(paperbill = yes) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.44 * $\log_2$ (0.44) + 0.56 * $\log_2$ (0.56)]

= –[0.44 * – 1.18 + 0.56 * – 0.84]

= 1

entropy(paperbill = no) = –p(ใช้ต่อ)*$\log_2$ p(ใช้ต่อ) + p(ไม่ใช้ต่อ)*$\log_2$ p(ไม่ใช้ต่อ)

= –[0.34 * $\log_2$ (0.34) + 0.66 * $\log_2$ (0.6)]

= –[0.34 * – 1.56 + 0.66 * – 0.6]

= 1

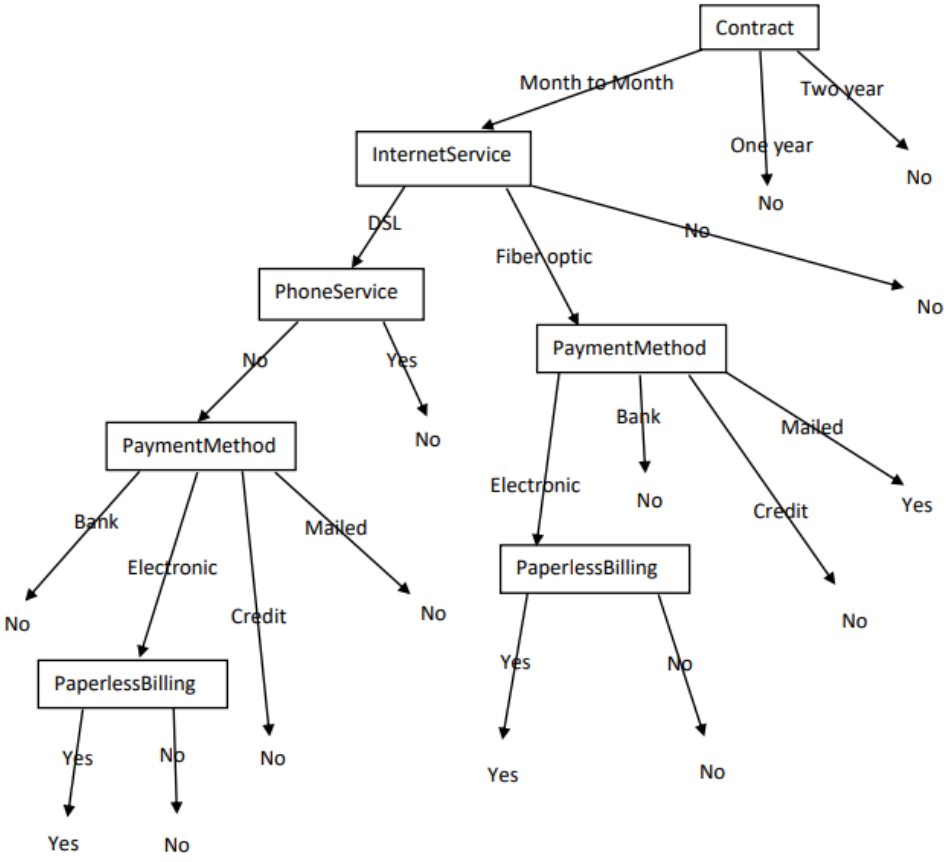
IG(contract & internet &phone & paperbill) = entropy(contract & internet &phone & paperbill) – [p(paperbill =yes)* entropy(paperbill =yes) + p(paperbill = no) * entropy(paperbill = no)]

= 1 – [0.6 * 1 + 0.36 * 1]

= 1 – 1

= 0

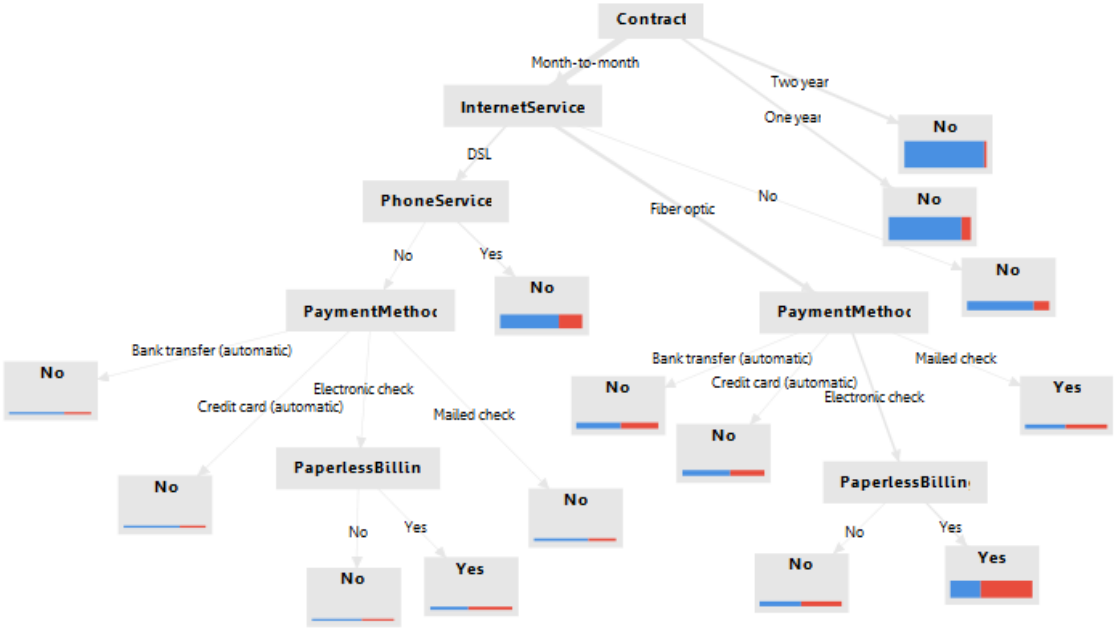
จากการคำนวณค่า IG ของแอตทริบิวต์ Contract(ประเภทสัญญา) และ InternetService(บริการสมัครอินเตอร์เน็ต) และ PhoneService(บริการสมัครโทรศัพท์) และ PaymentMethod(วิธีชำระเงิน) และ PaperlessBilling(การเรียกเก็บเงินแบบไร้กระดาษ) มีค่ามากที่สุด (0.01) ดังนั้นจึงเลือกแอตทริบิวต์ PaymentMethod ขึ้นมาเป็นโหนดในระดับที่ 4 ต่อจากโหนดระดับที่ 3 และ โหนดสุดท้ายคือแอตทริบิวต์ PaperlessBilling ทำการแตกกิ่งจากโหนดในระดับที่ 3 ออกไป พบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ผลลัพธ์เป็น Churn (การตัดสินใจความคงอยู่ของลูกค้าในการใช้บริการต่อ)



ภาพที่ 3.6 รูปแบบโมเดล Decision Tree จากการคำนวนมือ

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไป ใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรกว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่ผู้วิเคราะห์ได้ทำการทดลองโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (Testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Repaid Miner 9.0 และโปรแกรม Weka 3.9.6 ได้ดังนี้

