บทที่ 3
วิธีการดำเนินงานโครงงาน

โครงงานเรื่อง การวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วย เทคนิคทางดาต้ามายนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

3.2 แผนภาพบริบท (Context Diagram)

3.3 แผนภาพกระแสข้อมูล Data Flow Diagram

3.4 ความสัมพันธ์ของข้อมูล (ER-Diagram)

3.5 การออกแบบเว็บไซต์

3.6 บทสรุป

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

ภาพที่ 3.1 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM

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

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการ ที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

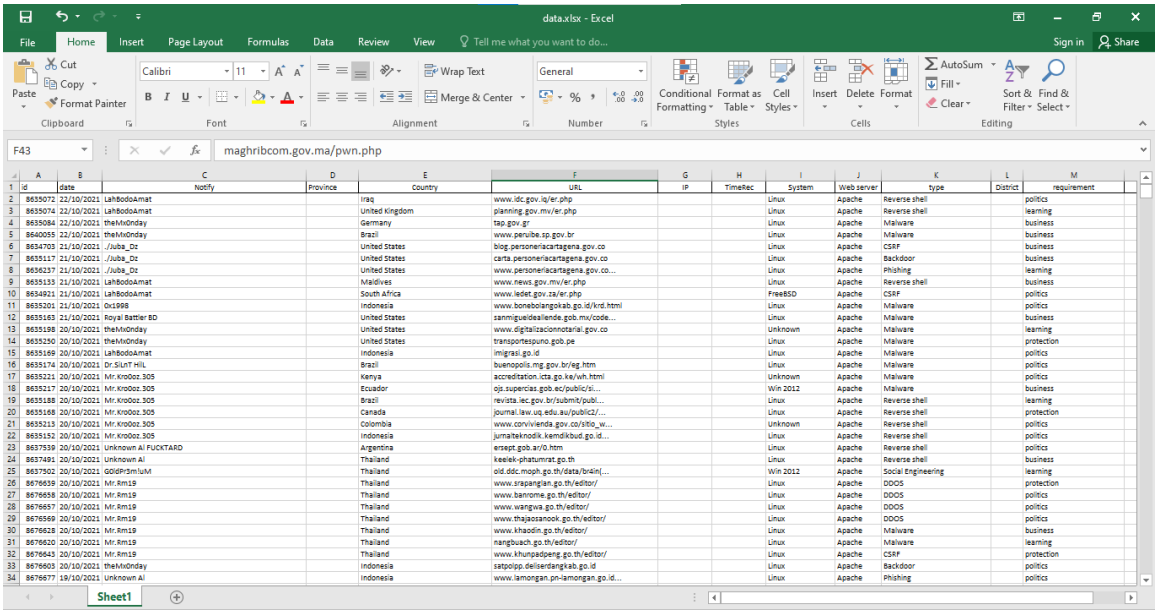
ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่งโดยการวิเคราะห์ข้อมูลรายงานการโจมตีทางเว็บไซต์ ปี 2019 - 2021 ซึ่งมีข้อมูลทั้งหมด 209,216 รายการ ซึ่งเป็นข้อมูลที่มีจำนวนมากและทำให้ซับซ้อน

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บ และรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลได้ทำการเก็บรวบรวมข้อมูลจากเว็บไซต์ osclab.hksecurity.net จากนั้นจะทำการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้ เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าข้อมูลการโจมตีเว็บไซต์ จะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

ภาพที่ 3.2 เว็บไซต์ osclab.hksecurity.net ที่ให้ข้อมูลภัยคุกคามบนเว็บไซต์

ซึ่งข้อมูลการโจมตีเว็บไซต์ มีจำนวน 209,216 รายการ ประกอบด้วย 13 แอตทริบิวท์ ข้อมูลหลัก ๆ จะประกอบด้วย ปี ประเภทภัยคุกคาม ประเทศ URL กลุ่มผู้โจมตี IP เว็บเซิร์ฟเวอร์ วัตถุประสงค์ของการโจมตี ระบบปฏิบัติการที่โดนโจมตี เป็นต้น

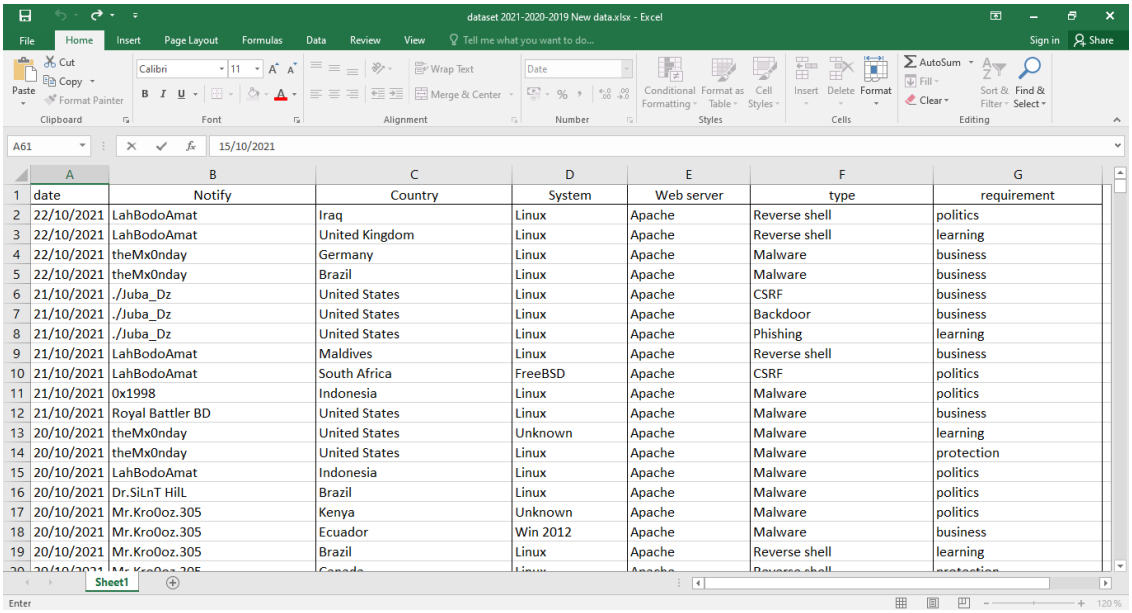


ภาพที่ 3.3 ข้อมูลดิบการโจมตีเว็บไซต์ทั้งหมดที่ได้จากเว็บไซต์ osclab.hksecurity.net

3.1.3 เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก โดยมีขั้นตอนดังนี้

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูลโดยการทำ Data Cleaning ข้อมูลรายงานภัยคุกคามทางไซเบอร์บนเว็บไซต์ โดยแยกข้อมูลออกและตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 7 แอตทริบิวท์ ได้แก่ ปี กลุ่มแฮกเกอร์ ประเทศ ระบบปฏิบัติการ เว็บเซิร์ฟเวอร์ รูปแบบภัยคุกคาม และ วัตถุประสงค์ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล



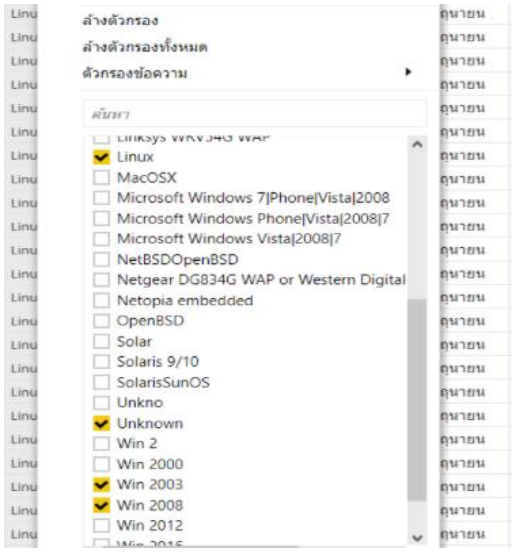
	A	B	C	D	E	F	G
1	date	Notify	Country	System	Web server	type	requirement
2	22/10/2021	LahBodoAmat	Iraq	Linux	Apache	Reverse shell	politics
3	22/10/2021	LahBodoAmat	United Kingdom	Linux	Apache	Reverse shell	learning
4	22/10/2021	theMx0nday	Germany	Linux	Apache	Malware	business
5	22/10/2021	theMx0nday	Brazil	Linux	Apache	Malware	business
6	21/10/2021	./Juba_Dz	United States	Linux	Apache	CSRF	business
7	21/10/2021	./Juba_Dz	United States	Linux	Apache	Backdoor	business
8	21/10/2021	./Juba_Dz	United States	Linux	Apache	Phishing	learning
9	21/10/2021	LahBodoAmat	Maldives	Linux	Apache	Reverse shell	business
10	21/10/2021	LahBodoAmat	South Africa	FreeBSD	Apache	CSRF	politics
11	21/10/2021	0x1998	Indonesia	Linux	Apache	Malware	politics
12	21/10/2021	Royal Battler BD	United States	Linux	Apache	Malware	business
13	20/10/2021	theMx0nday	United States	Unknown	Apache	Malware	learning
14	20/10/2021	theMx0nday	United States	Linux	Apache	Malware	protection
15	20/10/2021	LahBodoAmat	Indonesia	Linux	Apache	Malware	politics
16	20/10/2021	Dr.SiLnT HilL	Brazil	Linux	Apache	Malware	politics
17	20/10/2021	Mr.Kro0oz.305	Kenya	Unknown	Apache	Malware	politics
18	20/10/2021	Mr.Kro0oz.305	Ecuador	Win 2012	Apache	Malware	business
19	20/10/2021	Mr.Kro0oz.305	Brazil	Linux	Apache	Reverse shell	learning

ภาพที่ 3.4 ข้อมูลรายงานภัยคุกคามทางไซเบอร์บนเว็บไซต์ที่ทำการคัดเลือกข้อมูลแล้ว

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

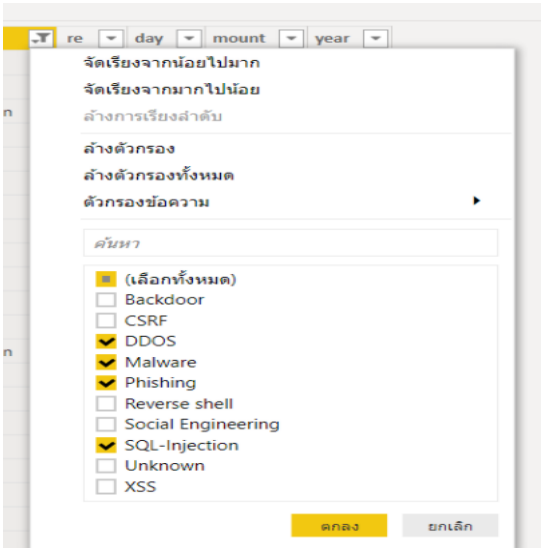
1) ข้อมูลรายงานภัยคุกคามทางไซเบอร์บนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ทำการแก้ไขและลบข้อมูล ซึ่งผู้วิเคราะห์ข้อมูลพบว่า ข้อมูลทั้งหมดนั้นมีจำนวนที่เยอะจึงเลือกส่วนหัวข้อย่อยในแต่ละแอตทริบิวท์มาเท่านั้น ดังนั้นผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

- Os (ระบบปฏิบัติการ) มี 85 หัวข้อย่อย ผู้วิเคราะห์ พบว่าควรตัด หัวข้อย่อยที่มีรายการน้อยกว่า หลักหมื่น จึงทำการลบทิ้ง และมีหัวข้อย่อย 3 หัวข้อย่อยที่มีมากเกินหลักหมื่น คือ Linux, Win2003, Win2008 ดังนี้



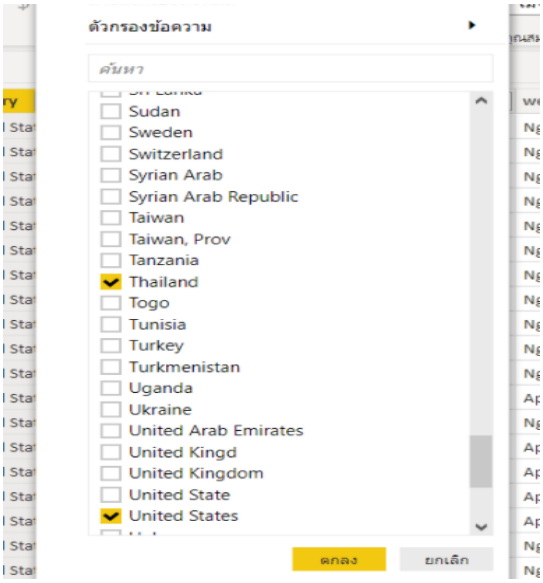
ภาพที่ 3.5 ทำการกลั่นกรองข้อมูล Os เพื่อนำไปวิเคราะห์

- Type (ประเภทภัยคุกคาม) มี 10 หัวข้อย่อย ผู้วิเคราะห์ พบว่าควรตัด หัวข้อย่อยที่มีรายการน้อยกว่า 17,000 รายการ จึงทำการลบทิ้ง และมีหัวข้อย่อย 4 หัวข้อย่อยที่เลือกมา คือ Phishing, Malware, DDOS และ SQL-Injection ดังนี้



ภาพที่ 3.6 ทำการกลั่นกรองข้อมูล รูปแบบการโจมตี เพื่อนำไปวิเคราะห์

- country (ประเทศ) มีทั้งหมด 156 หัวข้อย่อย ผู้วิเคราะห์ จึงทำการเลือกประเทศที่ มีจำนวนการโจมตีที่มากอันดับต้นๆ และประเทศไทย ซึ่งมีหัวข้อย่อย 4 หัวข้อย่อยที่เลือกมา คือ United States, Spain, Germany และ Thailand ดังนี้



ภาพที่ 3.7 ทำการกลั่นกรองข้อมูล ประเทศ เพื่อนำไปวิเคราะห์

3.1.3.3 แปลงรูปแบบของข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลในรูปแบบตาราฐานข้อมูลให้อยู่ในรูป item set เพื่อใช้สำหรับการนำมาวิเคราะห์ด้วยวิธีการของ data mining ผู้วิเคราะห์ข้อมูลได้ดำเนินการกับข้อมูลการโจมตีที่เกิดขึ้นบนเว็บไซต์ ปี 2019 ถึง ปี 2021 ดังนี้

1) ผู้วิเคราะห์ข้อมูลทำการแปลงรูปแบบข้อมูลด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ วัตถุประสงค์การโจมตี (Requirement) เนื่องจากปกติแล้วจะมีข้อมูลในแอตทริบิวต์นี้หลากหลาย ได้แก่ ด้านธุรกิจ (Business) ด้านการเมือง (Politics) ซึ่งทางผู้วิเคราะห์ข้อมูลจะขอรวมกลุ่มของวัตถุประสงค์การโจมตี ดังกล่าว ให้เหลือเพียง “การโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี” (Atk) อย่างเดียว

2) ผู้วิเคราะห์ข้อมูลทำการแปลงรูปแบบข้อมูลด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ วัตถุประสงค์การโจมตี (Requirement) เนื่องจากปกติแล้วจะมีข้อมูลในแอตทริบิวต์นี้หลากหลาย ได้แก่ ด้านการป้องกัน (Protection) ด้านการเรียนรู้ (Learning) ซึ่งทางผู้วิเคราะห์ข้อมูลจะ

ขอรวมกลุ่มของวัตถุประสงค์การโจมตี ดังกล่าว ให้เหลือเพียง “การโจมตีเพื่อทดสอบระบบและพัฒนาระบบ” (Def) อย่างเดียว

3) ผู้วิเคราะห์ข้อมูลทำการแปลงรูปแบบข้อมูลด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ ระบบปฏิบัติการ (Os) เนื่องจากปกติแล้วจะมีข้อมูลในแอตทริบิวต์นี้หลากหลาย ได้แก่ Win 2003, Win2008 ซึ่งทางผู้วิเคราะห์ข้อมูลจะขอรวมกลุ่มของระบบปฏิบัติการ ดังกล่าว ให้เหลือเพียง “Windows” อย่างเดียว

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

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิ่ง แบบการจำแนกประเภทข้อมูล (Classification) โดยการใช้โมเดลการตัดสินใจแบบต้นไม้ (Decision Tree) ซึ่งในขั้นตอนนี้จะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือก ดังนี้

	A	B	C	D	E
1	country	os	web	type	result
2	United States	Linux	Apache	SQL-Injection	Def
3	United States	Linux	Apache	Phishing	Atk
4	United States	Windows	Apache	Malware	Def
5	United States	Linux	Apache	Phishing	Def
6	United States	Linux	Nginx	Phishing	Atk
7	Spain	Linux	Apache	SQL-Injection	Atk
8	Germany	Linux	Apache	Phishing	Atk
9	United States	Linux	Apache	Malware	Atk
10	United States	Linux	Apache	DDOS	Def
11	United States	Linux	Nginx	DDOS	Def
12	Thailand	Windows	Nginx	DDOS	Def
13	Thailand	Linux	Apache	DDOS	Def
14	United States	Linux	Apache	Malware	Atk
15	United States	Windows	Apache	SQL-Injection	Atk
16	Germany	Linux	Nginx	Phishing	Atk
17	Spain	Linux	Apache	Phishing	Atk
18	United States	Linux	Apache	SQL-Injection	Def
19	Thailand	Windows	Nginx	Phishing	Atk
20	United States	Linux	Apache	DDOS	Atk
21	Spain	Linux	Apache	SQL-Injection	Atk
22	United States	Linux	Apache	DDOS	Atk
23	United States	Linux	Nginx	Phishing	Def
24	United States	Linux	Nginx	Phishing	Def

ภาพที่ 3.8 ข้อมูลแอตทริบิวต์ที่คัดเลือกมาวิเคราะห์ข้อมูลทั้งหมด

จากรูปภาพที่ ประกอบด้วย 5 แอตทริบิวต์ คือ

- Country (ประเทศ) ประกอบด้วย 4 ค่า คือ United States, Spain, Germany และ Thailand
- Os (ระบบปฏิบัติการ) ประกอบด้วย 2 ค่า คือ Linux, Windows
- Web server (เว็บเซิร์ฟเวอร์) ประกอบด้วย 2 ค่า คือ Apache, Nginx
- Type (ประเภทภัยคุกคาม) ประกอบด้วย 4 ค่า คือ Phishing, Malware, DDOS และ SQL-Injection
- Requirement (วัตถุประสงค์) ประกอบด้วย 2 ค่า คือ Atk และ Def

การสร้างโมเดล Decision Tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จาก

IG (parent, child) = entropy(parent) = – [p(c1) × entropy(c1) + p(c2) × entropy(c2) + ...]

