

ชื่อโครงการ	การวิเคราะห์ข้อมูล ภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์
โดย	นาย ศิริพงษ์ จินตุง รหัสนักศึกษา 60541207050-6 นาย ณัฐวัตร ชุ่มเชื้อ รหัสนักศึกษา 60541207022-5
อาจารย์ที่ปรึกษา	อาจารย์ จตุพร ศิลพรชัย
หลักสูตร	ระบบสารสนเทศทางธุรกิจ สาขาบริหารธุรกิจ คณะบริหารธุรกิจ และศิลปศาสตร์
ปีการศึกษา	2564

บทคัดย่อ

การจัดทำโครงการในครั้งนี้มีวัตถุประสงค์ เพื่อการวิเคราะห์ข้อมูล ภัยคุกคามการโจมตีทางด้านไซเบอร์บนเว็บไซต์ในปี 2019-2021 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ มาประมวลผลข้อมูล แยกแยะและพิจารณาข้อมูลผ่านโปรแกรมให้ตรงตามความถูกต้อง โดยคณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM (Cross Industry Standard Process For Data Mining) จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ด้วย เทคนิค Decision tree ต้นไม้ตัดสินใจ โดยใช้โปรแกรม RapidMiner Studio และโปรแกรม Weka 3.8.5

โดยเว็บไซต์ที่สร้างขึ้น เป็นลักษณะของการเผยแพร่ข้อมูลที่ได้จากการวิเคราะห์ข้อมูลและ การแสดงผลในรูปแบบของแผนภาพโดยใช้โปรแกรม Tableau Public และ Google Charts API ซึ่งเว็บไซต์นี้ใช้ภาษา HTML CSS PHP และ JavaScript และมีส่วนประกอบต่าง ๆ ดังนี้ 1) บทความภัยคุกคามการโจมตีบนเว็บไซต์ 2) กระบวนการวิเคราะห์ข้อมูล 3) ส่วนของการแสดงผลรูปแบบของแผนภาพ 4) ส่วนการแสดงผลการวิเคราะห์ต้นไม้ตัดสินใจ 5) การจัดการข้อมูลภัยคุกคาม 6) ส่วนการเพิ่มข้อมูลภัยคุกคาม 7) ส่วนของการอัปโหลดชุดข้อมูล 8) ส่วนการดาวน์โหลดชุดข้อมูล 9) ส่วนของการแนะนำวิธีการใช้งานเว็บไซต์ 10) ส่วนของผลการวิเคราะห์ 11) ส่วนของการประเมินคุณสมบัติจากการวิเคราะห์ข้อมูล

จากผลการวิเคราะห์ข้อมูล ด้วยการจำแนกแบบ Decision Tree การคำนวณด้วยมือนี้ ข้อมูลที่ได้ผลลัพธ์ว่า 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 จากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อนำเสนอ คือ โปรแกรม Weka ในการสร้างโมเดล Decision Tree และตรวจสอบข้อมูลผ่านทางโปรแกรม RapidMiner Studio เพื่อความแม่นยำ เมื่อนำไปเปรียบเทียบกับผลการคำนวณ โมเดล Decision Tree ด้วยตัวเอง ทางผู้วิเคราะห์ข้อมูลพบว่าทั้ง 3 โมเดลได้ผลลัพธ์ความแม่นยำของโมเดลที่ เหมือนกัน และสามารถนำโมเดลไปใช้งานได้ และประเมินประสิทธิภาพของโมเดลซึ่งมีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 75.55% จำนวนทั้งสิ้น 42,414 รายการ เมื่อเปรียบเทียบกับงานวิจัยการพัฒนาแบบจำลองของ วนิดา พงษ์สงวน , ทิพยา ถินสูงเนิน , มาโนช ถินสูงเนิน (2559) การพัฒนาแบบจำลองปัจจัยที่มีผลต่อการเป็นโรคเบาหวานด้วยเทคนิคต้นไม้ตัดสินใจ จำนวนทั้งสิ้น 44,002 รายการ พบว่าโมเดลแบบจำลองมีค่าความถูกต้องเฉลี่ยที่ใกล้เคียงกันที่ 76.14% ซึ่งเป็นค่าที่สามารถยอมรับได้ในการ ทดสอบประสิทธิภาพ และทุกโมเดลพบว่ามีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี ดังนั้นจึงเหมาะนำมาใช้งานกับการวิเคราะห์ข้อมูลสารสนเทศได้เป็นอย่างดี กระบวนการวิเคราะห์ข้อมูลนี้ ทำให้เกิดประสิทธิภาพในการศึกษาค้นคว้าข้อมูลตามที่ใช้ต้องการ

จากคะแนนในแต่ละด้านที่กลุ่มผู้เข้าใช้งานเว็บไซต์ ได้ประเมินผ่านแบบสอบถามความพึงพอใจในการวิเคราะห์ข้อมูล ได้คะแนนเฉลี่ย รวม 3.93 คะแนน อยู่ในระดับพึงพอใจมาก ซึ่งสามารถที่จะนำผลการวิเคราะห์แนวโน้มของการโจมตีบนเว็บไปใช้งานในด้านของการเฝ้าระวังและป้องกันภัยคุกคาม และคะแนนในแต่ละด้านการเข้าเยี่ยมชมเว็บไซต์ ได้ประเมินผ่านแบบสอบถามความพึงพอใจในการเข้าเยี่ยมชมเว็บไซต์ ได้คะแนนเฉลี่ย รวม 4.01 คะแนน อยู่ในระดับพึงพอใจมาก ซึ่งเหมาะสมกับการนำเว็บไซต์ไปเผยแพร่และสร้างชุมชนให้กับผู้ที่มีความสนใจในเรื่องของภัยคุกคามที่เกิดขึ้นบนเว็บไซต์