1. โปรแกรม RapidMiner Studio 9.0



ภาพที่ 3.7 รูปแบบโมเดล Decision Tree ในโปรแกรม RapidMiner

Tree

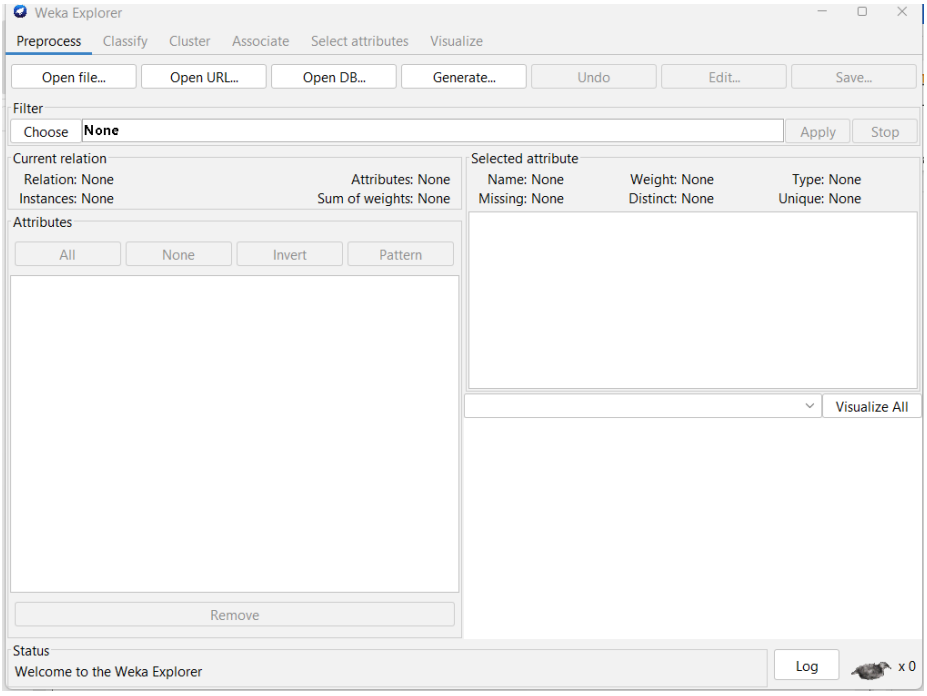
```
Contract = Month-to-month
|   InternetService = DSL
|   |   PhoneService = No
|   |   |   PaymentMethod = Bank transfer (automatic): No {No=35, Yes=17}
|   |   |   PaymentMethod = Credit card (automatic): No {No=39, Yes=18}
|   |   |   PaymentMethod = Electronic check
|   |   |   |   PaperlessBilling = No: No {No=28, Yes=18}
|   |   |   |   PaperlessBilling = Yes: Yes {No=60, Yes=69}
|   |   |   PaymentMethod = Mailed check: No {No=61, Yes=31}
|   |   PhoneService = Yes: No {No=606, Yes=241}
|   InternetService = Fiber optic
|   |   PaymentMethod = Bank transfer (automatic): No {No=178, Yes=149}
|   |   PaymentMethod = Credit card (automatic): No {No=171, Yes=122}
|   |   PaymentMethod = Electronic check
|   |   |   PaperlessBilling = No: No {No=125, Yes=122}
|   |   |   PaperlessBilling = Yes: Yes {No=393, Yes=667}
|   |   PaymentMethod = Mailed check: Yes {No=99, Yes=102}
|   InternetService = No: No {No=425, Yes=99}
Contract = One year: No {No=1307, Yes=166}
Contract = Two year: No {No=1647, Yes=48}
```

ภาพที่ 3.8 คำบรรยายลักษณะงาน Decision Tree ของ RapidMiner

2. โปรแกรม Weka 3.9.6 ซึ่งมีขั้นตอนการทำงาน ดังนี้

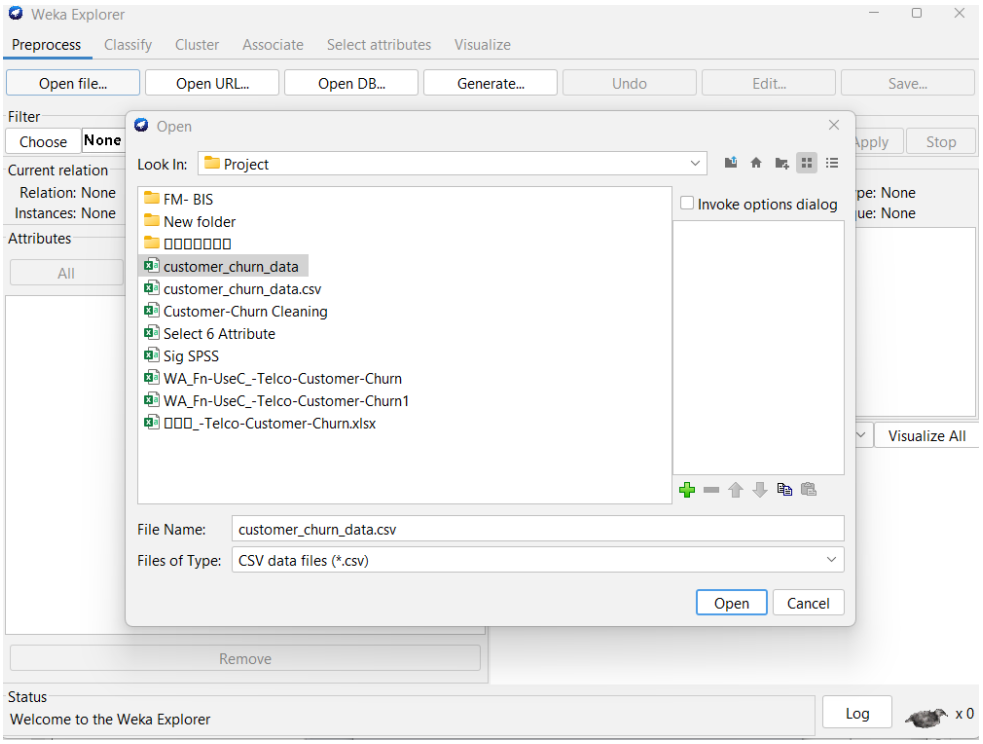
ขั้นตอนที่ 1 เปิดโปรแกรม Weka 3.9.6

ภาพที่ 3.9 แสดงการเปิดโปรแกรม Weka 3.9.6

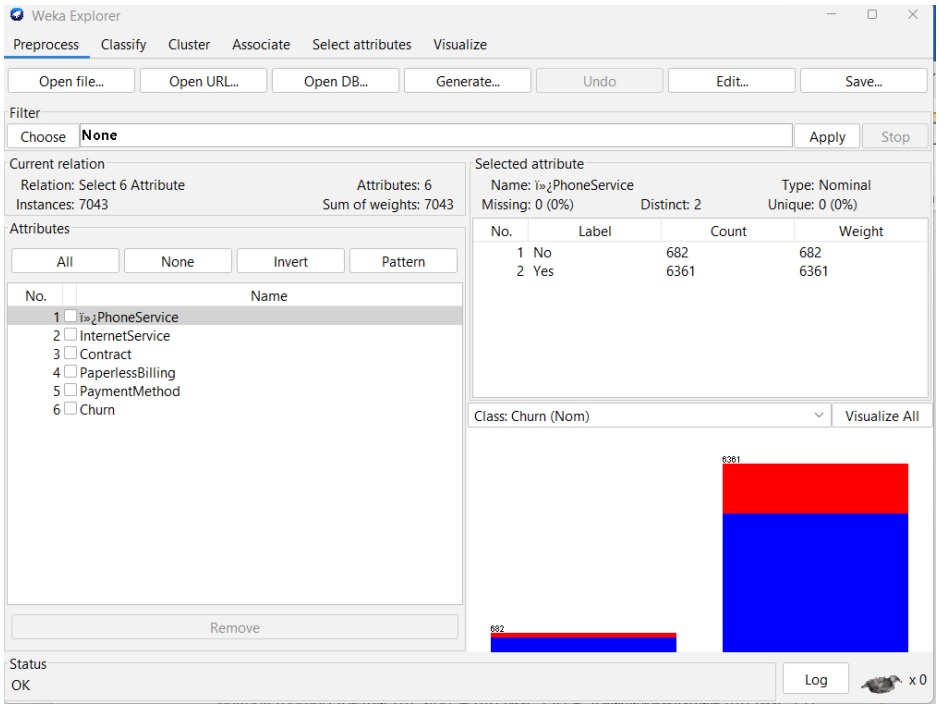


ภาพที่ 3.10 แสดงการเข้าใช้โปรแกรม Weka 3.9.6

ขั้นตอนที่ 2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Application > Explorer > Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ และหลังจากนั้นโปรแกรมแสดง หน้าจอข้อมูลตามภาพที่

