Journal of Engineering and Digital Technology (JEDT)
Vol.11 No.1 January - June 2023

การพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศด้วยการเรียนรู้ของเครื่อง

สัพพัญญู ชูแก้ว[1*] ประสงค์ ปราณีตพลกรัง[2] พายัพ ศิรินาม[3]

[1,2,3]หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีป้องกันประเทศ สำนักบัณฑิตศึกษา*
โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช, กรุงเทพมหานคร, ประเทศไทย

*ผู้ประพันธ์บรรณกิจ อีเมล : sappanyou@gmail.com

รับต้นฉบับ : 23 กุมภาพันธ์ 2566; รับบทความฉบับแก้ไข : 12 มีนาคม 2566; ตอบรับบทความ : 28 เมษายน 2566
เผยแพร่ออนไลน์ : 29 มิถุนายน 2566

บทคัดย่อ

ปัจจุบันภัยคุกคามทางไซเบอร์ได้ส่งผลกระทบเป็นวงกว้างต่อหน่วยงานด้านความมั่นคงของประเทศ จึงมีความจำเป็นอย่างยิ่งที่จะต้องมีระบบตรวจหาการบุกรุกทางไซเบอร์ หนึ่งในปัจจัยที่ส่งผลต่อประสิทธิภาพของการตรวจหาการบุกรุก คือ ต้องมีการรวบรวมข้อมูลภัยคุกคามทางไซเบอร์หรือมีชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในหน่วยงานสำหรับใช้ในการฝึกสอน และพัฒนาโมเดลภัยคุกคามและการตรวจหาการบุกรุกทางไซเบอร์ ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อทำการศึกษา รวบรวม และ วิเคราะห์ภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ และพัฒนาโมเดลภัยคุกคามทางไซเบอร์โดยใช้เทคนิคการเรียนรู้ของเครื่อง จากนั้นจึงทำการประเมินโมเดลเพื่อหาค่าความแม่นที่มีต่อการตรวจหาภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ โดยใช้เครื่องมือที่ชื่อ RapidMiner Studio ในการวิเคราะห์ ทั้งนี้ ผู้วิจัยได้ใช้ 5 โมเดล ได้แก่ Naïve Bayes, Decision Tree, Random Forest, Gradient Boosted Trees และ Support Vector Machines ผู้วิจัยใช้ชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศซึ่งประกอบด้วยการโจมตีภายในเครือข่ายของกองทัพอากาศ ซึ่งมีภัยคุกคามหลักจากมัลแวร์ (malware) หรือซอฟต์แวร์ประสงค์ร้าย จำนวน 7 ประเภท รวมการโจมตีทั้งสิ้น 38,641 ข้อมูล โดยข้อมูลดังกล่าวเป็นข้อมูลการจราจรทางคอมพิวเตอร์ (traffic log) ซึ่งใช้เป็นข้อมูลนำเข้าในการฝึกสอนโมเดล เมื่อทำการทดลองและเปรียบเทียบผลการวิเคราะห์โมเดลทั้งหมดแล้วจึงทำการคัดเลือกโมเดล Naïve Bayes และ Random Forest นำมาประกอบกันเพื่อปรับปรุงโมเดลให้มีประสิทธิภาพมากขึ้น ซึ่งพบว่าโมเดลแบบผสมนี้ ให้ค่าความแม่น (accuracy) มากถึง 98.01% ค่าความเที่ยง (Precision) 96.07% ค่าความระลึก (Recall) 98.17% และค่าเฉลี่ย F1 (F1 Score) 97.04%

คำสำคัญ **:** ความแม่น ภัยคุกคามทางไซเบอร์ ระบบตรวจหาการบุกรุก การเรียนรู้ของเครื่อง

Development of Cyber Threat Model in the Royal Thai Air Force Using Machine Learning

Sappanyou Chukaew[1*] Prasong Praneetpolgrang[2] Payap Sirinam[3]

[1,2,3]Master of Engineering Program in Defense Technology, Office of Graduate Studies, Navaminda Kasatriyadhiraj Royal Air Force Academy, Bangkok, Thailand*

*Corresponding Author. E-mail address: sappanyou@gmail.com

Received: 23 February 2023; Revised: 12 March 2023; Accepted: 28 April 2023
Published online: 29 June 2023

Abstract

Current cyber threats have a wide impact on security agencies. Therefore, it is absolutely necessary to have an intrusion detection system. One of the factors that affect the efficiency of an intrusion detection is that the Royal Thai Air Force (RTAF) must have his own cyber threat dataset used in training and develop the model. Therefore, the purposes of this research were to present studying, collecting and analyzing of cyber threats within the RTAF in order to respond to cyber threats and to develop a cyber threats model by using machine learning techniques imported into the process of valuing accuracy of cyber threat detection within the RTAF by using RapidMiner Studio to analyze with five models: Naïve Bayes, Decision Tree, Random Forest, Gradient Boosted Trees and Support Vector Machines. The researchers used the cyber threat data set which consists of attacks within the RTAF network in which the main threats were caused by 7-type malicious softwares, totaling 38,642 attacks, each contains computer traffic data (Traffic Log) used as the training data for the model. The Naïve Bayes and Random Forest models were chosen to increase efficiency. Both models gave the highest accuracy of 98.01% and a detailed assessment of the mixed model (Hybrid) gave the accuracy of 98.01 %, the precision of 96.07%, the recall of 98.17 % and the mean (F1 Score) of 97.04 %.

***Keywords*:** Accuracy, Cyber threat, Intrusion detection systems, Machine learning

1) บทนำ

ในยุคปัจจุบันความก้าวหน้าทางเทคโนโลยีเครือข่ายโดยเฉพาะอินเทอร์เน็ต ทำให้เกิดภัยคุกคามรูปแบบใหม่นั้นคือ ภัยคุกคามทางไซเบอร์ (cyber threat) ที่นับวันจะทวีความรุนแรงขึ้นตามลำดับ ทำให้หน่วยงานที่เป็นโครงสร้างพื้นฐานสำคัญทางสารสนเทศ (Critical Information Infrastructure : CII) รวมถึงกองทัพอากาศ ได้รับผลกระทบจากภัยคุกคามทางไซเบอร์อย่างหลีกเลี่ยงไม่ได้ การบุกรุกระบบเครือข่ายปัจจุบันมีจำนวนเพิ่มมากขึ้น รวมทั้งมีวิธีการหลบเลี่ยงการตรวจหาการบุกรุกจากเครื่องมือต่าง ๆ และความหลากหลายของรูปแบบการโจมตีที่ซับซ้อนขึ้น ซึ่งส่วนมากมาจากการขยายตัวอย่างรวดเร็วในการประยุกต์ใช้เครือข่ายคอมพิวเตอร์และอินเทอร์เน็ตในองค์กรหรือกองทัพอากาศในการทำธุรกรรมต่าง ๆ เช่น ด้านการเงิน ด้านสาธารณูปโภค ด้านการสื่อสาร จึงทำให้ระบบวิเคราะห์และตรวจหา จำเป็นต้องพัฒนาตามไปด้วย จึงมีความจำเป็นอย่างยิ่งที่ต้องมีการบริหารจัดการข้อมูลภัยคุกคามทางไซเบอร์ เพื่อไม่ให้เกิดปัญหาความเสี่ยงที่ส่งผลกระทบร้ายแรงและป้องกันการถูกโจมตีจากภัยคุกคามทางไซเบอร์