โดยที่ entropy(c1) = -p(c1) log p(c1) และ p(c2) คือ ค่าความน่าจะเป็นของ c1

สมการการคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุด

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของ Decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น Requirement (วัตถุประสงค์) Atk และ Def ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ Country จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= - [0.6167 × -0.6974 + 0.3833 × -1.3835]

= - [-0.4301 + -0.5303]

= 0.9603

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= - [0.5932 × -0.7534 + 0.4068 × -1.2976]

= - [-0.4469 + -0.5279]

= 0.9748

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= - [0.6529 × -0.6151 + 0.3471 × -1.5266]

= - [-0.4016 + -0.5299]

= 0.9315

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6683 × log2(0.6683) + 0.3317 × log2(0.3317)]

= - [0.6683 × -0.5814 + 0.3317 × -1.5920]

= - [-0.3886 + -0.5281]

= 0.9167

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= - [0.6434 × -0.6362 + 0.3566 × -1.4876]

= - [-0.4093 + -0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 0.9603 - [0.6221 * 0.9748 + 0.1196 * 0.9315 + 0.1370 * 0.9167 + 0.1213 * 0.9398]

= 0.9603 – [0.6064 + 0.1114 + 0.1256 + 0.1140]

= 0.9603 – 0.9571

= 0.0029

2) คำนวณค่า IG ของแอตทริบิวต์ Os จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= - [0.6167 × -0.6974 + 0.3833 × -1.3835]

= - [-0.4301 + -0.5303]

= 0.9603

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6000 × log2(0.6000) + 0.4000 × log2(0.4000)]

= - [0.6000 × -0.7370 + 0.4000 × -1.3219]

= - [-0.4422 + -0.5288]

= 0.9710

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7002 × log2(0.7002) + 0.2998 × log2(0.2998)]

= - [0.7002 × -0.5142 + 0.2998 × -1.7379]

= - [-0.3600 + -0.5210]

= 0.8810

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

= 0.9603 - [0.8333 * 0.9710 + 0.1667 * 0.8810]

= 0.9603 – [0.8091 + 0.1469]

= 0.9603 – 0.9560

= 0.0043

3) คำนวณค่า IG ของแอตทริบิวต์ Type จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= - [0.6167 × -0.6974 + 0.3833 × -1.3835]

= - [-0.4301 + -0.5303]

= 0.9603

entropy (ผล = DDOS) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= - [0.3333 × -1.5851 + 0.6667 × -0.5849]

= - [-0.5283 + -0.3899]

= 0.9183

entropy (ผล = Malware) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9336 × log2(0.9336) + 0.0664 × log2(0.0664)]

= - [0.9336 × -0.0991 + 0.0664 × -3.9127]

= - [-0.0925 +-0.2598]

= 0.3523

entropy (ผล = Phishing) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7000 × log2(0.7000) + 0.3000 × log2(0.3000)]

= - [0.7000 × -0.5146 + 0.3000 × -1.7370]

= - [-0.3602 + -0.5211]

= 0.8813

entropy (ผล = SQL-Injection) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5999 × log2(0.5999) + 0.4001 × log2(0.4001)]

= - [0.5999 × -0.7372 + 0.4001 × -1.3216]

= - [-0.4422 + -0.5288]

= 0.9710

IG (parent, child) = entropy(parent) – [p (ผล = DDOS) * entropy (ผล = DDOS) + p (ผล = Malware) * entropy (ผล = Malware) + p (ผล = Phishing) * entropy (ผล = Phishing) + p (ผล = SQL-Injection) * entropy (ผล = SQL-Injection)]

= 0.9603 - [0.2500 * 0.9183 + 0.1250 * 0.3523 + 0.4167 * 0.8813 + 0.2083 * 0.9710]

= 0.9603 – [0.2296 + 0.0440 + 0.3672 + 0.2023]

= 0.9603 – 0.8431

= 0.1172

4) คำนวณค่า IG ของแอตทริบิวต์ Web server จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= - [0.6167 × -0.6974 + 0.3833 × -1.3835]

= - [-0.4301 + -0.5303]

= 0.9603

entropy (ผล = Apache) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6750 × log2(0.6750) + 0.3250 × log2(0.3250)]

= - [0.6750 × -0.5670 + 0.3250 × -1.6215]

= - [-0.3828 + -0.5270]

= 0.9097

entropy (ผล = Nginx) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.5000× -1 + 0.5000 × -1]

= - [-0.5000 + -0.5000]

= 1

IG (parent, child) = entropy(parent) – [p (ผล = Apache) * entropy (ผล = Apache) + p (ผล = Nginx) * entropy (ผล = Nginx)]

= 0. 9603 - [0.6667 * 0.9097 + 0.3333 * 1]

= 0. 9603 – [0.6065 + 0.3333]

= 0. 9603 – 0.9398

= 0.0205

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ type มีค่ามากที่สุด (0.1172) ดังนั้นจึงเลือกแอตทริบิวต์ type ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งจากโหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ type (Malware) และ type (Phishing) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์การโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตีและการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ ดังนั้นผู้วิเคราะห์ข้อมูลจึงสร้างโหนดในระดับถัดไปของแอตทริบิวต์ type DDOS และ SQL-Injection

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 ต่อจากโหนด root เพื่อหาค่า IG ที่มากที่สุด ของแอตทริบิวต์ type กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Country ของ DDOS จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= - [0.3333 × -1.5851 + 0.6667 × -0.5849]

= - [-0.5283 + -0.3899]

= 0.9183

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.3746 × log2(0.3746) + 0.6254 × log2(0.6254)]

= - [0.3746 × -1.4166 + 0.6254 × -0.6771]

= - [-0.5306 + -0.4235]

= 0.9541

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.2469 × log2(0.2469) + 0.7531 × log2(0.7531)]

= - [0.2469 × -2.0180 + 0.7531 × -0.4091]

= - [-0.4982 + -0.3081]

= 0.8063

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.0371 × log2(0.0371) + 0.9629 × log2(0.9629)]

= - [0.0371 × -4.7524 + 0.9629 × -0.0545]

= - [-0.1763 + -0.0525]

= 0.2288

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.4054 × log2(0.4054) + 0.5946 × log2(0.5946)]

= - [0.4054 × -1.3026 + 0.5946 × -0.7500]

= - [-0.5281 + -0.4460]

= 0.9740

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 0.9183 - [0.6628 * 0.9541 + 0.1070 * 0.8063 + 0.0941 * 0.2288 + 0.1361 * 0.9740]

= 0.9183 – [0.6324 + 0.0863 + 0.0215 + 0.1326]

= 0.9183 – 0.8727

= 0.0456

1.1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Country ของ Malware จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9336 × log2(0.9336) + 0.0925 × log2(0.0925)]

= - [0.9336 × -0.0991 + 0.0925 × -3.9127]

= - [-0.0925 + -0.2598]

= 0.3523

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9477 × log2(0. 9477) + 0.0523 × log2(0. 0523)]

= - [0. 9477 × -0.0775 + 0. 0523 × -4.2570]

= - [-0.0734 + -0.2226]

= 0.2961

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.8942 × log2(8942) + 0.1058 × log2(0.1058)]

= - [0. 8942 × -0.1613 + 0. 1058 × --3.2406]

= - [-0.1613 + -0.3429]

= 0.4871

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [0 + 0]

= 0

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.8555 × log2(0.8555) + 0.1445 × log2(0.1445)]

= - [0.8555 × -2252 + 0.1445 × -2.7909]

= - [-0.1926 + -0.4033]

= 0.5959

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 0.3523 - [0.6056 * 0.2961 + 0.1069 * 0.4871 +
0.1256 * 0 + 0.1618 * 0.5959]

= 0.3523 – [0.1793 + 0.0766 + 0 + 0.0964]

= 0.3523 – 0.3523

= 0

1.2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Country ของ Phishing จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7000 × log2(0.7000) + 0.3000 × log2(0.3000)]

= - [0.7000 × -0.5146 + 0.3000 × -1.7370]

= - [-0.3602 + -0.5211]

= 0.8813

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7270 × log2(0.7270) + 0.2730 × log2(0.2730)]

= - [0.7270 × -0.4600 + 0.2730 × -1.8730]

= - [-0.3344 + -0.5113]

= 0.8457

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6653 × log2(0.6653) + 0.3347 × log2(0.3347)]

= - [0.6653 × -0.5879 + 0.3347 × -1.5791]

= - [-0.3911 + -0.5285]

= 0.9197

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6686 × log2(0.6686) + 0.3314 × log2(0.3314)]

= - [06686 × -0.5808 + 0.3314 × -1.5934]

= - [-0.3883 + -0.5280]

= 0.9164

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6684 × log2(0.6684) + 0.3316 × log2(0.3316)]

= - [0.6684 × -0.5812 + 0.3316 × -1.5925]

= - [-0.3885 + -0.5281]

= 0.9166

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 0.8813 - [0.5465 * 0.8457 + 0.1432 * 0.9197 + 0.1649 * 0.9164 + 0.1454 * 0.9166]

= 0.8813 – [0.4622 + 0.1341 + 0.1511 + 0.1333]

= 0.8813 – 0.8753

= 0

1.3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Country ของ SQL-Injection จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5999 × log2(0.5999) + 0.4001 × log2(0.4001)]

= - [0.5999 × -0.7372 + 0.4001 × -1.3216]

= - [-0.4422 + -0.5288]

= 0.9710

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.4554 × log2(0.4554) + 0.5446 × log2(0.5446)]

= - [0.4554 × -1.1348 + 0.5446 × -0.8767]

= - [-0.5168 + -0.4775]

= 0.9943

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [0 + 0]

= 0

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [0 + 0]

= 0

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 0.9710 - [0.7346 * 0.9943 + 0.0952 * 0 + 0.1393 * 0 + 0.0309 * 0]

= 0.9710 – [0.7143 + 0 + 0 + 0]

= 0.9710 – 0.7143

= 0.2040

2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Os ของ DDOS จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= - [0.3333 × -1.5851 + 0.6667 × -0.5849]

= - [-0.5283 + -0.3899]

= 0.9183

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.4000 × log2(0.4000) + 0.6000 × log2(0.6000)]

= - [0.4000 × -1.3219 + 0.6000 × -0.7370]

= - [-0.5288 + -0.4422]

= 0.9710

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0 × log2(0) + 1 × log2(1)]

= - [0 × -0 + 1 × -0]

= - [0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

= 0.9183 - [0.8333 * 0.9710 + 0.1667 * 0]

= 0.9183 – [0.8091 + 0]

= 0.9183 – 0.8091

= 0.1092

2.1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Os ของ Malware จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9336 × log2(0.9336) + 0.0925 × log2(0.0925)]

= - [0.9336 × -0.0991 + 0.0925 × -3.9127]

= - [-0.0925 + -0.2598]

= 0.3523

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × 0 + 0 × 0]

= - [-0 + -0]

= 0

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.8009 × log2(0.8009) + 0.1991 × log2(0.1991)]

= - [0.8009 × -0.3203 + 0.1991 × -2.3204]

= - [-0.2565 + -0.4636]

= 0.7201

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

= 0.3523 - [0.6665 * 0 + 0.3335 * 0.7201]

= 0.3523 – [0 + 0.2402]

= 0.3523 – 0.2402

= 0.1121

2.2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Os ของ Phishing จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7000 × log2(0.7000) + 0.3000 × log2(0.3000)]

= - [0.7000 × -0.5146 + 0.3000 × -1.7370]

= - [-0.3602 + -0.5211]

= 0.8813

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6667 × log2(0.6667) + 0.3333 × log2(0.3333)]

= - [0.6667 × -0.5849 + 0.3333 × -1.5851]

= - [-0.3899 + -0.5283]

= 0.9183

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [-0 + -0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

= 0.8813- [0.9000 * 0.9183 + 0.1000 * 0]

= 0.8813– [0.8265 + 0]

= 0.8813– 0.8265

= 0.0548

2.3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Os ของ SQL-Injection จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5999 × log2(0.5999) + 0.4001 × log2(0.4001)]

= - [0.5999 × -0.7372 + 0.4001 × -1.3216]

= - [-0.4422 + -0.5288]

= 0.9710

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.6667 × -0.5849 + 0.3333 × -1.5851]

= - [-0.3899 + -0.5283]

= 1

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [1 × log2(1) + 0 × log2(0)]

= - [1 × -0 + 0 × -0]

= - [-0 + -0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

= 0.9710 - [0.8000 * 0.9710 + 0.2000 * 0]

= 0.9710 – [0.7768 + 0]

= 0.9710 – 0.7768

= 0.1942

3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Web server ของ DDOS จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= - [0.3333 × -1.5851 + 0.6667 × -0.5849]

= - [-0.5283 + -0.3899]

= 0.9183

entropy (ผล = Apache) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.5000 × -1 + 0.5000 × -1]

= - [-0.5000 + -0.5000]

= 1

entropy (ผล = Nginx) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0 × log2(0) + 1 × log2(1)]

= - [0 × -0 + 1 × -0]

= - [0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Apache) * entropy (ผล = Apache) + p (ผล = Nginx) * entropy (ผล = Nginx)]

= 0.9183 - [0.6667 * 1 + 0.3333 * 0]

= 0.9183 – [0.6667 + 0]

= 0.9183 – 0.6667

= 0.2516

3.1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Web server ของ Malware จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9336 × log2(0.9336) + 0.0925 × log2(0.0925)]

= - [0.9336 × -0.0991 + 0.0925 × -3.9127]

= - [-0.0925 + -0.2598]

= 0.3523

entropy (ผล = Apache) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.9336 × log2(0.9336) + 0.0664 × log2(0.0664)]

= - [0.9336 × -0.0991 + 0.0664 × -3.9127]

= - [-0.0925 + -0.2598]

= 0.3523

entropy (ผล = Nginx) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0 × log2(0) + 0 × log2(0)]

= - [0 × -0 + 0 × -0]

= - [0 + 0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Apache) * entropy (ผล = Apache) + p (ผล = Nginx) * entropy (ผล = Nginx)]

= 0.3523 - [1 * 0.3523 + 0 * 0]

= 0.3523 – [0.3523 + 0]

= 0.3523 – 0.3523

= 0

3.2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Web server ของ Phishing จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7000 × log2(0.7000) + 0.3000 × log2(0.3000)]

= - [0.7000 × -0.5146 + 0.3000 × -1.7370]

= - [-0.3602 + -0.5211]

= 0.8813

entropy (ผล = Apache) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.7499 × log2(0.7499) + 0.2501 × log2(0.2501)]

= - [0.7499 × -0.4152 + 0.2501 × -1.9994]

= - [-0.3114 + -0.5001]

= 0.8114

entropy (ผล = Nginx) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6667 × log2(0.6667) + 0.3333 × log2(0.3333)]

= - [0.6667 × -0.5849 + 0.3333 × -1.5851]

= - [-0.3899 + -0.5283]

= 0.9183

IG (parent, child) = entropy(parent) – [p (ผล = Apache) * entropy (ผล = Apache) + p (ผล = Nginx) * entropy (ผล = Nginx)]

= 0.8813 - [0.4000 * 0.8114 + 0.6000 * 0.9184]

= 0.8813 – [0.3189 + 0.5510]

= 0.8813 – 0.8813

= 0

3.3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ Web server ของ SQL-Injection จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5999 × log2(0.5999) + 0.4001 × log2(0.4001)]

= - [0.5999 × -0.7372 + 0.4001 × -1.3216]

= - [-0.4422 + -0.5288]

= 0.9710

entropy (ผล = Apache) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5999 × log2(0.5999) + 0.4001 × log2(0.4001)]

= - [0.5999 × -0.7372 + 0.4001 × -1.3216]

= - [-0.4422 + -0.5288]

= 0.9710

entropy (ผล = Nginx) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0 × log2(0) + 0 × log2(0)]

= - [0 × -0 + 0 × -0]

= - [-0 + -0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Apache) * entropy (ผล = Apache) + p (ผล = Nginx) * entropy (ผล = Nginx)]

= 0.9710 - [1 * 0.9710 + 0 * 0]

= 0.9710 – [0.9710 + 0]

= 0.9710 – 0.9710

= 0

จากการคำนวณค่า IG ของแอตทริบิวต์ Type ต่อแอตทริบิวต์ Country แอตทริบิวต์ Os และแอทริบิวต์ Web server พบว่าค่า IG ของแอตทริบิวต์ Type (DDOS) ต่อแอตทริบิวต์ Web server มีค่ามากที่สุด (0.2516) และแอตทริบิวต์ Type (SQL-Injection) ต่อแอตทริบิวต์ Country มีค่ามากรองลงมาเป็น (0.2040) ดังนั้นจึงเลือกแอตทริบิวต์ Web server และแอตทริบิวต์ Country ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจากโหนด Root และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ Type (Malware, Phishing) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์การโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ ดังนั้นผู้วิเคราะห์ข้อมูลจึงทำการแตกกิ่งจากโหนดในระดับที่ 2 ของแอตทริบิวต์ Country, OS ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นเป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ ดังนี้

1)คำนวณค่า IG ของแอตทริบิวต์ Type (DDOS) และแอตทริบิวต์ Web server (Apache) ไปแอตทริบิวต์ Country (United States, Thailand, Spain, Germany) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.5000 × -1 + 0.5000 × -1]

= - [-0.5000 + -0.5000]

= 1

entropy (ผล = United States) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5293 × log2(0.5293) + 0.4707 × log2(0.4707)]

= - [0.5293 × -0.9178 + 0.4707 × -1.0871]

= - [-0.4858 + -0.5117]

= 0.9975

entropy (ผล = Thailand) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.4965 × log2(0.4965) + 0.5015 × log2(0.5015)]

= - [0.4965 × -1.0101 + 0.5015 × -0.9899]

= - [-0.5015 + -0.9484]

= 1

entropy (ผล = Spain) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.0550 × log2(0.0550) + 0.9450 × log2(0.9450)]

= - [0.0550 × -4.1844 + 0.9450 × -0.0816]

= - [-0.2301 + -0.0771]

= 0.3073

entropy (ผล = Germany) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.6818 × log2(0.6818) + 0.3182 × log2(0.3182)]

= - [0.6818 × -0.5526 + 0.3182 × -1.6520]

= - [-0.3767 + -0.5257]

= 0.9024

IG (parent, child) = entropy(parent) – [p (ผล = United States) * entropy (ผล = United States) + p (ผล = Spain) * entropy (ผล = Spain) + p (ผล = Germany) * entropy (ผล = Germany) + p (ผล = Thailand) * entropy (ผล = Thailand)]

= 1 - [0.7036 * 0.9975 + 0.0798 * 1 + 0.0952 * 0.3073 + 0.1214 * 0.9024]

= 1– [0.7018 + 0.0798 + 0.0293 + 0.1096]

= 1– 0.9204

= 0.0796

2)คำนวณค่า IG ของแอตทริบิวต์ Type (DDOS) และแอตทริบิวต์ Web server (Apache) ไปแอตทริบิวต์ Os (Linux, Windows) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.5000 × -1 + 0.5000 × -1]

= - [-0.5000 + -0.5000]

= 1

entropy (ผล = Linux) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0.5000 × log2(0.5000) + 0.5000 × log2(0.5000)]