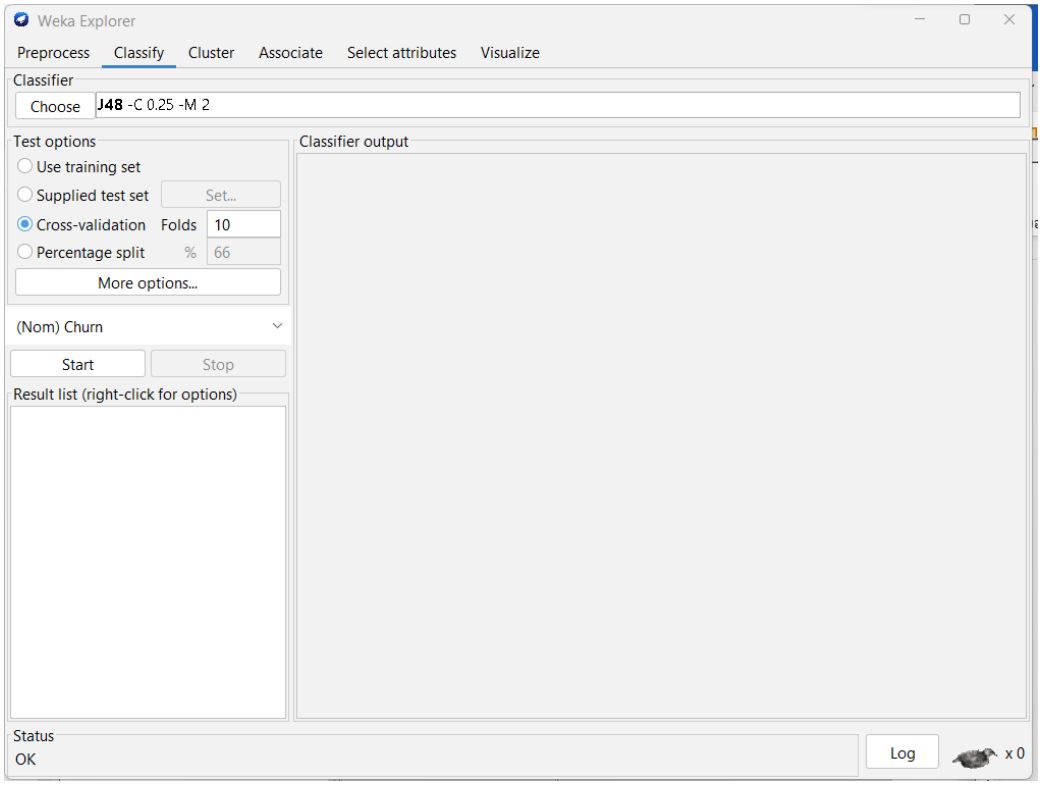


ภาพที่ 3.11 แสดงการนำไฟล์ข้อมูลเข้าโปรแกรม Weka 3.9.6

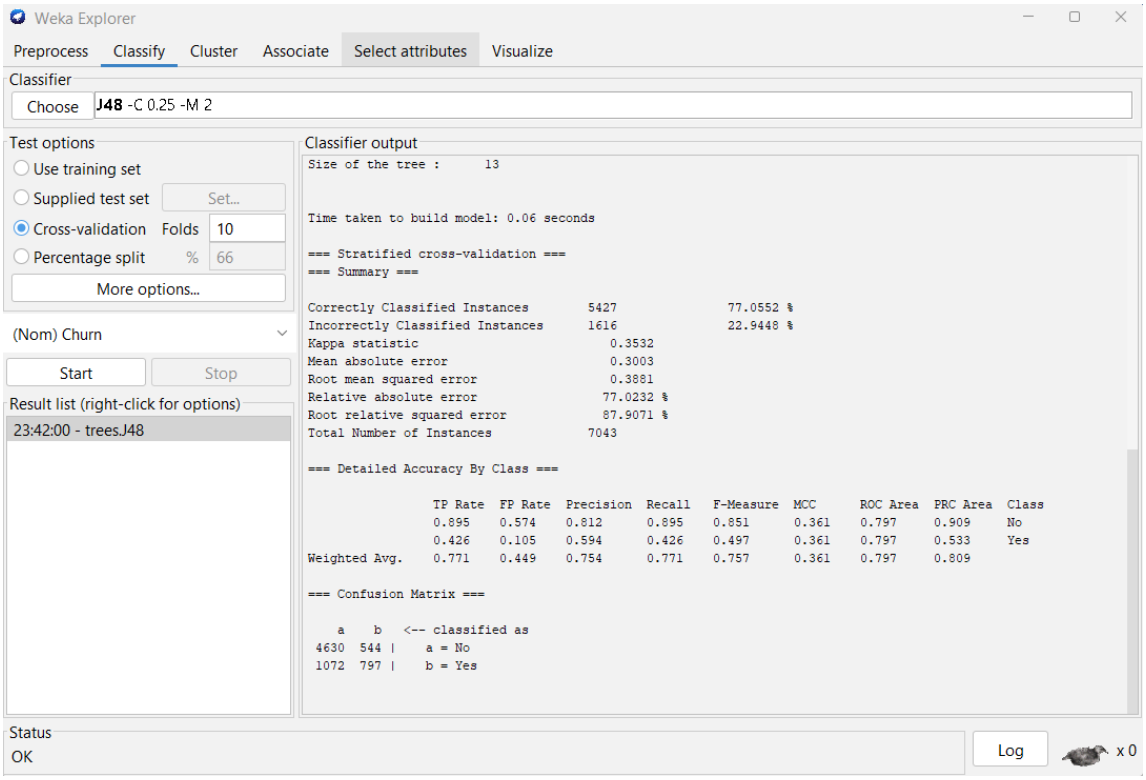


ภาพที่ 3.12 แสดงข้อมูลหลังจากนำเข้าโปรแกรม Weka 3.9.6

ขั้นตอนที่ 3 ดำเนินการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree โดย เลือกที่ Classification > Choose > tree และเลือกรูปแบบเป็น J48 ตามภาพที่ จากนั้นทำการเลือก entropy แล้วกด ปุ่ม Start ตามภาพที่ จะแสดงผลลัพธ์ที่ได้ตามภาพที่



ภาพที่ 3.13 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree



ภาพที่ 3.14 แสดงหน้าจอผลลัพธ์ของโมเดลการจัดกลุ่มข้อมูลแบบ Decision Tree: J48

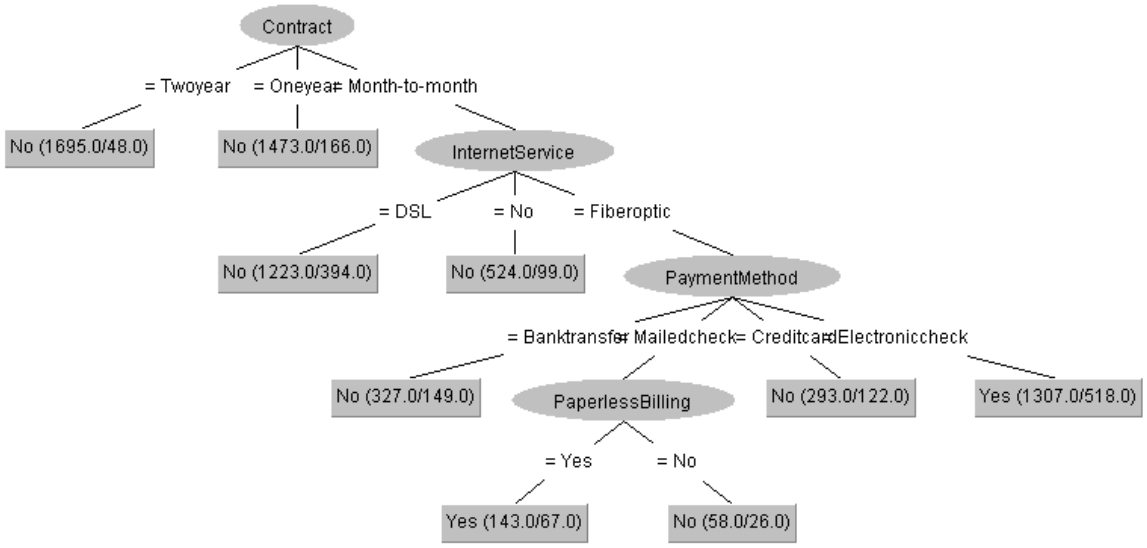
จากผลลัพธ์การทดลองพบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกความคงอยู่ของลูกค้า Yes และ No มีความถูกต้องถึง 77.0552%