กองทัพอากาศจึงได้ให้ความสำคัญ ด้านความมั่นคงปลอดภัยไซเบอร์ (cybersecurity) เป็นอย่างมากโดยสามารถเห็นได้จากแผนยุทธศาสตร์กองทัพอากาศ 20 ปี (พ.ศ.2560 – 2579) ในส่วนของมิติไซเบอร์ (cyber domain) กองทัพอากาศได้เล็งเห็นถึงเทคโนโลยีสารสนเทศและการสื่อสารด้านเครือข่ายและอินเทอร์เน็ตนั้น ได้รับการพัฒนาอย่างรวดเร็ว ภัยคุกคามทางไซเบอร์ก็ได้มีการพัฒนาในรูปแบบที่หลากหลายและรุนแรงมากยิ่งขึ้น จึงได้มีการนำระบบการตรวจหาการบุกรุก (Intrusion Detection System: IDS) [1]–[7] คือ ระบบในการตรวจสอบการใช้งานคอมพิวเตอร์และระบบเครือข่ายที่มีจุดประสงค์ร้าย (malicious information) ที่ส่งผลให้เกิดความเสียหายต่อความถูกต้องของข้อมูล ความสามารถในการให้บริการของระบบ หรือความน่าเชื่อถือขององค์กร ระบบตรวจหาการบุกรุกนั้นแบ่งออกได้เป็นประเภท ดังนี้ 1) ระบบการตรวจสอบการบุกรุกประเภทโฮสต์เบส (Host-based IDS) คือระบบตรวจจับการบุกรุกที่ถูกติดตั้งลงบนระบบคอมพิวเตอร์เพื่อตรวจสอบข้อมูลทั้งเข้าและออกจากคอมพิวเตอร์ ตรวจหาการทำงานของกระบวนการ เพื่อแจ้งเตือนและบันทึกข้อมูลการจราจรทางคอมพิวเตอร์ (traffic log) [8] เมื่อเกิดการบุกรุก หรือ ภัยคุกคาม 2) ระบบตรวจจับการบุกรุกเครือข่าย (Network-based IDS) คือระบบตรวจจับการบุกรุกภายในเครือข่ายว่ามีการใช้งานเครือข่ายที่เป็นการประสงค์ร้ายหรือไม่ โดยการตรวจจับการบุกรุกประเภทนี้จะตรวจข้อมูลการจราจรทางคอมพิวเตอร์ภายในระบบเครือข่ายที่ส่งไปมา เพื่อมาตรวจสอบค้นหารูปแบบที่น่าสงสัย และแจ้งเตือนในการระงับการบุกรุกนั้น

จากการตรวจหาภัยคุกคามทางไซเบอร์ผ่านระบบการตรวจหาการบุกรุกทางไซเบอร์ ทางผู้วิจัยได้นำชุดข้อมูลที่ได้จากการรวบรวมชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศมาจัดทำเป็นชุดข้อมูล (dataset) เพื่อนำมาทำการตรวจสอบหาค่าความแม่นภัยคุกคามทางไซเบอร์ ผ่านเทคโนโลยีการเรียนรู้ของเครื่อง (machine learning) ซึ่งก็คือรูปแบบหนึ่งของการวิเคราะห์ข้อมูลที่ดำเนินการวิเคราะห์ด้วยโมเดลการเรียนรู้ของเครื่องที่ตั้งอยู่บนรากฐานแนวคิดที่ว่า ระบบนั้นสามารถที่จะเรียนรู้และมีปฏิสัมพันธ์กับชุดข้อมูลต่าง ๆ รวมถึงสามารถระบุและทราบรูปภัยคุกคามที่เกิดขึ้นได้ และนำไปสู่การตัดสินใจ

เทคโนโลยีการเรียนรู้ของเครื่อง [2], [4] นั้น สามารถถูกแบ่งออกเป็นการเรียนรู้ได้เป็น 2 แบบ ดังนี้ 1) การเรียนรู้แบบมีผู้สอน (supervised learning) คือการเรียนรู้ โดยมีชุดข้อมูลที่ทำให้คอมพิวเตอร์รู้จักข้อมูลที่นำเข้าไป เมื่อใส่ข้อมูลเข้าไปจะทำการแบ่งชุดข้อมูลออกเป็น 2 ชุด (split test) คือ ชุดข้อมูลสำหรับฝึกสอนหรือข้อมูลชุดเรียนรู้ (training set) และชุดข้อมูลทดสอบ (testing set) เพื่อให้คอมพิวเตอร์แยกแยะตามประเภทข้อมูล หลังจากนั้นเราก็นำข้อมูลมาให้คอมพิวเตอร์ตรวจสอบ แล้วให้ตอบบอกว่าคืออะไร คอมพิวเตอร์ก็ตรวจจากประเภทข้อมูลใส่ไปให้ และสามารถนำเข้ากระบวนการทำนายผล (prediction) ต่อไป 2) การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) จะตรงข้ามกับการเรียนรู้แบบมีผู้สอน คือไม่มีข้อมูลมาฝึกสอน แต่ต้องแยกแยะข้อมูลเองโดยใช้ข้อมูลที่ซ้ำกันมาใส่รวมกันจนได้ผลลัพธ์ออกมา โมเดลที่ใช้กันคือ K Nearest Neighbour และ K Mean

ในงานวิจัยชิ้นนี้ผู้วิจัยได้เลือกใช้ประเภทของการเรียนรู้ของเครื่องแบบการเรียนรู้แบบมีผู้สอน โดยการสร้างชุดข้อมูลฝึกสอนและนำชุดข้อมูล โดยนำชุดข้อมูลทดสอบนำเข้ากระบวนการทำนายผล เพื่อหาค่าความแม่นการตรวจหาภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ โดยใช้งานเครื่องมือที่ชื่อ RapidMiner Studio ใช้โมเดล (model) ในการวิเคราะห์ 5 โมเดล ได้แก่ Naïve Bayes, Decision Tree, Random Forest, Gradient Boosted Trees และ Support Vector Machines เป็นต้น [9],

[10] โดยโปรแกรม Rapid Miner เป็นชุดซอฟต์แวร์สำเร็จรูปที่ศักยภาพสูงในการประยุกต์ใช้งานด้านวิทยาศาสตร์ข้อมูล (data science) และสามารถใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้เชิงลึก การทำเหมืองข้อความ และการวิเคราะห์การทำนาย (predictive analysis) โดยเหตุผลสำคัญที่ผู้วิจัยเลือกใช้เครื่องมือดังกล่าว เพื่อเป็นการเปิดโอกาสให้ผู้ปฏิบัติงานด้านไซเบอร์ของกองทัพอากาศสามารถนำไปใช้งานได้ทันที และลดช่องว่างทางทักษะ (skill gap) ในการเตรียมเครื่องมือเพื่อรองรับภัยคุกคามทางไซเบอร์ที่เกิดขึ้นอย่างรวดเร็วและเกิดขึ้นอย่างตลอดเวลา