= - [0.5000 × -1 + 0.5000 × -1]

= - [-0.5000 + -0.5000]

= 1

entropy (ผล = Windows) = - p(Atk) × $\log_2$p(Atk) + p(Def) × $\log_2$p(Def)

= - [0 × log2(0) + 0 × log2(0)]

= - [0 × -0 + 0 × -0]

= - [-0 + -0]

= 0

IG (parent, child) = entropy(parent) – [p (ผล = Linux) * entropy (ผล = Linux) + p (ผล = Windows) * entropy (ผล = Windows)]

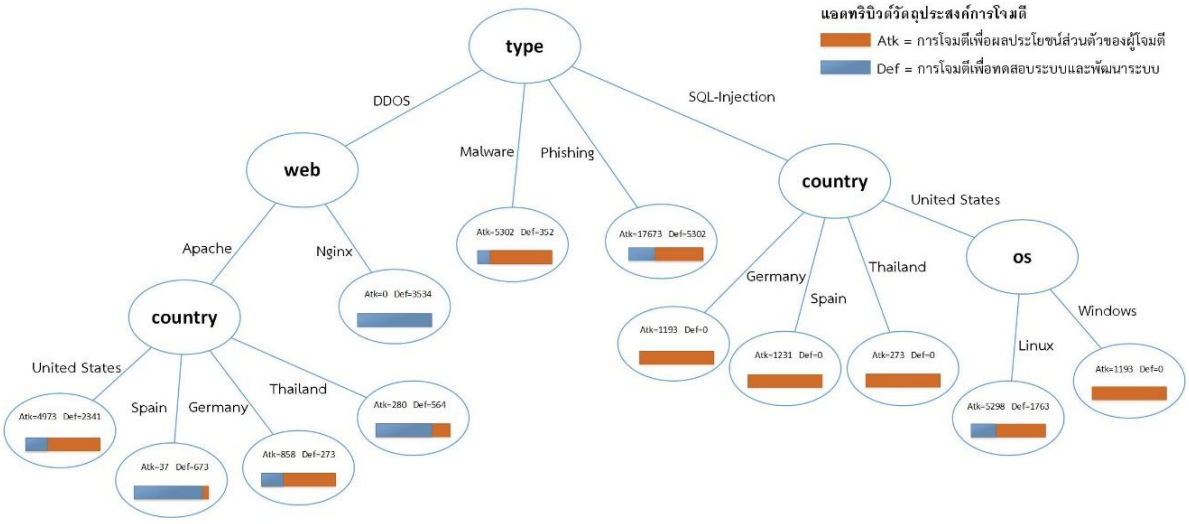
= 1 - [1 * 1 + 0 * 0]

= 1– [1 + 0]

= 1– 1

= 0

จากการคำนวณค่า IG ของแอตทริบิวต์ Type (DDOS) ต่อแอตทริบิวต์ Web server ไปแอตทริบิวต์ Os และแอตทริบิวต์ Country พบว่าค่า IG ของแอตทริบิวต์ Type ต่อแอตทริบิวต์ Web server ไปแอตทริบิวต์ Country มีค่ามากที่สุด (0.0796) ดังนั้นจึงเลือกแอตทริบิวต์ Country ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนด Root ต่อจากโหนดระดับที่ 2 และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ Web server (Nginx) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์การโจมตีเพื่อทดสอบระบบและพัฒนาระบบ และแอตทริบิวต์ Type (SQL-Injection) ต่อแอตทริบิวต์ Country ไปแอตทริบิวต์ Os เป็นแอตทริบิวต์สุดท้าย พบว่าแอตทริบิวต์ Country มีความสัมพันธ์กับแอตทริบิวต์ Os มากที่สุด ซึ่งพบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ การโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ ตามภาพที่ 3.9



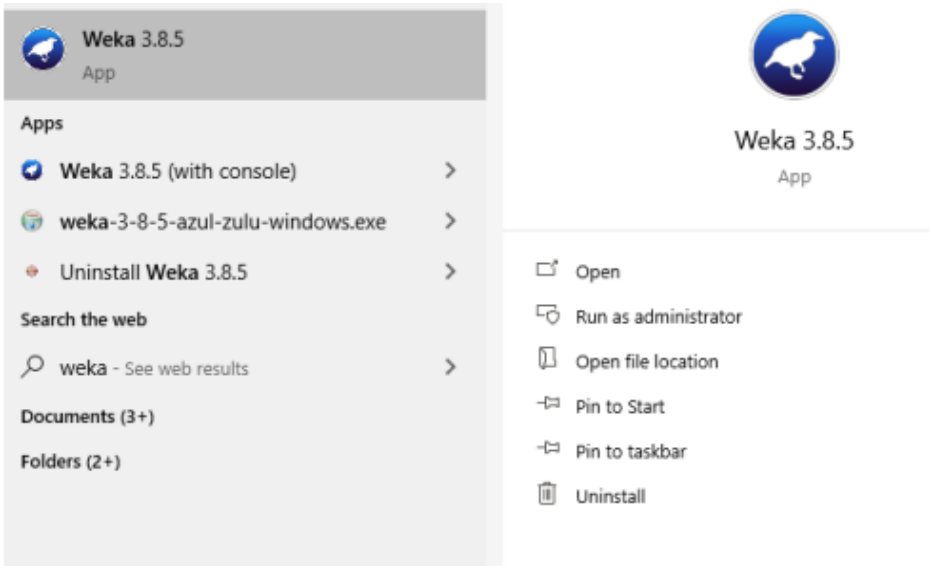
ภาพที่ 3.9 แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจ จากการคำนวณด้วยมือ

จากภาพที่ 3.9 โมเดลต้นไม้ตัดสินใจ จากการคำนวณด้วยมือนี้ ผู้วิเคราะห์ข้อมูลได้ผลลัพธ์ว่า โมเดลต้นไม้ตัดสินใจ Root node ที่คือ แอตทริบิวต์ Type และได้ interior node คือ แอตทริบิวต์ Country , แอตทริบิวต์ Web server และ leaf node คือ แอตทริบิวต์ Os ซึ่งไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด ก็จะได้ผลลัพธ์ที่ แอตทริบิวต์ Os linux เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ nginx เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี แอตทริบิวต์ Country United States เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี, Spain เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี, Germany เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ และ Thailand เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

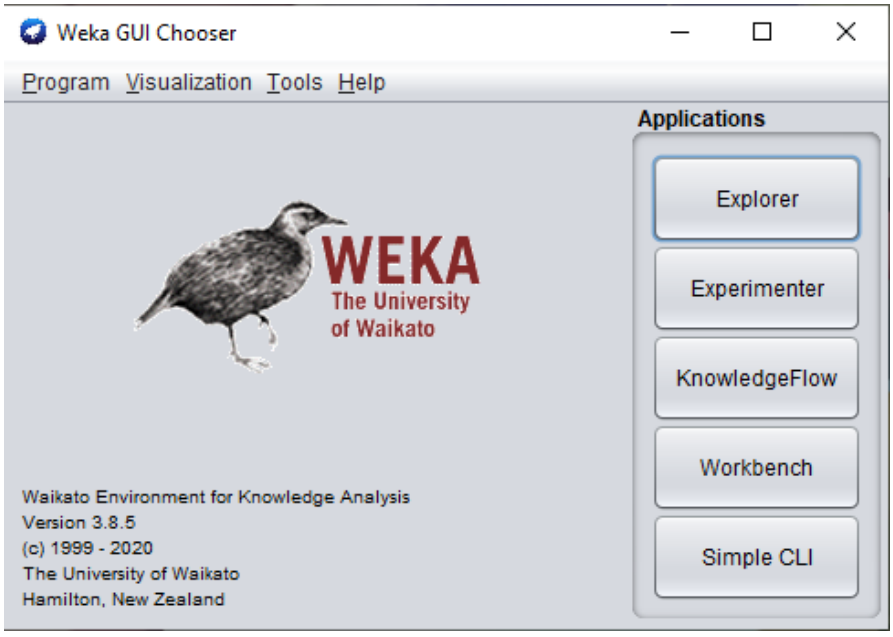
3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 3.1.4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

ผู้วิเคราะห์ข้อมูลได้ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพ เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (Testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการสร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka 3.8.5 ซึ่งมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 เปิดโปรแกรม Weka 3.8.5 ขึ้นมา

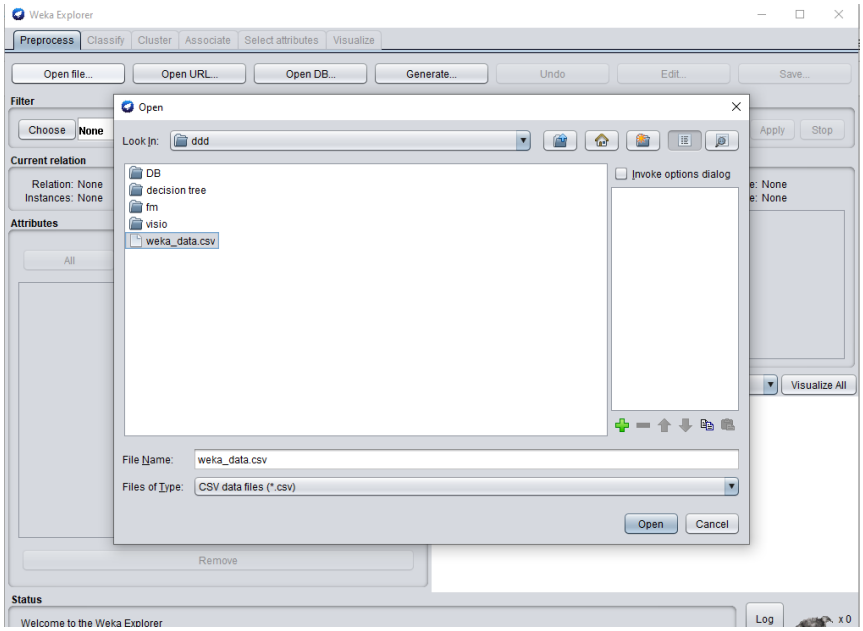


ภาพที่ 3.10 แสดงการเปิดโปรแกรม weka 3.8.5

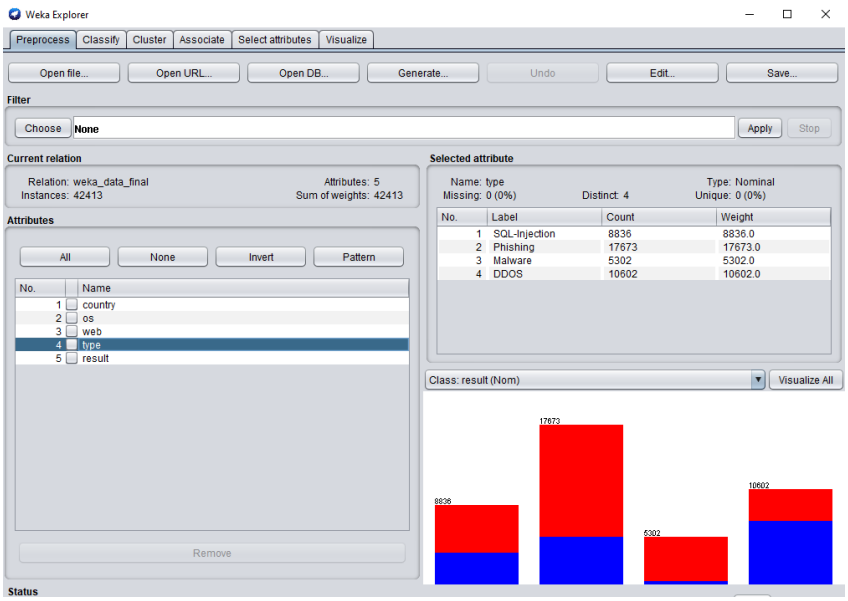


ภาพที่ 3.11 แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.5

ขั้นตอนที่ 2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Application>>Explorer>>Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ 3.11 และหลังจากนั้นโปรแกรมแสดงหน้าจอข้อมูลตามภาพที่ 3.12

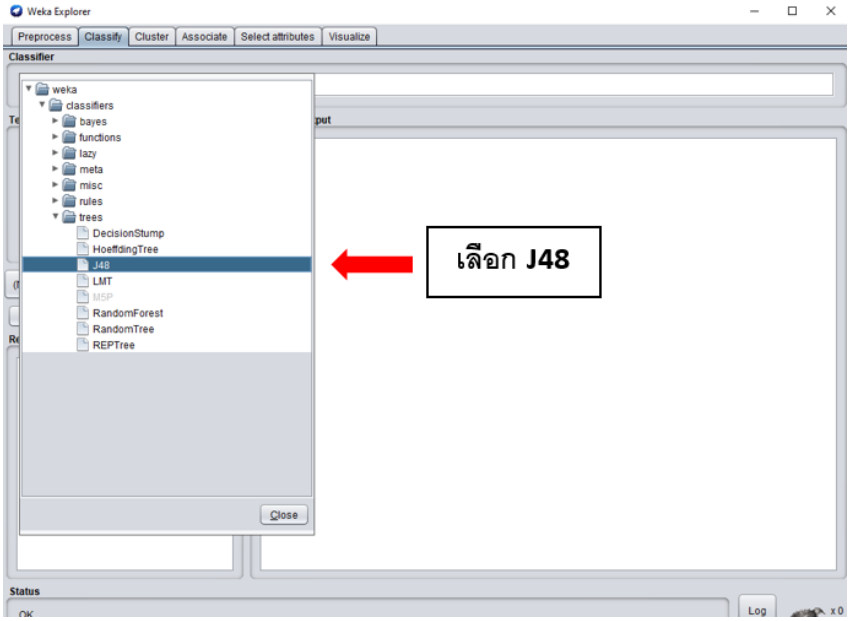


ภาพที่ 3.12 แสดงการนำเข้าข้อมูลเข้าในโปรแกรม Weka 3.8.5

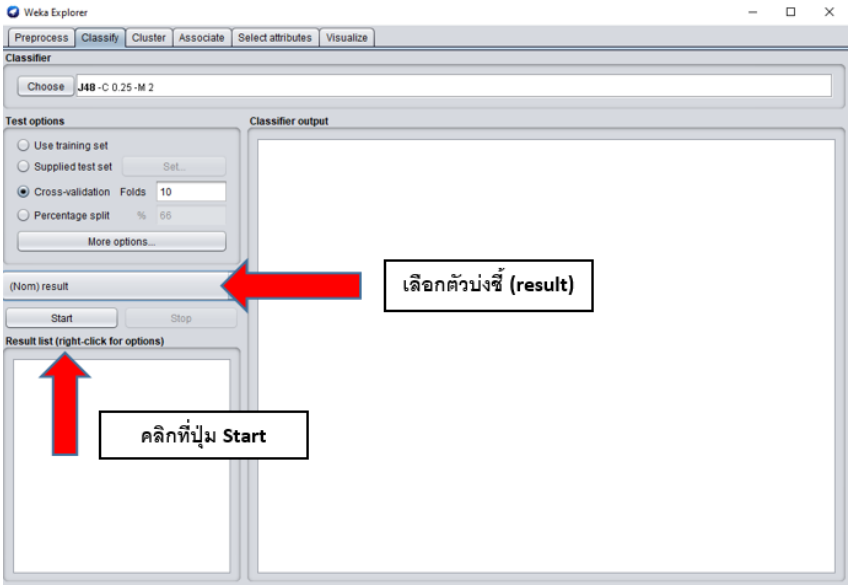


ภาพที่ 3.13 แสดงข้อมูลที่เข้าในโปรแกรม Weka 3.8.5

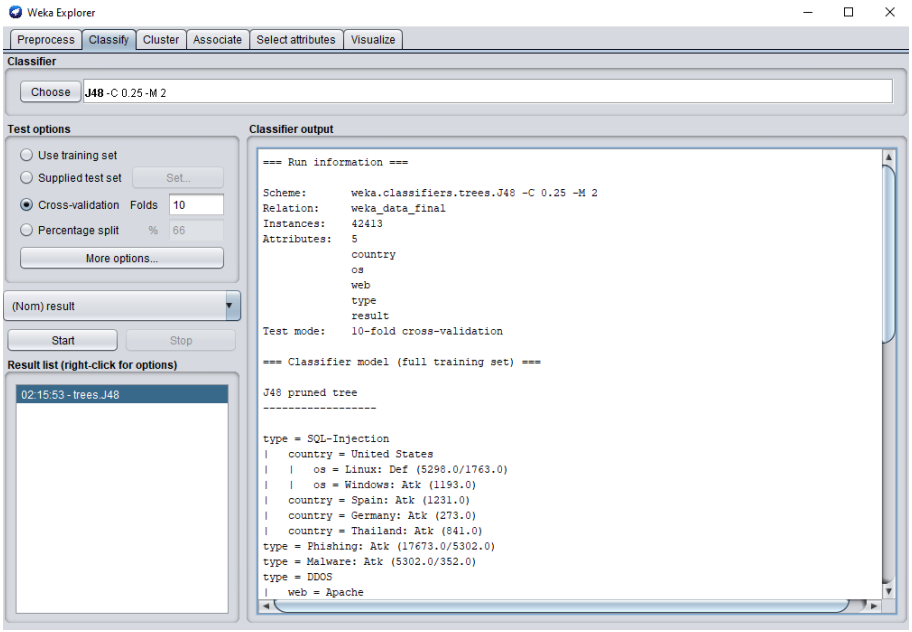
ขั้นตอนที่ 3 ดำเนินการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree โดยเลือกที่ Classification>>Choose>>tree และเลือกรูปแบบเป็น J48 ตามภาพที่ 3.14 จากนั้นเลือกตัวบ่งชี้ในที่นี้ใช้เป็นวัตถุประสงค์การโจมตี (requirement) ส่วนที่ใช้ในโปรแกรม ใช่ชื่อว่า (result) จากนั้นกดปุ่ม Start ตามภาพที่ 3.15 จะแสดงผลลัพธ์ที่ได้ตามภาพที่ 3.16



ภาพที่ 3.14 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree รูปแบบเป็น J48