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

Contract = Twoyear: No (1695.0/48.0)
Contract = Oneyear: No (1473.0/166.0)
Contract = Month-to-month
|   InternetService = DSL: No (1223.0/394.0)
|   InternetService = No: No (524.0/99.0)
|   InternetService = Fiberoptic
|   |   PaymentMethod = Banktransfer(automatic): No (327.0/149.0)
|   |   PaymentMethod = Mailedcheck
|   |   |   PaperlessBilling = Yes: Yes (143.0/67.0)
|   |   |   PaperlessBilling = No: No (58.0/26.0)
|   |   PaymentMethod = Creditcard(automatic): No (293.0/122.0)
|   |   PaymentMethod = Electroniccheck: Yes (1307.0/518.0)
```

ภาพที่ 3.15 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจที่แตกกิ่งออกมา



ภาพที่ 3.16 แสดงรูปแบบโมเดล Graph Decision Tree ในโปรแกรม Weka 3.9.6

หลังจากผู้วิเคราะห์เลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (model) และ

ข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลความคงอยู่ของลูกค้าที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งผู้วิเคราะห์เลือกโปรแกรมที่ใช้นำเสนอ คือ โปรแกรม RapidMiner พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้น มีค่าความถูกต้องเฉลี่ยในทุกโมเดล เท่ากับ 77.05% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 22.94% และมีค่าความคลาดเคลื่อน เท่ากับ 0.4118 และ เมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.17 พบว่าการหาค่า ของข้อมูลค่าจริง กับ จำนวนข้อมูลจากการทำนาย แบ่งตามความคงอยู่ของลูกค้า อยู่ต่อ (Yes) และ ไม่ใช้บริการต่อ(No) และนำมาหาค่าเฉลี่ยรวมของทุก Class ได้ค่าเฉลี่ยรวม เท่ากับ 0.2748 มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้

ดังนั้นผู้วิเคราะห์จึงเลือกใช้โมเดล RapidMiner ในการนำไปหาแนวทางการ วางแผนการจัดการและรับมือต่อปัจจัยเสี่ยงต่าง ๆ ที่เป็นปัจจัยชี้นำประสิทธิภาพการทำงานขององค์กร เพราะมีค่าความถูกต้องของโมเดล และค่าเฉลี่ย Confusion Matrix จากการแบ่งตามความคงอยู่ของลูกค้า อยู่ต่อ(Yes) และ ไม่ใช้บริการต่อ(No) อยู่ในระดับที่ค่อนข้างดี

จากผลสรุปจะเห็นได้ว่าการสร้างโมเดล Decision Tree จากโปรแกรม RapidMiner ผู้จัดทำได้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 14 กฎ และสามารถนำกฎที่ได้นำไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ ดังนี้

กฎข้อที่ 1 IF Contract = Month-to-month InternetService = DSL PhoneService = No PaymentMethod = Bank transfer หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์และไม่สมัครบริการโทรศัพท์และชำระเงินแบบเงินสด ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 2 IF Contract = Month-to-month InternetService = DSL PhoneService = No PaymentMethod = Credit card หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์และไม่สมัครบริการโทรศัพท์และชำระเงินแบบบัตรเครดิต ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 3 IF Contract = Month-to-month InternetService = DSL PhoneService = No PaymentMethod = Electronic check PaperlessBilling = No หมายความว่า ถ้าประเภทสัญญาเป็น

เดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์ และไม่สมัครบริการโทรศัพท์และชำระเงินแบบบัตรเครดิตและลูกค้าไม่เก็บเงินแบบไร้กระดาษ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 4 IF Contract = Month-to-month InternetService = DSL PhoneService = No PaymentMethod = Electronic check PaperlessBilling = Yes หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์ และไม่สมัครบริการโทรศัพท์และชำระเงินแบบบัตรเครดิตและลูกค้าเก็บเงินแบบไร้กระดาษ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าใช้บริการต่อ

กฎข้อที่ 5 IF Contract = Month-to-month InternetService = DSL PhoneService = No PaymentMethod = Mailed check หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์ และไม่สมัครบริการโทรศัพท์และชำระเงินแบบเช็คและลูกค้าเก็บเงินแบบไร้กระดาษ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 6 IF Contract = Month-to-month InternetService = DSL PhoneService = Yes หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายโทรศัพท์และสมัครบริการโทรศัพท์ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าใช้บริการต่อ

กฎข้อที่ 7 IF Contract = Month-to-month InternetService = Fiber optic PaymentMethod = Bank transfer หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายไฟเบอร์ออฟติก และชำระเงินแบบเงินสด ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 8 IF Contract = Month-to-month InternetService = Fiber optic PaymentMethod = Credit card หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายไฟเบอร์ออฟติก และชำระเงินแบบบัตรเครดิต ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 9 IF Contract = Month-to-month InternetService = Fiber optic PaymentMethod = Electronic check PaperlessBilling = No หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและ

สมัครอินเตอร์เน็ตแบบสายไฟเบอร์ออฟติก และชำระเงินแบบโอนเงินและลูกค้าไม่เก็บเงินแบบไร้กระดาษ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 10 IF Contract = Month-to-month InternetService = Fiber optic PaymentMethod = Electronic check PaperlessBilling = Yes หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายไฟเบอร์ออฟติก และชำระเงินแบบโอนเงินและลูกค้าเก็บเงินแบบไร้กระดาษ ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าใช้บริการต่อ

กฎข้อที่ 11 IF Contract = Month-to-month InternetService = Fiber optic PaymentMethod = Mailed check หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและสมัครอินเตอร์เน็ตแบบสายไฟเบอร์ออฟติก และชำระเงินแบบเช็ค ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าใช้บริการต่อ

กฎข้อที่ 12 IF Contract = Month-to-month InternetService = No หมายความว่า ถ้าประเภทสัญญาเป็นเดือนต่อเดือนและไม่สมัครอินเตอร์เน็ต ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 13 IF Contract = One year หมายความว่า ถ้าประเภทสัญญาเป็นหนึ่งปี ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