2) วัตถุประสงค์ของการวิจัย

1. เพื่อทำการศึกษา รวบรวม และวิเคราะห์ภัยคุกคามทางไซเบอร์ที่มีต่อกองทัพอากาศ

2. เพื่อพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศโดยใช้เทคนิคการเรียนรู้ของเครื่อง

3) วิธีดำเนินการวิจัย

3.1) ภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ

ผู้วิจัยได้ดำเนินการทำการศึกษารวบรวม และวิเคราะห์ประเภทของภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศผ่านการตรวจสอบหลักฐานการโจมตีที่เกิดขึ้นจากชุดข้อมูลการจราจรทางคอมพิวเตอร์ ซึ่งกองทัพอากาศได้ทำการเก็บรวบรวมข้อมูลจากชุดซอฟต์แวร์ Deep Instinct [11] ซึ่งถูกติดตั้งบนเครื่องลูกข่ายของกองทัพอากาศ ซึ่งหน้าที่รวบรวมข้อมูลภัยคุกคาม โดยผลการวิเคราะห์พบว่า จากข้อมูลการจราจรทางคอมพิวเตอร์ที่ใช้ระยะเวลาเก็บรวบรวมข้อมูลในรอบ 1 ปี (พ.ศ.2564) ในขั้นต้นตรวจพบภัยคุกคามทางไซเบอร์ จำนวน 7 ประเภท ดังแสดงในตารางที่ 1 ประกอบด้วย

- การโจมตีประเภท Backdoor
- การโจมตีประเภท Dropper
- การโจมตีประเภท Potentially Unwanted Applications (PUA)
- การโจมตีประเภท Ransomware
- การโจมตีประเภท Spyware
- การโจมตีประเภท Virus
- การโจมตีประเภท Worm

ตารางที่ 1 : แสดงประเภทและจำนวนการโจมตีทางไซเบอร์ต่อกองทัพอากาศซึ่งถูกเก็บรวบรวมข้อมูลผ่านซอฟต์แวร์ Deep Instinct ในช่วงปี พ.ศ.2564

ประเภทการโจมตี	จำนวน	อัตราส่วน
Backdoors	2,083	5.39%
Dropper	11,870	30.72%
PUA	10,227	26.47%
Ransomware	2,297	5.94%
Spyware	3,821	9.89%
Virus	7,461	19.31%
Worm	882	2.28%
รวมทั้งสิ้น	38,641	100.00%

ซึ่งการโจมตีทั้งหมดนั้น ถือเป็นการโจมตีที่อยู่ในประเภทมัลแวร์ (malware) หรือซอฟต์แวร์ประสงค์ร้าย (malicious software) ซึ่งอาจก่อให้เกิดความเสียหายตั้งแต่ระดับเล็กน้อย เช่น ก่อให้เกิดความรำคาญต่อผู้ใช้ ไปจนถึงเกิดความเสียหายระดับร้ายแรง กล่าวคือ อาจทำให้เป็นช่องทางในการเข้ามาจารกรรมข้อมูลที่มีความสำคัญต่อประเทศชาติ รวมถึงเป็นช่องทางในการโจมตีเพื่อให้เกิดความเสียหายต่อระบบงานด้านยุทธการของกองทัพ

จากผลการศึกษาดังกล่าว ทำให้ทราบถึงประเภทภัยคุกคามทางไซเบอร์ที่มีต่อกองทัพอากาศ และเป็นข้อมูลพื้นฐานสำคัญในการพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศโดยใช้เทคนิคการเรียนรู้ของเครื่องต่อไป

3.2) พัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศโดยใช้เทคนิคการเรียนรู้ของเครื่อง

หลังจากที่ผู้วิจัยเริ่มต้นจากการศึกษารวบรวม และวิเคราะห์ประเภทของภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศเรียบร้อยแล้ว ขั้นตอนต่อไปจะเป็นการนำข้อมูลที่ไปเพื่อพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศ โดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยผู้วิจัยได้ทำการทบทวนวรรณกรรมเพื่อศึกษางานวิจัยที่เกี่ยวข้องกับการประยุกต์การเรียนรู้ของเครื่อง ในการพัฒนาระบบตรวจหาการบุกรุก [2]–[7] รวมถึงการวิเคราะห์อัลกอริทึมประเภทต่าง ๆ ว่ามีข้อดี-ข้อเสียอย่างไร [10], [12], [13] และปัจจัยในการเพิ่มประสิทธิภาพของโมเดลอย่างไร เพื่อนำข้อมูลมาสรุปวิเคราะห์ และออกแบบพัฒนาโมเดลประเภท

ปรับตัวด้านภัยคุกคามทางไซเบอร์ได้อย่างถูกต้องโดยประกอบด้วยขั้นตอนต่าง ๆ [3], [14]–[16] แสดงดังรูปที่ 1

รูปที่ 1 : ขั้นตอนการดำเนินงานวิจัยและการออกแบบการทดลอง 5 ขั้นตอน [7]

จากรูปที่ 1 สามารถอธิบายขั้นตอนการดำเนินการวิจัย และการออกแบบการทดลองได้ดังต่อไปนี้

3.2.1) ชุดข้อมูล (Dataset) ในการพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศโดยใช้เทคนิคการเรียนรู้ของเครื่อง ผู้วิจัยได้ดำเนินการนำผลการศึกษาภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ เพื่อนำมาพัฒนาโมเดลที่มีความเหมาะสมโดยเริ่มต้นในการเตรียมชุดข้อมูลเพื่อฝึกสอนโมเดล ซึ่งในงานวิจัยนี้ได้จัดเตรียมชุดข้อมูลดังรูปที่ 2 โดยมีขั้นตอนการรวบรวมดังนี้ 1) ดักจับข้อมูลการจราจรทางคอมพิวเตอร์ [8] จากอุปกรณ์ภายในกองทัพอากาศ 2) นำข้อมูลการจราจรทางคอมพิวเตอร์ [8] เข้าวิเคราะห์ระบบตรวจหาผู้บุกรุก เพื่อเป็นข้อมูลตั้งต้นในการเตรียมชุดข้อมูล 3) ทำการจัดทำข้อมูลให้เป็นมาตรฐาน (normalization) 4) จำแนกข้อมูลภัยคุกคาม (classification) จากขั้นตอนที่กล่าวมาข้างต้นจึงได้ชุดข้อมูลเพื่อนำมาทำการตรวจสอบหาค่าความแม่นภัยคุกคามทางไซเบอร์ [1], [12]

รูปที่ 2 : ภาพรวมจัดเตรียมชุดข้อมูล

3.2.2) การเตรียมข้อมูล (Data Preparation) เมื่อจัดเตรียมชุดข้อมูลที่ได้จากการรวบรวมชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ จำนวน 38,642 ข้อมูล แสดงผลฟีเจอร์ที่นำมาคำนวณจากชุดข้อมูล [17] ดังตารางที่ 2 โดยใช้งานโปรแกรม RapidMiner Studio [18], [19] สำหรับขั้นตอนการตรวจสอบและประเมินประสิทธิภาพ โดยเริ่มจากการนำข้อมูลชุดฝึกฝนเข้าสู่โปรแกรมโดยกำหนดคุณลักษณะที่ต้องการทำนายเป็นฉลากหรือป้าย (label) ในที่นี้ได้ระบุเป็นในส่วนประเภทของภัยคุกคามทางไซเบอร์ที่อยู่ในฟีเจอร์ที่ชื่อ Deep Classification