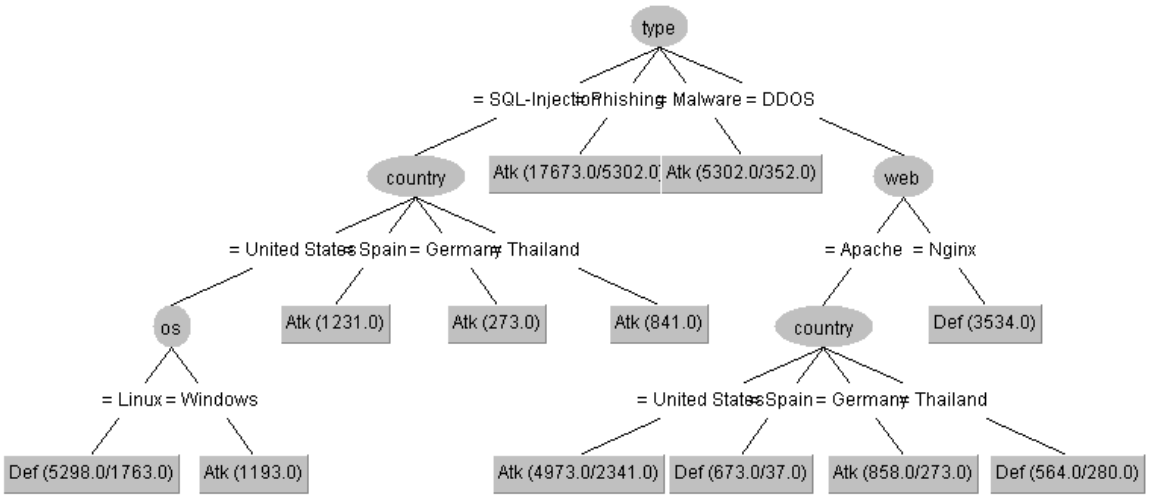


ภาพที่ 3.15 แสดงการเลือกตัวบ่งชี้เป็น result แล้ว คลิกที่ start



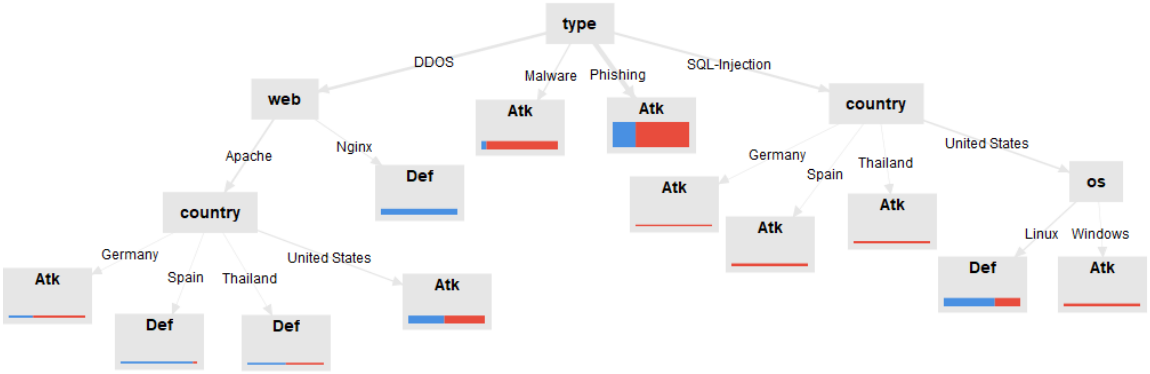
ภาพที่ 3.16 แสดงหน้าจอผลลัพธ์ของโมเดลการจัดกลุ่มข้อมูลแบบ Decision Tree : J48

จากผลลัพธ์การทดลองพบว่าเทคนิค Decision Tree : J48 ให้ผลลัพธ์การจำแนกประเภทการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี (Atk) และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ (Def) มีความถูกต้องถึง 75.55% แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจที่มีกิ่งแตกออกมา ดังภาพที่ 3.17

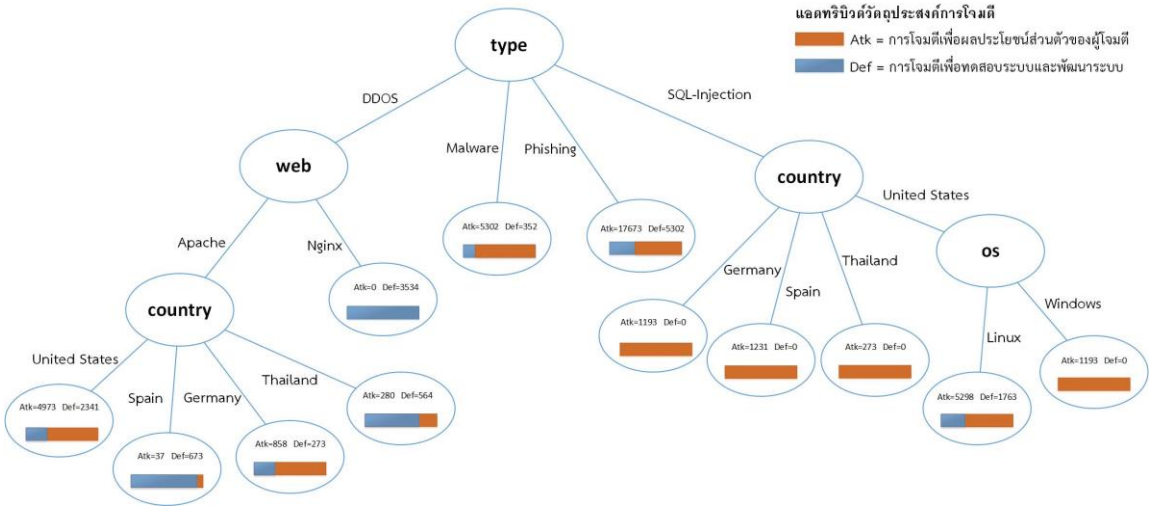


ภาพที่ 3.17 แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจ ในโปรแกรม Weka 3.8.5

ทางผู้วิเคราะห์ข้อมูลยังได้นำโมเดลของข้อมูลที่เลือกใช้ มาเปรียบเทียบกับโมเดลที่สร้างด้วยโปรแกรม RapidMiner Studio 9.5.1 เพื่อตรวจสอบความแม่นยำของโมเดล ซึ่งผู้วิเคราะห์ข้อมูลพบว่าได้ผลลัพธ์รูปแบบแผนภาพโมเดลที่ตรงกัน



ภาพที่ 3.18 แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจ ของโปรแกรม RapidMiner Studio



ภาพที่ 3.19 แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจ จากการคำนวณด้วยมือ

จากผลลัพธ์การคำนวณค่า IG ของโมเดลการโจมตีที่เกิดขึ้นบนเว็บไซต์แต่ละโหนดและจากการทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการด้วยวิธี Self-Consistency Test เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้นจากโปรแกรม Weka เวอร์ชั่น 3.8.5 และเมื่อนำไปเปรียบเทียบกับโปรแกรม RapidMiner Studio เวอร์ชัน 9.5.1 ทางผู้วิเคราะห์ข้อมูลพบว่าทั้ง 3 โมเดลได้ผลลัพธ์ความแม่นยำของโมเดลที่เหมือนกันและ สามารถนำโมเดลไปใช้งานได้

จากผลลัพธ์การสร้างโมเดลด้วยเทคนิค Decision Tree: J48 ในโปรแกรม Weka 3.8.5 ได้สร้างกฎจากการจำแนกกลุ่มต้นไม้การตัดสินใจแบบ Decision Tree มีกิ่งแตกออกมา ดังภาพที่ 3.20

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

type = SQL-Injection
|   country = United States
|   |   os = Linux: Def (5298.0/1763.0)
|   |   os = Windows: Atk (1193.0)
|   country = Spain: Atk (1231.0)
|   country = Germany: Atk (273.0)
|   country = Thailand: Atk (841.0)
type = Phishing: Atk (17673.0/5302.0)
type = Malware: Atk (5302.0/352.0)
type = DDOS
|   web = Apache
|   |   country = United States: Atk (4973.0/2341.0)
|   |   country = Spain: Def (673.0/37.0)
|   |   country = Germany: Atk (858.0/273.0)
|   |   country = Thailand: Def (564.0/280.0)
|   web = Nginx: Def (3534.0)

Number of Leaves  :     12

Size of the tree :      17
```

ภาพที่ 3.20 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจ ของโปรแกรม Weka 3.8.5

ซึ่งผู้วิเคราะห์ข้อมูลพบว่าเมื่อนำมาเปรียบเทียบกับกฎจากการจำแนกกลุ่มแบบ Decision Tree ในโปรแกรม RapidMiner Studio ได้ผลลัพธ์รูปแบบกฎจากการจำแนกกลุ่มที่ต่างกันเล็กน้อย

Tree

```
type = DDOS
|   web = Apache
|   |   country = Germany: Atk {Def=273, Atk=585}
|   |   country = Spain: Def {Def=636, Atk=37}
|   |   country = Thailand: Def {Def=284, Atk=280}
|   |   country = United States: Atk {Def=2341, Atk=2632}
|   web = Nginx: Def {Def=3534, Atk=0}
type = Malware: Atk {Def=352, Atk=4950}
type = Phishing: Atk {Def=5302, Atk=12371}
type = SQL-Injection
|   os = Linux
|   |   country = Germany: Atk {Def=0, Atk=273}
|   |   country = Spain: Atk {Def=0, Atk=938}
|   |   country = Thailand: Atk {Def=0, Atk=560}
|   |   country = United States: Def {Def=3535, Atk=1763}
|   os = Windows: Atk {Def=0, Atk=1767}
```

ภาพที่ 3.21 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจ ในโปรแกรม RapidMiner Studio

ดังนั้นผู้วิเคราะห์ข้อมูลจะใช้เทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้ในการศึกษา เนื่องจากให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 12 กฎ ซึ่งสามารถนำไปใช้ในการ

แบ่งกลุ่มได้ตามเงื่อนไขได้ชัดเจน และสามารถนำกฎที่ได้ไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ดังนี้

กฎข้อที่ 1 IF type = SQL-Injection AND country = United States AND os = Linux THEN requirement Def หมายความว่า ถ้ารูปแบบการโจมตีเป็น SQL-Injection ในประเทศสหรัฐอเมริกา โดยใช้ระบบปฏิบัติการ Linux เป็นส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ

กฎข้อที่ 2 IF type = SQL-Injection AND country = United States AND os = Windows THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น SQL-Injection ในประเทศสหรัฐอเมริกา โดยใช้ระบบปฏิบัติการ Windows เป็นส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 3 IF type = SQL-Injection AND country = Spain THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น SQL-Injection ในประเทศสเปน ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 4 IF type = SQL-Injection AND country = Germany THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น SQL-Injection ในประเทศเยอรมนี ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 5 IF type = SQL-Injection AND country = Thailand THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น SQL-Injection ในประเทศไทย ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 6 IF type = Phishing THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น Phishing ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 7 IF type = Malware THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น Malware ส่วนใหญ่ที่พบเจอ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 8 IF type = DDOS AND web = Apache AND country = United States THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น DDOS แล้วใช้ Webserver เป็น Apache ในประเทศสหรัฐอเมริกา ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 9 IF type = DDOS AND web = Apache AND country = Spain THEN requirement = Def หมายความว่า ถ้ารูปแบบการโจมตีเป็น DDOS แล้วใช้ Webserver เป็น Apache ในประเทศสเปน ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ

กฎข้อที่ 10 IF type = DDOS AND web = Apache AND country = Germany THEN requirement = Atk หมายความว่า ถ้ารูปแบบการโจมตีเป็น DDOS แล้วใช้ Webserver เป็น Apache ในประเทศเยอรมนี ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี

กฎข้อที่ 11 IF type = DDOS AND web = Apache AND country = Thailand THEN requirement = Def หมายความว่า ถ้ารูปแบบการโจมตีเป็น DDOS แล้วใช้ Webserver เป็น Apache ในประเทศไทย ส่วนใหญ่ ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ

กฎข้อที่ 12 IF type = DDOS AND web = Nginx THEN requirement = Def หมายความว่า ถ้ารูปแบบการโจมตีเป็น DDOS แล้วใช้ Webserver เป็น Nginx ผลการพิจารณาพบว่า การโจมตีที่เกิดขึ้นนี้ให้ผลลัพธ์ว่า เป็นการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ

หลังจากผู้วิเคราะห์ข้อมูลเลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self-Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลการโจมตีที่เกิดขึ้นบนเว็บไซต์ ในปี 2019 ถึง ปี 2021 ที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งผู้วิเคราะห์ข้อมูลเลือกโปรแกรมที่ใช้นำเสนอ คือ โปรแกรม Weka 3.8.5 พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree (J48) พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้น มีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 75.55% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 24.44% และมีค่าความคลาดเคลื่อนเท่ากับ 0.401 และเมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.22 พบว่าการหาค่าของข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนาย แบ่งตามประเภทของวัตถุประสงค์การโจมตี การโจมตีเพื่อผลประโยชน์ส่วนตัวของผู้โจมตี (Atk) และการโจมตีเพื่อทดสอบระบบและพัฒนาระบบ (Def) และนำมาหาค่าเฉลี่ยรวมของทุก class ได้ค่าเฉลี่ยรวมเท่ากับ 0.75 มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้

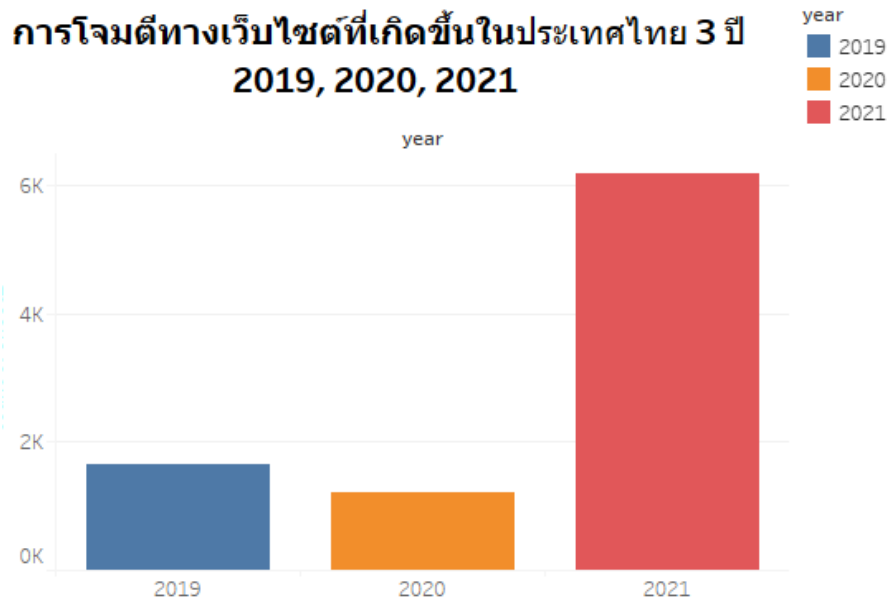
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       32044               75.5523 %
Incorrectly Classified Instances     10369               24.4477 %
Kappa statistic                          0.442
Mean absolute error                      0.3215
Root mean squared error                  0.401
Relative absolute error                 68.0079 %
Root relative squared error             82.473  %
Total Number of Instances            42413

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.488    0.078    0.795      0.488    0.605      0.470    0.808     0.758     Def
                 0.922    0.512    0.743      0.922    0.823      0.470    0.808     0.864     Atk
Weighted Avg.    0.756    0.346    0.763      0.756    0.739      0.470    0.808     0.823

=== Confusion Matrix ===

     a     b   <-- classified as
  7927  8330 |     a = Def
  2039 24117 |     b = Atk
```

ภาพที่ 3.22 แสดงผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree : J48 ในโปรแกรม Weka 3.8.5

3.1.6 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ

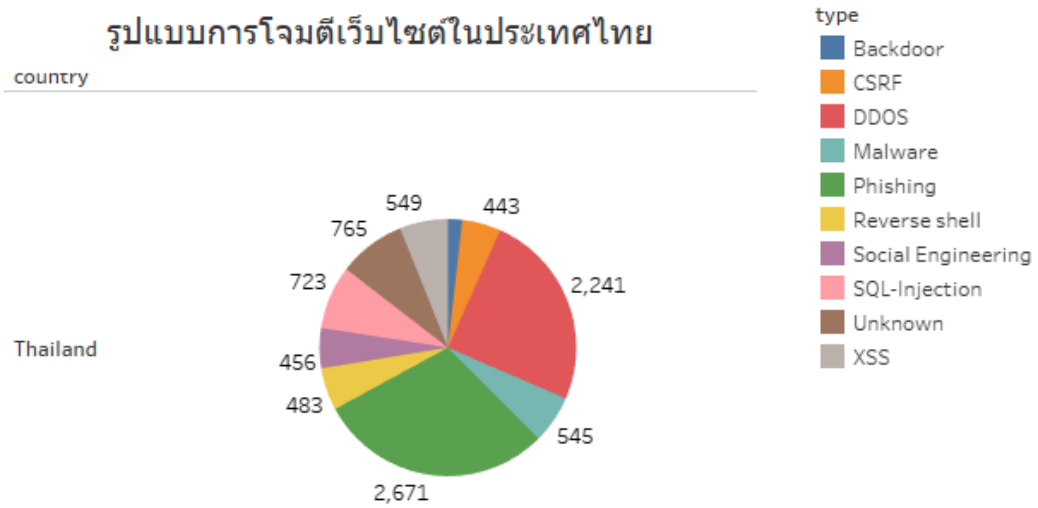
ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้าเว็บไซต์ ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม TableauPublic

1.) แสดงข้อมูลของประเทศไทยใน 3 ปี 2019, 2020, 2021 แบบ Bar Chart



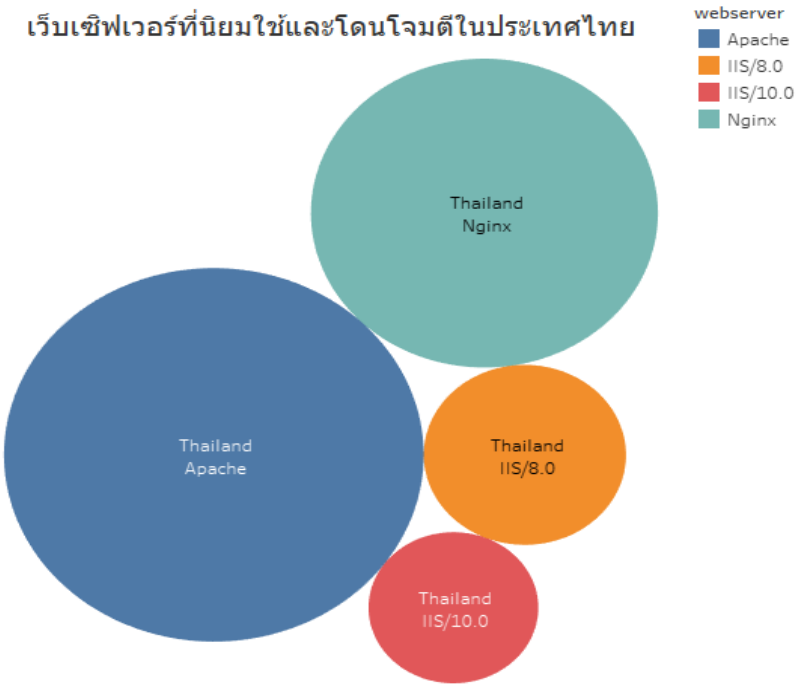
ภาพที่ 3.23 แสดงข้อมูลของประเทศไทยใน 3 ปี 2019, 2020, 2021 แบบ Bar Chart