กฎข้อที่ 14 IF Contract = Two year หมายความว่า ถ้าประเภทสัญญาเป็นสองปี ผลการพิจารณาพบว่า ความคงอยู่ของลูกค้าไม่ใช้บริการต่อ

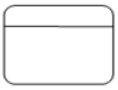
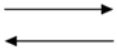
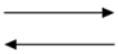
3.1.6 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผนกำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์แล้ว แสดงผลข้อมูลบน Web Browser โดยใช้ชุดคำสั่ง CSS3 ร่วมกับการนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพ

3.2 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูล (DFD) ย่อมาจาก Data Flow Diagram เป็นเครื่องมือเชิงโครงสร้างที่ใช้บรรยายภาพรวมของระบบโดยแสดงขั้นตอนการทำงานของระบบหรือโพรเซส(process) ระบุแหล่งกำเนิดของข้อมูล การไหลของข้อมูล ปลายทางข้อมูล การเก็บข้อมูลและการประมวลผลข้อมูล โดยจะช่วยแสดงแผนภาพ ว่าข้อมูลมาจากไหน จะไปไหน เก็บข้อมูลไว้ที่ไหน มีอะไรเกิดขึ้นกับข้อมูลระหว่างทางเรียกว่าแผนภาพกระแสข้อมูลหรือ แผนภาพแสดงความเคลื่อนไหวของข้อมูล

สัญลักษณ์ที่ใช้ในการเขียนแผนภาพกระแสข้อมูลนั้น ประกอบด้วย 4 สัญลักษณ์ คือ การประมวลผล (Process) เส้นทางการไหลของข้อมูล (Data Flows) ตัวแทนข้อมูล (External Agents) และแหล่งจัดเก็บข้อมูล (Data Store) โดยได้มีการศึกษาคิดค้นพัฒนาวิธีการอยู่หลายแบบ แต่ที่เป็นมาตรฐานมี 2 กลุ่ม ได้แก่ ชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย Gane and Sarson (1979) และชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย DeMarco and Yourdon (DeMarco, 1979; Yourdon and Constantine,1979) โดยมีสัญลักษณ์ ดังตารางที่ 3.2

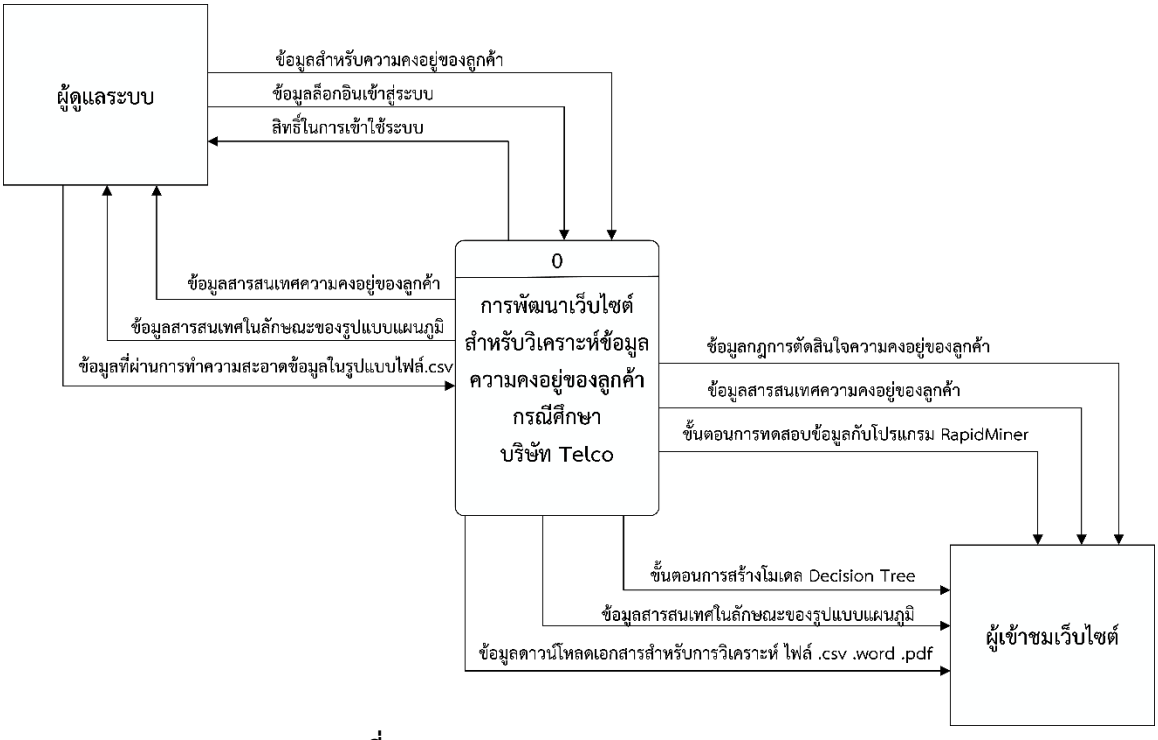
ตารางที่ 3.2 สัญลักษณ์ที่ใช้ในการเขียนแผนการไหลของข้อมูล

DeMarco & Yourdon	Gane & Sarson	ความหมาย
		Process : ขั้นตอนการทำงานภายในระบบ
		Data Store : แหล่งข้อมูลสามารถเป็นได้ทั้งไฟล์ข้อมูลและฐานข้อมูล (File or Database)
		External Agent : ปัจจัยหรือสภาพแวดล้อมที่มีผลกระทบต่อระบบ
		Data Flow : เส้นทางการไหลของข้อมูล แสดงทิศทางของข้อมูลจากขั้นตอนการทำงานหนึ่งไปยังอีกขั้นตอนหนึ่ง

ที่มา : https://sites.google.com/site/krittiyaporn8345/4-2-dfd (2551)

3.2.1 แผนภาพบริบท Context Diagram

แผนภาพบริบท (Context Diagram) จะแสดงภาพโดยรวมระบบของการพัฒนาระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า ดังภาพที่ 3.9



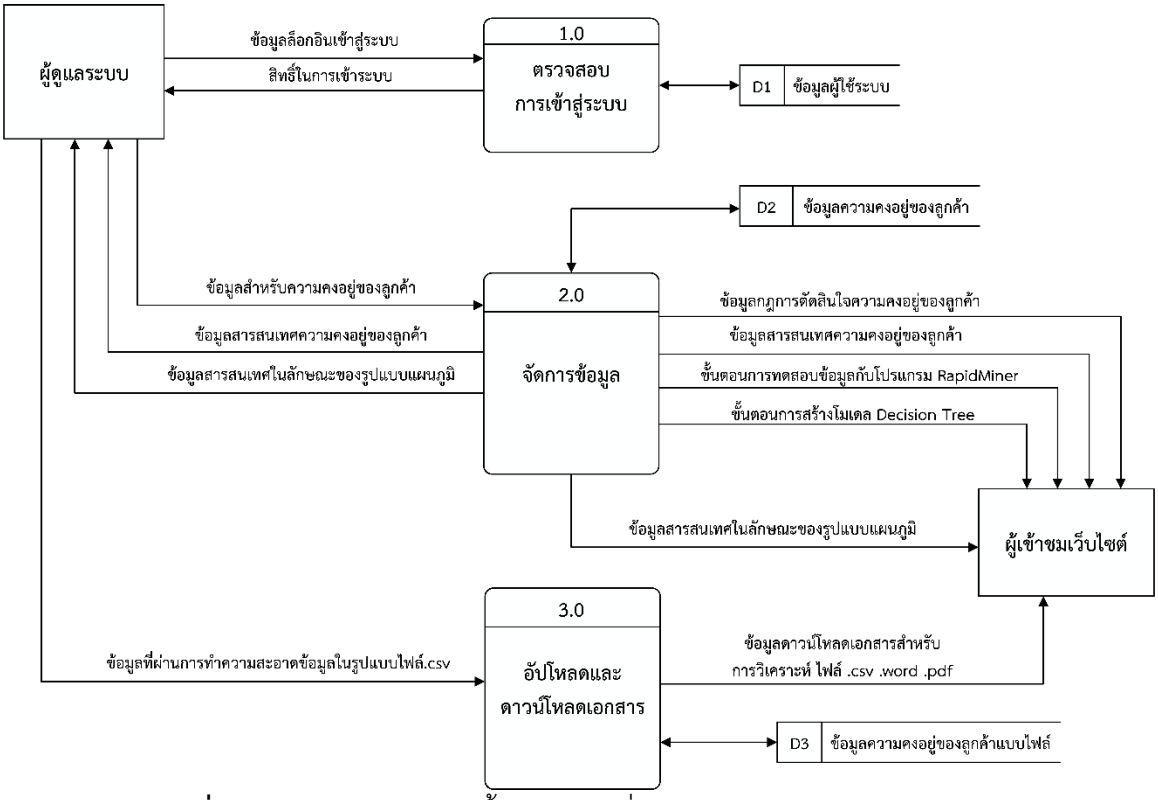
ภาพที่ 3.17 แผนภาพบริบท Context Diagram

จากรูปภาพที่ 3.9 เป็นแผนภาพบริบทระบบของระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้าโดยสามารถแบ่งผู้ใช้ออกเป็น 2 ประเภทดังนี้

1) **ผู้ดูแลระบบ (System Administrator)** สามารถล็อกอินเข้าสู่ระบบได้ สามารถเรียกดูข้อมูลความคงอยู่ของลูกค้าได้ สามารถอัปโหลดข้อมูลที่ผ่านการทำความสะอาดในรูปแบบไฟล์ .CSVได้ สามารถจัดการข้อมูลความคงอยู่ของลูกค้าได้ และสามารถเรียกดูข้อมูลในลักษณะของรูปแบบแผนภูมิได้

2) **ผู้เข้าชมเว็บไซต์ (Website visitor)** สามารถเรียกดูข้อมูลความคงอยู่ของลูกค้าได้ สามารถเรียกดูข้อมูลกฎการตัดสินใจความคงอยู่ของลูกค้าได้ สามารถดาวน์โหลดเอกสารบนเว็บไซต์ได้ สามารถดูขั้นตอนการทดสอบข้อมูลกับโปรแกรม Rapid Miner สามารถเรียกดูขั้นตอนการสร้างโมเดล Decision Tree และและสามารถเรียกดูข้อมูลในลักษณะของรูปแบบแผนภูมิได้