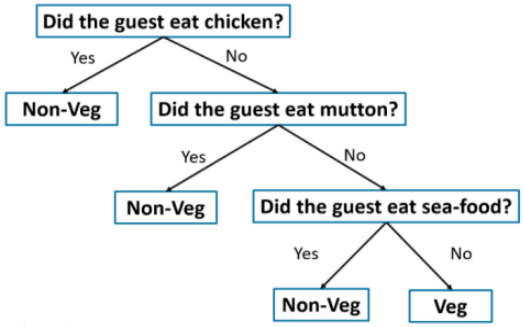
ตารางที่ 2 : แสดงผลฟีเจอร์สำหรับชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ

Feature	Usage
Status	

ตารางที่ 2 : แสดงผลฟีเจอร์สำหรับชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ (9jv)

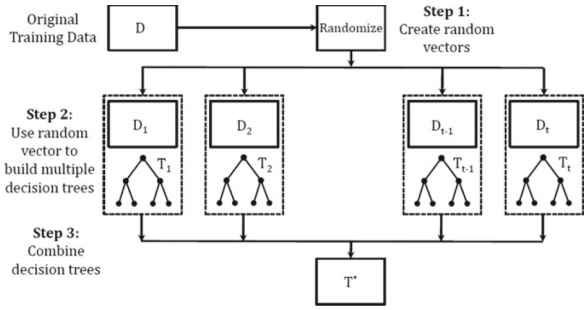
Feature	Usage
MAC Address	

จำแนกประเภท Decision Tree มักใช้เทคนิคการแบ่งกิ่งหลังที่ประเมินประสิทธิภาพของแผนผังการตัดสินใจ เนื่องจากถูกแบ่งโดยใช้ชุดการตรวจสอบความถูกต้อง แสดงดังรูปที่ 4 [5], [12]



รูปที่ 4 : การทำงานของ Decision Tree [12]

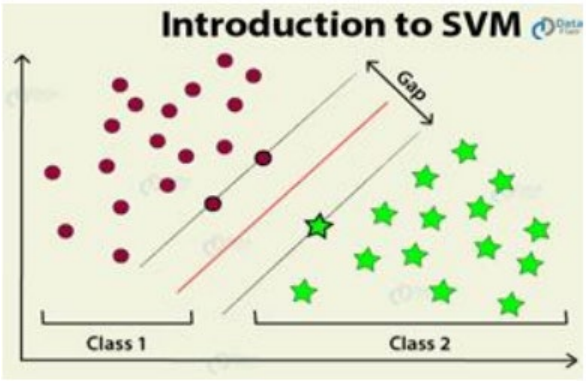
3.2.4.3) Random Forest [21] คือโมเดลที่มีอัลกอริทึมที่นำ Decision Tree หลาย ๆ Tree มา Train ร่วมกัน (ตั้งแต่ 10 ต้น ถึง มากกว่า 1000 ต้น) โดยที่แต่ละ Tree จะได้รับ Feature และ Data เป็น Subset ของ Feature และ Data ทั้งหมดแบบ Random ตอนทำ Prediction ก็ให้แต่ละ Decision Tree ทำ Prediction ของแต่ละอัน และเลือกผล Final Prediction จากค่า Prediction ที่ได้รับการโหวตมากที่สุดแสดงดังรูปที่ 5



รูปที่ 5: การทำงานของ Random Forest [21]

3.2.4.4) Gradient Boosted Trees [14] คือ โมเดลที่มีอัลกอริทึมที่มีพื้นฐานมาจาก Decision Tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของโมเดลให้มีค่าสูงขึ้น โดยการสุ่มสร้าง Decision Tree หลายร้อยโมเดล และประเมินผลแต่ละโมเดลจนกว่าจะได้ *Decision Tree* ที่สมบูรณ์

3.2.4.5) Support Vector Machines [4], [9] คือโมเดลที่มีอัลกอริทึมใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูลโดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการฝึกฝนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูล แสดงดังรูปที่ 6



รูปที่ 6 : การทำงานของ Support Vector Machine [12]

3.2.5) การประเมินตัวโมเดล (Model Evaluation) เมื่อผ่านขั้นตอนการสร้างโมเดลเรียบร้อยแล้วจะนำผลการทำนายเข้ากระบวนการประเมินประสิทธิภาพของตัวแบบโดยพิจารณาจากค่าความแม่น [13] ดังสมการที่ 2 ค่าความเที่ยง ดังสมการที่ 3 ค่าความระลึก ดังสมการที่ 4 และค่าเฉลี่ย F1 ของ ค่าความเที่ยง และ ค่าความระลึก ดังสมการที่ 5 [15]

$$Accuracy = \frac{TN+TP}{TP + FP + TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

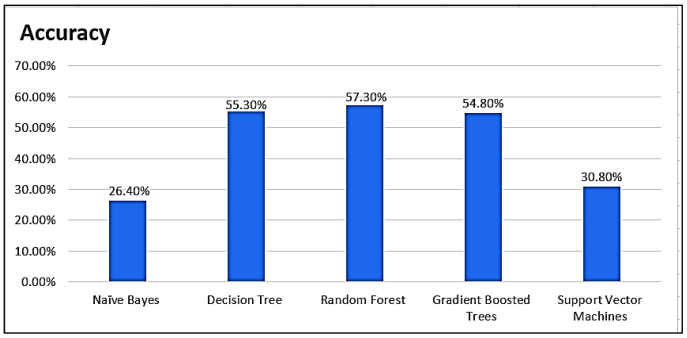
โดยที่ TP คือ ข้อมูลที่ทำนายถูกต้องเมื่อเทียบกับเฉลย
FP คือ ข้อมูลที่ทำนายแล้วไม่ถูกต้องเมื่อเทียบกับเฉลย
FN คือ ข้อมูลที่อยู่ในเฉลยแต่ไม่มีการทำนาย (ตรงข้ามกับ FN)
Precision คือ ค่าความเที่ยงเกิดจากการนำค่า TP มาเทียบกับ FP
Recall คือค่าความระลึกเกิดจากการนำค่า TP มาเทียบกับ FN
F1 Score คือ ค่าเฉลี่ยของ ค่าความเที่ยง และค่าความ