2.) แสดงข้อมูลของรูปแบบการโจมตีเว็บไซต์ในประเทศไทย แบบ Pie Chart



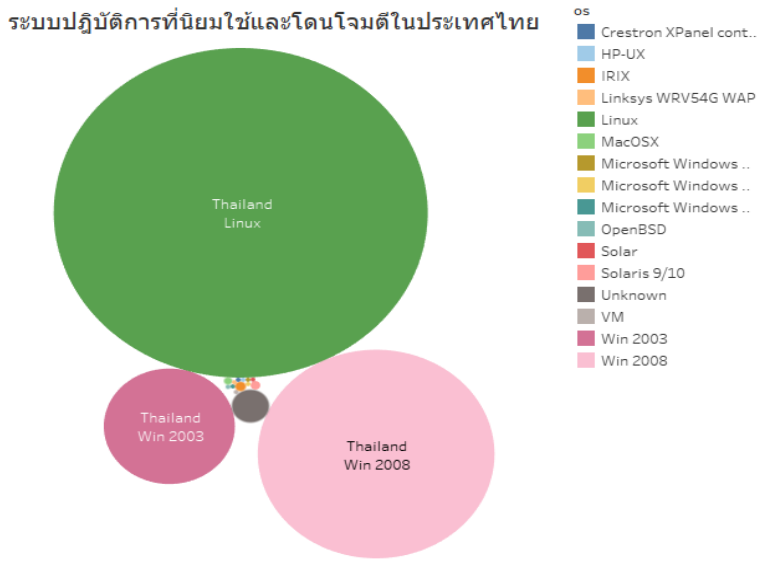
ภาพที่ 3.24 แสดงข้อมูลของรูปแบบการโจมตีเว็บไซต์ในประเทศไทย แบบ Pie Chart

3.) แสดงข้อมูลของเว็บเซิร์ฟเวอร์ที่นิยมใช้และโดนโจมตีในประเทศไทย แบบ Bubble Chart



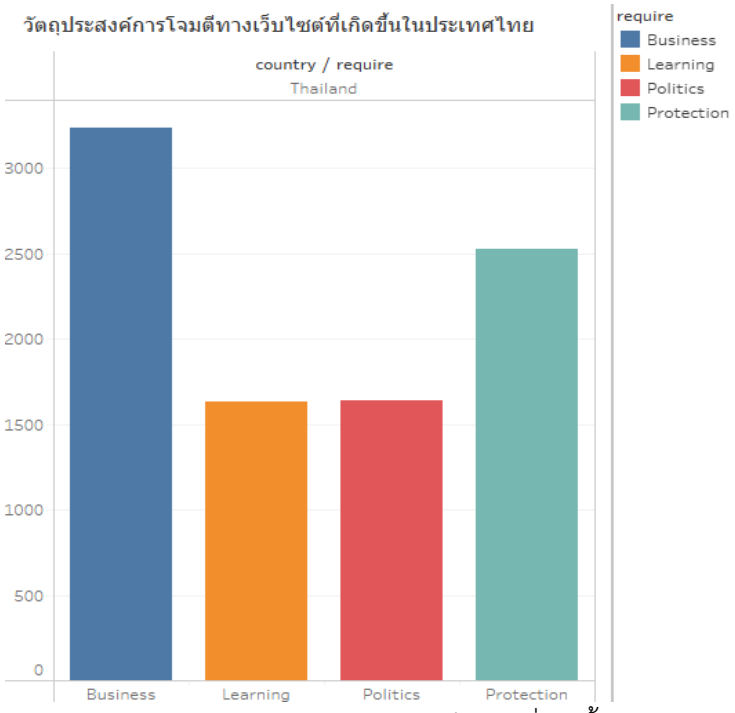
ภาพที่ 3.25 แสดงข้อมูลของเว็บเซิร์ฟเวอร์ที่นิยมใช้และโดนโจมตีในประเทศไทย แบบ Bubble Chart

4.) แสดงข้อมูลของระบบปฏิบัติการที่นิยมใช้และโดนโจมตีในประเทศไทย แบบ Bubble Chart



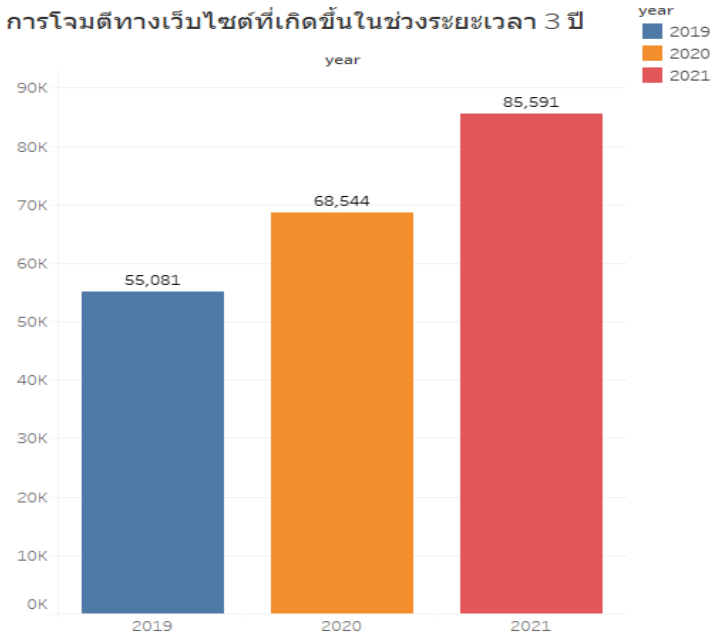
ภาพที่ 3.26 แสดงข้อมูลของระบบปฏิบัติการที่นิยมใช้และโดนโจมตีในประเทศไทย แบบ Bubble Chart

5.) แสดงข้อมูลของวัตถุประสงค์การโจมตีทางเว็บไซต์ที่เกิดขึ้นในประเทศไทย แบบ Bar Chart



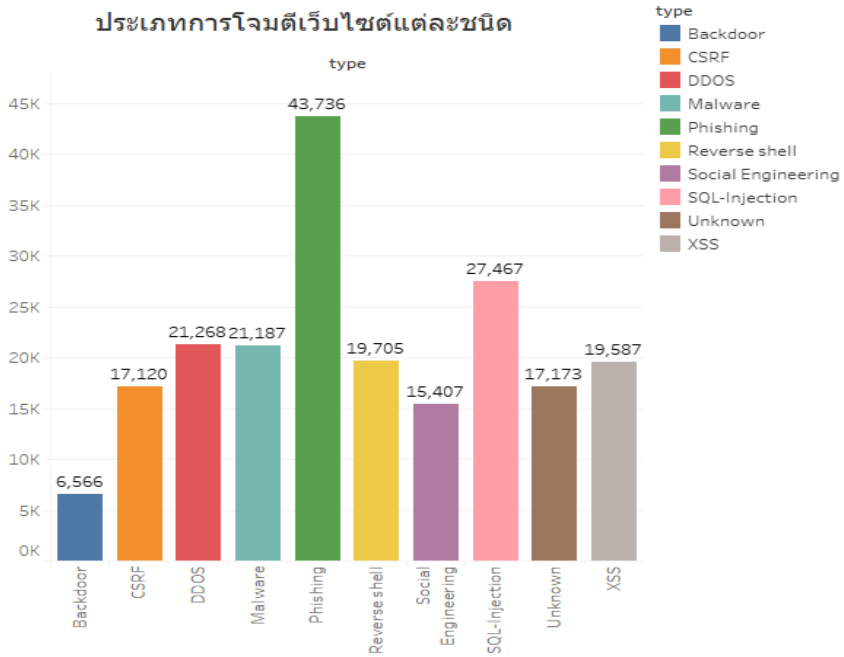
ภาพที่ 3.27 แสดงข้อมูลของวัตถุประสงค์การโจมตีทางเว็บไซต์ที่เกิดขึ้นในประเทศไทย แบบ Bar Chart

6.) แสดงข้อมูลของการโจมตีทางเว็บไซต์ที่เกิดขึ้นในช่วงระยะเวลา 3 ปี แบบ Bar Chart



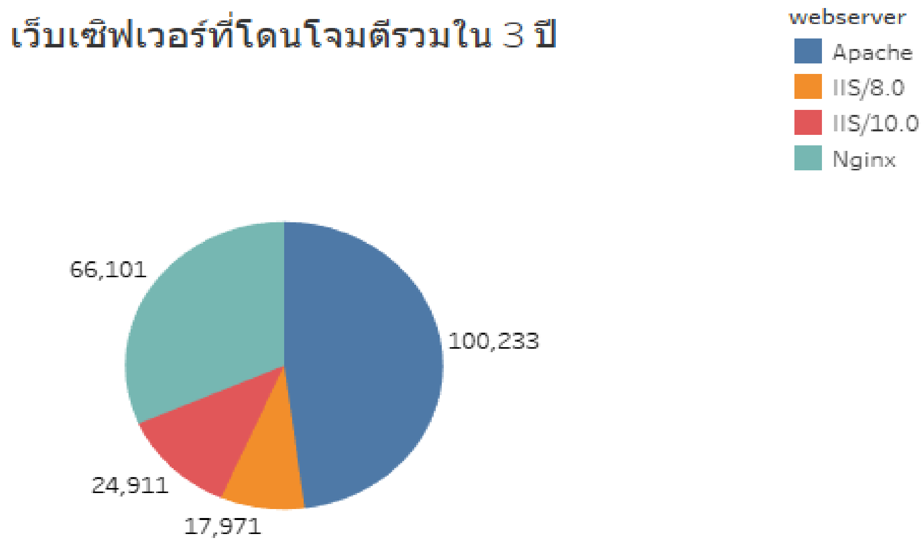
ภาพที่ 3.28 แสดงข้อมูลของการโจมตีทางเว็บไซต์ที่เกิดขึ้นในช่วงระยะเวลา 3 ปี แบบ Bar Chart

7.) แสดงข้อมูลของประเภทการโจมตีเว็บไซต์แต่ละชนิด แบบ Bar Chart



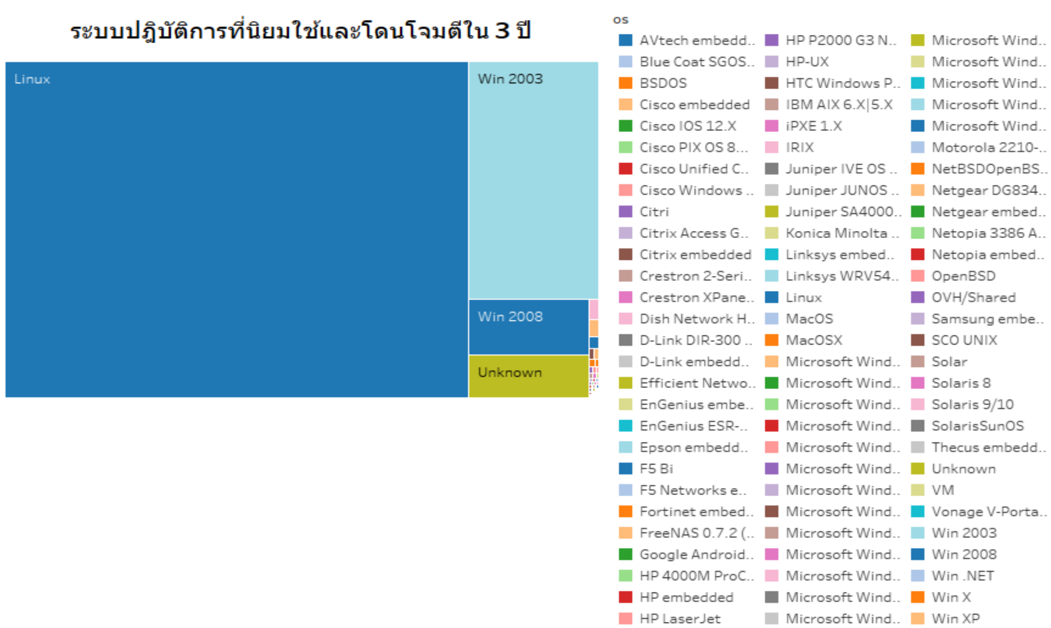
ภาพที่ 3.29 แสดงข้อมูลของประเภทการโจมตีเว็บไซต์แต่ละชนิด แบบ Bar Chart

8.) แสดงข้อมูลของเว็บเซิร์ฟเวอร์ที่โดนโจมตีรวมใน 3 ปี แบบ Pie Chart



ภาพที่ 3.30 แสดงข้อมูลของเว็บเซิร์ฟเวอร์ที่โดนโจมตีรวมใน 3 ปี แบบ Pie Chart

9.) แสดงข้อมูลของระบบปฏิบัติการที่นิยมใช้และโดนโจมตีใน 3 ปี แบบ Tree Maps Chart



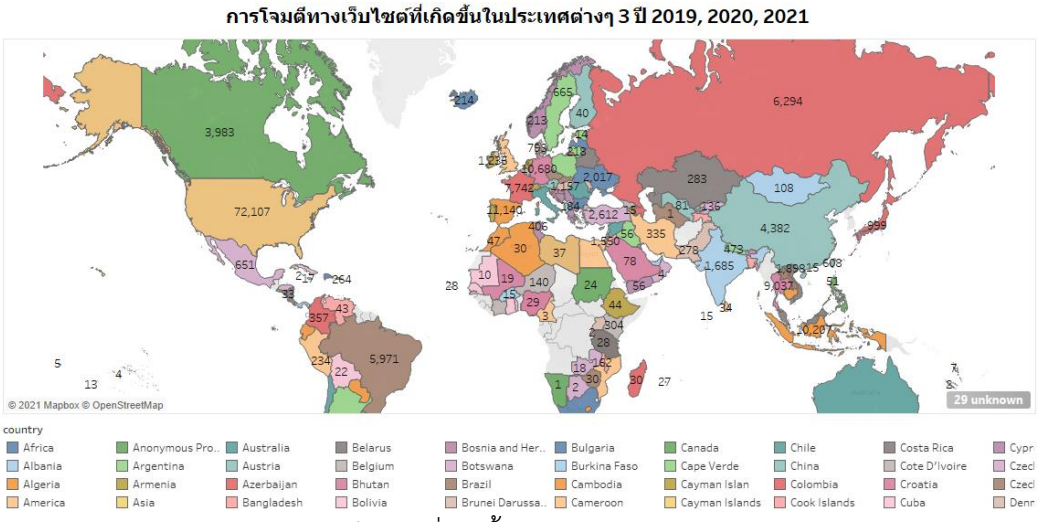
ภาพที่ 3.31 แสดงข้อมูลของระบบปฏิบัติการที่นิยมใช้และโดนโจมตีใน 3 ปี

แบบ Tree Maps Chart

10.) แสดงข้อมูลของวัตถุประสงค์การโจมตีทางเว็บไซต์ที่เกิดขึ้นใน 3 ปี แบบ Bubble Chart

ภาพที่ 3.32 แสดงข้อมูลของวัตถุประสงค์การโจมตีทางเว็บไซต์ที่เกิดขึ้นใน 3 ปี แบบ Bubble Chart

11.) แสดงข้อมูลของการโจมตีทางเว็บไซต์ที่เกิดขึ้นในประเทศต่างๆ 3 ปี 2019, 2020, 2021 แบบ GeoChart



ภาพที่ 3.33 แสดงข้อมูลของการโจมตีทางเว็บไซต์ที่เกิดขึ้นในประเทศต่างๆ 3 ปี 2019, 2020, 2021 แบบ GeoChart

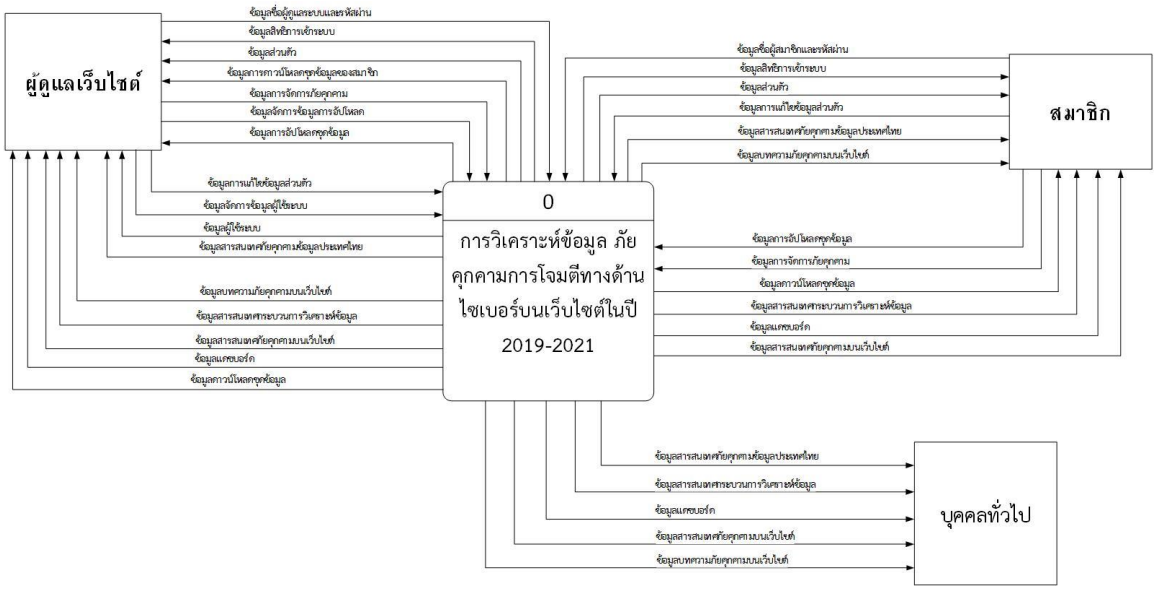
3.2 แผนภาพบริบท (Context Diagram)

แผนภาพบริบทเป็นแผนภาพที่แสดงถึงภาพรวมของระบบ และความสัมพันธ์ระบบกับสิ่งแวดล้อมที่เกี่ยวข้องกับระบบรวมถึงเหตุการณ์ต่าง ๆ ที่ใช้ในระบบการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 ซึ่งสามารถแบ่งออกมาได้ ดังนี้

ตารางที่ 3.1 ตารางรายละเอียดของเอ็กเทอร์นัลเอ็นทิตี้ และโปรเซสที่เกี่ยวข้อง

ผู้ใช้	รายการข้อมูล	รายการโปรเซส
1.) บุคคลทั่วไป 2.) สมาชิก 3.) ผู้ดูแลเว็บไซต์	1.) แฟ้มข้อมูลผู้ใช้งาน 2.) แฟ้มข้อมูลภัยคุกคามบนเว็บไซต์ 3.) แฟ้มข้อมูลการดาวน์โหลดไฟล์ 4.) แฟ้มข้อมูลการจัดเก็บไฟล์	1.) ตรวจสอบการเข้าสู่ระบบ 2.) จัดการข้อมูลส่วนตัว 3.) จัดการข้อมูลหน้าเว็บไซต์ 4.) รายงานข้อมูลสารสนเทศของเว็บไซต์