3.2.2 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)



ภาพที่ 3.18 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)

จากภาพที่ 3.18 แสดงแผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0) ระบบบริหารฐานข้อมูลความคงอยู่ของลูกค้า ประกอบด้วย 3 กระบวนการหลัก ดังนี้

ตารางที่ แสดงคำอธิบายกระบวนการตรวจสอบการเข้าระบบ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	1
Process Name	ตรวจสอบการเข้าระบบ
Input Data Flow	ชื่อผู้ใช้และรหัสผ่าน ข้อมูลสิทธิ์การเข้าใช้ระบบ ข้อมูลผู้ดูแลระบบ
Output Data Flow	สิทธิ์ในการเข้าถึงระบบ สิทธิ์การเข้าใช้ระบบ ข้อมูลผู้ดูแลระบบ
Data Store Used	ข้อมูลผู้ใช้และรหัสผ่าน ข้อมูลผู้ดูแลระบบ

Description	เป็นกระบวนการสำหรับตรวจสอบ และกำหนดสิทธิ์ในการเข้าใช้ระบบ โดยระบบจะทำการตรวจสอบชื่อผู้ใช้และรหัสผ่าน ผู้มีสิทธิ์เข้าใช้ระบบโดยมีชื่อผู้ใช้ระบบได้แก่ ผู้ดูแลระบบ

ตารางที่ แสดงคำอธิบายกระบวนการจัดการข้อมูลความคงอยู่ของลูกค้า

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	2
Process Name	จัดการข้อมูลความคงอยู่ของลูกค้า
Input Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Output Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Data Store Used	ข้อมูลความคงอยู่ของลูกค้า
Description	เป็นกระบวนการสำหรับจัดการข้อมูลความคงอยู่ของลูกค้าได้แก่ การเพิ่มข้อมูล ลบข้อมูล แก้ไขข้อมูล ผู้มีสิทธิ์ใช้งานกระบวนการนี้ ได้แก่ ผู้ดูแลระบบ และผู้เข้าชมเว็บไซต์

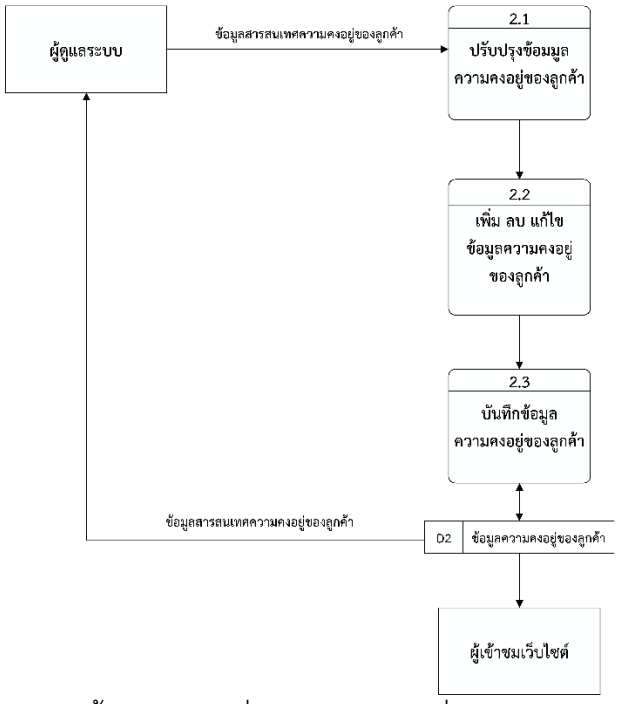
ตารางที่ แสดงคำอธิบายกระบวนการจัดการอัปโหลดไฟล์ข้อมูลความคงอยู่ของลูกค้า

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	3
Process Name	จัดการอัปโหลดไฟล์เอกสารข้อมูลความคงอยู่ของลูกค้า
Input Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Output Data Flow	ข้อมูลความคงอยู่ของลูกค้าเป็นไฟล์ .CSV และข้อมูลเอกสารที่เกี่ยวข้องที่เป็นไฟล์ .CSV
Data Store Used	ข้อมูลความคงอยู่ของลูกค้า

Description	เป็นกระบวนการสำหรับจัดการดาวน์โหลดไฟล์เอกสารต่าง ๆเช่น ข้อมูลความคงอยู่ของลูกค้าที่ผู้เยี่ยมชมเว็บไซต์สามารถดาวน์โหลดได้

จากแผนภาพกระแสข้อมูลระดับที่ 0 กระบวนการที่ 2 สามารถแยกย่อยเป็นกระบวนการย่อยระดับที่ 1 ได้ดังนี้

แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2.1, 2.2, 2.3 จัดการข้อมูลความคงอยู่ของลูกค้า



ภาพที่ 3.19 แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2.1, 2.2, 2.3 จัดการข้อมูลความคงอยู่ของลูกค้า

ตารางที่ แสดงคำอธิบายกระบวนการที่ 2.1 ปรับปรุงข้อมูลความคงอยู่ของลูกค้า

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	2.1
Process Name	ปรับปรุงข้อมูลความคงอยู่ของลูกค้า

Input Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Output Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Data Store Used	ข้อมูลความคงอยู่ของลูกค้า
Description	เป็นกระบวนการสำหรับการปรับปรุงข้อมูลความคงอยู่ของลูกค้าโดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