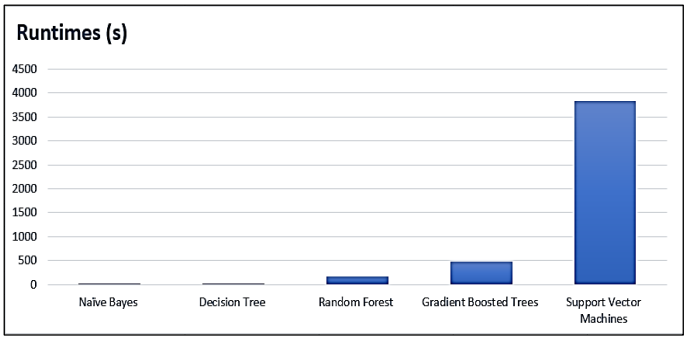
ระลึก

4) ผลการวิจัย

จากวัตถุประสงค์ข้อที่ 1 ที่ได้ทำการศึกษา รวบรวม และวิเคราะห์ภัยคุกคามทางไซเบอร์ที่มีต่อกองทัพอากาศนั้น ผู้วิจัยได้รวบรมชุดข้อมูลภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ จำนวน 38,642 ข้อมูล เพื่อนำมาวิเคราะห์ประสิทธิภาพ โดยใช้ทั้ง 5 โมเดล ได้แก่ Naïve Bayes โดยมีค่าความแม่น 26 % โดยใช้เวลารวมในการทดสอบ 11 วินาที, Decision Tree โดยมีค่าความแม่น 55.3 % โดยใช้เวลารวมในการทดสอบ 16 วินาที, Random Forest โดยมีค่าความแม่น 57.3 % โดยใช้เวลารวมในการทดสอบ 2 นาที 54 วินาที, Gradient Boosted Trees Forest โดยมีค่าความแม่น 54.8 % โดยใช้เวลารวมในการทดสอบ 7 นาที 52 วินาที และ Support Vector Machines โดยมีค่าความแม่น 30.8 % โดยใช้เวลารวมในการทดสอบ 1 ชั่วโมง 4 นาที แสดงดังรูปที่ 7 แสดงให้เห็นว่าโมเดล Naïve Bayes มีค่าที่ดีที่สุด คือ Fastest Scoring Time และ Fastest Total Time ในส่วนโมเดล Random Forest มีค่าที่ดีที่สุดคือ Best Performance และ Best Gain

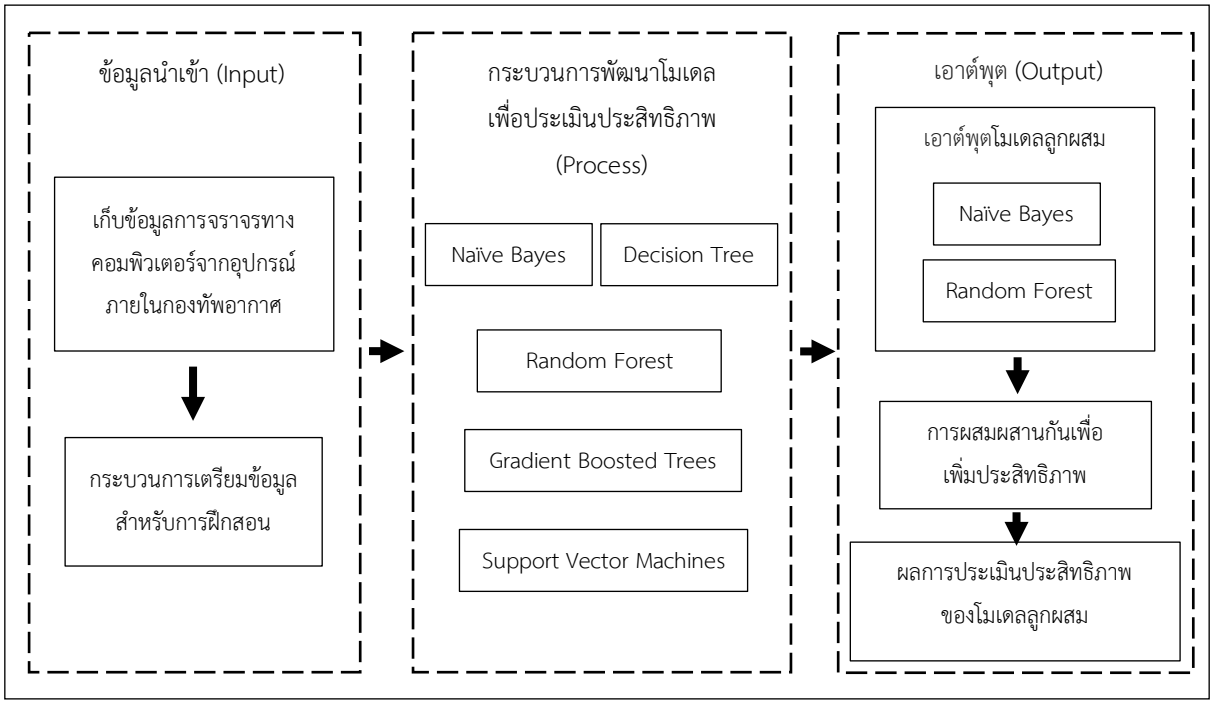
Overview





Model	Accuracy	Standard deviation	Gains	Total Time (s)
Naïve Bayes	26.40%	0.6	944	11
Decision Tree	55.30%	0.2	5460	16
Random Forest	57.30%	0.5	5572	174
Gradient Boosted Trees	54.80%	0.6	5365	477
Support Vector Machines	30.80%	0.6	0	3840

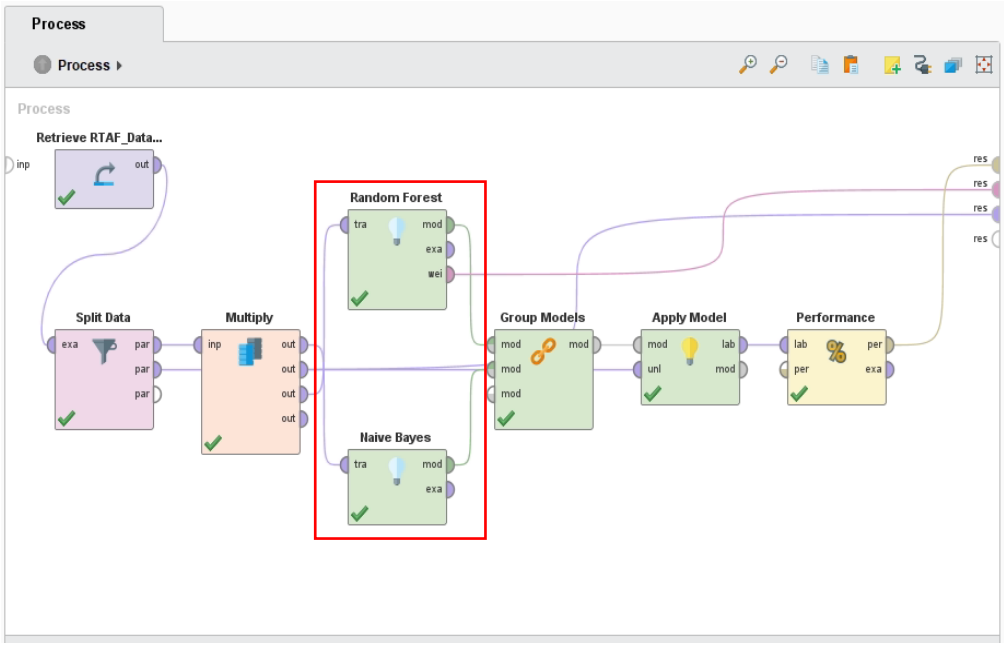
รูปที่ 7 : ผลการประเมินภาพรวมประสิทธิภาพของทั้ง 5 โมเดล