จากการกำหนดผู้ใช้และความต้องการ ที่ใช้ระบบ คือ บุคคลทั่วไป, สมาชิก และผู้ดูแลเว็บไซต์ สามารถแสดงความสัมพันธ์ด้วยแผนผังบริบทดังนี้



ภาพที่ 3.34 แผนภาพบริบท (Context Diagram)

จากรูปภาพที่ 3.34 เป็นแผนภาพบริบทระบบของการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 โดยสามารถแบ่งผู้ใช้ออกเป็น 3 ประเภทดังนี้

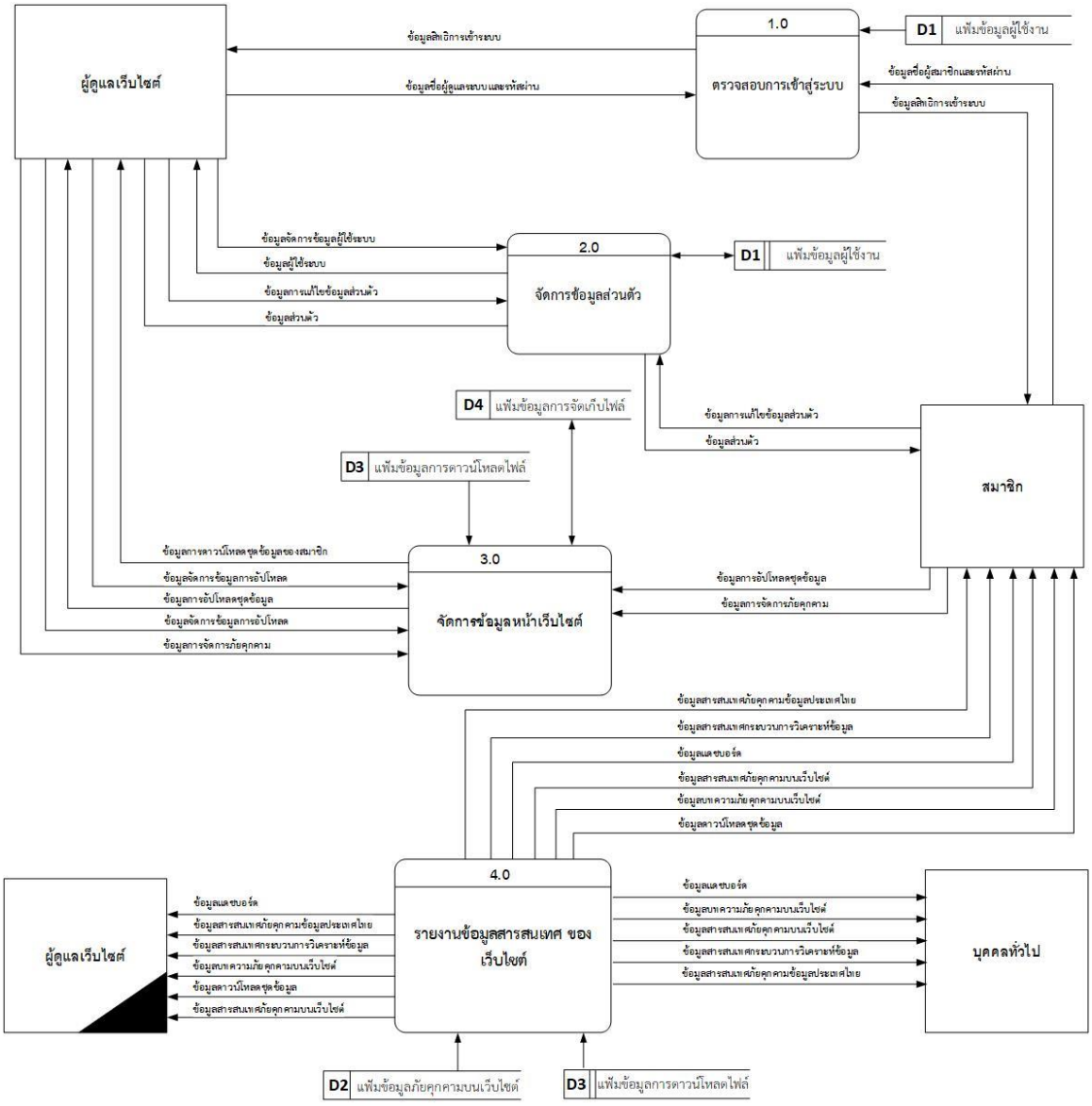
1.) บุคคลทั่วไป สามารถดูข้อมูลบทความเกี่ยวกับเว็บไซต์รูปแบบภัยคุกคามต่างๆ และเกี่ยวกับผู้จัดทำเว็บไซต์ได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ ของประเทศไทยแสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

2) สมาชิก สามารถกรอกข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ กรณีที่ ข้อมูลมีจำนวนน้อยได้ สามารถอัปโหลดชุดข้อมูลเกี่ยวกับภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ที่มีจำนวนมากๆ เพื่อให้ผู้ดูแลเว็บไซต์นำข้อมูลไปจัดการข้อมูลที่อัปเดตลงบนเว็บไซต์ สามารถดาวน์โหลดไฟล์ชุดข้อมูลที่ผู้ดูแลเว็บไซต์เปิดให้ดาวน์โหลดเพื่อนำไปศึกษาหรือใช้งานต่อได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ ของประเทศไทย แสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

3.) ผู้ดูแลเว็บไซต์ สามารถเพิ่ม ลบ และแก้ไขข้อมูลสารสนเทศการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ได้ สามารถจัดการสมาชิก ลบ แก้ไข ข้อมูล และดูข้อมูลสมาชิกได้ สามารถดาวน์โหลดไฟล์ชุดข้อมูลที่ ผู้ใช้งานอัปโหลดเข้ามา เพื่อนำไปทำ data cleaning ให้ข้อมูลนั้นพร้อมนำมาอัปเดตบนเว็บไซต์ได้ สามารถกรอกข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ กรณีที่ ข้อมูลมีจำนวนน้อยได้ สามารถอัปเดตชุดข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ที่มีจำนวนมากๆ เพื่อให้ข้อมูลที่อัปโหลดมานั้นนำใช้แสดงผลในรูปแบบของ data Visualization บนเว็บไซต์ได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ ของประเทศไทย แสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

3.3 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูลเป็นแผนภาพที่แสดงถึงกระบวนการทำงานต่าง ๆ ของระบบว่ามีผู้ใช้งานเกี่ยวข้องกับกระบวนการทำงานในด้านใดบ้าง และแสดงการไหลของข้อมูลในกระบวนการรวมถึงการจัดเก็บข้อมูล



ภาพที่ 3.35 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)

คำอธิบายกระบวยการ

ตารางที่ 3.2 DFD Number 1 ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์
DFD Number	1.0
Process Name	ตรวจสอบการเข้าสู่ระบบ
Input Data Flow	ข้อมูลชื่อผู้ดูแลระบบและรหัสผ่าน, ข้อมูลชื่อผู้สมาชิกและรหัสผ่าน
Output Data Flow	ข้อมูลสิทธิการเข้าระบบ
Data Store Used	แฟ้มข้อมูลผู้ใช้งาน
Description	เป็นกระบวนการตรวจสอบ ข้อมูลและรหัสผ่านข้อผู้ใช้ สำหรับ ผู้ดูแลเว็บไซต์และสมาชิก

ตารางที่ 3.3 DFD Number 2 จัดการข้อมูลส่วนตัว

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์
DFD Number	2.0
Process Name	จัดการข้อมูลส่วนตัว
Input Data Flow	ข้อมูลจัดการข้อมูลผู้ใช้ระบบ, ข้อมูลการแก้ไขข้อมูลส่วนตัว
Output Data Flow	ข้อมูลผู้ใช้ระบบ, ข้อมูลส่วนตัว
Data Store Used	แฟ้มข้อมูลผู้ใช้งาน
Description	เป็นการจัดการข้อมูลส่วนตัวของผู้ใช้ในระบบ

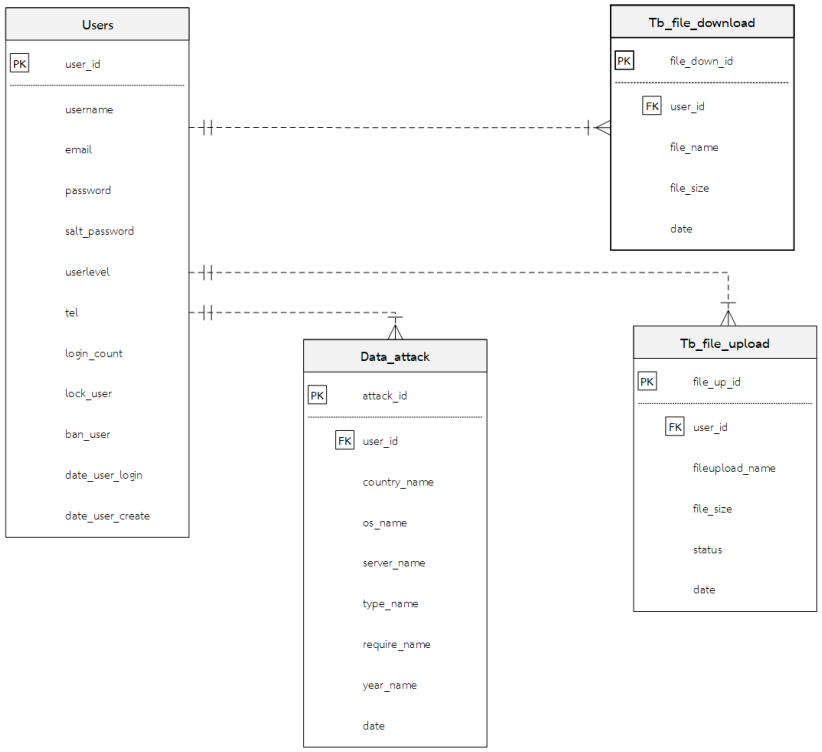
ตารางที่ 3.4 DFD Number 3 จัดการข้อมูลหน้าเว็บไซต์

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์
DFD Number	3.0
Process Name	จัดการข้อมูลหน้าเว็บไซต์
Input Data Flow	ข้อมูลจัดการข้อมูลการอัปโหลด, ข้อมูลจัดการข้อมูลการอัปโหลด, ข้อมูลการจัดการภัยคุกคาม, ข้อมูลการอัปโหลดชุดข้อมูล
Output Data Flow	ข้อมูลการดาวน์โหลดชุดข้อมูลของสมาชิก, ข้อมูลการอัปโหลดชุดข้อมูล
Data Store Used	แฟ้มข้อมูลการดาวน์โหลดไฟล์, แฟ้มข้อมูลการจัดเก็บไฟล์
Description	เป็นกระบวนการจัดการหน้าเว็บไซต์ ของผู้ดูแลเว็บไซต์ และสมาชิกเข้าใช้งานเว็บไซต์ เพิ่มกรอกข้อมูลภัยคุกคาม โดยจะมีผู้ดูแลเว็บไซต์ที่สามารถแก้ไขอัปเดตข้อมูลสารสนเทศภายในเว็บไซต์ได้

ตารางที่ 3.5 DFD Number 3.1 รายงานข้อมูลสารสนเทศ ของเว็บไซต์

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์
DFD Number	4.0
Process Name	รายงานข้อมูลสารสนเทศ ของเว็บไซต์
Input Data Flow	แฟ้มข้อมูลภัยคุกคามบนเว็บไซต์, แฟ้มข้อมูลการดาวน์โหลดไฟล์
Output Data Flow	ข้อมูลสารสนเทศภัยคุกคามข้อมูลประเทศไทย, ข้อมูลสารสนเทศกระบวนการวิเคราะห์ข้อมูล, ข้อมูลบทความภัยคุกคามบนเว็บไซต์, ข้อมูลดาวน์โหลดชุดข้อมูล, ข้อมูลสารสนเทศภัยคุกคามบนเว็บไซต์, ข้อมูลแดชบอร์ด
Data Store Used	แฟ้มข้อมูลภัยคุกคามบนเว็บไซต์, แฟ้มข้อมูลการดาวน์โหลดไฟล์
Description	เป็นกระบวนการแสดงผลลัพธ์แดชบอร์ดเกี่ยวกับภัยคุกคามบนเว็บไซต์ให้กับเว็บไซต์ โดยแสดงผลในแบบ data visualization โดยสิทธิสมาชิกสามารถดาวน์โหลดข้อมูลได้

3.4 ความสัมพันธ์ของข้อมูล (ER-Diagram)



ภาพที่ 3.36 แสดงภาพความสัมพันธ์ของข้อมูลการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021

จากภาพที่ 3.36 แสดงภาพความสัมพันธ์ของข้อมูลระบบการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 ได้ดังนี้

1.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง Data_attack คือ ผู้ใช้งานหนึ่งคนสามารถเพิ่มข้อมูลภัยคุกคามได้หลายข้อมูลภัยคุกคาม และแต่ละข้อมูลภัยคุกคามจะมีผู้ใช้งานจัดการข้อมูลเพียงคนเดียวเท่านั้น

2.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง Tb_file_download คือ ผู้ใช้งานหนึ่งคนสามารถดาวน์โหลดชุดข้อมูลได้หลายไฟล์ และแต่ละไฟล์ชุดมูลจะมีผู้ใช้งานดาวน์โหลดชุดข้อมูลเพียงคนเดียวเท่านั้น

3.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง Tb_up_download คือ ผู้ใช้งานหนึ่งคนสามารถอัปโหลดชุดข้อมูลได้หลายไฟล์ และแต่ละไฟล์ชุดมูลจะมีผู้ใช้งานอัปโหลดชุดข้อมูลเพียงคนเดียวเท่านั้น

พจนานุกรมข้อมูล (Data Dictionary)

การวิเคราะห์เพื่อให้ได้มาซึ่งแผนภาพอีอาร์ หรืออีอาร์ไดอะแกรมนั้นจะให้พื้นฐานหลักอยู่ 3 ประการด้วยกัน ได้แก่

3.3.1 เอ็นติตี้ (Entity) คือ บุคคล วัตถุ สถานที่ และรวมถึงเหตุการณ์ที่ทำให้เกิดกลุ่มของข้อมูลที่ต้องการจัดเก็บ ซึ่งบ่งชี้ถึงความเป็นเอกลักษณ์เฉพาะตัวได้ (Uniquely identifiable)

3.3.2 ความสัมพันธ์(Relation) คือ ค่าความสัมพันธ์ระหว่างเอ็นติตี้

3.3.3 แอททริบิวต์(Attribute) คือ คุณสมบัติของเอ็นติต

ตารางที่ 3.6 แสดงเอ็นติตี้ทั้งหมดภายในกระบวนการของการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021

ลำดับ	ชื่อตาราง	ประเภท	รายละเอียด
D1	Users	Master	แฟ้มข้อมูลผู้ใช้งาน
D2	Data_attack	Transaction	แฟ้มข้อมูลภัยคุกคามบนเว็บไซต์
D3	Tb_file_download	Master	แฟ้มข้อมูลการดาวน์โหลดไฟล์
D4	Tb_file_upload	Master	แฟ้มข้อมูลการจัดเก็บไฟล์

คำอธิบาย: ประเภทของตาราง ได้แก่

master หมายถึง ตารางข้อมูลหลัก

transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล

reference หมายถึง ตารางที่มีการอ้างอิง

คำอธิบาย: ประเภทของตาราง ได้แก่ master หมายถึง ตารางข้อมูลหลัก transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล reference หมายถึง ตารางที่มีการอ้างอิง

ตารางที่ 3.7 แสดงรายละเอียดของตาราง users

ชื่อตาราง : Users ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูลรวม คีย์หลัก : user_id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
user_id	VARCHAR(6)	รหัสผู้ใช้งาน	4
username	VARCHAR(100)	ผู้สมาชิก	admin
e-mail	VARCHAR(100)	อีเมล์	admin@gmail.com
password	VARCHAR(50)	รหัสผ่าน	1234
salt_password	VARCHAR(250)	ตัวต่อรหัสผ่าน	1fa6f421a8410d00
userlevel	INT(1)	สิทธิ์	1,2
tel	VACHAR(10)	เบอร์โทร	0891234567
login_count	INT(1)	จำนวนการเข้าสู่ระบบ	3
lock_user	INT(1)	ล็อคการเข้าสู่ระบบ	1
ban_user	datetime	เวลาการล็อค	2021-12-28 08:11:04
date_user_login	datetime	เวลาการเข้าสู่ระบบ	2021-12-31 23:15:44
date_user_create	datetime	เวลาการสมัคร	2021-11-13 09:55:04

คำอธิบาย: สิทธิ์ ได้แก่

หมายเลข 1 หมายถึง admin ผู้ดูแลเว็บไซต์

หมายเลข 2 หมายถึง member ผู้สมาชิกเว็บไซต์

คำอธิบาย: ตัวต่อรหัสผ่าน คือ

1fa6f421a8410d00 หมายถึง รหัสผ่านชั้น 2 ที่ถูกเข้ารหัสด้วย HEX แล้วไปต่อกับรหัสผ่านจริง ตัวอย่างเช่น **1234**1fa6f421a8410d00

คำอธิบาย: จำนวนการเข้าสู่ระบบ

หมายถึง เมื่อมีการเข้าสู่ระบบที่ไม่สำเร็จ จะนับค่าขึ้นไปทีละ 1

คำอธิบาย: ล็อคการเข้าสู่ระบบ ได้แก่

หมายเลข 0 หมายถึง ยังไม่มีการระงับการเข้าสู่ระบบ

หมายเลข 1 หมายถึง มีการระงับการเข้าสู่ระบบ

ตารางที่ 3.8 แสดงรายละเอียดของตาราง Data_attack

ชื่อตาราง : Data_attack ประเภทตาราง : Master คำอธิบาย : เก็บข้อมูลผู้ใช้งาน คีย์หลัก : attack_id คีย์รอง : user_id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
attack_id	INT(11)	ลำดับข้อมูล	489
country_name	VARCHAR(150)	ชื่อประเทศ	thailand
os_name	VARCHAR(100)	ชื่อระบบปฏิบัติการ	linux
server_name	VARCHAR(100)	ชื่อเซิร์ฟเวอร์	apache
type_name	VARCHAR(100)	ชื่อรูปแบบการโจมตี	DDOS
require_name	VARCHAR(100)	ชื่อวัตถุประสงค์	learning
year_name	VARCHAR(4)	ปี	2021
date	DATE	วัน/เดือน/ปี	31/12/2021

ตารางที่ 3.9 แสดงรายละเอียดของตาราง Tb_file_download

ชื่อตาราง : Tb_file_download ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูลการจัดเก็บไฟล์ คีย์หลัก : file_down_id คีย์รอง : user_id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
file_down_id	INT(11)	ลำดับไฟล์	21
file_name	varchar(100)	ชื่อชุดข้อมูล	Data_2021.csv
file_size	varchar(100)	ไฟล์ไซต์	10 MB
date	DATE	วัน/เดือน/ปี	10/01/2022