ตารางที่ แสดงคำอธิบายกระบวนการที่ 2.2 เพิ่ม ลบ แก้ไขข้อมูลความคงอยู่ของลูกค้า

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	2.2
Process Name	เพิ่ม ลบ แก้ไขข้อมูลความคงอยู่ของลูกค้า
Input Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Output Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Data Store Used	ข้อมูลความคงอยู่ของลูกค้า
Description	เป็นกระบวนการสำหรับการ เพิ่ม ลบ แก้ไขข้อมูลความคงอยู่ของลูกค้าโดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

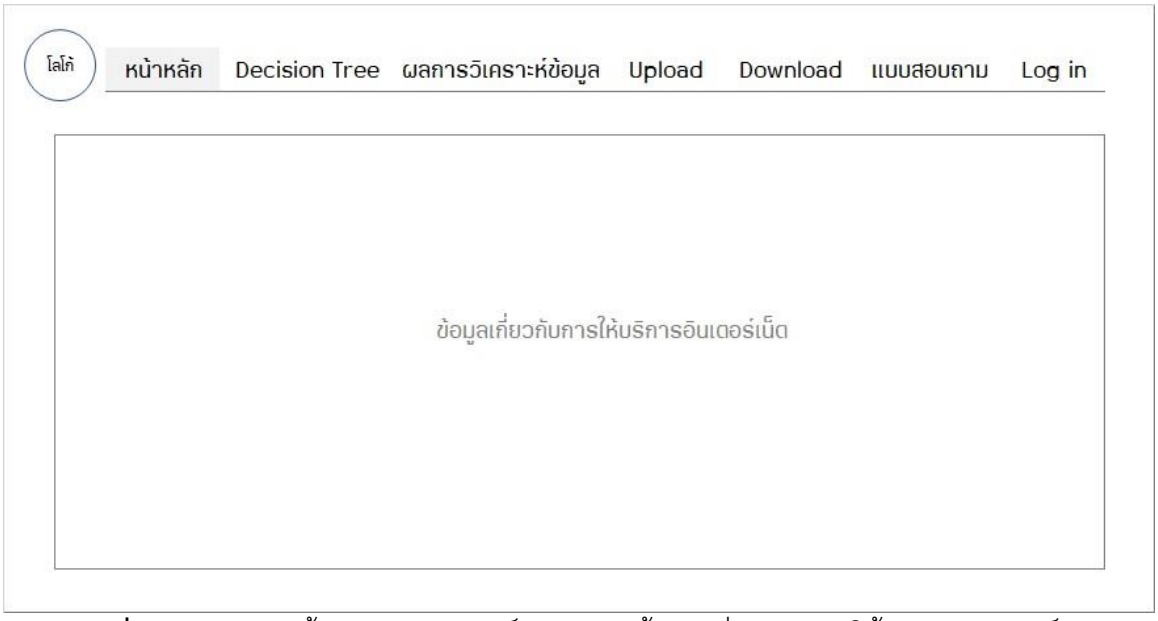
ตารางที่ แสดงคำอธิบายกระบวนการที่ 2.3 บันทึกข้อมูลความคงอยู่ของลูกค้า

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลความคงอยู่ของลูกค้า
DFD Number	2.3
Process Name	บันทึกข้อมูลความคงอยู่ของลูกค้า
Input Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Output Data Flow	ข้อมูลความคงอยู่ของลูกค้า
Data Store Used	ข้อมูลความคงอยู่ของลูกค้า

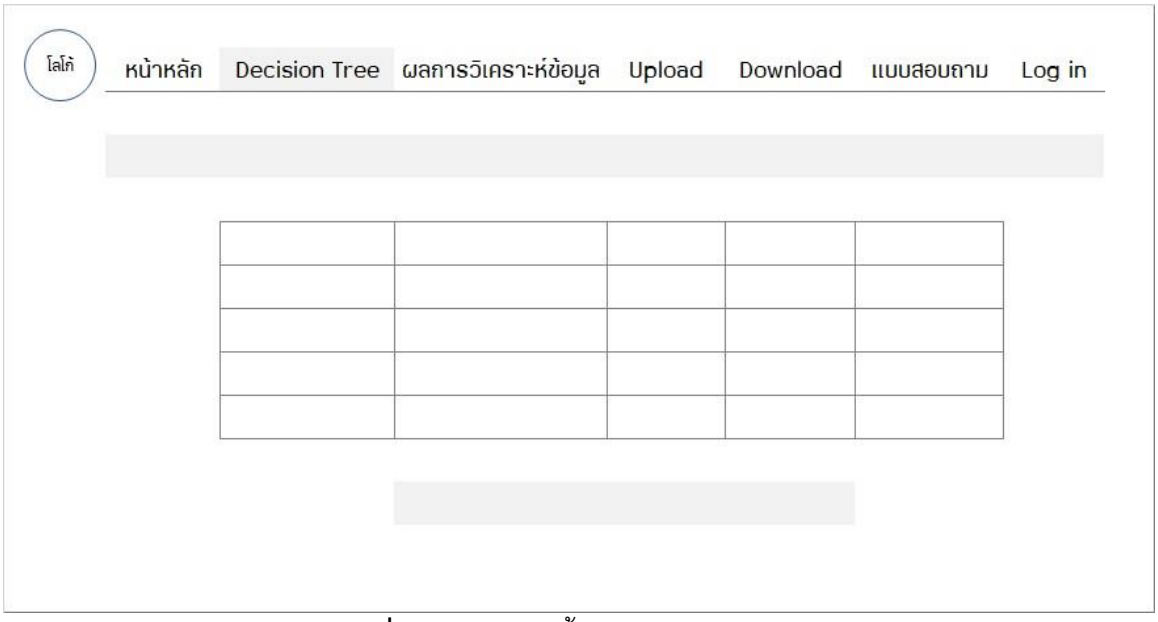
Description	เป็นกระบวนการสำหรับการบันทึกข้อมูลความคงอยู่ของลูกค้าเป็นไฟล์.CSV โดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

3.3 การออกแบบเว็บไซต์

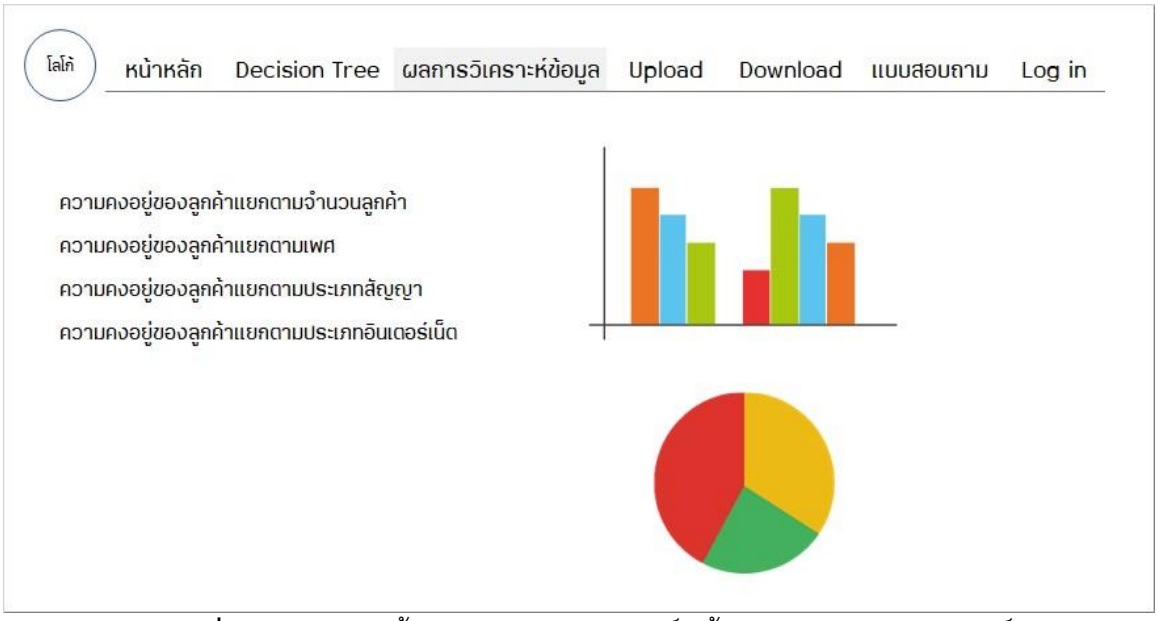
3.3.1 การออกแบบ Wireframe หน้าจอเว็บไซต์



ภาพที่ 3.20 แสดงหน้าแรกของเว็บไซต์และแสดงข้อมูลเกี่ยวกับการให้บริการอินเตอร์เน็ต



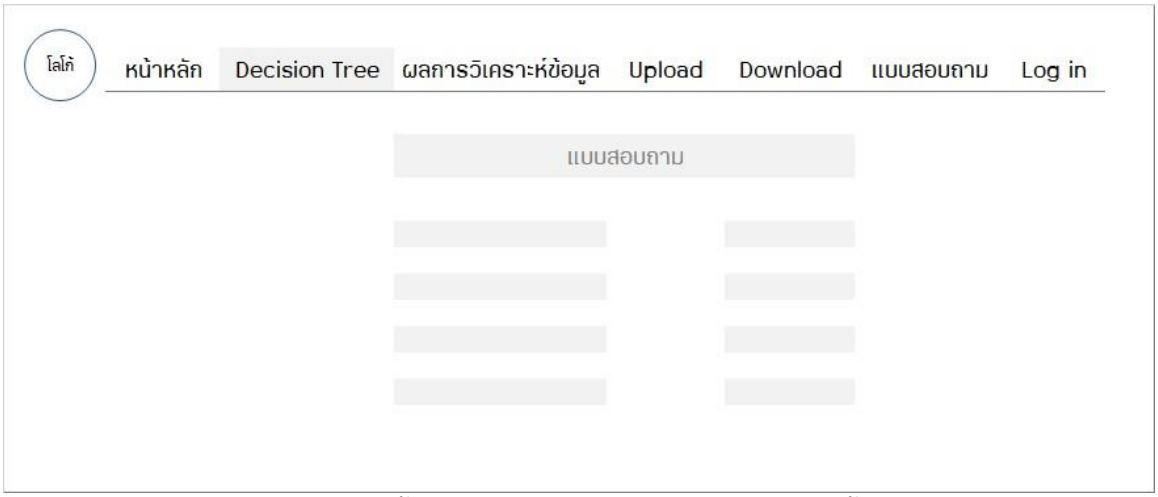
ภาพที่ 3.21 แสดงหน้าการทำ Decision Tree



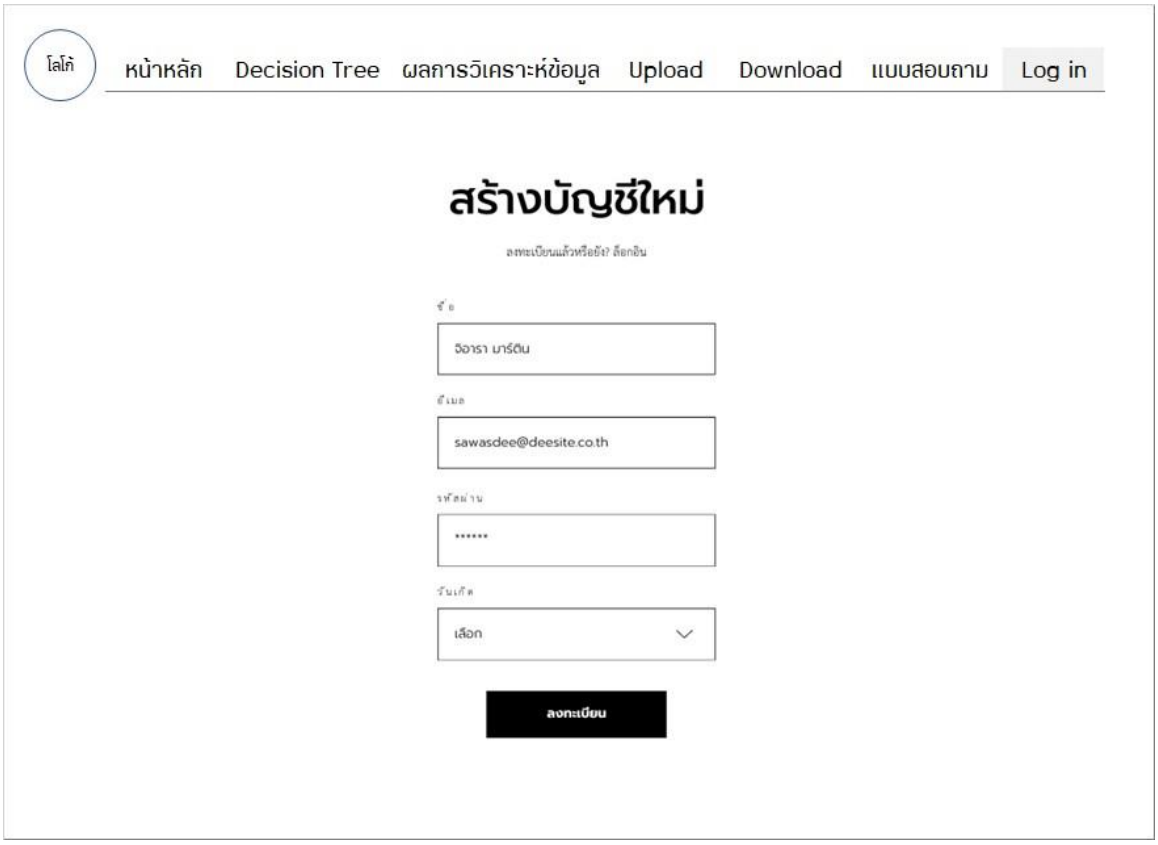
ภาพที่ 3.22 แสดงหน้าแสดงผลการวิเคราะห์มีทั้งหมด 4 ผลการวิเคราะห์

ภาพที่ 3.23 แสดงหน้าการอัปโหลดข้อมูลลงเว็บไซต์

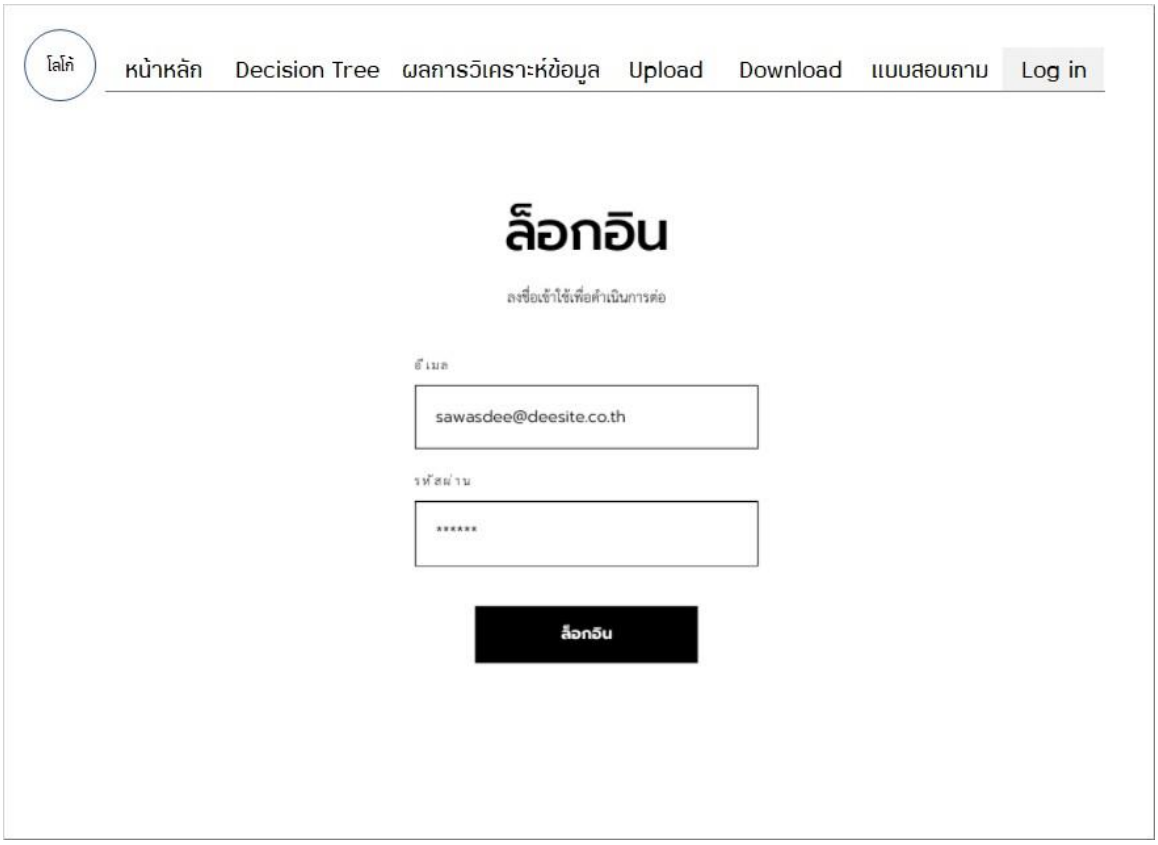
ภาพที่ 3.24 แสดงหนาเอกสารสำหรับดาวน์โหลด



ภาพที่ 3.25 หน้าแบบสอบถามความพึงพอใจในการใช้เว็บไซต์



ภาพที่ 3.26 แสดงหน้าสมัครสมาชิกสำหรับผู้เยี่ยมชมเว็บไซต์



ภาพที่ 3.27 แสดงหน้าล็อกอินสำหรับผู้วิเคราะห์

3.4 บทสรุป

จากวิธีการดำเนินงานโครงงานการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลความคงอยู่ของลูกค้า กรณีศึกษาบริษัท Telco ในข้างต้นทั้งหมดนี้ ผู้วิเคราะห์ข้อมูลทำการวิเคราะห์ความคงอยู่ของลูกค้าเพื่อเผยแพร่ข้อมูลบนเว็บไซต์ ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียด รวมถึงการสร้างโมเดล Decision Tree จากการเลือกแอทริบิวต์ด้วยโปรแกรม SPSS และโปรแกรมที่ใช้ทำเหมืองข้อมูล โปรแกรม Weka 3.9.6 และ โปรแกรม RapidMiner Studio 9.5.1 ในการสร้างโมเดล Decision Tree ซึ่งได้ผลลัพธ์ที่เหมือนกัน และนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ Visualization และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บน Web Browser ด้วยโปรแกรม Adobe XD อีกทั้งผู้วิเคราะห์ได้นำข้อมูลสารสนเทศมาทำการแสดงผลและเผยแพร่ข้อมูลสารสนเทศผ่านเว็บแอปพลิเคชันจะพัฒนาโดยใช้ภาษา PHP, JavaScript และชุดคำสั่ง CSS3 เพื่อนำเข้าวิเคราะห์ข้อมูลด้วยโมเดลที่เลือกใช้และผ่านการทดสอบประสิทธิภาพของการวิเคราะห์ข้อมูลแล้ว และนำเสนอข้อมูลภาพด้วยเฟรมเวิร์ค เพื่อสร้างกราฟและ Dashboard สำหรับแสดงผลข้อมูลภาพจากข้อมูลที่ผ่านการวิเคราะห์แล้วภายในเว็บแอปพลิเคชันที่พัฒนาขึ้น