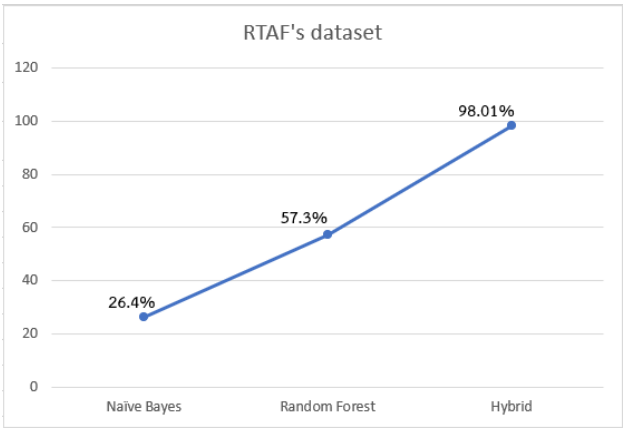
รูปที่ 8 : กระบวนการพัฒนาโมเดลลูกผสม

นอกจากนั้น จากวัตถุประสงค์ข้อที่ 2 เพื่อพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศโดยใช้เทคนิคการเรียนรู้ของเครื่องนั้น ผู้วิจัยได้พัฒนาโมเดลลูกผสม (hybrid) [16] เพื่อเพิ่มประสิทธิภาพในการหาค่าความแม่นการตรวจหาภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศ รูปที่ 8 แสดงถึงกระบวนการพัฒนาโมเดลลูกผสม ซึ่งการสรุปแนวทางในการพัฒนาโมเดลซึ่งประกอบด้วย 3 ขั้นตอนประกอบด้วย ขั้นตอนที่ 1 การจัดการข้อมูลนำเข้าซึ่งเกิดจากการวิเคราะห์ และหาภัยคุกคามที่ส่งผลต่อกองทัพอากาศและเก็บรวบรวมข้อมูลการจราจรทางคอมพิวเตอร์จากอุปกรณ์ภายในกองทัพอากาศ และทำการเตรียมข้อมูล เพื่อให้พร้อมสำหรับกระบวนการฝึกสอนโมเดล ขั้นตอนที่ 2 ขั้นการพัฒนาโมเดลเพื่อประเมินประสิทธิภาพ โดยขั้นตอนนี้เป็นการทดสอบในการฝึกสอนโมเดลจำนวน 5 โมเดล ในขั้นต้นจากนั้นทดสอบรูปแบบการใช้ประโยชน์จากโมเดลต่าง ๆ ตั้งแต่การใช้โมเดลเดี่ยว (1 โมเดลต่อการประเมิน) และการใช้โมเดลแบบลูกผสม (การผสานโมเดล 2 โมเดล) เพื่อทำการวิเคราะห์หารูปแบบที่มีความเหมาะสมมากที่สุด โดยผลการทดสอบพบว่าจากการทดสอบ ทั้งในทุก ๆ ความเป็นไปได้ พบว่ารูปแบบของโมเดลที่มีความเหมาะสม คือ การใช้โมเดลลูกผสมจำนวน 2 โมเดล คือ Naïve Bayes และ Random Forest ซึ่งให้ประสิทธิภาพของการตรวจจับการบุกรุกสูงสุดจากผลการทดลอง ซึ่งรูปแบบโมเดลแบบผสมดังกล่าว เป็นเอาต์พุตที่สำคัญที่กองทัพอากาศสามารถนำไปใช้เป็นต้นแบบของการสร้างโมเดล รวมถึงการเพิ่มประสิทธิภาพของโมเดลได้ในอนาคต โดยการพัฒนาต่อยอดสามารถทำการประยุกต์ใช้ซอฟต์แวร์ RapidMiner Studio ที่มีความสะดวก รวดเร็ว ต่อการพัฒนาโมเดล รวมถึงลดช่องว่างด้านทักษะของผู้พัฒนาที่ไม่จำเป็นต้องเชี่ยวชาญด้านโปรแกรมคอมพิวเตอร์มากนัก ดังแสดงรูปที่ 9

ทั้งนี้ เมื่อนำมาผสมผสานกันจนเป็นลูกผสมของทั้งสองโมเดลแล้วจะพบว่าสามารถเพิ่มประสิทธิภาพความถูกต้องได้สูงสุดอยู่ที่ 98.01% ดังแสดงในรูปที่ 10 และมีรายละเอียดของการประเมินโมเดลลูกผสมดังแสดงโดยละเอียดในตารางที่ 3 ที่แสดงผลโดยละเอียดผ่าน Confusion Matrix และเมื่อสรุปผลในภาพรวมพบว่าโมเดลดังกล่าวสามารถให้ค่าความแม่นมากถึง 98.01% ค่าความเที่ยง 96.07% ค่าความระลึก 98.17% และ ค่าเฉลี่ย F1 97.04% ตามลำดับ สรุปไว้ดังแสดงไว้ในดังตารางที่ 4



รูปที่ 9: โมเดลลูกผสมของ Naïve Bayes และ Random Forest ในซอฟต์แวร์ RapidMiner Studio



รูปที่ 10 : กราฟแสดงประสิทธิภาพของโมเดลลูกผสม

ตารางที่ 3 : ผลการแสดงการประเมินผลลัพธ์การทำนายผ่าน Confusion Matrix

Accuracy : 98.01%								
	True Worm	True Backdoor	True PUA	True Virus	True Dropper	True Spyware	True Ransomware	Class Precision
Pred. Worm	614	15	3	3	12	2	1	94.32%
Pred. Backdoor	1	1395	24	8	51	7	8	93.37%
Pred. PUA	1	2	7083	7	32	7	6	99.23%
Pred. Virus	0	0	28	5184	61	8	5	98.07%
Pred. Dropper	0	0	0	0	8002	0	0	100.00%
Pred. Spyware	0	3	10	18	42	2647	3	97.21%
Pred. Ransomware	1	43	11	3	108	4	1585	90.31%
Class recall	99.51%	95.68%	98.94%	99.25%	96.31%	98.95%	98.57%	

ตารางที่ 4 : ผลการประเมินประสิทธิภาพของโมเดลลูกผสม

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Hybrid	98.01	96.07	98.17	97.04

5) อภิปรายผลการวิจัย

จากผลการวิจัยในการทดสอบเพื่อหาโมเดลที่มีความเหมาะสมในการรับมือต่อภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศพบว่าการผสมผสานโมเดล 2 ประเภทได้แก่ Naïve Bayes และ Random Forest สามารถเพิ่มประสิทธิภาพของการตรวจหาภัยคุกคามได้อย่างมีนัยสำคัญ เมื่อเทียบกับการใช้งานโมเดลเพียงชนิดเดียว มากไปกว่านั้น การเลือกใช้ Naïve Bayes และ Random Forest มีความโดดเด่นในแง่ของสถาปัตยกรรมโมเดลที่ไม่ซับซ้อนมากเกินไป อันส่งผลให้ใช้ระยะเวลาน้อยในการฝึกฝนหรือฝึกสอนโมเดล และใช้ทรัพยากรในการประมวลผลที่น้อย ซึ่งเป็นประโยชน์อย่างยิ่งต่อการประยุกต์ใช้ในหน่วยงานทุกระดับ โดยเฉพาะในกองทัพอากาศ รวมถึงสามารถลดข้อจำกัดด้านเวลาในการฝึกฝนข้อมูลการโจมตีในรูปแบบใหม่เพื่อทำให้โมเดลมีความพร้อมสำหรับรับมือกับการถูกโจมตีจากภัยคุกคามทางไซเบอร์ที่เกิดขึ้นใหม่อยู่ตลอดเวลา