ตารางที่ 3.10 แสดงรายละเอียดของตาราง Tb_file_download

ชื่อตาราง : Tb_file_upload ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูลการจัดเก็บไฟล์ คีย์หลัก : file_up_id คีย์รอง : user_id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
file_up_id	INT(11)	ลำดับไฟล์	21
fileupload_name	varchar(100)	ชื่อชุดข้อมูล	Data_2021.csv
file_size	varchar(100)	ไฟล์ไซต์	10 MB
File_status	INT(1)	สถานะไฟล์	1
date	DATE	วัน/เดือน/ปี	03/01/2022

คำอธิบาย: สถานะไฟล์ได้แก่

หมายเลข 0 หมายถึง รอการอนุมัติดำเนินการ

หมายเลข 1 หมายถึง อยู่ระหว่างดำเนินการ

หมายเลข 2 หมายถึง ดำเนินการเสร็จสิ้นแล้ว

3.5 ออกแบบหน้าเว็บไซต์

โครงสร้างระบบการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021

ภาพที่ 3.37 แสดงภาพโครงสร้างระบบการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021

การออกแบบ Wireframe หน้าจอเว็บไซต์ มีการออกแบบโดยแบ่งตามประเภทของผู้ใช้งาน ซึ่งประกอบด้วย 3 ผู้ใช้งาน ดังนี้

3.5.1.) บุคคลทั่วไป

3.5.2.) สมาชิก

3.5.3.) ผู้ดูแลเว็บไซต์

3.5.1.) บุคคลทั่วไป

1.) หน้าแรกเว็บไซต์แสดงเมนูต่างๆ บนหน้าเว็บไซต์

ภาพที่ 3.38 หน้าแรกเว็บไซต์แสดงเมนูต่างๆ บนหน้าเว็บไซต์

2.) หน้าแสดงบทความเกี่ยวกับภัยคุกคามบนเว็บไซต์

ภาพที่ 3.39 หน้าแสดงบทความเกี่ยวกับภัยคุกคามบนเว็บไซต์

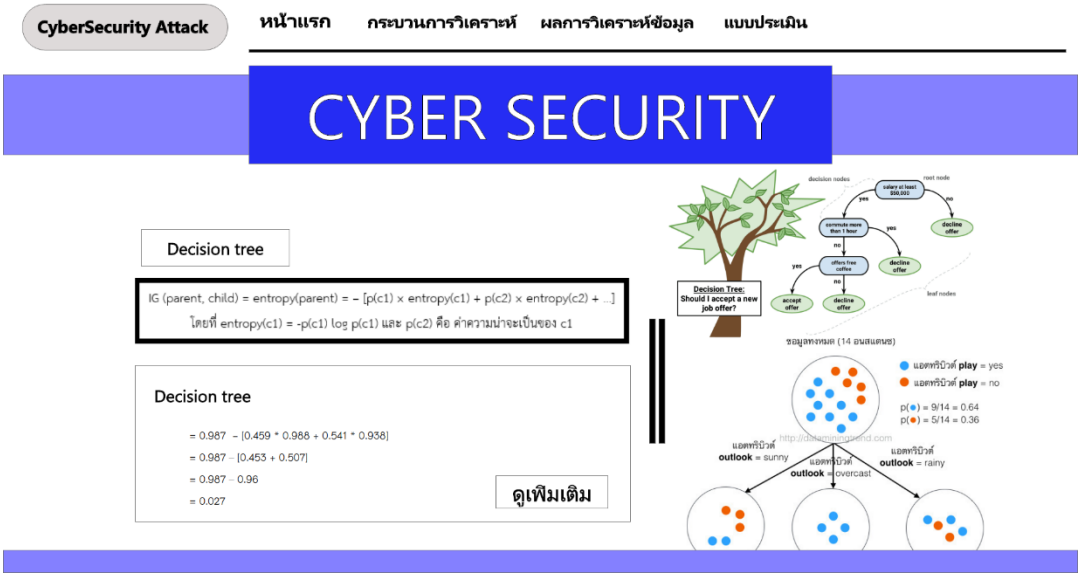
3.) หน้าแสดงรูปแบบการโจมภัยคุกคามบนเว็บไซต์ที่เกิดขึ้น

ภาพที่ 3.40 หน้าแสดงรูปแบบการโจมภัยคุกคามบนเว็บไซต์ที่เกิดขึ้น

4.) หน้าแสดงกฎต้นไม้ตัดสินใจ decision tree

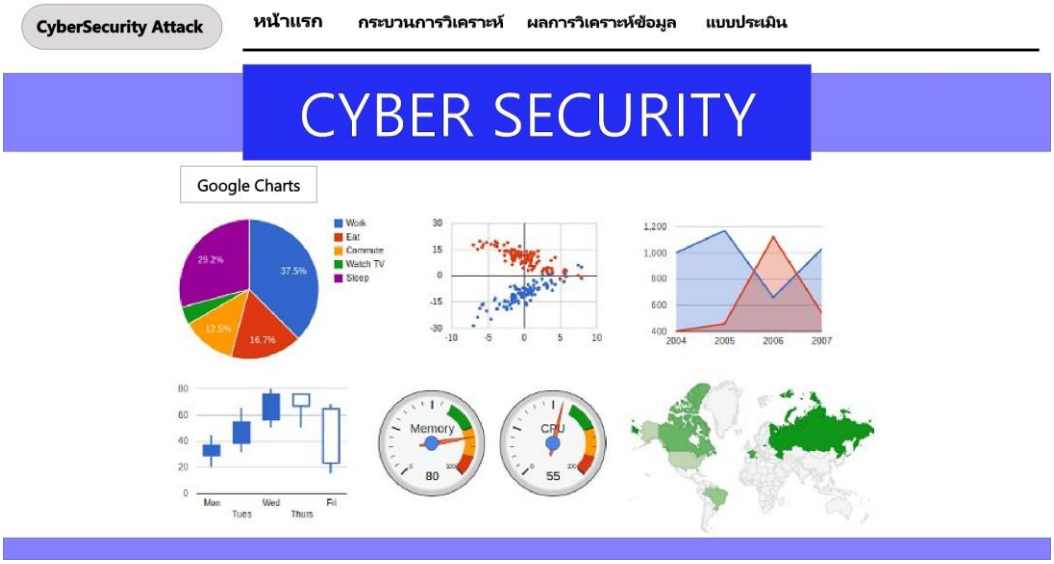
ภาพที่ 3.41 หน้าแสดงกฎต้นไม้ตัดสินใจ decision tree

5.) หน้าการวิเคราะห์ข้อมูล เมนูกระบวนการวิเคราะห์ข้อมูล ด้วย decision tree



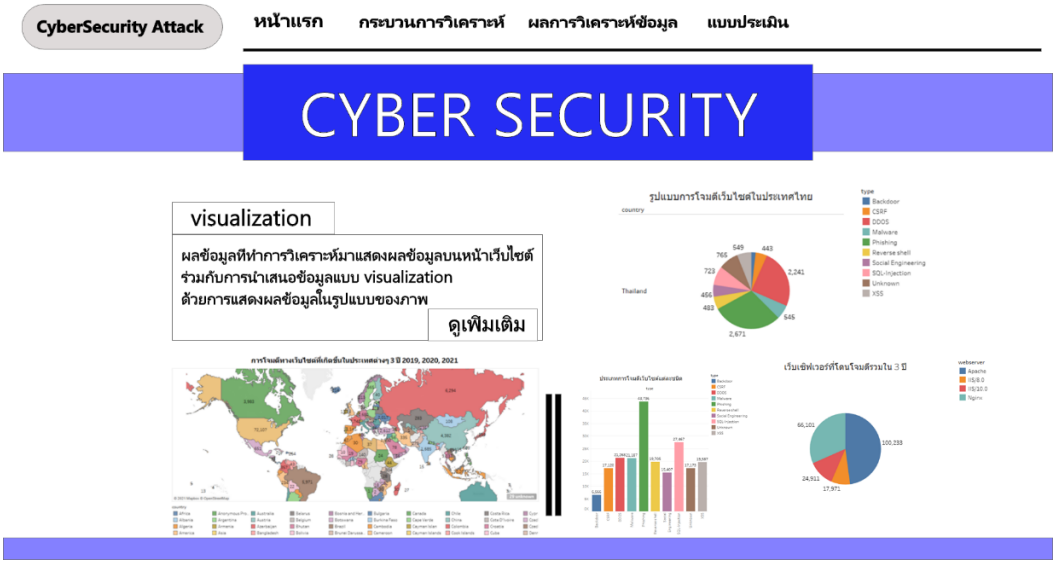
ภาพที่ 3.42 หน้าการวิเคราะห์ข้อมูล เมนูกระบวนการวิเคราะห์ข้อมูล ด้วย decision tree

6.) หน้าแสดงข้อมูล แบบแดชบอร์ดโดยแสดงผลแบบ visualization ด้วย Google charts



ภาพที่ 3.43 หน้าแสดงข้อมูล แบบแดชบอร์ดโดยแสดงผลแบบ visualization ด้วย Google charts

7.) หน้าหน้าแสดงข้อมูล แบบแดชบอร์ดโดยแสดงผลแบบ visualization Tableau Public



ภาพที่ 3.44 หน้าหน้าแสดงข้อมูล แบบแดชบอร์ดโดยแสดงผลแบบ visualization Tableau Public

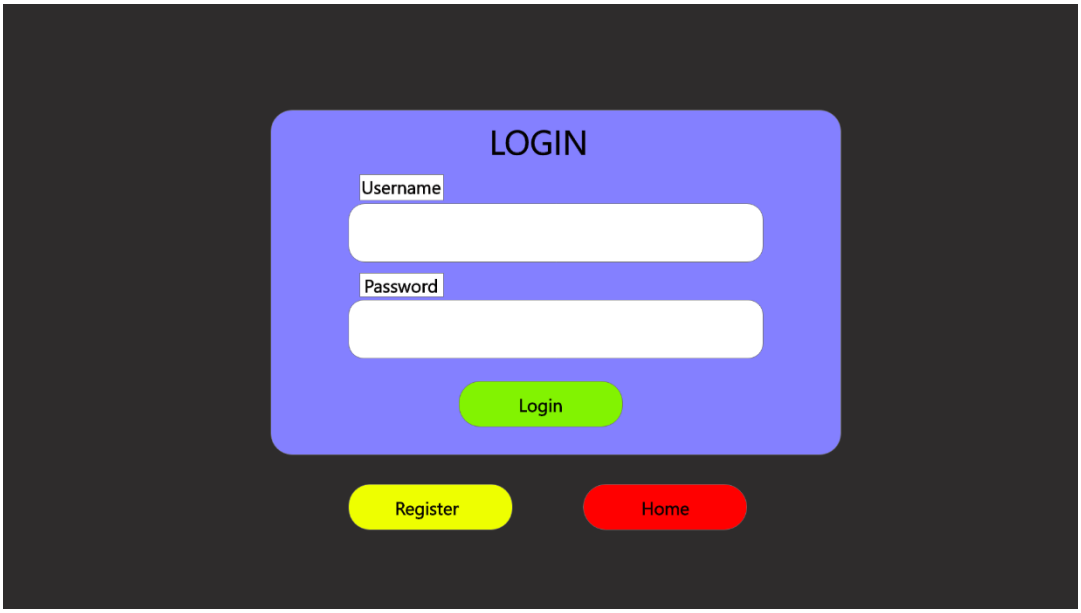
8.) หน้ากรอกแบบประเมิน ความพึงพอใจหน้าเว็บไซต์

ภาพที่ 3.45 หน้ากรอกแบบประเมิน ความพึงพอใจหน้าเว็บไซต์

9.) หน้าเกี่ยวกับ ประวัติของผู้จัดทำโครงงาน

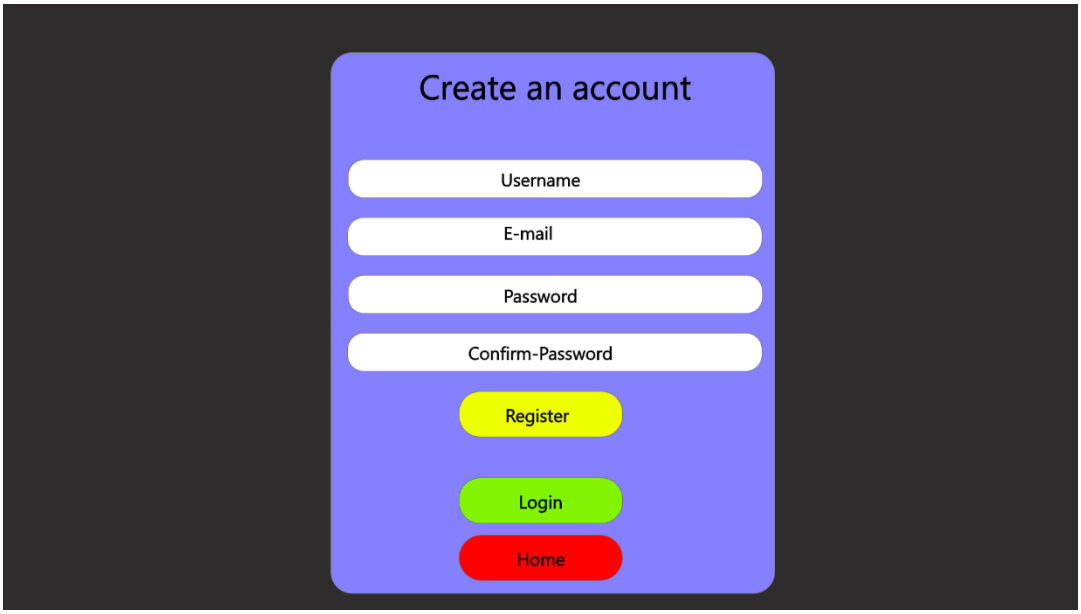
ภาพที่ 3.46 หน้าเกี่ยวกับ ประวัติของผู้จัดทำโครงงาน