6) สรุปผลการวิจัย

งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อ 1) ทำการศึกษา รวบรวม และวิเคราะห์ภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศสำหรับรับมือกับภัยคุกคามทางไซเบอร์ จากชุดข้อมูลที่ผ่านการศึกษา รวบรวม และวิเคราะห์ จากนั้นได้ทำจัดเตรียมชุดข้อมูล สำหรับการตรวจสอบและประเมินประสิทธิภาพ โดยเริ่มจากการกำหนดคุณลักษณะที่ต้องการทำนายเป็นฉลาก หรือป้าย ในที่นี้ระบุเป็นในส่วนของชนิดของภัยคุกคามทางไซเบอร์ที่อยู่ในฟีเจอร์ชื่อ Deep Classification จากนั้นนำชุดข้อมูลทำการหาค่าความแม่น โดยใช้โมเดลในการวิเคราะห์ 5 โมเดล ได้แก่ Naïve Bayes, Decision Tree, Random Forest, Gradient Boosted Trees และ Support Vector Machines 2) พัฒนาโมเดลประเภทปรับตัวด้านภัยคุกคามทางไซเบอร์ที่สามารถส่งเสริมการพึ่งพาตนเองและรักษาอธิปไตยไซเบอร์ภายในกองทัพอากาศ โดยคัดเลือกโมเดล Naïve Bayes ที่มีค่าที่ดีที่สุดคือ Fastest Scoring Time และ Fastest Total Time และ โมเดล Random Forest มีค่าที่ดีที่สุดคือ Best Performance และ Best Gain นำมาผสมเพื่อเพิ่มประสิทธิภาพโดยเมื่อผสมโมเดลทั้งสองชนิดแล้วให้ค่าความแม่นสูงที่สุดที่ 98.01 % ความเที่ยง 96.07 % ค่าความระลึก 98.17 % และ ค่าเฉลี่ย F1 97.04 %

7) ข้อจำกัดของงานวิจัย

1) งานวิจัยนี้ มุ่งเน้นการศึกษาเฉพาะการโจมตีประเภทหลักที่มีต่อเครือข่ายภายในของกองทัพอากาศ ดังนั้น โมเดลจึงถูกสร้างให้มีความเหมาะสมกับประเภทภัยคุกคามเฉพาะด้านเท่านั้น

2) งานวิจัยนี้มุ่งเน้นการพัฒนาโมเดลแบบผสมผสานที่เป็นลูกผสม ในการนำไปใช้งานจึงมีความจำเป็นที่ต้องรู้ประเภทของภัยคุกคามล่วงหน้าสำหรับเตรียมฝึกฝนโมเดลให้มีความเหมาะสม ดังนั้น หากมีการโจมตีผ่านช่องโหว่ที่ยังไม่เคยถูกตรวจพบมาก่อน (zero-day attacks) อาจต้องทำการพัฒนาโมเดลให้เป็นแบบปรับตัวได้

8) ข้อเสนอแนะ

จากที่ได้ทำการวิจัยการพัฒนาโมเดลภัยคุกคามทางไซเบอร์ในกองทัพอากาศด้วยการเรียนรู้ของเครื่อง ผู้วิจัยมีข้อเสนอแนะดังนี้

8.1) ข้อเสนอแนะทั่วไป

1. เนื่องจากข้อมูลภัยคุกคามทางไซเบอร์มีการพัฒนารูปแบบวิธีการโจมตีที่รวดเร็ว การศึกษา รวบรวม และวิเคราะห์ภัยคุกคามทางไซเบอร์ภายในกองทัพอากาศนั้น ควรมีการทำให้ข้อมูลเป็นปัจจุบันอยู่เสมอ

2. ในการพัฒนาโมเดล สามารถนำโมเดลที่อัปเดตของเครื่องมือที่ใช้มาทำการทดสอบเพื่อเพิ่มประสิทธิภาพให้กับการหาค่าความแม่นได้

3. ให้ความสำคัญในการเก็บชุดข้อมูลจริงเพราะมีส่วนสำคัญในการวิเคราะห์ สำหรับการตรวจสอบและประเมินประสิทธิภาพค่าความแม่นของงานวิจัย

8.2) ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

1. จากการทดลองจะเห็นได้ว่าการโจมตีมีรูปแบบที่เฉพาะและหลากหลายเพิ่มมากยิ่งขึ้นในปัจจุบัน อีกทั้งยังมีการโจมตีทางไซเบอร์ที่ใช้ปัญญาประดิษฐ์อีกด้วย ดังนั้น การตรวจหาภัยคุกคามทางไซเบอร์ภายใต้การวิเคราะห์ข้อมูลขนาดใหญ่ (big data) จึงเป็นสิ่งจำเป็นที่ต้องดำเนินการ

2. กระบวนการในการพัฒนาความมั่นคงปลอดภัยทางไซเบอร์ที่มีประสิทธิภาพนั้น นอกเหนือจากมีเครื่องมือที่มีประสิทธิภาพสูงแล้ว สิ่งที่สำคัญคือการศึกษาวิจัยและพัฒนาทักษะความตระหนักรู้ทางไซเบอร์ให้แก่บุคลากร เพื่อเป็นการป้องกันตั้งแต่ต้นทางก็จะช่วยลดความเสี่ยงต่อองค์กรได้ ดังนั้น การศึกษาและวิจัยในการสร้างความตระหนักรู้ด้านไซเบอร์จึงต้องทำควบคู่กันไป

3. การวิจัยแบบผสมผสานในเชิงจิตวิทยาไซเบอร์ (cyber psychology) รวมทั้งกรรมวิธีการเปลี่ยนทัศนคติและสร้างวัฒนธรรมทางไซเบอร์ที่หยั่งรากลึกในระดับบุคคลหน่วยงานให้เกิดภูมิคุ้มกัน และมีจริยธรรมทางไซเบอร์ ทั้งนี้เพื่อสร้างความพร้อมในการรับมือกับภัยคุกคามทางไซเบอร์ อีกทั้งยังทำให้บุคคลและหน่วยงานมีขีดความสามารถด้านไซเบอร์ ในอันที่จะป้องกันต่อต้าน ตรวจจับ และตอบสนองต่อการบุกรุก การจารกรรม หรือการหลอกลวง ที่จะทำให้เกิดความเสียหายได้ ผู้วิจัยจึงมองว่าควรทำวิจัยให้เป็นแบบสหวิทยาการต่อไปในอนาคต

8.3) ข้อเสนอเกี่ยวกับความสมดุลของข้อมูล

จากตารางที่ 1 จะเห็นว่า ชุดข้อมูลดังกล่าวเมื่อพิจารณาตามประเภทของการโจมตี (class of attack) จำนวนข้อมูลที่ใช้ในการฝึกฝนมีความแตกต่าง และไม่สมดุลกัน (imbalance) เนื่องจากเป็นข้อมูลที่มาจากการเก็บข้อมูลการโจมตีจริง ไม่ได้เกิดจากการจำลองขึ้นมา ดังนั้นจำนวนครั้งของการโจมตี จึงบ่งบอกถึงความถี่ของภัยคุกคามนั้น ผู้วิจัยตระหนักดีถึงปัญหาที่อาจเกิดจากข้อมูลที่ไม่สมดุลอาจกระทบโดยตรงต่อประสิทธิภาพของโมเดลและการแปลความในแง่ประสิทธิภาพของโมเดล แต่ในกรณีนี้จะเห็นว่า ข้อมูลของคลาสที่มีจำนวนน้อยที่สุดมีจำนวนถึง 882 ข้อมูล ซึ่งมากเพียงพอต่อการฝึกสอนโมเดลให้สามารถจำแนกประเภทการโจมตีดังกล่าว ดังจะเห็นได้จากค่าความแม่นกว่า 90% สำหรับการตีความเรื่องประสิทธิภาพ ผู้วิจัยตระหนักดีว่า นอกเหนือจากการตีความในภาพรวมของโมเดลที่เป็นค่าความแม่นเฉลี่ย ที่อาจได้รับผลเชิงบวกจากประเภทการโจมตีอื่น ๆ ที่มีจำนวนข้อมูลฝึกฝนที่มากกว่า ดังนั้น ผู้วิจัยจึงนำเสนอข้อมูลผลการแสดงการประเมินผลลัพธ์การทำนายผ่าน Confusion Matrix เพื่อเป็นข้อมูลอย่างละเอียดให้แก่ผู้ที่นำโมเดลดังกล่าวไปใช้งานต่อ

REFERENCES

[1] D. Kapil, N. Mehra, A. Gupta, S. Maurya, and A. Sharma, “Network security: Threat model, Attacks, and IDS using machine learning,” in *Int. Conf. Artif. Intell. and Smart Syst. (ICAIS)*, Coimbatore, India, 2021, pp. 203–208.

[2] U. S. Musa, S. Chakraborty, M. M. Abdullahi, and T. Maini, “A review on intrusion detection system using machine learning techniques,” in *Int. Conf. Comput., Commun., and Intell. Syst. (ICCCIS)*, Greater Noida, India, 2021, pp. 541–549.

[3] F. Hossain, M. Akter, and M. N. Uddin, “Cyber Attack Detection Model (CADM) based on machine learning approach,” in *2nd Int. Conf. Robotics, Elect. and Signal Process. Techn. (ICREST)*, Dhaka, Bangladesh, 2021, pp. 567–572.

[4] A. Halimaa and K. Sundarakantham, “Machine learning based intrusion detection system,” in *3rd Int. Conf. Trends in Electron. and Inform. (ICOEI)*, Tirunelveli, India, 2019, pp. 916–920.

[5] S. Biswas, “Intrusion detection using machine learning: A comparison study,” *Int. J. Pure Appl. Math.*, vol. 118, no. 19, pp. 101–114, Feb. 2018.

[6] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, “Supervised machine learning algorithms: Classification and comparison,” *Int. J. Comput. Trends Technol. (IJCTT)*, vol. 48, no. 3, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P126.

[7] A. Handa, A. Sharma, and S. K. Shukla, “Machine learning in cybersecurity: A review,” *WIREs Data Mining and Knowl. Discovery*, vol. 9, no. 4, 2019, doi: 10.1002/widm.1306.

[8] C. G. Cordero, E. Vasilomanolakis, A. Wainakh, M. Mühlhäuser, and S. N. Tehrani, “On generating network traffic datasets with synthetic attacks for intrusion detection,” *ACM Trans. Privacy Secur. (TOPS)*, vol. 24, no. 2, pp. 1–39, Dec. 2020.

[9] R. -F. Hong, S. -C. Horng, and S. -S. Lin, “Machine learning in cyber security analytics using NSL-KDD Dataset,” in *Int. Conf. Technol. and Appl. Artif. Intell. (TAAI)*, Taichung, Taiwan, 2021, pp. 260–265.

[10] A. O. David and U. J. Joseph, “A novel immune inspired concept with neural network for intrusion detection in cybersecurity,” *Int. J. Appl. Inf. Syst. (IJAIS)*, vol. 12, no. 30, pp. 13–17, Jun. 2020.

[11] A. Mathew, “Cybersecurity infrastructure and security automation,” *Adv. Comput.: An Int. J. (ACIJ)*, vol. 10, no. 6, pp. 1–7, 2019.

[12] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. and Res. (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020.

[13] S. Bagui, E. Kalaimannan, S. Bagui, D. Nandi, and A. Pinto, "Using machine learning techniques to identify rare cyber-attacks on the UNSW-NB15 dataset," *Secur. Privacy*, vol. 2, no. 6, 2019, doi: 10.1002/spy2.91.

[14] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 1, pp. 1104–1116, Mar. 2021, doi: 10.1109/TNSM.2020.3032618.

[15] N. F. Rusland, N. Wahid, S. Kasim, and H. Hafit, "Analysis of naïve bayes algorithm for email spam filtering across multiple datasets," in *Proc. Int. Res. Innov. Summit (IRIS2017)*, Melaka, Malaysia, May 2017, doi: 10.1088/1757-899X/226/1/012091.

[16] Y. Hamid, M. Sugumaran, and V. Balasaraswathi, "IDS using machine learning - current state of art and future directions," *British J. Appl. Sci. & Technol.*, vol. 15, no. 3, pp. 1–22, 2016, doi: 10.9734/bjast/2016/23668.

[17] I. Sharafaldin, A. H. Lashkar, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, 2018, pp. 108–116, doi: 10.5220/0006639801080116.

[18] M. H. Hameed, U. Rasheed, and A. Rehan, "Spam profile detection on Facebook and performance evaluation of machine learning Algorithms," *Int. J. Innov. Sci. Res. Technol.*, vol. 7, no. 8, pp. 870–875, Aug. 2022.

[19] T. Vincent and U. Prince, "Implementation of critical information infrastructure protection techniques against cyber attacks using big data analytics," Jun. 2021. [Online]. Available: https://www.researchgate.net/publication/352682216_IMPLEMENTATION_OF_CRITICAL_INFORMATION_INFRASTRUCTURE_PROTECTION_TECHNIQUES_AGAINST_CYBER_ATTACK_USING_BIG_DATA_ANALYTICS

[20] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: An overview from machine learning perspective," *J. Big Data,* vol. 7, 2020, Art. no. 41.

[21] A. B. Shaik and S. Srinivasan, "A brief survey on random forest ensembles in classification model," in *Proc. Int. Conf. Innov. Comput. and Commun.*, New Delhi, India, May 2018, pp. 253–260.

[22] M. Abdullahi, "Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review," *Electron.*, vol. 11, no. 2, p. 198, 2022, doi: 10.3390/electronics11020198.

[23] N. Baharun, N. F. M. Razi, S. Masrom, N. A. M. Yusri, and A. S. A. Rahman, "Auto modelling for machine learning: A comparison implementation between RapidMiner and Python," *Int. J. Emerging Technol. Adv. Eng.*, vol. 12, no. 5, pp. 15–27, 2022, doi: 10.46338/ijetae0522_03.