10.) หน้าล็อคอินเข้าสู่ระบบ เพื่อใช้งาน



ภาพที่ 3.47 หน้าล็อคอินเข้าสู่ระบบ เพื่อใช้งาน

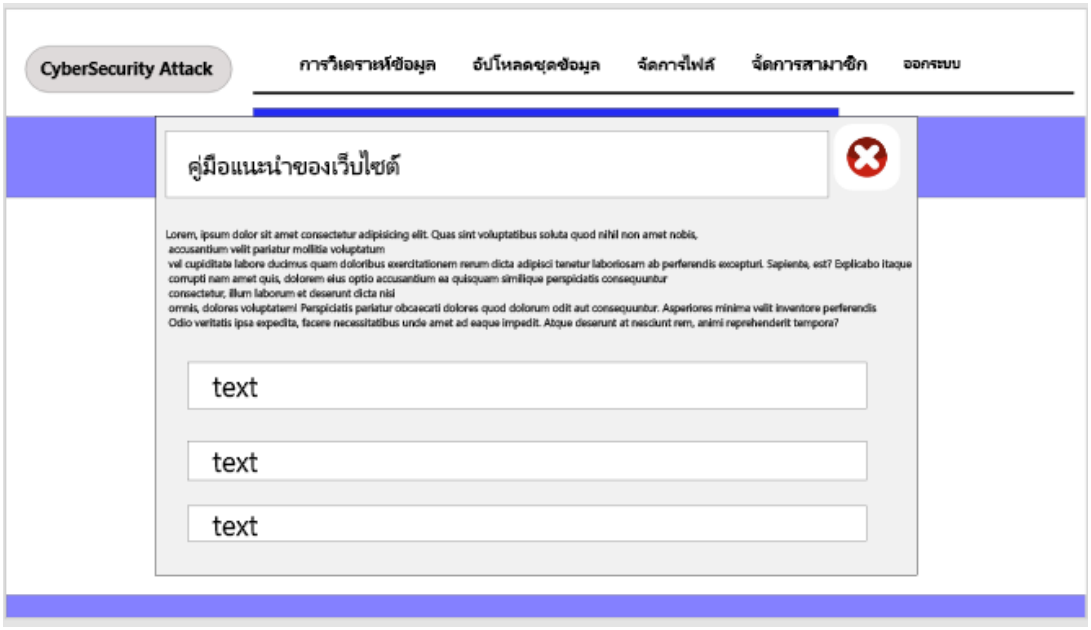
11.) หน้ากรอกข้อมูลเพื่อสมัครสมาชิก



ภาพที่ 3.48 หน้ากรอกข้อมูลเพื่อสมัครสมาชิก

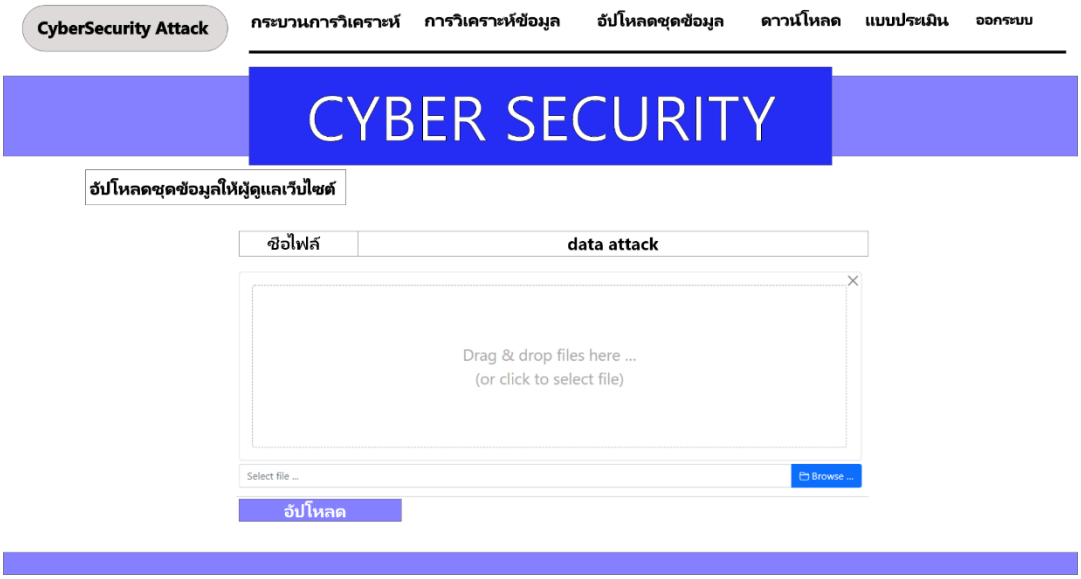
3.5.2.) สมาชิก

1.) หน้าแสดงคู่มือแนะนำของเว็บไซต์



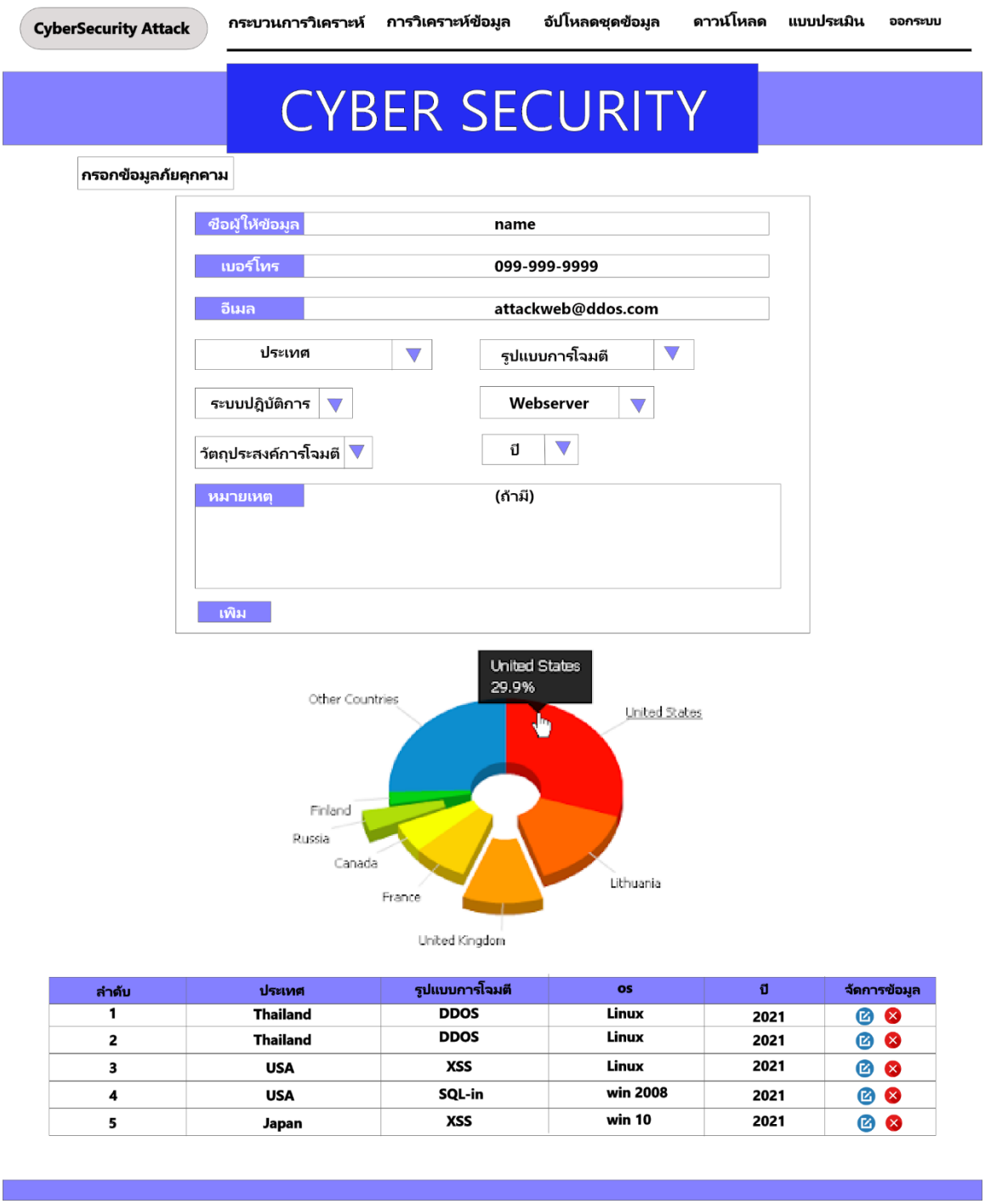
ภาพที่ 3.49 หน้าแสดงคู่มือแนะนำของเว็บไซต์

2.) หน้าอัปโหลดชุดข้อมูลให้กับผู้ดูแลเว็บไซต์



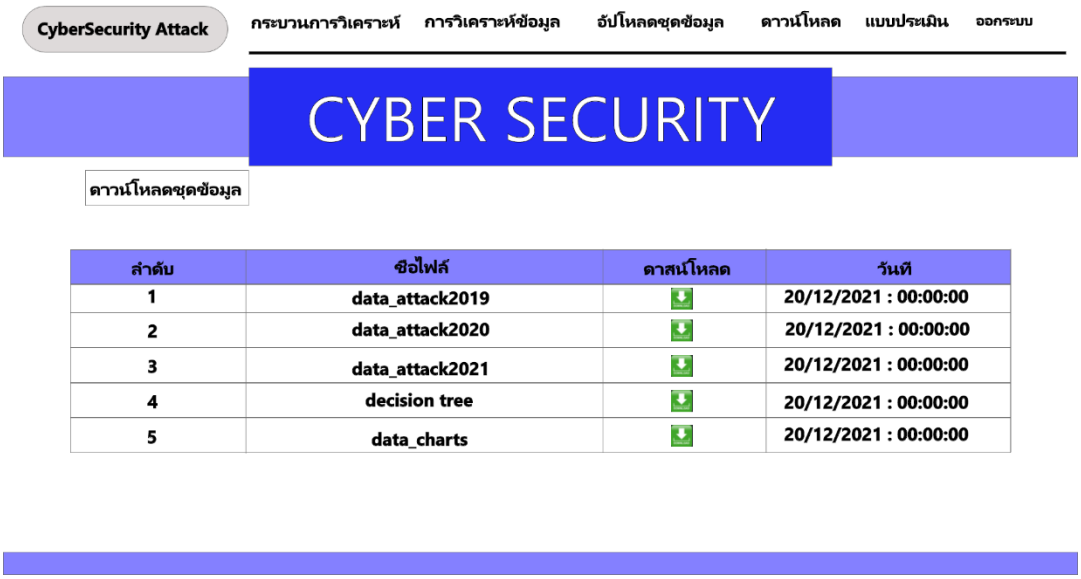
ภาพที่ 3.50 หน้าอัปโหลดชุดข้อมูลให้กับผู้ดูแลเว็บไซต์

3.) หน้ากรอกข้อมูลภัยคุกคามบนเว็บไซต์เพิ่มเติม



ภาพที่ 3.51 หน้ากรอกข้อมูลภัยคุกคามบนเว็บไซต์เพิ่มเติม

4.) หน้าตารางข้อมูล ชื่อไฟล์ต่างๆ เพื่อให้ดาวน์โหลด



ภาพที่ 3.52 หน้าตารางข้อมูล ชื่อไฟล์ต่างๆ เพื่อให้ดาวน์โหลด

5.) หน้าแก้ไขข้อมูลส่วนตัวของผู้ใช้งาน

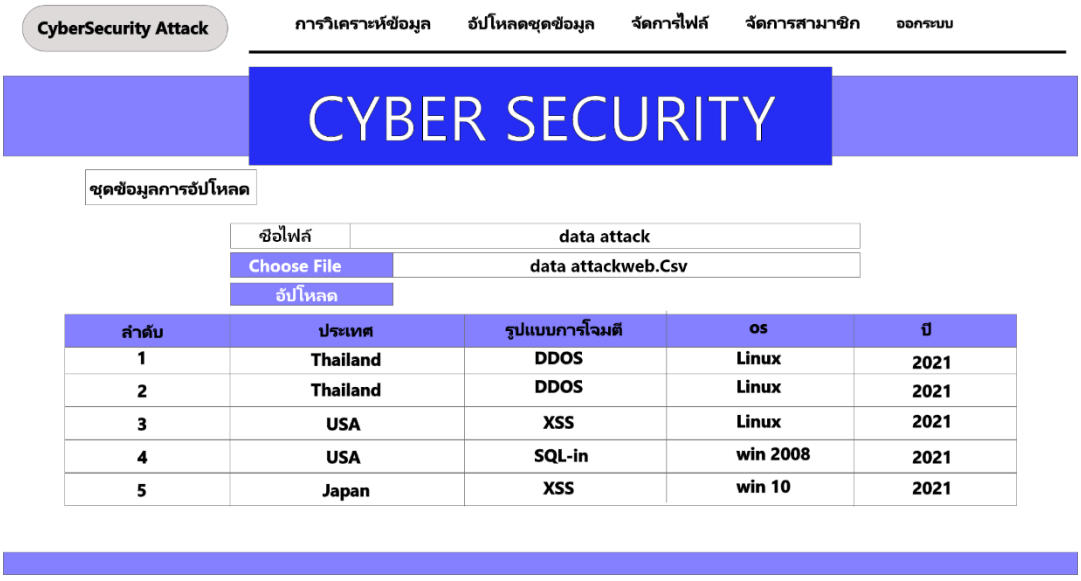
ภาพที่ 3.53 หน้าแก้ไขข้อมูลส่วนตัวของผู้ใช้งาน

3.5.3.) ผู้ดูแลเว็บไซต์

1.) หน้าแรกของผู้ดูแลเว็บไซต์

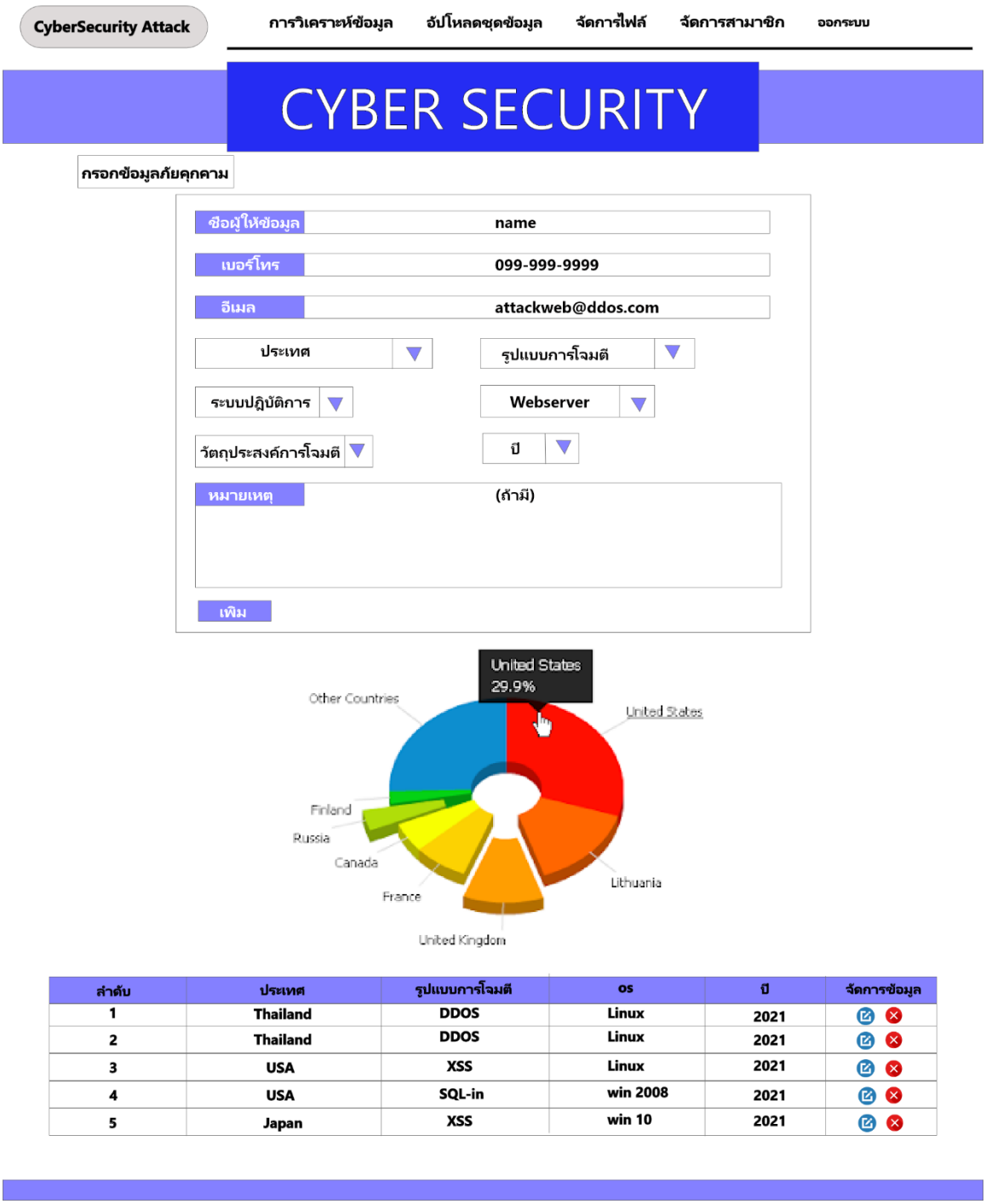
ภาพที่ 3.54 หน้าแรกของผู้ดูแลเว็บไซต์

2.) หน้าของการอัปโหลดชุดข้อมูล เพื่อให้สมาชิกสามารถเข้ามาดาวน์โหลดไปได้



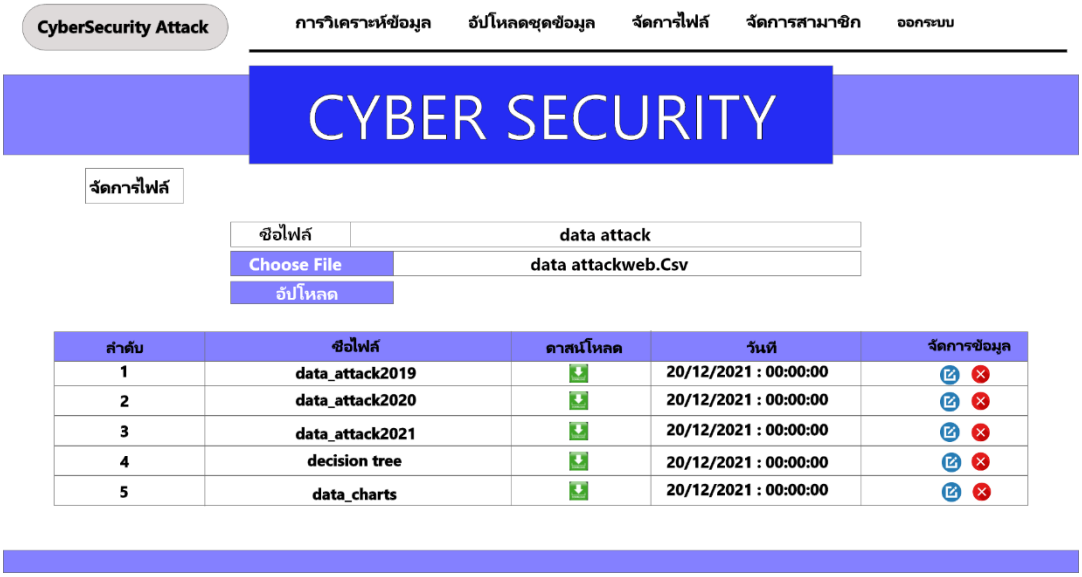
ภาพที่ 3.55 หน้าของการอัปโหลดชุดข้อมูล เพื่อให้สมาชิกสามารถเข้ามาดาวน์โหลดไปได้

3.) หน้ากรอกข้อมูลภัยคุกคามบนเว็บไซต์เพิ่มเติม



ภาพที่ 3.56 หน้ากรอกข้อมูลภัยคุกคามบนเว็บไซต์เพิ่มเติม

4.) หน้าของการจัดการไฟล์ ลบ แก้ไข ต่างๆ



ภาพที่ 3.57 หน้าของการจัดการไฟล์ ลบ แก้ไข ต่างๆ

5.) หน้าของการจัดการสมาชิก ลบ แก้ไข ต่างๆ

ภาพที่ 3.58 หน้าของการจัดการสมาชิก ลบ แก้ไข ต่างๆ

6.) หน้าแก้ไขข้อมูลส่วนตัวของผู้ดูแลเว็บไซต์

ภาพที่ 3.59 หน้าแก้ไขข้อมูลส่วนตัวของผู้ดูแลเว็บไซต์

3.6 บทสรุป

จากวิธีการดำเนินงานโครงงานในการคำนวน โมเดลต้นไม้ตัดสินใจ จากการคำนวณด้วยมือนี้ ผู้วิเคราะห์ข้อมูลได้ผลลัพธ์ว่า โมเดลต้นไม้ตัดสินใจ Root node ที่คือ แอตทริบิวต์ Type และได้ interior node คือ แอตทริบิวต์ Country , แอตทริบิวต์ Web server และ leaf node คือ แอตทริบิวต์ Os ซึ่งไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด ก็จะได้ผลลัพธ์ที่ แอตทริบิวต์ Os linux , nginx และ แอตทริบิวต์ Country United States , Spain , Germany และ Thailand ทางผู้วิเคราะห์ข้อมูลได้ทำการแสดงวิธีในการจัดการกับข้อมูลของการโจมตีที่เกิดขึ้นบนเว็บไซต์ในปี 2019 ถึง ปี 2021 ด้วยขั้นตอนการวิเคราะห์ข้อมูลด้วย CRISP-DM อย่างละเอียดรวมไปถึงการสร้างโมเดล Decision Tree จากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อนำเสนอ คือ โปรแกรม Weka ในการสร้างโมเดล Decision Tree และตรวจสอบข้อมูลผ่านทางโปรแกรม RapidMiner Studio เพื่อความแม่นยำ เมื่อนำไปเปรียบเทียบกับผลการคำนวณ โมเดล Decision Tree ด้วยตัวเอง ทางผู้วิเคราะห์ข้อมูลพบว่าทั้ง 3 โมเดลได้ผลลัพธ์ความแม่นยำของโมเดลที่ เหมือนกัน และสามารถนำโมเดลไปใช้งานได้ และประเมินประสิทธิภาพของโมเดล ซึ่งมีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 75.55% จำนวนทั้งสิ้น 42,414 รายการ เมื่อเปรียบเทียบกับงานวิจัยการพัฒนาแบบจำลองของ วนิดา พงษ์สงวน , ทิพยา ถินสูงเนิน , มาโนช ถินสูงเนิน (2559) การพัฒนาแบบจำลองปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ จำนวนทั้งสิ้น 44,002 รายการ พบว่าโมเดลแบบจำลองมีค่าความถูกต้องเฉลี่ยที่ใกล้เคียงกันที่ 76.14% ซึ่งเป็นค่าที่สามารถยอมรับได้ในการ ทดสอบประสิทธิภาพของ Model ด้วยวิธี Self-Consistency Test และทุกโมเดลพบว่ามีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี ทางผู้วิเคราะห์ข้อมูลได้นำข้อมูลสารสนเทศนั้นมาทำการแสดงผลแบบ Visualization โดยใช้โปรแกรม Tableau Public และ Google Charts API มาใช้งานร่วมดด้วย จากนั้นออกแบบ Wireframe ของเว็บไซต์ที่จะนำข้อมูลมาเผยแพร์บน เว็บไซต์ ด้วยโปรแกรม Adobe XD